Real Time Measurement of Household Electronic Financial Transactions in a Population Representative Panel

Marco Angrisani

(University of Southern California, United States)

Arie Kapteyn

(University of Southern California,United States)

Swaroop Samek

(University of Southern California, United States)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 4B-2: Improving the Measurement of Household Finances in Surveys

WEDNESDAY, AUGUST 22, AFTERNOON

Real Time Measurement of Household Electronic Financial Transactions in a Population Representative Panel

Marco Angrisani,[a] Arie Kapteyn,[a,b] and Swaroop Samek[a]

[a] *University of Southern California, Center for Economic and Social Research*

[b] *National Bureau of Economic Research*

July 31, 2018

Abstract

Since July 2016, we have been collecting data on household financial transactions, balances, and expenditures from electronic records in the Understanding America Study (UAS), a population-representative household Internet panel managed by the Center for Economic and Social Research at the University of Southern California. Unlike previous studies relying on electronic transaction information from financial aggregators, our data allow us to explore heterogeneity in consumer behavior as driven by demographics, health, cognitive ability and financial literacy. In this paper, we first document respondents' participation in the study. We then examine spending behavior during pay cycles. Finally, we compare self-reported expenditure measures to actual measures obtained from transactional data for a variety of spending categories.

1 Introduction

To date, the measurement of financial and consumer behavior mainly relies on self-reports. Such reports are burdensome for respondents and tend to suffer from serious measurement and selectivity issues. Since most individual financial and spending behaviors (use of payment instruments, investments, assets and liabilities) are recorded in electronic form, direct access to electronic records may both improve accuracy of measurement and greatly reduce respondent burden.

In this paper, we present first results of a project in which we collect financial electronic records directly through the financial institutions of respondents in a household panel. We have designed and implemented a data collection system, using a contract with a major financial information aggregation firm, and asked the panel members' consent to share with us their financial information through the aggregator platform.

Respondents in our study are members of the Understanding America Study (UAS) Internet panel (https://uasdata.usc.edu/index.php). The UAS is a population-representative household panel, implemented by the Center for Economic and Social Research (CESR) at the University of Southern California (USC), of currently approximately 6,000 respondents. Panel members are recruited by address-based sampling. Respondents without prior Internet access are provided tablets and broadband Internet access if needed. Panel members answer questions once or twice a month. Fundamental to this project, the surveys collect a large amount of information on all individuals, which can be combined with newly collected data. For every respondent, we collect about four hours worth of information every two years on their health, economic status, cognitive capabilities and personality, among others. By relying on the UAS, we are able to combine survey information with information from electronic records. Hence, unlike existing studies relying on electronic transaction information from financial aggregators, our data allow us to explore heterogeneity in consumer behavior as driven by a wide range of demographics.

In what follows, we describe (Section 3) the main features of this innovative data collection project and document respondents' recruitment and participation in the study. In Section 4, we provide descriptive statistics of collected transaction data and compare them to those reported by Gelman et al. (2014), a pioneering contribution exploiting data from a financial aggregator. In Section 5, we examine individuals' spending behavior over time. Specifically, we investigate whether expenditure exhibits pay-cycle patterns and identify for which spending categories such patterns are more pronounced. Finally, in Section 6 we compare the distributions of self-reported monthly expenditure and actual monthly expenditure obtained from transactional data. This exercise allows us to assess the accuracy of self-reports across spending categories.

2 Background

A number of ongoing U.S. studies collect financial and expenditure information from households, including the Survey of Consumer Finances (SCF), the Health and Retirement Study (HRS), the Panel Study of Income Dynamics (PSID), the Consumer Expenditure Survey (CE), and the Survey of Income and Program Participation (SIPP). These studies are used in a wide variety of research projects. Typically, respondents are asked a large number of questions about their expenditures, financial holdings, and/or financial behavior. Generally, questions about their finances or spending are among the least popular topics for respondents to answer. For instance, in an "end of year survey" we conducted in the American Life Panel, 59% of the respondents chose economic and financial surveys as the category of surveys they liked least, while only 10% found them the most interesting. Detailed questions about finances take a long time to answer. For example, the median interview length was about 90 minutes in the 2010 SCF, but some interviews took more than three hours (Bricker et al., 2012). Despite the importance and value of such data, several studies have assessed the quality of wealth information in major U.S. surveys with generally less than satisfactory conclusions (Bound et al., 2001).

2.1 Problems with measurement of economic variables in surveys

Psychologists and survey methodologists have studied how individuals answer survey questions. They have developed theories about cognitive processing and social interaction that model the chain from question to answer. Based on these insights, they have provided practical guidance that, when implemented, has the potential to greatly improve the quality of survey answers. Overviews of this literature have been given by Groves et al. (2009) and Tourangeau and Bradburn (2010), among others. The typical model for the process of answering a question consists of four stages: (1) question comprehension; (2) information retrieval; (3) judgment and estimation; and (4) reporting. In each of these stages, there is a potential for error. An example of error in comprehension is when a respondent is asked about expenditures on restaurant meals and fails to realize that this includes fast food restaurants. An example of error in retrieval is that the respondent understands that fast food restaurants are included, but forgets about the lunch on a Tuesday three weeks ago. Estimation happens when the number of items (income sources, expenditures) asked about is large. Respondents will use heuristics such as estimating a typical amount and a typical frequency and estimate a total from this. Errors in this stage may occur if "typical" is not "average" or if there are computational errors. An example of an error in the reporting stage is a typo (by the interviewer, or by the respondent in self-interview mode). If the question is too hard or ambiguous, it is also possible that respondents refuse to answer or give a "don't know" answer.

Economic variables often differ from many other types of variables that surveys may try to measure. First, they are mostly factual concepts, involving exact amounts of money, instead of opinions, attitudes, or other subjective concepts. Thus, in principle there exists a true value that economic surveys aim to measure. Second, these economic variables are often quantitative numerical concepts, which, for all practical purposes, can be considered continuous (e.g., the monetary value of an asset), and for which we are interested in the exact number. Third, there are typically issues of aggregation and categorization. Annual income consists of a stream of income amounts during shorter periods (e.g., months) and from multiple sources (e.g., earnings from one or more jobs; Social Security or unemployment benefits; asset returns). Wealth and expenditures also consist of many components. Fourth, there is a large amount of potential conceptual ambiguity (e.g., whether pre-tax pension contributions should be excluded from income, or whether food stamps are part of income). Finally, many individuals simply do not know the amounts to any degree of accuracy, even if they understand the concepts.

Thus, some of the challenges that the survey measurement of financial and expenditure behavior poses can be summarized as follows:

- **Definition of concepts**

  The respondent may not accurately perceive the concept the investigator is interested in. For example, are we interested in before-tax or after-tax income, is alimony considered part of income, should retirement accounts be included in total wealth, and so forth? Defining concepts precisely is easier with narrow categories, but this implies asking more questions and thereby increasing respondent burden. Furthermore, narrow categories may lead to double counting (e.g., stocks held in IRAs included both in the IRA component and the stocks component). With access to the account information, the investigator is able to define concepts precisely and extract the relevant information without undue respondent burden.

- **Failure to recall all components**

  During the interview, a respondent may not think about rarely used accounts or other unusual assets, fail to report incidental income components, or not remember certain expenses. By asking about many categories, components that would otherwise likely be missed are included. Studies that ask income, expenses, or wealth in multiple categories have higher averages than studies that ask only broad totals (Kapteyn et al., 1988; Browning et al., 2014). This is often interpreted as evidence that multiple categories are "better", and sometimes there is direct evidence for this (Winter, 2014). However, Menon (1997), studying frequencies of various types of behavior, found that decomposition into narrow categories made reports of irregular frequent behavior more accurate, but reports of regular frequent behavior less accurate. Conversely, Belli et al. (1997) found that decomposition leads to over-reporting of non-distinctive frequent events.

