

Consumer Payment Choice in the Fifth District: Learning from a Retail Chain

Zhu Wang and Alexander L. Wolman

The U.S. payments system has undergone fundamental changes over the past several decades. Perhaps the most significant trend is the shift from paper payment instruments, namely cash and check, to electronic ones such as credit and debit cards. Understanding this shift is important, as it affects billions of transactions worth trillions of dollars each year.¹ For many years, experts on payments systems have forecast the arrival of a completely electronic, paperless payments system, but it has not yet happened. Cash and check still play a large role in the economy, particularly in some sectors.

In this context, a sizable body of empirical literature has developed to study consumer payment choice. Most of the studies rely on data from consumer surveys.² While this research has improved our understanding of how consumers choose to pay, consumer survey data have their limitations, including small sample size and imperfect reporting.

Our paper reports and analyzes new evidence on consumer payment choice in retail transactions, including the use of cash, credit card, debit card, and check, based on a comprehensive dataset comprising

■ We thank Joseph Johnson for excellent research assistance. The views expressed in this article are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System. All errors are our own.

DOI: <http://doi.org/10.21144/eq1020102>

¹ According to the latest Federal Reserve Payments Study (2014), the estimated number of noncash payments alone, excluding wire transfers, was 122.8 billion in 2012, with a value of \$79.0 trillion.

² For example, Borzekowski et al. (2008), Borzekowski and Kiser (2008), Zinman (2009), Ching and Hayashi (2010), Arango et al. (2011), Cohen and Rysman (2012), Schuh and Stavins (2012), and Koulayev et al. (2016).

merchant transaction records. The data, provided by a large retail chain, cover every transaction in each of its stores over the five-year period from April 2010 through March 2015. We focus on hundreds of stores located across the Fifth Federal Reserve District. The purpose of our study is to provide a better understanding of payment variation for retail transactions in this region.³

Our study has several important findings. First, the fraction of cash transactions decreases in transaction size and is affected by location-specific variables that reflect consumers' preferences and the opportunity costs of using cash relative to noncash means of payments.

Second, based on the estimation results, we evaluate the relative importance of different groups of variables in explaining the payment variation across locations in our sample. We find that median transaction size, demographics, education levels, and state fixed effects are the top factors related to consumer payment choice. Taking these into consideration, we project the payment variation across the entire Fifth District for retail outlets similar to those in our sample.

Finally, we identify interesting time patterns of payment variation. In particular, the shares of cash and check transactions decline steadily over our five-year sample period, 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, largely replaced by debit. We show that the decline in cash at this particular retailer was likely not driven by transitory factors, and only a relatively small fraction could be explained by changes in median transaction size and zip-code-level variables. 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.

The structure of the paper is as follows. Section 1 describes the data used in our analysis as well as the empirical approach. Section 2 introduces the regression model and presents an overview of the estimation results. Section 3 evaluates the relative importance of different variables in explaining payment variation across locations in our sample, and projects payment variation across the entire Fifth District. Section 4 discusses the longer-run decline of cash. Finally, Section 5 concludes.

³ See Wang and Wolman (2016) for a study covering the entire chain's thousands of stores across the country between April 2010 and March 2013. That study mainly explores payment variation across transaction sizes and time frequencies. In contrast, this paper focuses on decomposing the relative importance of different local variables and projecting cross-sectional payment patterns in the Fifth District.

1. DATA AND EMPIRICAL APPROACH

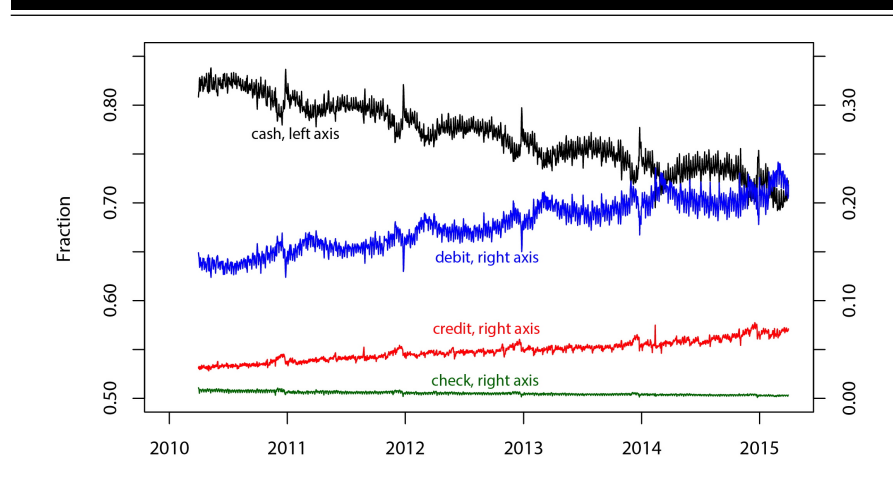
The transactions data used in our study is from a large discount retailer with hundreds of stores across the Fifth Federal Reserve District, which covers Maryland, North Carolina, South Carolina, Virginia, Washington, DC, and West Virginia. The stores sell a wide variety of goods in various price ranges, 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, 2015. For each transaction, the data include means of payment, time, location, and amount. 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.⁴ The retailer also provides cash-back services, and the purchase components of cash-back transactions are included in our analysis. In contrast, transactions made with special-purpose means of payment such as electronic benefit transfer (EBT), coupons, and store return cards are excluded. All told, our analysis covers 86 percent of the total transactions in the sample period. Our summary of the data in this section will refer to all stores located in the Fifth District; 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 retail outlets from that analysis because the zip-code-level data are unavailable.⁵

Payment Variation

The purpose of our paper is to explain payment variation across locations and time in the Fifth District. Figure 1 presents payment variation across time in our sample. The data are plotted at the daily level, displaying the fraction of all the transactions accounted for by each payment type. Note that while cash is measured on the left axis, and debit, credit, and check are all measured on the right 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,

⁴ Data limitations prevent us from distinguishing credit cards from signature debit and prepaid cards. However, our estimates reveal 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 consumers' account balances rather than borrowed funds, we can reasonably assume the estimated patterns are primarily driven by the true credit cards.

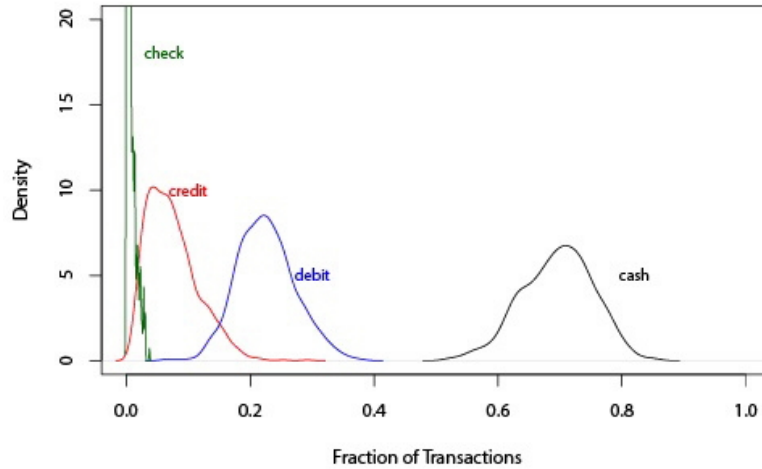
⁵ We omit Washington, DC, from the regression analysis due to lack of zip-code-level crime data.

