Working Papers

FEDERAL RESERVE BANK OF BOSTON™

No. 13-16

Optimal Recall Period Length in Consumer Payment Surveys

Marcin Hitczenko

Abstract:

Surveys in many academic fields ask respondents to recall the number of events that occurred over a specific period of time with the goal of learning about the mean frequency of these events among the population. Research has shown that the choice of the recall period, particularly the length, affects the results by influencing the cognitive recall process. We combine experimental recall data with use data to learn about this relationship in the context of consumer payments, specifically for the mean frequency of use of the four most popular payment instruments (cash, credit card, debit card, check).Overall, our analysis suggests that day-based recall is inefficient, with mean-squared errors of population estimates minimized for longer recall periods, although the optimal recall period differs among payment instruments. In addition, for cash, we develop a model relating recalled values to individual frequency of use in order to study the relationship between demographic variables and accuracy at different recall lengths. We find little link between demographic characteristics and accuracy of different recall periods for an individual.

Keywords: consumer payments, stochastic recall models, survey design, mean-squared error, bias

JEL Classifications: C83

Marcin Hitczenko is a survey methodologist and a member of the Consumer Payments Research Center in the research department of the Federal Reserve Bank of Boston. His e-mail address is marcin.hitczenko@bos.frb.org.

I would like to thank Scott Schuh and the CPRC for their guidance, feedback, and support of this work.

This version: December 2013

CONSUMER PAYMENTS Research Center

1 Introduction

Many academic fields, especially in the social sciences, rely on surveys to collect data to be used for inference about a population of interest. While the surveys themselves naturally vary in topic and mode, a particular class of question common to many is one that targets the frequency of a particular behavior among individuals. Indeed, questions of this form arise in surveys for datasets relevant to economic research, for example datasets for the following fields (with the acronyms for the dataset(s) in each field indicated in parentheses immediately after): economic consumption (BHPS; CES; SCA; PSID), health (NHIS), media (BMCS), and crime (BRFSS), among others. (The full names of these datasets appear in the reference list.)

With respect to behavioral frequency, there are two general approaches to data collection. The first approach asks for an occurrence rate over a generic period of time: the number of purchases made in a typical week. The second approach asks for the number of times an event occurred in a specific period of time: the number of purchases made in the past week. The properties and relative advantages of each technique have been studied and debated (Babor, Brown, and Boca 1990; Chang and Krosnick 2003). In this work, we consider only the latter, which appeals more directly to an individual's ability to recall.

What a respondent reports in a survey is not necessarily the truth, but rather the truth filtered through that individual's cognitive process. It is a given that there will be errors associated with recall. Survey responses have been shown to be influenced by virtually every aspect of the questionnaire design (see Sudman, Bradburn, and Schwarz (1996); Tourangeau, Rasinski, and Bradburn (1991) for an overview), but we focus here on the choice of the recall period, the period of time over which respondents report behavior. If different recall periods tend to have different effects on the reported values, choosing one recall period over another could lead to less reliable results. Our research focuses on the relationship between the length of the recall period and the accuracy of estimates based on the resulting survey data, in the context of consumer payment surveys.

Initial research on accuracy of recall predominantly worked under the assumption that the events to be recalled were sufficiently memorable or infrequent that a respondent would try to catalogue and count each one (referred to as a "recall-and-count" or "enumeration" procedure). In such cases, the two main sources of recall error were hypothesized to be omission, the forgetting of events, and telescoping (Neter and Waksberg 1964), wrongly counting events that occurred outside the period in question. The former naturally leads to under-reporting while the latter causes over-reporting. There is a fair amount of literature quantifying the relationship of telescoping to elapsed time (Huttenlocher, Hedges, and Bradburn 1990; Rubin and Baddeley 1989). In particular, Sudman and Bradburn (1973) used mathematical models to determine the optimal recall period length for surveying days taken off for sick leave by minimizing the net relative error. More recently, Clarke, Fiebig, and Gerdtham (2008) studied the optimal recall period when asking about the length of hospital visits within the past three months.

Over time, however, developments in cognitive research made it apparent that the mental process involved in recall varied significantly depending on the topic being surveyed. Specifically, frequent and not particularly memorable events tend to blend together and do not lend themselves to episodic recall (Bradburn, Rips, and Shevell 1987; Strube 1987). As the relative frequency or the length of the recall period increases, respondents rely less on enumeration and more on a variety of other techniques, most notably rate-based recall (Blair and Burton 1986; Eisenhower, Mathiowetz, and Morganstein 1991; Menon 1994). Increased understanding of recall has fueled the development of survey techniques, such as bounding (see Means et al. (1989); Sudman, Bradburn, and Schwarz (1996) for discussion), to improve respondent recall.

It is our view that many researchers will continue to rely on surveys for data collection, as it remains, generally speaking, a simpler and much less expensive option than other forms of learning about behavior. In consumer payments, for example, official records either do not exist, as for cash, or are tedious to parse and raise privacy concerns, such as bank or credit card statements. Recall-based measurement error is a reality of behavioral frequency analysis. In addition, many of the recall-improving survey strategies have mixed levels of effectiveness and, perhaps more importantly, require a prohibitively large expenditure of resources, in time and money, to implement. The issue then is how a researcher should choose the recall period length in order to optimize the quality of data-based estimates.

We hope to shed some light on this subject within the narrow framework close to the research interests of the Consumer Payment Research Center (CPRC) at the Federal Reserve Bank of Boston, namely the frequency of purchases by consumers. The analysis in this work involves combining data from an experiment on recall of payment use with data gathered by having consumers track payment activity in a diary for each of the four most frequently used payment instruments: cash, credit cards, debit cards, and checks. Studies comparing recall-based data with diary-based data in the context of household expenditure on goods already exist (Battistin and Padula 2010; McWhinney and Champion 1974), and at least a few discuss the effect of the recall period on survey-based statistics (Ahmed, Brzozowski, and Crossley 2010; Deaton and Grosh 2000; Hurd and Rohwedder 2009; NSSO Expert Group on Sampling Errors 2003). The consensus among this research seems to be that the optimal recall period varies with the behavior being surveyed. Therefore, we believe that it is informative to extend the analysis to the case of payment instrument use among U.S. consumers. In addition, we study the consistency of individuals' diary-reported cash activity with recalled cash activity for different recall periods with the hope of using demographic information to predict optimal recall periods for each individual and, in turn, improving the efficiency of population estimates.

The layout of this paper is as follows. In Section 2, we introduce the notation and outline the general framework of our statistical analysis. Section 3 provides details of the datasets used in this analysis, and Section 4 analyzes the diary-based data for any systematic behavioral patterns one would need to incorporate into any recall-based estimate. Analysis of the effect of the recall period length on survey-based estimates for mean population use is described in Section 5. Section 6 uses a Markov Chain Monte Carlo (MCMC) algorithm to fit a model that relates the recall period length to the consistency of reported values with individual frequencies of use. The section includes a discussion of the model, how it is fit, and the implications of the results on population parameter estimation. Finally, Section 7 provides a discussion of our findings.

2 Framework of Analysis

2.1 Notation

We begin by introducing the notation for variables relating to the use and recall of use of a payment instrument. In this work, analysis for the four payment instruments is done disjointly, meaning that the same procedures and models are applied to each mode of payment separately. Therefore, there is no notational need to distinguish between payment instruments, and all variables and equations below relate to a generic payment instrument.

In the context of payments, it seems reasonable to assume that the shortest unit of time over which a respondent will be asked to recall his behavior is a day. Indeed, we are unaware of surveys that ask for recall over a set of hours. On the other hand, days (used in the MPS and the DCPC), weeks (used in the BHPS), months (used in the CES), and years (used in the PSID and the SCA) have all been used as recall periods. As a result, we measure the recall period lengths in terms of the number of days, and each payment is defined by the day of its occurrence. By doing this, we implicitly assume an exchangeability in the use of and ability to recall payments within a single day. Days are indexed chronologically, starting with day $t = 1$ chosen to be some arbitrary date far enough in the past to precede any potential day of recall.

Individuals, indexed by a subscript i, are assumed to come from an infinite or a large, finite population. The number of payments made by individual i on day t is represented by X_{it}. Many, if not most, consumer payment surveys ask about the aggregate number of payments made over a series of days, so we define

$$A_i(t_1, t_2) = \sum_{t=t_1}^{t_2} X_{it}$$

to be the total number of payments made in the $(t_2 - t_1 + 1)$-day period between day t_1 and day t_2, $t_1 \leq t_2$.

The aggregates reported in a survey are themselves random variables, whose distribution and observed outcomes depend not only on the recall period but also on the particular implementation of the survey. It is possible for an individual to report a different number of payments for the same period of time on different occasions. To simplify matters, we assume that an individual will be asked to provide recalled aggregates at most once for any recall period on a given day. This allows us to index the survey by the day on which it occurred, and we define $R_{is}(t_1, t_2)$ to be the reported number of payments made between day t_1 and day t_2, as reported on survey day $s > t_2$.

It is fairly rare, especially in consumer payment research, for a survey to ask for behavior over a period that does not directly precede the survey. In other words, usually $s = t_2 + 1$. For this reason, we restrict our analysis to cases in which there is no gap between the survey and the end of the recall period. Working under this construct, it is simpler to re-parameterize the variables in terms of the length of the recall period, ℓ:

$$\begin{aligned} A_{is\ell} &= A_i(s - \ell, s - 1) \\ R_{is\ell} &= R_{is}(s - \ell, s - 1). \end{aligned} \tag{1}$$

2.2 Evaluating Recall Effect

The collection of survey data is motivated primarily by a desire to estimate a set of parameters relating to a population. As a result, a survey methodologist is fundamentally interested not in the accuracy of the recall data, but rather in the accuracy of an estimator based on the recall data. The former refers to the properties of the reported microdata alone and is characterized by the relationship between $A_{is\ell}$ and $R_{is\ell}$. While the accuracy of individual-level recall is obviously related to the accuracy of an estimator, the two concepts, in general, are not the same.

First, accuracy of recall data in a certain context is useful only if these data help to provide insight into a parameter of interest. For example, error-free recall for a period of a year might not shed much light on behavior over a particular month if there is seasonal variation. At the same time, it is possible for data with poor recall properties at the individual level to have desirable sample properties. Consider a situation in which each individual reports a value that is a weighted average of his or her true behavior and the overall population mean, as might occur if individuals adjust their responses to better coincide with a norm. In such a scenario, despite poor correspondence between an individual's reported values and the truth, the sample mean of reported values will be an unbiased estimate for the population mean under some mild sampling assumptions.

As a result, we believe that the parameter of interest should motivate the choice of the recall period, and we discuss the effect of recall period lengths in the context of a particular estimator. With this in mind, it is useful to formalize how the recall period fits with other aspects of the survey design and the modeling assumptions made by the researcher to define an estimator. Consider a parameter of interest, ω, and a hypothetical researcher who is interested in generating a survey-based estimate, $\hat{\omega}$. The form of the estimate depends on the model assumed by the researcher for the data-generating process as well as on the survey design used to collect the data. The two should be viewed as a pair, since the assumed model heavily influences the choice of survey design.

We let $\mathcal{S}$ encompass all aspects of the survey design *except* the assignment of the recall period length, ℓ. Thus, $\mathcal{S}$ defines the sample size, a methodology for selecting individuals from the population, a system of assigning survey days to each respondent, and even the design of the survey questionnaire. Combined with a recall period length, ℓ, the data collection process results in a vector of data: $\mathbf{A}(\mathcal{S}, \ell) = \{A_{is_i\ell}\}$ represents the collection of error-free values while $\mathbf{R}(\mathcal{S}, \ell) = \{R_{is_i\ell}\}$ represents the corresponding reported values.

The modeling assumptions made by the researcher are very important, as a poor understanding of the data-generating process could lead to biased estimates even with error-free recall. The form of the estimator will almost surely depend on the length of the recall period, so we let $\mathcal{M}_\ell(\cdot)$ represent a mapping from the collected data for a recall length ℓ to the estimate $\hat{\omega}$. We define an unbiased estimator to be one based on a pair $(\mathcal{M}_\ell, \mathcal{S})$ such that

$$\mathrm{E}\left[\mathcal{M}_\ell(\mathbf{A}(\mathcal{S}, \ell))\right] \quad = \quad \omega. \tag{2}$$

That is, the estimator based on error-free data is unbiased. In this work, we are primarily interested in evaluating the effect of potentially erroneous recall values on otherwise unbiased estimators. Specifically, we consider two general forms of estimators:

$$\hat{\omega} = \mathcal{M}_\ell(\mathbf{R}(\mathcal{S}, \ell)) \quad \text{and} \quad \hat{\omega} = \mathcal{M}_\ell\left(\frac{\ell}{\ell'}\mathbf{R}(\mathcal{S}, \ell')\right). \tag{3}$$

The first estimator in (3) measures the effect of relying on recall data rather than on true values. By definition, if there is no recall error, the estimate will be unbiased. The second estimator in (3) allows one to see how scaling results from different recall period lengths affects estimates. In practice, only certain combinations of (ℓ, ℓ') are reasonable, most notably those where there is an expectation that $\mathrm{E}\left[A_{is\ell}\right] = \frac{\ell}{\ell'}\mathrm{E}\left[A_{is\ell'}\right]$.

For a given estimate $\hat{\omega}$, we evaluate its accuracy through the mean-squared error

$$\begin{aligned} \mathrm{MSE}(\hat{\omega}) &= \mathrm{E}\left[\hat{\omega} - \omega\right]^2 \\ &= \mathrm{Var}(\hat{\omega}) + \mathrm{Bias}^2(\hat{\omega}), \end{aligned}$$

which decomposes into the variance of the estimator and the square of the bias. The framework established in this section is very general, and the ensuing analysis in this paper will naturally be restricted to the particulars of the data available to us. In Section 3, we introduce the data, and Section 4 defines some potential frameworks, $(\mathcal{M}_\ell, \mathcal{S})$, for our data.

