Working Paper Series

Payment Choice and Currency Use: Insights from Two Billion Retail Transactions

WP 14-09R

Zhu Wang
Federal Reserve Bank of Richmond

Alexander L. Wolman
Federal Reserve Bank of Richmond

FEDERAL RESERVE BANK OF RICHMOND®
Richmond ▪ Baltimore ▪ Charlotte

Payment Choice and Currency Use: Insights from Two Billion Retail Transactions*

Zhu Wang, Alexander L. Wolman†

September 2016
Federal Reserve Bank of Richmond Working Paper No. 14-09R

Abstract

Using three years of transactions data from a discount retailer with thousands of stores, we study payment variation along three dimensions: transaction size and location; weekly and monthly frequencies; and longer time horizons. In each case, we connect empirical patterns to theories of money demand and payments. We show that cross-sectional and time-series payment patterns are consistent with a theoretical framework in which individual consumers choose between cash and non-cash payments based on a threshold transaction size, and we evaluate factors that may account for the variation in threshold distributions across locations and time.

Keywords: Payment choice; Money demand; Consumer behavior
JEL Classification: E41; D12; G2

*Forthcoming, *Journal of Monetary Economics*. The data used in this study is proprietary and has been provided to us by the Payment Studies Group at the Federal Reserve Bank of Richmond. We are grateful to the members of that group for their efforts in putting together this data. In addition, we would like to thank members of FRB Richmond's IT department for their critical role in putting the data in manageable form. Sam Marshall and Joe Johnson have provided outstanding research assistance, and Liz Marshall and Sabrina Pellerin provided us with valuable input. We would like to thank the editor and two anonymous referees for valuable comments and suggestions. For helpful comments, we also thank Fernando Alvarez, Dave Beck, Evren Damar, Itay Goldstein, Tom Holmes, Boyan Jovanovic, Bob Lucas, Marc Rysman, Joanna Stavins, Matti Viren, Mark Watson, and participants at several conferences and at seminars at New York University, Kyoto University, and the Federal Reserve Banks of Philadelphia and Richmond.

†Research Department, Federal Reserve Bank of Richmond; zhu.wang@rich.frb.org and alexander.wolman@rich.frb.org. The views expressed here are those of the authors and do not represent the views of the Federal Reserve Bank of Richmond, the Board of Governors of the Federal Reserve, or the Federal Reserve System.

1 Introduction

The U.S. payments system has been undergoing fundamental changes in the past several decades, migrating from paper instruments, namely cash and check, to electronic forms such as debit and credit cards. Amidst these changes, a sizeable empirical literature has developed to study consumer payment choice, with the broader goals of understanding payments system functioning and the transaction demand for currency. For researchers and policymakers working on these issues, one major impediment is the difficulty of tracking consumers' use of cash. To gauge cash use, most studies rely on data from consumer surveys, and while this research has improved our understanding of how consumers choose to pay, consumer survey data has its limitations: Most surveys have relatively small samples (typically hundreds or thousands of participants) and lack broad coverage of locations and time.

Our paper helps to fill the gap. We report new evidence on cash use in retail transactions, as well as on credit, debit, and check use, based on a large merchant transaction dataset. The data, provided by a discount retail chain, covers every transaction in the three years starting in April 2010, in each of the chain's thousands of stores across most of the country. In total, we have about 2 billion transactions, which involve millions of consumers.[1] With this rich dataset we explore three themes: (i) payment variation across transaction sizes and locations, (ii) payment variation at weekly and monthly frequencies, and (iii) payment variation over the longer term. In each case, we link our empirical findings to theoretical work on money demand and payment choice, indicating which of our findings are consistent with existing theories, and which suggest directions for extending those theories.

Traditional money demand theories emphasize opportunity cost, especially foregone interest, as a factor in households' decisions of how much cash to hold. The early models have only one means of payment, but Prescott (1987), Whitesell (1989), Freeman and Kydland (2000), and Lucas and Nicolini (2015) among others, have considered multiple means of payment in models where cash payments mainly incur proportional costs while non-cash payments require a fixed per-transaction cost. Essentially, those models offer a "threshold framework," in which each consumer has a threshold transaction size below (above) which they use cash (non-cash means of payments), with the threshold determined by those costs. According to this framework, the share of cash transactions is the fraction of shoppers whose transaction sizes are below their thresholds, so the cash share

[1] If a consumer visits a store once a week, the data would cover more than twelve million consumers; even if we assume daily shopping, it would still cover almost two million consumers.

decreases in transaction size as more consumers cross their thresholds.

The threshold framework guides our study of payment variation across transaction sizes, locations, and time. Motivated by the threshold framework, we include explanatory variables in our empirical model that proxy for the costs of using cash relative to other means of payment. Given that our data only identifies transactions, not customers, the demographic and economic characteristics of the zip code in which each store is located are assumed to reflect the characteristics of the store's customers and the economic environment in which they live. In addition, considering that a nontrivial fraction of the U.S. population is unbanked or underbanked, we also include zip-code-level variables in the empirical model that are likely to be correlated with consumers' adoption of non-cash payments. Finally, state and time dummies are included to control for state and time fixed effects.

We start with two sets of regression analysis. The first one includes transactions of all sizes in a single regression, where the dependent variables are shares of each payment type for each zip-code day. The explanatory variables include median transaction size in a zip-code day, which controls for the transaction size distribution, and other variables that capture the threshold distribution. In the second set of regressions, we group the data by transaction size and estimate separate models for each group. In both sets of regressions, our findings confirm the implications of the threshold framework: The share of cash transactions decreases in transaction size, and conditional on transaction size, zip-code-level variables reflecting a higher opportunity cost of holding cash are negatively associated with the cash share, as are variables proxying for access to, or adoption of non-cash payments.

The estimation results provide the basis for exploring the three aforementioned themes of payment patterns. The first theme is payment variation across transaction sizes and locations. In addition to finding that the fraction of cash (non-cash) payments decreases (increases) in transaction size in a given zip code, we also find that the cross-location dispersion of the payment mix increases with transaction size, which indicates that the threshold distributions across locations exhibit more variation for larger transaction sizes. A quantitative decomposition reveals that the estimated threshold distributions cannot be explained by transaction-size fixed effects. They are instead primarily determined by location-specific characteristics, which in part proxy for the degree of access to non-cash payments.

The second theme is the day-of-week and day-of-month patterns. According to the threshold framework, these patterns could be driven by factors that affect the threshold distributions at weekly and monthly frequencies. Over the course of the week, the cash and debit fractions are nearly mirror images of each other, whereas over the month, credit comes closer to mirroring cash. In addition, the *number* of transactions shows time patterns similar to those for the cash share. We interpret the high correlation between the number of transactions and the cash share of transactions as indicating that consumers are subject to time-varying financial or cash constraints through the week and month (likely related to their frequency of pay), which then affect their shopping patterns and payment choices.

The last theme involves the seasonal cycles and longer-run trends in the payment mix identified by the month-of-sample dummies in our regressions. Over the longer term, the shares of cash and check transactions decline steadily, while debit and credit's shares rise. The overall cash fraction of transactions is estimated to have declined by 2.46 percentage points per year in our three-year sample period, largely replaced by debit. We argue that the decline in cash (in other words, the leftward shift of the threshold distribution) at this particular retailer was likely not driven by transitory factors, and if the decline were to continue, only a relatively small fraction could be explained by forecasted changes in the zip-code-level variables, including age-cohort composition. This leaves a large fraction of the time trend to be explained, with prime candidates being technological progress in debit and changing consumer perceptions of debit relative to cash.

Besides the theoretical works mentioned above, our paper is closely related to the growing empirical literature on payment choice. Klee (2008) also used merchant transaction records to study consumer payment choices. However, with a different empirical strategy and a much broader dataset, we are able to investigate cross-sectional and time-series payment variation that was not addressed in Klee's study. Unlike prior work, our analysis is also explicit about the threshold framework and its implications. Our findings on the negative relationship between cash use and transaction size are consistent with some recent papers that study payment choice using consumer survey data (e.g., Arango et al., 2011, Cohen and Rysman, 2014, Wakamori and Welte, forthcoming). Nevertheless, those studies typically focus on the effects of consumer-specific characteristics on payment choice, but ignore the effects of locationwide factors and time. Our work also complements recent research on how consumer payment choice is affected by attributes of payment instruments such as pricing, speed, security, and access to credit (e.g., Borzekowski et al., 2008, Borzekowski and Kiser, 2008, Zinman, 2009, Ching and Hayashi, 2010, Schuh and Stavins, 2012, Stavins, 2013, Koulayev et al., 2016). While our data does not contain direct information on payment attributes, we point out that they could be factors driving the longer-run payment trends observed in our data.

Finally, it is worth mentioning that compared with many existing studies, our data is especially informative about cash use. The stores in our sample are located in relatively low-income zip codes where the customer base is more likely to be unbanked or underbanked. The stores also have a large share of small-dollar transactions, for which cash has remained stubbornly popular. While our data likely overstates the *proportion* of cash use in U.S. retail transactions, this very fact means that it provides valuable insights into the *nature* of cash use.

The structure of the paper is as follows. Section 2 describes the data and the threshold framework. Section 3 introduces the regression model and estimation method, and presents an overview of the estimation results. The next three sections explore the three themes introduced above: payment variation across transaction sizes and locations in Section 4; weekly and monthly payment patterns in Section 5; and longer-run payment variation in Section 6. Section 7 concludes and suggests directions for future research.

2 Data and Threshold Framework

The transactions data is from a large discount retailer with thousands of stores, covering most U.S. states. The stores sell a wide variety of goods in various price ranges, with household consumables such as food and health-and-beauty aids accounting for a majority of sales. The unit of observation is a transaction, and the time period is April 1, 2010 through March 30, 2013. For each transaction, the data includes means of payment, time, location, and amount. In the remainder of this section we will provide summary figures about payment variation in the data; we will describe the theoretical framework that guides our analysis of the data; and we will discuss the explanatory variables used to conduct that analysis.

2.1 Payment Variation

Figures 1 and 2 provide an overview of the transactions data, in terms of payment variation across time, across locations, and across transaction sizes. To construct those figures and for our empirical analysis, we include only transactions that consist of a sale of goods, with one payment type used, where the payment type is cash, credit card, debit card, or check – the four general-purpose means of payment.[2] The retailer also provides cash-back services, and the purchase components of cash-

[2] Data limitations prevent us from distinguishing credit cards from signature debit and prepaid cards. However, our estimates reveal time variation in what we report as "credit cards" that is significantly different from the variation in PIN debit. Because signature debit and prepaid cards are close substitutes for PIN debit, in that they rely on

back transactions are included in our analysis. In contrast, transactions made with special-purpose means of payment such as EBT, coupons and store return cards are excluded. All told, our empirical analysis covers 94% of the total transactions in the sample period. Our summary of the data in Figures 1 and 2 will refer to all stores; the zip-code-level data introduced below and used in the empirical analysis covers most of those stores' zip codes, but we will need to omit a small fraction of the retail outlets from that analysis because the zip-code-level data is unavailable.

Figure 1A presents payment variation across time in our sample. The data are plotted at the daily level, displaying the fractions of all transactions accounted for by each payment type. Note that while only cash is measured on the left axis, both axes vary by 0.35 from bottom to top, so fluctuations for each payment type are displayed comparably. The figure shows that cash is the dominant payment instrument at this retailer, followed by debit, credit and check. Over the sample period, the fractions of cash and check are trending down, with debit and credit trending up. There seems to be a weekly pattern in both cash and debit shares, with the two moving in opposite directions. Credit displays a monthly pattern, rising over the course of the month. The econometric model will allow for these patterns through day-of-week, day-of-month, and month-of-sample dummies.

We turn now to payment variation across locations. Figure 1B restricts attention to the last full month of the sample, March 2013, aggregates the data by zip code, and displays smoothed estimates of the density functions for fraction of transactions conducted with cash, debit, credit, and check. Restricting to only one month is necessary because of the time trend evident in Figure 1A. The ranking from Figure 1A is also apparent in Figure 1B: Cash is the dominant form of payment, followed by debit, credit, and check.[3] The main message of Figure 1B, however, is that there is significant variation across zip codes in cash and debit use, and to a lesser extent in credit use as well. This variation highlights the need for including location-specific variables in our econometric model.

consumers' account balances rather than borrowed funds, we can reasonably assume the estimated time patterns are primarily driven by the true credit cards.

