Panel Design Effects on Response Rates and Response Quality*

Rene Segers† Philip Hans Franses
Econometric Institute and Tinbergen Institute
Erasmus University Rotterdam

Econometric Institute Report EI 2007-29

August 29, 2007

Abstract

To understand changes in individuals' opinions and attitudes it would be best to collect data through panels. Such panels, however, often cause irritation among respondents, resulting in low response rates and low response quality. We address whether this problem can be alleviated by designing a panel survey in an alternative way. For this purpose, we perform two field studies where we measure the effects of several panel design characteristics on response rates and response quality. These characteristics include the number of waves and the time between subsequent waves, which may either be fixed or random.

Our findings suggest that response rates and response quality can be improved significantly by surveying at random time intervals. It is then crucial that panel members are not informed about the dates they will be surveyed, because in this case respondents are less likely to develop expectations as to when they will be surveyed again. The methodology we put forward can be used to improve the efficiency of a panel study by carefully calibrating the studies' panel designs parameters.

Keywords: Panel design, randomized sampling, time sampling, nonresponse, panel conditioning
JEL Classification Codes: C33, C42, C81

*We thank participants of the 2005 INFORMS Marketing Science Conference in Atlanta, Georgia and the 2007 INFORMS Marketing Science Conference in Singapore for their helpful comments.

†Corresponding author. Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Tel.: +31-10-4082524, fax: +31-10-4089162. E-mail addresses: `rsegers@few.eur.nl` (R. Segers), `franses@few.eur.nl` (P. H. B. F. Franses).

1 Introduction

Understanding changes in individuals' opinions and attitudes is of fundamental interest in the social sciences. To measure such changes, it is desirable to conduct a longitudinal, or panel study, where the same individuals are surveyed at multiple points in time. This allows the researcher to study changes in opinions and attitudes over time at the individual level and to capture the dynamic relationships between events and behavior. Secondly, it is desirable to conduct the survey on a frequent basis. This allows the researcher to establish whether certain changes are permanent or transitory over time.

One can easily appreciate however that frequent surveying of the very same individuals likely deteriorates the quality of the survey. People get irritated and they disconnect from the panel, thereby making the panel less efficient. Or perhaps worse, respondents' (reported) opinions and attitudes may change due to being a member of a panel, which is called panel conditioning.

In sum, monitoring individuals at a high frequency is useful, but proper data collection is not trivial. It is precisely this topic that we address in our paper, that is, can we design ways of better data collection? We do believe we can, as we will argue in theory and as we show with various experiments.

The main contribution of our paper is that we propose a general framework to study the effects of panel design characteristics on response rates and response quality. We demonstrate that response rates and response quality can be significantly improved if the design of the panel is carefully calibrated. This makes the survey more efficient. As such, our approach helps to design efficient panels that are cost effective.

Although nonresponse bias can be handled quite effectively a posteriori using model-based procedures, obviously, preventing nonresponse at the data collection stage is to be preferred as nonresponse due to irritation may harm the relationship with the respondents. This approach has been successful in the field of *questionnaire* design, where the effects of design characteristics such as incentives, the length, and the presentation of the questionnaire have been examined, see for example the work on mail surveys by Adams and Darwin (1982) and Yu and Cooper (1983), and the more recent books by Dillman (2000) and Couper *et al.* (2004), among others. It also relates to the literature on consumer psychology, where the effect has been studied that respondents are often induced by the measurement process to form judgements, see Simmons *et al.* (1993), Dholakia and Morwitz (2002) and Morwitz and Fitzsimons (2004), among others. Other interesting contributions have been made in optimal experimental design theory, where the aim is to design experiments such that statistical inference is most efficient, see for example Atkinson and Donev (1992) for a general exposition, the work in psychological research by Allison *et al.* (1997) and McClelland (1997), and the work in econometrics by De Stavola (1986), Nijman *et al.* (1991) and Ouwens *et al.* (2002). In contrast, to the best of our knowledge, the effects of *panel* design characteristics on response and response quality have not been studied extensively yet, and that is what we do here.

Our paper is organized as follows. In the next section we discuss various types of response bias relevant to panel surveys, and review several panel designs that have been proposed in the literature to reduce these biases. In Section 3 we formulate our hypotheses regarding the effects of panel design on response rates and response quality. In Section 4 we introduce a flexible panel design. Together with a dynamic panel Tobit-II model the design is employed to test our hypotheses. Section 5 presents our empirical results. Finally, Section 6 concludes.

2 Background

In order to provide context for our discussion, this section provides the relevant background on types of survey nonresponse, response quality, and panel design. Finally, we touch upon the issue of modelling incomplete panels.

2.1 Nonresponse and Response Quality

Survey nonresponse

Once an individual has agreed to become a member of a panel, three different types of nonresponse may occur, see Verbeek (1991) and Schafer and Graham (2002), among others. The weakest form is *item nonresponse*, where panel members do not answer one or more particular questions of the survey. More serious is *wave nonresponse*. In this case, panel members do not participate in the survey during one or more particular waves. Ultimately, *attrition* or *dropout* occurs when panel members disconnect from the survey prematurely.

Response quality

Besides response rates, also response quality may be at stake if a panel design demands much from participants. Response quality is used as an all-embracing term that covers the desire to obtain survey data that is not biased by the survey environment. For example, an often demonstrated threat to response quality is the mere-measurement effect, which is the effect that measuring an individual's preferences may change his or her subsequent behavior, see Dholakia and Morwitz (2002) and Morwitz and Fitzsimons (2004). In our paper we are particularly concerned with potential biases in response due to panel participation. This form of response bias is usually referred to as panel conditioning or time-in-sample bias. Following Trivellato (1999), we distinguish two different types of panel conditioning bias. First, as a consequence of being a panel member, respondents may change their reporting behavior. For example, because panel members are typically asked to complete similar surveys repeatedly, they tend to get less involved in completing them. For example they may incorrectly report exactly the same now as they did during the previous wave. Alternatively, respondents may give socially desirable responses as they start to be aware of being monitored, an effect usually referred to as the Big Brother effect. Second, a respondent may change his or her attitudes or opinions due to panel participation. For example, having to report an opinion repeatedly may cause a respondent to think more about his or her opinion, reconsider it, and even change it.

As opposed to nonresponse bias, measuring panel conditioning bias through revealed preference data is not trivial. It requires a careful comparison of the responses given in the first wave of data collection, which is free of panel conditioning bias by definition, and in the next waves. We extend the notion of Hansen (1980)[1] who argues that there should not be a difference in the response distribution of different subgroups of panel members who have been exposed to different methods of data collection. In our case, this implies that responses should not depend on the particular panel design chosen, nor on the wave of data collection.

[1]See Deutskens *et al.* (2004) for a recent application of this notion.

2.2 Panel Design

The extremes: complete panels and repeated cross-sections

Complete panel data sets, as illustrated in Figure 1, Panel (a), allow a researcher to study changes in individuals' opinions over time at the individual level. Typically the individuals are indexed by i, $i = 1, ..., N$, and time is index by t, $t = 1, ..., T$. Using complete panels one may capture the dynamic relationships between events and behavior, and control for time-varying and individual-specific characteristics. One can also account for unobserved heterogeneity by exploiting the time invariance of the unobserved individual characteristics. However, since such complete panels are most vulnerable to nonresponse and to a low quality of response, this is often not feasible. In this subsection we therefore explore several pseudo-panel designs, mostly designed with the aim to reduce costs and respondent burden. We discuss whether these designs are attractive from a statistical point of view.