3 Data

For the purpose of evaluating the quality of recall it would be ideal to pair recalled values with corresponding daily records of true behavior for the same periods of time, at least whenever possible (credit, debit, and check). While there is hope of constructing such a dataset, no such database exists currently. Instead, we rely on data combined from two different, but related, datasets, derived from the RAND Corporation's American Life Panel (ALP). The ALP is composed of individuals from other respondent pools who expressed interest in taking surveys on a variety of topics, and at the time of sampling (October 2012) consisted of around $5{,}000$ recruits. Details of the ALP population and the many surveys in which its members participate can be found at `www.RAND.org/ALP`. The following subsections introduce the two datasets used in this work in the context of the notation defined in Section 2.1.

3.1 2011 − 2012 Payment Recall Data

The Payment Recall Data (PRD), collected in an experiment co-designed by the CPRC and RAND, asks individuals to recall payment behavior over a variety of recall periods for each of the four major payment instruments. The experiment in its entirety incorporates several experimental factors into the design, so the resulting dataset allows for a wide range of analyses. A more detailed description of the PRD can be found in Angrisani, Kapteyn, and Schuh (2012). Below, we provide a brief overview of the subset of the dataset most relevant to this work.

The PRD recruited $3{,}285$ members of the ALP to complete an online survey five times, once every three months from July 2011 to September 2012 (each three-month period is called a phase). In each phase of the survey, the respondent was asked to assess the number as well as dollar value of payments made by each of the four major payment instruments for four different recall periods. The result is 32 observations per individual: 16 corresponding to the number of payments made and 16 corresponding to the dollar value of those payments. The recall periods correspond to a day ($\ell = 1$) , week ($\ell = 7$), month ($\ell = 30$), and year ($\ell = 365$). An additional component of the design was that the "recall framework" changed from phase to phase for each individual. In the first phase, respondents were randomly assigned one of two frameworks: the first of which framed the questions in terms of number of payments within a "typical" time period and the second of which asked the respondent for behavior in a specified period of time. For the most part, the framework of each survey then alternated in the following phases of the study; however, an error in coding resulted in many individuals taking the same versions of the framework in the fourth and fifth phase.

In this paper, we limit ourselves to the data relating to the number of payments reported within the specific framework. The temporal layout of the survey days and the associated recall periods within each phase of the specific framework are shown for one individual in Figure 1. In the specific framework, the survey asks each respondent to provide the total number of payments for the year, month, and week directly preceding the day of the survey, as well as for a day randomly chosen within the past week of the survey. Recall is done for each payment instrument sequentially, with the order of the instruments chosen at random. In addition, for each payment instrument the order of the daily, weekly, and monthly periods is randomized, with the yearly period always coming last.

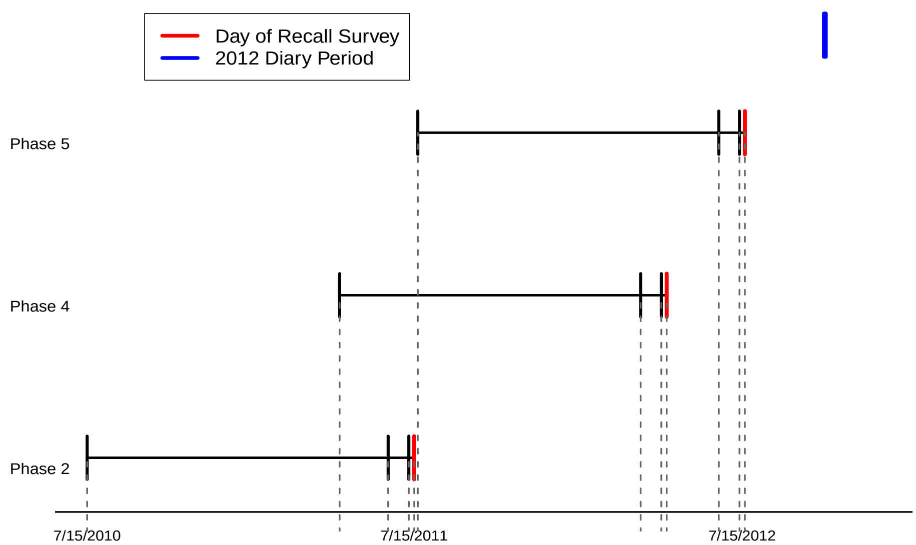

Figure 1: An example of the temporal distribution of the PRD survey and the DCPC for an individual who takes the specific framework in phases $p = 2, 4, 5$.

In the first phase, each individual in the sample is randomly assigned to receive the survey via email on the 15^{th} of July, August, or September of 2012. Surveys for all subsequent phases are sent to each respondent three months later. Although all surveys are received on the 15^{th} of each month, respondents can take the survey whenever they want. Figure 2 shows the number of individuals who completed each survey by the number of weeks after its receipt. Clearly, most individuals respond within the first week, although a substantial number did so later in the month. It should be noted that the specific recall periods are assigned at the commencement of the survey, so the reported values are relative to the day on which the survey was taken rather than to the day the survey link was emailed.

We organize the data by the phase of participation, letting $\mathcal{I}^p$ represent the set of individuals who took the specific framework survey in phase p for $p = 1, 2, 3, 4, 5$. The data collected from an individual from each phase correspond to four recalled values for each payment instrument. Three of these values, the responses for a week, month, and a year, fall into the framework of interest defined in (1), namely that the period of recall directly precedes the day of the survey. For individual i in phase p, we define these

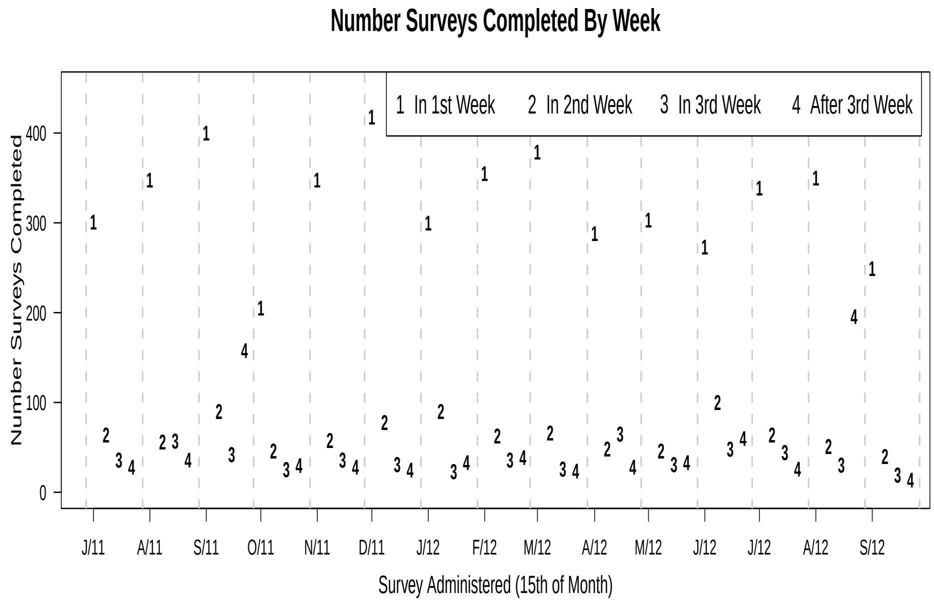

Figure 2: The number of individuals who completed the survey in each phase by week after survey release.

values as $R_i^p = \left\{ R_{is_i^p \ell} \mid \ell = 7, 30, 365 \right\}$, where s_i^p represents the survey day. As this notation is somewhat tedious, we simplify matters by letting $R_{i\ell}^p = R_{is_i^p \ell}$ for all ℓ. We let $\mathbf{R}^p = \{R_i^p \mid i \in \mathcal{I}^p\}$ be the collection of survey data for all individuals participating in the p^{th} phase.

The reported daily estimate, however, does not necessarily fall into the desired framework, because there is a randomly assigned survey lag of up to seven days. The day of recall in phase p, which we define as t_i^p, is drawn from a $\mathrm{Uniform}(s_i^p - 7, s_i^p - 6, \ldots, s_i^p - 1)$ distribution, independently of all other variables. The daily recall reported by individual i in phase p is thus $R_{is_i^p}(t_i^p, t_i^p)$, but we refer to it as R_{i1}^p in our shorthand notation.

For the purposes of this analysis we consider almost all reported sequences, even if they are not internally consistent (for example, if the number of payments reported for a week is greater than that for the month). The only exception made is for those individuals whose reported number of payments is greater than or equal to the reported dollar value. On one hand, the presumed reporting of dollar values as opposed to the number of payments made could be viewed as a natural source of measurement error. However, we believe the likelihood of such an error is drastically increased in the PRD data by explicitly asking for the dollar values along with the number of payments made in the survey.

3.2 2012 DCPC

While the PRD dataset provides information about consumers' recall, it does not provide any sense of consumers' true spending activity. For this information, we rely on the 2012 Diary of Consumer Payment Choice (DCPC). The 2012 DCPC, the first nonpilot rendition of the diary, invited 2,505 individuals, also from the ALP, to record various aspects of their payment behavior for three consecutive days in the fall of 2012.

Each participant was randomly assigned a three-day period between September 30^{th} and November 2^{nd} of 2012 over which to track and record all purchases and bill payments, along with other types of transactions including cash withdrawals and deposits. For each transaction, the individual provided identifying details such as the dollar value, time, purpose, and location of the transaction, and, whenever relevant, the type of payment instrument used in the transaction. Because the diary days are predominantly in October, they generally do not overlap with the recall periods in the PRD surveys, of which only a few extended into October 2012. An example of the relative timing of the DCPC to the recall periods in the PRD survey can be seen in Figure 1. Because this particular individual took the fifth phase of the PRD survey in July and participated in the diary in the middle of October, the gap between the diary period and the closest recall period is relatively large.

In order to make the DCPC data comparable to the "payments" reported in the PRD, we consider only purchases and bill payments. Of course, there is no way to assess the accuracy of the diary data itself, in the sense of how closely it mirrors true payment behavior. Respondents have access to a portable diary, but they do not have to use it. In practice, any responses entered online within five days of the day in question are accepted as valid. Past research has shown that diaries are not completely accurate, with possible omission and diary fatigue known to occur (Silberstein and Scott 1991). In addition, it has been hypothesized that the act of recording one's behavior itself may result in unusual behavior on the part of the individual, although there has been no conclusive evidence to verify this hypothesis (Kemsley and Nicholson 1960; McKenzie 1983). Despite these potential sources of error, the diary data currently provide the best insight we have into the true data patterns of individuals, and we believe it is worthwhile to use these data as a proxy for true payment behavior.

In terms of notation, we let $(t_{i1}^{\text{dcpc}}, t_{i2}^{\text{dcpc}}, t_{i3}^{\text{dcpc}})$ represent the three days of participation for individual i. Then, the DCPC data correspond to $\left(X_{it_{i1}^{\text{dcpc}}}, X_{it_{i2}^{\text{dcpc}}}, X_{it_{i3}^{\text{dcpc}}}\right)$. For simplicity of notation, we define this vector of responses as $X_i^{\text{dcpc}} = \left(X_{i1}^{\text{dcpc}}, X_{i2}^{\text{dcpc}}, X_{i3}^{\text{dcpc}}\right)$. We let $\mathbf{X}^{\text{dcpc}}$ represent the entire collection of data from the DCPC for all $N^{\text{dcpc}} = 2{,}468$ individuals who participated for all three days.

4 Defining $(\mathcal{M}_\ell, \mathcal{S})$

4.1 Temporal Trends in DCPC Data

In this section, we study the DCPC data with a goal of identifying any prevalent trends or patterns in individual payment behavior. Because the data from the DCPC span October as well as the last two days of September and the first two days of November, it is impossible to recognize any long-term patterns. Recognizing seasonal cycles with a low frequency or month-to-month variation is outside the scope of the data.

We begin with the assumption that $X_{it} \sim \text{Poisson}(\mu_{it})$, where μ_{it} represents the expected number of payments made on day t by individual i. We assume that μ_{it} is a product of an individual-specific base daily frequency and a day effect, so that $\log \mu_{it} = \mu_i + f(t)$. In addition, we assume that, conditional on the individual-specific parameters, the number of daily payments made by each of two individuals is independent of the number of daily payments made by the other: $X_{it} \mid \mu_i \perp\!\!\!\perp X_{i't} \mid \mu_{i'}$ for all t and $i \neq i'$. Similarly, we assume that the number of payments made on one day is independent of the number of payments made on a different day for each individual: $X_{it} \perp\!\!\!\perp X_{it'}$ for $t \neq t'$.

Figure 3 shows the empirical daily averages and 95 percent confidence intervals for the 35 days in 2012 for which DCPC data are collected. The confidence intervals represent ± 2 standard errors from the sample average for all observations taken on each calendar day. At least visually, these plots do not indicate a strong monthly trend or cycle. However, there is evidence of day-of-week effects, with the most notable changes in behavior involving the weekend. For cash, it seems that there is a slow, but steady increase in the average number of payments from Sunday to Saturday. Credit and debit, on the other hand, show a steady number of payments from Sunday to Thursday, with a jump in use on Friday and Saturday. Finally, checks are less likely to be written on Saturday or Sunday, but show a fairly constant frequency of use across the rest of the week.

