Working Papers

FEDERAL RESERVE BANK OF BOSTON™

No. 13-15

Modeling Anchoring Effects in Sequential Likert Scale Questions

Marcin Hitczenko

Abstract:

Surveys in many different research fields rely on sequences of Likert scale questions to assess individuals' general attitudes toward a set of related topics. Most analyses of responses to such a series do not take into account the potential measurement error introduced by the context effect we dub "sequential anchoring, " which occurs when the rating for one question influences the rating given to the following question by favoring similar ratings. The presence of sequential anchoring can cause systematic bias in the study of relative ratings. We develop a latent-variable framework for question responses that capitalizes on different question orderings in the survey to identify the presence of sequential anchoring. We propose a parameter estimation algorithm and run simulations to test its effectiveness for different data-generating processes, sample sizes, and orderings. Finally, the model is applied to data in which eight payment instruments are rated on a five-point scale for each of six payment characteristics in the 2012 Survey of Consumer Payment Choice. We find consistent evidence of sequential anchoring, resulting in sizable differences in properties of relative ratings for certain instruments.

Keywords: survey bias, latent variable models, EM algorithm, SCPC

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.

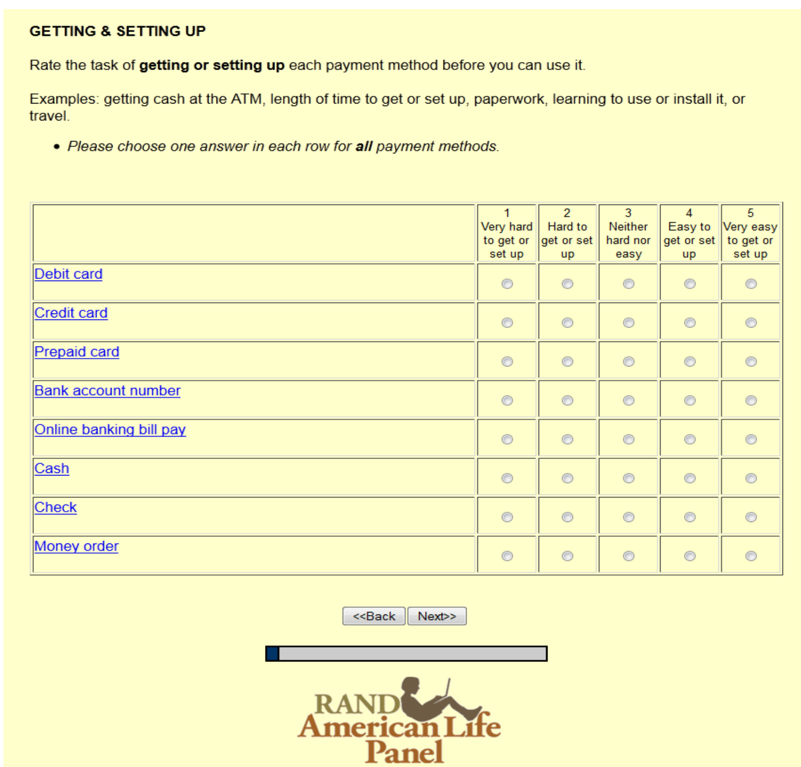
I would like to thank Scott Schuh and the CPRC for their guidance, feedback, and support of this work.

This version: December 9, 2013

CONSUMER PAYMENTS Research Center

1 Introduction

Prevailing attitudes among the individuals of a population are of interest to researchers in a variety of fields, from economics to psychology to public opinion research. Examples include a consumer's opinion of a product or a potential voter's stance on government policy. Such quantifications of individual beliefs are most often measured in surveys through Likert-scale questions (Likert 1932), which ask respondents to map their opinions onto a discrete set of polytomous, and usually ordinal, responses. For example, one might be asked to assess the degree to which one agrees with a statement on a scale of five response options ranging from "disagree strongly" to "agree strongly." It is often the case that a survey asks respondents to provide Likert-scale responses to a consecutive series of related questions. As an example, Figure 1, taken from the 2012 Survey of Consumer Payment Choice, prompts each individual to rate eight payment instruments on their ease of set-up. A continuous block of Likert-scale responses provides insight into attitudes relating to one question within the context of the related questions.



GETTING & SETTING UP

Rate the task of **getting or setting up** each payment method before you can use it.

Examples: getting cash at the ATM, length of time to get or set up, paperwork, learning to use or install it, or travel.

- *Please choose one answer in each row for **all** payment methods.*

	1 Very hard to get or set up	2 Hard to get or set up	3 Neither hard nor easy	4 Easy to get or set up	5 Very easy to get or set up
Debit card					
Credit card					
Prepaid card					
Bank account number					
Online banking bill pay					
Cash					
Check					
Money order					

<<Back Next>>

RAND American Life Panel

Figure 1: A screenshot of from the 2012 SCPC asking the respondent to rate the ease of setting up for each of the eight different payment instruments.

Cognitive science research has produced an impressive body of work showing that virtually all aspects of a Likert-scale question influence the survey responses. Wording choices in the questions (Schuman and

Presser 1996), response options (Friedman, Herskovitz, and Pollack 1994; Schwarz et al. 1991), and the number of ratings made available to the respondent (Dawes 2008) have all been shown to be important factors. In addition, survey methodologists have long been aware of "context effects," or survey-response biases that result from the interaction of the survey instrument with a respondent's cognitive process (see Daamen and de Bie (1992); Mason, Carlson, and Tourangeau (1994); Schuman (1992); Schwarz and Hippler (1995); Tourangeau, Rasinski, and Bradburn (1991) for examples and discussions). As evidenced by the range of topics and applications in the referenced papers and chapters, context effects take on many forms. One type of context effect, generally referred to as an "anchoring effect," occurs when initial information is subsequently used by an individual, usually subconsciously, to inform judgments. Changes in the initial information tend to change the response outcomes.

