Exploring customer Spending Behavior and Payday Effect using Prepaid Cards Transaction Data

Su Zhou[1], Alan Montgomery[2,1], and Geoff Gordon[1]

[1]Machine Learning Department, Carnegie Mellon University
[2]Tepper Business School, Carnegie Mellon University

August 12, 2016

Abstract

Background: Prepaid cards are reloadable charge cards typically designed for customers that would not be able to have a traditional credit card due to no or poor credit history, and hence unable to open checking account or credit card at a bank. Many prepaid card customers would have their salary credited to their account and make most withdrawals and purchases on this card since they do not have many other options. Their data are therefore very suitable for studying customer spending behaviors.

Aim: In this work, we study how income affects customer spending patterns. Specifically, we aim to determine whether or not a "payday effect", i.e. payroll incomes affecting customers' spending behaviors, exists in the prepaid card customers and if so how does it vary over time and transaction categories.

Data: We use prepaid card transaction data collected by a large national bank over the course of 28 months which covers customers across the United States. The subset we considered contains 41,610 customers. We processed the data so that there are 6 categories of spending and 4 categories of deposits. The length of the transaction history of each customer varies between 11 days to 28 months.

Method: We use a variant of the linear regression model from prior studies on payday effects and carefully control for possible confounding factors, specifically balance features.

Results: Strong (and statistically significant) responses are detected for all transaction categories on the coefficients corresponding to indicator features for days on and after payday. The explained variances on non-essential spending (restaurants and others) are larger than that on essential spending (grocery, gas, etc...).

Conclusion: We conducted our analysis on the new prepaid card transaction data with additional control for balance. A strong payday effect exists on prepaid card customers despite the predictable nature of payroll income. The strengths may vary across different types of spending.

1 Introduction

A growing segment of the credit card industry is the use of prepaid cards. Prepaid cards are reloadable charge cards that are used and look just like a typically charge card. The primary difference with a charge card is that no credit is extended, and prepaid card customers can only spend up to the balance of their account. In addition to functioning like a charge card, many prepaid cards also incorporate traditional banking features such as cash withdrawals from ATMs, direct deposit of paychecks, online bill payments, and writing traditional paper checks that draw funds on the card. Prepaid cards are especially attractive to customers that would not be able to have a traditional credit card due to no or poor credit history. These customers are frequently referred to as the *underbanked*. This is quite advantageous since these prepaid card customers get access to traditional banking systems that previously were not available to them. Since these underbanked individuals are less likely to have other accounts, data collected from prepaid cards represent a more complete picture of their financial activities. As a result, we find such data suitable for studying customers' spending behaviors. In particular, we aim to study how spending responds to income and answer the following questions:

1. Does a "payday effect" exist?
2. How does it differ across different categories of spending?

Similar problems were previously studied on mobile phone financial app data sets recently [Gelman et al., 2014, Pagel and Vardardottir, 2015]. To the best of our knowledge, no such study has been conducted for prepaid cards, which has a significant representation of arguably the most interesting subgroup in population for this problem — the low-income households who live hand-to-mouth and need to be cautiously planning and optimizing their consumption given limited resources. This is not surprising because despite being the fastest growing section, the total market share of prepaid card is still as small as 4% as of 2014 [CUNA, 2015]. This paper aims to partially fill in this knowledge gap concerning this segment of consumers.

1.1 Problem definition

We plan to investigate the problem "whether payday effect exists" by answering a sequence of questions.

1. Do people spend more around payday, than not around payday?
2. Do people spend differently in different categories before and after payday?
3. Do some categories (non-essential products) have larger number of transactions/larger spending than other categories (essential products)?
4. How do we know that larger number of transactions/larger spending is not due to the other factors, e.g., seasonal, holiday, day of week, individual, balance and so on?

1.2 Related work and our contributions

Classic economic theory predicts that consumption should not respond to predictable changes in disposable income (see e.g., Friedman, 1957, Hall, 1979). Empirical evidence seems to suggest otherwise. There had been various reports indicating that consumption is "excessively sensitive" to income changes [Flavin, 1981, Parker, 1999, Shapiro and Slemrod, 2003, 2009], as surveyed by Jappelli and Pistaferri [2010].

More recently, Gelman et al. [2014] collected data from a smartphone app that tracks personal financial activities and studied the response of spending to income. They considered total-spending, and non-recurrent spending and established strong pattern of increased spending in response to

Table 1: Transaction Snippet

CardKey	Amount	Date	Post time	MCC	Merchant Name	Transaction Description
10000344	-79.79	10/17/2015	16:25:29	5311	TJMAXX #0479	Purchase of Goods or Services
10000344	-4.23	10/18/2015	12:31:45	5814	STARBUCKS #02900	Purchase of Goods or Services
10000344	+284.77	10/22/2015	23:59:21	-	NYACS	ACH Load Credit

income. They also compared over families with different income levels and argue that the effects are stronger for those who have liquidity constraints.

Pagel and Vardardottir [2015] confirmed their findings in an unpublished manuscript, by conducting a very similar analysis on Icelandic population. They also describes a more careful analysis that suggests the payday effects are also prominent in wealthy families.

Our work is different as we are the first to study prepaid card customers. [Gelman et al., 2014] consider a more diverse set of consumers, but their sample likely overrepresents consumers interested in tracking their finances, while our panel is likely to be more representative of low-income consumers. Also, [Pagel and Vardardottir, 2015] only considers non-recurrent incomes such as tax rebates, while we consider mainly the effect of payroll incomes (ACH load credit). Moreover, we carefully control for the possible confounding factor — balance.

2 Data description

The raw data contains 18,533,520 transaction records for approximately 57,613 prepaid card customers with uniquely identifiable card key. Each transaction record contains 7 data fields: card key, transaction amount in dollar, transaction date, transaction post time, merchant category code [see e.g., Visa Commercial Solutions, 2007], merchant name, transaction description. A synthetic snippet of the transaction record data is shown in Table 1

To include more details: "A merchant category code is a four-digit number used by the bankcard industry to classify suppliers into market segment. There are approximately 600 MCCs that denotes various types of business (e.g., 5111 Office Supplies, 7299 Dog Grooming Services, 5722 Household Appliance Stores). "[Visa Commercial Solutions, 2007]. A "merchant name" is a textual description of the entity (business, agency, organization) that this transaction is associated with. Unfortunately, "merchant name" does not uniquely identify merchants. Different merchant names can correspond to the same entity and the same merchant name (sometimes acronyms) can potentially refer to several different entities. Transaction description is a more structured textual description of the nature of the transaction, e.g., "Purchase of Goods or Services", "ACH Load Credit"[1] or "Monthly Maintenance Fee" and so on. In particular we use "Purchase of Goods or Services" to identify transactions that corresponds to customer spendings and "ACH Load Credit" (after some additional screening) to identify payroll deposits into the account.