To quantify these visual trends, we consider several models for μ_{it}. Before introducing the models, we

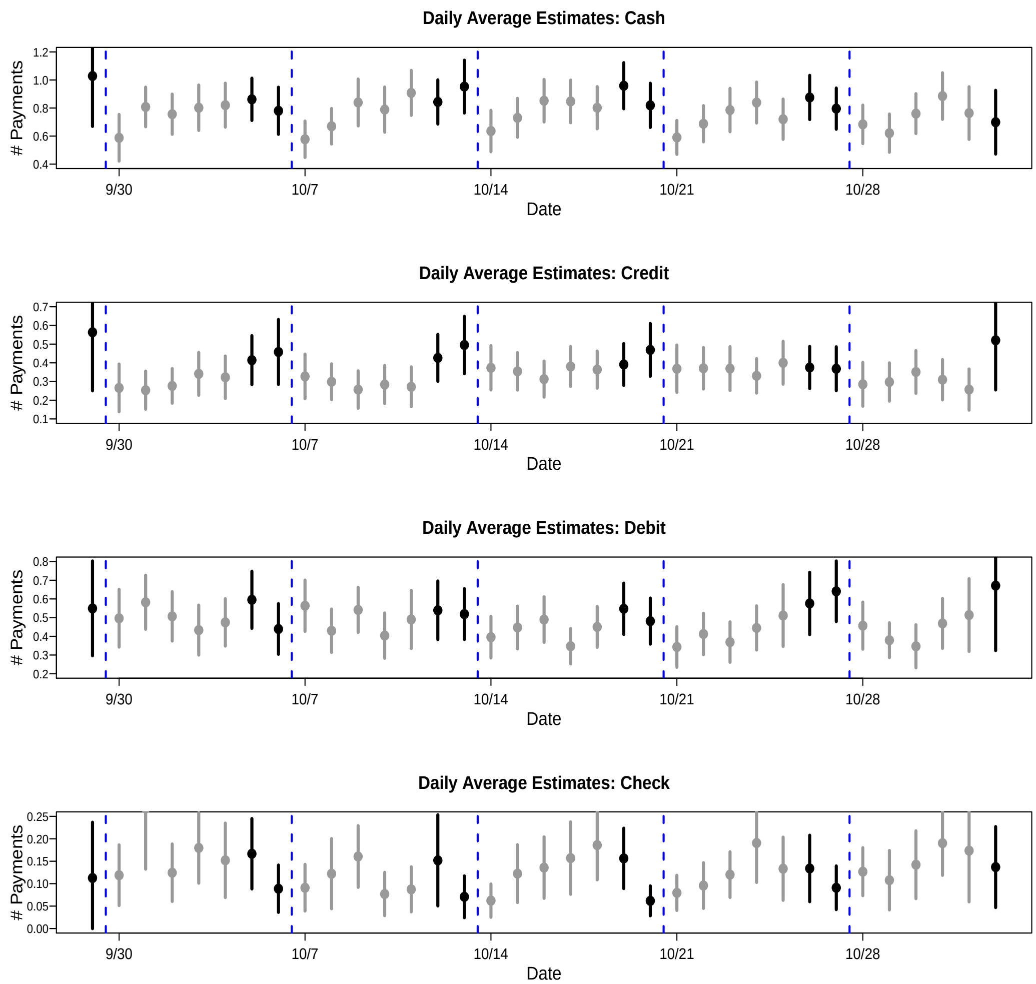

Figure 3: Means and 95 percent confidence intervals for mean daily frequencies for each survey day in the 2012 DCPC. Darker lines correspond to Fridays and Saturdays.

define two functions used in the mean structure. The first,

$$\texttt{dow}(t) = \begin{cases} 1 & t \text{ is a Sunday} \\ 2 & t \text{ is a Monday} \\ \vdots & \vdots \\ 7 & t \text{ is a Saturday,} \end{cases}$$

indexes the day of the week of a particular day, while the second,

$$\texttt{pom}(t) \quad = \quad \frac{\sum_{t'} 1\left[t' \leq \ t, \text{ and } t', t \text{ in same month}\right]}{\sum_{t'} 1\left[t', t \text{ in same month}\right],}$$

represents the relative location of a day within the month containing it. Thus, if the day in question is

October 15^{th}, then $\texttt{pom}(t) = 15/31$. Using these definitions, we consider three potential models for $f(t)$:

$$\begin{aligned}
&\textbf{Model A:} \qquad f(t) = \sum_{j=1}^{7} \beta_j 1\left[\texttt{dow}(t) = j\right] + \alpha_1 \texttt{pom}(t) + \alpha_2 \texttt{pom}^2(t) \\
&\textbf{Model B:} \qquad f(t) = \sum_{j=1}^{7} \beta_j 1\left[\texttt{dow}(t) = j\right] \\
&\textbf{Model C:} \qquad f(t) = 0.
\end{aligned} \tag{4}$$

Model A not only includes day-of-week effects but also allows for a trend with a monthly cycle. By including a quadratic term with respect to $\texttt{pom}(t)$, the model allows for monotonic trends as the month progresses as well as for patterns in which the middle of the month has relatively low or high use rates. By contrast, Model B only assumes periodicity on the scale of a week. Finally, Model C assumes stationarity across days, with identical distributions for all days. Each of these models corresponds to a mixed-effect model with the day-related effects fixed across all individuals and an individual component that varies for each individual in the population (and is thus commonly referred to as a random effect).

The assumed form of the mean daily number of payments for an individual, $\mu_{it} = \exp(\mu_i + f(t))$, ensures that $\mu_{it} > 0$. While it is obvious that the daily mean must be non-negative, it would be desirable for our model to allow $\mu_{it} = 0$ for individuals who are nonadopters of a particular instrument and thus lack the option of using it. Other than cash, which is adopted by essentially everyone in the population, the other three major instruments have adoption rates ranging from 75 percent to 90 percent (Foster, Schuh, and Zhang Forthcoming), significant fractions of the population. The assumed models compensate by allowing μ_i to be small enough that $\exp(\mu_i)$ and even yearly means, given by $365\exp(\mu_i)$, are such that the probability of nonzero payments is small.

The three models defined in (5) can be fit to the DCPC data with the function glmer in the R package lme4, specifically designed to fit generalized linear mixed-effects models. The default assumption is that $\mu_i \sim \text{Normal}(0, \sigma^2)$ with independence across individuals. The condition of independence is likely false, as we expect relative homogeneity within certain demographic strata. In effect, however, these assumptions are best viewed as a prior distribution, and the estimates from the model correspond to posterior estimates, which no longer carry an assumption of strict independence. Instead, the model implies independence conditional on the observed data, and thus that $\mu_i \mid X_i^{\text{dcpc}}$ and $\mu_{i'} \mid X_{i'}^{\text{dcpc}}$ are independent for all $i \neq i'$. This means that any inherent similarity in behavior due to nonmodeled variables provides no extra information once the diary data are observed. One would certainly consider this to be true for longer periods of behavior observation, and we are willing to to make this assumption for a diary of three days.

The three models are nested, allowing us to use an ANOVA procedure (done with the anova command in R) to compare the models' ability to explain variation in the DCPC data and thus determine the most appropriate model. For model m, we define θ_m to be σ as well as any parameters relating to $f(t)$ as given in (5). The corresponding parameters estimates based on $\mathbf{X}^{\text{dcpc}}$ are labeled $\hat{\theta}_m$ and the deviance for two models takes the form:

$$\texttt{Dev}(m_1, m_2) \quad = \quad -2\left[\log \mathrm{P}\left(\mathbf{X}^{\text{dcpc}} \mid \hat{\theta}_{m_1}\right) - \log \mathrm{P}\left(\mathbf{X}^{\text{dcpc}} \mid \hat{\theta}_{m_2}\right)\right],$$

where $\mathrm{P}(\cdot)$ represent the likelihood function. A likelihood-ratio test for the deviances produces p-values based on a Chi-squared random variable with degrees of freedom defined by the difference in parameters for the two models, allowing for model selection. Summary statistics for the models are shown for all four payment instruments in Table 1.

	Cash				**Credit**			
Model	**Log Lik.**	**Δ Deviance**	**Δ df**	**p-value**	**Log Lik.**	**Δ Deviance**	**Δ df**	**p-value**
C	-4870.7	–	–	–	-3351.6	–	–	–
B	-4843.6	54.3	6	0	-3340.2	22.7	6	0
A	-4843.4	0.3	2	0.87	-3338.0	4.5	2	0.10
	Debit				**Check**			
Model	**Log Lik.**	**Δ Deviance**	**Δ df**	**p-value**	**Log Lik.**	**Δ Deviance**	**Δ df**	**p-value**
C	-4064.6	–	–	–	-2082.1	–	–	–
B	-4054.3	20.7	6	0	-2069.3	25.6	6	0
A	-4051.4	5.7	2	0.06	-2065.7	7.3	2	0.03

Table 1: Results of model fits for Models A-C for the 2012 DCPC data.

Table 1 indicates that for all four payment instruments, Model B, which includes day-of-week effects, fits considerably better than Model C. Model A naturally improves the quality of fit by allowing for month-long effects, but, except in the case of checks, the improvement is not statistically significant. This discrepancy between checks and the other payment instruments is perhaps not too surprising. While cash, credit, and debit are routinely used for "everyday" purchases, checks tend to be used for more regular and larger-valued payments (Foster et al. 2011; Foster, Schuh, and Zhang Forthcoming). These types of payments, most notably bills, often have a monthly cycle. Nevertheless, the p-value associated with the inclusion of this monthly effect is 0.03, which is not overwhelmingly large, especially when the fact that multiple comparisons are being made is taken into account. Plots of the multiplicative day effects under Model B, which are proportional to the $\exp(\beta_j)$, for cash, credit, and debit are shown in Figure 4 and largely confirm the patterns noticed in Figure 3. The multiplicative day effects for check use across the month of October are shown under Model A and Model B in Figure 5, with a majority of the variation captured by the day-of-week effects. Again, a weekend effect, in which fewer checks are written on Saturday and Sunday, is reflected in the parameter estimates.

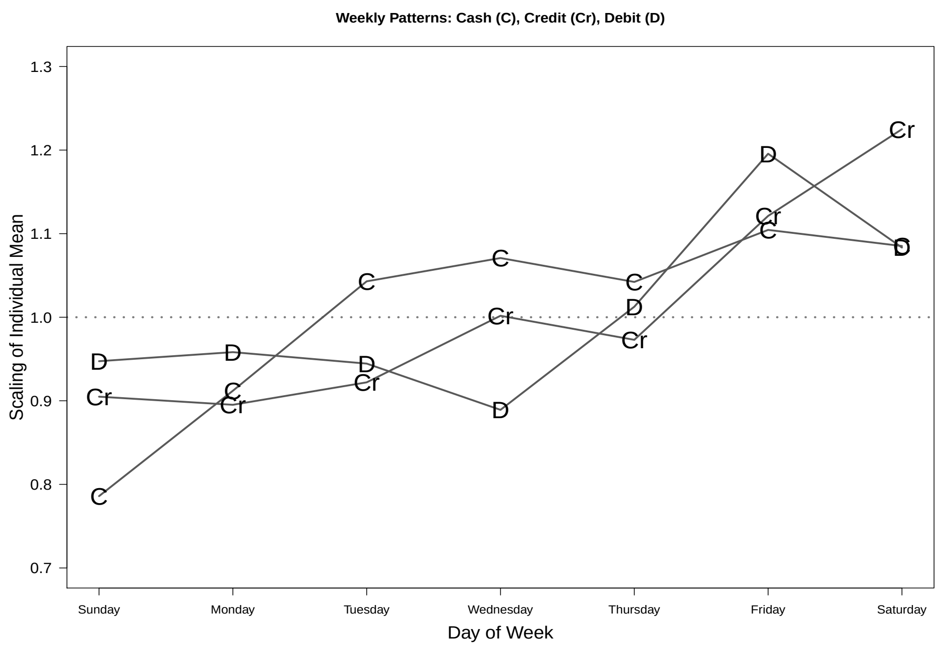

Figure 4: Estimated multiplicative effects for each day-of-week by Model B for cash, credit cards, and debit cards.

Based on this analysis, we assume a model for payment behavior based on Model B, with stationarity across weeks and multiplicative day-of-week effects, for all four instruments. Naturally, payment behavior is influenced by an assortment of complicated factors, and the assumed model is a simplification of this truth. One interesting aspect of the model is its reference period, or the degree to which its validity extends past the span of the DCPC data. The DCPC is centered around the month of October by design, since the month has seemingly few potential breaks from stationarity; it has no major holidays and is comfortably settled after the summer holidays and before the American holiday season based on Thanksgiving and Christmas. Without data from different portions of the year, it is impossible to determine the appropriate reference period for the assumed model. Extrapolations beyond October 2012 should be done with caution.

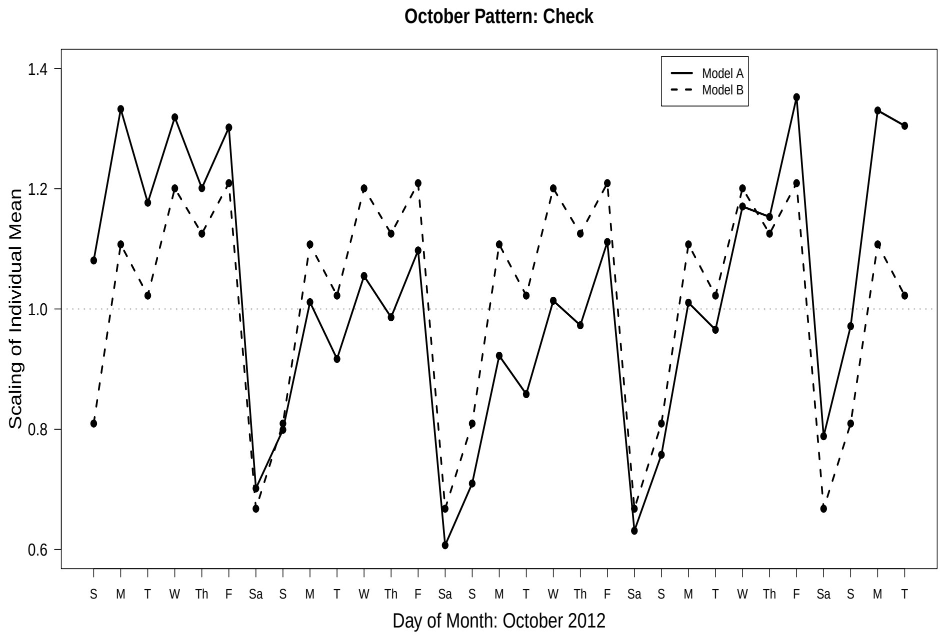

Figure 5: Estimated multiplicative effects for each survey day for Model A and B for checks.

