Research Data Reports

FEDERAL RESERVE BANK OF BOSTON™

No. 18-2

The 2015 and 2016 Diaries of Consumer Payment Choice: Technical Appendix

Marco Angrisani, Kevin Foster, and Marcin Hitczenko

Abstract:
This document serves as the technical appendix to the 2015 and 2016 editions of the Diary of Consumer Payment Choice (DCPC) administered by the Center for Economic and Social Research. The DCPC is a study designed primarily to collect data on financial transactions over a three-day period by U.S. consumers ages 18 and older. In this data report, we detail the technical aspects of the survey design, implementation, and analysis.

Keywords: survey design, sample selection, raking, survey cleaning, poststratification estimates

JEL codes: D12, D14, E4

Marco Angrisani is an associate economist at the University of Southern California Dornsife Center for Economic and Social Research (CESR). Kevin Foster is a survey methodologist, and Marcin Hitczenko is a statistician; when this report was written, they were members of the Consumer Payments Research Center (CPRC) in the research department of the Federal Reserve Bank of Boston. Both are now with the Federal Reserve Bank of Atlanta's research department. Their e-mail addresses are marco.angrisani@usc.edu, kevin.foster@atl.frb.org, and marcin.hitczenko@atl.frb.org, respectively. This paper, which may be revised, is available at http://www.bostonfed.org/economic/wp/index.htm.

The authors thank their colleagues and management in the CPRC and the Boston Fed research department, and we thank the management and staff at CESR. From the Boston Fed: Tamás Briglevics, Sean Connolly, Claire Greene, Vikram Jambulapati, Adam Karabatakis, Suzanne Lorant, William Murdock, Jason Premo, Scott Schuh, Oz Shy, Joanna Stavins, and Bob Triest. From CESR: Tania Gursche, Arie Kapteyn, Bart Orriens, and Bas Weerman. In addition, we acknowledge John Sabelhaus and the staff of the Survey of Consumer Finances at the Federal Reserve Board of Governors for their advice and mentorship. Geoff Gerdes and May Liu from the Board also shared advice and knowledge.

This version: March 2018

Contents

1 Introduction

This document serves as the technical appendix for the 2015 and 2016 editions of the Diary of Consumer Payment Choice (DCPC). The DCPC is a survey that was designed and sponsored by the Consumer Payment Research Center (CPRC) at the Federal Reserve Bank of Boston in collaboration with the Federal Reserve Bank of Richmond and the Cash Product Office at the Federal Reserve Bank of San Francisco. The DCPC is the second payments survey developed by the CPRC; the first was the Survey of Consumer Payment Choice (SCPC). The SCPC has been fielded annually since 2008, and annual reports, data, and tools for data use are available on the SCPC website.[1] The DCPC was conducted in 2010 and 2011 as pilot studies. The first official DCPC was conducted in 2012, although with a survey vendor and respondent panel that were different from those used in 2015 and 2016. Questionnaire programming, respondent recruitment, survey management, and data collection for the 2015 and 2016 DCPCs were contracted to the Center for Economic and Social Research (CESR) at the University of Southern California, and a majority of respondents came from the center's Understanding America Study (UAS).

All materials related to the DCPC, including the public-use dataset and the questionnaires, are available on the DCPC website.[2] Among the materials is the "Guide to the Diary of Consumer Payment Choice," which provides details for every variable in the DCPC dataset. Greene and Schuh (2017) summarizes key findings in the 2016 DCPC and how they compare with those in the 2015 DCPC. Greene, O'Brien, and Schuh (2017) describes the diary questionnaire in detail. Additional academic papers on consumer payment diaries and results from the 2012 and 2015 DCPCs have been published (Samphantharak, Schuh, and Townsend 2017 and Greene, O'Brien, and Schuh 2017).

The organization of this work follows the natural chronological progression of considerations involved in conducting and analyzing a survey. We establish the context and goals of the diary in Section 2. Section 3 details the sample selection strategy for the DCPC and presents statistics relating to diary response and completion. Section 4 delineates the methodology that generates the sample weights we used to make inferences about the entire population of U.S. consumers. Section 5 describes the preprocessing that produced the public-use dataset. Finally, in Section 6, we present the statistical methodology that we employed to generate population estimates and standard errors based on the DCPC data.

[1] https://www.bostonfed.org/publications/survey-of-consumer-payment-choice.aspx

[2] https://www.bostonfed.org/publications/diary-of-consumer-payment-choice.aspx

2 Survey Objective, Goals, and Approach

In this section we describe the overall objectives, goals, and general approach to the survey design of the DCPC. In particular, we explain our selections for the unit of observation and the interview mode. In both cases, the CPRC seeks to comply with best survey practices, subject to constraints imposed by budgets and resources.

2.1 Survey Objective and Goals

The DCPC was conceived as a natural extension of the SCPC. The SCPC is a 30-minute online survey fielded at the end of September and early October that asks respondents about payment instrument preferences, adoption, and use in terms of their typical number of monthly transactions. The DCPC is also a study of payments, but it focuses on the dynamics of payment instrument choice at the transaction level. It is dubbed a "diary" because respondents record details of their financial transactions, including all payments, on a daily basis over a three-day period. As discussed later in more detail, the measurement period during which observations are gathered is generally the month of October, though the 2015 DCPC stretches from the middle of October to the middle of December.

One major benefit of the DCPC is that it offers a new line of insight into payment instrument choice by collecting an array of information that is different from what the SCPC gathers. Perhaps the most useful feature of the DCPC distinguishing it from the SCPC is that it links information such as merchant and dollar value to each observed transaction. However, the information about a consumer's personal preferences and adopted set of payment instruments that only the SCPC collects is crucial to understanding consumer behavior. It is reasonable to assume that a complete model of payment instrument choice benefits from the SCPC information as well as from the transactional-level data the the DCPC gathers. Note that the DCPC does not give a full picture of payment instrument adoptions, because respondents' failure to use a payment instrument during the diary period does not necessarily mean they have not adopted that instrument.

The DCPC also enables an assessment of the reliability of SCPC data, especially data relating to the frequency of payment instrument use. For such variables, the SCPC and the DCPC represent two alternative methods of collecting information and making population estimates. Perhaps the most prominent example involves the average number of payments made in a month. The SCPC asks for a self-reported rate through recall, while the DCPC simply observes a portion of each individual's behavior. Each dataset requires a different type of pooling of the information to generate a population-wide estimate. Therefore, comparisons of SCPC-based estimates with DCPC-based estimates provide some insight into the accuracy and efficacy of different data-collection and estimation strategies. Findings will be used to identify data-collection strategies for future surveys.

2.2 Unit of Observation

The DCPC unit of observation is the consumer-day. The DCPC is designed primarily to make population estimates for a specific period of time (usually the month of October). However, the suspected loss in data quality that would occur due to respondents' fatigue or their dropping out of the survey makes it impractical to ask them to track all daily transactions for an entire month. Research has found that diary fatigue generally settles in after a few days (Ahmed, Brzozowski, and Crossley 2006; Jonker and Kosse 2009; and Schmidt 2011). Partly for this reason, most consumer diaries last from a day (Netherlands) to a week (Germany, France, Austria, and Australia), although the UK Payments Council Survey lasts four weeks. Of course, even if data quality were not a potential concern, the cost of incentives involved in tracking respondents for an entire month could preclude a larger, more diverse sample of respondents when the survey is administered under a fixed budget. The DCPC relies on three-day diaries in an attempt to balance the quality of data collected with respondent burden.

Just as it is important to sample a broad swath of consumers, it is also important to have data from all parts of the measurement period, because there may be temporal heterogeneity of economic behavior. It is best to observe a representative mix of consumers uniformly throughout the month. Therefore, collecting data involves selecting not only who participates but also when they participate. Overlapping waves of respondents, described in more detail in Section 3.2, are distributed throughout the entire month for complete temporal coverage.

To easily link the SCPC and the DCPC, we continue the practice established in the SCPC of relying on the consumer within the unit of observation. This choice means that each respondent reports only his or her own transactions in the DCPC, not the household's, which is the fundamental unit of observation in the Survey of Consumer Finances and the Consumer Expenditure Survey. In addition to maintaining consistency with the SCPC, the choice of individual-consumer resporting is appropriate for logistical and theoretical reasons.

Household surveys, which collect data on every transaction made by anyone in the household, require a greater coordination of survey responses. Ensuring that all household members participate likely takes considerable effort, and facilitation through in-person interviews would drive up data-collection costs significantly. Again, under a fixed budget, sampling households rather than consumers would mean that fewer households are included in the sample, potentially reducing the diversity of observed behavior.

In addition, for many economic concepts covered in the DCPC, we argue that a per-consumer questionnaire is likely to yield more accurate data. Asking one household member to collect information about all transactions from other household members could lead to under-counting, because individuals may not feel comfortable sharing the details of every purchase with one another, or simply because gathering reliable information requires consistent communication. Transactions of a certain kind, such as small-value cash payments, may be affected disproportionately, because they are easier to forget. Cash use, in particular, is of great interest to the CPRC and the Cash Product Office, and no other survey measures it.

For other variables, most notably the payment of bills or other expenses more closely associated with a household than an individual, the individual consumer may not be the ideal observation unit. Many such payments are made automatically and often come out of joint accounts or pooled resources. As a result, attributing responsibility for such payments often leads to measurement bias in the form of under-counting, if they are not reported at all, or double-counting, if several household members each claim responsibility for the same payment. Therefore, in such cases, the framework of a household payment might make more sense.

To enable the study of intra-household dynamics, the DCPC and the SCPC samples include some respondents from the same household. The presence of households with several sampled members may provide considerable insight into forms of under- and over-counting. Research based on the SCPC's multi-sample households suggests that survey respondents are more likely to have a greater share of financial responsibility within the household than would be expected if household members were selected at random, and thus they tend to be more likely than an average sample of the population to make certain types of payments (Hitczenko 2015). Treating such a sample as representative of all consumers may lead to overestimation of the number of bills paid. For accuracy, it might be better to ask about the entire household's bill-payment behavior. Nevertheless, for consistency within the survey instrument, the DCPC asks respondents to estimate only the number of bills that they physically pay themselves, either by mail, phone, online, or in person.

2.3 Interview Mode

The DCPC is a computer-assisted web interview (CAWI). This mode of interview is the best for the sampling frame, the internet-based Understanding America Study (UAS). The survey instrument is programmed in the NubiS survey system, developed by the CESR and compatible with all web browsers.[3]

The design of the DCPC makes an online survey a natural choice for implementation. In-person interviews, besides being more expensive to conduct, would be difficult to implement because respondents are expected to record certain data, such as cash holdings, at the end of the day. Using a CAWI allows respondents to log on at the end of their day and enter all relevant information. In addition, some evidence suggests that respondents are more likely to answer sensitive questions, such as those about certain purchase details, if the questions are administered through an online survey (De Leeuw 2005).

A second alternative, a paper survey, also is not practical, because the DCPC is designed to ask follow-up questions about transactions based on the initial information the respondent provides. For example, the use of cash for a payment may lead to a set of questions that would be different if the respondent instead had used a credit card. This type of skip logic fits easily within the CAWI framework but would be difficult to implement with a paper survey.

Although official data are collected online, respondents are encouraged to keep track of and record details of daily transactions through other means. To this end, they are sent two paper memory aids and a pouch in which they can keep receipts. The larger memory aid is the size of a folded 8.5-by-11-inch sheet of paper and includes instructions, examples, and response categories for every type of transaction. The second memory aid is the size of a checkbook and provides only enough space to record basic information about the transaction. The use of the memory aids is not required, and they are not collected after participants complete the DCPC. Nevertheless, 31 percent of respondents used the memory aids. Because the information collected through either memory aid does not include answers to follow-up questions about the transactions, they would not be suitable replacements for the online questionnaire. However, the statistics on memory aid use, along with the fact that 42 percent of respondents said they at least partially relied on the collection of receipts, suggest that much of the data entry in the online questionnaire is based on more than just recall.

[3] More information on NubiS is available at https://cesr.usc.edu/nubis/.

2.4 Public-use Datasets

The original, unprocessed datasets can be downloaded from the CESR's website. The Boston Fed DCPC website contains a link to the UAS data download site. Downloading the data requires the creation of a username and password. The data contain only the survey variables found directly in the survey instrument itself, though a few open-ended responses have been omitted for privacy reasons. The included survey variables have not been edited or processed. The simplest way to identify variables is by finding them directly in the DCPC questionnaire, which can be downloaded as a pdf document from the Boston Fed's DCPC website.[4]

The DCPC website also offers a version of the data processed by the CPRC in Stata, SAS, and CSV formats. The processed dataset is restricted to the respondents who completed all three days of the diary, which is a large majority of the participants. It is also organized differently from the raw dataset. Each row in the processed dataset corresponds to an indicator of a respondent's participation on a particular day or a single transaction (the participation indicator is necessary to highlight days on which a respondent had no reported transactions). In the original dataset, one row contains information about all transactions made by a particular consumer on a particular day. The processed data are easier to sort and manipulate by transaction characteristics, such as the type of extraction or payment instrument used. We edited certain variables in the processed dataset to improve the data quality. This process is detailed in Section 5.4. Demographic variables collected by the CESR's My Household Questionnaire, and not featured in the DCPC itself, are appended in the processed dataset. We recommend referencing the companion document, "Guide to the Diary of Consumer Payment Choice," which is also available on the DCPC website. The user's guide describes each variable, the diary question or questions that define it, and all response options.

One variable of note is `prim_key`, which serves as the unique identifier for each respondent. It is the primary key for the Boston Fed datasets and can be used to merge the data with all other SCPC and DCPC data from 2014 on. The variable `prim_key` is identical to the UAS variable `uasid`, which can be used to merge data with any other survey completed by UAS members.