Our study makes use of a carefully selected meaningful subset of the data containing 41,610 customers who satisfy the following criteria:

1. An "ACH Load Credit" has been made within 45 days of the account opening.

[1] Short for Automatic Clearing House (for more information, see IPPay, 2003). The majority of "ACH Load Credit" transactions are direct deposits of salary payments.

date	6 spending categories	4 income categories	imputed balance
06/03/2014	[23.14, 45.23,0,0,0,11.23]	[0,0,23.57,0]	378.25
06/05/2014	[0,0,0,0,0,0]	[0,0,0,0]	378.25
06/06/2014	[0,0,10.00,0,0,0]	[800, 0,0,0]	1168.25
06/07/2014	[0, 31.33, 0,0,0,299.99]	[0,0,0,0]	836.93
⋮	⋮	⋮	⋮

Table 2: Example of processed data of a customer.

2. There are at least 6 "ACH Load Credit" transactions.

These criteria are chosen to make sure the customers are sufficiently active and indeed use this card as a major (if not the only) medium of financial activities. In particular, our conjecture is that our data has all salaries deposited by those customers, and we have checked some random samples of selected customers' deposit transactons to make sure they are indeed payroll incomes.

We group the customer transactions into daily aggregates over 10 different categories of which 6 are spending categories and 4 are deposit categories. The $[25\%, 50\%, 75\%]$ quantiles of the customers' length of transaction history are about $[229, 323, 451]$ days respectively.

All spending transactions must have a transaction description of "Purchase of Goods or services". The 6 spending categories are selected based on Merchant Category Codes including: "restaurant", "auto-related spending", "grocery", "drug store/alcohol" and "wholesale/department Store". All other transactions are grouped into a separate category called "other spending".

The four deposit categories include: "Payroll", "Manual deposit" (including ATM, teller or internet deposit using cash, debit cards or other methods), "Purchase Return" and "other income".

Lastly, based on all the transaction records (including those that we filtered out, e.g., service charges, cash withdrawals) we impute the daily running balance for each customer and this will be an important factor to consider.

The processed customer data is illustrated in Table 2.

3 Exploratory data analysis

In this section, we present our exploratory data analysis and collect intuitive evidence on the existence of a payday effect.

The first thing we realized is that the customers in the dataset can be categorized into four groups according to their frequency of payroll deposits. Specifically, there are 15066 customers who are weekly-paid, 19406 customers who are biweekly-paid, 2071 customers who are monthly paid, and 5067 customers who do not display clear regularity in their payroll deposits. The grouping were constructed manually through the median gaps between adjacent payroll deposits of each customer (see Figure 1). The monthly salaries of customers in all groups vary over a large range from a few hundreds to up to 10,000 (see Figure 2) and they all have a similar positively skewed distributions. Perhaps a bit surprisingly, people who are paid more frequently on average earn more than those who paid less frequently. We further look into the distributions of daily aggregates of all transaction categories for each customer group and the results are summarized in Figure 3. While there are

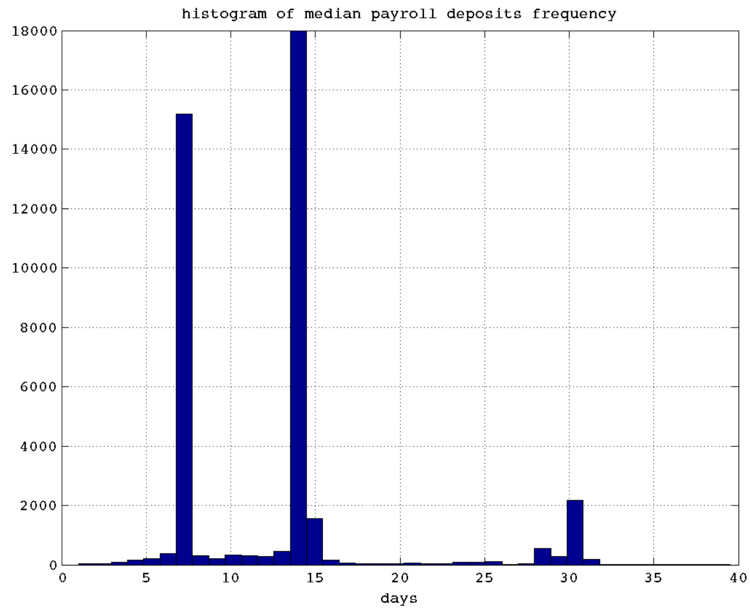

Figure 1: Histogram of median payroll deposits frequency. It is clear from the figure that there are three groups. Customers that falls into Bin 7 are considered "weekly-paid"; Bin 14 and 15 are considered "biweekly-paid"; and those in Bin 30 and 31 are considered "monthly-paid". Everyone else that we are not uncertain with this metric goes into a separate category that we call "others".

both variations over categories and customer groups, the variations are seemingly additive, and the relative patterns over categories are very similar over customer groups.

Moreover, spending seems to be affected by "day of the week" in all four customer groups as illustrated in Figure 4. In particular, people spend much more on weekend than weekdays in all categories.

In order to check whether payroll income plays an important role despite the "weekend" effects, we compare the spending patterns of the three customer groups over a two-week periods where we ensure that there is a "payday" in the first 7 days. The results are shown in Figure 5. When we compare the two figures in the first row, we see a clearly distinctive patterns between the two. There is significant increase of spending amount in both weekends in the weekly-paid customers while such increases for the biweekly-paid customers only appear in the weekend after the payday. Similar patterns appear in the monthly-paid customers, although the increase of spending appears to be more gradual. This is possibly due to that monthly payroll deposits are issued at calendar dates and are therefore equally likely on any day of the week, while weekly or biweekly payments tend to occur on Fridays more often than other days.

Assuming that the frequency that a customer receives payroll deposits is marginally independent to whether they change their spending behaviors near paydays, then Figure 5 can be thought of as a naturally-appearing randomized experiment that reveals a very strong payday effects in all spending categories.