4.2 Estimating a Weekly Frequency

Based on the analysis in Section 4.1, a natural parameter of interest is the weekly mean

$$\omega \quad = \quad \sum_{t=t_1}^{t_1+6} \mathrm{E}\left[\mu_{it}\right], \tag{5}$$

which is constant for all t_1 such that $(t_1, t_1{+}6)$ exist in the reference period. We now imagine a hypothetical researcher who accepts the assumed model and is designing a methodology to estimate ω such that the estimator $(\mathcal{M}_\ell, \mathcal{S})$ satisfies the unbiased criteria defined in (2). Perhaps the simplest, and most common strategy, is to rely on a simple random sample (SRS) within the population and to ask each respondent for recall over the past ℓ days. Under such a sample design, the only aspects of $\mathcal{S}$ to be determined by the researcher are the sample size, N, and the assignment of survey days, $\{s_i\}$, to the selected sample. Based on the above assumptions, we consider several estimators for ω. To emphasize our focus on the recall period length, the estimators will be identified as $\hat{\omega}(\ell, N)$.

The first option, under the assumptions of Model B in (5), is to decompose the weekly frequency into the sum of daily frequencies, $\omega = \sum_{j=1}^{7} \delta_j$, where $\delta_j = \mathrm{E}\left[\mu_{it} \mid \texttt{dow}(t) = j\right]$. In order to be able to produce estimates for each day of the week, the researcher must ensure that some fraction of the sample provides recall for each day of the week. We let $N_j = \sum_{i=1}^{N} 1\left[\texttt{dow}(s_i = j+1)\right]$ represent the number of individuals whose reported daily value corresponds to day-of-week j, with $\sum_{j=1}^{7} N_j = N$. A natural estimate for the average frequency on day j is then

$$\hat{\delta}_j(N) \quad = \quad \frac{1}{N_j}\sum_{i=1}^{N} R_{is_i 1} 1\left[\mathtt{dow}(s_i = j)\right].$$

For the purposes of this exercise, we assume that $N_j = N_{j'}$ for all j, j', which is the optimal design under the assumption that the variance of responses is the same regardless of the day of the week. It should be noted that if there is evidence of different degrees of variation across the days of the weeks, efficiency can be improved by assigning individuals to days of the week proportionally to the square root of the variances for each day. Regardless, the estimate of ω will be

$$\hat{\omega}(1, N) = \sum_{j=1}^{7} \hat{\delta}_j(N), \tag{6}$$

and it is easy to see that under error-free recall, $\hat{\omega}(1, N)$ is an unbiased estimator of ω.

A second approach avoids decomposing the weekly mean into the sum of daily means and simply asks for recall over seven-day periods. Within the reference period, the stationarity across seven-day periods means that the researcher can assign survey days at random. The simplest estimator that satisfies the unbiased condition under error-free recall is

$$\hat{\omega}(7, N) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} R_{is_i 7}. \tag{7}$$

The estimate in (7) simply estimates the mean weekly frequency by averaging the reported behavior for weekly periods.

The weekly periodicity provides motivation for a different form of estimator, in which estimates of weekly behavior are estimated by scaling down estimates for longer recall periods. In the case where the recall period is an integer multiple of seven days, conditional on the periods remaining in the reference period of the model, error-free recall provides unbiased estimates, since $\mathrm{E}\left[A_{is7}\right] = \frac{1}{k}\mathrm{E}\left[A_{is,(7k)}\right]$ for integer k. We consider a more general estimate of ω that takes the form

$$\hat{\omega}(\ell, N) \quad = \quad \frac{7}{\ell\, N}\sum_{i=1}^{N} R_{is_i \ell} \tag{8}$$

for $\ell = 30$ and $\ell = 365$. Both recall periods, by virtue of being noninteger multiples of seven days, fail to preserve the weekly proportions. For example, in a period of 30 days it is guaranteed that two days of the week will feature five times while all others feature only four times. Thus, even with error-free recall the estimator is biased because $\frac{7}{\ell}\mathrm{E}\left[A_{is\ell}\right] \neq \mathrm{E}\left[A_{is7}\right]$ for $\ell = 30, 365$. Yet, because the differences between days are not drastic and respondents can be assigned months or years with different daily compositions to average out any systematic bias, it seems that ignoring this necessary adjustment, especially for a period as long as a year, will have only a minor effect on final estimates.

We are fundamentally interested in the mean-squared errors of the estimators. With the assumption of independence across individuals, the mean-squared error for estimates based on daily-recall, as given in (6), is given by

$$\mathrm{MSE}(1, N) = \frac{7}{N}\sum_{j=1}^{7}\mathrm{Var}\left[R_{is_i1} \mid \texttt{dow}(s_i) = j\right] + \left(\sum_{j=1}^{7}\mathrm{E}\left[R_{is_i1} \mid \texttt{dow}(s_i = j)\right] - \omega\right)^2, \tag{9}$$

while that based on weekly estimates, as in (7) and (8), is given by

$$\mathrm{MSE}(\ell, N) = \frac{49}{\ell^2 N}\mathrm{Var}\left[R_{is_i\ell}\right] + \left(\frac{7}{\ell}\mathrm{E}\left[R_{is_i\ell}\right] - \omega\right)^2 \tag{10}$$

for $\ell = 7, 30, 365$.

Insight into how $\mathrm{MSE}(\ell, N)$ behaves as ℓ changes for different N allows researchers designing consumer payment surveys to answer important questions pertaining to survey methodology. Most importantly, one can determine the optimal choice of the recall period length, ℓ, for a given sample size. Generalizing the results, for fixed ℓ,

$$\mathrm{MSE}(N \mid \ell) \propto N^{-1}\mathrm{Var}\left[R_{is_i\ell}\right] + \mathrm{Bias}^2\left[R_{is_i\ell}\right]. \tag{11}$$

Equation (11) makes it clear that as the sample size increases ($N \to \infty$), the mean-squared error term will be increasingly dominated by the squared bias, which does not depend on the size of the sample. For this reason, a consistent bias in recalled values inherently related to the length of the recall period could result in relatively inefficient estimates of ω no matter the size of the sample. In the following section, we use the PRD and DCPC data to estimate these mean-squared errors and biases.

5 Estimating $\mathrm{MSE}(\ell, N)$

Estimates of the mean-squared errors in (9) and (10) require the first two moments of $R_{is\ell}$ and an estimate of ω. We use the PRD data for the former and the DCPC for the latter. While the DCPC data correspond to the month of October, the PRD data are generally collected from earlier periods of the year. Therefore, a potential concern is that the recall in the PRD data corresponds to a different reference period with different behavioral schemes than those observed in the DCPC.

Sample averages, scaled to a weekly frequency, for individuals who responded within the first week and the second week of receiving the survey (corresponding to groups "1" and "2" in Figure 2) are shown in Figure 6 for all 15 months of data collection. Averages based on daily recall also show standard errors for the weekly frequencies. These plots reveal some variation in the sample means, although the standard errors are relatively large. Further analysis, however, shows little evidence of seasonal trends, with a great deal of the variation attributable to the survey phase. Similarly, observations from July to September in 2011 and 2012 suggest there is no consistent trend over these months. Nevertheless, in order to minimize seasonal effects, we limit our analysis to only the 944 individuals who completed the survey after August 15^{th} of 2012, a subset of $\mathcal{I}^5$.

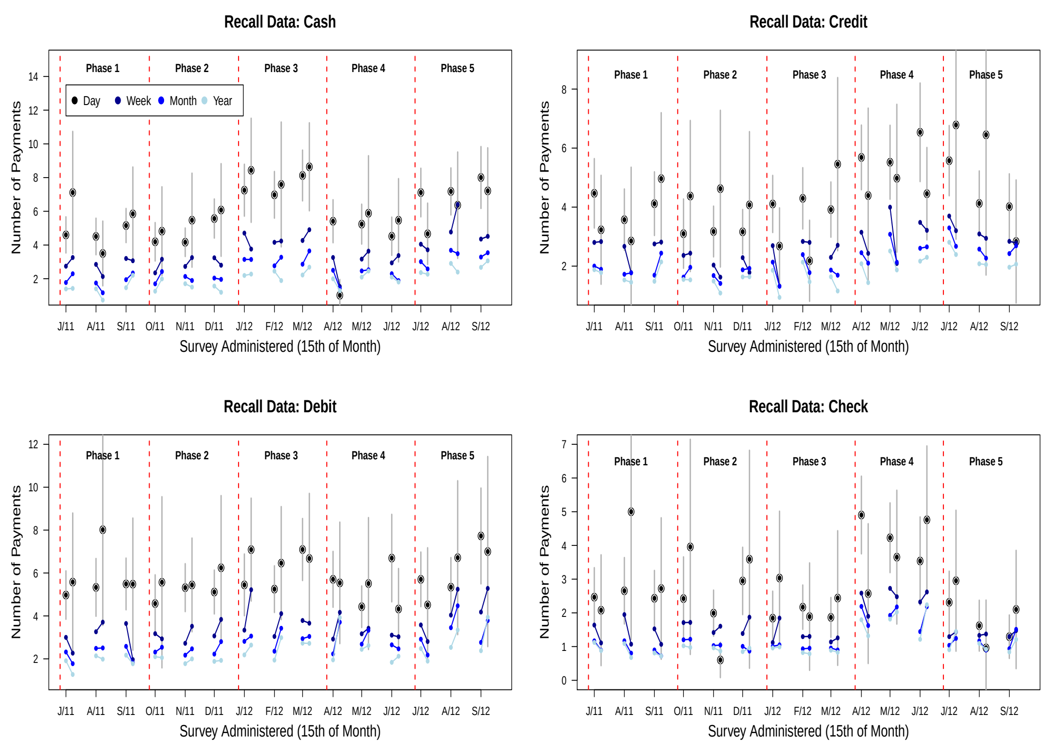

Figure 6: Average reported values (scaled to a week) for each of the four recall periods by respondents who answered within a week of survey release and those that answered in the second week. Estimates based on daily recall include 95 percent confidence intervals.

For recall periods of $\ell = 7, 30, 365$, the PRD data correspond exactly to the data of interest to our hypothetical researcher, so estimates of the moments, $\hat{\mathrm{E}}\left[R_{is_i\ell}\right]$ and $\hat{\mathrm{Var}}\left[R_{is_i\ell}\right]$ correspond to sample averages and variances, respectively. The corresponding estimates of $\mathrm{E}\left[R_{is_i1} \mid \texttt{dow}(s_i) = j\right]$ and $\mathrm{Var}\left[R_{is_i1} \mid \texttt{dow}(s_i) = j\right]$ are more difficult to estimate because the daily recall in the PRD data does not necessarily correspond to the desired form, as a result of the survey lags. One option is to simply use the cases in which $t_i^p = s_i^p + 1$, which would involve ignoring the roughly 86 percent of the data in which the survey lag is greater than a day. To measure the effect of the survey lag on the reported values, we fit a model of the form

$$R_{i1}^p \quad \sim \quad \text{Poisson}\left(\sum_{j=1}^{7} 1\left[\texttt{dow}(t_i^p) = j\right] + \sum_{j=1}^{7} 1\left[(s_i^p - t_i^p) = j\right]\right)$$

We find weak evidence of a survey lag effect, and the removal of a few outliers suggests that the survey lag does not have much influence on the values. As a result, we simply ignore the survey lag and use all available respondents in $\mathcal{I}^5$ to estimate the moments of the day-based recall estimators. Indeed, estimates based only on daily recall within the desired framework established in (1) are not much different. It is important to stress that this analysis does not mean that survey lags never have any effect on recalled values. We expect that, in general, they do, and we find only little evidence of it in the context of daily recall with a lag of at most seven days.

A second concern with the PRD data is that recalled values are collected for all four periods from each individual sequentially. As described in Section 3, the order of the recall periods is randomized for the three shorter periods, with a yearly estimate always coming last. Thus, the recalled values may not be independent and may be influenced by order effects, especially since a reported value for a week should limit the number reported for the day and set a lower bound for the value reported for the month. In fact, there seems little consistent evidence that this ordering makes much difference, and averaging over different orderings should eliminate many of these effects.

In order to estimate ω, we turn to the DCPC data, again estimating the weekly frequency as the sum of seven daily frequencies. Therefore, letting N_j^{dcpc} be the number of individuals who had an assigned diary day corresponding to the j^{th} day of the week, the estimate of δ_j based on the diary results will be

$$\hat{\delta}_j^{\text{dcpc}} = \frac{1}{N_j^{\text{dcpc}}} \sum_{i=1}^{N^{\text{dcpc}}} \sum_{t=1}^{3} x_{it}^{\text{dcpc}} 1\left[\texttt{dow}(t) = j\right],$$

and the estimate of the weekly frequency will be $\hat{\omega}^{\text{dcpc}} = \sum_{j=1}^{7} \hat{\delta}_j^{\text{dcpc}}$.

Combining all estimates and plugging back in to (9) and (10), we produce estimates: $\hat{\text{MSE}}(\ell, N)$ and $\hat{\text{Bias}}(\ell)$ for $\ell = 1, 7, 30, 365$. Besides the point estimates generated, we make use of a bootstrapping procedure to estimate a sampling distribution for $\hat{\text{Bias}}(\ell)$ and $\hat{\text{MSE}}(\ell, N)$. By bootstrapping, we address the uncertainty in our point estimates, while taking into consideration that the fair degree of overlap in the DCPC and PRD samples leads to a nonstandard dependence structure between $\hat{\omega}^{\text{dcpc}}$ and $\hat{\omega}(\ell, N)$. In fact, 656 of $1{,}426$ individuals participating in phase five of the PRD also took the 2012 DCPC. In each iteration of the bootstrap, we first sample with replacement from the 656 individuals who appear in both the fifth phase of the PRD and the 2012 DCPC. We complete the bootstrap by sampling with replacement from the individuals who took only the PRD and those who took only the DCPC. For each bootstrap-generated sample, we proceed as above and estimate the bias and the variance, combining the two for a mean-squared error.