[4] https://www.bostonfed.org/publications/diary-of-consumer-payment-choice.aspx

3 Data Collection

Once the survey instrument has been finalized, the collection of data involves two general steps: sample selection and administration of the survey. The strategies and philosophies adopted in each step for the 2015 and 2016 DCPCs are described below. In addition, summary statistics related to survey completion are detailed. Similar expositions focusing on the 2012 DCPC can be found in Angrisani, Foster, and Hitczenko (2017a). Angrisani, Foster, and Hitczenko (2017b) contains the technical appendix for the 2015 SCPC; a technical appendix for the 2016 SCPC is forthcoming.

3.1 Understanding America Survey

Prior to 2014, the CPRC used the Rand Corporation and its panel, the American Life Panel (ALP), exclusively to collect data, but in recent years, it has begun using the Center for Economic Social Research (CESR) and the Understanding America Study (UAS). A majority of respondents in the 2015 DCPC and all respondents in the 2016 DCPC were recruited from among UAS panelists. As a collection of individuals who have agreed to participate in online surveys, the UAS panel is similar in spirit to the ALP. However, the CESR uses the best practice of address-based sampling and the UAS began with the goal of representing all adults living in the United States (Dillman, Smyth, and Christian 2014). By contrast, the ALP started as a panel studying senior members of the population and has used a motley of recruitment strategies. Largely due to these differences, the UAS has representative qualities generally associated with better inference for population parameters.

The CESR has conducted address-based recruitment into the UAS in waves that started in 2014. The pool of potential respondents grew from about 1,900 individuals in fall 2015 to more than 4,000 individuals in fall 2016. A majority of panel members belong to the "nationally representative" cohort, which is designed to best match proportions for selected demographic variables among all U.S. adults. The UAS panel also includes a cohort that represents Native American households (recruiting started in 2014) and one that corresponds to households with young children in Los Angeles County (recruiting started in 2015). Details of the recruitment process are provided on the UAS website.[5] The person who was the primary point of contact for about 11 percent of the households reached by the CESR eventually joined the UAS panel. The CESR made a concerted effort to recruit additional household members, and as a result about 15 percent of UAS panelists have a fellow household member

[5] https://uasdata.usc.edu/index.php

in the panel.

Panelists who did not have reliable internet access were provided with a tablet and broadband internet. Survey opportunities were emailed to respondents and appeared on their account dashboard as a link. The surveys themselves were programmed in the propriety NubiS software developed by the CESR. The software allows online entry on all computer types as well as tablets and smartphones.

3.2 DCPC Sample Selection

Even more so than the 2012 edition, the 2015 and 2016 DCPC surveys are strongly linked to the SCPC, and thus the desire was to have respondents participate in both surveys. In fact, the DCPC uses responses about the adoption of payment instruments from the SCPC to help provide more context for recorded transactions. It was strongly recommended that the SCPC precede the DCPC, though this was not necessary.

Annual budgeting, considerations of sample sizes, and survey implementation all inherently involved both the SCPC and the DCPC. The general goal was to establish a longitudinal structure to both surveys in which the set of respondents in a given year would be largely the same as that from previous years. The benefits of a longitudinal panel, particularly the added power associated with tracking trends at the individual level, have been well discussed (Baltagi 2008, Duncan and Kalton 1987, Frees 2004, and Lynn 2009). For many research agendas, it is advantageous to base results on a longitudinal panel, rather than on a sequence of cross-sectional studies.

3.2.1 Sample Selection in 2015

The limited size of the UAS panel at the time of recruitment for the 2015 DCPC created a need for additional observations. The CESR secured additional diaries by having a subset of UAS respondents take the survey twice, as discussed below, and by recruiting respondents from third-party vendors. To minimize the need for vendor-supplied respondents, the CESR gave all 1,973 UAS panelists who did not belong to the Los Angeles County cohort the opportunity to participate in both surveys. They included 1,408 respondents who had participated in the 2014 SCPC, and an additional 565 who had not. Most of the latter had joined the UAS panel within the prior 12 months.

The joint-recruitment process for the SCPC/DCPC involved releasing a short "consent survey" in late September to the selected UAS panelists. The consent survey introduced the survey opportunity, along with a $70 incentive for completion of the DCPC and a $20 incentive for the SCPC, and asked respondents if they were willing to participate in both. The consent survey was answered by 1,482 individuals, representing only 75 percent of the invited panelists. However, of those who responded, 1,291 agreed to participate in the surveys, yielding an overall consent rate of 65 percent, but an 87 percent consent rate conditional on viewing the SCPC/DCPC invitation. This overall consent rate is significantly lower than the 75 percent to 85 percent observed in previous years, including the 2014 UAS. We hypothesize that a major reason for this decline is the fact that the consent survey was offered without the enticement of an incentive.

Of those individuals who tentatively agreed to participate, 652 individuals were asked if they would be willing to participate in the diary twice. This subset was chosen to match population strata according to the set of demographics listed in Table 1. The idea of dual surveys represents an experimental approach that the CPRC developed to simultaneously increase the number of diary days observed and study responses when data from more days are collected for each individual.

In addition to recruiting participants from the UAS sample, the CPRC sought to include 250 individuals from the Qualtrics panel (more information is available at www.qualtrics.com) and 500 from the GfK panel (more information is available at www.gfk.com). Both of those third parties recruited participants to represent the U.S. population of adults and provided the CESR with email contacts for each individual, so that they could correspond directly with the CESR and take all surveys using the NubiS software.

Participation in the DCPC is defined as the completion of online questionnaires for each diary day. The DCPC features data entry of all transactions on three days (Days 1, 2, and 3) and a short introductory module (Day 0) intended to be completed before Day 1. In 2015, 1,597 unique respondents (comprising 1,155 UAS panelists, 357 GfK panelists, and 85 Qualtrics panelists) logged on to at least one of the four daily questionnaires. The significantly lower response rate for the Qualtrics subsample, which was less than 30 percent, was largely due to failed attempts to contact its members through the provided email addresses. Because the response rate for this subsample was so low and the quality of the data was poor, we discarded Qualtrics observations from the analysis and did not include them in the public-use dataset. All future discussion will involve only the 1,512 individuals from the UAS and GfK panels.

Attrition in the DCPC generally increased across diary days. Whereas participants (including those who took the survey twice) commenced 2,028 Day 0 modules, the numbers for Days 1, 2, and 3 were 1,973, 1,932, and 1,901, respectively, representing a roughly 5 percent attrition rate from the first day to the final day. However, for those who started a diary day, the completion rates are very high. Completion of a diary day can be measured in a variety of ways, but we focus on those participants who logged into the online questionnaire and then logged out. Of course, it is possible that individuals entered information and simply did not log out, so the measures of completion are conservative. Three percent of respondents who began Day 1 failed to finish that day, representing the highest incompletion rate among the days. As a result, the set of respondents who completed all four daily questionnaires totals 1,392 individuals, including 1,076 from the UAS panel and 316 from the GfK panel. It is important to note that 60 respondents from the UAS subset came from the Native American cohort. Overall, 509 UAS respondents completed two diary periods. The subset of respondents in the public-use dataset represents 1,317 households: 1,251 households with one representative, 58 with two, seven with three, and one with four.

3.2.2 Sample Selection in 2016

Due to the massive growth of the UAS panel from fall 2015 to fall 2016, sample selection in 2016 more closely mirrors the approach used in the 2012 DCPC and most SCPC surveys. All respondents are UAS panelists, and each respondent participated in only one diary.

To prioritize representation, we chose 15 demographics categories with which to match our sample to U.S. proportions as measured by the Current Population Survey (CPS), administered in March of each year. The demographic strata are shown in Table 1. A set of 4,000 respondents that had stratum composition as similar as possible to known U.S. population values was selected from the nationally representative cohort of the ALP. To maximize the longitudinal component, all respondents from the nationally representative cohort who had participated in the 2015 DCPC or SCPC were included.

Table 1: Strata used in DCPC sample selection.

Stratum	Race	Age	Income	Stratum	Race	Age	Income
1	White	18−39	<$30K	10	Non-white	18−39	<$30K
2	White	18−39	$30K-$60K	11	Non-white	18−39	≥$30K
3	White	18−39	≥$60K	12	Non-white	40−55	<$30K
4	White	40−55	<$30K	13	Non-white	40−55	≥$60K
5	White	40−55	$30K-$60K	14	Non-white	56+	<$30K
6	White	40−55	≥$60K	15	Non-white	56+	≥$30K
7	White	56+	<$30K				
8	White	56+	$30K-$60K				
9	White	56+	≥$60K				

To improve upon the previous year's participation rate, the consent survey was fielded with a $5 incentive for completion. Indeed, of the 4,000 consent surveys that the CESR distributed, it received responses from 3,572, a significant improvement over the 2015 rate. Attrition rates at each step largely mirror those from 2015. Again, willingness to participate conditional on viewing the consent survey was high, with 3,361 individuals agreeing to take part. Overall, 3,048 participated on at least one day. The public-use dataset comprises information from the 2,848 respondents who completed all days of the diary. Of these respondents, 867 participated in the 2015 DCPC. The set of 2016 respondents in the public-use dataset represents 2,474 households; 2,146 have one member in the sample, 287 have two members, 36 have three members, and five have four members.

3.3 Diary Day Assignment

The ideal sample design for the DCPC has respondents uniformly taking the DCPC throughout the month of October. The desire to standardize this response period is threefold. First, from an analytical point of view, trends from year to year are more easily identified if differences in behavior are not attributable to seasonal behavioral variation. Second, from an economic perspective, the month of October is a reasonably representative month with respect to certain economic variables such as employment and sales volumes; it includes no major holidays and falls between summer and winter. Third, responses from the DCPC can be linked more naturally to the SCPC if both surveys correspond to the same period of economic activity. Because the SCPC has traditionally been fielded in the fall, the month of October is a natural time for fielding the DCPC.

An additional goal in diary-day assignment is to have uniformity across the measurement period. We want the number of diarists to be uniformly distributed throughout the measurement period, and we want the number of diarists responding on each of the three diary days to be uniform. This makes comparing results across days methodologically simpler, and it is useful because different diary days collect different types of information in addition to the daily transactions. For example, information about checking account balances is collected on Day 0 and Day 3 only, so not fielding these modules on October 1, a date on which individuals often receive income and many bill payments are likely scheduled, would be a mistake.

Diary days are assigned randomly so that the expected composition of respondents for each day is the same. The only exception involves international travel, because the goal of the CPRC surveys is to study domestic payments. To achieve this goal, the consent survey asks respondents who have agreed to take the DCPC if they are traveling internationally during the general diary period, and, if so, to provide the dates of travel. These respondents are given diary dates outside of their travel period. In 2016, only 89 respondents claimed to have plans for international travel.

Once diary days are assigned, the CPRC discourages respondents from changing them, because if they switch to a relatively more convenient time it will introduce bias into the responses. It is important to note that capturing travel within the United States is desirable, so respondents are encouraged to participate even if they are going to travel domestically. In fact, the CESR offers to extend deadlines for online data entry to participants who are going to be traveling domestically during their assigned diary days.

Because diary days are assigned in advance, reminder emails are sent to participants several days before the first day of the diary. These reminders include a link to an instruction video that explains the basic goals and methodology of the diary. Three days before the beginning of the diary period, the CESR mails packages containing two types of paper memory aids. The use of memory aids is not required, but they can be helpful and serve as a second reminder of the upcoming diary period.

3.3.1 Diary Day Assignment in 2015

Because of delays involved with organizing joint SCPC and DCPC participation from three different survey vendors, it was impossible to field the 2015 DCPC entirely in the month of October. So instead, diary data were collected from October 13 through December 17. Potential seasonal influences on payment behavior make it more challenging to compare data from the 2015 DCPC with data from the 2012 DCPC, which was fielded in October. However, information about the nature of these seasonal effects, especially during the holiday period from the end of November into December, is interesting in its own right.

To balance the number of individuals who would be participating on Diary Days 0, 1, 2, and 3 on any given day in October, we started diary waves on October 13 and continued them through December 14, so that the last of the Day 3 questionnaires was scheduled for December 17. This design yields observation of all modules from October 16 to December 14.

Individuals asked to take the DCPC twice in 2015 were assigned a first diary in the first four weeks of the measurement period, and the second diary period was randomly assigned in the final week, so that at least two weeks had passed since the completion of the first period. This was done partly to avoid diary fatigue and partly to spread this subset of surveys throughout the measurement period. Figure 1 shows a plot of the dates of the first and second diaries, and it reveals a block design in which participants taking the first diary in the first quarter (two weeks) of the measurement period take the second diary in the third quarter, while those who participate for the first time in the second quarter of the measurement period take the diary again in the fourth quarter.

Respondents from the GfK panel were given random diary days that stretch from December 3 to the end of the measurement period. Figure 2 shows the number of respondents by vendor logging on for each day over the diary period in fall 2015. The optimal survey design calls for the same number of respondents for each diary day on all days in October. Under ideal sampling and participation, with 2,500 respondents, each of the 63 diary waves should have about 32 people per wave, or slightly fewer than 100 individuals recording transactions (Days 1, 2, and 3) on each date in the measurement period. As is apparent from Figure 2, this ideal was not quite met. First, the GfK sample skews the distribution so that substantially more individuals are reporting on each day near the final quarter of the measurement period. Even in the UAS sample, there is a noticeable bump in participation in the middle of November. A likely cause of this phenomenon is that individuals forgot to fill out the diary on their given day and, upon receiving a reminder from the CESR, filled it out for later days. While this is not ideal, because it corresponds to a nonrandom change to individuals' diary day assignments, it is preferable to collecting data from fewer respondents. A more gradual bump also appears at the end of the measurement period.