Perhaps the most rigorous way to reduce respondent burden is to collect repeated cross-sectional data instead of complete panel data, as illustrated in Figure 1, Panel (b). In this design, at each point in time, a unique group of individuals, to be denoted by j, where $j = 1, ..., J$, is requested to complete a survey. Since in this case individuals are surveyed only once, this ensures that individuals' current opinions are not biased by previous panel participation. As a consequence, there is no panel conditioning bias. Several scholars have argued that the estimation of dynamic models at the individual level is possible on the basis of repeated cross-sections, see for example Deaton (1985), Verbeek and Nijman (1993), Moffitt (1993) and Collado (1997). However, the identification conditions for these estimators are very strong, and potentially unrealistic in many empirical applications (Verbeek and Vella, 2005). Alternatively, one may use matching techniques to match similar individuals and form a pseudo panel, but their assumptions are no less strong.

Rotation

In general, one wants to preserve the possibility to study individual dynamics. Still, individuals can only be monitored for a short period of time. A natural compromise is therefore to apply a panel refreshment strategy. This pertains to requesting each panel member to only join the panel for a fixed period of, say T^* time periods.

Second, each wave (or block of waves), a new group of individuals is invited to join the panel, with the aim to keep the total number of panel members constant. The latter strategy is also referred to as rotation, see Patterson (1950) and Kish and Hess (1959). An example of a rotating panel is shown in Figure 1, Panel (c). For analysis-of-variance models, Nijman *et al.* (1991) demonstrated how to set up a rotating panel to maximize estimation efficiency. They showed that the efficiency gains from using an optimal rotation strategy can be quite substantial, even if the costs of a reinterview equal the costs of acquiring a new panel member.

- Insert Figure 1 about here -

Continuous sampling and time sampling

Once one has decided upon the number of time periods T^* that an individual will be requested to join the panel and possibly upon a particular rotation strategy, naturally, a next step would be to decide upon the number of survey requests within this period, to be labelled n. Note that T^* and n together constitute the sampling frequency $f = n/T^*$ of the survey, which is equal to the reciprocal of the time between subsequent waves. Often the sampling frequency is set equal to the desired data frequency f^d. We will refer to this situation as continuous sampling. For example, suppose that the aim is to ask individuals to join a panel for two months and the desired sampling frequency f^d is weekly, but the maximum accepted number of survey requests n is four. A continuous sampling strategy would then imply that each panel member is interviewed weekly, but only during the first four weeks. See Figure 2, Panel (a), for an illustration.

In many cases, however, panel members are better observed over the longest horizon possible to detect possible changes in opinions in the longer run. This suggests to space the observations as much as possible, which in our example would imply to interview biweekly. In order to still be able to monitor on the desired weekly basis, one may opt to ask one-half of the panel members to complete a survey in the even weeks and the other half to complete a survey in the odd weeks, as illustrated in Figure 2, Panel (b). This approach, known as time sampling, has gained attention in recent years. It has been accepted as the natural way to lower the sampling frequency while keeping the data frequency unchanged.

- Insert Figure 2 about here -

Randomized sampling

Recall that we are interested in measuring individuals' attitudes and opinions and their changes over time. Hence we are interested in the autocorrelations of these attitudes and opinions. From this perspective it seems also appealing to consider choosing the n survey occasions at random, independently for each panel member, as illustrated in Figure 2, Panel (c). We argue as follows. To reveal the underlying correlation structure in the data it is important to measure autocorrelations at many different lags. This facilitates the identification of any type of individual dynamics in the data and efficient estimation.

For the three different sampling strategies given in Figure 2 we obtain the following (expected) numbers of observations of the $k-$th autocorrelation that are collected, $A(k)$:

- For continuous sampling (CS):

$$A_{CS}(k) = \begin{cases} n-k & \text{for } k = 1, 2, ..., n-1 \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

- For time sampling (TS) with frequency $f = n/T$:

$$A_{TS}(k) = \begin{cases} n-kf & \text{for } k = 1/f, 2/f, ..., (n-1)/f \\ 0 & \text{elsewhere} \end{cases} \tag{2}$$

- For randomized sampling (RS)[2]:

$$E[A_{RS}(k)] = \begin{cases} \frac{\binom{T-2}{n-2}}{\binom{T}{n}}(T-k) = \frac{\binom{n}{2}}{\binom{T}{2}}(T-k) & \text{for } k = 1, 2, ..., T-1 \\ 0 & \text{elsewhere} \end{cases} \tag{3}$$

The three functions are depicted in Figure 3. In the case of randomized sampling, data is collected to measure every possible autocorrelation up to $T-1$ lags, where the lower lag orders are sampled most frequently. Conceptually, it is convenient to

[2]To see that this result is correct, note that we can position a pair of surveys, being k time periods from one another to obtain one autocorrelation of the k-th order, at $(T-k)$ different points in time. The remaining $(n-2)$ surveys can be positioned in any of the $(T-2)$ remaining time periods. If we now divide this by the total number of different time-series possible, which is $\binom{T}{n}$, we obtain the result as stated.

interpret the sampling frequency f in this case as the participation request probability. In the next section we will argue that the randomization approach may not just seem attractive from an estimation point of view but also from the respondent's point of view.

- Insert Figure 3 about here -

Obviously, different survey designs may be combined or applied to different subsamples of individuals. One may combine repeated cross sections with complete panel data, see Nijman and Verbeek (1990) and Hirano *et al.* (2001). This approach facilitates testing for possible sampling biases, as will be discussed later. One may also decide to split the questionnaire and expose different groups of respondents to different subsets of questions in an efficient way. This approach is commonly known as matrix sampling. See, for example, Shoemaker (1973), Johnson (95-110), Raghunathan and Grizzle (1995) and the recent paper by Graham *et al.* (2006). In our paper, however, we restrict our attention to pure panel design aspects, exposing every panel member to the same questionnaire.

2.3 Modelling incomplete panels

As a result of using any of the above advanced panel designs, typically one is confronted with two types of missing data. First, parts of the panel are missing intentionally due to the design of the panel. Second, parts of the panel may be missing due to nonresponse.

Missingness by design

Observations that are missing intentionally due to the design of a panel are not related to the variables to be collected in any of the cases mentioned above. Therefore, following the typology of Rubin (1976), they are *missing completely at random* (MCAR). As a result, missingness by design can be ignored, in the sense that analyzing the incomplete panel will not bias our results. Still, especially in the case of randomization, the panel that is collected may seem rather intractable, as the individual time series are unequally spaced. Modern techniques are readily available however to deal with such data sets.