We generate 200 bootstrapped draws and for each draw, $d = 1, \ldots, 200$, let $\hat{\text{Bias}}_d(\ell)$ be the estimate for recall period of length ℓ. The means and 95 percent confidence intervals at the four observed recall period lengths are plotted in Figure 7. In addition, we take advantage of the fact that each of our generated sequences holds a monotonic property in which $\hat{\text{Bias}}(\ell) \leq \hat{\text{Bias}}(\ell')$ for all $\ell > \ell'$, in order to interpolate estimates of the bias for other recall periods. We perform the interpolation through monotonic splines (Fritsch and Carlson 1980) using the splinefun function in R. The 200 generated curves are shown in Figure 7. Any conclusions of recall length optimality based on these curves should be made with caution, as there are potential drawbacks to asking for recall over an unfamiliar or unnatural number of days. The bias curves are primarily a theoretical exercise. All four curves reveal an over-estimation of the number of weekly payments if estimation is based on recall for one day. While in the case of cash, the three longer recall periods lead to under-estimation, credit, debit, and checks appear to minimize the absolute bias somewhere between a week and a year. For checks, it seems that yearly estimates correspond most closely to the truth, as defined in the diary.

As seen in (11), the square of the bias will serve as the limit point of $\text{MSE}(\ell, N)$ as $N \to \infty$. To get a sense of the rate of this convergence, we look at the mean-squared error for five different sample sizes: $N = 100, 200, 500, 1000, 2000$. Again, bootstrapping of individuals is done to generate 200 samples, with each sample used to determine point estimates $\hat{\text{MSE}}(\ell, N)$ for $\ell = 1, 7, 30, 365$. Figure 8 shows the means and 95 percent confidence intervals for all 16 combinations of ℓ and N. As expected, for a fixed value of ℓ, as N increases, the sample means converge in probability and there is less variation in the estimates. The overall patterns and values are similar for $N = 100$ and $N = 2000$, suggesting a very small sample size would need to be chosen for the variance component to overwhelm the bias term. For each drawn

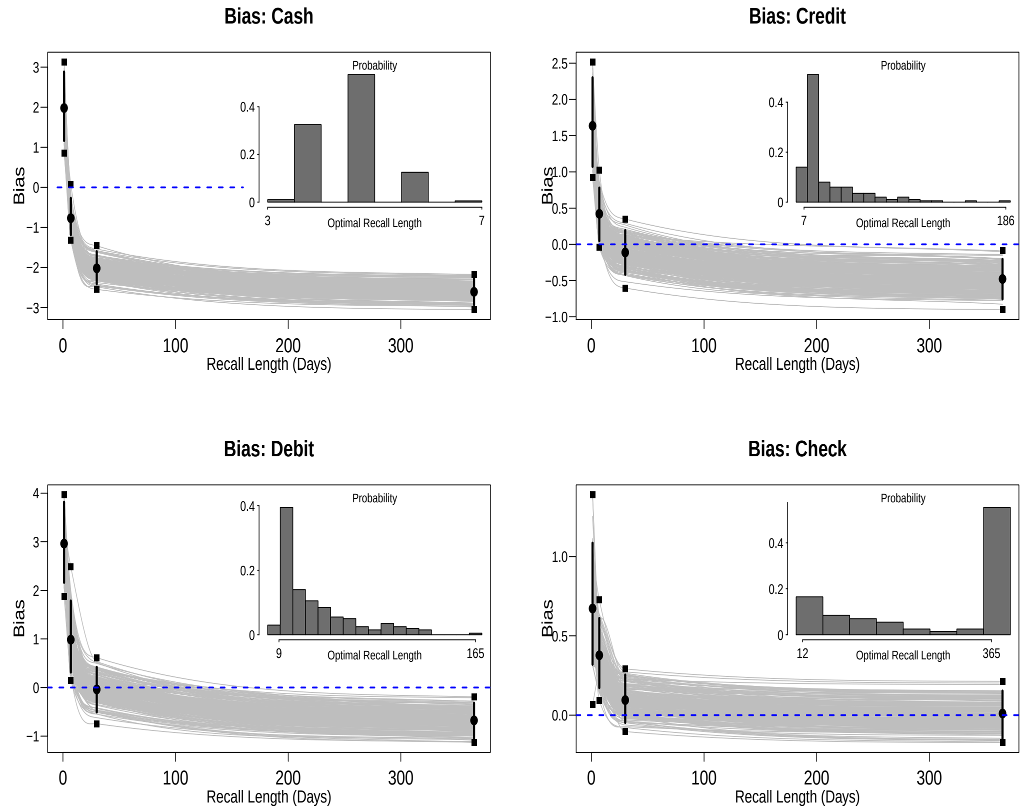

Figure 7: Bootstrapped means and 95 percent confidence intervals for $\mathrm{Bias}(\ell)$ along with fitted bias curves for each bootstrap sample (gray). Based on each curve, the recall length that minimizes the absolute bias can be determined.

sample in the bootstrap, we can determine the recall period at which the minimum mean-squared error was achieved. The distributions of these statistics are shown in Figure 8 as well; and they reflect the patterns observed for the mean-squared error estimates. For all five sample sizes, the conclusions are essentially the same. Namely, cash estimates are most likely to be best when based on weekly recall, while credit and debit estimates are generally best for monthly recall. Estimates of mean check use are predominantly best with yearly recall. In general, the optimality of different recall periods seems to mirror the frequency of use of each payment instrument, with longer recall periods better serving instruments with less frequent use.

6 Analysis: Individual-Level Inference for Cash

To gain further insight into the generation of efficient sample estimates, we study the accuracy of recall-based estimates at the individual level, comparing scaled empirical rates $\frac{7R_{is_i\ell}}{\ell}$ to weekly mean frequencies

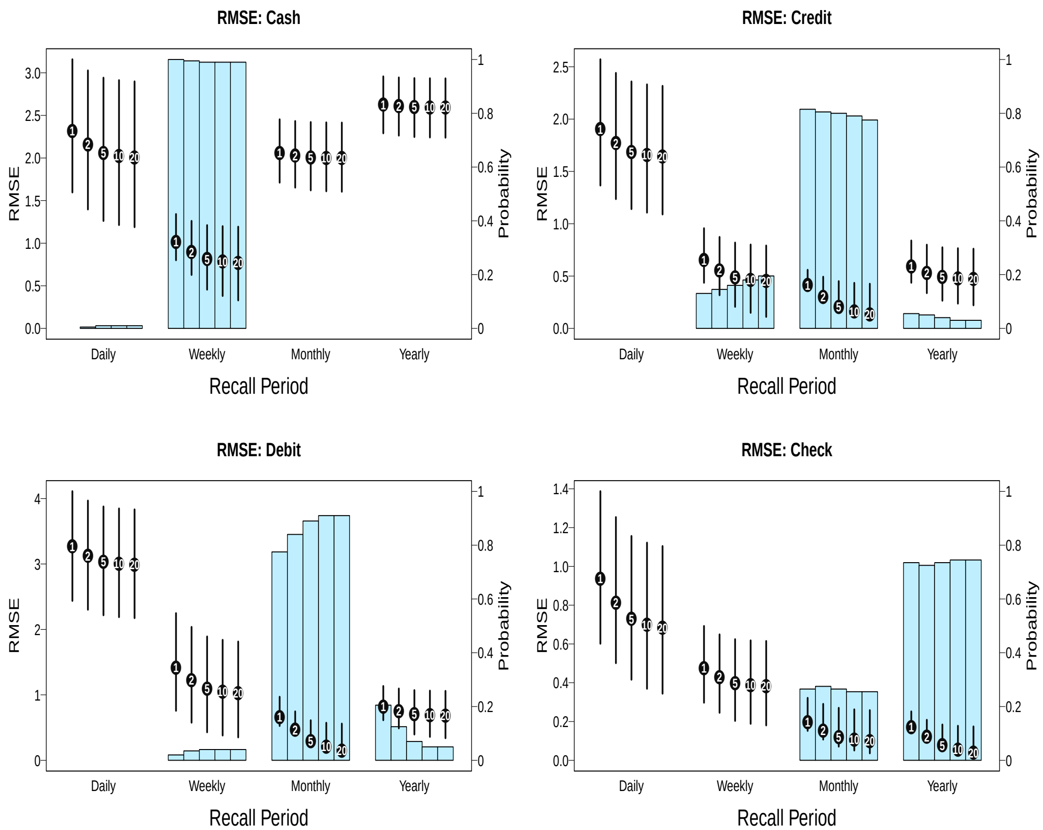

Figure 8: Means and 95 percent confidence intervals for root mean-squared errors along with estimated probabilities that each recall period provides the smallest mean-squared error. Distributions are based on bootstraps for each recall period for sample sizes of $N = 100, 200, 500, 1000, 2000$ (coded by "1","2","5","10", and "20").

ω_i. Accuracy at the individual level is useful if population parameters other than the mean, such as a percentile or the number of individuals who average more than five purchases a week, are of interest. However, our interest in this level of analysis stems mainly from a motivation to improve mean-squared errors of recall-based estimates of ω via a survey design in which respondents report behavior according to a recall period chosen to be optimal for each particular individual.

Consider that for a sample of N individuals, we do not necessarily need to enforce the same recall period for each individual. Instead, we can consider $\{\ell_i\}_{i=1}^N$ where ℓ_i is the recall period assigned to individual i. The absolute deviation of our estimator can be written as

$$\begin{aligned} |\hat{\omega}(\{\ell_i\}, N) - \omega| &= \left| \frac{1}{N} \sum_{i=1}^{N} R_{is_i\ell_i} - \omega \right| \\ &= \left| \frac{1}{N} \sum_{i=1}^{N} (R_{is_i\ell_i} - \omega_i) + \frac{1}{N} \sum_{i=1}^{N} (\omega_i - \omega) \right|, \end{aligned} \tag{12}$$

where ω_i represents the mean weekly frequency of use by individual i. The second term in (12) will converge in probability to 0 as $N \to \infty$ under appropriate sampling conditions, and it is known that

$$N^{-1} \left| \sum_{i=1}^{N} R_{is_i\ell_i} - \omega_i \right| \quad \leq \quad N^{-1} \sum_{i=1}^{N} \left| R_{is_i\ell_i} - \omega_i \right| .$$

Therefore, the mean of the individual-level absolute deviations serves as an upper bound for the absolute deviation of the same mean from ω. While it is no guarantee, it seems likely that reducing the individual-level errors would lead to estimates for the population parameter with lower mean-squared errors. One way to potentially decrease the individual-level errors is to assign different recall periods to different individuals. If it is known prior to the survey that certain types of individuals tend to be more accurate under certain recall periods, individuals can be assigned the recall period most likely to produce the best estimates. Hopefully, such a strategy will lead to a general improvement in the population-wide estimate.

Naturally, in order to assess the degree of correspondence between recalled values and the true individual frequency, it is necessary to understand the dependence between $R_{is_i\ell}$ and ω_i. An assumption of independence between individual tendencies and recall would make calculations of the mean-squared errors much simpler, but ignores a likely positive correlation between the two variables. Namely, as ω_i increases so do the expected number of payments $A_{is_i\ell}$, and individuals who make more payments are likely to report a higher number of payments, even with imperfect recall. We rely on a parametric model and take advantage of the sample overlap between the DCPC and the PRD to infer the nature of this relationship by studying the paired diary data and recalled use data.

As this type of modeling can be quite complicated, we make a few simplifications. First, we consider only cash, since cash is the only one of the four main payment instruments for which adoption is not a major issue. The nonadoption of an instrument quite obviously affects use and, almost surely, recall, meaning that an in-depth analysis necessitates an adoption component in the model. Second, to avoid consideration of day-of-week effects, we limit ourselves to recall for a week or longer. Finally, to avoid the complexities introduced by missing values, we consider only those individuals who participated in both the PRD experiment and the 2012 DCPC. Because there is less temporal variability in the reported counts for the three longer recall periods, as evidenced by Figure 6, we consider a larger span of the PRD data than in Section 5. If we use only the respondents from the fifth phase, this would involve 656 individuals. By also including the fourth phase (PRD surveys taken in April, May, and June of 2012), this number increases to 804 individuals, 581 of whom provide recall data in both phases. From a notational point of view, we let $\mathcal{I}^{\mathrm{DP}}$ represent the collection of individuals who participated in both the DCPC and the PRD experiment in either phase four or five. Then, $\mathbf{R}^{\mathrm{DP}} = \left\{R_i^{\mathrm{DP}} \mid i \in \mathcal{I}^{\mathrm{DP}}\right\}$ with $R_i^{\mathrm{DP}} = \{R_i^p \mid i \in \mathcal{I}^4, i \in \mathcal{I}^5\}$ representing the set of observations for individual i from phases four and five.

In the following section, we describe the assumed model as well as the methodology used in fitting it to the data. Section 6.2 then analyzes the results and discusses how the results are used to search for individual demographic information linked to the efficiency of the three different recall periods.