In this paper, we focus on a particular form of anchoring effect specific to a sequence of Likert-scale questions. The effect, which we dub "sequential anchoring," manifests itself by having the response to one question serve as an anchor for the response to the subsequent question. Tourangeau, Couper, and Conrad (2004) found evidence of such a phenomenon in the context of binary assessments (such as, expensive or not expensive) of a relatively unfamiliar item among a list of related, but more familiar, items. In a majority of cases, respondents tended to assimilate the response toward the rating of the surrounding items. Under the confirmatory hypothesis testing theory (Chapman and Johnson 1999; Strack and Mussweiler 1997), anchor values often serve as plausible responses and thus induce a search for similarities in consequent responses. Sequential anchoring may also result from a conscious decision to respond as quickly as possible and thus minimize the variability of the responses. As a result, we posit that the sequential anchoring effect skews responses to tend to be more similar to previous responses. For example, a response of "agree strongly" for one question makes it more likely the next response will be on that side of the spectrum than if the response had been "neither agree or disagree." This directional effect of anchors has been noted in other contexts, such as negotiations, where individuals tend to assimilate final values toward an initial offer (Galinsky and Mussweiler 2001). To our knowledge, however, there has been little discussion of this source of bias with respect to sequences of Likert-scale questions.

In the presence of sequential anchoring, the order of the questions matters, since a different series of anchors likely leads to different results. Sequential anchoring, like many other forms of anchoring, is a source of measurement error, which could result in a systematic bias in sample results. Much of the work that has identified the various sources of bias in survey questions has also provided insight into the effective design of such questions in order to best eliminate, or at least minimize, the bias. Virtually all of these efforts, which include providing incentives, explicit warnings, and specific question structures (see Furnham and Boo (2011) for a comprehensive list and discussion), focus on surveying techniques and data collection. Overall, the effectiveness of these techniques is uncertain and seems to depend on the particular context (Furnham and Boo 2011). Interestingly, there is little research on quantitative methods to identify and measure the extent of the measurement bias *after* the data have already been collected. Though useful in practice, conducting such analysis is often difficult because context effects are hard to quantify. The nature of sequential anchoring, however, makes it well suited for statistical analysis, as it induces different distributions of responses for different question orderings.

The overall goal of this paper is to develop a stochastic model for a set of responses to a sequence of Likert-scale questions. More specifically, within this goal, the primary objective is to identify the presence of sequential anchoring and a secondary objective is to measure the magnitude of its effect. To this end, we develop a latent Gaussian variable framework that is suitable for a large number of Likert-scale questions and naturally accommodates a component meant to mimic the anchoring effect. Ultimately, we are interested in applying this model to the data from the 2012 Survey of Consumer Payment Choice (SCPC) regarding the assessment of various payment characteristics. We begin in Section 2 by introducing the relevant portion of the 2012 Survey of Consumer Payment Choice. Section 3 develops a latent variable model for a sequence of Likert-scale questions and a model of the anchoring effect. In Section 4, we discuss the methodology for fitting this model through an adapted Expectation-Maximization (EM) algorithm and discuss the results of doing so on simulated data. Section 5 follows this methodology to fit the model to the SCPC data. A discussion of the results is given in Section 6.

2 Data

In this paper, we analyze data from the 2012 version of Survey of Consumer Payment Choices (SCPC), an online survey conducted annually since 2008 by the Consumer Payment Research Center at the Boston Federal Reserve. A portion of the 2012 SCPC asks each respondent to rate a sequence of eight payment instruments on six different payment characteristics. In 2012, the characteristics to be rated were: acceptance, cost, convenience, security, ease of setting up, and access to payment records. Each characteristic is presented on a separate screen with the instruments listed vertically in a table as shown by the screenshot in Figure 1. Each instrument is to be rated on a five-point ordered scale.

In all years of the SCPC, the order in which the six payment characteristics are presented to the respondent are randomized, but prior to 2012 the order of the instruments was always fixed. However, the 2012 version of the survey randomizes the order of the instruments themselves. The eight payment instruments are grouped into three general types of payment instruments: paper, plastic, and online. The top panel in Table 1 lists the eight instruments by type. The randomization of the survey instruments was done by permuting the order of the three general groups of instruments while maintaining the same order within each group. Therefore, there are six possible orderings for the instruments, all shown in the second panel of Table 1. The 2012 SCPC was taken by 3,177 individuals, and the instrument orderings were assigned randomly to each respondent (and maintained for all six characteristics for that individual), meaning we have around 500 sequences of ratings for each ordering. It is this randomization, rare in consumer surveys, that allows us to study patterns in ratings under different orderings and look for asymmetries attributable to anchoring.

While the SCPC, which samples from RAND's American Life Panel (ALP), offers a wealth of information about each respondent including weights matching each annual sample to the population of adult consumers in the United States, we focus exclusively on the assessment of instrument characteristics data. We are less interested in making population-based estimates than we are in identifying a surveying phenomenon, so we treat the sample of respondents in the SCPC as representative of the population of survey-takers. It should be noted, however, that any inferences made in this work about general attitudes toward characteristics of payment instruments is limited to the population behind the ALP and may not be representative over broader populations of interest. More information about the ALP can be found at `http://mmic.rand.org/alp`.

3 Model

Likert-scale questions take many forms, but for simplicity of discussion we refer to each question as an "item" to be "rated," just like the eight payment instruments in the SCPC. As defined conceptually in the introduction, a sequential anchoring effect introduces bias by affecting the joint distribution of ratings for a sequence of items. In particular, the correlation of ratings for consecutive items increases. To identify the presence of sequential anchoring, it is necessary to view responses from different item orderings to assess whether the dependence structure changes.