[3] Note that the estimated kernel density for checks is truncated in Figure 1B. The check fractions are concentrated near zero, so the figure would be uninformative about the other payment instruments if we extended the y-scale to include the entire check density.

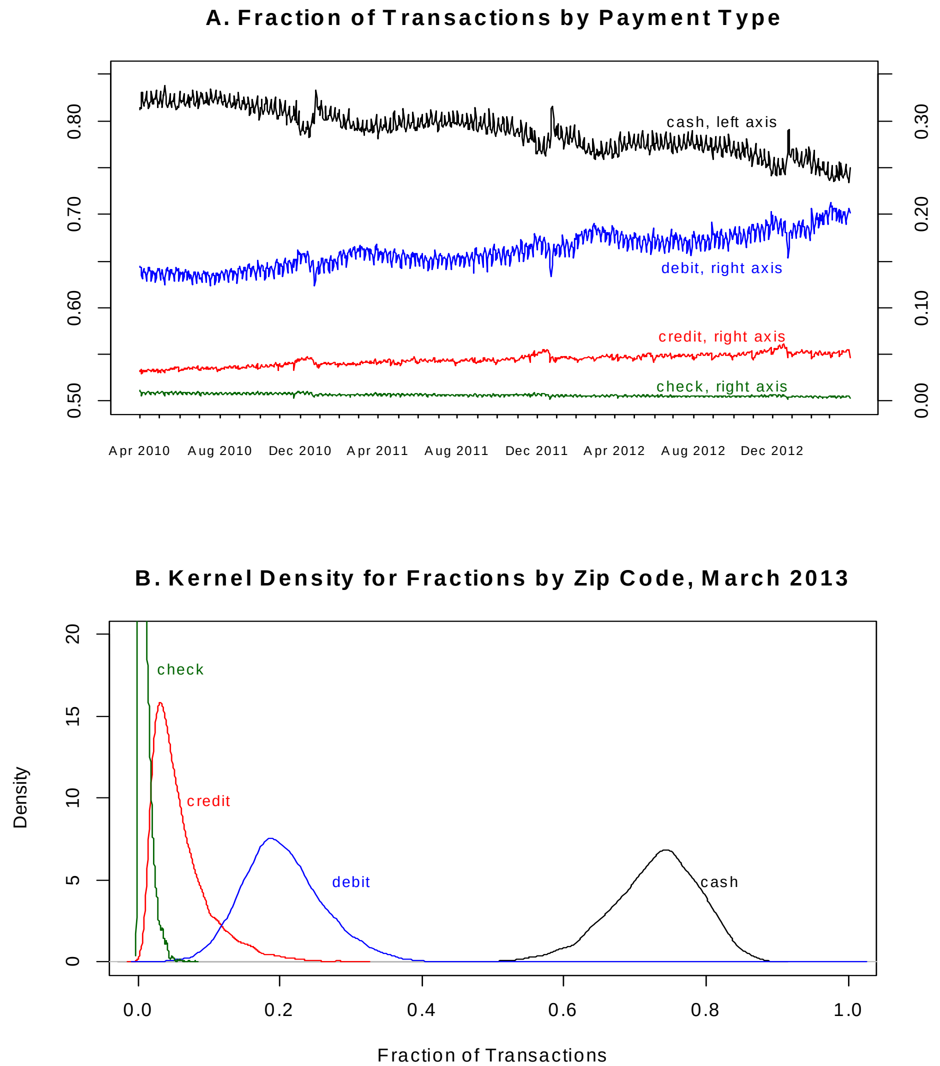


Figure 1. Payment variation across time and locations.

In Figure 2 we show how the payment mix varies with transaction size, again restricting attention to March 2013. To construct Figure 2, for each zip-code day we group the data by transaction size, using $1 bins between $1 and $15, $5 bins between $15 and $50, and combining all transactions above $50 into one bin. These categories were chosen to ensure a sufficient number of transactions in each bin. For transactions in a given size bin, we calculate the shares of the four payment types on each zip-code day. The solid lines represent the median across zip-code days of the payment shares, and the dashed lines represent the 5th and 95th percentiles of the distribution. The overall message of Figure 2 is twofold: Cash is relatively more important for small transactions; and the distribution of payment shares across zip-code days exhibits increasing dispersion for higher transaction sizes, as reflected by the fanning out of the 5th and 95th percentiles. Our analysis will show how these patterns relate to the threshold framework, and what factors may determine the empirical distributions of the thresholds.

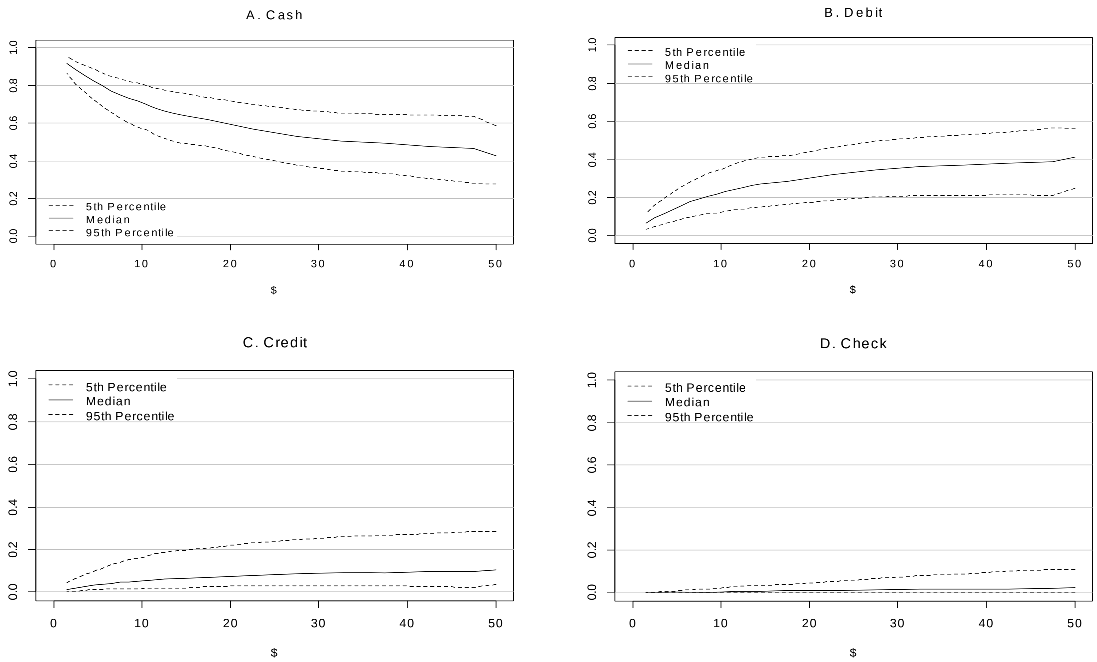


Figure 2. Payment variation across transaction sizes.

2.2 Threshold Framework

Figures 1 and 2 reveal sizable variation in payment shares across time, locations, and transaction sizes. The relationship between payment shares and transaction size in Figure 2 leads us to adopt the threshold framework as our organizing principle for analyzing the data. We conjecture that each consumer has a threshold transaction size below which they use cash, and above which they use a non-cash means of payment. Under certain conditions, it follows that the fraction of cash transactions at a given transaction size represents the fraction of consumers with a cash threshold at least as large as that transaction size.[4] The threshold framework arises in each of the money demand and payment choice models referenced in the introduction. The common features of those models are that (1) there is a fixed per-transaction cost of using non-cash means of payment, and (2) the cost of using cash is mainly proportional, but there is a fixed cost of replenishing cash inventory. Examples for (1) include fees, restrictions or risks of identity theft that are related to checks and payment cards. Examples of the proportional costs of using cash include nominal interest rates and

[4]The main requirement is that the transaction size distribution for an individual consumer not vary too strongly with their cash threshold.

the risks of robbery or theft. Although there was little time-series variation in short-term interest rates over the period of our sample, the interest rate opportunity cost of holding cash nonetheless varies across locations according to the availability and cost of banking services.

The threshold framework yields several important implications. First, the fraction of cash (non-cash) transactions decreases (increases) in transaction size. Everything else equal, as transaction size increases, more consumers would cross their cash thresholds and switch to non-cash payments. Second, consumer/location characteristics ought to be related to payment choice. For a given transaction size, the threshold framework implies that consumer/location variables proxying for the costs of using cash relative to other means of payment would be negatively associated with cash thresholds and thus the share of cash transactions. Finally, time effects can play a role in payment choice through an impact on cash thresholds and shopping patterns.

While we do not directly observe individual consumers' behavior, our data can be used to verify or quantify the implications of the threshold framework listed above. First, if the cash fraction of transactions were not decreasing in transaction size this would seem to contradict the framework: along this dimension, both the observed cash fractions in Figure 2A and the estimated empirical model provide support for the threshold framework. Second, the estimated model will be used to confirm and quantify the relationships between payment shares and variables that proxy for the relative costs of using cash. Finally, systematic time variation in payment shares will be quantified and interpreted as largely representing shifts over time in the distribution of cash thresholds.

2.3 Explanatory Variables

Guided by the threshold framework, our analysis includes transaction size, location-specific variables, and time dummies in an econometric model of payment shares. The location-specific variables include variables meant to describe the characteristics of shoppers in a given location – for example demographic variables; other variables that describe the socioeconomic environment in a location and are related to the relative costs of different means of payment – for example the degree of banking competition and the robbery rate; and state dummies. Table 1 lists summary statistics for the zip-code-level explanatory variables used in the regressions, fixed at their 2011 values.[5] We

[5] Most of our zip-code-level variables come from the U.S. Census's American Community Survey (ACS) and the FDIC's Summary of Deposits. The robbery data is from the FBI's Uniform Crime Report. We fix zip-code-level explanatory variables at their 2011 values (5-year estimates), because the ACS provides only 5-year estimates for areas with less than 20,000 residents, and the median zip code in our sample has less than 18,000 residents. In Section 6, on longer-run payment variation, we will explicitly account for the effects of time variation in zip-code-level explanatory variables.

also list summary statistics for zip-code-level explanatory variables for the entire country in Table A1 in the online Appendix to contrast our sample to the United States as a whole.

2.3.1 Transaction Size

Figure 3 provides information about the size distribution of transactions in March 2013. Figure 3A displays a smoothed density function for all transactions in the month, and Figure 3B plots the distribution of median transaction sizes across zip-code days. Figure 3B shows that there is substantial heterogeneity across locations with respect to size of transaction. Our empirical analysis will incorporate a role for transaction size in payment choice.

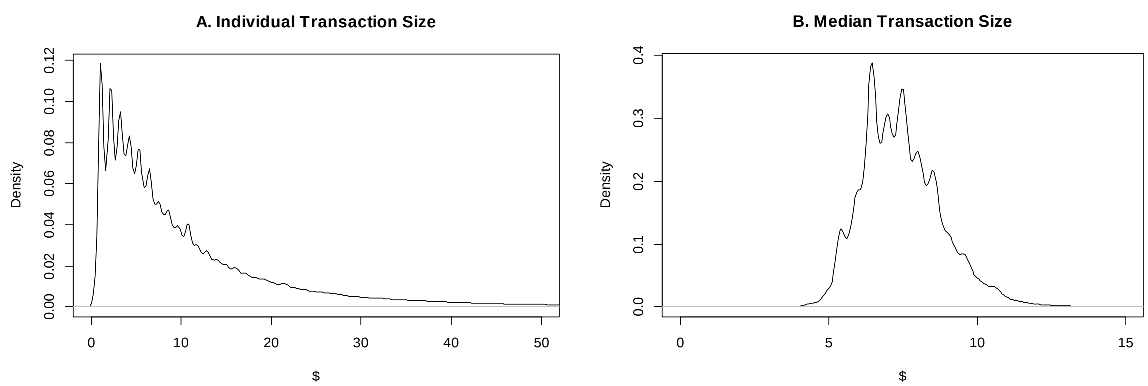


Figure 3. Kernel Densities of transaction size in March 2013.