Figure 1 Payment Variation Across Time

and check. Over the sample period, the fractions of cash and check are trending down, with debit and credit trending up. There are higher frequency patterns as well, with cash and debit again moving in opposite directions. We will allow for these patterns in the econometric model by including day-of-week, day-of-month, and month-of-sample dummies.

Figure 2 presents payment variation across locations, restricting attention to the last full month of the sample, March 2015. We aggregate the data by zip code and display smoothed estimates of the density functions for the fraction of transactions conducted with cash, debit, credit, and check.⁶ We use only one month because of the time trend evident in Figure 1. The ranking from Figure 1 is also apparent in Figure 2: cash is the dominant form of payment, followed by debit, credit, and check. Moreover, Figure 2 shows significant variation across zip-code locations in cash, debit, and credit use. This variation highlights the need for including location-specific variables in our econometric model.

⁶ Note that the estimated kernel density for checks is truncated in Figure 2. 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 2 Payment Variation Across Locations

Explanatory Variables

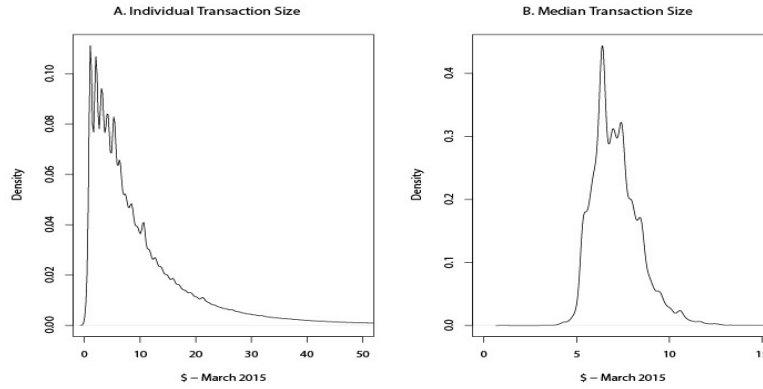
The payment variation identified in Figures 1 and 2 suggests the quantitative importance of including location- and time-specific variables in an econometric model of payment choice. In Wang and Wolman (2016), we discuss how theories of money demand and payment choice motivate the choice of particular variables. Here, we simply list the variables we use and explain informally why they may be associated with variation in the shares of different payment types across locations and time. Note that our data identify transactions but not customers, so we treat the characteristics of the zip code in which each store is located as representative of the characteristics of the store's customers and the economic environment in which they live. Table 1 lists summary statistics for the zip-code-level explanatory variables used in the regressions, fixed at their 2011 values.⁷

⁷ Most of our zip-code-level variables come from the U.S. Census' American Community Survey (ACS) and the FDIC's Summary of Deposits. The robbery data are from the FBI's Uniform Crime Report. We fix zip-code-level explanatory variables at their 2011 values (five-year estimates), because the ACS provides only five-year estimates for areas with fewer than 20,000 residents. In Section 4, where we study longer-run payment variation, we will discuss the effects of time variation in zip-code-level explanatory variables.

Table 1 Summary Statistics of Zip-Code Variables

Variable (unit)	Mean	Std. dev.	1%	99%
Banking condition				
HHI metro	0.211	0.172	0.070	0.735
HHI rural	0.273	0.110	0.125	0.561
Branches per capita ($1/10^3$)	0.44	0.44	0.06	2.59
Socioeconomic condition				
Robbery rate ($1/10^5$)	29.76	40.31	0.00	235.07
Median household income (\$)	41015	12666	22214	90078
Population density (per mile ²)	579	1024	20	5017
Family households (%)	67.06	5.84	43.44	79.05
Housing (%): Renter occupied	27.73	9.25	10.49	54.56
Owner occupied	59.08	9.85	28.05	80.18
Vacant	13.20	7.80	4.38	43.89
Demographics (%)				
Female	51.08	2.37	40.35	55.01
Age				
<15	18.74	2.79	11.20	24.80
15-34	25.27	5.06	16.25	46.17
35-54	27.08	2.45	19.43	32.43
55-69	18.69	3.49	10.65	29.08
>70	10.23	3.04	4.02	19.12
Race				
White	68.23	21.99	8.66	98.86
Black	24.22	20.16	0.25	82.09
Hispanic	6.20	6.34	0.36	32.16
Native	1.11	5.24	0.05	26.93
Asian	1.24	1.88	0.04	8.17
Pacific Islander	0.05	0.07	0.00	0.35
Other	3.22	3.93	0.04	19.38
Multiple	1.93	0.94	0.42	5.32
Education level (%)				
Below high school	19.26	6.90	5.70	36.90
High school	33.85	7.11	15.90	51.70
Some college	20.28	3.88	11.50	29.20
College	26.61	9.97	10.70	56.10

(I) Median Transaction Size We use the median transaction size for each zip-code day to capture the transaction size distribution. The theory outlined in Wang and Wolman (2016) suggests that higher transaction sizes will be associated, all else equal, with less cash use. Figure 3 provides information about the size distribution of transactions in March 2015 without regard to means of payment. Figure 3A displays a smoothed density function, by transaction size, for all transactions in the month. Figure 3B plots the distribution of median transaction sizes across zip-code days. Figure 3B complements Figure 2 in showing that there is substantial heterogeneity across locations with respect to size of transaction, as well as payment mix.

Figure 3 Kernel Densities of Transaction Size in March 2015

(II) Banking Variables Local banking condition matters for payment choice, but the effects are subtle. Cash use may be expected to decrease in banking-sector competition (which results in lower banking fees and/or better deposit terms that 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 Herfindahl-Hirschman Index (HHI) in each Metropolitan Statistical Area (MSA) or rural county.⁸ Bank branches per capita are measured at the zip-code level.

(III) Socioeconomic Variables We include the robbery rate, median household income, population density, fraction of family households, and fraction of homeownership as socioeconomic variables. The robbery rate is measured at the county level while other variables are measured at the zip-code level.

A higher robbery rate increases the cost of holding cash, which we would expect to reduce cash use. The other variables are likely to correlate with consumers' access to bank accounts or ownership of credit or debit cards. Note that population density is relevant for adoption

⁸ 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 an MSA in urban areas and by a county in rural areas. The most commonly used measure of market concentration is the HHI, calculated by squaring each bank's share of deposits in a market and then summing these squared shares.

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 adoption and usage of electronic payment instruments should be higher in areas with a high population density or more business activity.