6.1 Modeling Recall and Individual Weekly Frequencies

As alluded to above, information about the individual weekly frequency, ω_i, comes not only from the observed payment behavior in the diary but from the recalled values as well. Adopting a Bayesian framework, we are interested in the posterior distribution of the individual weekly frequencies conditional on observing both the diary data and the recall data for the chosen subsample. Thus, letting $\boldsymbol{\omega} = \{\omega_i \mid i \in \mathcal{I}^{\mathrm{DP}}\}$, the target distribution is

$$\mathrm{P}(\boldsymbol{\omega} \mid \mathbf{X}^{\mathrm{dcpc}}, \mathbf{R}^{\mathrm{DP}}) \quad \propto \quad \mathrm{P}(\boldsymbol{\omega})\mathrm{P}(\mathbf{X}^{\mathrm{dcpc}} \mid \boldsymbol{\omega})\mathrm{P}(\mathbf{R}^{\mathrm{DP}} \mid \boldsymbol{\omega}). \tag{13}$$

The simplification of the posterior distribution in (13) comes from an assumption of independent increments. Namely, because the diary periods do not overlap with the recall periods in any of the 804 selected individuals, use and recall of use are conditionally independent on the individual mean ω_i. Due to this decomposition, the right-hand side of equation (13) nicely decomposes into a prior distribution for $\boldsymbol{\omega}$, a likelihood for true behavior, and a likelihood for the recalled values. We discuss the former in Section 6.1.2, but begin by defining a model for $R_{is\ell} \mid \mu_i$ below in Section 6.1.1

6.1.1 Modeling Reporting Behavior

We assume that the likelihood for a recalled aggregate number of payments over a recall period of length $\ell > 7$, $R_{is\ell}$, takes the form

$$[R_{is\ell} \mid \omega_i, \lambda_{is\ell}] \sim \text{Poisson}\left(\lambda_{is\ell} \times \frac{\ell}{7}\omega_i\right), \tag{14}$$

where $\lambda_{is\ell} \geq 0$ corresponds to a scaling factor specific to the individual, recall period length, and particular implementation of the survey. The value of $\lambda_{is\ell}$ represents the degree of the expected bias of a recalled value since $\frac{\ell}{7}\omega_i$ is the expected number of payments made over $\frac{\ell}{7}$ weeks by an individual with weekly mean ω_i. The closer $\lambda_{is\ell}$ is to one, the less bias in recall. By contrast, if $\lambda_{is\ell}$ has a high degree of variability for an individual, this indicates little correlation between the true frequency and the reported behavior.

An important limitation of the model in (14) is that it is not well suited for a payment instrument with nonadopters. Instrument nonadoption by individual i means that $\omega_i = 0$, which implies that $\mathrm{P}(R_{is\ell} = 0 \mid \omega_i = 0) = 1$. Therefore, a nonadopter must always report zero payments under the model in (14). While this would be ideal, our data suggest that it is possible for a nonadopter to report using the payment instrument. Thus, for instruments in which nonadoption is significant, an additional stochastic component would need to be introduced to accurately model reporting of use. This is not the case with cash.

Before continuing, it is worth discussing the origins of and motivation behind the model in (14). As written, it is a purely statistical model in the sense that the parameters and equations serve only to link the individual weekly frequencies ω_i to the reported values on survey day s, $R_{is\ell}$, but do not in themselves carry any cognitive interpretation. For the purposes of this paper, this is sufficient. However, it is possible to relate (14) to a cognitive model based on evidence that recall is a composition of episodic enumeration and rate-based estimation (Blair and Burton 1986; Bradburn, Rips, and Shevell 1987).

As defined in (14), the distribution of $R_{is\ell} \mid \omega_i$ itself is unknown and will depend on the distribution of $\lambda_{is\ell}$. However, the first two moments, conditional on ω_i, will take the general forms

$$\mathrm{E}[R_{is\ell} \mid \omega_i] = g_1(i,s,\ell)\ell\omega_i \quad \text{and} \quad \mathrm{Var}\left[R_{is\ell} \mid \omega_i\right] = g_1(i,s,\ell)\ell\omega_i + g_2(i,s,\ell)\ell^2\omega_i^2 \tag{15}$$

for some functions $g_1(\cdot,\cdot,\cdot), g_2(\cdot,\cdot,\cdot)$. This structure is a special case of a model that links ω_i to $R_{is\ell}$ directly through a conditional relationship on the true number of payments made in the period in question, $A_{is\ell}$. The model assumes that the recalled value will be based on either an enumerative cognitive process or a rate-based estimate, with the likelihood of each depending on the length of the recall period. The probability of the latter is given by $c(\ell) \in [0,1]$, and based on the findings in (Blair and Burton 1986; Menon 1994), it is expected that $c(\ell)$ decreases as ℓ increases. Within the framework of enumeration, the recalled value is some scaled version of the truth, $\lambda_{is\ell}A_{is\ell}$. Within the framework of rate-based recall, the estimate does not depend on $A_{is\ell}$, but will, on average, be proportional to the individual's assessment of his weekly rate, $\gamma_i\omega_i$. If $\gamma_i = 1$, the individual correctly evaluates his weekly rate, and $\gamma_i > 1$ indicates an over-estimation. In order to account for expected variation in the rate-based recall estimate, we assume

$R_{is\ell} \sim \mathrm{Poisson}(\frac{\ell}{7} \times \gamma_i \omega_i)$ under this recall framework. Taken together, the overall model is

$$R_{is\ell} = \begin{cases} \lambda_{is\ell} A_{is\ell} & \text{w.p. } c(\ell) \\ \mathrm{Poisson}(\frac{\ell}{7} \times \gamma_i \omega_i) & \text{w.p. } 1 - c(\ell). \end{cases}$$

It can be shown that the first two moments of $R_{is\ell}$ will have the same structure as in (15), but the parameters in (16) have clearer interpretations with respect to the cognitive process. Although a conceptually simple model, it is possible that this model could be used to gain insight into human recall. Regardless, such considerations are beyond the scope of this paper.

With respect to the PRD data, we let $\boldsymbol{\lambda} = \{\lambda_i^{\mathrm{DP}}\}$, where $\lambda_i^{\mathrm{DP}} = \{\lambda_i^p \mid i \in \mathcal{I}^4, i \in \mathcal{I}^5\}$ and $\lambda_i^p = \{\lambda_{i\ell}^p \mid \ell = 7, 30, 365\}$ for all relevant phases and recall periods for individual i. Again, a simplification in the likelihood function arises through an assumption of conditional independence on $\boldsymbol{\omega}$ and $\boldsymbol{\lambda}$:

$$\begin{aligned} \mathrm{P}(\mathbf{R}^{\mathrm{DP}} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) &= \prod_{i \in \mathcal{I}^{\mathrm{DP}}} \mathrm{P}(R_i^{\mathrm{DP}} \mid \omega_i, \lambda_i^{\mathrm{DP}}) \\ &= \prod_{p=4}^{5} \prod_{i \in \mathcal{I}^{\mathrm{DP}} \bigcap \mathcal{I}^p} \mathrm{P}(R_i^p \mid \omega_i, \lambda_i^p) \\ &= \prod_{p=4}^{5} \prod_{i \in \mathcal{I}^{\mathrm{DP}} \bigcap \mathcal{I}^p} \prod_{\ell \in (7,30,365)} \mathrm{P}(R_{i\ell}^p \mid \omega_i, \lambda_{i\ell}^p). \end{aligned}$$

6.1.2 Fitting the Model

Having defined the likelihood function for recall, we are ready to define a full posterior distribution. We assume a priori that the $\boldsymbol{\lambda}$ follow a distribution defined by a set of hyper-parameters, θ, so that the posterior takes the form

$$\mathrm{P}(\boldsymbol{\omega}, \boldsymbol{\lambda}, \theta \mid \mathbf{X}^{\mathrm{dcpc}}, \mathbf{R}^{\mathrm{DP}}) \quad \propto \quad \mathrm{P}(\boldsymbol{\lambda} \mid \theta)\mathrm{P}(\theta)\mathrm{P}(\mathbf{X}^{\mathrm{dcpc}} \mid \boldsymbol{\omega})\mathrm{P}(\boldsymbol{\omega})\mathrm{P}(\mathbf{R}^{\mathrm{DP}} \mid \boldsymbol{\lambda}, \boldsymbol{\omega}). \tag{16}$$

The first two terms on the right-hand side of (16) correspond to the prior distributions assumed for $\boldsymbol{\lambda}$ and θ. We want to allow a great deal of flexibility in the $\lambda_{i\ell}^p$, as we know least about the mechanics of the recall process and hope most of the inference about ω_i comes from the diary data. As a result, we choose a fairly uninformative prior for $\lambda_{i\ell}^p$, while imposing a population structure via an assumed distribution. We therefore assume that $\lambda_{i\ell}^p \sim \mathrm{Gamma}(k^{(\ell)}, \tau^{(\ell)})$, with independence across individuals, recall period lengths, and phases. As before, this level of independence is unlikely and may reveal itself to be false in posterior analysis. However, because we are not generalizing results to a target population, there is no need to model the joint distribution of $(\omega_i, \lambda_{is\ell})$ for unsampled individuals.

With this choice of prior, the hyper-parameters are defined to be $\theta = \{(k^{(\ell)}, \tau^{(\ell)})\}$ for $\ell = 7, 30, 365$. In accordance with our desire for flexibility in the posterior distributions of $\boldsymbol{\lambda}$, we take $\mathrm{P}(\theta) \propto 1$. Thus, the data will completely drive the distribution of the scaling parameters.

The second two terms on the right-hand side of (16) can be expressed as

$$\mathrm{P}(\mathbf{X}^{\text{dcpc}} \mid \boldsymbol{\omega})\mathrm{P}(\boldsymbol{\omega}) \propto \mathrm{P}(\boldsymbol{\omega} \mid \mathbf{X}^{\text{dcpc}}). \tag{17}$$

The right-hand side of (17) corresponds to a posterior distribution of $\boldsymbol{\omega}$ conditional on the diary data, which we have estimated in the analysis of the DCPC data in Section 4. According to the assumed Model B,

$$\begin{aligned}
\omega_i &= \sum_{j=1}^{7} \exp(\mu_i + \beta_j) \\
&= \exp(\mu_i) \sum_{j=1}^{7} \exp(\beta_j),
\end{aligned}$$

where μ_i is assumed to have a Normal distribution with mean m_i and variance v_i. The estimates from the fits in Section 4 can be viewed as posterior estimates, $\hat{m}_i$ and $\hat{v}_i$, so that $\mu_i \mid \mathbf{X}^{\text{dcpc}} \sim \mathcal{N}(\hat{m}_i, \hat{v}_i)$ with $\mu_i, \mu_{i'}$ conditionally independent on $\mathbf{X}^{\text{dcpc}}$. This implies that $\omega_i \mid \mathbf{X}^{\text{dcpc}}$ will be proportional to a Log-Normal distribution, with density function $\frac{1}{K}\mathrm{LN}\left(\frac{w}{K} \mid \hat{m}_i, \hat{v}_i\right)$, where $K = \sum_{j=1}^{7} \exp(\hat{\beta}_j)$ and $\mathrm{LN}(\cdot \mid m, v)$ is the density function of a Log-Normal distribution with parameters m and v (corresponding to the mean and variance of the logarithm). As a result, ω_i will have posterior mean and variance given by

$$\mathrm{E}[\omega_i \mid \mathbf{X}^{\text{dcpc}}] = \left[\sum_{j=1}^{7} \exp(\hat{\beta}_j)\right] \exp\left(\hat{m}_i + \frac{\hat{v}_i}{2}\right)$$

and

$$\mathrm{Var}[\omega_i \mid \mathbf{X}^{\text{dcpc}}] = \left[\sum_{j=1}^{7} \exp(\hat{\beta}_j)\right]^2 \exp\left(2\hat{m}_i + \hat{v}_i\right)\left(\exp(\hat{v}_i) - 1\right).$$

We take advantage of the fact that the Log-Normal and Gamma distributions have the same domains and share similar density curves to use the latter to approximate the former. Corresponding distributions are generated by matching the first two moments. Thus, we assume

$$\omega_i \mid \mathbf{X}^{\text{dcpc}} \sim \text{Gamma}(k_i, \tau_i), \tag{18}$$

where $\tau_i = \frac{\text{Var}[\omega_i|\mathbf{X}^{\text{dcpc}}]}{\text{E}[\omega_i|\mathbf{X}^{\text{dcpc}}]}$ and $k_i = \frac{\text{E}[\omega_i|\mathbf{X}^{\text{dcpc}}]}{\tau_i}$. This change of distribution makes a minor difference in the inference about ω_i, but fits more naturally with the adopted model framework. Evidence of this can be found in Figure 9, which shows the distribution for $\omega_i \mid \mathbf{X}^{\text{dcpc}}$ under the Log-Normal (dashed) and the corresponding Gamma (solid, black) for six different individuals.