In principle, nonparametric procedures testing whether the observed frequencies for sets of item

Table 1: The three different types of instruments referenced in the SCPC along with the six different orderings presented at random to respondents. The different orderings reflect different permutations of the three instrument types.

Type of Instrument	Instruments		
Paper	Cash (C)	Check (Ch)	Money Order (MO)
Plastic	Credit (CC)	Debit (DC)	Prepaid Card (PC)
Online	Bank Acct. # Payments (BA)	Online Bill Payment (OB)	

Order 1	C	Ch	MO	CC	DC	PC	BA	OB
Order 2	C	Ch	MO	BA	OB	CC	DC	PC
Order 3	CC	DC	PC	C	Ch	MO	BA	OB
Order 4	CC	DC	PC	BA	OB	C	Ch	MO
Order 5	BA	OB	C	Ch	MO	CC	DC	PC
Order 6	BA	OB	CC	DC	PC	C	Ch	MO

ratings differ substantially under different orderings could be developed. However, as the number of items or the number of possible ratings increases, the number of possible response sequences grows quickly, requiring a very large sample size for each ordering to produce robust estimates of the distributions. It might be possible to affirm the presence of sequential anchoring by studying the marginal distribution of ratings for each item under different orderings, but this would not use all of the available information and a negative result would not necessarily indicate a lack of anchoring. Perhaps more importantly, in the context of a nonparametric approach, it is not clear how to quantify the degree of the sequential anchoring and thus measure its effect on sample-based inference.

Although responses to singular Likert-scale questions are often modeled in item-response theory (Clogg 1979; Masters 1985), or through multinomial regression (Agresti 2002) and its variants (most notably the proportional odds model (McCullagh 1980; McCullagh and Nelder 1989)), there is little history of modeling entire sequences of Likert-scale responses. This is perhaps due to a combination of difficulty and lack of motivation. First, a broad class of models that can easily capture a wide range of complicated response patterns based on modest sample sizes is virtually nonexistent. In addition, sequences of Likert-scale questions are most likely to appear in surveys, and analysis of such data most commonly relates to simple calculations that take the data at face value and do not require modeling. An exception might be the imputation of missing values, but the relative ease of techniques such as hot-deck or k-nearest neighbors imputation (Enders 2010; Rubin 1987) make those techniques much more appealing.

This work assumes that a parametric latent variable model underlies the reported Likert ratings. The model defines a deterministic mapping from a normal random variable to a set of ordered ratings for each item. Such a model is easily extended to the case in which respondents are asked to rate a sequence of items by considering a latent vector from a multivariate normal distribution. The model framework also allows for the introduction of a sequential anchoring component that affects the latent vector and thus biases the ratings. The following sections provide more detail about the model and its notation.

3.1 Latent Gaussian Variable Model

Consider a survey questionnaire in which J items are to be rated sequentially by each respondent, just as for the $J = 8$ instruments in the SCPC survey shown in Figure 1. The analysis in this work assumes that each item is rated with one of five possible ratings, represented by the integers from one to five. The model can be extended to a different number of possible rating choices, though its effectiveness in fitting the data generally decreases as the number of choices increases, as discussed below in Section 3.3. The ensuing results based on five ratings, however, should be of wide interest, as a five-point scale is common in survey literature (Dawes 2008).

For individual i, we let R_{ij} be the rating given to item j in some predetermined, standard ordering of all the items to be rated. In the case of the SCPC, the standard ordering for the eight payment instruments is taken as the first ordering given in Table 1. The collection of ratings given by individual i for all J items is then $\mathbf{R}_i = [R_{i1}, R_{i2}, \ldots, R_{iJ}]^T$. For each item rating, R_{ij}, we assume an underlying Gaussian random variable with a mapping from that variable to the five possible ratings given by the respondent:

$$\mathcal{R} : \mathbb{R} \to \{1, 2, 3, 4, 5\}.$$

Specifically, for the j^{th} item in the sequence, let $X_{ij} \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Then, the mapping $\mathcal{R}$ is as follows

$$R_{ij} = \begin{cases} 1 & \text{if } X_{ij} \in (-\infty, -3) \\ 2 & \text{if } X_{ij} \in [-3, -1) \\ 3 & \text{if } X_{ij} \in [-1, 1) \\ 4 & \text{if } X_{ij} \in [1, 3) \\ 5 & \text{if } X_{ij} \in [3, \infty). \end{cases} \tag{1}$$

Given the definition in (1) and the parameters (μ_j, σ_j^2) it is possible to determine the probability of each of the five possible ratings for item j. We first define the functions $\ell(r)$ and $u(r)$ as the lower and upper bounds that correspond to a rating of r,

$$\ell(r) = \inf\{x \mid \mathcal{R}(x) = r\} \quad \text{and} \quad u(r) = \sup\{x \mid \mathcal{R}(x) = r\},$$

with $\mathcal{R}$ defined in (1). For example, for a rating of $r = 3$, $\ell(3) = -1$ and $u(3) = 1$. Then, the probability of observing a rating of r for item j is defined as

$$\begin{aligned} P_j(r) &= \text{Prob}(X_{ij} \in [\ell(r), u(r)]) \\ &= \int_{\ell(r)}^{u(r)} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ -\frac{1}{2\sigma_j^2}(x - \mu_j)^2 \right\} dx. \end{aligned}$$

The generation of $P_j(r)$ through a density function in this fashion assures that $0 \leq P_j(r) \leq 1$ for all r and $\sum_{r=1}^{5} P_j(r) = 1$, necessary and sufficient conditions for a probability distribution on five outcomes. Three examples of underlying Gaussian random variables and their implied probability distributions of rating values are shown in Figure 2.