- **Incorrect timing**

  When asking about income or expenses in a certain reference period, the respondent may inadvertently include income that was received or expenses that were incurred shortly before or after the reference period. The potential for errors is generally larger with longer reference periods or reference periods in a more distant past. This would appear to be particularly problematic for panel studies like the PSID and HRS, which have a periodicity of two years. In that case, one faces a choice between shorter reference periods than two years, so that the history of a variable has gaps, or to ask for the whole two year history in each wave and accept the implied deterioration in measurement quality. For the purpose of accurately recording information, reference periods should ideally be short and recent (e.g., "yesterday"). However, such short reference periods lead to substantial random variation (Hurd and Rohwedder, 2009). Angrisani et al. (2014) document that the volatility of household expenditures decreases monotonically with the length of recall periods. They find that volatility is not the only outcome that varies across reporting periods. For instance, when confronted with specific time spans (e.g., last Wednesday, last 7 days, last 30 days, last 12 months), the implied annual amount spent is higher when individuals refer to day or week than to month or year. This pattern is stronger for debit cards and cash than for credit cards and checks. Without further information, it is not possible to decide which periodicity (day/week/month/year) is best to elicit individual/household expenditures.

- **Incorrect recall or report**

  Respondents may misremember amounts, make rough guesses about amounts that they do not precisely know, or give rounded amounts. Furthermore, respondents and interviewers may make typographical errors, or interviewers may mishear what respondents say.

Studies examining measurement error in economic data typically compare survey reports with validation sources: administrative records that have been matched to the survey data. According to (Bound et al., 2001), survey-reported annual earnings have less measurement error than hourly wages or weekly earnings, but they may be mean-reverting, that is, low earnings are overreported and high earnings underreported. However, Kapteyn and Ypma (2007) find that a small fraction of mismatch between the administrative data and the survey may account for this apparent mean reversion. Transfer income is typically underreported, due to both underreports of receiving the source of income at all and underreports of the amounts conditional on receiving. Finally, transfer income is subject to seam bias, in which there is an over-report of transitions in and out of programs coinciding with the start of the period that the current wave asks about (the "seam" between the previous wave and the current wave) and an underreport of such transition within the period asked about.

2.2 Consequences of measurement errors in economic variables

Due to these and other problems, the measurement of income in surveys is far from perfect. Moore et al. (1999) document consistent and often large shortfalls in nationally weighted survey estimates as compared to independent benchmarks, as well as a considerable level of nonresponse error in income reports. Inaccurate income measures may lead to biased estimates of key economic parameters, which, in turn, may be used to poorly inform policy interventions. For instance, survey data make it difficult to disentangle transitory variations in labor income from changes attributable to reporting error (Meghir and Pistaferri, 2004). This may prevent a precise assessment of the degree of income risk faced by individuals and hamper the design of programs aimed at reducing it.

Much less is known about measurement error in survey measures of assets, because fewer validation sources exist for assets. Nevertheless, from the existing evidence, Moore et al. (1999) conclude that assets are typically underreported, both due to failure to report owning an asset at all, and to underreports of the amounts. Attempts to aid respondents by providing ranges they can choose from lead to anchoring effects (Hurd, 1999): respondents tend to use the provided ranges as information about what a reasonable value is for their own amount, and thus tend to gravitate towards these "anchors." Juster et al. (1999)) show that even if the measurement of cross-sectional wealth is satisfactory, this may still need further improvement if the goal is to use the between-wave changes in wealth to study wealth accumulation (including saving) and decumulation, because the effects of measurement error in levels on empirical estimation of economic models are exacerbated by taking first differences.

Measurement errors in recall consumption data (Browning et al., 2003; Carroll et al., 2014) may greatly affect the estimation of consumption models and the testing of theoretical predictions (Attanasio, 2000). Typically, measurement errors in consumption imply that instruments lagged one period are invalid when estimating Euler equations. Conversely, instruments lagged two or more periods may be too weak.

2.3 Use of aggregator data

More recently, authors have used data obtained by financial aggregators. Gelman et al. (2014) illustrate the potential of financial aggregator data for improving our understanding of economic questions that are key to public policy. They use data collected by Check (https://check.me) on transactions and balances of a sample of more than 75,000 users, containing comprehensive information on income as well as spending. Exploiting such data, they study the responsiveness of spending to the arrival of anticipated income. Previous studies that have looked at this question using data from the Consumer Expenditure Survey lack precision (Souleles, 1999; Parker, 1999), while other studies using administrative data are not as comprehensive (e.g., Agarwal et al., 2007 exploiting data from a single credit card company only).

Although the work by Gelman et al. (2014) illustrates the enormous potential of these types of data, it also helps us understand the limitations. One of the challenges with the data is selectivity. For instance, individuals who voluntarily sign up to use Check's service are more likely to be male, to be 25-44 years of age, and are less likely to have a graduate degree than the average person living in the U.S. Another issue with data from financial aggregators is that they provide a partial picture of the financial transactions and balances of a couple or household. The individual signs up accounts that s/he has with different financial institutions, including joint accounts s/he may share with a partner. However, the data do not include the individual accounts of one's partner. Finally, financial aggregators do not collect or share demographic information. This greatly limits the range of issues that can be studied with such data. Our approach, exploiting rich background information and self-reports at both individual and household level, aims at addressing all three of these limitations.

After the pioneering work by Gelman et al. (2014), several other authors have relied on data from financial aggregators to study a range of economic questions, from spending responses to both regular and irregular income (Olafsson and Pagel, 2018) or predetermined payments (Kueng, 2015), to interaction between household balance sheets, income, and consumption during the Great Recession (Baker, 2014) and debt repayment behavior (Kuchler, 2015). The data limitations described above, which our project aims to overcome, all apply to these studies.

3 Our Approach: Combining Survey and Aggregator Data

As noted, the context for our proposed data collection effort and research project on financial behavior is the Understanding America Study (UAS), a household panel managed by USC of approximately 6,000 households representing the entire U.S. The study is an "Internet Panel," which means that respondents answer surveys on a computer, tablet, or smartphone, wherever they are and whenever it is convenient for them to participate. From the viewpoint of representativeness, it is important to note that the panel is recruited through address-based sampling. Anyone willing to participate, who does not have a computer or Internet access, is provided with a tablet and broadband Internet.

Panel members answer surveys about once or twice a month. Surveys are restricted to about 30 minutes per interview, but since all data can be linked across surveys, a large amount of information is available about panel members, including demographics, health, financial behavior and financial literacy, cognitive capability, and personality. Of specific interest for this study is that all panel members answer the full survey instrument of the HRS, which contains some 130 minutes worth of interview time (administered over a number of separate sessions to stay within the 30 minute limit per survey) with information on health, income, assets, labor market position, and expectations. Respondents receive compensation for their time spent answering questions at a rate of $20 per 30 minutes of interview time. Annual attrition rates are modest (on the order of 6-7% per year).[1]

The UAS is a probability-based Internet panel. That is, respondents are drawn from a well-defined sampling frame (U.S. Postal addresses) with known inclusion probabilities. Probability Internet panels have to be distinguished from convenience Internet panels, where respondents are recruited from among existing Internet users by placing banners on web-sites to invite respondents, and inclusion probabilities and sampling properties are unknown. Several studies have shown that probability Internet panels and convenience panels differ fundamentally in the quality of information they provide about the U.S. population. Chang and Krosnick (2009) administered the same questionnaire to a telephone sample, an Internet probability sample, and a non-probability sample of volunteers who do Internet surveys for money. They found that the telephone sample has the most measurement error, while the non-probability

[1] The UAS webpage (https://uasdata.usc.edu/index.php) provides full details about recruitment, response rates and attrition.

(convenience) sample exhibits most bias. On balance, the probability Internet sample produced the most accurate results. Yeager et al. (2011) reached similar conclusions. They also found that response rates are not critical for achieving accurate estimates. Even with relatively low response rates, probability samples yield unbiased estimates. Non-probability samples do much worse. Moreover, weighting and matching do not eliminate the differences between estimates based on samples of respondents with and without Internet access (Schonlau et al., 2009). A recent contribution by Angrisani et al. (2018) examines sample characteristics and elicited survey measures of HRS and UAS. Comparison of a variety of survey outcomes with population targets shows a strikingly good fit for both studies, with outcome distributions in the HRS only marginally closer to population targets than outcome distributions in the UAS.

3.1 Aggregator data

Firms such as Mint (https://www.mint.com), Yodlee (http://www.yodlee.com) and Check (https://check.me) provide financial aggregation services to consumers.[2] An individual signing up for any of these aggregators lists his/her various accounts with financial institutions and shares passwords with the aggregator. The software of the aggregator (which reflects agreements with financial institutions) then combines the information from the various accounts and provides overviews of spending in broad categories, use of payment instruments and balances. Users receive weekly overviews and alerts if spending or changes in balances exceed some pre-specified trigger levels. Most major banks provide similar services to their own customers, but by necessity that information is only based on the data contained in the bank's own records. The backend of banks' services is often provided by Yodlee, whose platform is used by 7 of the 10 top U.S. banks, more than 700 global financial institutions, and over 50 million consumers.