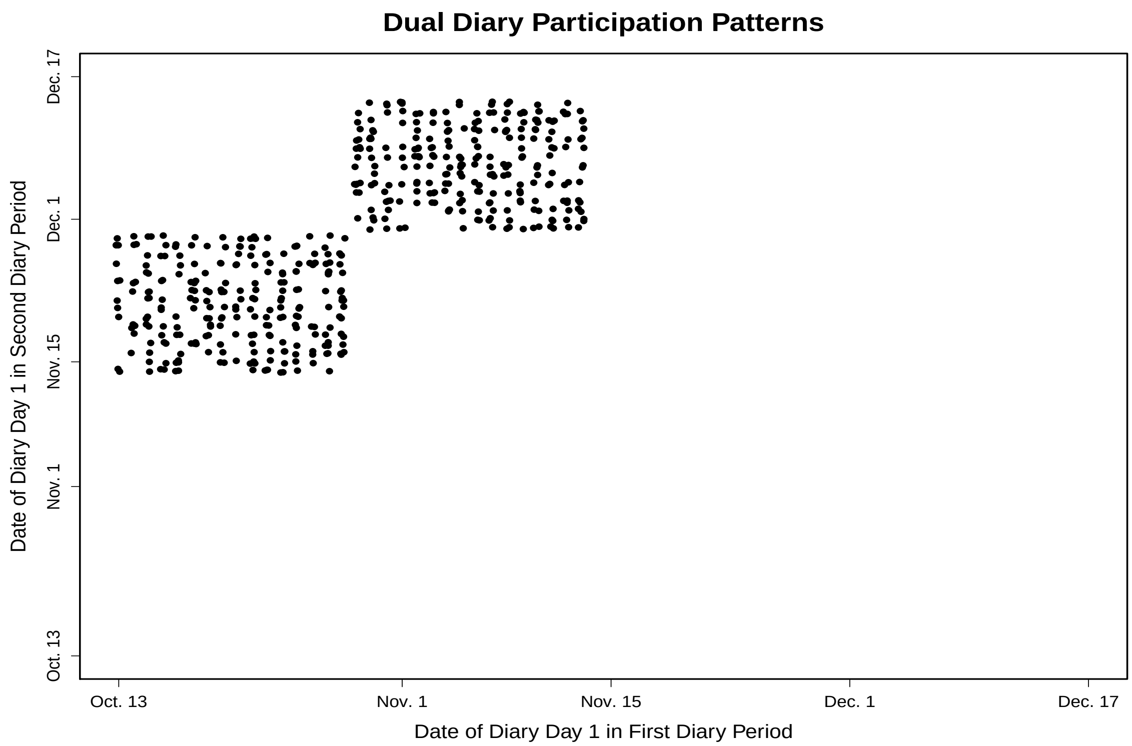

Figure 1: Dates for Diary Day 1 for respondents who participated in two diaries. Source: Authors' calculations.

3.3.2 Diary Day Assignment in 2016

As with sample selection, diary day assignment in 2016 was much more straightforward than it was in 2015. The first Day 0 was September 28 and the final Day 3 was assigned for November 2 to ensure uniformity within October. The shorter measurement period in 2016 yielded 33 waves, which, under the assumption that there were 2,800 respondents, resulted in roughly 340 unique individuals participating on each day. Figure 3 shows that the 2016 data are much more uniform than the previous year's.

3.4 Completion Data

Below, we consider various statistics on DCPC completion rates. Figure 4 depicts for the 2015 DCPC the distribution of the number of days between the dates respondents were supposed to fill out the online diary and when they actually did. Figure 5 shows that distribution for the 2016 DCPC. Note that if an individual were to log on at 1 a.m. on the morning after the assigned day, he or she would effectively be completing the diary as intended, but the

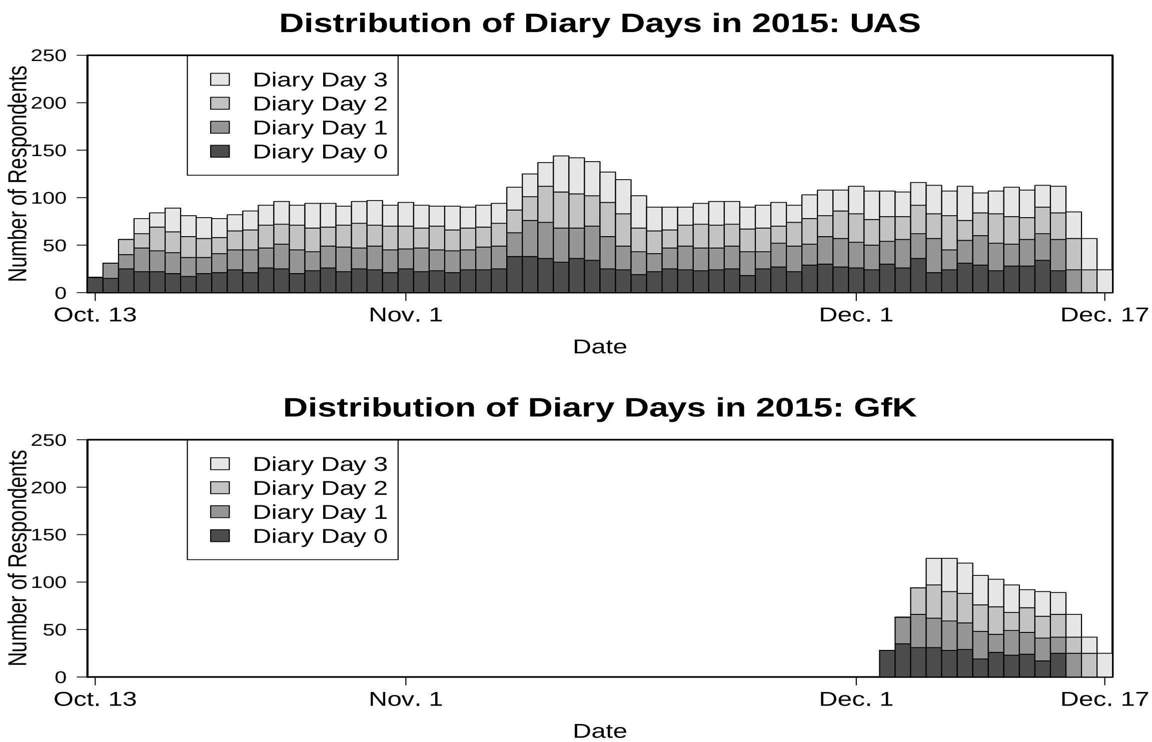

Figure 2: Number of respondents logging on to complete particular day of diary by date in 2015. Source: Authors' calculations.

metric in Figure 4 would register that recording as being delayed one day. Overall, a vast majority of respondents, more than 95 percent, took the survey on the assigned date. We find that, in both years, about 40 percent of respondents completed all days as assigned, while roughly 80 percent completed all days within one day. Long lags are very infrequent; in both years, only about 1 percent of diary data were entered online a week or more after the assigned diary day.

Figures 6 and 7 show, for 2015 and 2016, the distribution of the number of days between respondents taking the SCPC and the first day of the DCPC. The average absolute difference in days is 18 in 2015 and 21 in 2016, and nearly everyone took the SCPC before the DCPC (96.2 percent in 2015 and 98.6 percent in 2016), which the design of the surveys strongly encouraged. The lag appears to be much more uniform in 2016 than in 2015, for which the mode lag lies between one and two weeks difference. The release date of the 2016 SCPC may partly explain the generally longer lags in 2016. It was released on September 19, 10 days before the start of the diary measurement period; whereas, in 2015, the SCPC was released on October 6, seven days before the first diary wave.

Figures 8 and 9 show the distribution of completion times for each diary day in 2015 and 2016,

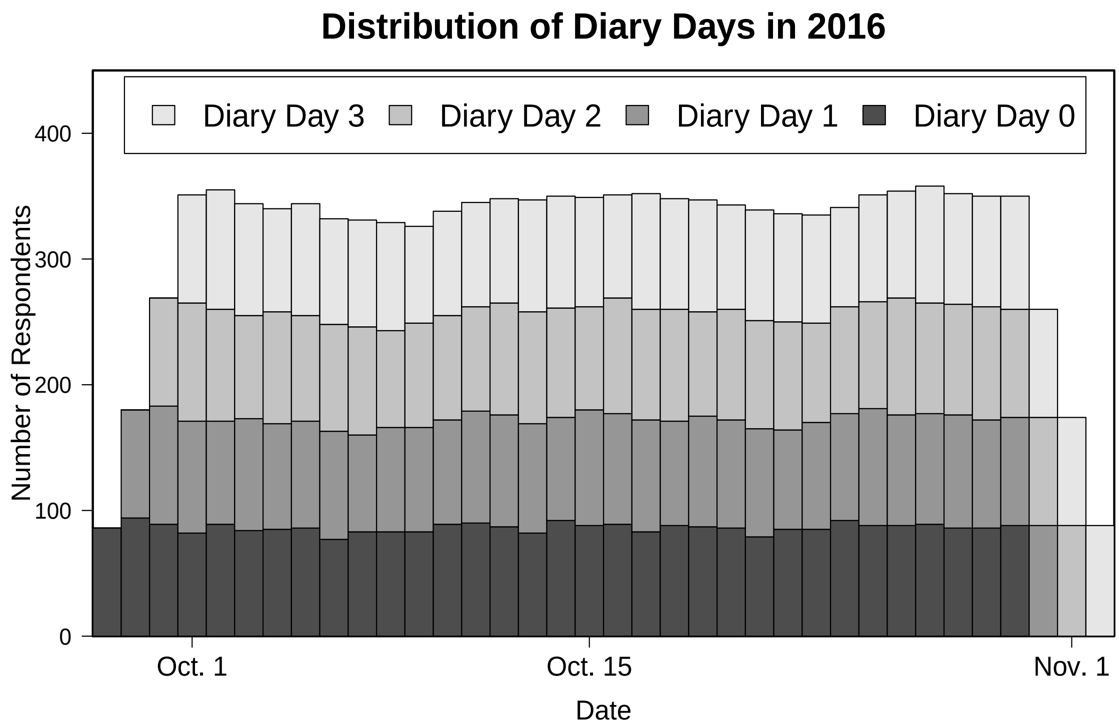

Figure 3: Number of respondents logging on to complete particular day of diary by date in 2016. Source: Authors' calculations.

with truncation occurring at 60 minutes. The times were measured as the difference between the first log in and final log out. Participants who took at least 60 minutes to complete a diary day might have stepped away from the computer, perhaps even logging on to enter one entry and doing so again later. Figures 8 and 9 reveal that in both years, Day 3, which features lengthy sets of questions (particularly the ones related to the payment of bills) that were not included on previous days, took substantially longer to complete than did the other days. Day 2 was consistently completed the fastest, likely because the corresponding questionnaire has the fewest question modules following the general daily transaction module. The overall completion times are generally longer for 2016 than 2015, about 55 minutes total versus 50 minutes total. This is largely driven by respondents' spending more time on Day 0, likely due to the changes to the structure of the questions for that day. Most notably, the 2016 diary collected more detail about account and payment instrument ownership, especially information relating to credit and debit cards.

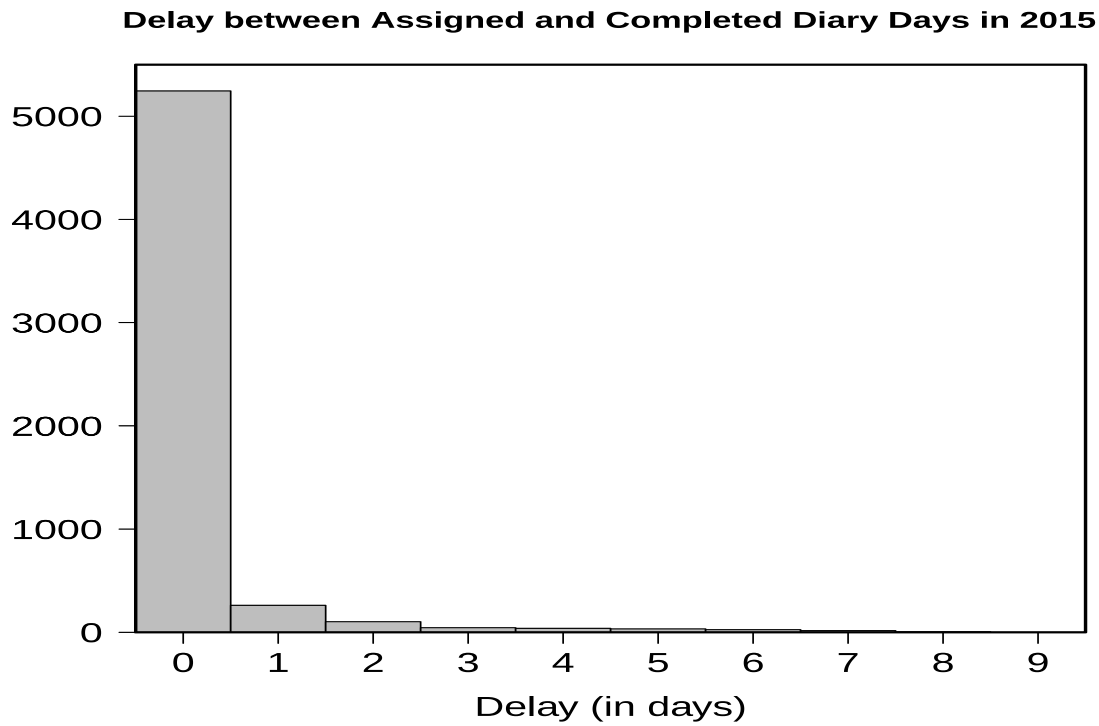

Figure 4: Distribution of the number of days between assigned diary day and day on which respondent logged on to enter data for all completed diary days (as many as eight per respondent) in 2015.

Source: Authors' calculations.