2.3.2 Cash Holding and Payment Choice

Several of the explanatory variables represent aspects of the economic environment that have direct bearing on the cash-holding and payment-choice behavior considered in the theoretical literature. These include the banking competition measure, bank branches per capita, and the robbery rate. According to the threshold framework, cash use should decrease in banking-sector competition and the robbery rate (which increase consumers' opportunity costs of using cash), but increase in bank branches per capita (which reduces consumers' costs of replenishing cash balances). Following the banking literature and antitrust tradition, we measure banking-sector concentration by the HHI index in each MSA or rural county.[6]

[6]Both the theoretical literature and antitrust practice typically assume that the relevant geographic banking market is a local area where banks compete to offer financial services to households and small businesses. That market area is often approximated by a metropolitan area (MSA) in urban areas and by a county in rural areas. The most commonly used measure of market concentration is the Herfindahl-Hirschman Index (HHI), calculated by squaring each bank's share of deposits in a market and then summing these squared shares.

Table 1. Summary statistics of zip-code variables

Variable (unit)	Mean	Std. dev.	1%	99%
Cash holding and payment choice				
Banking HHI Metro	0.155	0.097	0.059	0.615
Banking HHI Rural	0.283	0.145	0.105	0.809
Branches per capita ($1/10^3$)	0.47	2.14	0.05	1.66
Robbery rate ($1/10^5$)	13.88	29.60	0.00	179.15
Adoption of non-cash payments				
Median household income ($)	40,623	11,389	19,370	76,850
Deposits per capita ($)	2712	20,158	35.09	15,765
Population density (per mile^2)	1436	2643	4.2	12,021
Demographics (%)				
Family households	66.24	8.40	36.47	83.52
Housing: Renter-occupied	30.18	11.81	10.04	67.46
Owner-occupied	56.68	12.61	19.34	80.18
Vacant	13.14	8.18	3.93	46.96
Female	50.66	2.58	39.38	55.16
Age < 15	19.71	3.78	10.07	29.45
15-34	26.64	5.93	15.59	48.91
35-54	26.28	2.79	18.07	32.53
55-69	17.34	3.74	9.13	28.35
≥ 70	10.04	3.78	3.25	21.42
Race white	73.17	22.70	5.24	98.29
black	16.52	21.26	0.13	90.64
Hispanic	14.12	19.67	0.56	91.72
Native	1.22	4.53	0.07	17.56
Asian	1.55	2.43	0.06	12.50
Pac-Islr	0.07	0.22	0.00	0.68
other	5.07	7.03	0.07	32.87
multiple	2.39	1.31	0.55	6.77
Educ below high school	18.16	8.88	4.60	47.10
high school	34.07	7.48	15.30	50.90
some college	21.38	4.41	10.90	31.70
college	26.39	10.50	8.70	57.70

The distribution of the HHI index in our sample is comparable with the overall nation, and banking markets are significantly more concentrated in rural counties than in MSAs. However, the average number of bank branches per capita in our sample appears smaller than that in the entire U.S. We measure the robbery rate at the county level (and will have to discard some zip codes from our analysis because of missing robbery data), and the robbery rate in our sample is not appreciably different than in the nation as a whole.

2.3.3 Adoption of Non-cash Payments

We generalize the threshold framework slightly to incorporate a fraction of the population that is unbanked or underbanked, who have not adopted banking services or payment cards. The pure threshold framework can accommodate those consumers, but it interprets them as simply being part of the group of consumers with such a high cash threshold that it does not bind. There is a clear sense, however, in which adoption can represent a distinct extensive margin, compared to the intensive margin associated with cash holding and payment choice. This leads us think in terms of an adoption decision layered on top of the threshold framework.

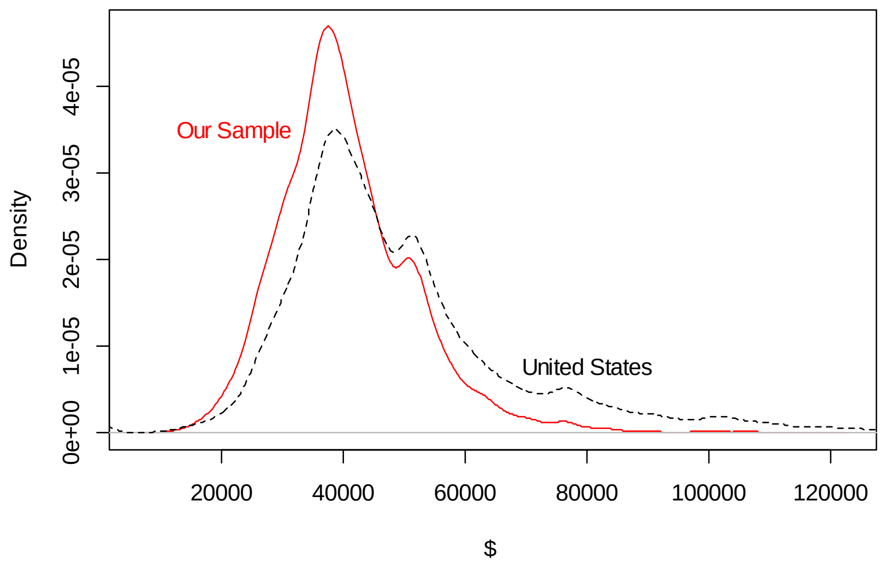


Figure 4. Distribution of median household income across zip codes.

Three of the variables we include, median household income, deposits per capita, and population density may be correlated with adoption, that is, the likelihood that consumers have bank accounts or own credit or debit cards.[7] The mean value of median household income is 20 percent lower in our sample than in the U.S. as a whole. Figure 4 provides additional detail, plotting kernel smoothed density functions for median income in our sample of zip codes and in the United States.

[7]Note that our classification of variables should not be taken as exclusive: Banking competition, prevalence of bank branches, and the robbery rate may also affect a household's choices of whether to adopt non-cash forms of payment. Likewise, while we classified deposits and income as "adoption" variables, to the extent that they proxy for the opportunity costs of households' time, they may also fall into the intensive margin category: Households with a high opportunity cost of time face higher costs of replenishing their cash balances, and will therefore use cash less often.

Mean deposits per capita are dramatically lower in our sample than in the entire country, but the nationwide value is likely driven by a small number of zip codes with extremely large bank branches. Population density is relevant for adoption because, as McAndrews and Wang (2012) point out, replacing traditional paper payments with electronic payments requires merchants and consumers to each pay a fixed cost but reduces marginal costs for doing transactions. Their work suggests that adoption and usage of electronic payment instruments should be higher in areas with a high population density or more business activity. The zip codes in our sample are somewhat less densely populated than in the broader U.S.

2.3.4 Demographics

We include a range of demographic variables in the regressions under the assumption that these variables may be systematically related to payment adoption and usage. Relative to the United States average, the zip codes in our sample have a low percentage of owner-occupied dwellings, with little variation. The racial composition of these zip codes also differs markedly from the rest of the country: There is a higher percentage of Blacks, Hispanics, and Native Americans and a lower percentage of Whites and Asians. Also, there is a relatively low percentage of college graduates. However, the age, gender and family profiles of our sample are not significantly different from the nation as a whole.

3 Empirical Analysis: Estimating Payment Shares

In this section we introduce an empirical model aimed at explicitly relating the location-specific variables, as well as time dummies, to payment shares. In our first specification we aggregate all transactions by zip-code day, and include median transaction size for each zip-code day as an explanatory variable. In the second specification we omit median transaction size, but estimate separate regressions for 22 different transaction-size bins.

3.1 Empirical Model

The data is analyzed using a fractional multinomial logit model (FMLogit). The dependent variables are the fractions of each of the four payment instruments used in transactions at stores in one zip code on one day between April 1, 2010, and March 31, 2013.[8] The explanatory variables fall

[8] In our sample, more than three quarters of zip codes have only one store. Because we measure the fraction of payment instruments at the zip code level, we do not distinguish locations with one store from those with multiple

into three groups, for the purposes of our discussion in Sections 4-6: (1) location-specific variables, including median transaction size; the economic and demographic variables listed above; and state-level dummies; (2) day-of-week and day-of-month dummies; and (3) month-of-sample dummies. All together, we will show that these explanatory variables account for substantial variation in the dependent variables. While our regression analysis does not aim to identify causality, it provides a convenient framework for studying payment variation, (i) across transaction sizes and locations, (ii) at weekly and monthly frequencies, and (iii) over the longer term.

The FMLogit model conforms to the multiple fractional nature of the dependent variables, namely that the fraction of payments for each instrument should remain between 0 and 1, and the fractions add up to 1. The FMLogit model is a multivariate generalization of the method proposed by Papke and Wooldridge (1996) for handling univariate fractional response data using quasi-maximum likelihood estimation. Mullahy (2010) provides more econometric details.

Formally, consider a random sample of $i = 1, ..., N$ zip-code-day observations, each with M outcomes of payment shares. In our context, $M = 4$, which correspond to cash, debit, credit, and check. Let s_{ik} represent the k^{th} outcome for observation i, and x_i, $i = 1, ..., N$, be a vector of exogenous covariates which are the explanatory variables listed above. The nature of our data requires that

$$s_{ik} \in [0, 1] \qquad k = 1, ..., M; \tag{1}$$

$$\Pr(s_{ik} = 0 \mid x_i) \geq 0 \quad \text{and} \quad \Pr(s_{ik} = 1 \mid x_i) \geq 0; \tag{2}$$

$$\text{and} \quad \sum_{m=1}^{M} s_{im} = 1 \quad \text{for all } i. \tag{3}$$

Given the properties of the data, the FMLogit model provides consistent estimates by enforcing conditions (4) and (5),

$$E[s_k|x] = G_k(x; \beta) \in (0, 1), \quad k = 1, ..., M; \tag{4}$$

$$\sum_{m=1}^{M} E[s_m \mid x] = 1; \tag{5}$$

stores. In the latter case, we simply sum up the transactions of all the stores in the zip code.

and also accommodating conditions (6) and (7),

$$\Pr(s_k = 0 \mid x) \geq 0 \qquad k = 1, ..., M; \tag{6}$$

$$\Pr(s_k = 1 \mid x) \geq 0 \qquad k = 1, ..., M; \tag{7}$$

where $\beta = [\beta_1, ..., \beta_M]$.[9] Specifically, the FMLogit model assumes that the M conditional means have a multinomial logit functional form in linear indexes as

$$E[s_k \mid x] = G_k(x; \beta) = \frac{\exp(x\beta_k)}{\sum_{m=1}^{M} \exp(x\beta_m)}, \qquad k = 1, ..., M. \tag{8}$$

As with the multinomial logit estimator, one needs to normalize $\beta_M = 0$ for identification purposes. Therefore, (8) can be rewritten as

$$G_k(x; \beta) = \frac{\exp(x\beta_k)}{1 + \sum_{m=1}^{M-1} \exp(x\beta_m)}, \qquad k = 1, ..., M-1; \tag{9}$$

and

$$G_M(x; \beta) = \frac{1}{1 + \sum_{m=1}^{M-1} \exp(x\beta_m)}. \tag{10}$$

Finally, one can define a multinomial logit quasi-likelihood function $L(\beta)$ that takes the functional forms (9) and (10), and uses the observed shares $s_{ik} \in [0, 1]$ in place of the binary indicator that would otherwise be used by a multinomial logit likelihood function, such that

$$L(\beta) = \prod_{i=1}^{N} \prod_{m=1}^{M} G_m(x_i; \beta)^{s_{im}}. \tag{11}$$

The consistency of the resulting parameter estimates $\hat{\beta}$ then follows from the proof in Gourieroux et al. (1984), which ensures a unique maximizer. In the following analysis, we use Stata code developed by Buis (2008) for estimating the FMLogit model.[10]

[9] To simplify the notation, the "i" subscript is suppressed in Equations (4)-(10).

[10] The FMLogit model we use is closely related to the Multinomial logit (MLogit) model, but it aggregates individual transactions up to the shares of transactions for each payment type on each zip-code day. Aggregation allows us to use all transactions (which the MLogit model would not be able to handle), and it smooths out the "noise" in individual transactions.