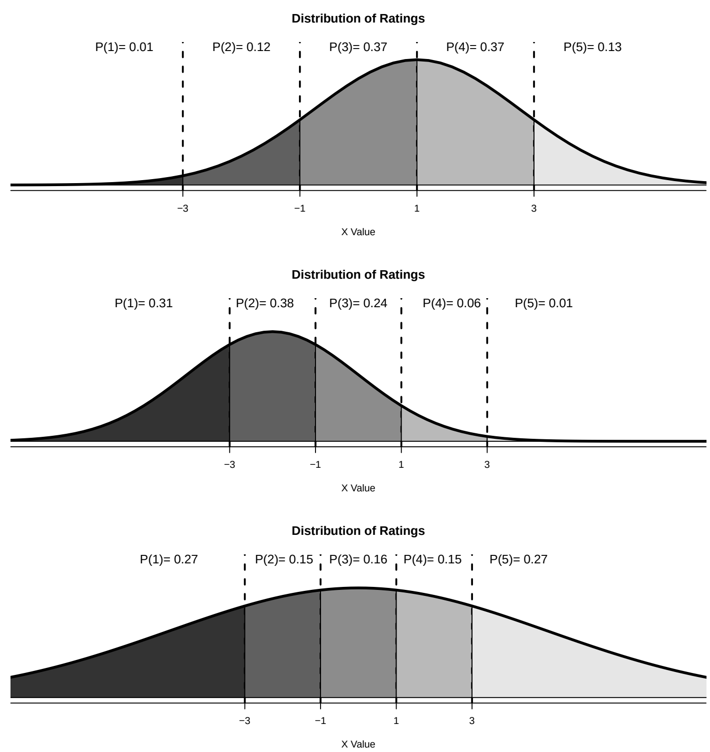


Figure 2: Distributions of the five ratings under the latent Gaussian model for $\mu_j = 1, -2, 0$, and $\sigma_j = 2, 2, 5$, respectively.

Before proceeding, it is important to discuss the implication of the interval choices in (1). While the intervals themselves may seem somewhat arbitrary, translations or proportional re-scaling of the intervals will not affect the span of the model, because such changes can be compensated for with changes in μ_j and σ_j^2, respectively. What does matter, however, is the relative size of the intervals to one another. With the intervals fixed, the Gaussian model has only two functional parameters, μ_j and σ_j^2, with which to define the set of five probabilities for the rating values, $P_j(r)$ with $r = 1, \ldots, 5$. This difference in degrees of freedom means that the set of rating distributions generated through the latent Gaussian variable is only a subset of all valid rating distributions.

In this sense, the choices for the relative lengths may seem somewhat arbitrary, since different selections lead to a different space of possible distributions. For any particular Likert question, a better fit may be induced by a change in the relative lengths of the intervals. However, our choice to have the middle three intervals be of the same length can be justified through a Bayesian argument. Namely, a uniform prior on μ_j that is symmetric around zero together with an uninformative prior on σ_j imply that $P_j(1) \leq P_j(2) \leq P_j(3)$ and $P_j(1) = P_j(5)$ and $P_j(2) = P_j(4)$. The implication is that, a priori, the expected distribution of ratings is symmetric, with the more neutral ratings being more common. Therefore, under the adopted choice of intervals, a natural and relatively uninformative prior on (μ_j, σ_j) corresponds to desirable prior assumptions about the ratings of a well-formulated Likert-scale question.

The model can be extended to the case of J sequential Likert-scale questions by considering a multivariate Gaussian distribution as the latent variable. Consider $\mathbf{X}_i = [X_{i1}, X_{i2}, \ldots, X_{iJ}]^T$ to be a multivariate normal vector with mean μ and variance Σ, $\mathbf{X}_i \sim \mathcal{MVN}(\mu, \Sigma)$. Then, given $\mathbf{X}_i$, we can map each individual component X_{ij} via the mapping $\mathcal{R}$ to determine the observed sequence of ratings

$$\begin{aligned} \mathbf{R}_i &= \mathcal{R}(\mathbf{X}_i) \\ &= \left[\mathcal{R}(X_{i1}), \ldots, \mathcal{R}(X_{iJ})\right]. \end{aligned}$$

The multivariate model accounts for the inherent relationships between attitudes toward related items through Σ. For example, if people tend to feel similarly about the items j and j', this will be reflected in a positive correlation between X_{ij} and $X_{ij'}$.

If μ and Σ are known, it is conceptually easy to determine the probability of any particular set of ratings by integrating the multivariate normal distribution over the subspace of $\mathbb{R}^J$ that maps to $\mathbf{R}_i$. Therefore, the probability of observing a particular sequence $\mathbf{r} = (r_1, \ldots, r_J)$ by any individual is given by

$$
\begin{aligned}
\mathcal{L}(\mathbf{R}_i = \mathbf{r} \mid \mu, \Sigma) &= \mathcal{L}(R_{i1} = r_1, \ldots, R_{iJ} = r_J \mid \mu, \Sigma) \\
&= \int_{\ell(r_1)}^{u(r_1)} \cdots \int_{\ell(r_J)}^{u(r_J)} \frac{1}{(2\pi)^{J/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\} d\mathbf{x},
\end{aligned}
$$

where $\mathcal{L}(\text{data} \mid \theta)$ is used to represent the likelihood of observing the data under a given set of parameters θ. Marginal and conditional distributions for item ratings can be calculated as well, since the underlying variables are also normally distributed with means and variances well-known functions of μ and Σ.