Based on Yodlee's capabilities and discussions about the services it can

[2] In 2014, Mint owner, Intuit, purchased Check. Since Spring 2017, Mint and Check are integrated in the Mint platform.

provide, we have invited UAS panel members to join Yodlee, enabling the UAS to gain access to the financial information that Yodlee collects. We have asked consent from the panel members to sign up to Yodlee and for their permission for Yodlee to share their financial information with us. Panel members are incentivized by receiving monetary compensation for their effort.

We obtain information on every electronically recorded financial transaction of UAS panel members who have signed up for our project directly from Yodlee. For the recruitment (from among UAS members) and maintenance of the sample that signs up with Yodlee, we have varied certain parameters experimentally to gauge their effect on participation and data quality. The Yodlee data are supplemented with information collected from surveys of the UAS respondents, both to assess and improve the quality of the collected data and to elicit information that is particularly powerful for answering basic research questions in combination with actual transactional data.

3.2 Recruiting respondents

Before starting the actual project, we fielded a consent survey among members of the UAS to gauge willingness to participate in the study. Results indicated that about 60% of the respondents were interested in participating. We also varied the level of promised incentives for participation, but found no significant effects. Given the budget available for the project and the anticipated 60% response rate, we decided to invite 1,110 panel members to join the study, expecting to eventually have a sample of about 600-650 respondents. It should be noted that participation involves a number of steps to be taken by the panel members; at each step there is potential for attrition. These steps are:

1. Provide consent to participate.
2. Create an account on the financial management web-site (the web-site was created by us, but connected to the financial aggregator system).
3. Once an account has been created, the respondent needs to add financial institutions to the account, which includes the sharing of passwords with

the financial aggregator.

4. The accounts need to be kept up-to-date (e.g, if the account password was changed with the bank, the password needs to be changed with the aggregator for it to continue accessing account data).

The 1,100 selected UAS respondents received the following invitation (the amounts varied experimentally across respondents, as will be explained below):

We are interested in how Americans spend their money and how they are doing financially. We would ask you to sign up with a custom made financial management web-site. The web-site has been developed in collaboration with one of the biggest financial management service companies in the world: Yodlee. For instance, Yodlee provides services to 12 of the 20 largest banks in the United States.

We will NOT have access to your passwords or any other identifying information*; this information will be safeguarded by Yodlee. We will use the data in the same way we use surveys you participate in: to make summary tables or graphs to better understand how Americans are doing. Just like the information you provide through surveys, you will be compensated for the information that you share with us.*

If you agree to participate, we will pay you $25 just for signing up with the financial management web-site, plus $5 for every one of your financial institutions that you add on the web-site.

For example, imagine you have a checking and a savings account with one financial institution, a credit card with another, a brokerage account with another financial institution, and a retirement account (such as a 401(K) or IRA) with yet another one. That means you have a total of 5 accounts at 4 financial institutions. You will earn $20 (4 x $5) if you sign all 4 of your financial institutions up.

Every month after that, we will pay you $2 per institution that you signed up to the web-site. That means the earlier you sign up all of your financial institutions, the sooner you can start earning money, just for letting Yodlee summarize information about your accounts for us. You'll get the monthly *amount as long as you keep information about each institution current in the system.*

Out of the 1,110 invitees, 1,107 answered the consent form and 508 stated they would be willing to take part in the study for a 46% consent rate, which is lower than the anticipated 60% obtained from an earlier hypothetical question. Possibly, some respondents, who initially were interested in potentially participating, changed their mind when they were actually asked to commit. Table 1 compares the characteristics of consenters with the characteristics of the entire UAS pool. As can be seen, consenters tend to be younger and to have lower incomes. The education composition is rather similar in the two groups.

Table 1: Study Participation: Consenters' Characteristics

	Consenters	All UAS
Age		
18-34	21.85	19.36
35-44	24.41	18.80
45-54	17.72	19.18
55-64	21.85	22.64
65+	14.17	20.02
Education		
High School or Less	24.02	24.10
Some College	40.75	38.49
Bachelor or More	35.24	37.41
Income		
$< \$30k$	32.87	25.03
$\$30k - \$60k$	24.80	26.77
$\$60k - \$99k$	22.44	23.38
$\$100k+$	19.88	24.82

Once respondents consented, they were asked to create an account on the financial management web-site. Out of the 508 respondents who consented, only 350 respondents actually created an account ("signed up"). Hence, the unconditional sign-up rate was about 32%. There may be several reasons why respondents who initially consent to participate do not create an account. One reason is that respondents find this more cumbersome than anticipated; another possibility is that the process of creating an account is another opportunity for reflection on perceived risks and, hence, for a decision to not proceed further.

Table 2: Determinants of Participation: Demographics

	Consent	Unconditional Sign-Up	Sign-up Conditional on Consent
Male	-0.003	-0.017	-0.042
Non-White	-0.047	-0.058*	-0.070
Age 35-44	0.016	-0.011	-0.035
Age 45-54	-0.071	-0.051	-0.010
Age 55-64	-0.128**	-0.151**	-0.130**
Age 65+	-0.214**	-0.187**	-0.162**
Some College	0.085**	0.108**	0.151**
Bachelor or More	0.088**	0.128**	0.198**
Working	-0.034	-0.021	-0.033
HH Income 30-60k	-0.083**	-0.040	0.026
HH Income 60-100k	-0.092**	-0.021	0.074
HH Income 100k+	-0.050	0.013	0.109
N	1,106	1,106	508

Table 2 shows results of regressions of consent and sign-up indicators on background variables. The consent equation confirms the earlier observation that younger respondents and respondents with lower incomes are more likely to consent. We now also see that education has a significant effect: the probability of consent goes up with education. Working has a non-significant negative effect. In itself, a negative sign of the working indicator would be consistent with the idea that respondents who work will have less time available to participate. Wenz et al. (2017) hypothesize that time constraints may reduce willingness to participate in extra tasks, but, similar to our finding, they also estimate only small and insignificant effects of time constraints. The regression results in the third column (sign-up conditional on consent) show similar patterns for age and education. Income is not significant in this regression.

Thus, the overall picture suggests that young and highly educated respondents are most likely to consent and, conditional on consent, to sign up. Income only has an effect on the likelihood to consent, but not on the likelihood to sign up given consent. In addition to demographics, one would expect the probability of consent or sign-up to be related with online banking experience. Table 3 confirms that to be true. Respondents who are active in internet banking are more likely to consent. However, conditional on consent, sign-up appears to be only related to whether one checks balances online.

Table 3: Determinants of Participation: Online Banking Indicators

	Consent	Unconditional Sign-Up	Sign-up Conditional on Consent
Account at Fin Inst	-0.137**	-0.050	0.029
Internet Bill Payment	0.126**	0.039	-0.078
Internet Balance Check	0.032	0.142**	0.277**
Internet Transfers	0.109**	0.107**	0.084
N	1,106	1,106	508

Once respondents have signed up, they need to add financial institutions to their dashboard. This is a critical step, without adding institutions, which implies sharing username and password with the financial aggregator, no transactional data can be retrieved. Out of the 350 respondents who had signed up (created an account), only 135 actually linked at least one financial institution. Table 4 compares the respondents who added financial institutions to the UAS as a whole. It confirms the earlier observations regarding consent and sign-up that younger respondents are more likely to link financial institutions than older respondents. We observe that the final sample of respondents who have linked financial institutions tend to be higher educated and to have higher incomes than the overall UAS.