3.2 Estimates for All Transaction Sizes Together

In Table 2 we report estimation results for the first specification, where we model payment shares at the zip-code-day level, including all transactions greater than $1. The coefficient estimates are expressed in terms of marginal effects.[11]

3.2.1 Cash Holding and Payment Choice

As suggested by theory, we assume that each consumer has a threshold transaction size (possibly time-varying), below which they only use cash. Aggregating transactions within a zip-code day, we then expect to find that a rightward shift in the size distribution of transactions corresponds to a lower share of cash transactions. Using median transaction size as a convenient summary of the size distribution, we find the expected result: Evaluating at the mean of median transaction size, $6.86, the marginal effects indicate that a $1 increase in median transaction size reduces the predicted cash share by 1.8 percentage points but raises debit by 1.2 percentage points, credit by 0.5 percentage points, and check by 0.1 percentage points.

We find that higher banking concentration corresponds to a higher cash share (lower card shares) in rural areas, but a lower cash share (higher card shares) in MSAs. We conjecture that in rural areas HHI does a good job proxying for banks' market power, whereas in metro areas it may not: In metro areas, banking is inherently competitive, and a high level of concentration (as measured by HHI) may simply indicate the presence of one or more especially efficient banks.[12] We also find the expected results for robbery rate and bank branches per capita.

3.2.2 Adoption of Non-cash Payments

For the variables that we classified as relating to the adoption decision, our coefficient estimates have the expected signs. The share of cash (card) transactions is negatively (positively) associated with median household income and deposits per capita. A higher population density is associated with higher shares of card payments, mainly offsetting a lower check share.

[11] For continuous variables, the marginal effects are calculated at the means of the independent variables. For dummy variables, the marginal effects are calculated by changing the dummy from zero to one, holding the other variables fixed at their means.

[12] Our findings on the rural and MSA markets correspond to two well-known hypotheses in the industrial organization and banking literature. The Structure-Conduct-Performance (SCP) hypothesis assumes that the ability of banks in a local market to charge high prices depends positively on market concentration. The Efficient-Structure (ES) hypothesis takes an opposite view and argues that a concentrated market may reflect the efficiency advantages of leading banks in the market, so it may instead be associated with lower prices for banking services. See Gilbert and Zaretsky (2003) for a comprehensive literature review.

Table 2. Marginal effects for zip-code-level variables

Variable	Cash	Debit	Credit	Check
Cash holding and payment choice				
Median transaction size	-0.018*	0.012*	0.005*	0.001*
Banking HHI	0.030*	-0.023*	-0.010*	0.003*
Banking HHI*Metro	-0.050*	0.032*	0.024*	-0.005*
Branches per capita	0.007*	-0.005*	-0.004*	0.001*
Robbery rate	-0.054*	0.063*	0.000	-0.010*
Adoption of non-cash payments				
Median household income	-0.033*	0.005*	0.036*	-0.008*
Deposits per capita	-0.006*	0.016*	0.000	-0.010*
Population density	-0.038*	0.079*	0.091*	-0.131*
Demographics				
Family households	-0.098*	0.089*	0.016*	-0.006*
Housing Owner-occupied	-0.020*	0.006*	0.006*	0.008*
Vacant	-0.040*	0.011*	0.026*	0.004*
Female	-0.052*	0.080*	-0.005*	-0.023*
Age 15-34	-0.184*	0.163*	0.034*	-0.013*
35-54	-0.152*	0.115*	0.053*	-0.016*
55-69	0.031*	0.000	-0.013*	-0.018*
$\geq$ 70	-0.024*	-0.038*	0.054*	0.007*
Race black	0.055*	-0.025*	-0.020*	-0.011*
Hispanic	0.024*	-0.019*	0.003*	-0.007*
Native	0.133*	-0.074*	-0.052*	-0.007*
Asian	-0.018*	0.008*	0.032*	-0.022*
Pac-Islr	-0.311*	0.550*	-0.203*	-0.036*
other	0.091*	-0.042*	-0.048*	-0.001*
multiple	-0.081*	0.109*	0.005*	-0.032*
Edu high school	-0.202*	0.138*	0.057*	0.007*
some college	-0.322*	0.233*	0.088*	0.001*
college	-0.225*	0.140*	0.079*	0.007*
Pseudo R^2(incl state, time)	0.59	0.58	0.59	0.57
R^2 (excl state)	0.46	0.20	0.31	0.42
R^2 (excl time)	0.52	0.51	0.56	0.51
R^2 (excl state, time)	0.36	0.14	0.28	0.36
Zip code-day observations	4,505,642	4,505,642	4,505,642	4,505,642

Note: *1% significance level based on robust standard errors. The dependent variables are the fractions of each of the four general payment instruments used in transactions at stores in a zip code on a day between April 1, 2010, and March 31, 2013. The independent variables take their values in 2011. Banking HHI index is calculated by squaring each bank's share of deposits in a market (an MSA or a rural county) and then summing these squared shares. Metro is a dummy variable taking the value one when the banking market is an MSA, otherwise equal to zero. Branches per capita is measured as the number of bank branches per 100 residents in a zip code. Robbery rate is defined as the number of robberies per 100 residents in a county. Median household income is measured in units of \$100,000 per household in a zip code. Deposits per capita is measured in units of \$10,000 deposits per resident in a zip code. Population density is measured in units of 100,000 residents per square mile in a zip code. All the demographic variables are expressed as fractions.

3.2.3 Demographics

We find that demographic characteristics such as age, race, and education are systematically related to consumer payment choices, in a manner largely consistent with previous studies based on consumer surveys. Age effects are especially notable. A higher presence of older age groups is associated with greater use of payment cards relative to the baseline age group, under 15. This might be simply because minors do not have access to non-cash payments, or because families with children tend to use more cash or checks. However, the age profile with respect to cash and checks is non-monotonic. A higher presence of the age group 55-69 is associated with a significantly higher cash fraction, while a higher presence of people at age 70 and older is associated with a higher check fraction. These findings suggest that the age variables may be standing in primarily for cohort effects: Older people tend to be cash or check users not because they are older but because they did not have access to cards when they first reached adulthood. For the purpose of projecting future cash use as we do in Section 6, the distinction between cohort and age interpretations is important, and we will adopt the cohort interpretation except for the youngest age group.

3.2.4 State and Time Effects

The regression also includes dummy variables for state, for day of week, day of month, and month of sample. The estimates for state dummies reveal substantial payment variation across states (Figure A1 in the online Appendix plots histograms of state dummies for each payment type). Conditioning on the other variables, the cross-state variation appears largest in the fraction of debit, followed by credit, cash, and check. We defer discussion of the time effects until Sections 5 and 6 below.

Table 2 reports pseudo R^2 values when the time and state dummies are either included or excluded. Our full model explains 59 percent of the variation in cash fractions, and that number falls by 23 percentage points when both state and time effects are removed. The contributions of state and time dummies are nearly additive; that is, removing "state and time" effects is close to removing state effects plus removing time effects.

3.3 Estimates by Transaction Size Class

We now turn to an analysis based on individual transaction sizes, estimating separate regressions for the 22 transaction size bins used in Figure 2 at the zip-code-day level. This will explain payment variation across transaction sizes and locations, and yield an estimated counterpart to Figure 2.

3.3.1 Estimation Approach

Before aggregating to the zip-code-day level the sample is subdivided by transaction size class. This allows us to use FMLogit regressions as before, but based on subsamples according to transaction sizes. In the background, the threshold framework continues to motivate our analysis: The fraction of cash payments at a particular transaction size for a given zip-code day represents the fraction of shoppers whose threshold for cash use lies above that transaction size. Conditioning on transaction size puts the focus here on "marginal" payment shares instead of "total" payment shares. With heterogeneous consumers, the payment share at a particular transaction size still depends on the distribution of consumer characteristics, the economic environment, and calendar time.

All coefficient estimates are allowed to vary across transaction size regressions. A more restrictive approach would impose common coefficients on zip-code-level variables, allowing only the constant terms to vary across each transaction size regression. We will see in Section 4 that the data do not appear consistent with common coefficients. The sensitivity of both level and dispersion of payment shares, shown in Figure 2, is attributed overwhelmingly to variation across transaction sizes in the coefficients on zip-code-level variables.

3.3.2 Findings: Marginal Effects and Amplification

For the sake of space, we plot marginal effects for cash in Figure 5, leaving the others to the online Appendix. Most zip-code-level explanatory variables show a sign consistent with our estimates for the overall zip-code-day shares. In fact, our marginal-effect estimates from the overall regression in Section 3.2 closely match those from the \$6-\$7 transaction size regression (Recall that for our overall sample, the mean value of zip-code-day median transaction size is \$6.86). Moreover, as transaction size increases, the marginal effects for most explanatory variables are increasing in absolute value (we refer to this pattern as amplification). Such patterns are found for debit, credit and check as well, as shown in Figures A3-A5 in the online Appendix.[13]

The state and time effects also get amplified with transaction size. We provide histograms of state fixed effects in Figure A6 in the online Appendix, and again defer discussion of the time effects until Sections 5 and 6.

[13] In a few cases, the amplification of marginal effects is more subtle. For example, in Figure 5, population density shows a small and decreasing relationship with the cash share as transaction size increases. However, a careful look into the results in the online Appendix (Figures A3-A5) shows that population density mainly affects the substitution between cards and check, while cash only captures a small residual effect. In fact, the marginal effects of population density clearly get amplified with transaction size for debit, credit and check.

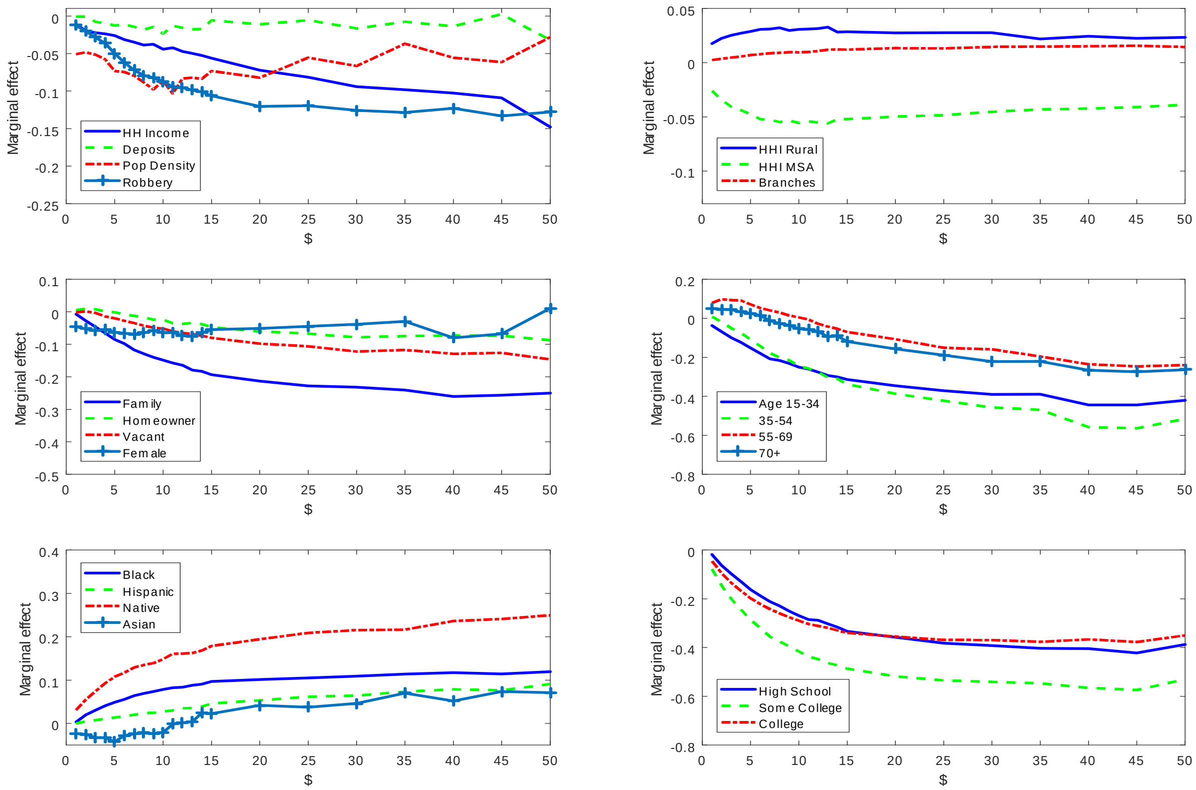


Figure 5. Cash marginal effects by transaction size

4 Payment Variation across Transaction Sizes and Locations

Figure 2 showed that the fraction of cash (non-cash) transactions is decreasing (increasing) in transaction size, and that the dispersion of these fractions across locations is increasing in transaction size. In this section we investigate the counterpart to Figure 2 implied by the estimated model. We begin by confirming that the estimated counterpart replicates the patterns in Figure 2, and then provide an interpretation based on the cash thresholds of individual shoppers. Finally, a decomposition shows that the estimated relationship between transaction size and payment shares (level and dispersion) is accounted for not by a fixed transaction-size effect, but rather by the amplifying effect of the zip-code-level variables.