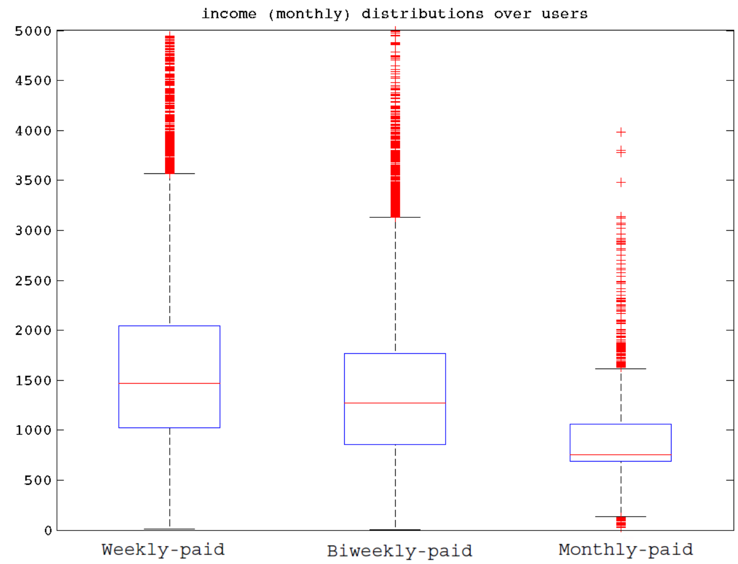

Figure 2: Box plots of salaries of customers in three customer groups (scaled to monthly amount so that they are comparable). From left to right, the plots correspond to weekly-paid, biweekly-paid and monthly-paid (blank font on server needs to be fixed later). On average, people who are paid more often earn slightly more than people who are paid less frequently. A positive skewness appears in all customer groups.

4 Method

In the last section, we have gathered intuitive evidence about the existence of the payday effects. In this section, we describe a principled way to formalize the observation, and at the same time control over many other possible confounding variables.

4.1 Model description

We use the standard linear regression as in Gelman et al. [2014], Pagel and Vardardottir [2015], which finds the ordinary least square fits of the following linear model:

$$y = \boldsymbol{x}^T \boldsymbol{\beta} + \beta_0 \tag{1}$$

Here y denotes the daily total spending in a particular category, and $\boldsymbol{x}$ gathers the features associated to that day. This is equivalent to the maximum likelihood estimator for finding

$$y \sim N(\boldsymbol{x}^T \boldsymbol{\beta} + \beta_0, \sigma^2) \tag{2}$$

for any σ^2.

The question is of course, what are the features used? As in Gelman et al. [2014], Pagel and Vardardottir [2015], we considered fixed effects variables including:

1. Day of the Week

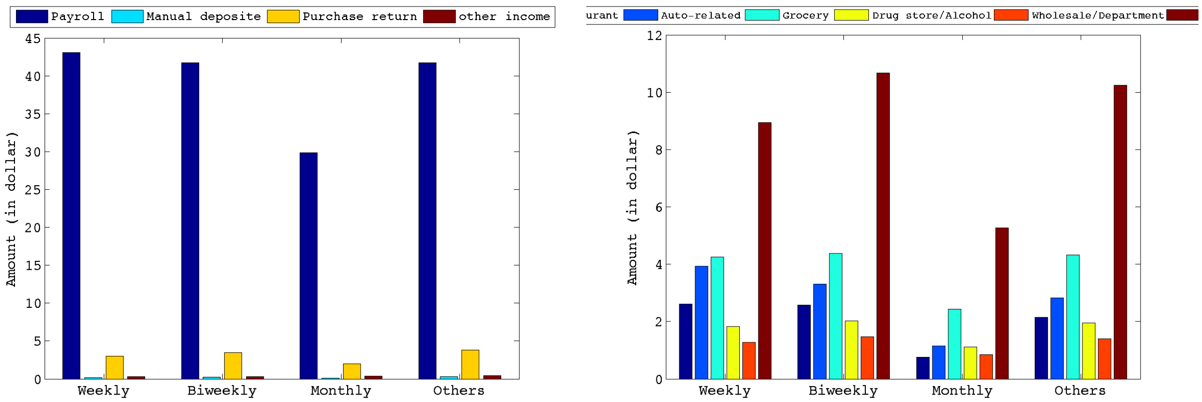

Figure 3: Average daily income and spending over transaction categories and the four customer groups. There are clear variations over categories and customer groups, but the variations over groups seem more or less homogeneous over categories.

2. Month of the Year
3. Year (2013 - 2016)
4. Holiday
5. Individual effect (one variable associated with every unique customer)

These are all binary indicator variables that we use as controls. Then we consider the variables of interest: "payday" features, in the form of a 15-dimensional indicator vector:

Payday -7 days,	...,	Payday -1 day,	Payday,	Payday +1 day,	...	Payday +7 days,

Lastly, we move beyond what was previously studied in Gelman et al. [2014], Pagel and Vardardottir [2015] and consider an additional group of control variables derived from the "account balance", which include

1. Balance in dollar
2. logarithm of Balance + 1.
3. An indicator vector of whether balance is in $[0, 50), [50, 100), [100, 500), [500, 2000), [2000, \infty)$

We allow for various non-linear transformation of the balance so as to capture the possible non-linear dependence of spending on balance.

We also considered the mixed effect model, where we model individual effects as a random draw from a Gaussian prior. The maximum a posteriori estimate of this mixed effect model translates into a Tikhonov regularization on the subset of $\boldsymbol{\beta}$ corresponding to individual effects. All results we present will be based on this version of the model with $\lambda = 1/\sqrt{n}$, but we note that the results are insenstive to the choices of λ, and the difference between OLS and the optimally-tuned mixed effect model is negligible.