(IV) Demographic Variables Gender, age/cohort group, and race are included to reflect the fact that payment behavior may vary systematically with demographic characteristics. These variables are each measured as a fraction of the population at the zip-code level.

(V) Education Variables We specify four education levels: below high school, high school, some college, and college and above. Higher education is often associated with better financial literacy and higher opportunity time cost of using cash, so it may be associated with a higher adoption and usage of noncash payments. The education variables are each measured as a fraction of the population at the zip-code level.

(VI) and (VII) State and Time Dummies We also include state dummies as well as day-of-week, day-of-month, and month-of-sample dummies.

2. ESTIMATION RESULTS

We turn now to an empirical model aimed at explaining the variation in payment shares through the behavior of the explanatory variables. The data are analyzed using the fractional multinomial logit model (FMLogit).⁹ 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, 2015.¹⁰ The explanatory variables are those introduced above.¹¹

⁹ The FMLogit model addresses the multiple fractional nature of the dependent variables, namely that the fraction of payments for each instrument should remain between zero and one, and the fractions add up to one. More details of the FMLogit model are provided in the Appendix.

¹⁰ In our sample, most 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 stores. In the latter case, we simply sum up the transactions of all the stores in the zip code.

¹¹ Note that the local characteristics data are from a single year, 2011, while the dependent variables and the median-transaction-size variable come from multiple years,

Table 2 reports the estimation results, expressed in terms of marginal effects.¹² We summarize the findings as follows.

(I) Median Transaction Size Aggregating transactions within a zip-code day, we expect to find that a rightward shift in the size distribution of transactions corresponds to a lower share of cash transactions, as consumers are less likely to use cash for larger 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, \$6.65, 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 1.3 percentage points, credit by 0.4 percentage points, and check by 0.1 percentage points.

(II) Banking Variables We find that higher banking concentration corresponds to a higher cash share (lower card shares) in rural areas. However, higher concentration corresponds to 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.¹³ In contrast, more bank branches per capita are associated with a higher cash share, mainly at the expense of debit and credit. These findings are consistent with our discussion in Section 1.

(III) Socioeconomic Variables As expected, a higher robbery rate is found to be associated with less cash use and more debit use. Our estimates show that a one-standard-deviation increase in the robbery rate (i.e., four more robbery incidences per 10,000 residents) reduces

2010-15. For robustness checks, we also ran regressions only on 2011 data as well as on data from other sample years. The results are largely consistent.

¹² 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.

¹³ When interpreting the relationship between market performance and HHI, two hypotheses are often tested. One is the Structure-Conduct-Performance (SCP) hypothesis, which assumes that the ability of banks in a local market to set relatively low deposit rates or high fees depends positively on market concentration. The other is the Efficient-Structure (ES) hypothesis, which 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. The empirical evidence on these two hypotheses is mixed (Gilbert and Zaretsky [2003] provides a comprehensive literature review). Our findings suggest that both hypotheses are relevant for our sample, with the SCP hypothesis supported by the rural market evidence and the ES hypothesis supported by the MSA evidence.

Table 2 Marginal Effects for Zip-Code Variables

Variable	Cash	Debit	Credit	Check
Median transaction size	-0.018*	0.013*	0.004*	0.001*
Banking condition				
HHI	0.035*	-0.027*	-0.010*	0.002*
HHI*metro	-0.051*	0.042*	0.011*	-0.003*
Branches per capita	0.069*	-0.038*	-0.029*	-0.002*
Socioeconomic condition				
Robbery rate	-0.126*	0.121*	0.020*	-0.014*
Median household income	0.003*	-0.013*	0.019*	-0.009*
Population density	-0.450*	0.470*	0.077*	-0.097*
Family households	-0.089*	0.104*	-0.006*	-0.009*
Housing: Owner occupied	-0.006*	-0.030*	0.025*	0.011*
Vacant	-0.021*	-0.029*	0.043*	0.006*
Demographics				
Female	-0.043*	0.131*	-0.079*	-0.010*
Age 15-34	-0.272*	0.285*	0.000	-0.013*
35-54	-0.366*	0.416*	-0.033*	-0.016*
55-69	0.070*	-0.037*	-0.011*	-0.022*
>70	-0.172*	0.161*	0.008*	0.004*
Race Black	0.055*	-0.040*	-0.007*	-0.007*
Hispanic	0.049*	-0.168*	0.114*	0.005*
Native	0.105*	-0.060*	-0.040*	-0.004*
Asian	0.037*	-0.018*	-0.018*	-0.001
Pacific Islander	0.986*	0.811*	-1.595*	-0.202*
Other	0.129*	0.111*	-0.220*	-0.019*
Multiple	-0.019	-0.136*	0.251*	-0.096*
Education level				
High school	-0.280*	0.169*	0.108*	0.003*
Some college	-0.275*	0.184*	0.089*	0.002*
College	-0.271*	0.162*	0.106*	0.003*
Pseudo R ²	0.604	0.534	0.607	0.559
Zip-code-day Observations	1,021,764	1,021,764	1,021,764	1,021,764

Note: *1 percent 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, 2015. The explanatory 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 1 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. Population density is measured in units of 100,000 residents per square mile in a zip code. All the other variables are expressed as fractions.

the predicted cash share by 0.5 percentage points but raises debit by 0.49 percentage points.

High median household income in a zip code is associated with high credit use, mainly at the expense of debit. We find that for a one-standard-deviation increase (\$12,666) in the median household income from its mean, the predicted credit share increases by 0.24 percentage points, but the debit and check shares drop by 0.16 and 0.11 percentage points, respectively. The effect on the cash share is small – it rises by 0.04 percentage points. The results suggest that median household income in our sample may largely proxy for access to credit.

We find that higher population density is associated with lower shares of paper payments and higher shares of card payments. This is consistent with McAndrews and Wang’s (2012) theory of the scale economies of adopting relatively new payment instruments. A one-standard-deviation increase in population density (1,024 residents per square mile) reduces the predicted cash share by 0.46 percentage points and check by 0.10 percentage points, but it raises debit by 0.48 percentage points and credit by 0.08 percentage points. Although the stores in our sample accept both credit and debit cards, consumers’ adoption decisions should be related to the policies of other stores, and those may vary systematically with population density.

(IV) Demographic Variables Consistent with some existing payments studies (e.g., Klee [2008]), we find that demographic characteristics such as gender, age, and race are systematically related to consumer payment choice.

We find that a higher female ratio is associated with less cash use and more debit use. A higher presence of older age groups is associated with greater use of debit but less use of cash and check relative to the baseline age group, under 15. This might be because minors do not have access to noncash payments or because families with children tend to use more cash and check. However, the age profile with respect to cash is nonmonotonic. A higher presence of the age group 55-69 is associated with a significantly higher cash fraction. These findings suggest that the age variables may capture a combination of age and cohort effects. We also find that compared to white, minority groups tend to be associated with higher cash shares but lower debit shares.