4.1 Threshold Distributions

Figure 6 displays the estimated counterpart to the raw data of Figure 2. For each size class, the median, 5th, and 95th percentiles of the distribution of predicted values are plotted for the four payment shares. Comparing the two figures, it is clear that the estimated models for each transaction size are successful at replicating both (i) the relationship between transaction size and the level of payment composition, and (ii) the relationship between transaction size and the dispersion of payment composition across zip-code days.

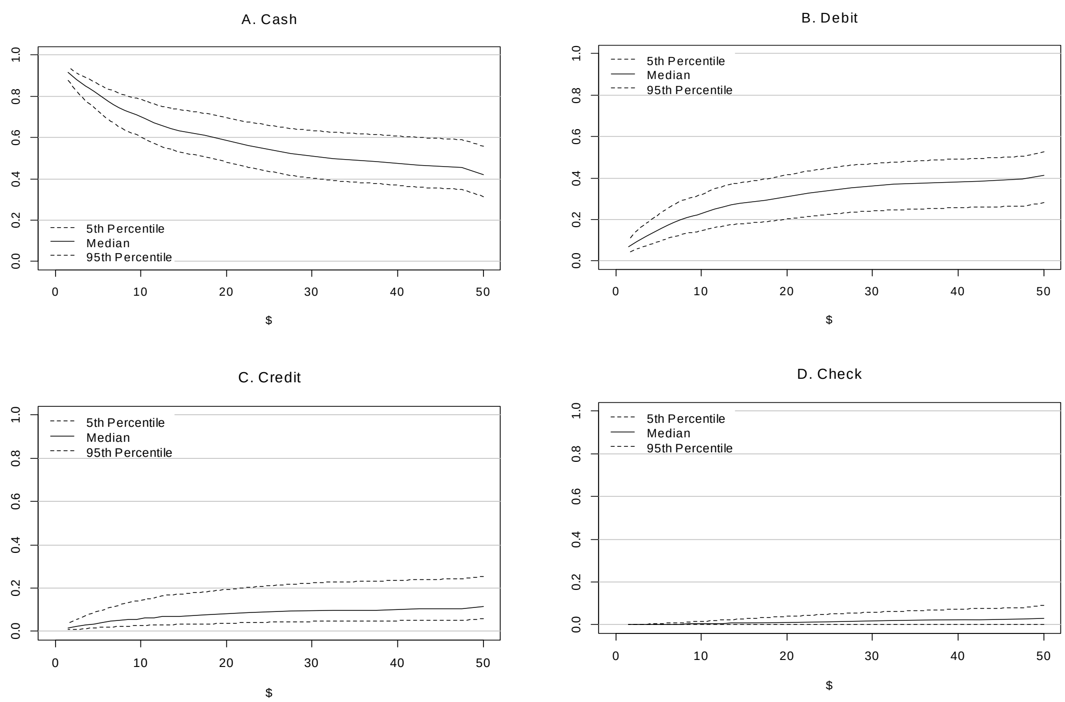


Figure 6. Predicted payment variation across transaction sizes.

Under the assumption that the transaction size distribution for an individual consumer does not vary too strongly with their cash threshold, our model estimates uncover the empirical distributions of thresholds implied by theory. Specifically, the sum of all three non-cash payment fractions shown in Figure 6 (or one minus the cash fractions) provides the empirical bounds for location-specific cumulative distribution functions of the cash thresholds in our sample. The figures show that for $1 transactions, about 90 percent of shoppers are under their cash thresholds in most locations. As transaction size increases, more shoppers pass their cash thresholds, and the dispersion of cdfs increases across locations. Eventually, for transactions $50 and above, 58 percent of shoppers have passed their cash thresholds in the median location, while the fraction is 45 percent in the 5th percentile location and 70 percent in the 95th percentile location.[14]

[14] It is useful to compare our model-predicted payment choice probabilities with those in Klee (2008) and Briglevics and Schuh (2014). Those two studies focus on grocery transactions, one using data from a grocery store chain in 2001 and the other using 2012 Diary of Consumer Payment Choice (DCPC) data. Their results show that in the past decade, there has been a dramatic decline in cash and check use in grocery transactions, and a substantial increase

4.2 Determinants of Threshold Distributions

We now discuss how those threshold distributions are related to the amplification of marginal effects. The explanatory variables are first divided into two groups: One comprises constant terms, which include the intercept and time and state-level fixed effects, and the other comprises all zip-code-level variables. Our goal is to quantify the relative contributions of the two groups of variables to the levels and dispersion of the payment mix across transaction sizes. Figure 6 shows that larger transaction sizes are associated with a lower share of cash payments, but that finding is consistent with either the constant terms or zip-code-level variables driving the shares in the transaction size-class regressions. Likewise, our theoretical framework of individual-specific threshold transaction sizes does not tell us which effect should dominate.[15]

With this division of variables, we decompose the level and dispersion of the payment mix across transaction sizes using the \$1-\$2 regression as a benchmark. First, the constant terms are allowed to take on their estimated values in each of the size-class regressions, holding fixed the coefficients on zip-code-level variables at the \$1-\$2 benchmark. Then the coefficients on the zip-code-level variables are allowed to take on their estimated values in each of the size-class regressions, holding fixed the constant terms at their \$1-\$2 benchmark. The results of this decomposition are shown in Figure 7. For each size class the median, 5th, and 95th percentiles of the distribution of counterfactual values are plotted for each payment fraction. The lines marked with "x" come from the first exercise described above, and the lines marked with "o"s come from the second exercise.

Because of the nonlinearity inherent in the FMLogit model, the decomposition is not additive. In addition, there is no guarantee that it will unambiguously assign the change in the payment mix as transaction size changes to one or the other set of coefficients. However, Figure 7 shows that the decomposition turns out to be relatively clean: Overwhelmingly, it is changes in the coefficients on zip-code-level variables, rather than changes in constants, that account for changes in the level and dispersion of each payment type.

---

in credit and debit use (See Figure A2 in the online Appendix for the comparison). For example, considering a \$50 transaction, the probability of cash use was 30% in 2001 but dropped to 15% in 2012, while check use dropped from more than 30% to almost zero. Meanwhile, credit use (including signature debit) increased from less than 20% to 55%, and PIN debit use increased from 20% to 30%. Comparing with their results, estimates based on our sample show a much higher probability of cash and PIN debit use, both about 40% for a \$50 transaction in the median location.

[15] Parametric assumptions about (1) the function matching characteristics to the threshold transaction size, and (2) the distribution of characteristics would imply restrictions on the roles of constant terms and zip-code level variables, with the latter standing in for the zip-code-level distribution of characteristics. However, we choose to view our FMLogit model as a low-order approximation to arbitrary threshold functions and distributions of characteristics.

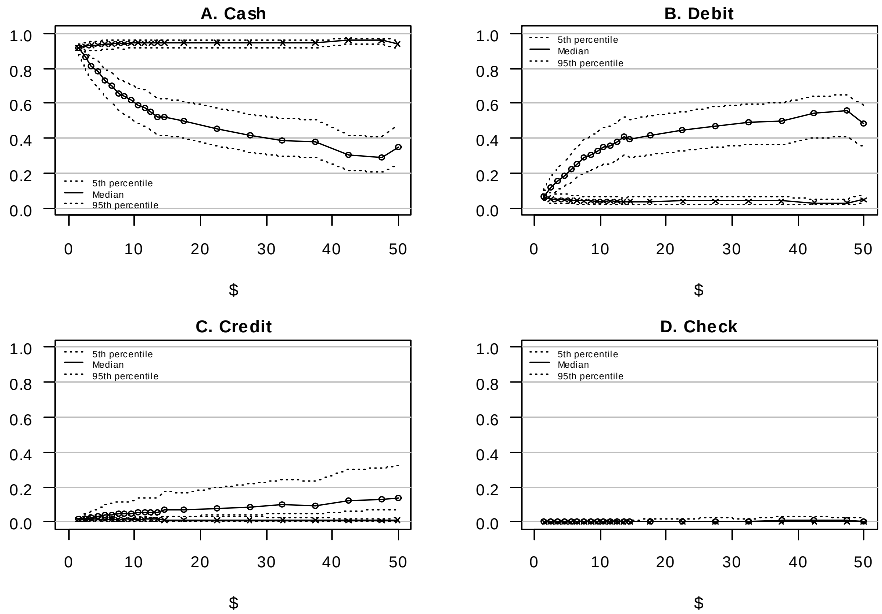


Figure 7. Decomposition of payment variation.

(Fixed zip-code-level coefficients (x) Fixed constants(o))

As mentioned, the negative relationship between transaction size and the level of cash fractions is consistent with the threshold framework. The increasing relationship between dispersion of payment fractions and transaction size is a new empirical finding. Our decomposition suggests an intuitive explanation for the relationship: For higher transaction sizes, the fixed per-transaction costs of using non-cash instruments become less important, and thus consumers in locations with better access to those instruments behave increasingly differently than consumers in locations with worse access. According to this explanation, the zip-code-level variables in part proxy for the degree of adopting non-cash payments. Therefore, it is useful to incorporate an adoption margin into the threshold framework in order to explain the dispersion in threshold distributions.

5 Payment Variation at Weekly and Monthly Frequencies

The analysis in the previous section concerned payment variation across transaction sizes and locations, holding time constant. In this section we discuss the estimated weekly and monthly time effects, holding location (and transaction size) fixed. After providing an overview of those effects, we offer an explanation based on consumers' time-varying financial positions and, for the weekly pattern, a "convenience" cycle. Both elements of the explanation involve time variation in the distribution of shoppers' cash thresholds.

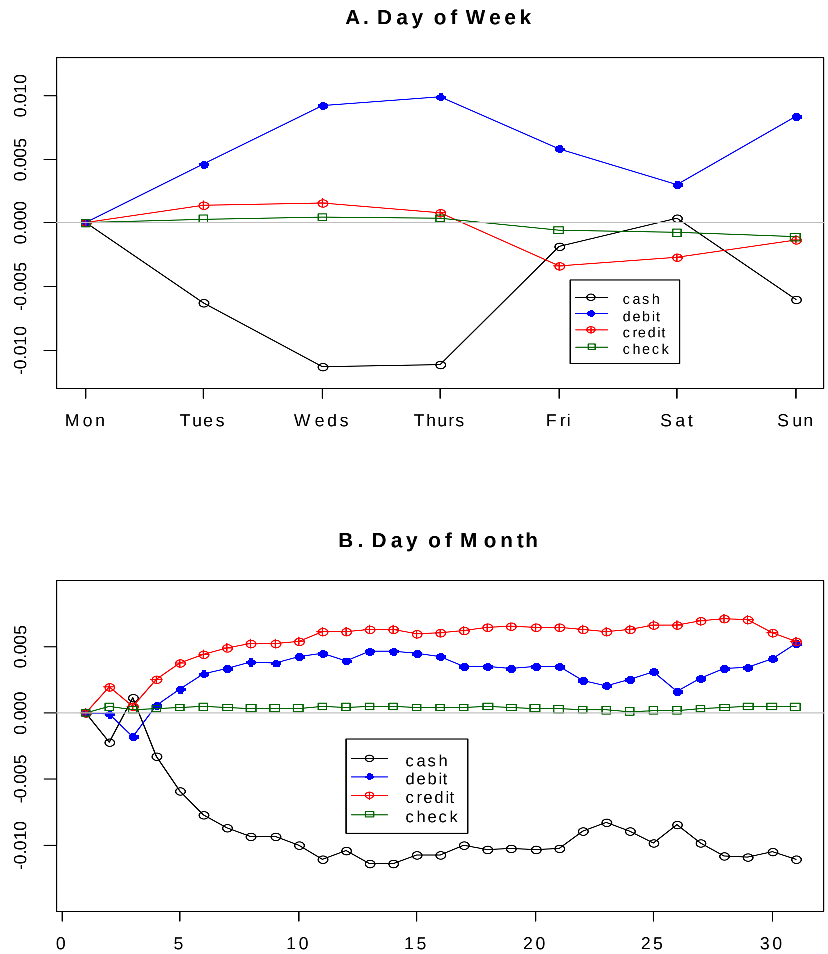


Figure 8. Daily marginal effects from overall regression.