First of all, autoregressions which include only one lag order can be rewritten such that they match the unequally spaced data. This amounts to rewriting the model for the panel $y_{i,t}$ in terms of $y_{i,t}$ and $y_{i,t-d_t}$. Here $y_{i,t-d_t}$ denotes the previous observation, which is measured d_t time periods before $y_{i,t}$. To illustrate this, consider the basic first-order panel autoregressive model with fixed effects

$$y_{i,t} = \rho y_{i,t-1} + (\alpha_i + \varepsilon_{i,t}), \qquad \varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2), \tag{4}$$

where we treat the α_i's, for $i = 1, ..., N$, ρ and σ_ε^2 as fixed unknown parameters. This model can be rewritten as

$$\begin{aligned}
y_{i,t} &= \rho\big(\rho y_{i,t-2} + (\alpha_i + \varepsilon_{i,t-1})\big) + (\alpha_i + \varepsilon_{i,t}) \\
y_{i,t} &= \rho\Bigg(\rho\big(\rho y_{i,t-2} + (\alpha_i + \varepsilon_{i,t-2})\big) + (\alpha_i + \varepsilon_{i,t-1})\Bigg) + (\alpha_i + \varepsilon_{i,t}) \\
&\dots \\
y_{i,t} &= \rho^{d_{i,t}} y_{t-d_{i,t}} + \left(\sum_{\tau=0}^{d_{i,t}-1} \rho^\tau\right)(\alpha_i + \varepsilon_{i,t}) \\
y_{i,t} &= \rho^{d_{i,t}} y_{i,t-d_{i,t}} + \frac{1-\rho^{d_{i,t}}}{1-\rho}(\alpha_i + \varepsilon_{i,t}) \\
y_{i,t} &= \rho^{d_{i,t}} y_{1-d_{i,t}} + (\alpha_i^* + \varepsilon_{i,t}^*)
\end{aligned}$$

The factor $\rho^{d_{i,t}}$ is a finite duration adjustment of the geometric lag or Koyck model, see Ansari *et al.* (2006) and Van Diepen *et al.* (2006) for recent applications. Using this representation, the unequally spaced time series can be analyzed directly.

To estimate more complex dynamic models, one may use multiple imputation or likelihood-based techniques, see the excellent survey of Schafer and Graham (2002), and the books by Schafer (1997) and Little and Rubin (2002). In the first case, each missing value in the panel is substituted by a set of estimated or predicted values based on the available data. The resulting complete panels are then analyzed using conventional complete-data techniques and the results are combined. A widely used likelihood-based approach is maximum likelihood coupled with the EM algorithm of Dempster *et al.* (1977). In this case, the E-Step imputes the best predictors of the missing values, using current estimates of the parameters. It also calculates the adjustments to the estimated covariance matrix needed to allow for imputation of the missing values Little and Rubin (2002). A second likelihood-based approach is to reformulate the model as a state space model, where in the observation equation the incomplete panel is transformed into a latent complete panel. In the state equation the complete panel is then regressed on its past values and additional regressors. By alternating between Kalman filtering, smoothing recursions and maximum likelihood estimators, estimation is relatively straightforward, see Palma and Chan (1997), Shumway and Stoffer (1982) and Shumway and Stoffer (2006).

Missingness due to nonresponse

In contrast to missingness due to the panel design, missingness as a result of nonresponse can generally not be ignored, as the tendency to drop out is often systematically related to the variables of interest. In some cases, however, the distribution of missingness only depends on observed data and not on missing data. In these cases, the missing data are said to be *missing at random* (MAR), and multiple imputation or likelihood-based techniques can still be applied. If the variables of interest also depend on the unseen responses then the missing data are *missing not at random* (MNAR), see Schafer and Graham (2002) and the references cited there for a more thorough discussion.

In case one does not want to make the MAR assumption, one usually employs either selection models or pattern mixture models. The first amounts to incorporating the response decision explicitly in an econometric model, as a form of non-random selection (Hausman and Wise, 1979). The usual approach is to formulate a two-step model. In the first step, the response decision is modelled. Conditional on response, in the second step, the dependent variables of interest are modelled, see Amemiya (1984) and Heckman (1976, 1979) for details. Pattern mixture models stratify the sample by the pattern of missing data and then model differences in the distribution of the variables of interest over these patterns, see Little (1995), among others.

3 The Effects of Panel Design Characteristics

In this section we hypothesize the effects of panel design characteristics on response rates and response quality. In our empirical section below we will collect data to examine the empirical plausibility of these hypotheses.

The sampling frequency and the number of waves

We assume that participating in a survey is a social exchange, similar to responding to a mailing, for example, of a charity organization. Therefore, to hypothesize the effects of the sampling frequency and the number of waves of a panel survey we adopt the well-known Recency, Frequency, Monetary value (RFM) framework. The sampling frequency, and thus the (average) time between subsequent waves, is a recency variable. The higher the sampling frequency, the higher the (expected) recency of the last participation request. As, in general, recency has a negative effect on the probability of (high quality) response, we hypothesize

> H_1: The higher the sampling frequency, the lower (a) the response rate and (b) the response quality.

In the same vein, the number of waves can be seen as a frequency effect. The higher the number of waves, the higher the frequency in an RFM context, and hence the lower the probability of (high quality) response. We hypothesize

> H_2: As the number of participation requests increases, (a) the response rate and (b) the response quality decrease.

Randomized sampling

To increase response rates it has been argued that potential respondents should not be requested directly to join a panel. Instead, they should only be requested to complete a first survey and to grant permission to be contacted again for follow-up surveys. If respondents are persuaded initially to comply with this smaller request, subsequent requests may less likely be declined (Reingen and Kernan, 1977). This effect is referred to as the foot-in-the-door effect. As a result, overall response may be higher in the first waves. Nevertheless, as the survey evolves, respondents may quickly learn they are members of a panel with a particular sampling frequency and disconnect from the survey after all. In the case of randomized sampling, learning the (average) sampling frequency of the panel may take long, however. As a result, in this case, we expect respondents to participate longer.

An often reported source of panel conditioning bias is negativity bias. This is the tendency of individuals to be more negative in their judgements if they expect to be evaluated, see, for example, Ofir and Simonson (2001). If we survey at random time intervals, panel members are less likely to develop expectations as to when they will be surveyed again. As a result, we expect less panel conditioning bias in this case, which may contribute to a higher quality of response.

Thus we hypothesize the following

> H_3: Randomized sampling increases (a) the response rate and (b) the response quality.

4 Research Design

Setup

In order to be able to study the effects of panel design characteristics on response rates and response quality, we choose a flexible panel design. Our design is characterized by the collective answers to the following questions:

(i) How many times n should we request a panel member to be surveyed at the maximum?

(ii) Within which time span T^* should this occur?

(iii) Given the number of requests and the time span chosen, at which dates should we survey each panel member? In particular, should we apply a time sampling or a randomized sampling approach?

Note that when the average sampling frequency $f = n/T^*$ is equal to one, both time sampling and randomized sampling reduce to continuous sampling. We therefore focus only on time sampling and randomized sampling, and consider different values of f.

We perform two field studies. In the first we follow a foot-in-the-door approach by not informing our subjects about the panel design to be employed. We request each potential panel member to fill in one questionnaire. Additionally, we ask whether they agree to be contacted in the future for further research. To each respondent who reacts positively, we assign a sampling strategy (time sampling or randomized sampling) and a sampling frequency f. We do not fix the total number of survey requests n nor the number of time periods T^*. Obviously, these increase as the survey evolves. We expect our panel members to learn about the design through experience, and we assume that they base their further participation decision on this experience.