3.2 Modeling Anchoring Effects

As of now, the latent Gaussian model does not incorporate any anchoring effects, and the ordering of items in the survey does not affect the distribution of ratings. A "true" rating for item j is defined as one based on the underlying variable X_{ij} alone. By construction, sequential anchoring has a Markovian structure in which the reported rating for the j^{th} item in a sequence is based partly on the "true" rating for that question and partly on the reported rating for the $(j-1)^{th}$ item in the sequence, which in turn depends on that of the previous one. Below we incorporate this component into the stochastic model defined in the previous section.

Any particular reordering of the J items from the standard ordering that is given to individual i can be expressed by $o_i = \{o_{i1}, \ldots, o_{iJ}\}$, a permutation of the integers 1 through J corresponding to the order in which the items are rated. As an example, if $o_{i1} = 3$, then the third item in the standard ordering is rated first by individual i. In addition, we let O_i be a $J \times J$ matrix such that in the j^{th} row the o_{ij}^{th} element is one and all other elements are zero. Therefore, O_i is a permutation matrix so that $O_i\mathbf{R}_i$ refers to the ordered sequence of ratings from respondent i and $O_i\mathbf{X}_i$ is the sequence of corresponding latent variables. It should be noted that $O_i^T = O_i^{-1}$, and thus O_i^T can be used to reorder a sequence back to the standard ordering. In theory, there are $J!$ possible orderings, but in practice a survey will be limited to a much smaller subset of potential orderings. We let $\mathcal{O}$ denote the collection of possible orderings and define $\#_o$ to be the number of unique orderings in $\mathcal{O}$. In the case of the SCPC payment characteristic questions, $\#_o = 6$, and all orderings are shown in Table 1.

With no measurement bias, the set of ratings given by individual i to the items in ordering o_i is given by $\mathcal{R}(O_i\mathbf{X}_i)$. Thus, the set of ratings in the standard ordering is $O_i^T\mathcal{R}(O_i\mathbf{X}_i)$, which, due to the linearity of the mapping, is equivalent to $\mathcal{R}(\mathbf{X}_i)$. The sequential anchoring effect is incorporated by having the ratings for individual i not be a mapping of $\mathcal{R}(\mathbf{X}_i)$, but rather a mapping of $\mathcal{R}(\mathbf{Y}_i)$ where $\mathbf{Y}_i = O_i^T h(O_i\mathbf{X}_i)$ for some function $h(\cdot)$. If $h(\cdot)$ has a different effect on each element of the input vector, the particular ordering of the items, as defined by O_i, influences the latent vector and thus the final ratings. To motivate the form of $h(\cdot)$, we let $X_{i,o_i(j)}$ be the underlying variable of the j^{th} item in the ordering defined by o_i. Thus, $X_{i,o_i(1)}$ is the value of X_{ij} such that item j is rated first in the ordering given by o_i. $Y_{i,o_i(j)}$ is the biased variable corresponding to $X_{i,o_i(j)}$ through $h(\cdot)$.

We assume that the very first question is rated without anchoring or bias and therefore that $Y_{i,o_i(1)} = X_{i,o_i(1)}$. However, the value of the second ordered variable will now be a weighted average of $X_{i,o_i(2)}$ and the previous response $Y_{i,o_i(1)}$: $Y_{i,o_i(2)} = wX_{i,o_i(2)} + (1-w)Y_{i,o_i(1)}$ for some $0 \leq w \leq 1$. This pattern continues and thus $h(\cdot)$ is defined by

$$\begin{aligned} Y_{i,o(1)} &= X_{i,o(1)} \\ Y_{i,o(j)} &= wX_{i,o(j)} + (1-w)Y_{i,o(j-1)}, \text{ for } j = 2, \ldots, J. \end{aligned}$$

The exact distribution of $O_i\mathbf{Y}_i$ can be derived by letting W be a $J \times J$ lower-triangular matrix of the following form

$$W = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1-w & w & 0 & \ldots & 0 \\ (1-w)^2 & w(1-w) & w & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-w)^{J-1} & w(1-w)^{J-2} & w(1-w)^{J-3} & \ldots & w \end{bmatrix}.$$

Then, $O_i\mathbf{Y}_i = WO_i\mathbf{X}_i$. The linear nature of $h(\cdot)$, combined with the fact that $\mathbf{X}_i$ is a multivariate normal vector, implies that the random vector $\mathbf{Y}_i$, conditional on the assigned ordering, is itself normally distributed:

$$\mathbf{Y}_i \mid o_i \sim \mathcal{MVN}\left(O_i^T W O_i \mu, O_i^T W O_i \Sigma\, O_i^T W^T O_i\right). \tag{2}$$

The derived distribution in (2) makes clear that the distribution of $\mathbf{Y}_i$, and thus $\mathbf{R}_i$ depends on the anchoring effect captured by w and the ordering of the questions. The value of w corresponds to the degree of anchoring, with a value of $w = 1$ corresponding to no anchoring and the degree of bias increasing as w approaches zero. Only if $w = 1$, is W the identity matrix, and $\mathbf{Y}_i = \mathbf{X}_i$, meaning that the ratings will be given according to the true distribution defined by μ and Σ, without anchoring bias.

3.3 Discussion of the Latent Gaussian Model

The main drawback of the latent Gaussian model, as alluded to earlier, is that it cannot generate all valid distributions of ratings for questions with five response choices. The proportion of valid sets of probabilities that can be generated through the latent Gaussian model decreases as the number of ratings increases, because of a growing difference in the number of free parameters with which to define the probability distribution. Conversely, if there are only three possible ratings, the latent Gaussian model will be able to match any set of probabilities perfectly. In the case of five possible responses, the most extreme divergences generally correspond to patterns lacking unimodality; for example, a high frequency of twos and fours, but low frequencies of ones, threes, and fives. From a survey methodological standpoint, we hope that many of these distributions are rarely observed in well-calibrated questions. At the same time, the last plot in Figure 2 shows that more regular bimodal patterns, especially those with a high frequency of ones and fives, can be accommodated with the latent Gaussian model. Nevertheless, even many unimodal rating distributions will not be perfectly matched by the latent Gaussian model, due to its limited number of parameters. The general lack of flexibility contrasts with the flexibility of a multinomial model, in which any assortment of rating frequencies is possible and maximum likelihood estimates will simply be the empirical frequencies of each set of ratings.