3.3 Incentive experiments

There are various considerations in designing an incentive scheme for UAS panel members to join our study and share information with the financial

Table 4: Characteristics of Those Who Linked Financial Institutions

	Linked Fin.Inst.	All UAS
Age		
18-34	24.44	19.36
35-44	25.19	18.80
45-54	18.52	19.18
55-64	20.74	22.64
65+	11.11	20.02
Education		
High School or Less	12.59	24.10
Some College	37.78	38.49
Bachelor or More	49.63	37.41
Income		
$< \$30k$	20.00	25.03
$\$30k - \$60k$	24.44	26.77
$\$60k - \$99k$	27.41	23.38
$\$100k+$	28.15	24.82

aggregator. The incentives should be large enough to make it attractive for panel members to participate, with minimal distortion of reporting behavior. In order for respondents to sign up for Yodlee, a one-time incentive is needed. However, we also want to make sure that respondents provide as complete information as possible, which suggests that the incentive should be higher as more accounts (or rather more financial institutions) are involved. A third consideration has to do with attrition vs. change. For instance, if respondents change a password for any of their accounts, if they move an account to another, new institution, or if they open a new account with a new institution, those changes need to be communicated to the financial aggregator. Hence, we need to incentivize respondents to keep the information up-to-date. Based on these considerations, we have tested two incentive schemes, each characterized by different incentive combinations as summarized in Table 5. We adopted Scheme I for the first batch of invited members. Under this scheme, the sign-up incentive is relatively generous and can be \$10, \$25 or \$50, while adding an institution was rewarded with either \$2 or \$5. The incentive to maintain the information up-to-date was either $1 or $2. Given the low fraction of individuals who had linked a financial institution after creating an account within the first batch of respondents, we decided to move to Scheme II, by decreasing the sign-up incentive to $5 and increasing the reward for adding an institution up to $10. Incentive Scheme II is the one currently in place.

Table 5: Incentive Schemes

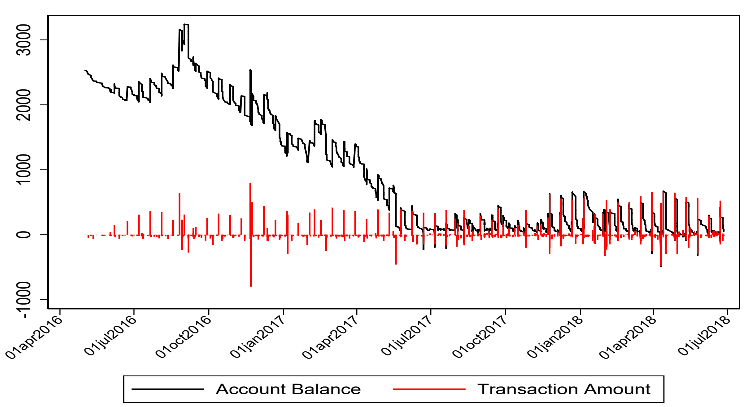
Scheme I (old)			
	Sign-Up	Add Institution	Monthly Payment
Treatment 1	$10	$2	$1
Treatment 2	$25	$2	$1
Treatment 3	$50	$2	$1
Treatment 4	$10	$5	$2
Treatment 5	$25	$5	$2
Treatment 6	$50	$5	$2
Scheme II (current)			
	Sign-Up	Add Institution	Monthly Payment
Treatment 1	$5	$3	$1
Treatment 2	$5	$5	$1
Treatment 3	$5	$10	$1
Treatment 4	$5	$5	$2
Treatment 5	$5	$10	$2
Treatment 6	$5	$15	$2

We have checked whether different monetary incentives have produced different participation behavior and found no clear patterns. For instance, among the 99 respondents subject to incentive Scheme I, the probability of creating an account with Yodlee does not vary with whether the sign-up monetary reward is $10, $25, or $50 (it is 36%, 36% and 33%, respectively). For the 35 respondents in Scheme I who created an account with Yodlee, the likelihood of linking a financial institution is higher (53%) when the monetary incentive to add an institution if $5 rather than $2 (39%), but this difference is not statistically significant. Considering the 315 participants who signed-up with Yodlee among those subject to incentive Scheme II, the probability of adding a financial institution is 37% when the incentive to add an institution is \$3, 40% when the incentive is \$5, 34% when the incentive is \$10, and 42% when the incentive is \$15. We have tested all possible combinations pair-wise and found no evidence of statistically significant differences across different monetary rewards.

4 Descriptive Statistics of Transaction Data

Before proceeding with the analysis of the transaction data, we take the following sample selection steps. We start with the 135 respondents who have added financial institutions. We drop all account entries with missing transaction date and transaction amount (0.25% of the entire data set). In doing this, we lose one respondent who linked accounts for which the data were never retrieved. Finally, we eliminate accounts with less than 10 observations or with data covering less than a month. After this, we are left with 130 individuals whose accounts are observed from September 1, 2016 through June 27, 2018. The remainder of the analysis is based on this sample, unless otherwise noted. Figure 1 illustrates what individual data look like. It shows the balance and the transaction amount of a checking account of a study participants over the observation period.

Figure 1: A Respondent's Checking Account Over Time



We begin by comparing the features of key outcome variables in our sample with those reported by Gelman et al. (2014). In Table 6, the total number of accounts per respondent is similar across the two studies, albeit slightly lower in our study. The breakdown by account type reveals that the number of checking accounts is similar. The main differences are observed in the number of savings accounts, which is larger in our sample than in the Gelman et al. (2014)'s data set, and in the number of credit cards, which is substantially larger in Gelman et al. (2014) than in our sample. As far as the number of daily transactions is concerned, it is apparent that individuals in the Gelman et al. (2014)'s data set are more active. These observed differences are likely due to selection. The Gelman et al. (2014) study is based on data from participants in Check.me. These individuals, who decide to sign up with a financial aggregator, are likely more financially savvy and probably feel a greater need for the services of an aggregator, due to the complexity of their financial portfolio and number of transactions.

Table 6: Number of Accounts and Daily Transactions

	Gelman et al. (2014)		Our Study	
	Mean	Median	Mean	Median
Accounts				
Total	5.84	5	5.19	4
Checking	1.35	1	1.71	1
Savings	0.79	1	1.76	1
Credit card	3.58	3	2.47	2
Daily Transactions				
Total	4.54	3	2.17	2
Checking	3.03	2	1.06	1
Savings	0.22	0	0.08	0
Credit card	1.23	1	0.76	0

The comparison of account balances in Table 7 reflects the similarities and differences described above. Specifically, checking account balances appear comparable in the two studies, while individuals in our sample keep twice as much in savings accounts compared to those in Gelman et al. (2014)'s data set. Both credit card balances and limits are substantially larger in Gelman et al. (2014).

Table 7: Account Balances

	Gelman et al. (2014)		Our Study	
	Mean	Median	Mean	Median
Checking	6,969	1,400	6,678	1,273
Savings	4,476	400	8,919	1,056
Credit card	7,228	3,600	3,073	1,537
Credit limit	23,019	11,900	11,246	9,000

Unlike Gelman et al. (2014), we have individual-level background information at our disposal. We take advantage of that and present, in Table 8, a breakdown of account balances by education. As one would expect, balances increase monotonically with education, with the exception of credit card balances that appear rather homogenous across education groups (slightly lower among individuals with a Bachelor's degree than among those without or more than a Bachelor's degree).[3]

Table 8: Account Balances by Education

	< Bachelor		Bachelor		> Bachelor	
	Mean	Median	Mean	Median		
Checking	-504	737	4,052	1,517	21,906	2,249
Savings	2,779	1,410	12,407	627	10,921	2,005
Credit card	3,055	1,517	2,864	1,417	3,424	1,759
Credit limit	8,973	6,000	11,570	10,000	14,390	12,000

Table 9 presents means and medians of monthly income, monthly total expenditure and a few expenditure components. The reported statistics indicate that individuals tend to spend their entire regular income. The saving rate out of total income is about 18% $(1 - 4,558/5,534)$. This implausibly high saving rate suggests that individuals may not link all their accounts, so that some

[3]The negative mean of checking account balance among individuals with no Bachelor's degree reported in Table 8 is due to one account outlier. This is a checking account which is first observed at a balance of -$115,000 and remains negative for most of the observation period. After excluding this account, average (median) balance in the sample is $1,564 ($728).

expenditures may not be captured. It should be noted that we do observe ATM withdrawals and check payments, so our definition of total expenditure does take into account non-electronic payments.

Table 9: Monthly Income and Expenditures

	Mean	Median
Salary and Regular Income	4,447	3,794
Total Income	5,534	4,375
Total Expenditures	4,558	1,904
Grocery Expenditures	321	189
Restaurant Expenditures	262	144
Merchandise Expenditures	492	291
Automotive Expenditures	292	153

Total Income includes: salary and regular income, investment and retirement income, interest income, and other income.
Total Expenditures include: all expenditures except service fees associated with retirement and investment accounts, mortgages and loans.