In the second study, we first ask our potential panel members to complete the same questionnaire as before. We then explain again that we would like to contact our respondents in the future for further research. In order for us to learn what they find an acceptable way to do this, we sequentially present ten randomly generated sampling strategies. Here we do not only randomize over the sampling strategy and the sampling frequency f, but also over the number of requests n. The respondents are requested to indicate which of the sampling strategies are acceptable to them, if any. We then select and use one of their accepted strategies at random. In personal following up mailings the respondents are informed about the design chosen. In this second study, the respondents thus have full information about the panel design used and we assume that possible effects through learning are eliminated.

Model and estimation

To study the effects of panel design characteristics on the response decision and subsequent responses separately, it is convenient to use a selection model. We summarize the responses as a result of our participation requests in the response indicator matrix R, where its elements, $r_{i,t}$, for $i = 1, ..., N$ and $t = 1, ..., T$, register:

$$r_{i,t} = \begin{cases} 1 & \text{Member } i \text{ is requested to participate at time } t \text{ and did so} \\ 0 & \text{Member } i \text{ is requested to participate at time } t \text{ but did not so} \\ \cdot & \text{Member } i \text{ is not requested to participate at time } t \end{cases} \quad (5)$$

For those values of i and t for which $r_{i,t}$ equals 1, we additionally observe our questionnaire data, summarized in the variable $\boldsymbol{Y} = (Y_1, ..., Y_Z)$. For ease of presentation, we denote any panel Y_z, $z = 1, ..., Z$ in $\boldsymbol{Y}$, by Y, with elements $y_{i,t}$. To allow for the possibility that respondents resume their participation after one or more waves of nonresponse, our model below seeks to explain wave response rather than attrition. To explain the measurement process R as well as the questionnaire data $\boldsymbol{Y}$, our workhorse model will be a dynamic panel Tobit-II model, which consists of a Probit model for R being 0 or 1 and a standard regression model for each censored panel Y. It can be written as

$$\begin{aligned} r_{i,t} = 0 \qquad & \text{if } y^*_{i,t} = f(\boldsymbol{X}, \boldsymbol{\theta}_1, u_{1,i,t}) \leq 0 \qquad (6) \\ r_{i,t} = 1 \quad \text{and} \quad y_{i,t} = f(\boldsymbol{X}, \boldsymbol{\theta}_2, u_{2,i,t}) \qquad & \text{if } y^*_{i,t} = f(\boldsymbol{X}, \boldsymbol{\theta}_1, u_{1,i,t}) > 0 \end{aligned}$$

where $\boldsymbol{X}$ represents all regressors, $u_{s,i,t}$ are the error terms and $\boldsymbol{\theta}_s$ the parameter vectors, for $s = 1, 2$. The Tobit-II model allows $\boldsymbol{X}$ to have a different effect on R and on Y. The model parameters can conveniently be estimated by maximum likelihood.

5 Empirical Results

Our two field studies have been conducted among students at Erasmus University Rotterdam over the period September 2004 to March 2007. Our desired frequency of data collection f^d was weekly, and the sampling frequencies ranged from bimonthly to weekly ($f = 0.125$ to 1). The experiments were conducted online through an interactive website. All correspondence, including personalized participation requests, was generated automatically and sent by e-mail. The general task we asked the respondents to complete was a test for their knowledge of recent news events. In the test we presented 20 news headlines, among which 10 were literally quoted from newspapers published in the previous week and 10 were slightly altered. The respondents had to indicate which 10 headlines were indeed literally quoted and which were not. The headlines were selected from five different news categories (domestic news, foreign news, politics, economics, sports & culture) and were carefully pretested to ascertain that they were equally familiar in each week. On average, respondents completed the test in three minutes.

Additional to the news test, we surveyed the respondents about their attitude towards the experiment, with the aim to measure possible signs of irritation and panel conditioning bias. For this purpose, at each wave we posed three statements per construct, randomly selected out of ten, which had to be rated on a seven-point Likert scale. As a consequence, the overall size of the questionnaire remained constant, but the statements were different. We summarized the scores in two variables, which measure self-reported irritation and self-reported panel conditioning bias[3]. As an incentive to continue participation, we raffled out a $25 gift voucher at each wave of data collection among the respondents of that particular wave.

[3]For this purpose, we performed a principal components analysis on the scores obtained on the ten different statement which supposedly measure the same construct. Next, we selected the first component. This component explains over 75% of the variation in the data in both cases.

To complete the model in (6) for our field studies, we specify the error terms $u_{s,i,t}$ in (6) as

$$u_{s,i,t} = (\alpha_{s,i} + \tau_{s,t} + \varepsilon_{s,i,t}), \quad \alpha_{s,i} \sim N(0, \sigma^2_{\alpha_s}), \quad \varepsilon_{s,i,t} \sim N(0, \sigma^2_{\varepsilon_s}) \quad \text{for } s = 1, 2 \tag{7}$$

where $\alpha_{s,i}$ are individual-specific random effects to account for unobserved heterogeneity among respondents. The parameters $\tau_{s,t}$ are time-specific fixed effects to account for possible variation in response due to the particular week of data collection and the particular news test. For example, these parameters may capture lower response rates during examination periods or higher news scores due to a relatively easy test. The random variables $\varepsilon_{s,i,t}$ are idiosyncratic shocks.

Next we have to specify the regression part $f(\boldsymbol{X}, \boldsymbol{\theta}_s, u_{s,i,t})$, for $s = 1, 2$. We assume that our respondents learn about the design parameter n through the number of times they have been requested to be surveyed previously, which is $n - 1$. The sampling frequency f may be learned through the number of weeks since the previous request, which is $1/f$ in expectation. Therefore we include these two variables in our model, together with a dummy variable indicating whether time sampling or randomized sampling has been applied. We summarize the three design variables in the vector $\boldsymbol{P}$. Additional to the design parameters, we include an AR(1) term which captures possible dynamics in response, through a finite duration adjustment as discussed in Section 2.3. Finally we include several individual specific regressors such as demographics, self-reported irritation and panel conditioning, to be denoted by $\boldsymbol{V}$. In sum, the regression equations can now be written as:

$$f(\boldsymbol{x}_{i,t}, \boldsymbol{\theta}_s, u_{s,i,t}) = \rho_j^{d_{i,t}} \boldsymbol{x}_{i,t-d_{i,t}} + \boldsymbol{\zeta} \boldsymbol{p}_{i,t} + \boldsymbol{\beta} \boldsymbol{v}_{i,t} + (\alpha_{s,i} + \tau_{s,t} + \varepsilon_{s,i,t}) \quad \text{for } s = 1, 2 \tag{8}$$

Study 1: No information provided about the design

Among the 623 students we contacted, 290 agreed to participate in the first study. The total sample period of the study was 26 weeks. In Table 1 we summarize the response rates, where we classify the respondents to four groups according to the sampling frequency assigned to them. It is already apparent from this table that the sampling frequency negatively influences response. For example, after six waves of data collection, the response rate among those respondents who have been surveyed monthly to biweekly is 23% (0.68 – 0.45) higher as compared to the group that has been surveyed three times per month to weekly.

- Insert Table 1 about here -

Table 2 shows the estimation results of the dynamic panel Tobit-II model. In the first panel the results of the Probit part of the model are presented. The second panel presents the results of the regression part. In this part we consider four different explanatory variables Y, and as a consequence we estimated our model four times. The estimates within the Probit part are the same across all variables Y.