3.5 Item Nonresponse

The item response rate for each question must be high for a survey to provide a valid picture of the overall population. High nonresponse rates yield less information on which to base estimates, and they raise concerns about potential bias in the estimates. If an observation is missing independent of the value of the observation, a condition known as "missing at random" (Little and Rubin 2002), imputation procedures can be used to generate estimates of sample statistics. However, if a confounding variable relates to both the value of a variable and the likelihood of nonresponse, it is impossible to adjust for the effects on sample statistics. Certain payment details are potentially sensitive topics, and the willingness of respondents to provide answers could relate to the answers themselves. Naturally, variables with low nonresponse rates are less susceptible to this type of bias.

While we cannot determine if entire transactions were omitted from online entry, we can look at the fraction of details provided for those transactions that were recorded. Table 2 looks at the nonresponse rates for key attributes requested in the three most commonly used modules

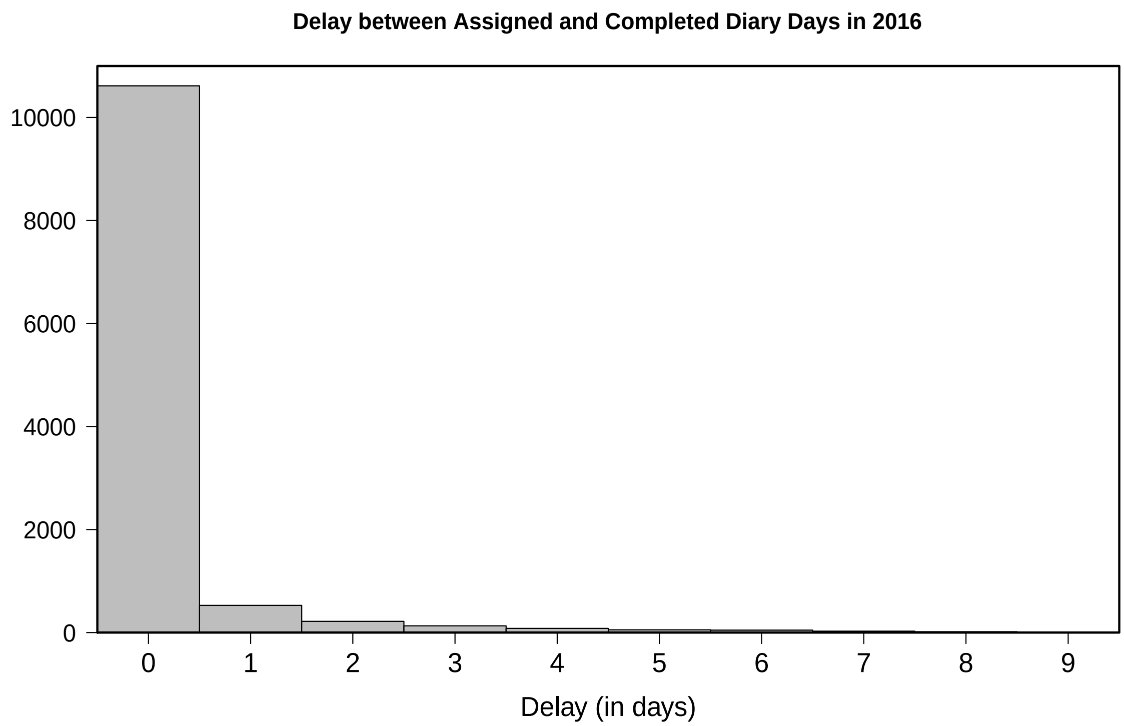

Figure 5: Distribution of the number of days between assigned diary day and day on which respondent logged on to enter data for all completed diary days (as many as four per respondent) in 2016.

Source: Authors' calculations.

for data entry: purchases, bills, and cash withdrawals. Overall, the item response rates are very high; time of transaction is the most commonly omitted field at about 6 percent. The general trends are consistent across both years.

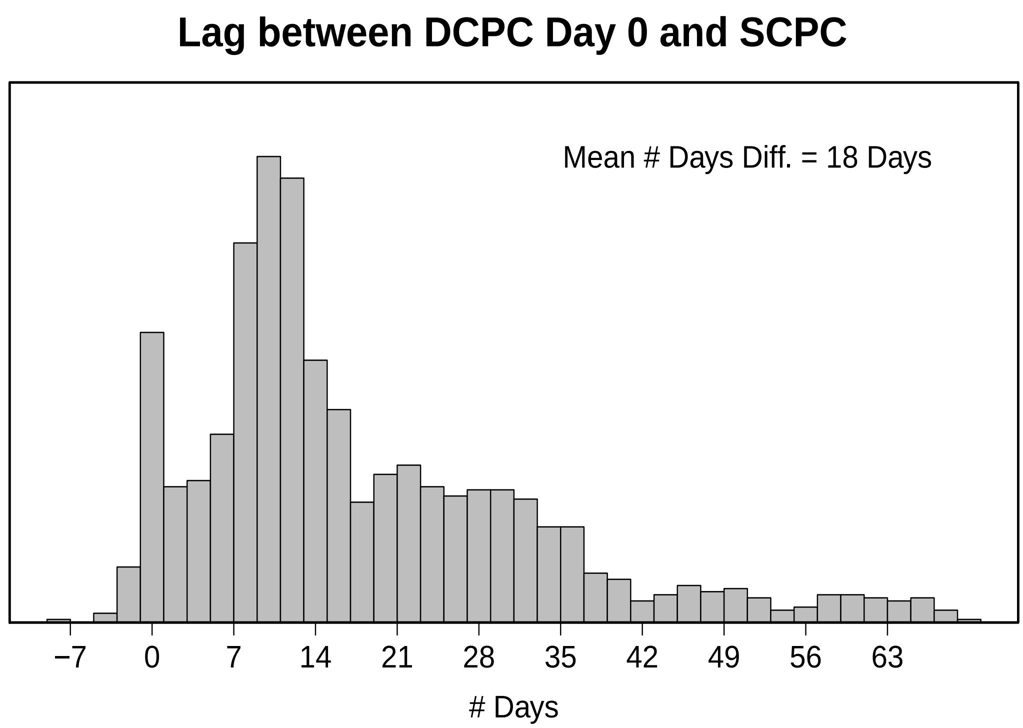

Figure 6: Distribution of the number of days between completion of diary day 0 and completion of the SCPC (Date of diary−Date of SCPC) in 2015.

Source: Authors' calculations.

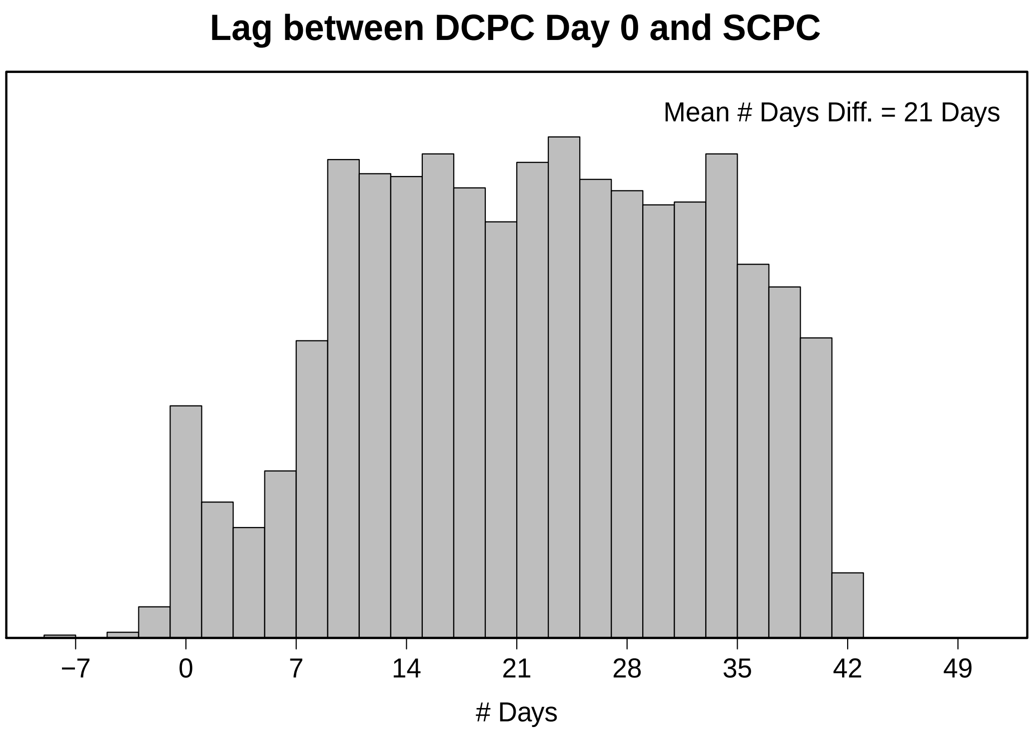

Figure 7: Distribution of the number of days between completion of diary day 0 and completion of the SCPC (Date of diary−Date of SCPC) in 2016.

Source: Authors' calculations.

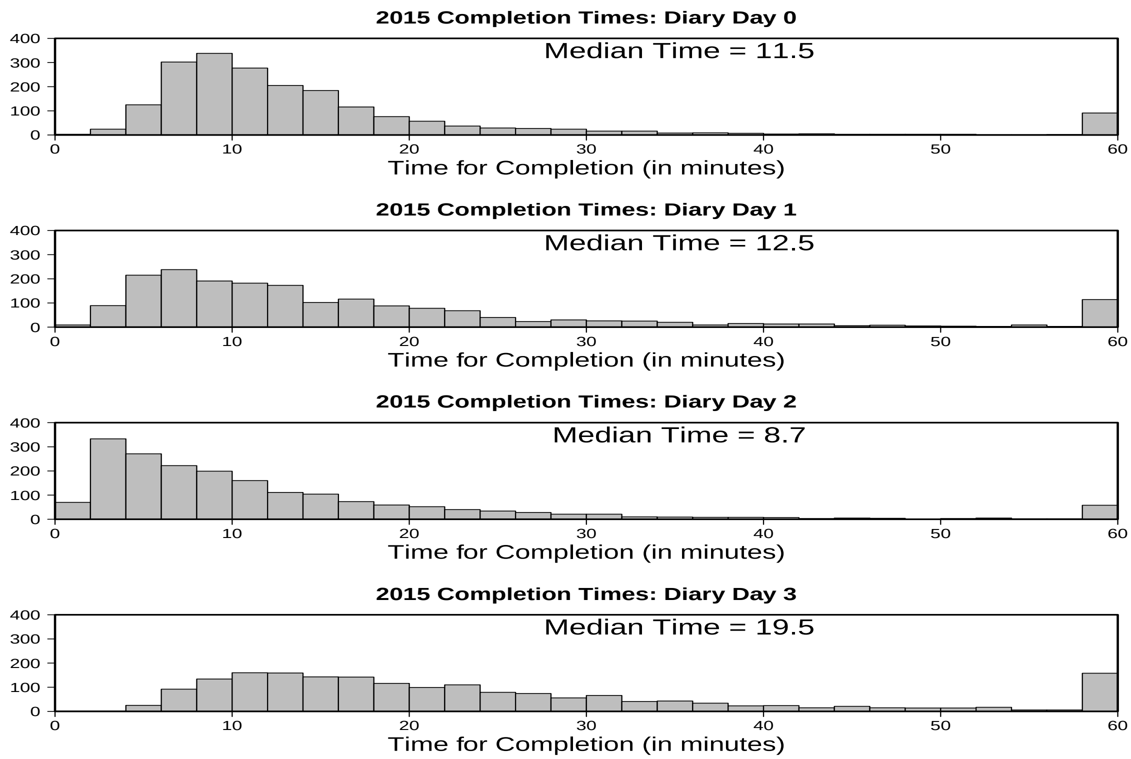

Figure 8: Distribution of the number of minutes to complete each diary day in 2015, with truncation to 60 minutes.

Source: Authors' calculations.

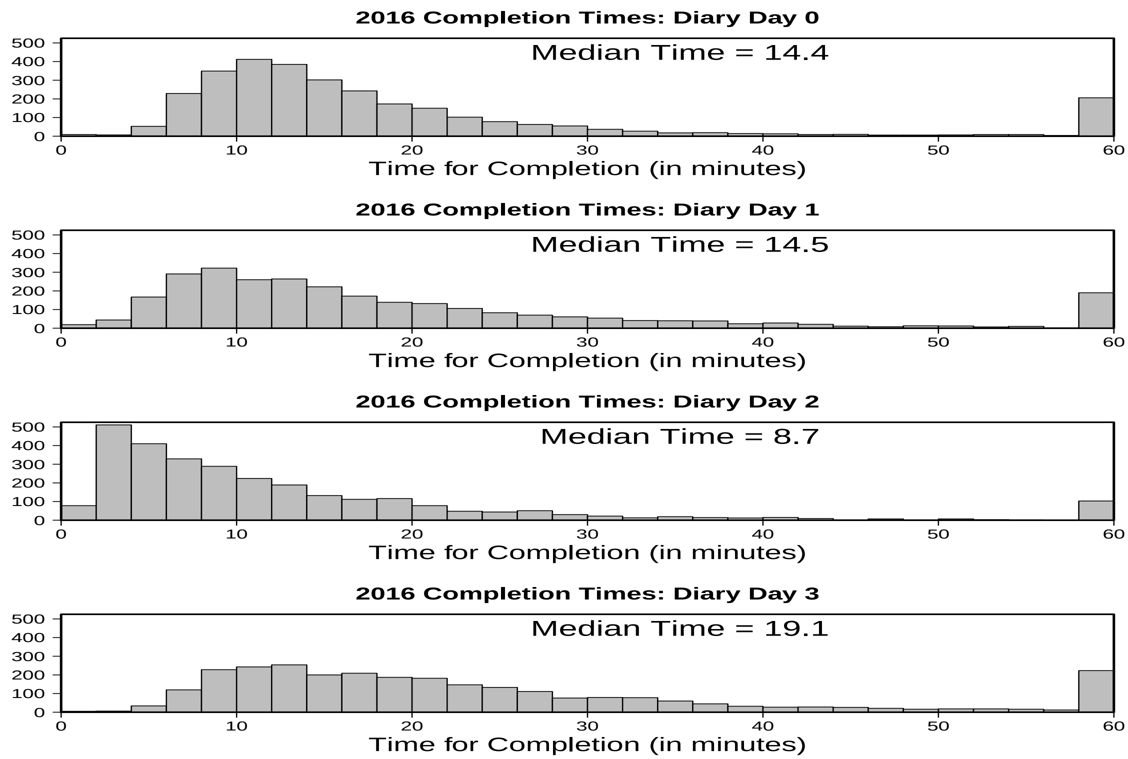

Figure 9: Distribution of the number of minutes to complete each diary day in 2016, with truncation to 60 minutes.