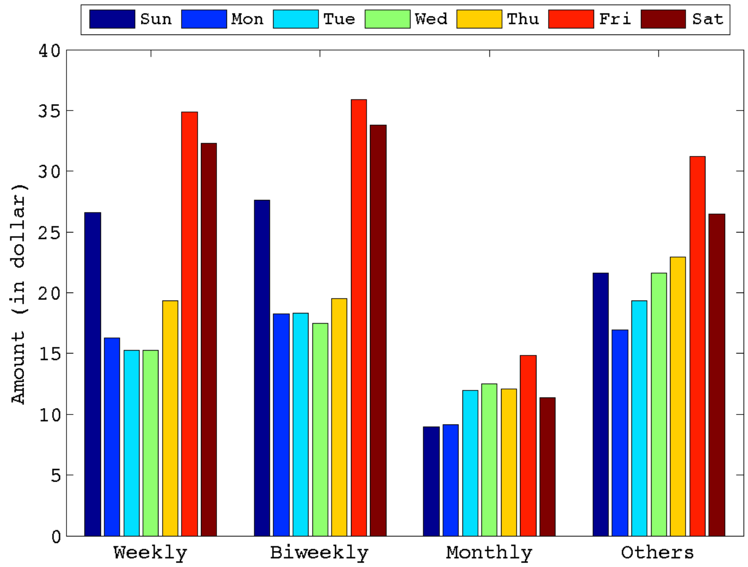

Figure 4: Average total daily spending on each day of the week, by customer groups.

4.2 Evaluation

We evaluate our fitted coefficient vector $\boldsymbol{\beta}$ to draw inferences about payday effects. We are particularly interested in the part of the coefficients that correspond to the payday (after everything else is jointly taken into consideration).

Moreover, we evaluate the explained variance measured by the R^2 value for the following four configurations:

1. $\boldsymbol{x}$ contains fixed-effects only.
2. $\boldsymbol{x}$ contains fixed-effects + balance features.
3. $\boldsymbol{x}$ contains fixed effects + payday features (This is what is used in Gelman et al. [2014], Pagel and Vardardottir [2015]).
4. $\boldsymbol{x}$ contains fixed effects + balance features + payday features.

We will then see that to what extent the changes in the spending amount can be explained by either payday or balance.

4.3 Statistical inference

Under the assumption that the linear regression model is correct. Denote the fixed design matrix by X and the true coefficient vector by β and noise be iid Gaussian with variance σ^2. In this case, $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^T X)^{-1})$. And we can conduct statistical inference on coordinate subsets of β by estimating σ^2 using the mean prediction error, which gives an estimate of the covariance

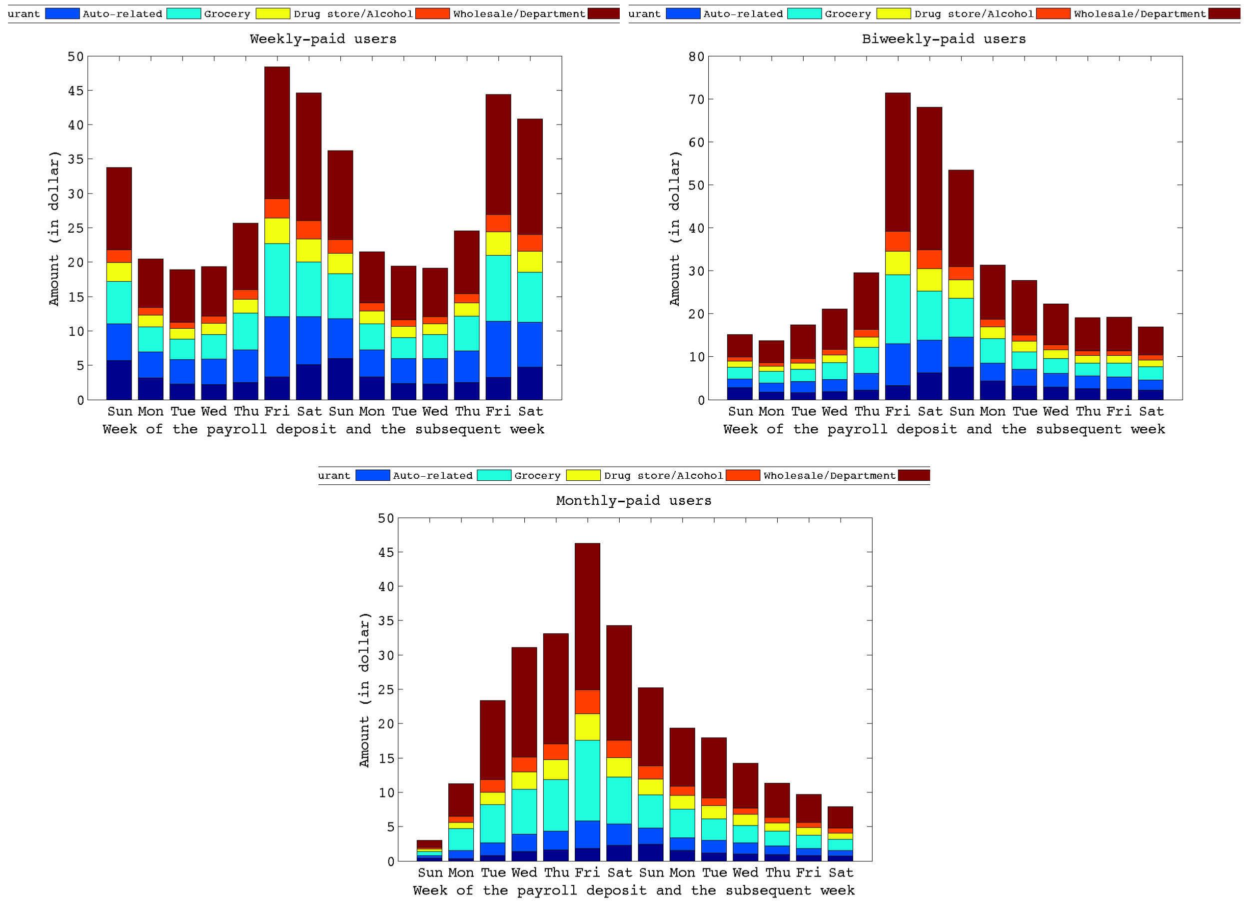

Figure 5: Customer spending over a fortnight (Sunday - Saturday, Sunday - Saturday) starting from the week with at least one payroll payment and extending to the week after that. The weekly-paid customers have a clear and almost identical surge of spending in both weekends, while biweekly-paid and monthly-paid customers only have significant surge in the weekend right after the payday.

matrix

$$\widehat{\mathrm{Cov}}(\hat{\beta}) = \left[\frac{1}{N}\sum_{i=1}^{N}(y_i - \boldsymbol{x}_i\hat{\beta})^2\right](X^TX)^{-1},$$

and for coordinate k, the (marginal) standard error:

$$se(\hat{\beta}_k) = \sqrt{\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - \boldsymbol{x}_i\hat{\beta})^2\right][(X^TX)^{-1}]_{k,k}}.$$

This allows to construct pointwise confidence interval and conduct significance test for β_k under the null hypothesis that $\beta_k = 0$ is a t-test with the t-statistic being

$$T = \hat{\beta}_k / se(\hat{\beta}_k).$$

Under the model assumption, this follows a student t-distribution with degree of freedom $N - 2$. Note that the feature dimension grows with the number of users, thanks to the individual fixed effect. Evaluating $(X^TX)^{-1}$ quickly becomes computationally burdensome. Leveraging upon the computational resources we have, we decide to conduct the hypothesis tests on a random sample of $10,000$ users. This is about a quarter of the entire data set and we still have 3.5 million data points. The test has the valid α-level and retains a fraction of statistical power (which means that we are more conservative and the results hold with higher confidence than we claim).