(V) Education Variables We find a more educated population (i.e., high school and above) is associated with a lower cash fraction relative to the baseline education group (i.e., below high school). For education levels at high school and above, however, the difference is quite small between the sub-groups.

Table 3 State Fixed Effects

State Dummies	Cash	Debit	Credit	Check
North Carolina	-0.069*	0.095*	-0.025*	-0.001
South Carolina	-0.058*	0.086*	-0.027*	-0.000
Virginia	-0.063*	0.067*	-0.006*	0.002*
West Virginia	-0.033*	0.042*	-0.010*	0.001*

Note: *1 percent significance level based on robust standard errors.

(VI) State Dummies Our results reveal some interesting state fixed effects, as shown in Table 3. Compared with the benchmark state, Maryland, other states show lower shares of cash use and higher shares of debit use. This is particularly significant for North Carolina, South Carolina, and Virginia. They each have a cash share that is 5.8-6.9 percentage points lower than Maryland but a debit share that is 6.7-9.5 percentage points higher. West Virginia is the intermediate case, of which cash share is 3.3 percentage points below Maryland, and debit share is 4.2 percentage points higher. These states also show a lower share of credit use than Maryland but the magnitude is fairly small compared with debit, and the state fixed effects on check are quantitatively negligible.

(VII) Time Dummies Figure 4 plots the marginal effects associated with our estimated day-of-week dummies. The cash and debit effects are nearly mirror images of each other: cash falls and debit rises from Monday through Thursday, then cash rises and debit falls on Friday and Saturday, and the pattern reverses again on Sunday. Although credit displays less variation than cash or debit, there are noticeable movements in credit from Friday through Sunday.

Figure 5 plots the marginal effects associated with our day-of-month dummies. Whereas most of the “substitution” within the week occurred between cash and debit, within the month the substitution with cash comes from both credit and debit, especially credit. Early in the month, cash is at its highest and credit and debit are at their lowest. Over the month, cash generally falls and credit rises. Debit has a similar pattern to credit, although the variation is smaller.

Figure 4 Day-of-Week Marginal Effects

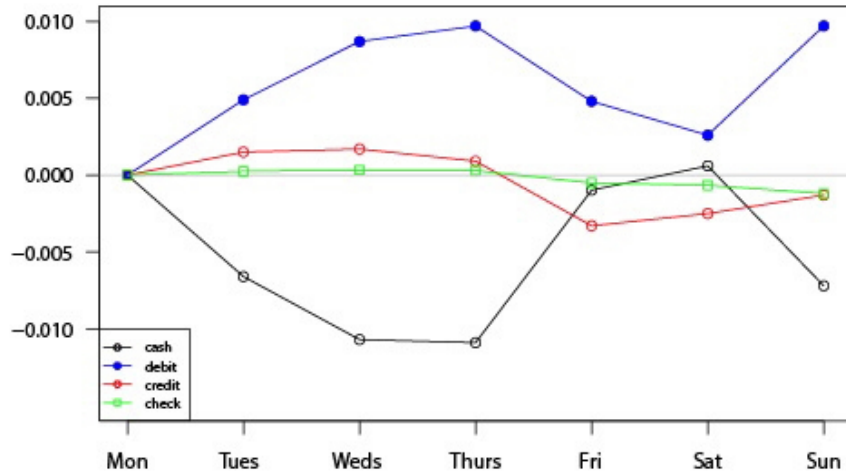
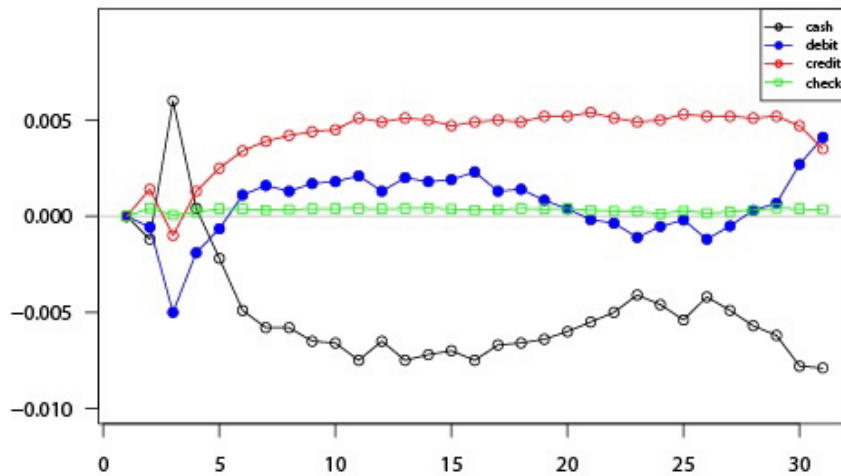
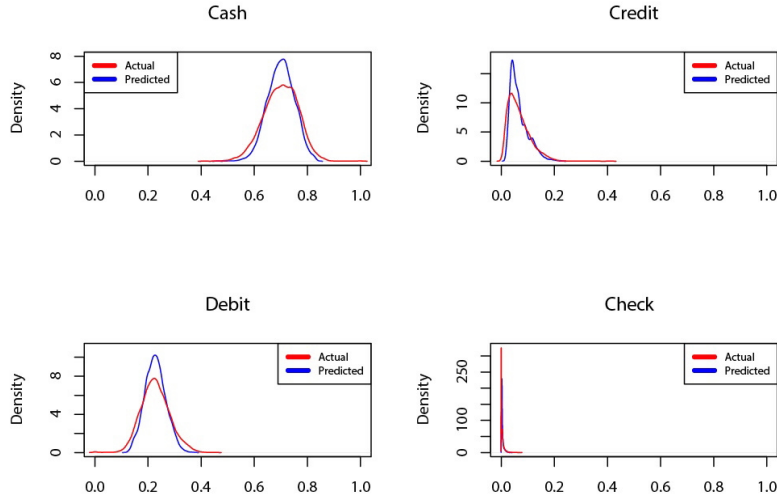


Figure 5 Day-of-Month Marginal Effects



A natural explanation for the day-of-week and day-of-month effects is consumers' changing financial or cash-holding positions during the period. Presumably, the weekly pattern could be driven by consumers

Figure 6 Fitted and Actual Payment Shares for March 2015

who receive weekly paychecks, while the monthly patterns are likely driven by those who receive monthly pay, including those who receive certain government benefits. One notable feature of the monthly pattern is a transitory reversal of the broad trends on the third day of the month. In fact, many recipients of Social Security and Supplemental Security Income are usually paid on the third of the month. Early in the month, these customers may be financially unconstrained, and thus spend cash, whereas late in the month they rely more on credit while anticipating the next paycheck. In Wang and Wolman (2016), we provide more extensive discussions of the weekly and monthly patterns.

The month-of-sample dummies in our regression identify the seasonal cycles and longer-run trends in the payment mix, but we will defer that discussion to Section 4.