5.1 Day-of-week and Day-of-month Effects

Figures 8A and 8B display the day-of-week and day-of-month marginal effects for the Section 3.2 regression, where all transaction sizes are pooled. The weekly cycle shows cash and debit to be nearly mirror images of each other, with credit displaying similar behavior to debit but with much smaller variation. In contrast, credit displays greater variation than debit over the month, although again their behavior is qualitatively similar and in opposition to cash. Note that while the patterns are quite striking, the variation in payment shares is relatively small, on the order of one percentage point per week and per month.

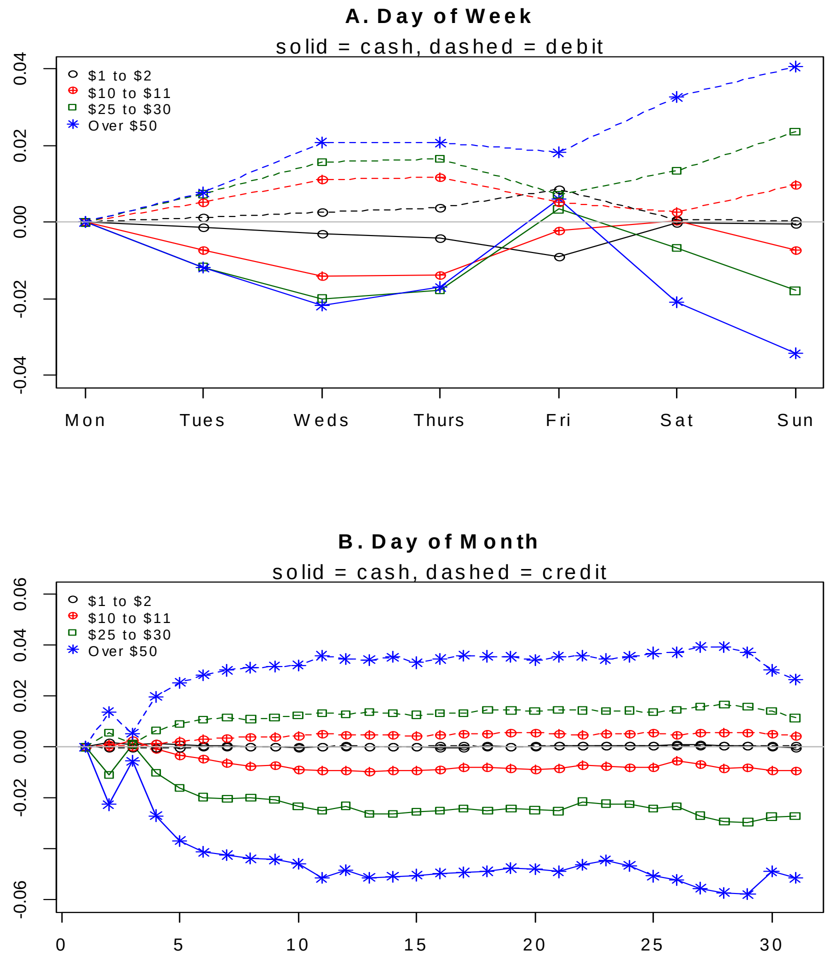


Figure 9. Daily marginal effects by transaction size.

Figures 9A and 9B display estimated weekly and monthly cycles for four different transaction size regressions included in Section 3.3. To conserve on space, and because debit (credit) exhibits a greater cycle over the week (month), Figure 9A includes only cash and debit, and Figure 9B includes only cash and credit. In both figures, the patterns from the overall regression generally appear again in the transaction size regressions, and those patterns are amplified as transaction size increases. The only exception is that in Figure 9A, the smallest transactions display somewhat different payment behavior than other transaction sizes from Friday to Sunday.

5.2 Interpreting Weekly and Monthly Patterns

A potential explanation for the weekly and monthly patterns is that consumers' financial positions vary systematically over the week and the month, because of factors such as paydays, regular bill payments, and cash withdrawals. These fluctuations may affect payment choices, especially for consumers who are financially constrained. If that is true, we would expect to find supporting evidence in the behavior of transaction volumes over the week and month. For example, financially-constrained consumers would drop out of the pool of shoppers as time passed since their last payday. If paydays are correlated across consumers then the cash share would decrease because of systematically lower cash thresholds for the consumers who stay in the pool.

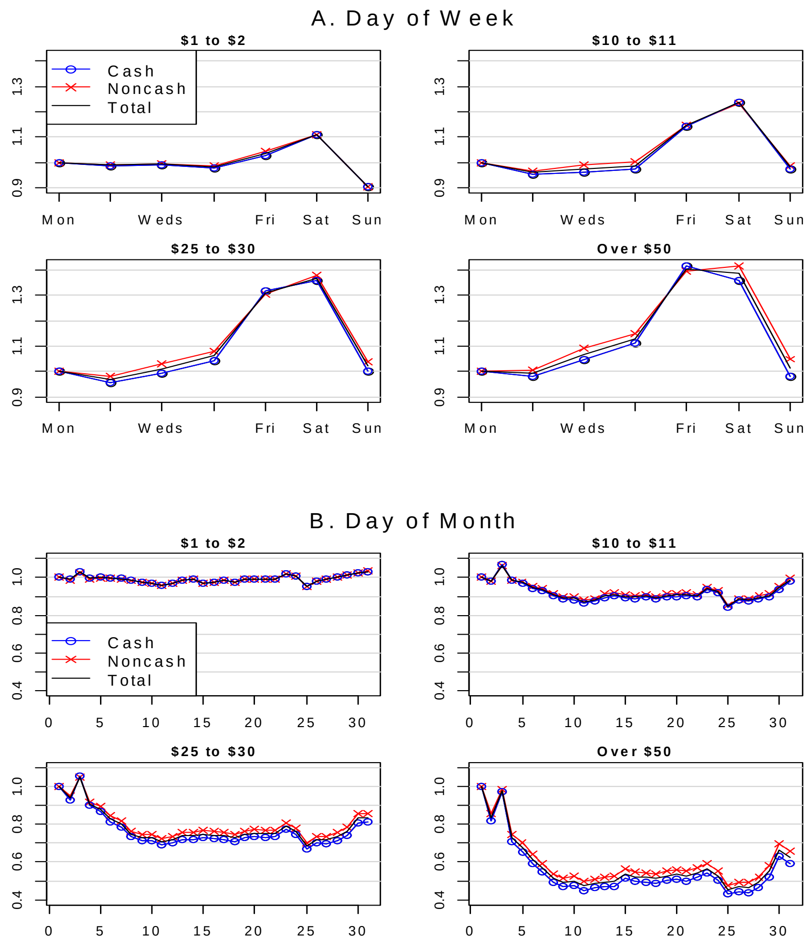


Figure 10. Daily marginal effects for number of transactions.

To explore these conjectures, we analyze the time patterns of transaction volumes analogously to the time patterns in payment shares: We regress the log number of transactions in each of the 22 transaction size bins for each zip-code day on the same explanatory variables used in the above regressions of payment shares. The regressions are conducted by OLS, so the estimated coefficients equal the marginal effects. Figures 10A and 10B plot the estimated day-of-week and day-of-month dummy coefficients from the regressions (black line without symbols) together with the implied patterns of number of cash and non-cash transactions (blue and red lines with circles). The latter are calculated by combining the estimated day-of-week and day-of-month effects from the log transactions regressions with the estimated payment fraction effects in Figures 9A and 9B. Essentially, these figures plot the predicted number of total transactions as well as cash and non-cash transactions over the week and the month, with the first day normalized to 1.

There are several striking features of Figures 10A and 10B. First, the variation in number of transactions is enormous relative to the variation in payment shares; and, as with payment shares the variation tends to be amplified for larger transaction sizes. Over the week, for \$1-\$2 transactions volume varies by roughly 20 percent with the peak being Friday and Saturday, and for transactions over \$50 this variation is almost 40 percent. Over the month, transaction volume is even more sensitive to transaction size: For \$1-\$2 transactions there is less than a ten percent difference between the highest and lowest transaction volume, whereas for transactions over \$50 the difference is more than 50 percent. Second, the broad pattern of transactions across the week and month is shared by both cash and non-cash transactions. This must be the case, given that the variation in number of total transactions is much larger than the variation in cash fractions. Finally, and crucially for our hypothesis about the role of financial position, the monthly pattern of transactions volume looks quite similar to the pattern of cash fractions (comparing Figures 9B and 10B). This similarity is also present, though not as strong, in the weekly pattern (Figures 9A and 10A). Table 3 quantifies these relationships; it displays correlations for each size class between the cash marginal effects from Figures 9A and 9B and the volume marginal effects from Figures 10A and 10B. Apart from transactions below \$3 and, over the week, transactions above \$45, all the correlations are above 0.4. Over the week most of the correlations are above 0.6, and over the month most are above 0.75.

We interpret the strong correlation between cash fractions and transaction volumes over the week and month as indicating that financial position is likely driving both transaction volumes and the share of cash transactions. For financially-constrained consumers, regular financial fluctuations (e.g., paychecks, rent payments) may lead to shopping being concentrated in certain parts of the week or month. To the extent that those consumers tend to have high cash thresholds, as they exit the pool of shoppers the fraction of cash payments will decline (the distribution of cash thresholds will shift to the left), yielding a positive correlation between volume and cash fraction.[16] For some other consumers low cash inventory restricts the size of their cash purchases, but without necessarily affecting their shopping behavior; we refer to such consumers as "cash-constrained": Their cash threshold may fall as their cash balances decline over the week and month, reinforcing the positive correlation between the cash share and the number of transactions.

Table 3. Correlations: Cash Shares and Transaction Volumes

Transaction Size	Days of Week	Days of Month
\$1 to \$2	-0.13	0.20
\$2 to \$3	0.29	0.31
\$3 to \$4	0.47	0.53
\$4 to \$5	0.54	0.55
\$5 to \$6	0.62	0.69
\$6 to \$7	0.63	0.65
\$7 to \$8	0.62	0.74
\$8 to \$9	0.65	0.73
\$9 to \$10	0.63	0.78
\$10 to \$11	0.66	0.77
\$11 to \$12	0.65	0.75
\$12 to \$13	0.66	0.80
\$13 to \$14	0.65	0.86
\$14 to \$15	0.62	0.82
\$15 to \$20	0.62	0.85
\$20 to \$25	0.54	0.83
\$25 to \$30	0.52	0.87
\$30 to \$35	0.51	0.88
\$35 to \$40	0.47	0.91
\$40 to \$45	0.41	0.93
\$45 to \$50	0.37	0.91
Over \$50	0.31	0.93

The correlations between cash fractions and transaction volumes are lower for weekly patterns than for monthly patterns, and the explanation likely has to do with "convenience": Many consumers have a regular leisure-time cycle to their week associated with work, school, or other activities. Although convenience undoubtedly plays some role in payment patterns–to the extent that it affects ATM visits, for example–our view is that convenience is a more important determinant of the weekly transactions pattern, and thus works against a high weekly correlation between the cash share and the number of transactions.