5 Results

We conducted the analysis as described in the previous section. The coefficients corresponding to payday features are shown in Figure 6, where different lines indicate each of the six spending categories. On the other hand, the coefficients corresponding to the balance features are shown in Figure 8. The explained variance of the four configurations are presented in Figure 9. Lastly, we show in Table 3 and 4 the p-values of the significant tests on the fitted coefficients in the case when both payday and balance features are used. These results allow us to draw the following conclusions.

The payday effect exists. Figure 6 shows that there are strong payday effects in all six spending categories that we considered. Specifically, the plot on the top agrees very well with the results presented in [Gelman et al., 2014, Pagel and Vardardottir, 2015]. There is a sharp increase in spending on payday, which reach their peak at either the payday itself or within the following two days, then slowly dies down in the subsequent couple of days. The nearly 0 or negative coefficients on the days before pay day seems to suggest that instead of anticipating a regular income and start spending more a few days before the pay day, the pre-paid card users often wait until the salary is physically in hand/or in their bank account before spending. The additional control over the balance features do not seem to affect the general structure of the coefficients.

Payday effects differ subtly across categories. While payday features seem to have an effect on all spending categories, they differ by the amount and how they evolve over time. The effects are stronger for "others" and "groceries" than other categories, but that may well be due to that "others" have relatively higher daily spending amount in that category in the first place. It is more interesting to look at the relative effects within each category. The coefficients for "restaurants", "alcohol" and "others" last longer and reach their mode in the day after the payday rather than the payday itself.

The balance effect exists. Interestingly, our results suggest that the balance interacts with people's spending behaviors more strongly than just in the form of liquidity constraints as studied before. Since the raw-balance feature, $\log(1+\cdot)$ balance features and the indicators of income bins have different units (and scale), their coefficients are not directly comparable across blocks. As a result, we compare their T-scores in Figure 7, and then the coefficients of only the bin-indicators in Figure 8. These results suggest that the effect is monotonic albeit non-linear w.r.t. balance amount. The monotonicity is clear from Figure 7 as the coefficients are mostly positive. Moreover, we see that the T-scores of the logarithmic transformed balance has a much higher significance level than that of the raw balance. Figure 8 provides a more descriptive view of the balance effect. When the balance is smaller than 100, the effects on spending is negative, which means that when all other conditions are the same, small balance would decrease the chance that people spend. When the balance is larger than 500, it appears that it grows very quickly for all categories instead of saturating at a certain level, which would be the case if a liquidity constraint is the only reason why spending depends on balance. What we see from the results is that there is a steady increase in the balance effect coefficient beyond 500 even for categories like groceries shopping where balance is no longer a constraint. This suggests that the balance effect is really based upon relative than absolute consumer perceptions, in that people feel more at ease to spend when there is a larger cushion in their bank account.

Explained variances. In Figure 9, we show the explained variance in each of the six spending categories. Adding balance features and payday features both significantly increase the explained variances in almost all categories. Instead of having their effects more or less correlated with each other, they jointly increase the R^2 score even more. This shows that the balance effect and payday effect might be largely complementary to each other, namely, they work in almost orthogonal dimensions, instead of being highly correlated.

Statistical inference. The pointwise confidence intervals of the linear regression coefficients for payday effects and balance effects are already given in Figure 6 and Figure 8 respectively. More quantitatively, the results in Table 3 and 4 suggest that with overwhelmingly high confidence, we can conclude that the detected non-zero coefficients in Figure 6 are statistically significant, and these effects are especially strong on the pay day and those days that immediately follow the pay day.

6 Discussion

Model mismatch and statistical inference. A caveat of our approach (also of those in Gelman et al., 2014, Pagel and Vardardottir, 2015) is the two model assumptions: linearity of $\mathbb{E}(Y|X)$ and the Gaussianity of the noise $Y|X - \mathbb{E}(Y|X)$. These assumptions are required for the confidence intervals and hypothesis tests to be valid. However, we would like to point out that there is an interpretation, even if $\mathbb{E}(Y|X)$ is not linear and noise being non-Gaussian. Let's assume X are also iid random variables [2], so that (X_i, Y_i) pair are random. In this case, at the limit when $n \to \infty$, the linear regression estimator converges to a "linear approximation" of $\mathbb{E}(Y|X)$. Moreover, the estimated coefficient $\hat{\beta}$ is asymptotically normal (under regularity conditions) with the covariance matrix given by the "sandwich" formula [White, 1982], although now the covariance depends on unknown quantities that relies on the unknown true nonlinear mapping from X to Y and the distribution of X, which makes it very hard to perform valid statistical inferences.

Nonparametric methods, black box machine learning tools. These methods predict better, but are computationally infeasible and most importantly they are usually not very interpretable. We used approximate kernel ridge regression on a random subset of the data set. We tune the parameter using cross-validation and then evaluate on a hold-out set. The results provide comparable R^2 values using only a fraction of the data. This results (analogous to Figure 9) are provided in the Appendix. The drawbacks are, however, they are orders-of-magnitude slower than linear methods; and more critically, they do not provide interpretations for the effects of individual features.

[2] A fixed design X can be thought of as random design with a discrete uniform distribution over the fixed set of X.