3. PAYMENT VARIATION ACROSS LOCATIONS

Our regression analysis helps shed light on payment variation across locations. In this section, we will first evaluate the relative importance of the explanatory variables in accounting for such variation, and then project payment variation across the entire Fifth District for retail outlets similar to those in our sample.

Relative Importance of Explanatory Variables

In Figure 6, we plot the actual and model-predicted distributions of payment fractions for March 2015, a counterpart to Figure 2.¹⁴ The figure shows that our regression model does a good job of capturing observed payment variation. With many explanatory variables included in the regression analysis, an immediate question is what factors account for most of the variation. To answer this question, we conduct the following decomposition exercise. We first calculate the pseudo- R^2 statistics, defined as the square of the correlation between the model-predicted value and the actual data, for the March 2015 sample. We then fix each subgroup of explanatory variables one by one at the sample mean values and recalculate the pseudo- R^2 statistics. The reduction of the model fit is then used as a measure of explanatory power of the controlled explanatory variables. Finally, we compare the relative importance across all the subgroups of explanatory variables.

Table 4 reports the comparison results for cash and debit, the two most used means of payment in our data. The table shows that the day-of-week and day-of-month dummies account for little of the data variation (1 to 2 percent), so the payment variation in the one-month data is mostly cross-location variation. For cash, it is median transaction size, education levels, demographics, and state fixed effects that rank as the top four factors in explaining the variation in cash fractions, each accounting for 44 percent, 19 percent, 17 percent, and 14 percent, respectively. These are also the top four factors that explain the variation in debit fractions, though the ranking is a little different, with state fixed effects ranking first (44 percent), followed by median transaction size (23 percent), demographics (14 percent), and education levels (9 percent).

The decomposition exercise above takes the median transaction size as given and shows that it explains a large share of payment variation across locations. However, it is possible that median transaction size is not independent of other location-specific variables. This will in turn affect the interpretation of the decomposition. To account for that, we conduct an alternative exercise. First, we regress median transaction size for each zip-code day on all the other explanatory variables using a linear model and calculate the model-predicted median transaction sizes and the residuals. Second, we re-run the FMLogit model as before but replace the median transaction sizes with the residual median

¹⁴ Note that the data plots in Figure 2 and Figure 6 are slightly different because a small fraction of stores is omitted from the regression analysis due to missing zip-code-level information.

**Table 4 Relative Importance of Explanatory Variables
(March 2015)**

Scenarios	Cash			Debit		
	R ²	ΔR ²	$\frac{\Delta \bar{R}^2}{\text{sum}(\Delta \bar{R}^2)}$	R ²	ΔR ²	$\frac{\Delta \bar{R}^2}{\text{sum}(\Delta \bar{R}^2)}$
All variables included	0.531			0.374		
Constant median transaction size	0.342	0.189	44%	0.264	0.110	23%
Constant banking condition	0.521	0.010	2%	0.366	0.008	2%
Constant socioeconomic factors	0.522	0.009	2%	0.347	0.027	6%
Constant demographics	0.456	0.075	17%	0.306	0.068	14%
Constant education levels	0.450	0.081	19%	0.330	0.044	9%
Constant state fixed effects	0.472	0.059	14%	0.163	0.211	44%
Constant day of week & month effects	0.525	0.006	1%	0.367	0.007	2%

transaction sizes. Finally, we redo the decomposition exercise above based on this new FMLogit regression.¹⁵

Table 6 in the Appendix reports the results of the new FMLogit regression. Note that the new FMLogit model contains the same information as the original one, so it yields the same marginal effects for median transaction size, as well as the same model fit in terms of the pseudo- R^2 values as in Table 2. The only difference is that the new model attributes some additional payment variation to the location-specific variables through their impact on median transaction size, which results in different estimated marginal effects for those variables. Comparing Tables 2 and 6 confirms this, but the qualitative results found in Table 2 remain largely unchanged.

Based on the alternative regression model, we redo the decomposition exercise and report the results in Table 5. For cash, median transaction size, education, demographics, and state fixed effects remain the top four factors driving cash fractions, though the ranking and relative shares differ slightly from Table 4: demographics now comes in first (36 percent) followed by median transaction size (23 percent), education levels (16 percent), and state fixed effects (13 percent). A similar case is found for debit.

¹⁵ For the purpose of estimating the effects of the other explanatory variables, the alternative model where we use residual median transaction size instead of median transaction size is equivalent to running a regression without the median transaction size. Also, in principle, we could run the alternative model for each subgroup of variables other than median transaction size, but we chose not to do so. One consideration is that median transaction size is likely to be affected by other, more fundamental variables (such as income and race) but not the other way around.

Table 5 Relative Importance of Explanatory Variables (An Alternative Model)

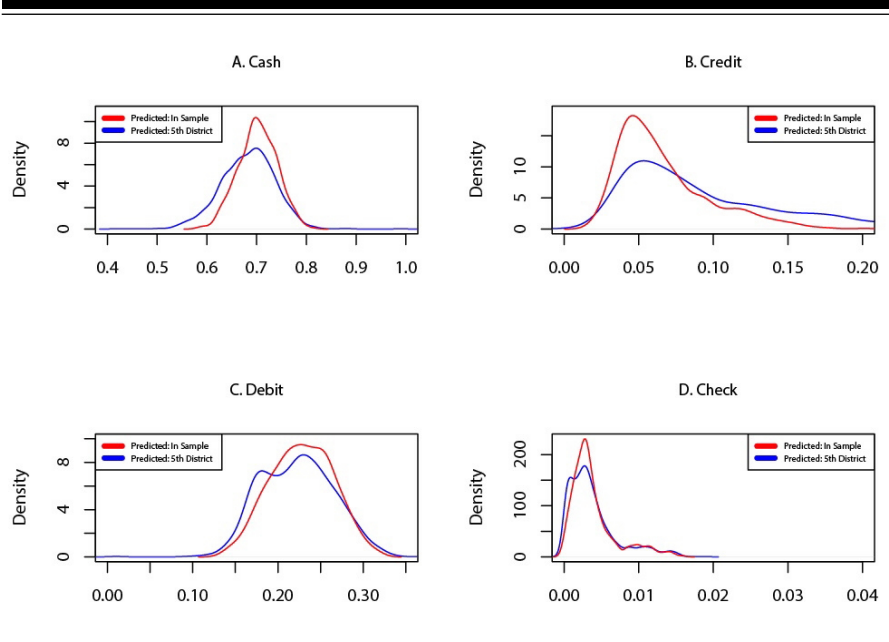
Scenarios	Cash			Debit		
	R ²	ΔR^2	$\frac{\Delta R^2}{\text{sum}(\Delta R^2)}$	R ²	ΔR^2	$\frac{\Delta R^2}{\text{sum}(\Delta R^2)}$
All variables included	0.531			0.374		
Constant median transaction size	0.425	0.106	23%	0.316	0.058	13%
Constant banking condition	0.527	0.004	1%	0.367	0.007	1%
Constant socioeconomic factors	0.490	0.041	9%	0.330	0.044	10%
Constant demographics	0.364	0.167	36%	0.259	0.115	25%
Constant education levels	0.457	0.074	16%	0.333	0.041	9%
Constant state fixed effects	0.472	0.059	13%	0.201	0.173	38%
Constant day of week & month effects	0.522	0.009	2%	0.359	0.015	3%

Payment Variation across the Entire Fifth District

The estimation results above allow us to project payment variation across the entire Fifth District for similar retail outlets. Comparing our data with the entire Fifth District, we notice that the store locations in our sample are not fully representative (Table 7 in the Appendix provides summary statistics for zip-code-level explanatory variables for the entire Fifth District). On average, store locations in our sample have fewer bank branches per capita, lower median household income, lower population density, and a smaller percentage of college graduates. The racial composition also differs from the rest of the Fifth District: there is a higher percentage of blacks, Hispanics, and Native Americans and a lower percentage of whites and Asians.