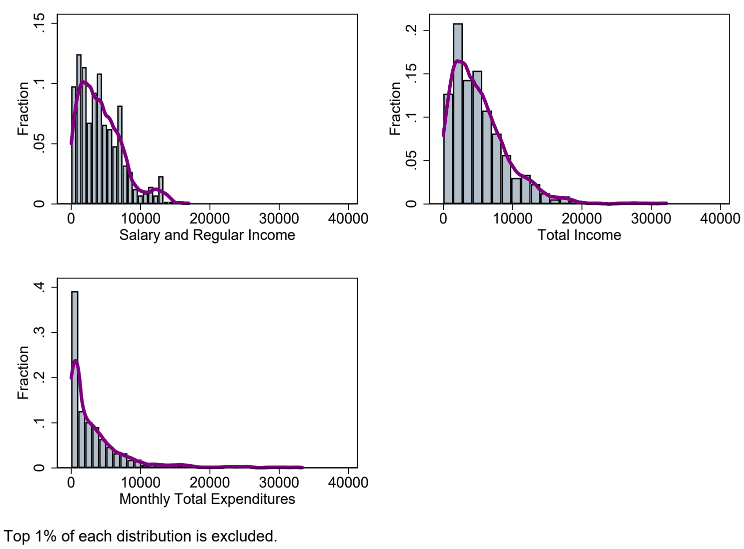
Table 10: Monthly Income and Expenditures by Education

	< Bachelor		Bachelor		> Bachelor	
	Mean	Median	Mean	Median	Mean	Median
Salary and Regular Income	2,715	1,984	4,591	4,055	5,818	4,960
Total Income	3,626	2,602	5,280	4,577	7,599	5,524
Total Expenditures	2,365	992	4,528	2,339	9,423	4,086
Grocery Expenditures	316	146	281	170	397	338
Restaurant Expenditures	204	100	267	165	363	237
Merchandise Expenditures	453	238	470	287	614	444
Automotive Expenditures	199	120	367	151	325	183

Total Income includes: salary and regular income, investment and retirement income, interest income, and other income.
Total Expenditures include: all expenditures except service fees associated with retirement and investment accounts, mortgages and loans.

Table 10 shows monthly income and expenditures by education. Income and expenditures are higher on average for respondents who have a Bachelor's degree or more. The implied saving rate out of total income for respondents without a college degree is 35%, while for respondents with a Bachelor's degree is virtually zero. This may suggest that the higher educated group may be more likely to add accounts than the lower educated group. The group with more than a Bachelor's degree spends, on average, about 24% more than income. This result is mainly driven by one respondent for whom large check payments are recorded over the observation period. If we exclude this respondent, average (median) total income among those with more than a Bachelor's degree is \$9,307 (\$4,877) and average (median) monthly expenditure is \$7,285 (\$3,941). It is worth noting that for total expenditures there is a much larger difference between mean and median than for income, indicating that the distribution of expenditures is much more skewed than that of income. The data also show substantially larger variation for expenditures than income. The distributions of monthly salary/regular income, total income, and total expenditure are reported in Figure 2.

Figure 2: Monthly Income and Expenditure Distribution



5 Expenditure Smoothing

An issue of perennial interest in economic models of consumption is the extent to which consumers smooth consumption over time, as they are assumed to do in life-cycle models. To take a first cut at this we consider the subset of respondents with clearly identifiable pay days and investigate to which extent expenditures are different before and after pay day. For now, this is just an illustrative exercise, as expenditures do not need to coincide with consumption. Moreover, for larger items (e.g. rent), payment patterns may be synchronized with income receipt, without affecting consumption (housing services, in this example).

The subsequent analysis requires a number of assumptions and data manipulations. We focus on transactions classified as "salary and regular income." First, for each individual we identify a "payment date." That is, a date when a transaction classified as salary/regular income is recorded. If multiple payment dates exist within a two-week period, we ignore those for which the amount of salary/regular income received is less than 25% than the maximum amount received within the two-week period. Second, we compute the distance between two subsequent payment dates. Third, we define the individual-specific frequency of payment as the within-individual median of distances between two successive payment dates.

Out of 41 respondents for whom we observe salary and regular income payments over time, 24 are paid every 14 days, 2 every 13 days and 3 every 3 days. In our analysis, we will only use these 29 individuals, assuming they are paid bi-weekly. As far as the others are concerned, 6 are paid every 7 days, 2 every 8 days, 1 every 10 days, 1 every 12 days, and 2 every 30 days.

We observe these 29 individuals over the period September 1, 2016 – June 27, 2018. We consider four types of expenditures: grocery, restaurant, merchandise, and automotive.[4] For each individual, we compute average expenditures over the observation period and then take the ratio of each date's expenditure to the average expenditure. This ratio, defined for each expenditure category, is the dependent variable in our analysis.

We run regressions of expenditure ratios on a set of dummies for the 7 days before the payment date, the payment date itself, and the 6 days after

[4]The category "merchandise" may comprise a variety of goods. The transaction description for this category typically features big retailers such as Amazon, Walmart, etc.

the payment date (and before the next pay cycle starts). The omitted dummy is the one for the day before payment day. We also include day-of-the-week dummies. Standard errors are clustered at the individual level. The figures in this section show the estimated coefficients from these regressions alongside with their 95% confidence intervals.

Figure 3: Grocery Expenditure
Fraction of Daily Average Spending

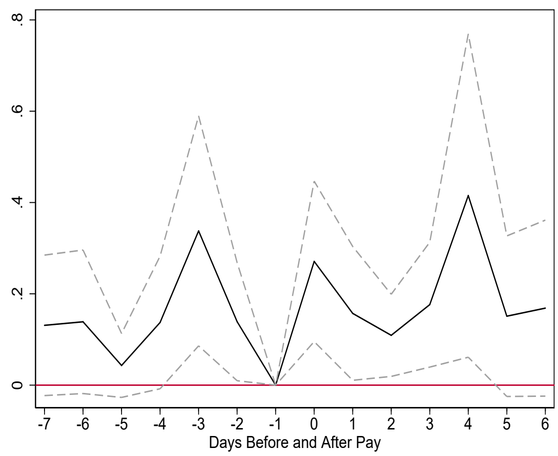
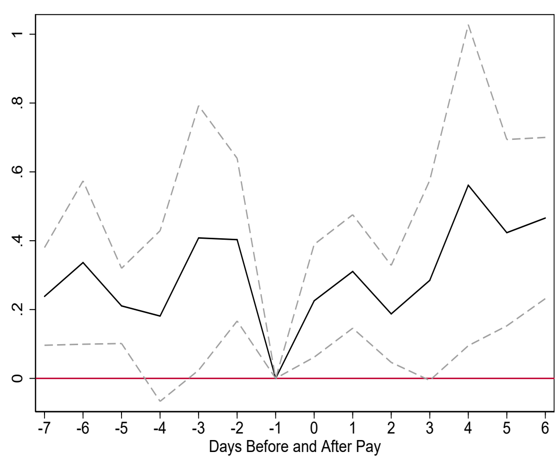


Figure 4: Restaurant Expenditure
Fraction of Daily Average Spending



Figures 3-6 report the estimates for the four expenditure categories. Given the way the coefficients are estimated, the dependent variable is always exactly zero on day -1. It is striking to see that for all four expenditure categories, the day before pay day shows the lowest spending level.