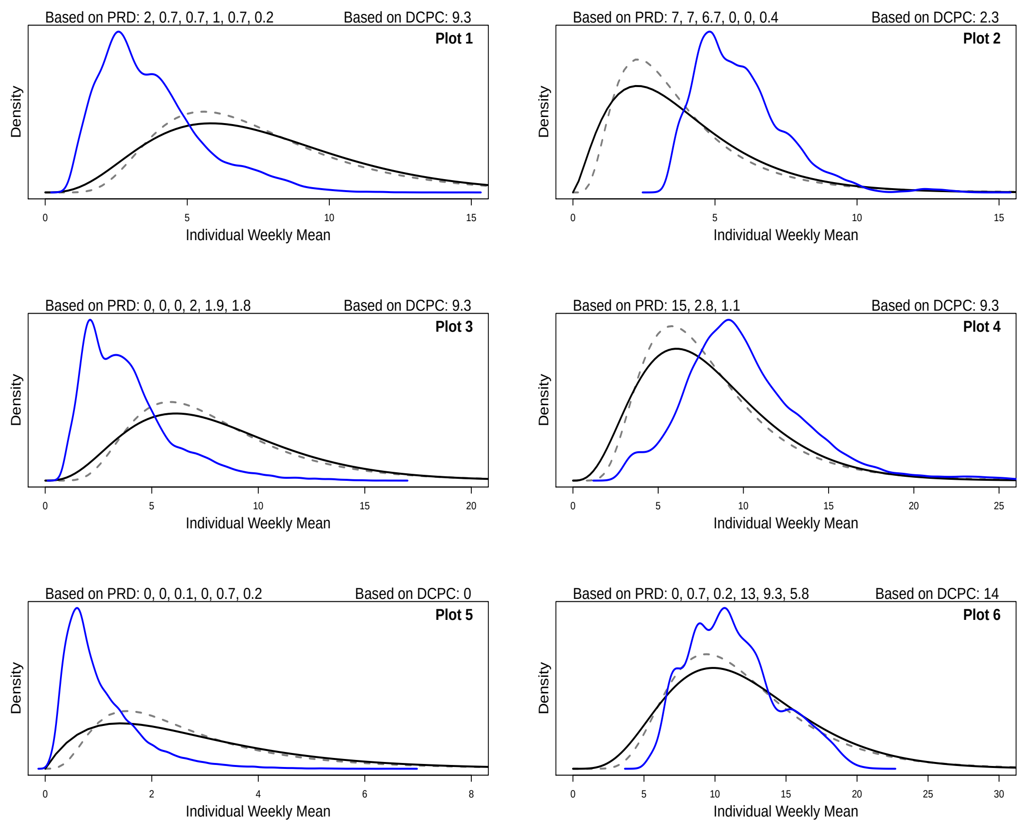

Figure 9: Prior distributions for ω_i for six different individuals based on the Log-Normal distribution as derived from Model B fits (dashed, gray) and the corresponding Gamma distributions (solid, black), as well as kernel density estimates of the posterior distribution of ω_i (solid, blue). Estimates based on the PRD data are simply reported values scaled to a weekly frequency and ordered by phase and by week, month, year within each phase. Estimates based on the DCPC are the three-day totals scaled to a weekly frequency.

With the full posterior distribution determined, we proceed by fitting the model to the data with an MCMC iteration algorithm. In each iteration of the MCMC algorithm, parameters are either updated with a Gibbs sampler or with a Metropolis-Hasting step. The former is used whenever the posterior distribution has a well-known form, such as in the case of the shape parameters of the Gamma distribution that define $\boldsymbol{\lambda}$, which can be shown to follow an Inverse-Gamma distribution. To assure convergence before sampling from the posterior distribution, we impose a burn-in period of 1000 iterations. In order to decrease autocorrelation, we thin by keeping only every 20^{th} posterior sample.

The use of a Gamma distribution for $\mu_i \mid \mathbf{X}^{\text{dcpc}}$ not only makes the MCMC algorithm easier, but also means that the posterior distribution of $\omega_i \mid \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}, \boldsymbol{\lambda}$ will also follow a Gamma distribution. Specifically,

$$\omega_i \mid \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}, \boldsymbol{\lambda} \sim \text{Gamma}(k_i', \tau_i'), \tag{19}$$

where

$$k_i' = k_i + \sum_{p=4}^{5} \sum_{\ell \in (7,30,365)} R_{i\ell}^{p} 1\left[i \in \mathcal{I}^p\right] \quad \text{and} \quad \tau_i' = \left[\frac{1}{\tau_i} + \frac{402}{7} \sum_{p=4}^{5} \lambda_{i\ell}^{p} 1\left[i \in \mathcal{I}^p\right]\right]^{-1}.$$

Posterior sampling of ω_i, therefore, only involves sampling the pairs (k_i', τ_i') upon convergence, as these parameters can then be used to sample the individual weekly frequencies.

6.2 Results

Having run the MCMC algorithm, we begin with an assessment of the effect of incorporating the PRD recall data into inference about the individual weekly frequencies, ω_i. The MCMC algorithm can be conceptually viewed as updating distributions of ω_i based only on the DCPC data, with additional information provided by the PRD data. The former is defined by the parameters (k_i, τ_i) in (18), while the latter by the sampled posterior parameters (k_i', τ_i') in (19). While we obviously expect changes in the estimated distributions for the individual frequencies with the added recall information, it would be worrisome if the changes were drastic. After all, we believe the DCPC data correspond to the true behavior of each individual over the three-day period. Posterior estimates in which likely values of ω_i were inconsistent with the observed X_i^{dcpc} would be troubling.

Based on the generated posterior estimates of (k_i', τ_i') or the prior estimates of (k_i, τ_i), we can sample ω_i for $i \in \mathcal{I}^{\text{DP}}$ and, by repeating this exercise, generate estimates of the sample averages $\frac{1}{804} \sum_{i \in \mathcal{I}^{\text{DP}}} \omega_i$. One set of estimates will be based only on the DCPC, while the other will be based on both the DCPC and the PRD. Kernel density estimates for this sample average for each are shown in Figure 10. This plot reveals a slight downward shift in the sample average when recall data are included. This is to be expected, as the analysis in Section 5 showed that for recall periods of a week or longer, the recall-based estimates were lower than those based on the diary. Overall, however, the shift is not significant, with a great deal of overlap between the two distributions.

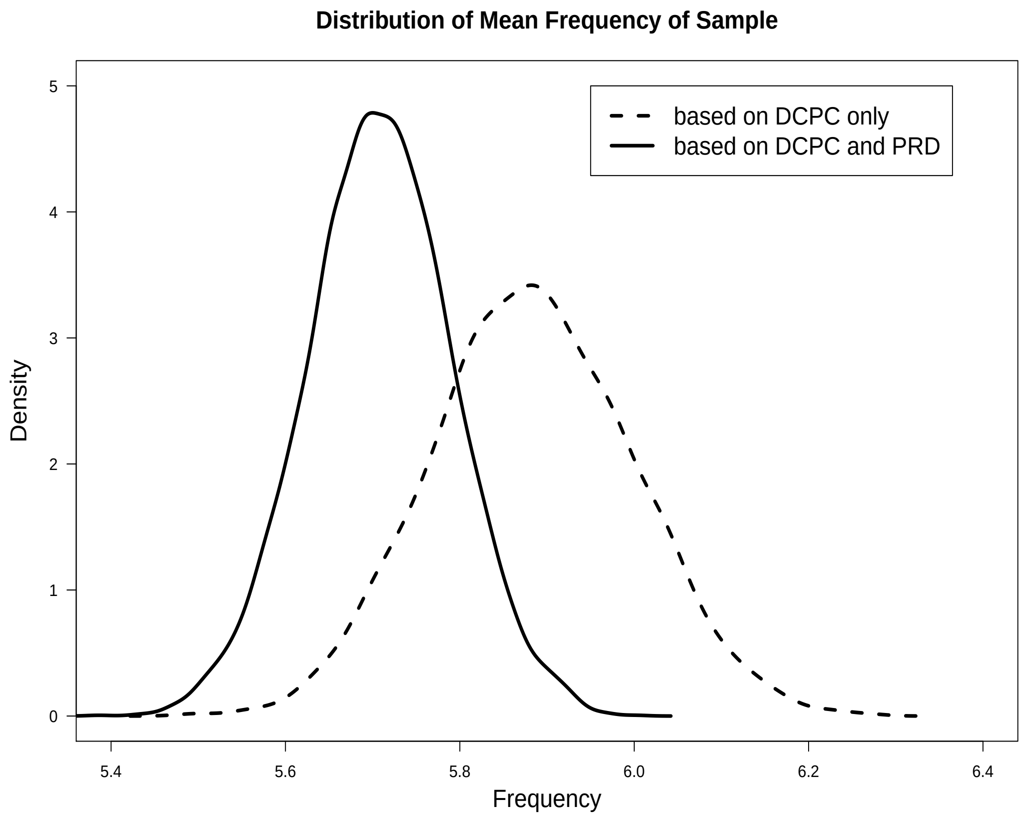

Figure 10: Kernel density estimates of the sample average of ω_i among the 804 individuals participating in either phase four or five of the PRD and the 2012 DCPC based on the prior distributions of ω_i and posterior distributions of ω_i.

Figure 9 shows the influence of the PRD recall data on the inference of ω_i at the individual level for six respondents. As expected, the estimated density curve for ω_i shifts in the direction of the recall-based estimates, with larger recalled estimates driving the mean value of ω_i up (as in Plots 2 and 4) and lower values of recall driving the mean down (as in Plots $1, 3$ and 5). In addition, in cases where the PRD data and the DCPC data are consistent with one another (as in Plot 5), the uncertainty in the value of ω_i decreases. Conversely, Plot 6 shows a case where the recall-based estimates vary significantly, and this lack of clarity means that the overall distribution for ω_i does not change much from that based on the DCPC alone.

6.2.1 Predicting Individual Accuracy of Recall Periods

Cognitive processes associated with different recall periods may lead to systematic under- or overestimation of the truth, and it is possible that these effects vary from individual to individual. Thus, for a given individual, we assume that different recall periods carry different levels of accuracy. For individual i, this idea is captured by

$$B_{is} = \text{argmin}_{\ell \in (7,30,365)} \left| \frac{7}{\ell} R_{is\ell} - \omega_i \right|,$$

which is the recall period length that minimizes the absolute deviance of the empirically derived weekly rate, $\frac{7}{\ell} R_{is\ell}$, from the true weekly rate ω_i. In practice, this random variable is not observable; we neither know the true value of ω_i nor do we generally observe responses for more than one recall period in one survey. However, in the case of the PRD data, we do observe the latter. As a result, we can combine this data with the estimated posterior distributions for ω_i to determine probability distributions on

$$B_i^p \quad = \quad B_{is_i^p} \mid \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}$$

and thus $\mathbf{B}^{\text{DP}} = \left\{ B_i^p \mid i \in \mathcal{I}^{\text{DP}}, p = 4, 5 \right\}$. Given our fitted model, we define

$$\begin{aligned} q_i^p(\ell) &= \mathrm{P}(B_i^p = \ell \mid \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}) \\ &= \int_0^\infty \mathrm{P}(B_i^p = \ell \mid \omega_i, \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}) \mathrm{P}(\omega_i \mid \mathbf{X}^{\text{dcpc}}, \mathbf{R}^{\text{DP}}) \mathrm{d}\omega_i \end{aligned} \tag{20}$$

as the estimated probability of the recall period of length ℓ being the closest to the true weekly mean for individual i in the survey taken in phase p. The first term within the integral in (20) takes values of 1 or 0 only, as given the individual weekly frequency ω_i and the reported values $R_{i\ell}^p$, we can determine B_i^p with absolute certainty. To estimate $q_i^p(\ell)$, we can thus sample ω_i from its derived posterior and determine which reported value is most consistent with the sampled value of ω_i. Repeatedly doing so should converge to the desired probabilities, whose estimates we call $\hat{q}_i^p(\ell)$. Given these probabilities, we can in turn sample $\mathbf{B}^{\text{DP}}$, assuming independence across individuals and phases, according to $\mathrm{P}(B_i^p = \ell) = \hat{q}_i^p(\ell)$. We refer to the d^{th} draw of $\mathbf{B}^{\text{DP}}$ as $\mathbf{b}_d^{\text{DP}}$.

In doing so, we gain some insight into the distribution of B_i^p among the sample. We estimate that approximately 56 percent of surveys have closest correspondence between reported behavior and weekly frequency occurring for a week-based recall period, with 26 percent doing better at a month and 18 percent at a year. Although, as a whole, the weekly recall period performs best in terms of estimating the population mean, there is a fair amount of variation in terms of the optimal period at the individual level. This is interesting as, under unbiased recall, longer recall periods would presumably produce better estimates due to the Law of Large Numbers.

The hope is that $\mathrm{P}(B_{is} = \ell)$ is related to some set of demographic variables, which we generically refer to as $\texttt{demo}_i$. Alternatively, the consistency of responses with true behavior might not be predictable based on demographic insights for the individual. These two options correspond to the following models:

Model 0: $P(B_{is} = \ell) \propto \exp\left(\psi_{0\ell}\right)$

Model 1: $P(B_{is} = \ell) \propto \exp\left(\texttt{demo}_i^T \psi_{1\ell}\right)$

for some set of parameters $\psi_0 = \{\psi_{0\ell}\}$ and $\psi_1 = \{\psi_{1\ell}\}$.

For a given sample, $\mathbf{b}_d^{\mathrm{DP}}$, we can estimate the parameters, $\hat{\psi}_m$ for $m = 0, 1$, via a maximum-likelihood procedure using the multinom function in R and calculate the deviance

$$\texttt{Dev}(\mathbf{b}_d^{\mathrm{DP}}) \quad = \quad -2\left[\log \mathcal{P}(\hat{\psi}_1 \mid \{\texttt{demo}_i\}, \mathbf{b}_d^{\mathrm{DP}}) - \log \mathcal{P}(\hat{\psi}_0 \mid \mathbf{b}_d^{\mathrm{DP}})\right].$$

Sampling $\mathbf{B}^{\mathrm{DP}}$ 500 times and calculating the deviance for both models for each draw allows us to estimate the expected deviance of the two models for a particular draw:

$$\hat{\mathrm{E}}[\texttt{Dev}(\mathbf{B}^{\mathrm{DP}})] \quad = \quad \frac{1}{500}\sum_{d=1}^{500} \texttt{Dev}(\mathbf{b}_d^{\mathrm{DP}}).$$

Based on this expected deviance, we can calculate a p-value based on the degrees of freedom corresponding to the difference in the number of free parameters in the two models. For demographic variables we consider age, education, and gender. Age and education are treated as a numerical variables, with education coded as the number of grades completed; gender is treated as a categorical variable. This analysis leads to an estimate of the deviance of 9.16 with 6 degrees of freedom, corresponding to a p-value of 0.16. This suggests little relationship between the optimal size of the recall period and individual demographics. Indeed, Figure 11 shows the estimated probabilities of $B_i^p = 7$ for six different draws of $\mathbf{B}^{\mathrm{DP}}$. These plots suggest a slight trend in which older individuals are more likely to perform better with weekly recall, but the change as age increases is rather small. Despite this relationship, in all draws, the fitted probabilities uniformly suggest that a weekly recall period is the most likely to well represent true behavior, no matter the individual demographics. Therefore, a survey methodologist is best served by using a weekly recall period for all individuals, regardless of age, education, or gender.