Based on the estimates from our regression model, we now address a counterfactual question: if the retail chain were to locate stores equally across the entire Fifth District, what would be the payment pattern? To answer the question, we first use the benchmark model to predict payment shares across the Fifth District with the assumption that all the zip-code locations in the Fifth District have the same median transaction size as the mean of the regression sample. The results are shown in Figure 7. We find that comparing with our regression sample, the entire Fifth District would show a similar pattern of payment variation: cash is being used most at this type of retail outlets, followed by debit, credit, and finally check. However, the relative share of these payment means would differ. We find that cash as well as debit and check would be used less in the rest of the Fifth District, while credit would be used more. This is consistent with the location bias of the stores in our sample, as discussed above.

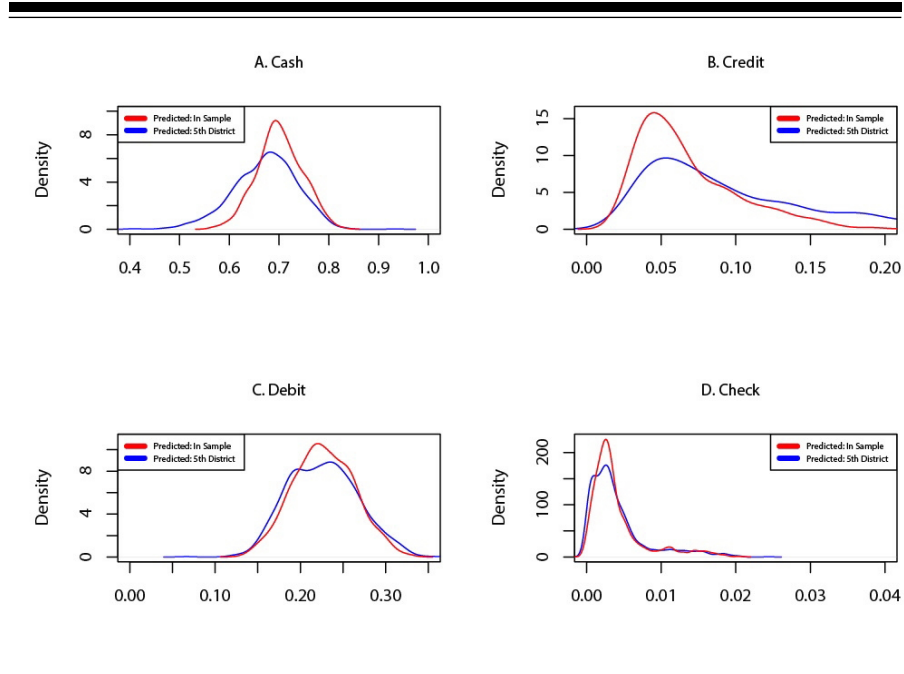
Figure 7 Predicted Payment Variation: Sample Locations vs. Entire Fifth District



As a robustness check, we also redo the counterfactual exercise using the alternative regression model, in which we replace the median transaction size with the residual median transaction size. This takes into account that location-specific variables may also affect payment variation through their effects on transaction sizes. The results are plotted in Figure 8. As it turns out, Figures 7 and 8 are not very different.

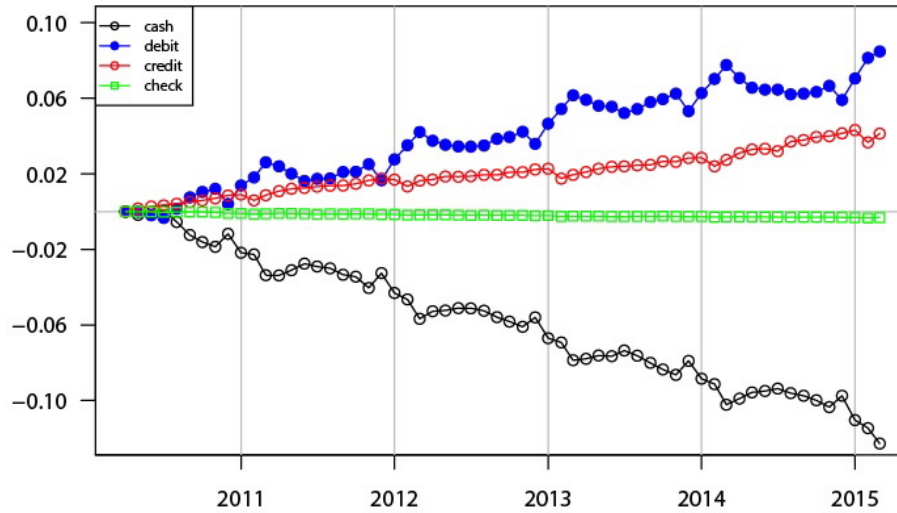
4. PAYMENT VARIATION OVER THE LONGER TERM

The month-of-sample dummies in our regression identify changing payment mix over the longer term. Figure 9 plots the marginal effects for month-of-sample dummies. These effects combine seasonality with a time trend and idiosyncratic monthly variation. The vertical lines indicate each January in our sample years. The estimated annual time trends are -2.46 percentage points for cash, 1.69 percentage points for debit, 0.83 percentage points for credit, and -0.06 percentage points for check. This suggests a longer-term trend of declining cash shares at this retailer, largely replaced by debit.