Inspecting the results for the Probit part, we find that the number of weeks between subsequent participation requests has a positive effect on the probability of response. Hence the higher the sampling frequency f, the lower the response rate, as hypothesized in H_{1a}. Second, the number of waves n has the expected significant negative effect. Perhaps more surprising, the dummy variable which distinguishes between randomized sampling (1) and time sampling (0) is significantly positive. This indicates that it is beneficial to request panel members to be surveyed at random points in time rather than at fixed time intervals. We thus find support for both hypotheses H_{2a} and H_{3a}. As can be seen from the hitrate, the model predicts 70% of the responses correctly. Moreover, it does not seem to seriously overpredict either response or nonresponse.

Next, we inspect the results of the regression part with the purpose to test our hypotheses regarding the effects of panel design characteristics on response quality. The first two variables we consider here are the individuals' scores on the news test and the time needed to complete this test. There is a positive effect of the number of weeks since the previous request both on the score and on the time needed to complete the test. This implies that as the time between subsequent waves gets shorter, respondents tend to score lower on the test and spend less time to complete it, which is a clear signal of panel conditioning bias. The second two variables are our measures for the stated level of irritation of an individual and the extent to which he or she feels his or her (reporting) behavior has changed due to participation in the panel study. As these variables are self-reported, they have to be interpreted with a bit more care. First of all, although there is a weak effect on irritation, the sampling frequency f does seem to influence respondents' stated level of panel conditioning bias. We did find an effect on the news test score however. This suggests that, even though respondents are biased in their (reporting) behavior due to the sampling frequency, they do not perceive this effect. In contrast, the number of waves n does seem to drive stated panel conditioning bias. Randomization seems to lower both irritation and panel conditioning bias. These two variables did not significantly influence the test performance however. In sum, we find support for hypothesis H_{1b} in the revealed data, but support for H_{2b} and H_{3b} only in the stated data.

Finally, we discuss the results for the additional regressors and demographics. The highly significant estimates of ρ suggest there is a strong relationship between current and past responses, which cannot be ignored. The effects of respondents' news consumption levels on their performance on the news test are positive, as expected. Looking at the effects of demographics, we observe that women tend to be slightly more loyal to the experiment, as can be seen from the negative effect of gender on response and irritation. Also they tend to score slightly higher on the news test.

- Insert Table 2 about here -

Study 2: Full information provided about the design

For the second study, we found 148 students willing to participate out of a sample of 292. In the first part of the experiment, we presented ten panel designs to each respondent and requested them to state which of these designs would be acceptable to them. Among the accepted designs, we found the median of the maximum number of waves to be 5.5, and the median of the maximum frequency to be biweekly, which is high considered the actual participation to Study 1. There were 22 respondents who did not accept any design. To study the effects of panel design parameters on stated acceptance, we first estimate the acceptance decision using the Probit part of our Tobit-II model. The estimation results are shown in Table 3. Again, the effects of the sampling frequency and the number of waves are prominent. It is interesting to see, however, that a respondent's probability of acceptance is not influenced by the choice for a randomized or a time sampling strategy. Respondents state to be indifferent between the two strategies.

- Insert Table 3 about here -

Next, we check whether our respondents' stated participation matches their revealed participation, by indeed surveying them according to one of their accepted designs. The results are shown in Table 4. Clearly, there is a big distinction between respondents' promised participation and actual participation as there still is considerable nonresponse in this experiment ($1 - 0.755 = 24.5\%$). Finally, we examine whether there are differences in behavior between the informed respondents of Study 2 and the uninformed respondents of Study 1. The informed respondents' probability of response, performance on the news test, stated level of irritation and panel conditioning bias still seem to be influenced by the design parameters n and f in the same manner as they did in Study 1. This indicates that providing information about the design of the experiment does not alleviate response or panel conditioning bias. In contrast, the influence of randomized sampling on any of our dependent variables disappeared in the informed case. This supports our conjecture that randomization has an effect on response rates and response quality as a result of learning about the design. Response rates are improved as panel members actually have low awareness of being such a member when they are uninformed about the panel design. Response quality gets improved as respondents are less likely develop expectations as to when they will be surveyed again.

We summarize the testing results for our hypotheses in Table 5, where we distinguish between the informed and uninformed case.

- Insert Tables 4 and 5 about here -

Panel calibration

One of the advances of our approach is that it can be used to calibrate the design of a new panel survey. To illustrate this, suppose that we conducted Study 1 as a pilot study with the aim to calibrate a continuous monitor for students' knowledge of recent news events. Using our parameter estimates, we can now simulate data from the model for various values of the design parameters $\boldsymbol{P}$. These data can be used to compute the expected response rates under different scenario's. In Figure 4 we plot the response curves, which are the expected response rates as a function of the sampling frequency f. Windows (a)-(c) show the curves for the 2nd, 6th and 12th wave, respectively, where we distinguish between the expected response rates obtained using a time sampling and a randomized sampling strategy.

Since the sampling frequency has an effect on a student's news score, which is undesirable, but the sampling strategy and the number of waves have not, it is advisable to adopt randomized sampling, to choose a high number of waves, but with a low sampling frequency. Suppose now that our budget restricts us to have at least 50% expected response in the last wave. As can be read from the graph, in this case, if we want to collect as much as 12 waves of data, we should set the sampling frequency f equal to 0.47, which is close to biweekly, or lower. In case time sampling is to be preferred, then the frequency should be equal to 0.35 or lower.

- Insert Figure 4 about here -

6 Conclusions

Response rates of panel surveys may be low and responses given by panel members may depend highly on the panel design chosen. Therefore researchers must choose their design carefully. To facilitate this choice, in this paper we studied the effects of panel design characteristics on response rates and response quality. We hypothesized that response rates and response quality decrease as the sampling frequency and the number of waves of a panel increase. Further, we proposed a new sampling strategy, labelled randomized sampling, where at each wave only a random subsample of the panel is selected to be surveyed. We hypothesized that, as compared to the often applied time sampling strategy, randomized sampling improves response rates and response quality. To test our hypotheses, we proposed a two-step selection model, where in the first step we explained the response decision and in the second step we explained subsequent responses, conditional on response in the first step. Two empirical studies indeed confirmed the above hypotheses. We did find, however, that the effect of randomized sampling is only significant in a setting where panel members are not informed about the dates they will be surveyed, because in this case respondents are less likely to develop expectations as to when they will be surveyed again.

The main advantage of our approach is that it can be used to calibrate new panels. Careful calibration should lead to an efficient panel design where expected response is maximized and threats to response quality due to the design are suppressed. This in turn reduces the survey costs, since fewer respondents need to be acquired and higher quality data is collected. The methodology may also be used to capture residual response bias and panel conditioning bias, to the extent that these biases are explained by our model.

Further research could focus on explaining attrition rather than wave response, since attrition is a bigger concern than wave nonresponse in many cases. Finally, it would be interesting to understand more about the learning process of panel members over time.

References

Adams, L. M. and G. Darwin (1982), Solving the Quandary Between Questionnaire Length and Response Rate in Educational Research, *Research in Higher Education*, **17**, 231–240.

Allison, D. B., R. L. Allison, M. S. Faith, F. Paultre, and F. X. Pi-Sunyer (1997), Power and Money: Designing Statistically Powerful Studies While Minimizing Financial Costs, *Psychological Methods*, **2**, 20–33.