Our key findings in this section are that (i) the fraction of cash payments declines over the month and also has a regular weekly pattern; (ii) the monthly and weekly patterns are stronger for larger transaction sizes; and (iii) similar patterns in the number of transactions suggest that financial

[16] One notable feature of the monthly pattern is a transitory reversal of the broad trends in both cash share and number of transactions on the 3rd day of the month. In fact, many recipients of Social Security and Supplemental Security Income are usually paid on the 3rd of the month. Our rough estimates suggest that more than one million individuals in the U.S. fall into this category.

constraints and a changing pool of shoppers may play a large role in (i) and (ii). Further progress in understanding these findings requires advances in both measurement and theory. Regarding measurement, the first step would be to disentangle changes over time in the composition of shoppers from changes over time in the payment behavior of individual shoppers. Large marketing datasets such as Kilts-Nielsen that contain information on both consumers and retail transactions can help answer this question, by revealing how the characteristics of shoppers change over the week and month.

Regarding theory, our findings point in two directions. The first is to model changing individual payment choices over the week and month by incorporating multiple means of payment into the cash inventory framework set out by Baumol (1952) and Tobin (1956); consumers in these models are cash constrained in the sense we described above. Along these lines, the recent work of Alvarez and Lippi (2015) explicitly models the dynamic interaction between cash holding and payment choice. However, transaction size is fixed in their model. Incorporating variable transaction size in an inventory-theoretic model with a regular pattern of cash injections (i.e., paydays) would shed light on the extent to which time patterns in individual payment behavior can explain the time patterns in aggregated payment shares that we document.

Taking a broader perspective, our findings also point to the value of jointly modeling households' overall financial position (as well as their cash inventory), spending pattern, and payment choice. The comovement between time patterns of payment mix and transaction volumes that we document suggests that consumers' binding financial constraints play a potentially important role. A model with financial frictions, cash inventory considerations, and payment choice would be the ideal framework for exploring the joint determination of consumers' spending and payment patterns over time. In turn, parameterizing such a model would require data on households' overall financial positions, spending patterns, and payments behavior.

6 Payment Variation over the Longer Term

We now turn to payment variation over the longer term, which is captured by the month-of-sample dummies in our regressions. We first summarize the findings regarding seasonal cycles and long-run trends, and then use the findings to consider the future of currency use at this retailer.

6.1 Seasonal Cycles and Long-run Trends

Figure 11A plots the marginal effects for month-of-sample dummies from the overall regression in Section 3.2. These effects combine seasonality with a time trend and idiosyncratic monthly variation. The vertical lines lie between March and April, and thus divide our sample into three 12-month periods. Comparing these periods, both the seasonality and trend are striking. In terms of seasonality, the figure clearly shows a rise in cash mirrored by a fall in debit in December, the holiday season. Also, the credit fraction dips after the holiday season. In terms of long-run trends, cash and check fractions decline, while debit and credit rise.

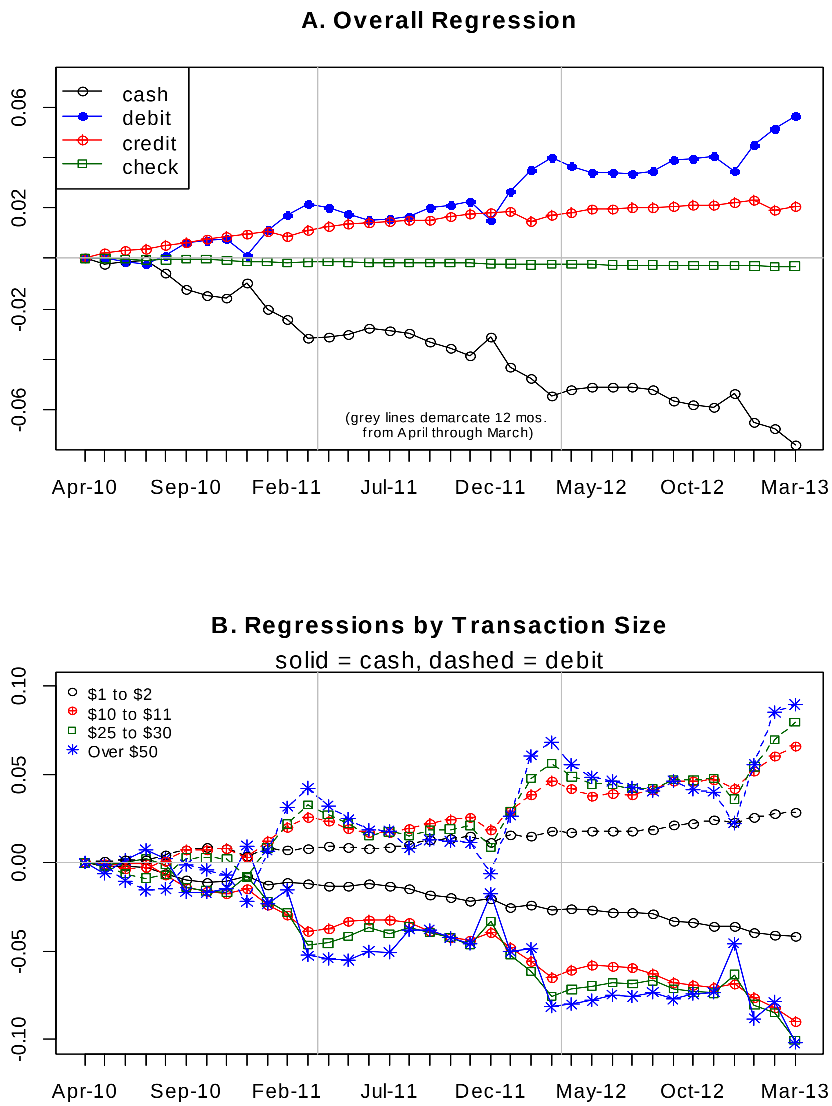


Figure 11. Month-of-sample marginal effects.

The patterns are similar as we turn to the transaction size class regressions. Figure 11B plots the month-of-sample marginal effects by transaction size for cash and debit. They basically replicate the pattern that we found in the overall regression, but the marginal effects are amplified with transaction size.

With the estimated month-of-sample effects from the transaction size-class regressions, we can discuss how the model-predicted payment mix varies from the beginning to the end of the sample period, as well as the implied time trends. Figure 12 compares the predicted payment mix at the mean values of the explanatory variables for the first and last months of our sample; the lines marked with x's represent April 2010, and the lines marked with o's represent March 2013. For each transaction size, the x's and the o's are from the same set of regressions, simply evaluated at different values of the time dummies. In contrast, the different transaction sizes represent different regressions. There is a marked downward shift in the predicted cash and check fractions, and corresponding upward shifts in the predicted debit and credit fractions. The magnitude of the shift is generally increasing in transaction size.

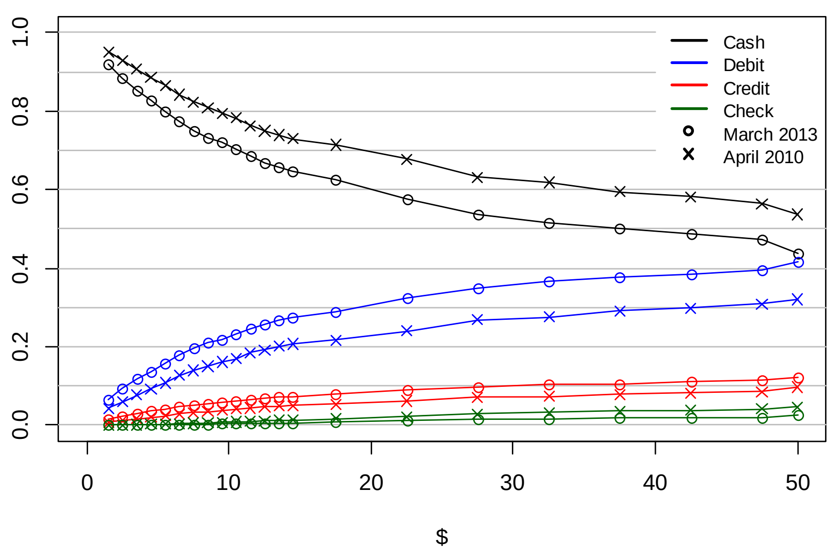


Figure 12. Predicted payment fractions by transaction size.

We then estimate linear time trends for each transaction size within each payment type. For cash, the time trends range from a decrease of 1.3 percentage points per year for $1-$2 transactions to a decrease of 3.32 percentage points per year for $20-$25 transactions; for debit, the trends range from an increase of less than 1 percentage point per year for $1-$2 transactions to an increase of

2.6 percentage points per year for transactions greater than \$50. The increase in credit is smaller, ranging from 0.45 to 1.13 percentage points per year across transaction sizes. Combining these estimated trends with the transaction size distribution, our model implies that the cash share of transactions declined by 2.46 percentage points per year at this discount retailer, with most of the decline in cash offset by an increase in debit.

6.2 The Decline in the Cash Share

Our findings show a trend decline in the share of cash transactions at this large discount retailer. Exploring the driving forces behind the trend would be useful for understanding the changing demand for currency in retail transactions more broadly. We evaluate several candidate factors below.

First, because the trend is estimated based on three years of data, we are naturally concerned that the decline in cash may be driven by some transitory factors. One candidate is the Great Recession. According to the Boston Fed's latest report on consumer payments (Schuh and Stavins, 2014), cash payments increased significantly after the financial crisis, replacing credit payments. Therefore, as the economy recovered from the recession, we may expect credit to have risen at the expense of cash. However, this is unlikely to explain the pattern in our sample where most of the cash decline was replaced by an increase in debit. As Figure 1 shows, credit accounts for only a very small fraction of transactions, roughly 3 percent at the beginning of the sample period and 5 percent at the end. And note that even 5 percent is an overestimate because our measure of credit includes signature debit and prepaid cards.

Another possibility is changes in the store's payment acceptance policy. However, as far as we know, there was a uniform payment policy in place across all the chain's stores during the sample period, with cash, debit, credit and checks accepted on equal terms. Still, because our sample covers the implementation of the Durbin regulation on debit card interchange fees (effective on October 1, 2011), one may wonder if the chain had an incentive to steer customers toward more debit use. Again, this was unlikely. The Durbin regulation established a 21-cent cap on the debit interchange fees that financial institutions with more than \$10 billion in assets can charge to merchants through merchant acquirers. However, we learned from the retailer that more than 50 percent of its debit transactions were exempt from the regulation because the debit cards used were issued by financial institutions under \$10 billion in assets. Moreover, the Durbin regulation is known for its unintended consequence of raising interchange fees to 21 cents for small-dollar transactions, which account for the vast majority of transactions at this retailer (Wang, 2016). Therefore, if the new regulation were to have any impact on the stores in our sample, it should have caused them to try to reduce debit use rather than promote it.

One may also wonder whether the store expanded or altered the range of retail goods it sold during the sample period so that it attracted a different clientele that had different payment preferences. While we cannot fully rule out this possibility without observing individual customers, the company's annual financial reports indicate that the composition of goods sold did not undergo major changes during the period.

Given that the transitory factors discussed above are unlikely to explain the decline in the cash share at this retail chain, attention turns to longer-term economic and demographic factors. We first forecast the retailer's payments mix in the next few years based on our estimated month-of-sample dummies, and then quantify how much of the implied cash decline can be explained by forecasted changes in zip-code-level variables, especially the age-cohort effect, included in our analysis.

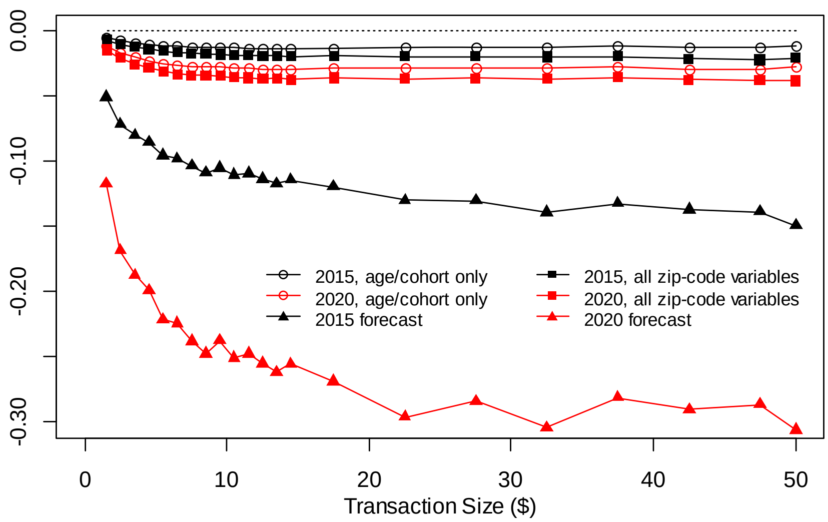
The forecasting exercise begins with the retailer's predicted payment mix for March 2013, evaluated at the means of the zip-code-level explanatory variables. A time trend is then incorporated by assuming the payment mix will change each year at an exogenous rate implied by the marginal effects associated with our estimated coefficients on month-of-sample dummies. The forecasted declines in cash fractions by transaction size in 2015 and 2020 compared with 2011 are the lines marked with triangles in Figure 13.