However, the latent Gaussian model does take into account the correlations in ratings for different items. Doing so is considerably more difficult with a multinomial model. On one extreme, treating each sequence of responses as a joint outcome requires a multinomial distribution on 5^J possible outcomes. A sample size that is much smaller than 5^J ensures that frequency estimates will not be robust, with plausible sequences that happened not to be observed in the sample assigned a probability of zero. On the other hand, an assumption of independence across items may also be untenable, as inherent characteristics of the items often lead to positive or negative correlations in ratings.

The utility of the latent Gaussian model when sequential anchoring is not an issue depends on the degree of mismatch between the empirical rating frequencies and the optimal fits within the class of latent Gaussian models as well as on the extent of dependence in the item ratings. Perhaps the simplest way to compare the latent Gaussian model to any other is to fit both and compare the quality of fit. From a practical point of view, we find that even if the probability distributions of item ratings are not well matched, the overall average ratings for each item based on the latent Gaussian model are quite close to the empirical ones.

In the context of this paper, the latent Gaussian model serves as a framework for assessing the extent of the measurement error due to sequential anchoring. Its appeal lies in the ability to model the joint distribution of ratings with a modest number of parameters. The addition of w produces a nested model in which a single parameter accounts for any potential sequential anchoring. Thus, fitting the model twice, once with $w = 1$ fixed and once with w to be estimated, serves as a hypothesis test in which the null hypothesis is that the sequential anchoring does not exist. A model in which w is set to one fits a model in the space of the latent Gaussian models that most closely matches the empirical distribution of ratings regardless of item ordering. With the parameter w estimated, $w \neq 1$ only if there is evidence that the joint distributions under different orderings are different, in other words $\mathcal{L}(\mathbf{R}_i \mid o_i = o) \neq \mathcal{L}(\mathbf{R}_i \mid o_i = o')$ for $o \neq o'$.

Much like the multinomial model, the latent Gaussian model supposes that, in the absence of sequential anchoring, rating sequences for individuals are independent and identically distributed from the population distribution defined by μ and Σ. The anchoring extension implies a fixed, population-wide value of w, suggesting the same degree of sequential anchoring for each individual. In reality, this may not be the case. Survey-specific causes, such as rushing to finish, and topic-specific causes, such as varying familiarity with the items, may lead to different levels of influence of the anchor values across individuals (Chapman and Johnson 1994; Wilson et al. 1996). However, the body of work on anchoring as a whole suggests it to be a robust phenomenon in which individual effects have a relatively minor effect (Furnham and Boo 2011). In addition, modeling individual anchoring effects is considerably more difficult, as a great deal of power to measure the extent of the asymmetries in joint distributions of ratings comes from the assumption of a common biasing term. Overall, we believe that our model provides a good first approximation into the extent of any sequential anchoring. Despite this simplification, we do allow for different levels of anchoring effects for different sets of Likert-scale sequences.

4 Parameter Estimation

In this section, we outline the algorithm used to estimate the parameters in the latent Gaussian model, $\theta = \{w, \mu, \Sigma\}$, introduced in Section 3. From a scientific point of view, all three sets of parameters could be of interest, albeit for different reasons. The parameters μ and Σ define the true distribution of ratings for the J items rated in the survey, and knowledge of these parameters allows one to make unbiased inferences about the distribution of ratings in the population. The parameter w, on the other hand, has survey methodological importance as a measure of the degree of sequential anchoring for a particular topic.

Our algorithm estimates all three parameters simultaneously. The procedure is a hybrid of the EM algorithm and Monte-Carlo-based simulated annealing. More on EM algorithms can be found in Dempster, Laird, and Rubin (1977), with Monte Carlo variants discussed in Booth and Hobert (1999); McCulloch (1994); simulated annealing is described in Kirkpatrick, Gelatt, and Vechhi (1983); Press et al. (2007). Section 4.1 provides an overview of the methodology with the details found in the appendix. Section 4.2 describes the results of the fitting procedure on simulated data.

4.1 Estimation Algorithm

We assume that a sequence of J items is rated by each of N individuals, with respondent i providing the ratings in ordering defined by o_i. We consider $\mathbf{R} = \{\mathbf{R}_1, \ldots, \mathbf{R}_N\}$ to be the collection of ratings in the standard ordering for all individuals and $\mathbf{o} = \{o_1, \ldots, o_N\}$ to be the set of question orderings shown to all respondents. Based on these data, the observed data likelihood function for the parameters conditional on the assigned orderings can be written as

$$\begin{aligned} \text{lik}(\theta \mid \mathbf{R}, \mathbf{o}) &= \prod_{i=1}^{N} \mathcal{L}(\mathbf{R}_i = \mathbf{r}_i \mid o_i, w, \mu, \Sigma) \\ &= \prod_{i=1}^{N} \mathcal{L}\left(R_{i,o_i(1)} = r_{i,o_i(1)}, \ldots, R_{i,o_i(J)} = r_{i,o_i(J)} \mid o_i, w, \mu, \Sigma\right). \end{aligned} \tag{3}$$

For individual i, the likelihood on the right-hand side of (3) takes the explicit form

$$\mathcal{L}\left(\mathbf{R}_i \mid o_i, \theta\right) = \int\limits_{\ell(r_{i,o_i(1)})}^{u(r_{i,o_i(1)})} \cdots \int\limits_{\ell(r_{i,o_i(J)})}^{u(r_{i,o_i(J)})} \frac{(2\pi)^{-J/2}}{|WO_i\Sigma\, O_i^T W^T|} \exp\left\{-\frac{1}{2}\mathbf{z}_i^T \left[WO_i\Sigma\, O_i^T W^T\right]^{-1} \mathbf{z}_i\right\} d\mathbf{y}_i, \tag{4}$$

where $\mathbf{z}_i = O_i\mathbf{y}_i - WO_i\mu$.