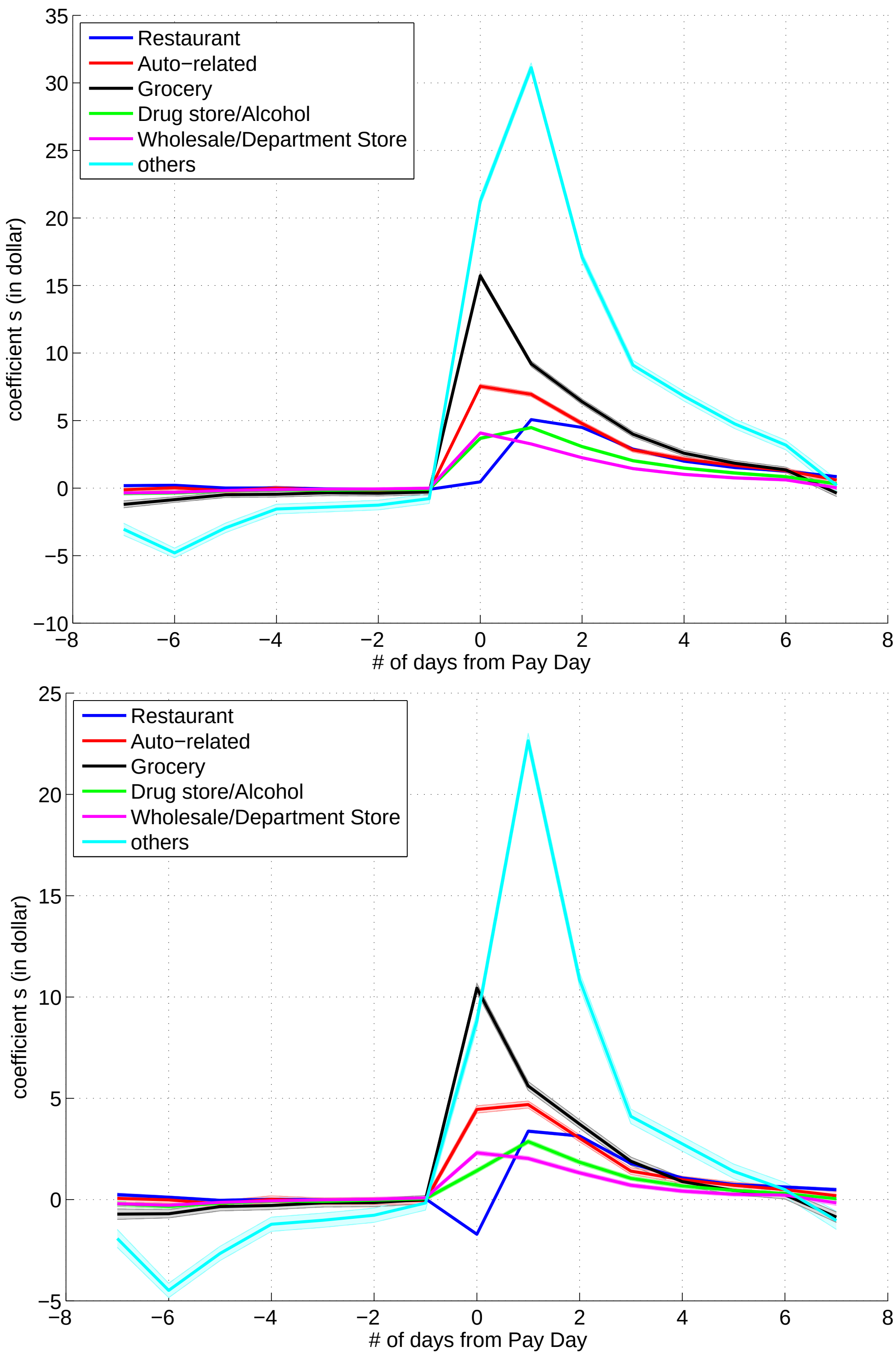

Figure 6: Linear regression coefficients of payday features. **Top pane**: Payday effects without controlling for balance. **Bottom pane**: Payday effects after controlling for balance features. The error bars are 3 standard deviation on each sides so the pointwise coverage is about 0.997.

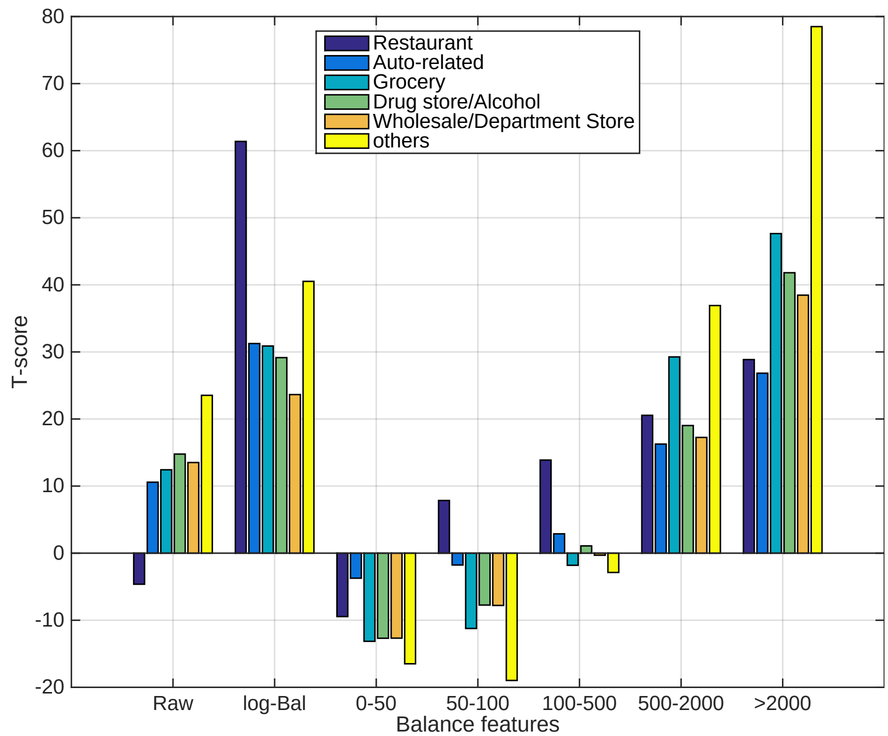

Figure 7: T-scores for linear regression coefficients of balance features. Note that the T-scores of the log-balance feature is much larger than that of the balance.