Figure 5: Merchandise Expenditure
Fraction of Daily Average Spending

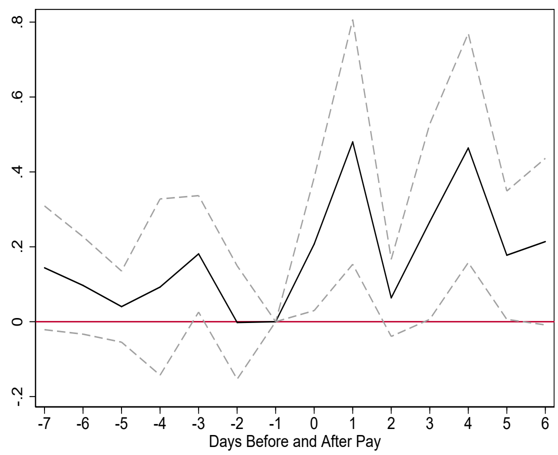
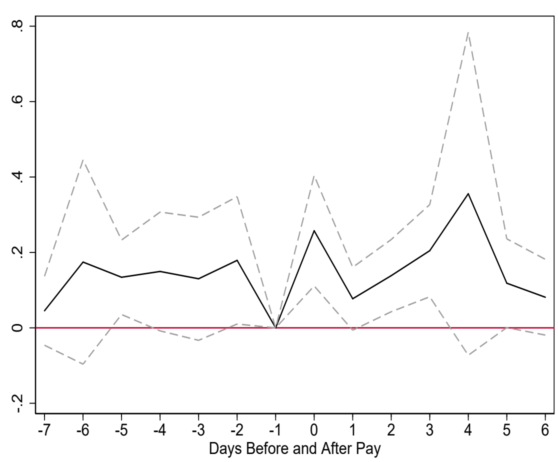


Figure 6: Automotive Expenditure
Fraction of Daily Average Spending



Also, for all spending categories there is a spike right after pay day, which is typically statistically different from zero. This, however, is generally not the only observed skipe as others are apparent (and often statistically significant) throughout the pay cycle.

To gain some further insight and take full advantage of our unique data set, we have estimated our regression models separately for different demographic groups with different age, education, and cognitive ability. For reasons of space, here we only present figures for grocery, restaurant, and merchandise expenditures broken down by education (Figures 7-9).

Figure 7: Grocery Expenditure by Education
Fraction of Daily Average Spending

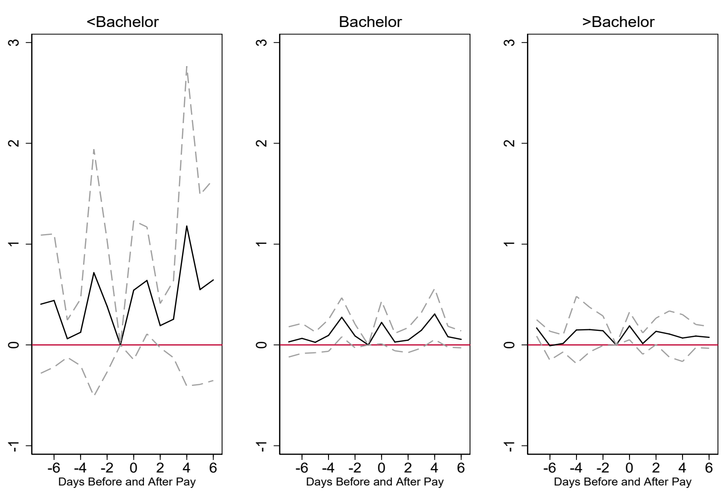


Figure 7 suggests that the lowest education category has most trouble smoothing grocery expenditures. However, the confidence intervals are wide so that expenditures on days before and after pay day are usually not significantly different from expenditures on day -1.

Figure 8: Restaurant Expenditure by Education
Fraction of Daily Average Spending

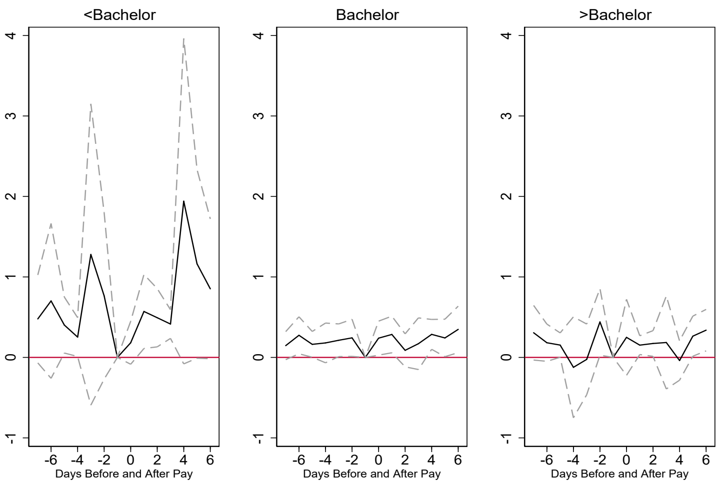
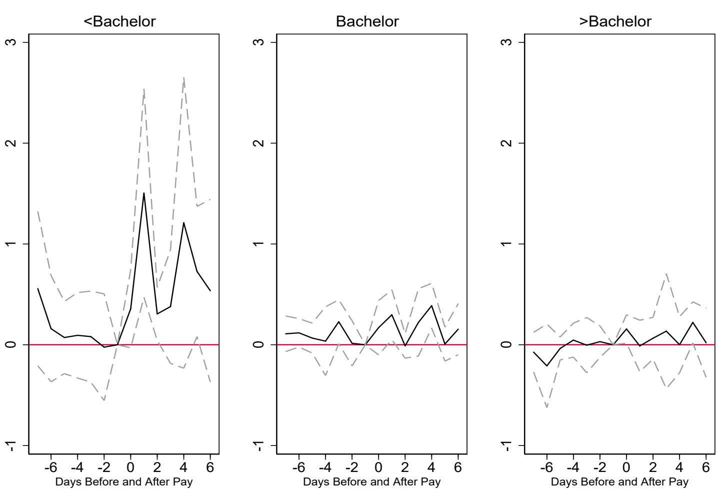


Figure 9: Merchandise Expenditure by Education
Fraction of Daily Average Spending



The pattern for restaurant expenditures in Figure 8 is not fundamentally different, although there is more of a suggestion of significant differences between the day before payday and other days for individuals with less than a Bachelor's degree. For merchandise expenditure (Figure 9) the difference between those without a Bachelor's degree and those with a Bachelor's degree or more is most pronounced in that the lowest education group appears to have clearest cyclical pattern over the pay cycle.

6 Comparison of Expenditure Measures: Self-Reports vs. Transactional Data

A comprehensive expenditure module was administered in the UAS between August and November 2016 to all active members at that time. The same respondents were invited to answer the module one more time throughout 2017. In the same year, the expenditure module was administered to all newly recruited UAS members. As a result of this, we have at our disposal self-reported expenditure measures for all 130 individuals in our sample.

The UAS expenditure module elicits expenditure amounts on a wide range of spending categories with reference to the previous calendar month. We do observe the date when the expenditure module was answered by each respondent. Hence, we match the self-reported expenditure measures with the transactional data of the calendar month prior to the month when the expenditure module was answered (provided that transactional data for that month exist). Out of the 130 respondents in our original sample, we are able to match self-reported and transactional expenditure measures for 114.[5]

We rely on these individuals to run a comparison exercise between self-reported and electronically recorded expenditures. The purpose is to assess

[5]Some individuals answered the UAS expenditure module prior to joining the electronic payments study. For a few respondents, no expenditure information could be retrieved from transactional data (e.g., they only linked retirement or brokerage accounts). About one-third of these 114 respondents are observed twice (i.e. we are able to match two self-reports). The results that follow are unaffected by whether we use weights that account for this or not.

how accurate self-reports are and the extent to which accuracy varies across expenditure categories. This information is particularly valuable to determine the reliability of self-reports in different spending domains.

Before commenting on the results of this analysis, it is worth pointing out the limitations of this exercise. First, we proceed under the assumption that the transactional data provide us with the "true" level of expenditure. Since we do not know how many accounts we are missing for each individual, we cannot evaluate what part of household spending is missing in our data. Second, the transaction categorization extracted from electronic records is coarser than the one available in the UAS expenditure module. For some spending categories, such as grocery, restaurant, utilities, and mortgage payments, there exists an immediate and exact correspondence between the two sources of data. For others, like automotive, entertainment, and insurance payments, it is harder to achieve a good match. Thus, observed differences between self-reports and transactional data may reflect differences in the composition of the categories that are being compared.

For each spending category, we compare the self-reported expenditure measure for a given month to the same-month expenditure measure taken from transactional data, as well as to the average monthly expenditure over the entire observation period, again calculated using transactional data. The idea is to shed light on the cognitive process behind self-reports, whether it is an episodic memory process, by which individuals try to recall what they actually spent in the previous calendar month, or a rate-based estimation process, by which they try to compute an average over a few months and apply it to the previous calendar month.