Part of the time trend is presumably attributable to the change in zip-code-level variables. Recall that all zip-code-level variables were treated as fixed at their 2011 values across time in the regressions. Therefore any time trend is picked up by the month-of-sample dummies, even if some of the trend is actually associated with time variation in the zip-code-level variables. Using forecasted changes for the zip-code-level variables, it is possible to estimate those variables' contributions to the change in forecasted cash fractions.[17]

The "contributions" of demographic and other zip-code-level changes to our forecasts of cash use are displayed in Figure 13. Each of the lines denoted by a square or a circle plots the difference between (i) a forecasted cash fraction that is based on a particular forecasted change in zip-code-level variables, and (ii), the estimated cash fractions for March 2011. There are two main messages from Figure 13. First, a majority of the decline in cash shares that we can attribute to changes

[17] Wherever possible, forecasts for the zip-code-level variables are the United States Census Department projections. The online Appendix describes the forecasts for all the zip-code-level variables.

in zip-code-level variables is due to the cohort effect: For 2015, between 53% and 75% of the zip-code-level effects represent cohort effects, across transaction sizes, and for 2020 these numbers rise to 71% and 79%, respectively. Second, while the cohort effects are important relative to other zip-code-level effects, the overall effects of zip-code-level variables are small relative to the time trends: For our 2020 forecasts, the effects of forecasted changes in zip-code-level variables represent between 11.6% and 15.2% of the changes in cash shares implied by our estimated time trends.



Note: Lines represent differences between forecasts based on indicated factors and estimated cash use fractions for March 2011.

Figure 13. Zip-code-level Variables and Forecasts of Cash Fractions.

Therefore, a large fraction of the time trend is still left to explain. Prime candidates are technological progress and changing consumer perceptions of the attributes of debit payments relative to others. These attributes include but are not limited to adoption costs, marginal cost of transactions, speed of transactions, security, record keeping, general merchant acceptance, and ease of use, which are not directly included in our regressions. In fact, the Survey of Consumer Payment Choice shows that there has been continuous improvement in consumers' perceptions of debit card security relative to cash in recent years.[18]

While our above exercise is informative for understanding future cash use in this retailer, we need to be cautious about the forecast per se or applying it to the overall economy. First of all, our reduced-form empirical model does not recover structural parameters. Second, our transaction data

[18] The Survey of Consumer Payment Choice is a longitudinal panel survey conducted by the Federal Reserve Bank of Boston every year since 2008, covering approximately 2,000 consumers.

do not keep track of individual consumers and their payment choices. Third, it is fundamentally difficult to forecast aggregate behavior from one firm given the selective nature of the sample. With that said, our exercise nevertheless highlights several factors that may be relevant for currency demand in the future, in particular the rise of debit and the role played by demographic, economic and technological factors: These features are likely to be shared in the broader retail sector. In fact, debit has seen tremendous overall growth in the past decade. According to the latest Federal Reserve Payments Study (2014), it has risen to be the top non-cash payment instrument in the U.S. economy: Debit accounted for 19 percent of all non-cash transactions in 2003 and its share doubled by 2012. Our study provides first-hand micro evidence that the increase in debit came at the expense of cash at a large cash-intensive retailer. In terms of the threshold framework, the increasing replacement of cash by debit represents a downward shift in the cash thresholds for shoppers at this particular retailer. Assuming that the shift from cash to debit is also occurring in retail more generally, and that it continues, it could eventually be manifested in a decline in currency in circulation.

7 Conclusion

Using three years of transactions data from a discount retailer with thousands of stores, we study payment variation across transaction sizes and locations; at weekly and monthly frequencies; and over longer time horizons. In each case, we identify interesting empirical patterns and connect them to a theoretical framework in which individual consumers choose between cash and non-cash payments based on a threshold transaction size, with the distribution of thresholds depending on the relative costs of using cash, which may vary by consumer, location, and time.

In terms of payment variation across locations, our empirical model provides support for the threshold framework by finding a negative relationship between the cash share of payments and location-specific variables that proxy for the relative costs of using cash. With respect to transaction size, we emphasize two salient features: First, the fraction of cash payments declines with transaction size at a given location, and second, the dispersion of the payment mix across locations increases with transaction size. The first feature is implied by the threshold framework, whereas the second feature is a new empirical finding that would require incorporating cross-location heterogeneity into a threshold model. A quantitative decomposition reveals that the threshold distributions cannot be explained by transaction-size fixed effects. They are instead primarily determined by location-specific characteristics, which in part proxy for the degree of access to non-cash payments.

We find weekly and monthly fluctuations in cash shares and associate them with weekly and monthly fluctuations in the distribution of shoppers' cash thresholds. We analyze the time series for number of transactions in our data and find similar patterns to those in cash shares. The results suggest that some consumers in our sample are subject to time-varying financial constraints through the week and month, which affect both their shopping patterns and the distribution of cash thresholds.

Finally, over the longer run, we identify a declining share of cash transactions, largely replaced by debit, at this retailer. The decline in cash shares (in other words, the leftward shift of the threshold distribution) was likely not driven by transitory factors, and only a relatively small fraction can potentially be explained by changes in the zip-code-level variables, including age-cohort composition. Technological progress and changes in consumer perceptions are possible explanations for the rise of debit at the expense of cash. Assuming that shift is also occurring in retail more generally, and that it continues, it can be expected to have an aggregate impact on the transaction demand for currency.

While our analysis sheds new light on empirical patterns of payment choice and currency use, the approach we take is nevertheless reduced-form. A challenge for future research is to construct formal structural models, in order both to provide a deeper understanding of the empirical patterns and to answer policy questions.

Focusing on the cross-sectional dispersion in payment shares, a first step in advancing theory could be to introduce heterogeneity within and across locations in a threshold model such as Whitesell (1989) or Lucas and Nicolini (2015). In addition, our findings regarding high-frequency time-series patterns point toward extending explicit inventory theoretic models as in Alvarez and Lippi (2015) in the manner discussed at the end of Section 5. Key elements of such an extension would involve variable transaction size, regularly scheduled cash or liquidity injections, and financial constraints. Ultimately, of course, one would want to address both the cross-section and time-series elements in a unified model, and to include an adoption decision for alternative means of payment. Such a model would also be well-suited for studying the longer-term shift from cash to debit seen in our data.

With an explicit model, one could better address important questions related to both monetary policy and payments policy. Given that households' cash thresholds are endogenously determined, what is the welfare cost of inflation, and what would be the optimal rate of inflation (See Wang, Wright and Liu, 2016 for a discussion)? What would be the costs and benefits of eliminating physical currency, as has been proposed by Rogoff (2015), for example, partly as a way of eliminating the lower bound on nominal interest rates? What are the welfare consequences of limits on debit- and credit-card processing fees when the adoption and usage margins are both important (e.g., Koulayev et al., 2016)? These are some prominent examples of policy questions that are underscored by our empirical analysis, and that further advances in theory can help to address.

References

[1] Alvarez, F.E., Lippi, F., 2015. Cash Burns: an Inventory Theoretic Model with a Cash-credit Choice. NBER Working Paper 21110.

[2] Arango, C., Huynh, K. P., Sabetti, L., 2011. How Do You Pay? The Role of Incentives at the Point-of-Sale. Working Paper 2011-23, Bank of Canada.

[3] Baumol, W.J., 1952. The Transactions Demand for Cash: An Inventory Theoretic Approach. Q. J. Econ. 66(4), 545–556.

[4] Borzekowski, R., Kiser, E. K., Ahmed, S., 2008. Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response. J. Money Credit Bank. 40(1), 149-172.

[5] Borzekowski, R., Kiser, E. K., 2008. The Choice at the Checkout: Quantifying Demand across Payment Instruments. Int. J. Ind. Organ. 26(4), 889–902.

[6] Briglevics, T., Schuh, S., 2014. This Is What's in Your Wallet... and Here's How You Use It. Federal Reserve Bank of Boston Working Paper 14-5.

[7] Buis, M. L., 2008. FMLogit: Stata module fitting a fractional multinomial logit model by quasi maximum likelihood. Statistical Software Components, Department of Economics, Boston College.

[8] Ching, A. T., Hayashi, F., 2010. Payment Card Rewards Programs and Consumer Payment Choice. J. Bank. Finance 34(8), 1773-1787.

[9] Cohen, M., Rysman, M., 2012. Payment Choice with Consumer Panel Data. Memo, Department of Economics, Boston University.

[10] Federal Reserve System, 2014. The 2013 Federal Reserve Payments Study. https://www.frbservices.org/communications/payment_system_research.html.

[11] Freeman, S., Kydland, F.E., 2000. Monetary Aggregates and Output. Am. Econ. Rev. 90(5), 1125-1135.

[12] Gilbert, A., Zaretsky, A., 2003. Banking Antitrust: Are the Assumptions Still Valid? Federal Reserve Bank of St. Louis Review 85(6), 29-53.

[13] Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo Maximum Likelihood Methods: Theory. Econometrica 52, 681-700.

[14] Klee, E., 2008. How People Pay: Evidence from Grocery Store Data. J. Monet. Econ. 55, 526-541.

[15] Koulayev, S., Rysman, M., Schuh, S., Stavins, J., 2016. Explaining Adoption and Use of Payment Instruments by U.S. Consumers. RAND J. Econ. 47(2), 293-325.

[16] Lucas, R.E. Jr., Nicolini, J.P., 2015. On the Stability of Money Demand. J. Monet. Econ. 73, 48-65.

[17] McAndrews, J., Wang, Z., 2012. The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage. Research Working Paper 12-06, Federal Reserve Bank of Richmond.

[18] Mullahy, J., 2010. Multivariate Fractional Regression Estimation of Econometric Share Models. NBER Working Papers 16354.

[19] Papke, L. E., Wooldridge, J.M., 1996. Econometric Methods for Fractional Response Variables with An Application to 401(k) Plan Participation Rates. J. Appl. Econometrics 11(6), 619-32.

[20] Prescott, E. C., 1987. A Multiple Means of Payment Model, in: Barnett, W.A., Singleton, K.N. (Eds.), New Approaches to Monetary Economics, Cambridge University Press: 42-45.

[21] Rogoff, K., 2015. Costs and Benefits to Phasing out Paper Currency. NBER Macroeconomics Annual 2014, 29, 445-456.

[22] Schuh, S., Stavins, J., 2014. The 2011 and 2012 Surveys of Consumer Payment Choice. Federal Reserve Bank of Boston Research Data Report No. 14-1.

[23] Schuh, S., Stavins, J., 2012. How Consumers Pay: Adoption and Use of Payments. Federal Reserve Bank of Boston Working Paper 12-2.

[24] Stavins, J., 2013. Security of Retail Payments: The New Strategic Objective. Federal Reserve Bank of Boston Public Policy Discussion Paper No. 13-9.

[25] Tobin, J., 1956. The Interest Elasticity of the Transactions Demand for Cash. Rev. Econ. Stat. 38 (3), 241–247.

[26] Wakamori, N., Welte, A., forthcoming. Why Do Shoppers Use Cash? Evidence from Shopping Diary Data. J. Money Credit Bank.

[27] Wang, Z., 2016. Price Cap Regulation in a Two-sided Market: Intended and Unintended Consequences. Int. J. Ind. Organ. 45, 28-37.

[28] Wang, L., Wright, R., Liu, L., 2016. Money and Credit: Theory and Applications. Working paper.

[29] Whitesell, W., 1989. The Demand for Currency versus Debitable Account. J. Money Credit Bank. 21(2), 246-251.

[30] Zinman, J., 2009. Debit or Credit? J. Bank. Finance 33(2), 358–366.