While the likelihood defined by (3) and (4) is fairly easy to evaluate, it is not as straightforward to optimize for θ. Instead, the EM algorithm is a natural choice for parameter estimation, since the objective function would be much easier to work with if the latent variables $\mathbf{Y}_i$ were observed. Letting $\mathbf{Y}$ represent the underlying Gaussian variables for all N individuals, the full data negative log-likelihood takes the general form

$$
\begin{aligned}
\mathrm{nll}(\theta \mid \mathbf{Y}, \mathbf{R}, \mathbf{o}) &= -\sum_{i=1}^{N} \log \mathcal{L}(\mathbf{Y}_i, \mathbf{R}_i \mid o_i, w, \mu, \Sigma) \\
&= - \sum_{i=1}^{N} \log \mathcal{L}(\mathbf{Y}_i \mid o_i, w, \mu, \Sigma).
\end{aligned} \tag{5}
$$

The simplification in the last step of (5) is due to the fact that $\mathbf{Y}_i$ necessarily determines the ratings $\mathbf{R}_i$. Utilizing the derived distribution for $\mathbf{Y}_i \mid o_i$ in (2), we can specify the full-data negative log-likelihood as

$$
\mathrm{nll}(\theta \mid \mathbf{Y}, \mathbf{R}, \mathbf{o}) \propto \sum_{i=1}^{N} \left\{ \log |W O_i \Sigma\, O_i^T W^T| + (O_i \mathbf{Y}_i - W O_i \mu)^T \left[W O_i \Sigma\, O_i^T W^T \right]^{-1} (O_i \mathbf{Y}_i - W O_i \mu) \right\},
$$

which simplifies to

$$
\begin{aligned}
\mathrm{nll}(w, \mu, \Sigma \mid \mathbf{Y}, \mathbf{R}) \propto\ & N \log |W \Sigma\, W^T| + N \mu^T \Sigma^{-1} \mu - 2 \sum_{i=1}^{N} \mu^T \Sigma^{-1} O_i^T W^{-1} \mathbf{Y}_i \\
& + \mathbf{Y}_i^T O_i^T W^{-T} O_i \Sigma^{-1} O_i^T W^{-1} O_i \mathbf{Y}_i.
\end{aligned} \tag{6}
$$

The variable $\mathbf{Y}_i$ is not actually observed, but the EM algorithm allows us to optimize the expectation of the log-likelihood function in (6) with respect to $\mathbf{Y}_i \mid \mathbf{R}_i, o_i$ based on the most recent estimates of the parameters θ, which in the k^{th} iteration we refer to as $\hat{\theta}^{(k)}$. This conditional expectation takes the form

$$
\begin{aligned}
\mathrm{E}\left[\mathrm{nll}(\theta \mid \mathbf{Y}, \mathbf{R}, \mathbf{o}) \mid \hat{\theta}^{(k)} \right] \propto\ & N \log |W \Sigma\, W^T| + N \mu^T \Sigma^{-1} \mu - 2 \sum_{i=1}^{N} \mu^T \Sigma^{-1} O_i^T W^{-1} \hat{\mathbf{M}}_i^{(k)} \\
& + \hat{\mathbf{M}}_i^{(k)} O_i^T W^{-T} O_i \Sigma^{-1} O_i^T W^{-1} \hat{\mathbf{M}}_i^{(k)} + \mathrm{tr}\left[O_i^T W^{-T} O_i \Sigma^{-1} O_i^T W^{-1} O_i \hat{\mathbf{S}}_i^{(k)} \right]
\end{aligned} \tag{7}
$$

where $\hat{\mathbf{M}}_i^{(k)} = \mathrm{E}\left[\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \hat{\theta}^{(k)} \right]$, $\hat{\mathbf{S}}_i^{(k)} = \mathrm{Var}\left[\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \hat{\theta}^{(k)} \right]$, and $\mathrm{tr}[\cdot]$ represents the trace of a matrix.

The calculation of $\hat{\mathbf{M}}_i^{(k)}$ and especially $\hat{\mathbf{S}}_i^{(k)}$, along with the optimization of (7) with respect to w, μ, and Σ, can be difficult. Therefore, we adapt the algorithm to break down the optimization into two parts, so that the general approach in the k^{th} iteration of the optimization algorithm is:

(i.) Given $\hat{w}^{(k-1)}, \hat{\mu}^{(k-1)}, \hat{\Sigma}^{k-1}$, determine new estimates: $\hat{w}^k$ and $\hat{\mu}^k$.

(ii.) Given $\hat{\Sigma}^{k-1}, \hat{w}^k, \hat{\mu}^k$, determine new estimate: $\hat{\Sigma}^{k+1}$.