Amemiya, T. (1984), Tobit Models: A Survey, *Journal of Econometrics*, **24**, 3–61.

Ansari, A., C. Mela, and S. Nesslin (2006), Customer Channel Migration, *Conditionally accepted, Journal of Marketing Research*.

Atkinson, A. C. and A. N. Donev (1992), *Optimum Experimental Designs*, Clarendon Press, Oxford Science Publications, Oxford, England.

Collado, M. D. (1997), Estimating Dynamic Models from Time Series of Independent Cross-sections, *Journal of Econometrics*, **82**, 37–62.

Couper, M. P., J. Rothgeb, J. Lessler, E. Martin, J. Martin, and E. Singer (2004), *Methods for Testing and Evaluating Survey Questionnaires*, Wiley, New York.

Deaton, A. (1985), Panel Data from Time Series of Cross-sections, *Journal of Econometrics*, **30**, 109–126.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), ML Estimation from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society*, **B39**, 1–38.

Deutskens, E., K. de Ruyter, M. Wetzels, and P. Oosterveld (2004), Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study, *Marketing Letters*, **15**, 21–36.

De Stavola, B. (1986), Sampling Designs for Short Panel Data, *Econometrica*, **54**, 415–424.

Dholakia, U. M. and V. G. Morwitz (2002), The Scope and Persistence of Mere-measurement Effects: Evidence from a Field-study of Customer Satisfaction, *Journal of Consumer Research*, **29**, 159–167.

Dillman, D. A. (2000), *Mail and Internet Surveys: The Tailored Design Method*, second edition edn., John Wiley & Sons, New York.

Graham, J. W., B. J. Taylor, A. E. Olchowski, and P. E. Cumsille (2006), Planned Missing Data Designs in Psychological Research, *Psychological Methods*, **11**, 323–343.

Hansen, R. A. (1980), A Self-Perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent behavior, *Journal of Marketing Research*, **17**, 77–83.

Hausman, J. A. and D. A. Wise (1979), Attrition Bias in Experimental and Panel Data: The Gary Income Maintainance Experience, *Econometrica*, **47**, 455–473.

Heckman, J. J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement*, **5**, 475–492.

Heckman, J. J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, **47**, 153–161.

Hirano, K., G. Imbens, G. Ridder, and D. B. Rubin (2001), Combining Panel Data Sets with Attrition and Refreshment Samples, *Econometrica*, **69**, 1645–1659.

Johnson, E. G. (95-110), The design of the National Assessment of Educational Progress, *Journal of Educational Measurement*, **29**.

Kish, L. and I. Hess (1959), A “Replacement” Procedure for Reducing the Bias of Nonresponse, *The American Statistician*, **13-4**, 17–19.

Little, R. J. A. and R. B. Rubin (2002), *Statistical Analysis with Missing Data*, second edn., John Wiley & Sons.

Little, R. J. A. (1995), Modeling the Drop-out Mechanism in Repeated-Measures Studies, *Journal of the American Statistical Association*, **90**, 1112–1121.

McClelland, G. H. (1997), Optimal Design in Psycological Research, *Psychological Methods*, **2**, 3–19.

Moffitt, R. (1993), Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-sections, *Journal of Econometrics*, **59**, 99–123.

Morwitz, V. G. and G. J. Fitzsimons (2004), The Mere-Measurement Effect: Why Does Measuring Intentions Change Actual Behavior?, *Journal of Consumer Psychology*, **14**, 64–74.

Nijman, T., M. Verbeek, and A. van Soest (1991), The Efficiency of Rotating-panel Designs in an Analysis-of-variance Model, *Journal of Econometrics*, **49**, 373–399.

Nijman, T. and M. Verbeek (1990), Estimation of Time Dependent Parameters in Linear Models using Cross-sections, Panels or Both, *Journal of Econometrics*, **46**, 333–346.

Ofir, C. and I. Simonson (2001), In Search of Negative Customer Feedback: The Effect of Expecting to Evaluate on Satisfaction Evaluations, *Journal of Marketing Research*, **38**, 170–182.

Ouwens, M. J. N. M., F. E. S. Tan, and M. P. F. Berger (2002), Maximin D-Optimal Designs for Longitudinal Mixed Effects Models, *Biometrics*, **58**, 735–741.

Palma, W. and N. H. Chan (1997), Estimation and Forecasting of Long-memory Time Series with Missing Values, *Journal of Forecasting*, **16**, 395–410.

Patterson, H. D. (1950), Sampling on Successive Occasions with Partial Replacement of Units, *Journal of the Royal Statistical Society Series B*, **49**, 241–255.

Raghunathan, T. and J. Grizzle (1995), A Split Questionnaire Survey Design, *Journal of the American Statistical Association*, **90**, 54–63.

Reingen, P. H. and J. B. Kernan (1977), Compliance with an Interview Request: A Foot-in-the-Door, Self-Perception Interpretation, *Journal of Marketing Research*, **14**, 365–369.

Rubin, D. B. (1976), Inference and Missing Data, *Biometrika*, **63**, 581–592.

Schafer, J. L. and J. W. Graham (2002), Missing Data: Our view of the State of the Art, *Psychological Methods*, **7**, 147–177.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York.

Shoemaker, D. M. (1973), *Priciples and Procedures of Multiple Matrix Sampling*, Ballinger Publishing, Cambridge.

Shumway, R. H. and D. S. Stoffer (1982), An Approach to Time Series Smoothing and Forecasting using the EM Algorithm, *Journal of Time Series Analysis*, **3**, 253–264.

Shumway, R. H. and D. S. Stoffer (2006), *Time Series Analysis and Its Applications*, second edn., Springer, New York.

Simmons, C. J., B. A. Bickart, and J. G. Lynch, Jr. (1993), Capturing and Creating Public Opinion in Survey Research, *Journal of Consumer Research*, **20**, 316–329.

Trivellato, U. (1999), Issues in the Design and Analysis of Panel Studies: A Cursory Review, *Quality & Quantity*, **33**, 339–352.

Van Diepen, M., B. Donkers, and P. H. Franses (2006), Dynamic and Competitive Effects of Direct Mailings, *ERIM Report Series Research in Management*, ERS–2006–050–MKT.

Verbeek, M. J. C. M. (1991), *The Design of Panel Surveys and the Treatment of Missing Observations*, Ph.D. thesis, Tilburg Univesity.

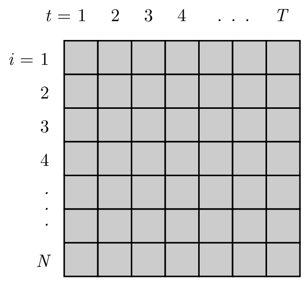
Verbeek, M. and T. Nijman (1993), Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-sections, *Journal of Econometrics*, **59**, 125–136.

Verbeek, M. and F. Vella (2005), Estimating Dynamic Models from Repeated Cross-sections, *Journal of Econometrics*, **127**, 83–102.

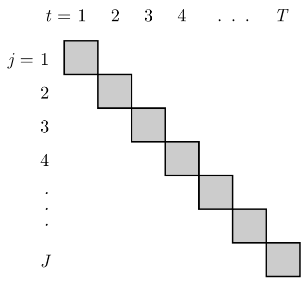
Yu, J. and H. Cooper (1983), A Quantitative Review of Research Design Effects on Response Rates to Questionaires, *Journal of Marketing Research*, **20**, 36–44.