Source: Authors' calculations.

Table 2: Item nonresponse rates (percentage of data items missing) for three different transaction type modules.

Purchases						
Year	# Transactions	PI	Device	Location	Merchant	Time
2015	8,215	0.06	0.19	0.14	2.78	5.75
2016	11,347	0.72	0.25	0.04	2.34	6.53

Bills				
Year	# Transactions	PI	Device	Location
2015	1,385	0.65	2.23	1.37
2016	1,893	0.42	2.06	1.27

Cash Withdrawals				
Year	# Transactions	Location	Source	Time
2015	524	0.00	0.19	6.30
2016	781	0.00	0.00	6.40

Source: Authors' calculations.

4 Sampling Weights

4.1 DCPC Sample Demographics

An important goal of the DCPC is to provide estimates of payment statistics for the entire population of U.S. consumers over the age of 18. Generally, such a goal is accomplished by using weights to scale the sample data so that they appropriately match the population of interest. For the DCPC, the selection of households was done according to calculable probabilities, but the selection of adults from each household was not. Consequently, the set of DCPC respondents is not a probability sample, making it impossible to generate population-wide inferences from probability-based weighting. Instead, we rely on poststratification weights that perform the scaling with respect to certain demographic variables.

Although much effort was devoted to making the UAS a representative sample of the U.S. population, the variations in recruitment rates across demographic groups created non-trivial discrepancies. For example, the proportion of females in the UAS is close to 60 percent, much higher than in the general population. Nevertheless, as the UAS continues to grow, the fact that we are not necessarily sampling the entire American Life Panel (ALP) means the SCPC and DCPC subsets can be selected to better match population demographics.

Table 3 shows the unweighted and weighted sample proportions for a set of chosen demographic categories for the 2015 and 2016 DCPC samples. Interestingly, it seems that the 2015 DCPC is more representative of the U.S. population, most notably in terms of gender and race. The 2015 sample might be helped by the GfK sample, which does a better job of matching the U.S. population and thus nudges the characteristics of the full sample closer to those of the population. Overall, both samples are older, more educated, and more likely to be male than the general population. Education level, in particular, shows the worst representation. This is because it is not one of the characteristics used to screen potential participants during the sample selection process.

4.2 Daily vs. Individual Weights

To enable better inference for the entire population of U.S. consumers, SCPC respondents are assigned post-stratified survey weights designed to align as much as possible the composition of the SCPC sample with that of a reference population. Specifically, each year the benchmark distributions against which SCPC surveys are weighted are derived from the Current Population Survey. This follows common practice in other social science surveys, such

Table 3: Unweighted percentages for various unweighted and weighted marginal demographics in the 2015 and 2016 DCPC samples. The weighted values are based on CPS data.

Demographics		Unweighted 2015 DCPC	Weighted 2015 DCPC	Unweighted 2016 DCPC	Weighted 2016 DCPC
Gender	Male	47.6	47.7	43.1	46.8
	Female	52.4	52.3	56.9	53.2
Age	18−24	4.9	9.2	3.9	5.6
	25−34	17.6	20.4	17.4	23.7
	35−44	20.8	15.8	20.0	16.6
	45−54	20.0	19.3	20.3	18.1
	55−64	20.2	16.2	22.9	17.1
	65 and older	16.6	19.1	15.5	19.0
Race	White	76.6	77.0	82.5	76.4
	Black	9.1	9.1	8.3	12.5
	Asian	4.2	4.1	2.1	3.0
	Other	10.1	9.9	7.1	8.1
Education	No HS diploma	5.4	10.4	4.5	7.6
	High School	17.3	32.3	18.3	32.4
	Some College	22.9	18.3	23.6	18.4
	College	36.3	25.7	37.5	28.3
	Post-graduate	18.1	13.2	16.1	13.4
Income	$< \$25K$	20.1	21.7	21.8	22.4
	$\$25K - \$49K$	22.0	24.2	23.1	24.2
	$\$50K - \$74K$	19.3	19.0	20.0	17.7
	$\$75K - \$99K$	13.6	11.7	13.4	11.7
	$\$100K - \$124K$	10.2	9.6	9.2	10.3
	$\$125K - \$199K$	11.4	10.7	9.6	10.6
	$\geq \$200K$	3.6	3.2	3.0	3.1

Source: Authors' calculations.

as the Consumer Expenditure Survey (CES). The 2015 and 2016 DCPCs generate weights on an individual basis for each respondent and on a daily basis for each respondent who participated on each day in October.

The two sets of weights, as opposed to one, offer more flexibility in generating population estimates. While daily weights naturally allow calculations for a larger variety of time periods, individual weights correspond to a specific period of time. For the 2012 and 2016 DCPCs, individual weights correspond to the month of October, while for the 2015 DCPC, the individual weights in the public-use dataset represent the measurement period, October

13 through December 17. Weights based on other periods of time, including only the days in October or the first 31 days of the measurement period, are available in the "*full_weights.dta*" dataset and described in the document "2015 Survey and Diary of Consumer Payment Choice Weighting Procedure," both of which can be found on the DCPC website.

Applied to the same reference period, the two sets of weights offer competing estimates. The details of these estimates are described in Section 6, but we discuss their general differences here. Individual weights are calculated for each participant, meaning they are based on the demographic composition of the entire sample (about 1,300 people in 2015 and 2,800 people in 2016). Daily weights are calculated separately for the set of respondents who participated on each day, for those days that feature all three diary days (October 16 through December 14 in 2015 and October 1 through October 31 in 2016). This difference in the number of people is important, because when a smaller number is available for daily calculations the demographic strata must be more coarse than they are for individual weights. Therefore, the daily weights are less equipped to adjust for heterogeneity due to demographic differences where individuals are grouped in broader categories. On the other hand, estimates based on individual weights do not adjust for different numbers of observations on different days of the measurement period, thus giving some daily means more weight than others by virtue of having more respondents on those days. The use of daily weights in estimation accounts for the non-uniform distribution of respondents across days by assigning equal weight to each daily mean (see Section 6 for details). Based on these principles, daily weights are more appropriate when heterogeneity in behavior comes from temporal changes, such as day effects, while individual weights are more appropriate when heterogeneity is attributed more to demographic differences.

The result of having individual and daily weights is that each respondent has one individual weight and multiple daily weights that correspond to the number of days of diary participation. For most respondents, this number is three, though the roughly 500 people in the subset who took the diary twice in 2015 have six sets of daily weights. In addition, individuals who did not participate on all assigned days or had some diary days at the very beginning or end of the measurement period have fewer than three daily weights. Figures 10 and 11 show a scatterplot of participants' individual weights versus the average of their daily weights in each year. Although there are differences between the two, due to the random fluctuations in daily assignments and different raking procedures for generating each set of weights, the general trend shows consistency of the weights for each respondent. A more in-depth comparison of daily and individual weights in the context of population estimates is found in Section 6.

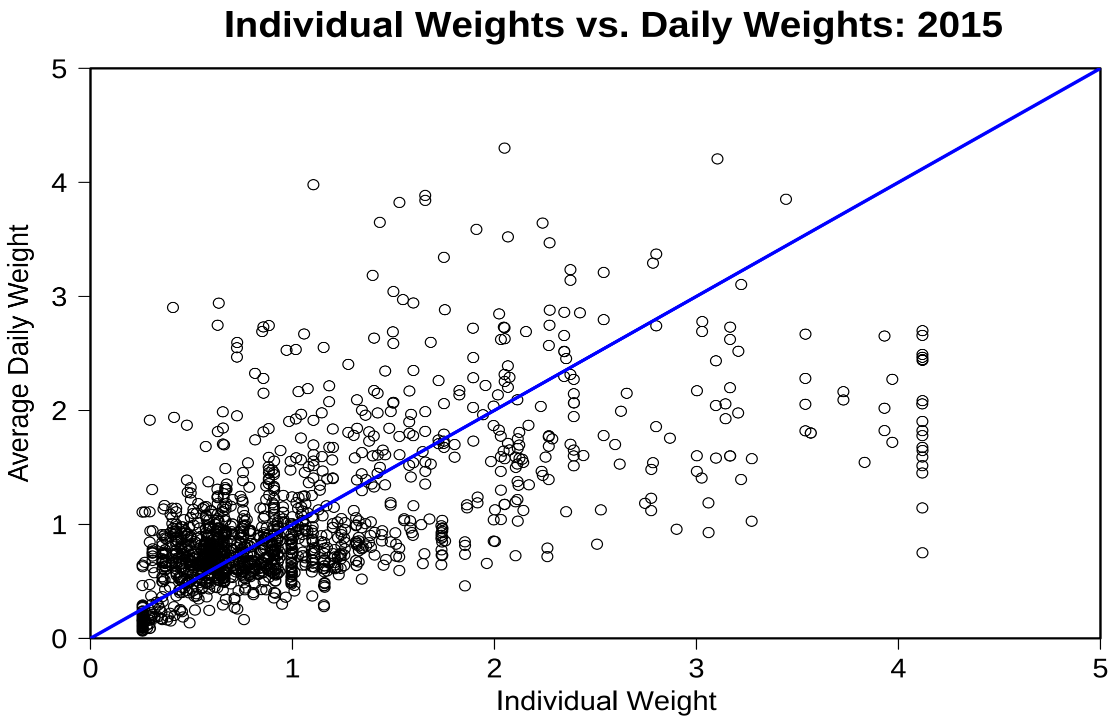

Figure 10: Individual weights vs. mean of daily weights for each respondent in 2015. Source: Authors' calculations.

4.3 Raking Algorithm

The CESR uses a raking algorithm (Deming and Stephan 1940 and Gelman and Lu 2003) to generate sampling weights. This iterative process assigns a weight to each respondent so that the weighted distributions of specific socio-demographic variables in the SCPC sample match their population counterparts (benchmark or target distributions). The weighting procedure involves two main steps. In the first, demographic variables from the CPS are chosen and mapped onto those available in the DCPC. Continuous variables such as age and income are recoded as categorical variables by assigning each to one of several disjoint intervals. For example, Table 4 shows five classifications for age and four classifications for income. The number of levels for each variable should be small enough to capture homogeneity within each level, but large enough to prevent strata containing a very small fraction of the sample, which could cause weights to exhibit considerable variability.

Table 4 shows the variables used in weighting as well as the levels within each variable for individual weights. Comparable information used for daily weights is shown in Table 5. In the second step, the raking algorithm is implemented and sample weights are generated by

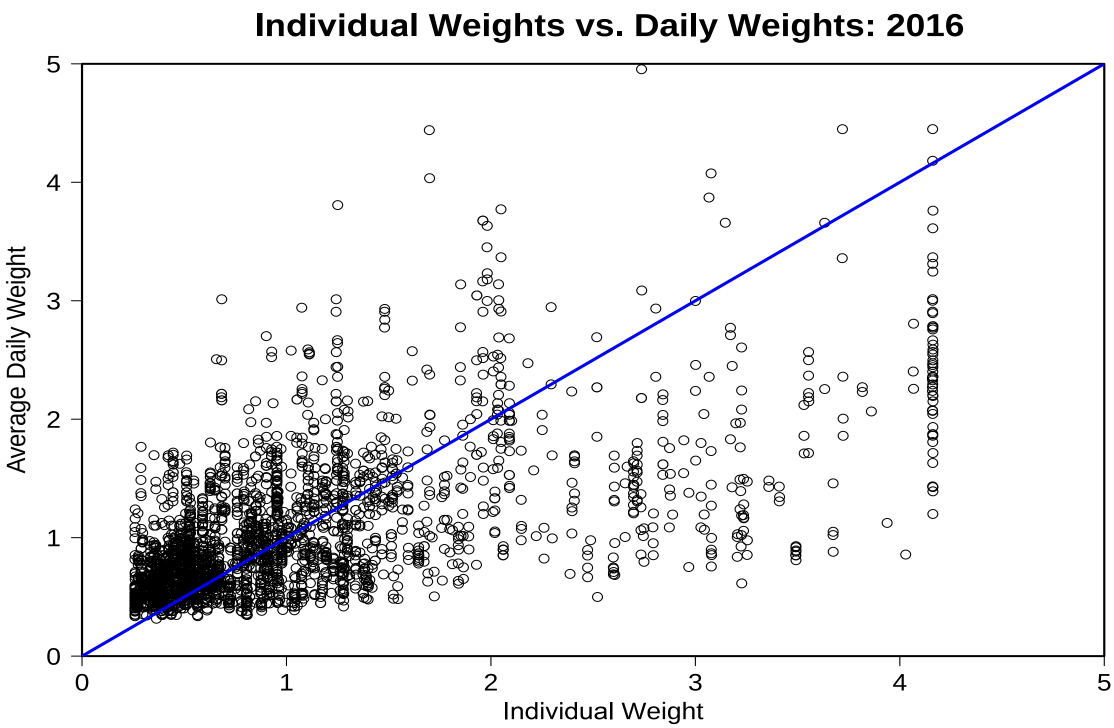

Figure 11: Individual weights vs. mean of daily weights for each respondent in 2016. Source: Authors' calculations.

Table 4: The set of weighting variables for the individual weights. "M" stands for male, and "F" stands for female. The highest income brackets for single households were combined to avoid small cell sizes.