Figure 8 Predicted Payment Variation: A Robustness Check

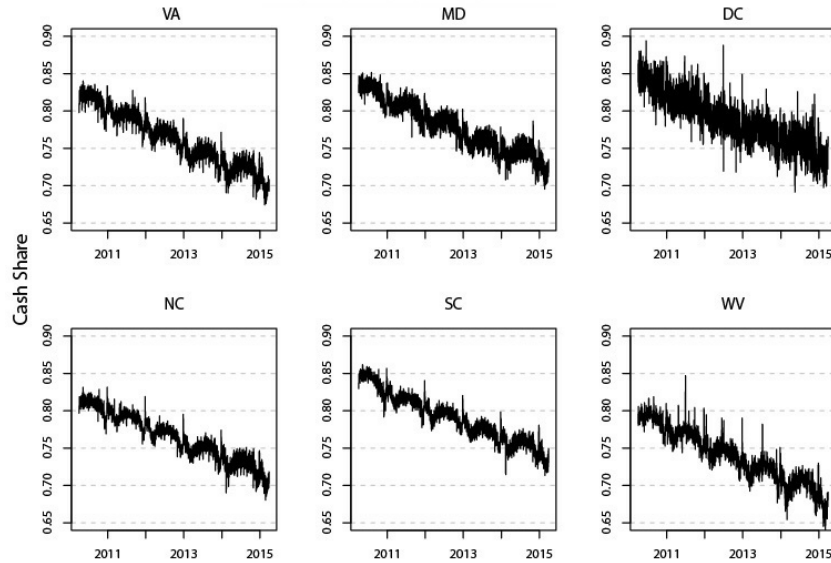
The trend decline in the share of cash transactions is striking. Moreover, we plot the raw transactions data in Figure 10, which shows that this trend is not driven by any particular subset of stores or regions but is universal for the Fifth District. Exploring the driving forces behind the trend would be useful for understanding the changing demand for currency in retail transactions more broadly. We discuss several candidate factors below.

First, one may wonder whether the decline in cash over the five years of our sample could be driven by transitory factors, such as 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, in our sample most of the cash decline was offset by an increase in debit. As Figure 1 shows, credit accounts for only about 4 percent of transactions at the beginning of the sample period and 7 percent at the end. And note that even 7 percent is an overestimate because our measure of credit includes signature debit and prepaid cards.

Figure 9 Month-of-Sample Marginal Effects

Another possible transitory factor is a change 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 company that more than 50 percent of its debit transactions were exempt from the regulation because the debit cards used were issued by financial institutions with 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.

Figure 10 The Decline in Cash Transactions in the Fifth District States and D.C.



Another question is whether the store altered the range of retail goods it sold during the sample period so that it attracted a clientele with different payment preferences. We cannot fully rule out this possibility given that we do not observe individual customers, but 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, we then turn to longer-term factors. First, there could be an increasing trend of transaction sizes. It is indeed true that the average median transaction size at this retailer increased from \$6.27 to \$7.07 from 2010 to 2015. However, according to our estimation results, this could only account for a decline of cash shares of 1.47 percentage points out of the overall decline of 12.28 percentage points over the five years. Second, part of the time trend is presumably attributable to the change in zip-code-level variables. Recall that we treated all zip-code-level variables 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. However, as shown in Wang and Wolman (2016), the forecasted changes for the zip-code-level variables can only explain a relatively small portion of the decline of cash shares.

This leaves a large fraction of the time trend still to be explained. Prime candidates are technological progress and changing consumer perceptions of the attributes of debit payments relative to others. These attributes include 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.

While our data is from one retail chain, our exercise highlights the rise of debit in place of cash. 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 noncash payment instrument in the U.S. economy: debit accounted for 19 percent of all noncash transactions in 2003, and its share doubled by 2012. Our study provides firsthand micro evidence that the increase in debit came at the expense of cash at a large cash-intensive 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.

5. CONCLUSION

Using five years of transactions data from a large discount retail chain with hundreds of stores across the Fifth District, we study payment variation across locations and time. We find that the fraction of cash (noncash) transactions decreases (increases) with median transaction size and is affected by location-specific variables reflecting consumers' preferences and the opportunity costs of using cash relative to non-cash means of payment. With the estimation results, we evaluate the relative importance of various factors in explaining the cross-location payment variation in our sample. We find that the median transaction size, demographics, education levels, and state fixed effects are the top factors. Taking those into consideration, we also project payment variation across the entire Fifth District for retail outlets similar to those in our sample.

We also identify interesting time patterns of payment variation. In particular, over the longer term, the shares of cash and check transactions decline steadily, while debit and credit shares rise. The overall cash fraction of transactions is estimated to have declined by 2.46 percentage points per year in our five-year sample period, largely replaced by debit. We show that the decline in cash at this particular retailer was likely not driven by transitory factors, and only a relatively small

fraction could be explained by changes in the median transaction size and the zip-code-level variables. 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.

APPENDIX: THE FMLOGIT MODEL

The regression analysis in the paper uses the fractional multinomial logit model (FMLogit). 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 $i = 1, \dots, N$ zip-code-day observations, each with M outcomes of payment shares. In our context, $M = 4$, which correspond to cash, debit, credit, and check. Letting s_{ik} represent the k^{th} outcome for observation i , and x_i , $i = 1, \dots, N$, be a vector of exogenous covariates. The nature of our data requires that

$$s_{ik} \in [0, 1] \quad k = 1, \dots, M;$$

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

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

Given the properties of the data, the FMLogit model provides consistent estimates by enforcing conditions (1) and (2),

$$E[s_k \mid x] = G_k(x; \beta) \in (0, 1), \quad k = 1, \dots, M; \quad (1)$$

$$\sum_{m=1}^M E[s_m \mid x] = 1; \quad (2)$$

and also accommodating conditions (3) and (4),

$$\Pr(s_k = 0 \mid x) \geq 0 \quad k = 1, \dots, M; \quad (3)$$

$$\Pr(s_k = 1 \mid x) \geq 0 \quad k = 1, \dots, M; \quad (4)$$

where $\beta = [\beta_1, \dots, \beta_M]$.¹⁶ Specifically, the FMLogit model assumes that the M conditional means have a multinomial logit functional form in linear indexes as

$$E[s_k \mid x] = G_k(x; \beta) = \frac{\exp(x\beta_k)}{\sum_{m=1}^M \exp(x\beta_m)}, \quad k = 1, \dots, M. \quad (5)$$

As with the multinomial logit estimator, one needs to normalize $\beta_M = 0$ for identification purposes. Therefore, Equation (5) can be rewritten as

$$G_k(x; \beta) = \frac{\exp(x\beta_k)}{1 + \sum_{m=1}^{M-1} \exp(x\beta_m)}, \quad k = 1, \dots, M-1; \quad (6)$$

and

$$G_M(x; \beta) = \frac{1}{1 + \sum_{m=1}^{M-1} \exp(x\beta_m)}. \quad (7)$$

Finally, one can define a multinomial logit quasi-likelihood function $L(\beta)$ that takes the functional forms Equations (6) and (7) 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

$$L(\beta) = \prod_{i=1}^N \prod_{m=1}^M G_m(x_i; \beta)^{s_{im}}. \quad (8)$$

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 our regression analysis, we use Stata code developed by Buis (2008) for estimating the FMLogit model.

¹⁶ To simplify the notation, the “ i ” subscript is suppressed in Equations (1)-(7).