A Tables and Figures

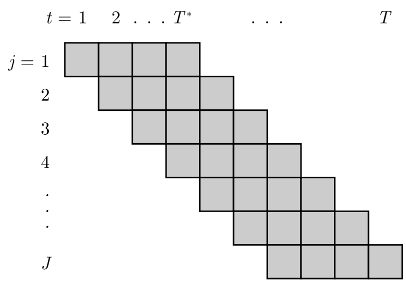
Figure 1: Alternative sampling strategies across individuals



(a) Complete panel



(b) Repeated cross-sections



(c) Rotating panel

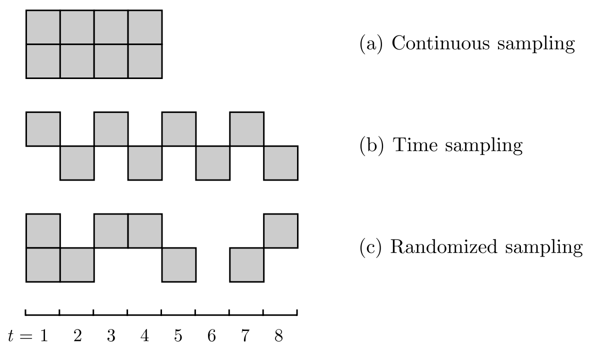
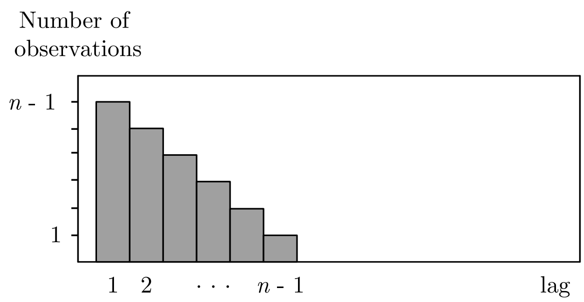
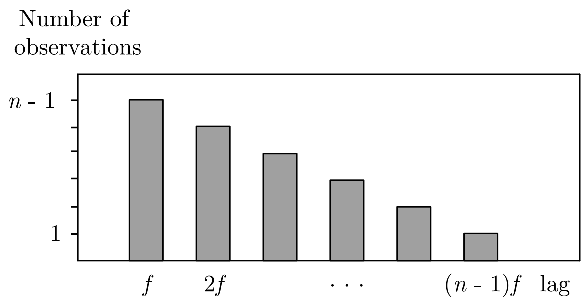


Figure 2: Alternative sampling strategies over time

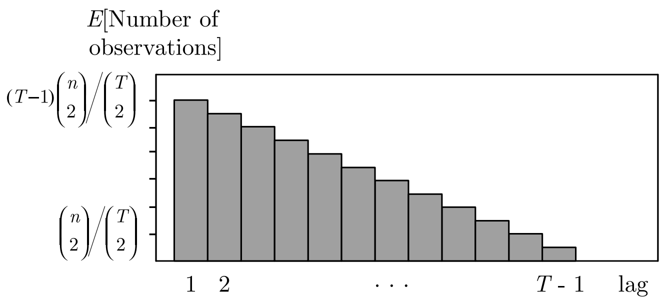
Figure 3: (Expected) number of autocorrelations observed under different sampling strategies



(a) Continuous sampling

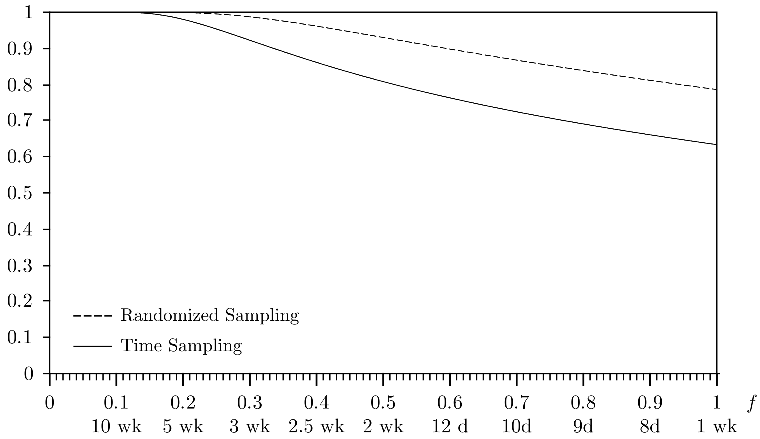


(b) Time sampling

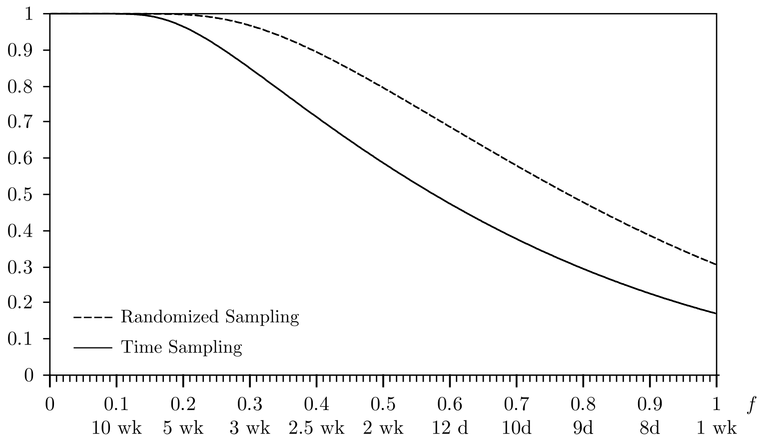


(c) Randomized sampling

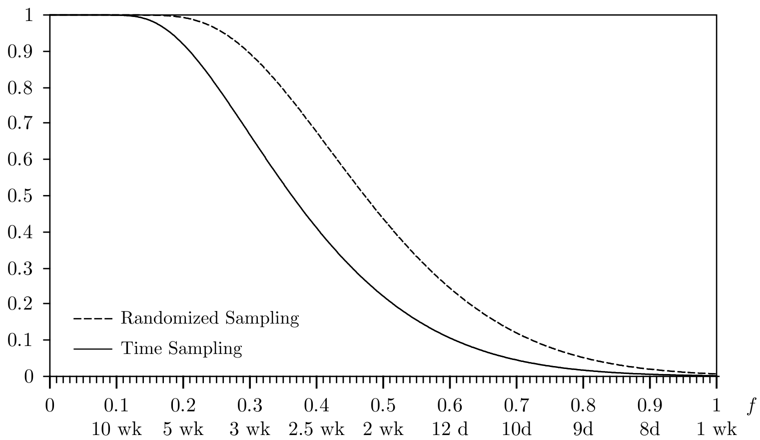
Figure 4: Response curves



(a) Expected response rates of the 2nd wave



(b) Expected response rates of the 6th wave



(c) Expected response rates of the 12th wave

Table 1: Response rates

Frequency	Wave				
	2	3	4	6	12
Three times per month - weekly	0.67	0.69	0.59	0.45	0.06
$(0.75 < f \leq 1.00)$	(0.50)	(0.46)	(0.49)	(0.50)	(0.24)
Biweekly - three times per month	0.80	0.75	0.65	0.59	0.22
$(0.50 < f \leq 0.75)$	(0.40)	(0.44)	(0.48)	(0.50)	(0.31)
Monthly - biweekly	0.83	0.82	0.76	0.68	–
$(0.25 < f \leq 0.50)$	(0.37)	(0.38)	(0.43)	(0.26)	
Once every eight weeks - monthly	0.88	0.84	0.77	–	–
$(0.125 \leq f \leq 0.25)$	(0.33)	(0.39)	(0.37)		

Sample standard deviations are in parentheses. The response rates for the bottom right cases are missing, due to the studies' time window of 26 weeks.