Gender × Age				
M, $18-32$	M, $33-43$	M, $44-54$	M, $55-64$	M, 65+
F, $18-32$	F, $33-43$	F, $44-54$	F, $55-64$	F, 65+

Gender × Ethnicity	
M, White	M, Other
F, White	F, Other

Gender × Education		
M, High School or Less	M, Some College	M, Bachelor's Degree or More
F, High School or Less	F, Some College	F, Bachelor's Degree or More

Household Size × Household Income			
Single, $< \$30K$	Single, $\$30K - \$59K$	Single, $\geq 60K$	
Couple, $< \$30K$	Couple, $\$30K - \$59K$	Couple,$\$60K - \$99K$	Couple, $\geq \$100K$
≥ 3 , $< \$30K$	≥ 3 , $\$30K - \$59K$	≥ 3 ,$\$60K - \$99K$	≥ 3 , $\geq \$100K$

matching the proportions of predefined demographic groups in the DCPC to those in the CPS. More precisely, the weighting algorithm is performed using the 31 pairs of demographic variables shown in Table 4 for individual weights and the 20 pairs of demographic variables shown in Table 5 for daily weights.

Table 5: The set of weighting variables for the daily weights. "M" stands for male, and "F" stands for female. The highest income brackets for single households were combined to avoid small cell sizes.

Gender $\times$ Age		
M, 18 – 39	M, 40 – 64	M, 65+
F, 18 – 39	F, 40 – 64	F, 65+

Gender $\times$ Education		
M, High School or Less	M, Some College	M, Bachelor's Degree or More
F, High School or Less	F, Some College	F, Bachelor's Degree or More

Household Size $\times$ Household Income		
Single, $< \$40K$	Single, $\geq 40K$	
Couple, $< \$40K$	Couple, $\$40K - \$75K$	Couple, $\geq \$75K$
≥ 3 , $< \$40K$	≥ 3 , $\$40K - \$75K$	≥ 3 , $\geq \$75K$

The socio-economic variables chosen for the raking procedure carry over from Rand's internal research on the sampling properties of weights based on different demographic factors. First, we developed a new imputation algorithm for all possible socio-demographic variables to allow for weights based on a wider range of consumer information. The procedure is sequential, so that variables with the fewest missing values are imputed first and, in turn, used as inputs to impute the variables with the most missing values. We perform imputations by ordered logistic regression for ordered categorical variables and by multinomial logistic regression for categorical variables. Sample weights produced by different combinations of variables are evaluated on the basis of how well they match the distributions of demographic variables not used as raking factors (test variables). To assess the robustness and accuracy of different combinations of weighting variables, we draw Monte Carlo samples and generate demographic distributions of the test variables based on the weights for that particular sample. We estimate mean deviation from the CPS-defined levels for test variables by averaging over the samples. The combination of variables in Table 4 consistently matches the target distributions of the CPS for a variety of different sample sizes.

Pairing gender with other socio-demographic variables allows us to better correct for discrepancies between distributions within each gender while avoiding the problem of small cell counts. In other words, implementing the raking algorithm on the sets of pairs shown in

Table 4 and Table 5 ensures that the distributions of age, ethnicity, and education in the DCPC are matched separately for men and women to their population counterparts in the CPS. Moreover, since bivariate distributions imply marginal distributions for each of the two variables, this approach also guarantees that the distributions of gender, age, ethnicity, and education for the entire DCPC sample are aligned with the corresponding benchmarks in the CPS. The same is true for household size and household income.

If necessary, we impute missing information about education, household size, and income using ordered logistic regression with gender and age as predictors. Race is imputed using logistic regression. The procedure is sequential, so that variables with the fewest missing values are imputed first and, in turn, used as inputs to impute the variables with the most missing values. We perform imputations by ordered logistic regression for ordered categorical variables and by multinomial logistic regression for categorical variables. The order of imputation moves from variables with the fewest missing entries to those with the most, and each level of imputation uses all previous demographic variables. In the DCPC, income is the most commonly missing variable, yet in 2015 only 13 respondents did not provide an income range in 2015, and 17 failed to do so in 2016. The imputations are used only to generate post-stratification weights and are left as missing in the dataset.

We describe the raking algorithm in greater detail below. Let $d = 1, \ldots, 4$ represent the four bivariate demographic categories presented in Table 4. For each demographic, we index the possible groups with the variable g, so that d_g represents group g for demographic category d. For example, $d = 1$ corresponds to the intersection of gender and age, and group $g = 1$ might correspond to males ages 18 to 32. We use the shorthand notation $i \in d_g$ to indicate that individual i belongs to group g in demographic d. Then, we define $w_i^{(k)}$ as the weight assigned to individual i after iteration k of the raking algorithm. Within each such iteration, the raking algorithm iterates across demographic groups d, so we let $w_i^{(k,d)}$ define the assigned weights after iterating over d.

The weighting algorithm begins by assigning base weights that are intended to account for the presence of the Native American subset, effectively down-weighting the contribution of its members so that the proportion of total weight assigned to the Native American sample matches the frequency in the U.S. adult population. For 2015, this means that

$$w_i^{(0,4)} = \begin{cases} 0.26, & \text{if Native American} \\ 1.04, & \text{if General Population,} \end{cases}$$

while for 2016, $w_i^{(0)} = 1$ for all i, because respondents from the Native American cohort were not included in the 2016 sample.

Within each iteration, $k = 1, 2, 3, \ldots$, we do the following, mirroring the algorithm found in Valliant, Dever, and Kreuter (2013):

1. $w_i^{(k,0)} = w_i^{(k-1,4)}$

2. Otherwise, for $d = 1, \ldots, D$, we let $w_i^{(k,d)} = w_i^{(k,d-1)} m_{k,s_d[i]}$, where

$$m_{k,s} = \frac{\sum w_i^{(k,d-1)} 1\left[i \text{ in stratum } s\right]}{\sum w_i^{(k,d-1)}} \times f_{d,g},$$

   where $f_{d,g}$ represents the proportion of the U.S. population that belongs in group g of demographic d, and $1\left[i \text{ in stratum } s\right]$ is 1 if individual i belongs in stratum s and is 0 otherwise. Therefore, after iteration d, the weighted marginal frequencies in the sample for demographic d will perfectly match those in the population.

3. Trim weights by letting $\bar{w}^{(k)}$ represent the average weight within the sample and then assign weight values according to:

$$w_i^{(k,4)} = \begin{cases} 0.25\bar{w}^{(k,4)}, & \text{if } w_i^{(k)} < 0.25\bar{w}^k \\ 4\bar{w}^{(k,4)}, & \text{if } w_i^{(k,4)} > 4\bar{w}^k \\ w_i^{(k,4)} & \text{else.} \end{cases} \tag{1}$$

Therefore, in each iteration, weights that are less than a quarter of the average or more than four times the average are trimmed. We trim to reduce large weights and increase small weights, thereby decreasing the variation in the weights. While this trimming may sacrifice the unbiasedness of the estimators, it does so by reducing the mean-squared error, which is adversely affected by high variations in weights. Although about 5 percent of the unweighted 2015 sample comprises members of the Native American cohort, the weighted proportion of 2.1 percent is much more in line with the group's 1.7 share of the general population.

The CESR runs 50 iterations of this algorithm or until all marginal weight matching and trimming specifications are achieved. Upon convergence, we let w_i represent the weight given to individual i. Weights are standardized to have a mean of 1.0, so the maximum weight is 4.0 and the minimum weight is 0.25. Of the possible 660 strata represented in each sample, there are, overall, 495 unique weights for 2015 and 484 unique weights for 2016. While more strata are represented in 2015, the per-stratum numbers for 2016 are superior. In 2016, only 95 strata have one representative in the sample, while 207 have one in the 2015 sample. The decrease in variability of the weights from 2015 to 2016 is greater than would be expected from the increase in sample size, suggesting improved efficiency of estimators. An increase in sample size from about 1,300 to about 2,800 would generally lead to an increase in standard errors of approximately 1.5. Instead, the variability increases by a factor of 1.8.

Because the UAS sample itself is not representative of the U.S. population, post-stratification is an important step for making inferences about the population. Because not all strata of interest are represented in the sample, raking is the natural method for assigning weights. However, raking introduces a few complications related to the statistical framework and analysis of the data. The first concerns the increased difficulty in calculating standard errors of population estimates, which are weighted averages of the sample values. In all tables and publications, the standard errors have been calculated by taking the weights as fixed values, thereby reducing the standard errors. The sampling weights, which are a function of the strata representation in the sample, are random variables, and their variation should be factored into the calculation of standard errors (Gelman and Lu 2003). However, due to high response rates and targeted sampling (as described in Section 3.2), the variability in the observed sample composition is small, which implies that the variability in the raked weights is small. Therefore, conditional on the chosen weighting scheme, the variance of our estimators can be largely attributed to the variation in the observed responses themselves and not in the sample composition.

The second concern involves the effects of the sampling scheme on the weights and on the estimates they produce. For the raking algorithm to be appropriate in the sense that the expected weights for each stratum equal those of the population, each stratum should be proportionally represented in the sample. To be precise, the expected proportion of the sample belonging to a specific stratum is directly proportional to the relative proportion of that stratum within the population. A sampling procedure lacking this property is likely to consistently produce weights for certain strata that do not reflect the true representation in the entire population. If strata properties correlate with payment behavior, this could lead to population-wide biased estimates. In a sampling procedure in which some strata tend to be over-represented and others under-represented, the raking algorithm, which strives to match marginal proportions rather than those of the cross-sections of all the variables, may generate sample weights that fail to align the sample composition with the reference population. Although the sample from the UAS does not perfectly reflect the U.S. population (for example, it tends to include more females than males), the differences between the panel and the broader population are relatively small for the demographics used in weighting. In addition, for many DCPC variables there is little evidence of strong correlations with the variables used in weighting, so any bias is likely to be small.

5 Data Preprocessing

Before conducting further statistical analysis, it is important to carefully examine and ensure data quality. For the DCPC, this quality check primarily involves properly identifying all unique, valid transactions and confirming details, or attributes, of those transactions. While this process mostly requires the reclassifying of details of reported transactions, the CPRC also assesses the validity of certain numerical values reported in the data. Below, we detail the individual steps taken to convert the raw data entries into those found in the public-use dataset.

5.1 Identifying Valid Transactions

The primary unit of measurement in the DCPC is the transaction; the final dataset is organized so that each row corresponds to a transaction. The public-use dataset includes two fundamental types of transactions: those that indicate participation of a respondent on a particular day and those that represent economic transactions. The latter are defined by some form of exchange or transfer of money and thus must have a dollar value. So, the first step is to limit the data to only those transactions associated with non-missing and positive dollar values. Specifically, we exclude from the final data the small number of transactions in which respondents recorded a dollar value of zero or omitted the value. The next step of the editing process is to ensure that each entered transaction is unique.

Data collection in the DCPC is defined by modules that correspond to different types of transactions. The 2015 and 2016 DCPCs have 14 such modules, each of which corresponds to a common transaction such as a purchase, cash withdrawal, or income received and to a less common one such as returned goods, prepaid card loading, or a cash-to-coin exchange. The set of modules is designed to be exhaustive, but not necessarily disjoint. Therefore, transactions can, and often are, entered in two different modules. For example, a cash deposit may also correspond to income received.

In identifying duplicates in the DCPC, we take a conservative approach, requiring transactions to match on most attributes available for each transaction module. This matching involves some subset of the respondent, diary day, dollar value, payment instrument used, device used, location of transaction, and source of money. Identified duplicates are consolidated into one transaction defined by information from both transaction modules. Table 6 shows the types and number of duplicates in each year. The most common form of duplicate results from bill payments also being entered as regular purchases, which is understandable since the bills-reminder section is featured on the final diary day. Although respondents are asked not to enter bills that they already included as a purchase, many ignored this instruction. The second most common duplicate occurs within the bills framework. This is because the bills module prompts specific bills, some of which may correspond to a single payment. The most prominent example is that of the telephone, cable, and internet bills, which the diary asks about individually and in different combinations. Therefore, someone who bundles cable and internet will make only one payment for both but may enter this payment as many as three times: once for cable payment, once for internet payment, and once for cable and internet together.

Table 6: The number of identified transaction duplicates across data entry modules.

Duplicate Type	**# in 2015**	**# in 2016**
Bill/Bill	31	68
Cash Deposits/Checking Deposits	26	37
Income/Cash Deposits	14	28
Income/Checking Deposits	199	330
Puchases/Bills	712	863

Source: Authors' calculations.

The final dataset available for public-use is based on individuals who participated and completed the diary on all days. Making this reduction, as well as removing duplicates, drops the dataset to 11,503 transactions for 2015. Of those, 8,959 are expenditures, 1,323 account transfers, and 1,221 income payments. For 2016, there are 12,407 expenditures, 1,888 transfers, and 1,916 income payments, for a total of 16,211 transactions.

5.2 Editing Details of Transactions

We also make certain edits to the fields that describe transaction attributes, although this is relatively rare. Most often, we use information provided in the text box for the "other (specify)" option (instead of the multiple-choice selection) to recode the response to an existing response option. In each such case, the entered text remains in the original data, but it is not part of the public-use dataset due to the possibility of including information that could identify the participant.

The most common form of this type of edit relates to classifying purchases as bills or nonbills. Specifically, information provided in open-ended questions often reveals that payments reported in the purchases module were bills, though the respondent did not classify them as such. In 2015, the CPRC identified 69 such payments, while in 2016, there were 98.

5.3 Editing Cash Holdings

The DCPCs asked respondents every day to enter the number of bills for each denomination of bill they owned ($1, $5, $10, etc.) and to confirm the total value of cash on hand, which was automatically calculated but could be overwritten. On rare occasions, a respondent entered a dollar value instead of a number of bills. In those cases, the dollar value was replaced by the number of bills implied by dividing the reported dollar value by the bill denomination value. When non-numeric responses were entered, we reset all related variables to zero.