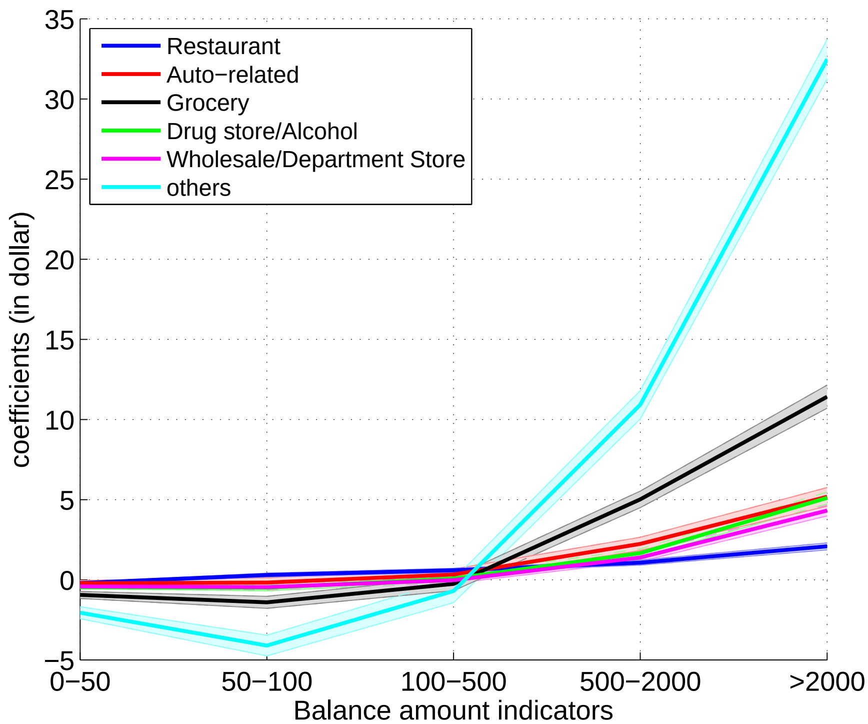

Figure 8: Linear regression coefficients of the indicator section of the balance features. The error bars are 3 standard deviation on each sides so the pointwise coverage is about 0.997.

Day from payday	Restaurant	Auto-related	Grocery	Drug store	Wholesale	others
-7	**0**	**0.044**	**4.8e-14**	**5.7e-6**	**7.7e-6**	**0**
-6	**0**	0.32	**1.7e-24**	**4.4e-32**	**2.2e-13**	**0**
-5	0.56	**8.2e-4**	**7.8e-4**	**2.6e-5**	**2.0e-9**	**0**
-4	**4.3e-4**	0.080	**1.7e-3**	0.46	**3.1e-3**	**3.1e-24**
-3	0.085	0.535	**0.026**	0.279	0.117	**1.8e-13**
-2	0.137	0.103	0.181	0.233	0.251	**1.4e-9**
-1	**3.4e-5**	0.392	0.423	**0.013**	0.240	0.90
0	**0**	**0**	**0**	**0**	**0**	**0**
+1	**0**	**0**	**0**	**0**	**0**	**0**
+2	**0**	**0**	**0**	**0**	**0**	**0**
+3	**0**	**0**	**0**	**0**	**0**	**0**
+4	**0**	**0**	**0**	**0**	**0**	**0**
+5	**0**	**0**	**2.2e-7**	**0**	**0**	**0**
+6	**0**	**0**	**0.921e-8**	**0**	**0**	**2.77e-8**
+7	**0**	0.054	**5.1e-27**	0.205	**5.6e-4**	**4.1e-11**

Table 3: p-values of each payday coefficients for each categories. Rejected null-hypotheses at 0.05 level are highlighted in boldface. We label anything smaller than 1e-32 by 0.

Balance	Restaurant	Auto-related	Grocery	Drug store	Wholesale	others
0-50	**4.0e-14**	**3.8e-6**	**7.9e-32**	**4.4e-27**	**0**	**8.7e-29**
50-100	**8.0e-15**	3.4e-4	**3.5e-23**	**5.1e-10**	**1.94e-15**	**0**
100-500	0	0.459	**0.848**	**0.011**	0.578	**0.037**
500-2000	**0**	**0**	**0**	**0**	**0**	**0**
$\geq$ 2000	**0**	**0**	**0**	**0**	**0**	**0**

Table 4: p-values of each balance coefficient for each category. Almost all balance features are statistically significant.

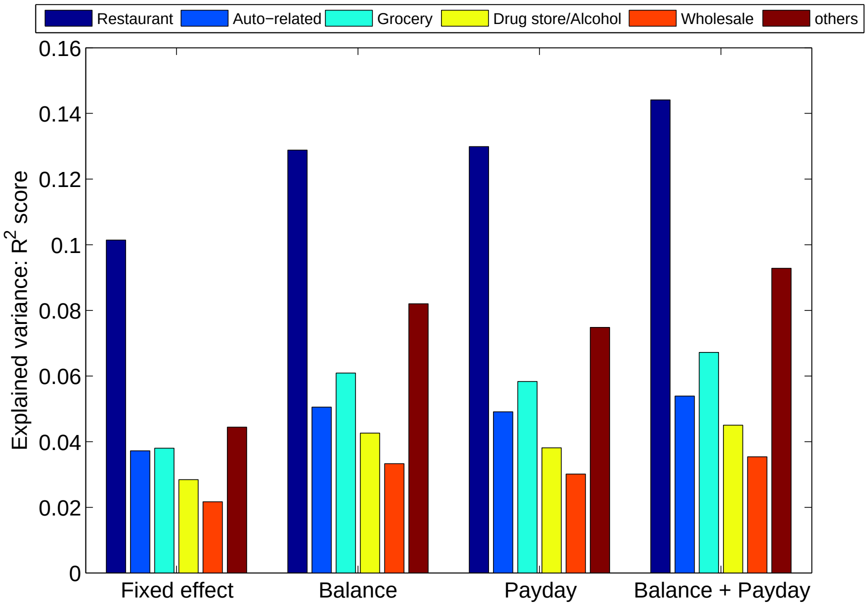

Figure 9: Explained variances of the four configurations. As we add payday features and balance features, the explained variance gets substantially larger in all categories. The R^2 score is calculated using the training data set, as per linear regression convention. Note that in our case, we have a large data set, thus evaluation on a holdout data set produces results that are almost the same.