In the interest of space, we only present the results of this exercise for three spending categories – groceries, restaurant, and utilities – and for total expenditure. Figure 10 shows a quantile-to-quantile plot for the distribution of self-reported grocery expenditure against the same-month grocery expenditure from transactional data (left panel) and average monthly grocery expenditure from transactional data (right panel). As can be seen, self-reports tend to be substantially larger than transactional data outcomes for most of the distribu-

Table 11: Monthly Income and Expenditures by Education

	1^{st}q	Median	3^{rd}q	Mean	Std Dev
Grocery Expenditures					
Self-Report	250	400	500	424	294
Same-Month Trans Data	59	151	534	323	445
Average Trans Data	109	236	523	330	287
Restaurant Expenditures					
Self-Report	100	150	300	198	168
Same-Month Trans Data	40	144	292	208	226
Average Trans Data	92	188	371	254	221
Utilities Expenditures					
Self-Report	150	207	258	214	104
Same-Month Trans Data	55	94	161	115	82
Average Trans Data	76	110	192	139	84
Total Expenditures					
Self-Report	2,400	3,739	5,340	4,506	3,709
Same-Month Trans Data	546	2,194	4,639	3,930	6,137
Average Trans Data	1,127	2,795	5,262	3,954	4,044

tion, although percentiles around the 75th seem to align well. As reported in Table 11, median self-reports are 2.5 times median same-month expenditures measured from the transaction data and 1.6 times average monthly expenditure measured from transaction data. Mean self-reports are about 1.3 times larger than same-month and average actual expenditures. The large discrepancy of the medians reflects the fact that measured expenditures are likely to vary by month (for instance, because some months will contain five typical shopping weekdays, and others only four). To the extent that self-reports are reflective of some kind of average spending per month, these will show less variation over time and median and mean will be closer to each other. As one would expect, median and mean of the transaction data are closer to each other, when we calculate averages over the whole observation period.

Figure 10: Grocery Expenditure Comparison: Quantile-to-Qunatile Plot

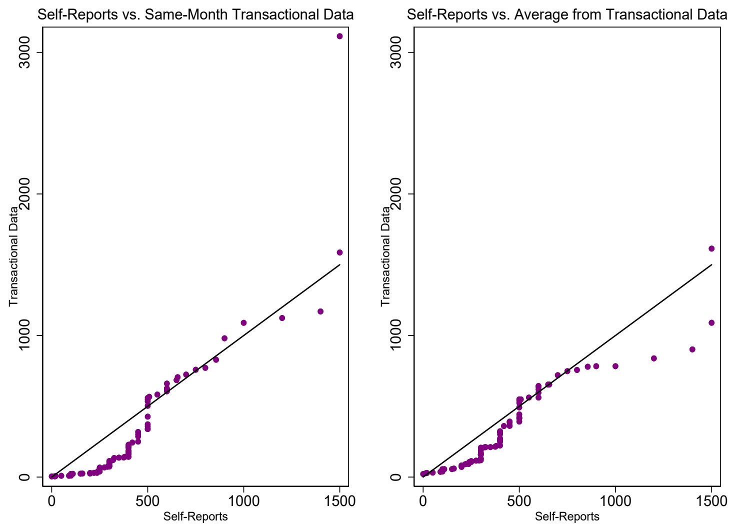
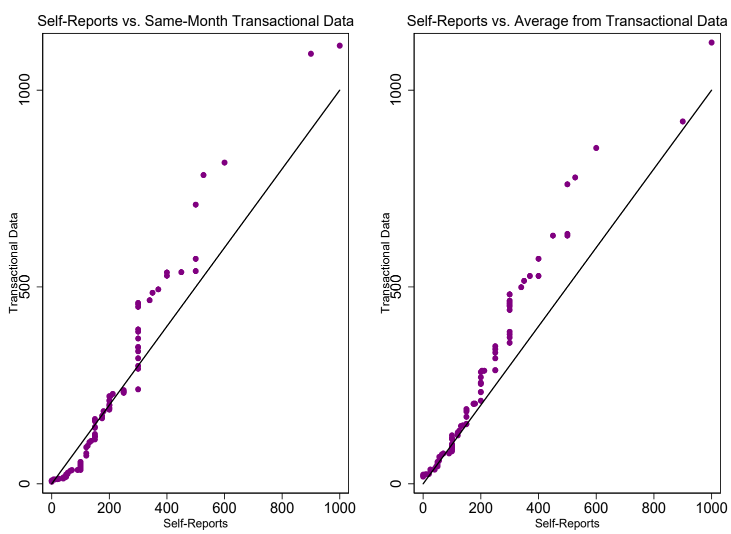


Figure 11: Restaurant Expenditure Comparison: Quantile-to-Qunatile Plot



For restaurant expenditures, the distributions of self-reports and transactional data look rather similar. Figure 11 shows a rather close alignment in the lower half of the distribution, while in the upper half of the distribution, values from transactional data exceed self-reported ones.

Figure 12: Utilities Expenditure Comparison: Quantile-to-Qunatile Plot

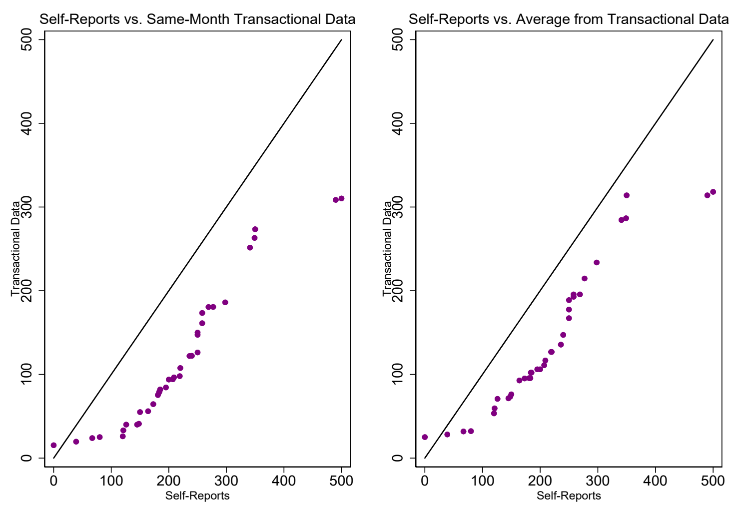
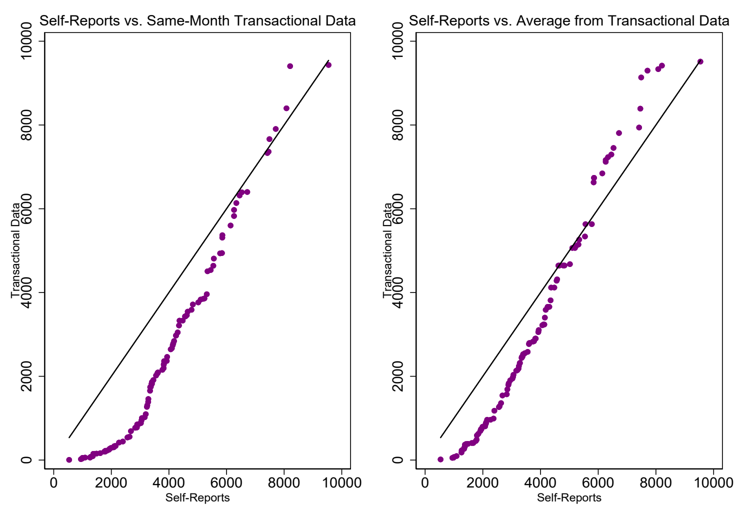


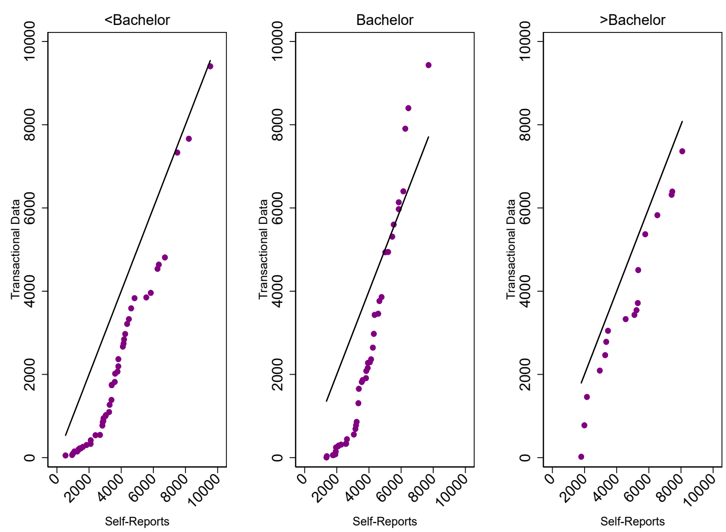
Figure 13: Total Expenditure Comparison: Quantile-to-Qunatile Plot



Unlike grocery expenditures, self-reported restaurant expenditures appear to be closer to the same-month actual expenditures than to the monthly average.