5.4 Editing Dollar Values

The greatest challenge in data preprocessing for the DCPC involves the dollar values of transactions. Measurement errors, defined as any incongruity between the data entry and the true value, can be attributed to sources ranging from recall error to rounding errors to data-entry errors. In practice, there is an asymmetry to identifying measurement error, as we can really hope to identify only the values that are too high (Chambers and Ren 2004).

In determining the editing philosophy, it is important to distinguish between influential and likely invalid data points. An influential point is one whose inclusion or exclusion in any inferential analysis causes a significant difference in estimates (Bollen and Jackman 1990 and Cook and Weisberg 1982), and thus the influence of a point depends on the statistical procedure being performed. Influential points can be identified through a variety of statistical measures, including Cook's distance (Cook 1977 and Cook and Weisberg 1982). An invalid data entry is, technically, any entry that does not represent the truth. An invalid data point does not have to be influential, and an influential point is not necessarily invalid. Our goal is to identify transaction values that are likely impossible.

Identification of data that are technically possible, but very unlikely, is difficult, because it involves comparing a data entry in the context of heterogeneity of behavior within the population. This is especially true for economic variables such as dollar values, which have fat right tails. In other words, it is possible that data entries that, by some numerical evaluations, are statistical outliers are actually accurate and valid. This issue is not unique to the DCPC. Many consumer surveys, including the Survey of Consumer Finances (SCF) and the Consumer Expenditure Survey (CES) must also tackle the cleaning of such fat-tailed variables. While the details of the preprocessing of outliers are not provided in either survey, the general strategy is similar to that adopted in the DCPC (Bricker et al. 2012 and U.S. Bureau of Labor Statistics 2013): use all available intuition and information about value distributions to assign a degree of plausibility. This process often involves making subjective decisions.

It is possible to use model-based methodologies to identify transaction values that are implausible, and the perceived objectivity of such an approach is certainly desirable. However, this is difficult to do in practice. Fitting parametric models is somewhat subjective to begin with, because we do not know how right tails of dollar values should behave. Choosing different distributions, such as a log-Normal or Gamma distribution, will lead to different conclusions, and the distribution parameters are largely determined by the data in the center of the distribution anyway. The question of how to best pool and analyze data is also complicated, because there is potentially a lot of information about the transaction that should inform the decision. The merchant, payment instrument used, and individual characteristics of the payer, such as household income, naturally relate to the degree of plausibility. Hospital bills are quite plausibly in the thousands of dollars, but such values are rare for convenience-store payments; large-value payments to financial advisers are less likely for those with low income than for those with high income. However, in separating data into finer categories with distinct distributions, we must rely on fewer data to make inferences about behavior in the right tail and thus lose some ability to identify extreme outliers.

Instead, the DCPC dollar-value cleaning relies on subjective evaluation by the CPRC of likely errors for the largest reported dollar values (over $1,000) and one negative-valued entry. Assessments were based on all available information, most notably from the merchant category and open-ended text responses, which often feature important details of the transaction. The CPRC's general assumption is that decimal points were likely misplaced or omitted, and consequently intended dollar values were recorded as much larger, and often implausible, transaction values. In the public-use dataset, the value provided by the respondent in the diary corresponds to the variable `amnt_orig`. Our editing approach identifies implausible transaction values and sets the variable `amnt` for those transactions to a better guess of the true value. Table 7 shows a set of 21 transactions, 15 from 2015 and six from 2016, for which changes to dollar amounts were made. We stress that the original data, absent of open-ended text responses, are found in the final dataset, so researchers can edit transaction values as they like.

Table 7: Large-value transactions for which edits were made.

Date	Type	Merchant/ Source	PI	Original Amount ($)	New Amount ($)
11/12	Purchase	Retail store	Prepaid	7,132	761.32
11/02	Bill	Services	Debit	6,650	66.50
11/02	Bill	Services	Debit	8,084	80.84
11/02	Bill	Services	Debit	8,495	84.95
11/15	Purchase	Retail store	Credit	2,400	24.00
11/09	Purchase	Retail store	Cash	1,400	14.00
11/26	Bill	Financial services	BANP	67,000	670.00
12/16	Purchase	Retail store	Credit	1,532	15.32
12/08	Bill	Financial services	OBBP	34,500	345.00
11/02	Bill	Financial services	BANP	1,358,086	1,358.86
11/07	Purchase	Retail store	Credit	5,125	51.25
11/03	Bill	Services	Cash	4,300	43.00
11/13	Purchase	Retail store	Debit	2,946	29.46
11/17	Bill	Financial services	OBBP	22,350	223.50
12/05	Transfer	Retail store	Credit	21,900	219.00
10/20	Purchase	Retail store	Check	19,500	195.00
10/11	Purchase	Retail store	Credit	-3.76	3.76
09/29	Bill	Services	Cash	73,942	739.42
10/09	Purchase	Retail store	Prepaid	12,100	121.00
10/25	Income	NA	NA	21,027	21.27
10/24	Transfer	NA	NA	14,000	140.00

6 Population Parameter Estimation

An important goal of the data collection in the DCPC is to produce estimates of consumer payment behavior for the entire population of U.S. consumers. This can be accomplished using a variety of methodologies that factor in different aspects of the collected data, such as effects associated with days of the week or monthly cycles. In the sections below, we present a simple methodology for generating point estimates and standard errors for two general types of variables: those that define an average consumer and those that define an average transaction. The adopted framework has the benefit of being applicable to a wide range of variables and does not require distributional assumptions. We provide formulas based on individual weights and daily weights and discuss how the various estimates relate to one another. For expository purposes, we assume the period of interest is the month of October, as it is for the 2012 and 2016 DCPCs, but the formulas can be generalized to any period of interest.

We begin with some notation. Every transaction is indexed by the triplet (i, d, j), where $i = 1, \ldots, N$ identifies the respondent, d defines the date of the transaction, and j enumerates the transactions made by individual i on day d. Each transaction has various associated attributes for which we use the general notation Y_{idj}. These attributes can take one of three forms: 1) a numerical value, such as the dollar value of the transaction; 2) a binary variable, for example, 1 if cash was used and 0 otherwise; or 3) a combination of the two, such as the dollar value of the transaction if cash was used and 0 otherwise. We refer to individual weights, corresponding to the variable `ind_weight`, as w_i and to daily weights, corresponding to the variable `day_weight`, as v_{id}.

6.1 Per-consumer Estimates

Per-consumer estimates define an average for all consumers over a designated period of time determined by the dates of data collection. Take $\theta(Y)$ as some population value corresponding to the attribute Y, such as the average amount spent in October or the average number of cash payments in October.

Our methodology yields estimates that are linear combinations of observed daily totals, given by $Y_{id} = \sum_{(i,d,j)} Y_{idj}$. The daily total is natural given that the fundamental sampling unit is the consumer-day. It also helps account for the fact that not all individuals participated for three days in October, either because some assigned days were in September or November or due to non-responses. We let D_i represent the number of individual i's diary days that are in October. Thus, $D_i = 1, 2,$ or 3.

To generate individual estimates for October, we restrict the analysis to data from that month only. For certain economic concepts it may be reasonable to assume that data collected at the end of September or the beginning of November have the same distribution as those from October. In such a case, it is beneficial to incorporate the additional data in the October estimates. However, for variables that have distributions at the beginning or end of a month that differ from those in the middle of the month, doing so would bias results. For example, if rent payments tend to be scheduled for the last day of the month, including data from the end of September as well as the end of October in our estimates would effectively count such payments twice. Therefore, all per-consumer estimates have the general form

$$\hat{\theta}(Y) = \sum_{i=1}^{N} \sum_{d \in Oct} k_{id} Y_{id}, \qquad (2)$$

where k_{id} represent scalar weights. The weights, k_{id}, can be constructed using individual or daily weights. Naturally, daily weights can be used to generate population estimates for time periods other than October, such as a particular day or a portion of the month. However, we do not detail how this would be done, because it is a trivial extension of the individual estimate discussed below.

6.1.1 Individual Weights

An estimate for an average for the month of October based on data from N respondents might take the general form

$$\hat{\theta}^{month}(Y) = \frac{\sum_{i=1}^{N} w_i \hat{\theta}_i(Y)}{\sum_{i=1}^{N} w_i},$$

where $\hat{\theta}_i(Y)$ is an estimate of the monthly mean for individual i. This estimate will be unbiased if $\mathrm{E}\left[\hat{\theta}_i(Y)\right] = \theta_i(Y)$ and weights are appropriately assigned. Perhaps the simplest estimate of $\theta_i(Y)$ is just the corresponding estimate for the days of participation in October scaled up to a monthly basis, which in the case of a month with 31 days takes the form

$$\hat{\theta}_i(Y) = \frac{31}{D_i} \sum_{d \in Oct} Y_{id}.$$

If the days of participation assigned to individual i span the month uniformly (as if they were chosen at random), then $\mathrm{E}\left[\hat{\theta}_i(Y)\right] = \theta_i(Y)$. Then, the population estimate takes the form

$$\hat{\theta}^{month}(Y) = \frac{\sum_{i=1}^{N} w_i \left[\frac{31}{D_i} \sum_{d \in Oct} Y_{id}\right]}{\sum_{i=1}^{N} w_i}, \tag{3}$$

and $k_{id}^{month} = \frac{31 w_i}{D_i \sum_{i=1}^{N} w_i}$. We discuss calculating standard errors for (3) in Section 6.3.

6.1.2 Daily Weights

The principle behind estimates of θ based on daily weights is identical to that for the monthly estimates, except the methodology is applied to each day and the monthly estimate is a summation over all days in October:

$$\hat{\theta}^{day}(Y) = \sum_{d \in Oct} \frac{\sum_{i=1}^{N} v_{id} Y_{id}}{\sum_{i=1}^{N} v_{id}}. \tag{4}$$

The inner sum in (4) is a population estimate for day d. The sequence of sums in (4) can be rearranged to sum over individuals first and days second:

$$\hat{\theta}^{day}(Y) = \sum_{i=1}^{N} \sum_{d \in Oct} \frac{v_{id}}{\sum_{i'=1}^{N} v_{i'd}} Y_{id}, \tag{5}$$

so that $k_{id}^{day} = \frac{v_{id}}{\sum_{i'=1}^{N} v_{i'd}}$.

It is worth comparing k_{id}^{month} with k_{id}^{day} to see that the relative contributions of Y_{id} to a monthly estimate are similar, especially in expectation. Within k_{id}^{day}, the relative ratio of v_{id} to $\sum_{i=1}^{N} v_{id}$ represents the relative contribution of individual i to the estimate for day d. On average, therefore, $\mathrm{E}\left[k_{id}^{day}\right] \approx \frac{1}{N_d}$, where N_d represents the number of respondents on day d. By design of the survey, 3 out of 33 diary waves are participating on each day, meaning about one-eleventh of the individuals, on average, are participating on day d. Thus, $N_d \approx \frac{N}{11}$ and $\mathrm{E}\left[k_{id}^{day}\right] \approx \frac{11}{N}$. Using similar logic for k_{id}^{month}, we have $\mathrm{E}\left[k_{id}^{month}\right] \approx \frac{31}{D_i N}$. For virtually all respondents, $D_i = 3$, so that $\mathrm{E}\left[k_{id}^{month}\right] \approx \frac{11}{N}$. Therefore, through this approximation, we can see that, on average, the contribution of each relevant data point in the individual and daily weights is the same. This is a desirable quality in the estimates. Discrepancies are largely due to the lack of representative sampling across the days of the month and differences in raking algorithms used for daily and individual weights.

The calculation of standard errors using daily weights is discussed in Section 6.3.

6.2 Per-transaction Estimates

The use of the DCPC data to make estimates on a per-transaction basis is perhaps less obvious, because it is less intuitive that employing weights associated with consumers can yield values that correspond to an average transaction. An example of such a parameter might be the average value spent at a gas station or the proportion of payments that are made for medical expenses with a credit card.

In fact, the per-transaction estimates are simply ratios of per-consumer estimates as can be seen from the identity

$$\text{Average value per transaction} \quad = \quad \frac{\text{Average value of transactions per consumer}}{\text{Average number of transactions per consumer}}. \quad (6)$$

Therefore, a per-transaction estimate represents a share. The CPRC has found that such shares are robust and economically meaningful economic quantities. We note that these shares are based on the macroeconomic definitions, meaning that we consider the share of the averages rather than the average of the individual shares. The microeconomic alternative, which calculates a share for each individual and takes the average of those, is not recommended. Especially with limited data from a three-day diary, such a calculation will likely yield undefined estimates $\left(\frac{0}{0}\right)$ for individuals, which are difficult to deal with. In addition, within many frameworks, the macroeconomic estimates correspond to the maximum likelihood estimates of the share. For example, if for $i = 1, \ldots, N$ individuals, the number of observed transaction is T_i of which $S_i \sim \text{Binomial}(T_i, p)$ are of interest, then the maximum likelihood estimate of the share, p, is $\frac{\sum_{i=1}^N S_i}{\sum_{i=1}^N T_i}$ rather than $N^{-1} \sum_i \frac{S_i}{T_i}$.

Both the numerator and denominator in (6) are per-consumer estimates based on two different attributes, which we generically label $Y^{(1)}$ and $Y^{(2)}$. The denominator can be expressed as $\theta(Y^{(1)})$, using the general notation of (2), where $Y^{(1)}$ is a binary attribute that identifies the transactions of interest. If one is interested in all transactions, then $Y_{idj}^{(1)} = 1$ for all (i, d, j), while in the second example given in the previous paragraph, $Y_{idj} = 1\,[\text{transaction } (i, d, j) \text{ is for medical expenses}]$. The numerator is $\theta(Y^{(2)} \times Y^{(1)})$, where $Y^{(2)}$ can be any type of attribute. We use the notation $Y_{idj}^{(2,1)} = Y_{idj}^{(2)} \times Y_{idj}^{(1)}$. In the example relating to medical expenses, $Y_{idj}^{(2)}$ is a binary variable that is 1 if the the transaction was made with a credit card and 0 otherwise. The per-transaction estimate is thus estimated by

$$\hat{\mu}(Y^{(1)}, Y^{(2)}) \quad = \quad \frac{\hat{\theta}(Y^{(2,1)})}{\hat{\theta}(Y^{(1)})}. \quad (7)$$

The numerator and denominator in (7) can be calculated using individual or daily weights according to (3) or (5), respectively.