7 Conclusion

In this paper, we studied the payday effects on prepaid card customers. We validated the existing research on payday effects by reproducing nearly the same result on a very different data set. We also improved the existing approach by adding more careful control of other variables, in particular, the balance features. The study reveals more complicated dependence on the balance than simply a "liquidity constraint". From the explained variance of the linear regression, we find that both payday effects and balance effects features can lead to significant improvements in the predictive power almost independently. The results might lead to interesting future directions in studying consumers' spending behaviors when stimulated by income, perceptual effects of balances, or expected versus unexpected changes in income. The major issue will be to identify either a valid instrumental variable or to somehow actively collect data with appropriate randomization.

Acknowledgments

I want to thank my committee members Alan Montgomery and Geoff Gordon for their support and advice that led to substantial improvement of the paper. I much appreciate supports from my family, peers and the program manager Diane Stidle.

A A preliminary investigation with nonparametric methods

As we discussed before the main drawback of a linear regression based approach is that the true conditional expectation might not be linear, and the confidence intervals as well as hypothesis tests become invalid (albeit we argued that they are still useful in gauging uncertainty). As a result, we conducted kernel ridge regression in place of the mixed effect hierarchical linear regression. Limited by the computational resources, we could not construct the full kernel matrix and solve the kernel ridge regression exactly. Instead, we used random fourier features [Rahimi and Recht, 2007] to explicitly approximate the feature expansions. Specifically, the feature map we use is

$$\phi(x) = \exp(iWx)$$

where W is an appropriately scaled (such that every coordinate has magnitude roughly 1) Gaussian random matrix, and $\exp(i\cdot)$ is the complex exponential, that can be decomposed into sines and cosines. The inner product of this finite dimensional complex vector is a uniform approximation of a Gaussian RBF kernel function evaluation. In our problem, we only mapped the low-dimensional part of the feature vector (everything except the the individual effects) to a Hilbert space, and then we concatenate the raw features with the explicitly represented random fourier features.

Usually, the number of random fourier features should be large (although still much smaller than the number of data points). But in our case, since we have about 12 million data points, representing a dense 5000×12 million dense matrix and manipulating it into the Gram matrix is not a feasible option. Therefore, for this preliminary analysis, we used only 200 random fourier features and to further reduce the dimension, we subsample the data set and use only a randomly selected 2000 customers. These allows the method run in a realistic amount of time.

Lastly, the data set is randomly split into training/validation/testing sets, each containing 50%, 25% and 25% of the data set respectively. The ridge regularization parameter is tuned on the validation set over an exponential grid of λ. Figure 10 illustrates the explained variance of the four configurations on the test set. As we can see, the explained variances are similar but slightly smaller than those of the linear model (see Figure 9). This could be due to the crude kernel approximation or the much smaller training data set. Further investigation along this line will be pursued as future work.

References

CUNA. Nilson: Credit outpaces debit, prepaid for card use. http://news.cuna.org/articles/106126-nilson-credit-outpaces-debit-prepaid-for-card-use, 2015.

Marjorie A Flavin. The adjustment of consumption to changing expectations about future income. *The Journal of Political Economy*, pages 974–1009, 1981.

Milton Friedman. *A theory of the consumption function.* Pickle Partner Publishing, 1957.

Michael Gelman, Shachar Kariv, Matthew D Shapiro, Dan Silverman, and Steven Tadelis. Harnessing naturally occurring data to measure the response of spending to income. *Science*, 345(6193): 212–215, 2014.

Robert E Hall. Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *NBER working paper*, 1979.

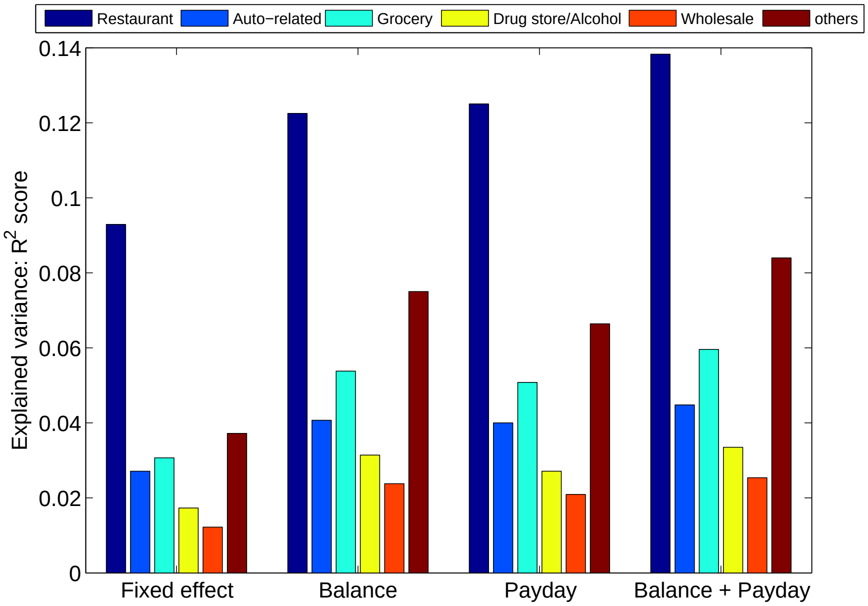

Figure 10: Explained variances of the four configurations using kernel ridge regression. The R^2 scores are calculated on the holdout data set.

IPPay. Understanding the ach network: An ach primer, 2003. URL http://www.ippay.com/downloads/ACH_101.pdf.

Tullio Jappelli and Luigi Pistaferri. The consumption response to income changes. Technical report, National Bureau of Economic Research, 2010.

Michaela Pagel and Arna Vardardottir. The only day better than friday is payday! https://server1.tepper.cmu.edu/Seminars/docs/Pagel_PaydayLiquidity.pdf, 2015.

Jonathan A Parker. The reaction of household consumption to predictable changes in social security taxes. *The American Economic Review*, 89(4):959–973, 1999.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *NIPS*, pages 1177–1184, 2007.

Matthew D Shapiro and Joel Slemrod. Consumer response to tax rebates. *The American Economic Review*, 93(1):381–396, 2003.

Matthew D Shapiro and Joel B Slemrod. Did the 2008 tax rebates stimulate spending? Technical report, National Bureau of Economic Research, 2009.

Visa Commercial Solutions. Merchant category codes for irs form 1099-misc reporting, 2007. URL https://web.archive.org/web/20070710202209/http://usa.visa.com/download/corporate/resources/mcc_booklet.pdf.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.