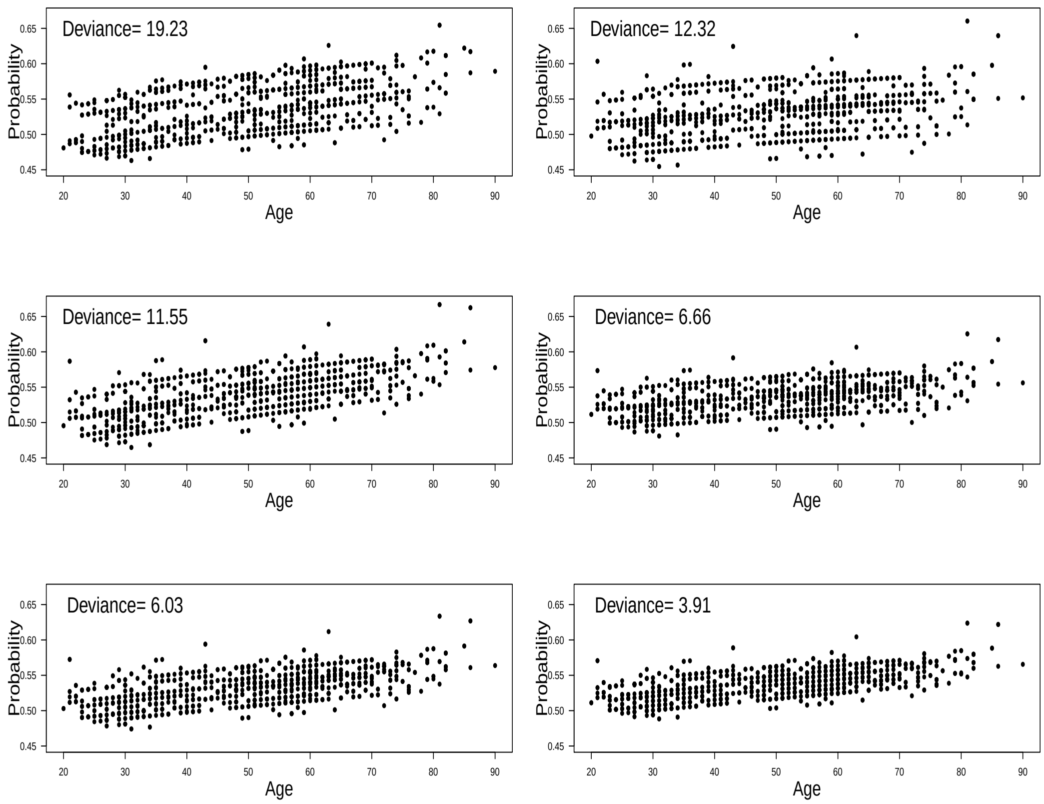

Figure 11: Fitted probabilities, $\hat{P}(B_i^p = 7 \mid \mathbf{X}^{\mathrm{dcpc}}, \mathbf{R}^{\mathrm{DP}})$ based on six draws $\mathbf{b}_d^{\mathrm{DP}}$.

7 Discussion

By comparing diary-based use data gathered in the 2012 DCPC with recalled estimates of the number of payments for various periods of time in the PRD experiment, we hope to learn about the efficiency of estimates of mean population frequency of use as a function of the recall period length. The analysis was done under the assumption of a simple random sample from the population of interest and according to a model of payment behavior with variation resulting from day-of-week effects. We compared mean-squared errors and biases for population estimates based on recall from four different recall periods.

Overall, the optimal recall periods differed across instruments. Perhaps most interesting, the daily recall proved not optimal for any of the instruments, and, other than for cash, was the worst by a sizeable margin. As the mean-squared error is dominated by the squared bias for large enough sample sizes, this is somewhat surprising, as one might expect shorter recall periods to yield less biased results, since it is easier to enumerate events. Nevertheless, this phenomenon is consistent with other studies that show that recalled data often over-estimate diary-based estimates on topics as diverse as household food consumption (Ahmed, Brzozowski, and Crossley 2010), length of hospital visits (Clarke, Fiebig, and Gerdtham 2008), and extent of exercise (Nusser et al. 2012).

The choice of optimal recall period seems related to the type of payments for which each payment instrument is used. Cash, which is most closely associated with small-valued and frequent payments, shows the highest efficiency when estimated by weekly recall. On the other end of the spectrum, there is strong evidence that average check use is best determined by yearly recall. Finally, recall of credit and debit card use seems most accurate on either the monthly or yearly scale. As noted in the introduction, analyses of a variety of consumer surveys show that different recall periods may be best for different variables being measured (Ahmed, Brzozowski, and Crossley 2010; Deaton and Grosh 2000; Hurd and Rohwedder 2009; NSSO Expert Group on Sampling Errors 2003). Hurd and Rohwedder (2009) explicitly suggest that the optimal recall period is related to the frequency of a behavior, with shorter recall periods better suited for more frequent behaviors. While we find consistent over-estimation with daily recall, which was not measured in the aforementioned authors' study, our results for the other three recall periods seem consistent with this hypothesis.

The second major component of this work involves inference about the relationship between recalled values and an individual's mean behavior, for different recall periods. The intention was to see whether expected deviations of estimates based on different recall periods for individual frequencies could be related to demographic information of the individuals. A strong relationship of this nature might suggest a sampling design, based on assigning individuals different recall periods, that could decrease the error for a sample-based estimate for the population frequency. This analysis was done only for cash. We found no evidence of a demographic variable that is highly correlated with the accuracy of a recall period. As a result, our results suggest that, at least for cash, a design in which everyone is asked for weekly recall is optimal. Although in a different context, Sudman and Bradburn (1973) found that in terms of recalling sick leave, only age, out of a set of measured demographic variables, related to the quality of memory.

While the individual-level analysis did not yield a methodology for improving efficiency of population frequency estimates, it does relate to a modeling framework that could potentially yield insights into the cognitive recall process. Specifically, it could illuminate the balance between reliance on enumeration of events and rate-based recall for different recall periods as well as the biases associated with each type of recall feature in the model. An extension of the model to payment instruments other than checks requires the incorporation of a new variable corresponding to adoption of a payment instrument, since other than cash, anywhere from 10 to 25 percent of individuals identify themselves as people who never use a payment instrument. This variable is sure to influence use (in an obvious way) as well as recall.

The results in this work are based on the use of two distinct data sets. While both are taken from the American Life Panel and with a significant degree of overlap in respondents, there is a possibility that a discrepancy in the samples contributes to differences in recalled averages and diary-reported averages, thus affecting our conclusions. Similarly, the fact that recall occurs and relates to different portions of time than the three-day diary period could have the same effect. For these reasons, it would be beneficial to conduct the analysis for one sample in which recall and true payment behavior are recorded for each individual over the same period of time. As noted, the use of the diary as a proxy for the truth could also lead to false conclusions if diary results are systematically introducing a bias as well. Data based on official records are being sought for those payment instruments for which they exist.

References

Ahmed, Naeem, Matthew Brzozowski, and Thoms F. Crossley. 2010. "Measurement Errors in Recall Food Consumption Data." *Institute for Fiscal Studies Working Papers*.

Angrisani, Marco, Arie Kapteyn, and Scott Schuh. 2012. "Measuring Household Spending and Payment Habits: The Role of 'Typical' and 'Specific' Time Frames in Survey Questions." In *Improving the Measurement of Consumer Expenditures*, eds. Christopher Carroll, Thomas Crossley, and John Sabelhaus, chap. 15. NBER.

Babor, Thomas F., Joseph Brown, and Frances K. Del Boca. 1990. "Validity of Self-Reports in Applied Research on Addictive Behaviors: Fact or Fiction?" *Behavioral Assessment* 12: 5–31.

Battistin, Erich, and Mario Padula. 2010. "Survey Instruments and the Reports of Consumption Expenditures: Evidence from the Consumer Expenditure Surveys." Centre for Studies in Economics and Finance Working Papers.

BHPS. Various Years. "British Household Panel Survey." https://www.iser.essex.ac.uk/bhps.

Blair, Edward, and Scott Burton. 1986. "Processes Used in the Formulation of Behavioral Frequency Reports in Surveys." *American Statistical Association Proceedings of the Section on Survey Methods* 481–487.

BMCS. Various Years. "Biennial Media Consumption Survey." http://www.cpanda.org/data/profiles/bmcs.html.

Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys." *Science* 236: 157–161.

BRFSS. Various Years. "Behavioral Risk Factor Surveillance System." http://www.cdc.gov/brfss/.

CES. Various Years. "Consumer Expenditure Survey." http://www.bls.gov/cex/.

Chang, Linchiat, and Jon A. Krosnick. 2003. "Measuring the Frequency of Regular Behaviors: Comparing the Typical Week to the Past Week." *Sociological Methodology* 33: 55–80.

Clarke, Philip M., Denzil G. Fiebig, and Ulf-G. Gerdtham. 2008. "Optimal Recall Length in Survey Design." *Journal of Health Economics* 27: 1275–1284.

DCPC. Various Years. "Diary of Consumer Payment Choices."

Deaton, Angus, and Margaret Grosh. 2000. "Consumption." In *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten years of LSMS Experience*, eds. Margaret Grosh and Paul Glewwe, chap. 17. The World Bank.

Eisenhower, Donna, Nancy A. Mathiowetz, and David Morganstein. 1991. "Recall Error: Sources and Bias Reduction Techniques." In *Measurement Errors in Surveys*, eds. Paul P. Biermer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. Wiley.

Foster, Kevin, Erik Meijer, Scott Schuh, and Michael A. Zabek. 2011. "The 2009 Survey of Consumer Payment Choice No. 11–1." Federal Reserve Public Policy Discussion Papers.

Foster, Kevin, Scott Schuh, and Hanbing Zhang. Forthcoming. "The 2010 Survey of Consumer Payment Choice." Federal Reserve Bank of Boston, Working Paper.

Fritsch, F.N., and R.E. Carlson. 1980. "Monotone Piecewise Cubic Interpolation." *SIAM Journal on Numerical Analysis* 17(2): 238–246.

Hurd, Michael, and Susann Rohwedder. 2009. "Methodological Innovations in Collecting Spending Data: The HRS Consumption and Activities Mail Survey." *Fiscal Studies* 30: 435–459.

Huttenlocher, Janellen, Larry V. Hedges, and Norman M. Bradburn. 1990. "Reports of Elapsed Time: Bounding and Rounding Processes in Estimation." *Journal of Experimental Psychology, Learning, Memory and Cognition* 16: 196–213.

Kemsley, William F. F., and J. L. Nicholson. 1960. "Some Experiments in Methods of Conducting Consumer Expenditure Surveys." *Journal of the Royal Statistical Society, Series A* 123(3): 307–328.

McKenzie, John. 1983. "The Accuracy of Telephone Call Data Collected by Diary Methods." *Journal of Marketing Research* 20: 417–427.

McWhinney, Isabel, and Harold Champion. 1974. "The Canadian Experience with Recall and Diary Methods in Consumer Expenditure Surveys." In *Annals of Economic and Social Measurement*, ed. Sanford V. Berg, vol. 3(2), 113–140. NBER Books.

Means, Barbara, Gary E. Swan, Jared B. Jobe, James L. Esposito, and Elizabeth F. Loftus. 1989. "Recall Strategies for Estimation of Smoking Levels in Health Surveys." Paper for the American Statistical Association Meetings.

Menon, Geeta. 1994. "Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies." In *Autobiographical Memory and the Validity of Retrospective Reports*, eds. Norbert Schwarz and Seymour Sudman, 161–172. Springer-Verlag.

MPS. Various Years. "Methods of Payment Survey." http://www.bankofcanada.ca/2012/09/publications/research/

Neter, John, and Joseph Waksberg. 1964. "A Study of Response Errors in Expenditure Data from Household Interviews." *Journal of the American Statistical Association* 59: 18–55.

NHIS. Various Years. "National Health Interview Survey." http://www.cdc.gov/nchs/nhis.htm.

NSSO Expert Group on Sampling Errors. 2003. "Suitability of Different Reference Periods for Measuring Household Consumption: Result of a Pilot Study." *Economic and Political Weekly* 37(4): 307–321.

Nusser, Sarah M., Nicholas K. Beyler, Gregory J. Welk, Alicia L. Carriquiry, Wayne A. Fuller, and Benjamin M. N. King. 2012. "Modeling Errors in Physical Activity Recall Data." *Journal of Physical Activity and Health* 9: 56–67.

PSID. Various Years. "Panel Study of Income Dynamics." http://psidonline.isr.umich.edu/.

Rubin, David C., and Alan D. Baddeley. 1989. "Telescoping is Not Time Compression: A Model of the Dating of Autobiographical Events." *Memory and Cognition* 17(6): 653–661.

SCA. Various Years. "Survey of Consumers." http://www.sca.isr.umich.edu/.

Silberstein, Adriana R., and Stuart Scott. 1991. "Expenditure Diary Surveys and Their Associated Errors." In *Measurement Errors in Surveys*, eds. Paul P. Biermer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. Wiley and Sons.

Strube, Gerhard. 1987. "Answering Survey Questions: The Role of Memory." In *Social Information Processing and Survey Methodology*, eds. Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman, 86–101. Springer-Verlag.

Sudman, Seymour, and Norman M. Bradburn. 1973. "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association* 68(344): 805–815.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.

Tourangeau, Roger, Kenneth Rasinski, and Norman Bradburn. 1991. "Measuring Happiness in Surveys: A Test of the Subtraction Hypothesis." *The Public Opinion Quarterly* 55: 255–266.