Table 6 Marginal Effects for Zip-Code Variables

Variable	Cash	Debit	Credit	Check
Residual median transaction size	-0.018*	0.013*	0.004*	0.001*
Banking condition				
HHI	0.036*	-0.028*	-0.010*	0.002*
HHI*metro	-0.042*	0.036*	0.009*	-0.003*
Branches per capita	0.040*	-0.017*	-0.022*	-0.001*
Socioeconomic condition				
Robbery rate	-0.177*	0.158*	0.032*	-0.012*
Median household income	-0.018*	0.003*	0.024*	-0.008*
Population density	-0.623*	0.595*	0.118*	-0.091*
Family households	-0.158*	0.154*	0.010*	-0.006*
Housing: Owner occupied	0.005*	-0.038*	0.023*	0.010*
Vacant	-0.033*	-0.020*	0.046*	0.007*
Demographics				
Female	-0.074*	0.154*	-0.072	-0.009*
Age 15-34	-0.364*	0.351*	0.022*	-0.009*
35-54	-0.485*	0.502*	-0.005	-0.012*
55-69	-0.018*	0.027*	0.010*	-0.019*
>70	-0.148*	0.144*	0.002	0.003*
Race Black	0.093*	-0.068*	-0.016*	-0.009*
Hispanic	-0.022*	-0.117*	0.131*	0.008*
Native	0.125*	-0.075*	-0.045*	-0.005*
Asian	0.115*	-0.074*	-0.036*	-0.004*
Pacific Islander	-0.153	1.637*	-1.324*	-0.159*
Other	0.257*	0.017*	-0.251*	-0.024*
Multiple	0.220*	-0.309*	0.194*	-0.105*
Education level				
High school	-0.271*	0.162*	0.106*	0.003*
Some college	-0.278*	0.186*	0.090*	0.002*
College	-0.257*	0.153*	0.102*	0.002*
Pseudo R ²	0.604	0.534	0.607	0.559
Zip-code-day observations	1,021,764	1,021,764	1,021,764	1,021,764

Note: *1 percent 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, 2015. The explanatory 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 1 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. Population density is measured in units of 100,000 residents per square mile in a zip code. All the other variables are expressed as fractions.

Table 7 Summary Statistics of Zip-Code Variables (Entire Fifth District)

Variable (unit)	Mean	Std. dev.	1%	99%
Banking condition				
HHI metro	0.192	0.148	0.059	0.735
HHI rural	0.326	0.171	0.125	1.000
Branches per capita (1/10 ³)	0.66	2.97	0.05	4.68
Socioeconomic condition				
Robbery rate (1/10 ⁵)	30.13	41.30	0.00	235.07
Median household income (\$)	50910	24859	22214	140093
Population density (per mile ²)	1157	2473	15	11514
Family households (%)	66.87	9.92	28.17	88.16
Housing (%): Renter occupied	27.17	13.73	6.52	77.14
Owner occupied	59.69	15.10	3.87	89.03
Vacant	13.14	10.70	2.41	64.89
Demographics (%)				
Female	50.73	3.22	35.64	55.18
Age				
<15	18.22	3.92	5.76	27.27
15-34	25.65	8.40	13.92	60.96
35-54	27.46	3.82	12.12	35.05
55-69	18.71	4.66	2.42	33.53
>70	9.97	3.86	0.79	22.71
Race				
White	72.18	21.73	10.56	98.95
Black	19.80	19.70	0.07	82.09
Hispanic	5.97	6.40	0.30	32.24
Native	0.69	3.54	0.00	6.67
Asian	2.44	4.27	0.00	23.34
Pacific Islander	0.06	0.09	0.00	0.46
Other	2.74	3.69	0.00	18.48
Multiple	2.10	1.14	0.38	5.60
Education level (%)				
Below high school	15.20	11.38	0.00	54.00
High school	34.60	13.18	0.00	70.60
Some college	20.91	8.89	0.00	49.60
College	29.30	16.71	0.00	80.40

REFERENCES

- Arango, Carlos, Kim P. Huynh, and Leonard Sabetti. 2011. "How Do You Pay? The Role of Incentives at the Point-of-Sale." Bank of Canada Working Paper 2011-23 (October).
- Borzekowski, Ron, Elizabeth K. Kiser, and Shaista Ahmed. 2008. "Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response." *Journal of Money, Credit, and Banking* 40 (February): 149–72.
- Borzekowski, Ron, and Elizabeth K. Kiser. 2008. "The Choice at the Checkout: Quantifying Demand Across Payment Instruments." *International Journal of Industrial Organization* 26 (July): 889–902.
- Buis, Maarten L. 2008. "FMLogit: Stata Module Fitting a Fractional Multinomial Logit Model by Quasi Maximum Likelihood." Statistical Software Components, Department of Economics, Boston College (June).
- Ching, Andrew T., and Fumiko Hayashi. 2010. "Payment Card Rewards Programs and Consumer Payment Choice." *Journal of Banking and Finance* 34 (August): 1773–87.
- Cohen, Michael, and Marc Rysman. 2012. "Payment Choice with Consumer Panel Data." Memo, Boston University Department of Economics.
- Federal Reserve System. 2014. "The 2013 Federal Reserve Payments Study." https://www.frbservices.org/communications/payment_system_research.html (July).
- Gilbert, R. Alton, and Adam M. Zaretsky. 2003. "Banking Antitrust: Are the Assumptions Still Valid?" Federal Reserve Bank of St. Louis *Review* 85 (November/December): 29–52.
- Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52 (May): 681–700.
- Klee, Elizabeth. 2008. "How People Pay: Evidence from Grocery Store Data." *Journal of Monetary Economics* 55 (April): 526–41.
- Koulayev, Sergei, Marc Rysman, Scott Schuh, and Joanna Stavins. 2016. "Explaining Adoption and Use of Payment Instruments by U.S. Consumers." *RAND Journal of Economics* 47 (Summer): 293–325.

- McAndrews, James, and Zhu Wang. 2012. "The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage." Federal Reserve Bank of Richmond Working Paper 12-06 (October).
- Mullahy, John. 2010. "Multivariate Fractional Regression Estimation of Econometric Share Models." National Bureau of Economic Research Working Paper 16354 (September).
- Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics* 11 (November/December): 619–32.
- Schuh, Scott, and Joanna Stavins. 2014. "The 2011 and 2012 Surveys of Consumer Payment Choice." Federal Reserve Bank of Boston Research Data Report 14-1 (September).
- Schuh, Scott, and Joanna Stavins. 2012. "How Consumers Pay: Adoption and Use of Payments." Federal Reserve Bank of Boston Working Paper 12-2.
- Wang, Zhu. 2016. "Price Cap Regulation in a Two-sided Market: Intended and Unintended Consequences." *International Journal of Industrial Organization* 45 (March): 28–37.
- Wang, Zhu, and Alexander L. Wolman. 2016. "Payment Choice and Currency Use: Insights from Two Billion Retail Transactions." *Journal of Monetary Economics* 84 (December): 94–115.
- Zinman, Jonathan. 2009. "Debit or Credit?" *Journal of Banking and Finance* 33 (February): 358–66.