Utilities are recurrent expenses, which tend to exhibit relatively limited variation. As such, one would expect individuals to be able to accurately report them. This does not appear to be the case. As shown in Figure 12, utilities self-reports are largely above the level recorded by electronic transactions. Both median and mean values of self-reported utilities expenditure are double the median and mean values computed for the same month from transactional data. Self-reports are marginally closer to the average utilities expenditure computed from transactional data.

Figure 14: Total Expenditure Comparison by Education Self-Reports vs. Same-Month Transactional Data



In Figure 13, we report the comparison for total expenditures. As can be seen, self-reports systematically exceed same-month actual expenditure, with distances that are more pronounced in the lower half of the distribution. Median self-reported expenditure is 1.7 times same-month actual expenditure. Mean self-reported expenditure is 1.15 times same-month actual expenditure. A legitimate question at this point is whether self-reports accuracy depends on individuals' characteristics, especially education and cognitive ability. Figure

14 reveals that "over-reporting" applies to all education groups, although it is more pronounced for individuals without a college degree.

7 Conclusions

The analyses presented in this paper represent a first exploration of the potential of using electronic transaction data in combination with a population representative household panel. The data collection effort we have reported on reveals a number of challenges. These include consent rates, the willingness or skills of respondents to actually register their accounts, and possible omission of some of the accounts. The advantages of recruiting respondents from the UAS are 1) the ability to determine selectivity of the resulting sample; 2) the ability to investigate spending behavior separately for different demographic groups; 3) the ability to compare electronically recorded expenditures with self-reports.

We document the extent to which the characteristics of those actively participating in our study differ from the characteristics of the whole UAS pool of respondents. We report how the features of key outcome variables, such as number of accounts per person and account balances, in our sample differ from those reported by Gelman et al. (2014), the first study exploiting transactional data.

As far as more substantive analysis is concerned, we examine expenditure patterns over pay cycles for different spending categories. We do find evidence of non-smoothing behavior especially among individuals with less than a Bachelor's degree. Due to our limited sample size, however, our estimates are rather noisy. The exercise comparing self-reports and transactional data shows less variation and less skewness in self-reported monthly expenditures than in measured expenditures. A plausible explanation of this observation is that self-reports over longer periods, like a month, partly reflect "normal" or "typical" expenditures, rather than precise fluctuations from month to month. We also observe that for larger expenditures, like spending on restaurants, self-reports and measured expenditures provide similar outcomes. For smaller or less stable expenditures (groceries or utilities), the distribution of self-reports and transaction measures can be quite different. This merits further investigation. One possible reason why self-reports on utilities seem to be off could be that these expenses are often paid by direct debit. As a result, these payments may escape attention and, hence, a self-report only provides a noisy estimate of the actual expenditures.

Apart from obvious next steps aiming at larger samples, higher response rates, and exploration of methods to induce respondents to sign up all their accounts, there are a number of further analyses that can be performed. These include analyses of selectivity, by comparing participants to the complete panel, analyses of individual discrepancies, and attempts to improve estimates of total expenditures by combining self-reports and transactional data.

References

Agarwal, S., C. Liu, and N. Souleles (2007). The reaction of consumer spending and debt to tax rebates–evidence from consumer credit data. *Journal of Political Economy 115*, 986–1019.

Angrisani, M., B. Finley, and A. Kapteyn (2018). Can internet match high quality traditional surveys? Comparing the Health and Retirement Study and its online version. In K. Huynh, D. Jacho-Chavez, and G. Tripathi (Eds.), *Advances in Econometrics, Vol. 39. The Econometrics of Complex Survery Data: Theory and Applications.* forthcoming.

Angrisani, M., A. Kapteyn, and S. Schuh (2014). Measuring household spending and payment habits: The role of "typical" and "specific" time frames in survey questions. In C. Carroll, T. Crossley, and J. Sabelhaus (Eds.), *Imrpoving the Measurement of Consumer Expenditures. Chapter 15.* Chicago: University of Chicago Press.

Attanasio, O. P. (2000). Consumption. In J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, pp. 741–812. Amsterdam: Elsevier Science.

Baker, S. R. (2014). Debt and the consumption response to household income shocks. *mimeo*, Stanford University.

Belli, R. F., N. Schwarz, E. Singer, and J. Talarico (1997). Decomposition can harm the accuracy of behavioral frequency reports. *Applied Cognitive Psychology 14*, 295–308.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. *Hanbook of Econometrics 5*, 3705–3843.

Bricker, J., A. B. Kennickell, K. B. Moore, and S. J. (2012). Changes in U.S. family finances from 2007 to 2010: Evidence from the Survey of Consumer Finances. *Federal Reserve Bulletin 98*, 1–80.

Browning, M., T. F. Crossley, and G. Weber (2003). Asking consumption questions in general purpose surveys. *The Economic Journal 113*, 540–567.

Browning, M., T. F. Crossley, and J. Winter (2014). The measurement of household consumption expenditure. *Annual Review of Economics 6*, 475–501.

Carroll, C., T. Crossley, and J. Sabelhaus (2014). *Imrpoving the Measurement of Consumer Expenditures*. Chicago: University of Chicago Press.

Chang, L. and J. A. Krosnick (2009). National surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly 73*, 641–678.

Gelman, M., S. Kariv, M. Shapiro, D. Silverman, and S. Tadelis (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science 345*, 212–215.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009). *Survey methodology*. Hoboken, NJ: Wiley.

Hurd, M. (1999). Anchoring and acquiescence bias in measuring assets in household surveys. *Journal of Risk and Uncertainty 19*, 111–136.

Hurd, M. and S. Rohwedder (2009). Methodological innovations in collecting spending data: The HRS Consumption and Activities Mail Survey. *Fiscal Studies 30*, 435–449.

Juster, F. T., J. P. Smith, and F. Stafford (1999). The measurement and structure of household wealth. *Labour Economics 6*, 253–275.

Kapteyn, A., P. Kooreman, and R. J. Willemse (1988). Some methodological issues in the implementation of subjective poverty definitions. *Journal of Human Resources 23*, 222–242.

Kapteyn, A. and J. Y. Ypma (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics 25*, 513–551.

Kuchler, T. (2015). Sticking to your plan: Empirical evidence on the role of present bias for credit card paydown. *mimeo*, New York University.

Kueng, L. (2015). Explaining consumption excess sensitivity with near-rationality: Evidence from large predetermined payments. *NBER Working Paper No. 21772*.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*, 1–32.

Menon, G. (1997). Are the parts better than the whole? The effects of decompositional questions on judgments of frequent behaviors. *Journal of Marketing Research 34*, 335–346.

Moore, J. C., L. L. Stinson, and E. J. Welniak (1999). Income reporting in surveys: Cognitive issues and measurement error. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research*. New York: Wiley.

Olafsson, A. and M. Pagel (2018). The liquid hand-to-mouth: Evidence from personal finance management software. *Review of Financial Studies hhy055, https://doi.org/10.1093/rfs/hhy055.*

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

Schonlau, M., A. van Soest, A. Kapteyn, and M. Couper (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods & Research 37*, 291–318.

Souleles, N. (1999). The response of household consumption to income tax refunds. *American Economic Review 89*, 947–958.

Tourangeau, R. and N. M. Bradburn (2010). The psychology of survey response. In P. V. Marsden and J. D. Wright (Eds.), *Handbook of survey research*, pp. 315–346. Bingley, UK: Emerald.

Wenz, A., A. Jäckle, and M. P. Couper (2017). Willingness to use mobile technologies for data collection in a probability household panel. *Understanding Society Working Paper 2017-10*.

Winter, J. (2014). Response bias in survey-based measures of household consumption. *Economics Bulletin 3*, 1–12.

Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang (2011). Comparing the accuracy of rdd telephone and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly 75*, 709–747.