Table 2: Results of the panel Tobit-II model for Study 1

Variable	Probit part (R)		Regression parts ($\boldsymbol{Y}$) News score (0 - 20)		Time needed (in mins.)		Irritation (stated; -1.0 - 1.0)		Panel conditioning (stated; -1.0 - 1.0)	
Design parameters										
No. of weeks since previous request ($1/f$)	0.322**	(0.143)	0.249**	(0.095)	0.670**	(0.305)	0.017*	(0.010)	0.012	(0.014)
No. of times requested before ($n-1$)	-0.728***	(0.054)	-0.014	(0.044)	0.001	(0.207)	0.001	(0.003)	0.011**	(0.005)
Randomized sampling	0.585***	(0.201)	0.567	(0.409)	1.070	(0.978)	-0.102***	(0.038)	-0.098*	(0.051)
Additional regressors										
Dynamics (ρ)	0.939***	(0.290)	0.397***	(0.036)	0.465***	(0.048)	0.547***	(0.041)	0.682***	(0.040)
No. of hours spent consuming news			0.139**	(0.065)	-0.091	(0.062)	-0.003	(0.002)	-0.004	(0.003)
Subscribing to a newspaper			0.459*	(0.259)	-0.524	(0.616)	-0.044	(0.044)	-0.052	(0.032)
Demographics										
Age	0.046	(0.034)	-0.003	(0.037)	-0.074	(0.085)	-0.001	(0.003)	0.008*	(0.004)
Gender (0: female; 1: male)	-0.389*	(0.237)	0.688*	(0.411)	-0.154	(0.961)	-0.120**	(0.059)	-0.110**	(0.051)
Additional model parameters										
Intercept	-5.557***	(0.908)	6.308***	(1.115)	5.190*	(3.155)	-0.232**	(0.099)	-0.014	(0.130)
Cross-section random st. dev.	0.362		2.306		5.385		0.163		0.233	
Idiosyncratic random st. dev.	1.000[1]		1.466		9.594		0.161		0.190	
Mean of the dependent	0.683		10.223		3.255		-0.646		-0.126	
McFadden / Adjusted R^2	0.555		0.583		0.379		0.756		0.768	
Hitrate	0.702									
Prop. of correctly predicted response	0.730									
Prop. of correctly predicted nonresponse	0.642									

*** Significant at the 1% level, ** at the 5% level, * at the 10% level. White heteroskedasticity-consistent standard errors are in parentheses. [1] Standardized for identification purposes. The estimates of the period fixed effects are not displayed for ease of presentation.

Table 3: Results of the panel Probit model for Study 2

Variable		
Design parameters		
No. of weeks since previous request ($1/f$)	0.135***	(0.016)
No. of times requested before ($n-1$)	−0.161***	(0.011)
Randomized sampling	0.162	(0.117)
Demographics		
Age	−0.024	(0.035)
Gender (0: female; 1: male)	0.131	(0.500)
Additional model parameters		
Intercept	0.777	(0.784)
Cross-section random st. dev.	0.750	
Idiosyncratic random st. dev.	1.000[1]	
Mean of the dependent	0.362	
McFadden / Adjusted R^2	0.569	
Hitrate	0.728	
Prop. of correctly predicted acceptance	0.626	
Prop. of correctly predicted rejection	0.785	

*** Significant at the 1% level, ** at the 5% level, * at the 10% level. White heteroskedasticity-consistent standard errors are in parentheses. [1] Standardized for identification purposes. The estimates of the period fixed effects are not displayed for ease of presentation.

Table 4: Results of the panel Tobit-II model for Study 2

Variable	Probit part (R)		Regression parts ($\boldsymbol{Y}$)							
			News score (0 - 20)		Time needed (in mins.)		Irritation (stated; -1.0 - 1.0)		Panel conditioning (stated; -1.0 - 1.0)	
Design parameters										
No. of weeks since previous request ($1/f$)	0.282***	(0.072)	0.329**	(0.167)	0.524**	(0.272)	0.020**	(0.011)	0.008	(0.031)
No. of times requested before ($n-1$)	-0.583***	(0.101)	-0.071	(0.062)	-0.167	(0.304)	0.003	(0.005)	0.013**	(0.007)
Randomized sampling	0.193	(0.197)	-0.777	(0.735)	-0.698	(1.910)	-0.022	(0.068)	-0.092	(0.097)
Additional regressors										
Dynamics (ρ)	0.931***	(0.296)	0.471***	(0.055)	0.698***	(0.063)	0.484***	(0.064)	0.575***	(0.063)
No. of hours spent consuming news			0.143***	(0.042)	-0.106	(0.095)	-0.006	(0.036)	-0.003	(0.005)
Subscribing to a newspaper			0.300	(0.639)	-0.102	(1.764)	0.172	(0.193)	0.173	(0.182)
Demographics										
Age	0.016	(0.029)	-0.101	(0.073)	-0.243	(0.197)	-0.018*	(0.015)	0.004	(0.010)
Gender (0: female; 1: male)	-0.287	(0.199)	1.118*	(0.599)	1.881	(1.651)	-0.113*	(0.060)	-0.136*	(0.085)
Additional model parameters										
Intercept	-4.712***	(1.066)	8.955***	(1.811)	8.353	(5.551)	-0.095	(0.168)	-0.154	(0.237)
Cross-section random st. dev.	0.276		2.214		4.502		0.166		0.233	
Idiosyncratic random st. dev.	1.000[1]		1.401		5.621		0.184		0.264	
Mean of the dependent	0.755		9.746		2.879		-0.611		-0.143	
McFadden / Adjusted R^2	0.639		0.594		0.555		0.741		0.780	
Hitrate	0.690									
Prop. of correctly predicted response	0.720									
Prop. of correctly predicted nonresponse	0.598									

*** Significant at the 1% level, ** at the 5% level, * at the 10% level. White heteroskedasticity-consistent standard errors are in parentheses. [1] Standardized for identification purposes. The estimates of the period fixed effects are not displayed for ease of presentation.

Table 5: Summary of hypothesis-testing results

Effect	Hypothesis	Information about the panel design	
		Not provided	Provided
The sampling frequency on response rates	H_{1a}	Supported	Supported
The sampling frequency on response quality	H_{1b}	Supported	Supported
The number of participation requests on response rates	H_{2a}	Supported	Supported
The number of participation requests on response quality	H_{2b}	Weak support[1]	Weak support[1]
Randomized sampling on response rates	H_{3a}	Supported	Not supported
Randomized sampling on response quality	H_{3b}	Weak support[1]	Not supported

[1] Only supported by stated preference data.