6.2.1 Individual Weights

Each per-consumer estimate can be generated using individual weights:

$$\hat{\mu}^{month}(Y^{(1)}, Y^{(2)}) \quad = \quad \frac{\sum_{i=1}^N \frac{w_i}{D_i} \sum_{d \in Oct} Y_{id}^{(2,1)}}{\sum_{i=1}^N \frac{w_i}{D_i} \sum_{d \in Oct} Y_{id}^{(1)}}. \quad (8)$$

Standard errors forms are described in Section 6.3.

6.2.2 Daily Weights

Estimates on a per-transaction basis can also be constructed using daily weights:

$$\hat{\mu}^{day}(Y^{(1)}, Y^{(2)}) = \frac{\sum_{d \in Oct} \frac{\sum_{i=1}^{N} v_{id} Y_{id}^{(2,1)}}{\sum_{i=1}^{N} v_{id}}}{\sum_{d \in Oct} \frac{\sum_{i=1}^{N} v_{id} Y_{id}^{(1)}}{\sum_{i=1}^{N} v_{id}}}. \tag{9}$$

Standard errors forms are described in Section 6.3.

6.3 Standard Errors

Standard errors are simply calculated as $\text{SE}\left(\hat{\theta}(Y)\right) = \sqrt{\text{Var}\left(\hat{\theta}(Y)\right)}$ or $\text{SE}\left(\hat{\mu}(Y^{(1)}, Y^{(2,1)}\right) = \sqrt{\text{Var}\left(\hat{\mu}(Y^{(1)}, Y^{(2,1)}\right)}$. We begin by discussing standard errors for per-consumer estimates and then for per-transaction estimates.

6.3.1 Per-consumer Estimates

Assuming independence across individuals, the general form of variances of per-consumer estimates of type (2) is

$$\text{Var}(\hat{\theta}(Y)) = \sum_{i=1}^{N} \text{Var}\left[\sum_{d \in Oct} k_{id} Y_{id}\right]. \tag{10}$$

Treating the weights, k_{id}, as fixed, the heart of the calculation involves determining $\text{Cov}(Y_{id}, Y_{id'})$ for all combinations of d, d'. Here again, various models and assumptions could be used. However, we rely on a relatively simple one. For any individual, let $Y_i = [Y_{id_1} Y_{id_2} Y_{id_3}]$ represent the vector of the measurements for individual i on three consecutive days. Then, the quantity of interest can be determined by modeling the $\text{Cov}(Y_{id}, Y_{id'})$ according to:

$$\text{Var}(Y_i) = \begin{bmatrix} \sigma^2 & \rho_1 & \rho_2 \\ \rho_1 & \sigma^2 & \rho_1 \\ \rho_2 & \rho_1 & \sigma^2 \end{bmatrix}.$$

Therefore, σ^2 represents the variance of the daily observations, ρ_1 represents the covariance of observations from consecutive days, and ρ_2 represents the covariance of observations from two days apart. This framework allows for the daily attributes' potentially being correlated. For example, we allow that an individual who makes a large number of purchases on one day may be more likely to make fewer purchases on the next day. The key simplification is that we assume stationarity, in that joint distributions depend only on the number of days that separate the observed data and not when those days occurred.

The weighted sum over observed days in October can be expressed in matrix notation as

$$\sum_{d \in Oct} k_{id} Y_{id} \quad = \quad P_i K_i Y_i^T,$$

where

$$K_i \quad = \quad \begin{bmatrix} k_{id_1} & 0 & 0 \\ 0 & k_{id_2} & 0 \\ 0 & 0 & k_{id_3} \end{bmatrix},$$

and P_i represents the three-dimensional vector, with 1 indicating that the data from that day were observed, and 0 indicating otherwise. Thus, if individual i participated on all three days, $P_i = [1\ 1\ 1]$, and if individual i participated only on the first assigned day, $P_i = [1\ 0\ 0]$. Then, the variance of the weighted sum is expressed as

$$\mathrm{Var}\left[\sum_{d \in Oct} k_d Y_{id}\right] \quad = \quad P_i K_i \Sigma\, K_i^T P_i^T. \tag{11}$$

We estimate the variance in (11) by replacing Σ with its estimate, $\hat{\Sigma}$, based on the following calculations:

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^N D_i - 1} \sum_{i=1}^N \sum_{d \in Oct} \left(Y_{id} - \bar{Y}\right)^2,$$

$$\hat{\rho_1} \quad = \quad \frac{1}{\sum_{i=1}^N (D_i - 1)} \sum_{i=1}^N \sum_{d_1, d_2 | d_1 - d_2 = 1} (Y_{id_1} - \bar{Y})(Y_{id_2} - \bar{Y}),$$

and

$$\hat{\rho_2} = \frac{1}{\sum_{i=1}^{N} 1[D_i = 3]} \sum_{i=1}^{N} \sum_{d_1,d_2|d_1-d_2=2} (Y_{id_1} - \bar{Y})(Y_{id_2} - \bar{Y}),$$

where

$$\bar{Y} = \frac{1}{\sum_{i=1}^{N} D_i} \sum_{i=1}^{N} \sum_{d \in Oct} Y_{id}.$$

$\bar{Y}$ is simply the unweighted, sample average, and all other calculations are sample covariances and variances for appropriate pairs of observations. Note that for ρ_1, individuals who participated on all three days have two pairs of observations that contribute to the estimate: (Y_{id_1}, Y_{id_2}) and (Y_{id_2}, Y_{id_3}). Thus, the standard error can be expressed in its most general form as

$$\mathrm{SE}(\hat{\theta}(Y)) = \sum_{i=1}^{N} P_i K_i \hat{\Sigma} K_i^T P_i^T. \tag{12}$$

It is worth considering the specific forms of the standard errors, especially for monthly estimates, in which there is simplification because $k_{id} = k_i$ does not depend on the day d. Table 8 shows the variances for individual sums when using individual weights. Notice that when cross-day covariances are $\rho_1 = \rho_2 = 0$, the variance is proportional to D_i, the number of days participated. For certain variables, this may be an adequate simplification.

Table 8: Variance of weighted daily sums based on individual weights for different participation patterns in October.

Participation Dates	D_i	P_i	Variance
(9/29, 9/30, 10/1)	1	[0 0 1]	$k_i^2\sigma^2$
(9/30, 10/1, 10/2)	2	[0 1 1]	$k_i^2\,(2\sigma^2 - 2\rho_1)$
All 3 days in October	3	[1 1 1]	$k_i^2\,(3\sigma^2 - 4\rho_1 - 2\rho_2)$
(10/30, 10/31, 11/1)	2	[1 1 0]	$k_i^2\,(2\sigma^2 - 2\rho_1)$
(10/31, 11/1, 11/2)	1	[1 0 0]	$k_i^2\sigma^2$

6.3.2 Per-transaction Estimates

For per-transaction estimates, the variances are more complicated, because the form of the estimator is a ratio of two random variables. Using properties of variances, we have

$$\mathrm{Var}(\hat{\mu}(Y^{(1)}, Y^{(2)}) \;=\; \mathrm{E}\left(\mathrm{Var}\left[\frac{\hat{\theta}(Y^{(2,1)})}{\hat{\theta}(Y^{(1)})}\middle|\{Y_{id}^{(1)}\}\right]\right) + \mathrm{Var}\left(\mathrm{E}\left[\frac{\hat{\theta}(Y^{(2,1)})}{\hat{\theta}(Y^{(1)})}\middle|\{Y_{id}^{(1)}\}\right]\right). \quad (13)$$

Because the expected value of $\frac{\hat{\theta}(Y^{(2,1)})}{\hat{\theta}(Y^{(1)})}$ does not depend on the number of observations and, therefore as some fixed value, has a variance of zero, the second term on the right-hand side of (13) is zero. Conceptually, this corresponds to the idea that the average value of a transaction is the same no matter how many transactions are observed. The variance estimate is thus

$$\mathrm{Var}(\hat{\mu}(Y^{(1)}, Y^{(2)}) \;=\; \mathrm{E}\left[\frac{1}{\hat{\theta}(Y^{(1)})}\right]^{-2} \mathrm{Var}\left[\hat{\theta}(Y^{(2,1)})\right].$$

While it is possible to estimate the expected value of the inverse of the square of $\hat{\theta}(Y^{(1)})$, especially using a Normal approximation and the delta method Casella and Berger 2002, it is common to approximate this term simply by the inverse square of the observed value itself, $\hat{\theta}(Y^{(1)})$. In practice, when the number of transactions observed is high, this simplification will have a trivial effect on the estimate. Thus, the variance estimate becomes

$$\hat{\mathrm{Var}}\left[\hat{\mu}\left(Y^{(1)}, Y^{(2)}\right)\right] \;\approx\; \left[\hat{\theta}(Y^{(1)})\right]^{-2} \mathrm{Var}\left[\hat{\theta}(Y^{(2,1)})\right].$$

This corresponds to a variance of the average value of a transaction conditional on the number of transactions observed. The variance on the right-hand side above can be calculated using the forms in (12).

References

Ahmed, Naeem, Matthew Brzozowski, and Thomas Crossley. 2006. "Measure Errors in Recall Food Consumption Data." Institute for Fiscal Studies Working Paper W06/21.

Angrisani, Marco, Kevin Foster, and Marcin Hitczenko. 2017a. "The 2012 Diary of Consumer Payment Choice: Technical Appendix." Federal Reserve Bank of Boston Research Data Reports No. 17-5.

Angrisani, Marco, Kevin Foster, and Marcin Hitczenko. 2017b. "The 2015 Survey of Consumer Payment Choice: Technical Appendix." Federal Reserve Bank of Boston Research Data Reports No. 17-4.

Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data.* Hoboken, NJ: John Wiley & Sons.

Board of Governors of the Federal Reserve System. Various Years. "Survey of Consumer Finances." http://www.federalreserve.gov/econresdata/scf/scfindex.htm.

Bollen, Kenneth A., and Robert W. Jackman. 1990. "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases." In *Modern Methods of Data Analysis*, eds. John Fox and J. Scott Long, 257–291. Newbury Park, CA: Sage.

Bricker, Jesse, Arthur B. Kennickell, Kevin B. Moore, and John Sabelhaus. 2012. "Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin* 98(2).

Casella, George, and Roger L. Berger. 2002. *Statistical Inference.* Pacific Grove, CA: Thomson Learning.

Chambers, Raymond L., and Ruilin Ren. 2004. "Outlier Robust Imputation of Survey Data." *The Proceedings of the American Statistical Association.*

Cook, R. Dennis. 1977. "Detection of Influential Observations in Linear Regression." *Technometrics* 19(1): 15–18.

Cook, R. Dennis, and Sanford Weisberg. 1982. *Residuals and Influence in Regression.* New York, NY: Chapman & Hall.

De Leeuw, Edith D. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21(5): 233–255.

Deming, W. Edwards, and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Tables are Known." *The Annuals of Mathematical Statistics* 11: 427–444.

Dillman, Don A, Jolene D Smyth, and Leah M Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.

Duncan, Greg J., and Graham Kalton. 1987. "Issues of Design and Analysis of Surveys Across Time." *International Statistical Review* 55: 97–117.

Frees, Edward W. 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge, UK: Cambridge University Press.

Gelman, Andrew, and Hao Lu. 2003. "Sampling Variances for Surveys with Weighting, Post-stratification, and Raking." *Journal of Official Statistics* 19(2): 133–151.

Greene, Claire, Shaun O'Brien, and Scott Schuh. 2017. "U.S. Consumer Cash Use, 2012 and 2015: An Introduction to the Diary of Consumer Payment Choice." Federal Reserve Bank of Boston Research Data Reports No. 17-6.

Greene, Claire, and Scott Schuh. 2017. "The 2016 Diary of Consumer Payment Choice." Federal Reserve Bank of Boston Research Data Reports No. 17-7.

Hitczenko, Marcin. 2015. "Identifying and Evaluating Selection Bias in Consumer Payment Surveys." Federal Reserve Bank of Boston Research Data Reports No. 15-7.

Jonker, Nicole, and Anneke Kosse. 2009. "The Impact of Survey Design on Research Outcomes: A Case Study of Seven Pilots Measuring Cash Usage in the Netherlands." De Nederlandsche Bank Working Paper No. 221.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons.

Lynn, Peter. 2009. *Methodology of Longitudinal Surveys*. Hoboken, NJ: John Wiley & Sons.

Samphantharak, Krislert, Scott Schuh, and Robert M. Townsend. 2017. "Integrated Household Financial Surveys: An Assessment of U.S. Methods and an Innovation." Federal Reserve Bank of Boston Research Data Reports No. 17-7.

Schmidt, Tobias. 2011. "Fatigue in Payment Diaries — Empirical Evidence from Germany." Deutsche Bundesbank Discussion Paper Series 1: Economic Studies No 11/2011.

U.S. Bureau of Labor Statistics. 2013. "Consumer Expenditures and Income." In *BLS Handbook of Methods*.

U.S. Bureau of Labor Statistics. Various Years. "Consumer Expenditure Survey." http://www.bls.gov/cex/home.htm.

U.S. Census Bureau and U.S. Bureau of Labor Statistics. 2010. "Current Population Survey." https://www.census.gov/programs-surveys/cps/html.

Valliant, Richard, Jill A Dever, and Famke Kreuter. 2013. *Practical Tools for Designing and Weighting Survey Samples.* New York: Springer.