First, conditional on Σ, the optimal value of μ for a given w has closed form depending only on $\hat{\mathbf{M}}_i^{(k)}$, which can be approximated reasonably well for all $i = 1, \ldots, N$. Thus, with a fixed covariance matrix, the optimization reduces to a one-dimensional search over $w \in [0, 1]$, which is relatively simple. Further details of step (i.) are provided in Appendix A. To estimate Σ, we take advantage of the fact that the likelihood function (3) is easy to evaluate and the distribution $\mathbf{Y}_i \mid \mathbf{R}_i, o_i$ is easy to sample from, in order to avoid estimating $\hat{\mathbf{S}}_i^{(k)}$. Namely, given the most recent estimates of w, μ, and Σ, we sample $\mathbf{Y}_i$ conditionally on all available information and use the empirical covariance matrices based on the sampled $\mathbf{Y}$ as candidates of Σ. Sampling is repeated until a covariance that improves the likelihood is found. More details about the Monte Carlo simulated annealing used in step (ii.) are found in Appendix B. Finally, Appendix C discusses other aspects of the optimization algorithm such as starting values and stopping conditions.

4.2 Simulated Data

We rely on simulated data to test two main features of our model and the assumed parameter estimation procedure. First, we want to assess how well the algorithm estimates the parameters in data generated according to the latent Guassian model. In doing so, we explore how changes in the survey paradigms affect the quality of those estimates. The second aspect, equally important, is to verify that rating sequences generated with no sequential anchoring do, in fact, reflect this in the model fits. This is done for data simulated both within and outside of the latent Gaussian framework. All simulations are based on $J = 8$ items and five possible response choices for each item.

4.2.1 Simulations with Sequential Anchoring

We generated rating sequences according to different processes, as defined by (w, μ, Σ), and with parameters estimated under different survey paradigms, as defined by $(N, \mathcal{O})$. Considering different processes gives insight into how our algorithm does with different levels of anchoring and different degrees of dependence between ratings of the instruments. Varying the survey paradigms, on the other hand, reveals how the sample size and the set of possible orderings influence the estimates. Appendix D also discusses the results of using LASSO penalties with respect to the estimation of Σ.

There are infinitely many combinations of processes and survey paradigms that one could consider. In order to remain faithful to the design in the SCPC survey, we take as a base case the case where $N =$ 3,000 and $\mathcal{O}$ represents the six orderings in Table 1. We consider two additional paradigms; the first has $N =$ 3,000 and $\#_o = 12$, and the second has $N =$ 6,000 and $\#_o = 6$. In the latter, the six orderings are the same as in the base case. In the former, we simply add six new orderings to those in the base case. The new orderings were constructed by continuing to keep the three general blocks of items together, as in Table 1, but permute the sequence of the blocks as well as the order of items within each block. In particular, with 12 orderings, six items are featured as first in the sequence as opposed to just three items in the original six orderings. The additional orderings also ensure that no one item is always followed by the same item, which is true for five items in the original six orderings.

In terms of the underlying process, we consider samples from four unique parameter sets corresponding to two different values of w and two different values of Σ. In general, we choose the parameters μ, Σ so that there is a fair amount of variation in the ratings provided for each item, as observed in the SCPC data and expected in most well-designed questions. If there is no variation in the ratings within each item (for example, if $|\mu_j|$ is large and σ_j small), the observed data will not carry evidence of sequential anchoring even if it is present. We sample the mean μ_j as independent draws from the $\mathrm{Unif}(-4, 4)$ distribution. The chosen values of $w = 0.8$ and $w = 0.95$ reflect different degrees of anchoring, with the former imposing more bias than the latter. Our two choices of the covariance matrix, Σ_1 and Σ_2 as illustrated in Figure 3, are also meant to reflect different dependence structures for the ratings. Σ_1 is sparse with three independent blocks, but high correlations within each block; Σ_2 is not sparse, but has weaker correlations between items.

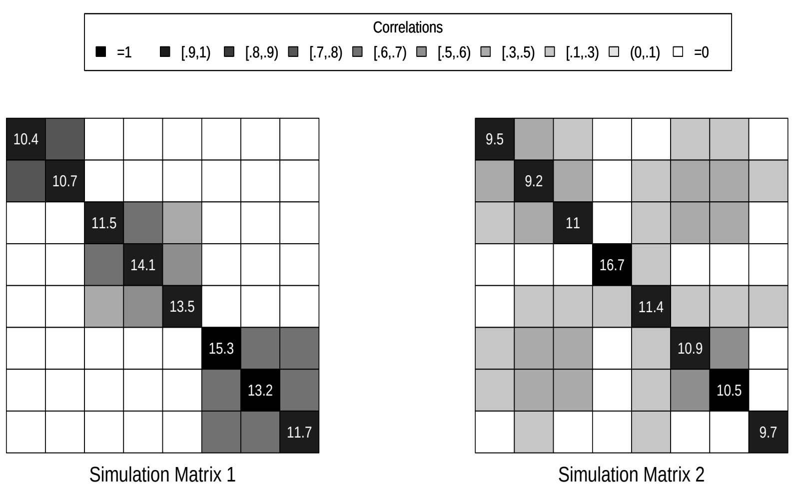


Figure 3: The two covariance matrices used in simulation: Σ_1 and Σ_2. The diagonal numbers represent the variances.

For each of the 12 combinations of process and survey paradigm, we generated multiple samples to account for sampling variability. We generated 10 independent samples for each, except in the case where $N = 6{,}000$, where only five samples were generated. This reduction was due to the fact that the computation time to fit our model increases with the sample size. The number of samples for each combination is shown in Table 2.

Table 2: The four processes used in simulations as well as the number of simulated datasets for each process under the different paradigms. Only five simulations were done for the case where $N = 6000$ because of a restriction on time to estimate the parameters, which grows with N.

	Process Parameters			**Survey Paradigm** $(N, \#_o)$		
Simulation	**w**	μ	$\boldsymbol{\Sigma}$	(3000,6)	(6000,6)	(3000,12)
A	0.8	iid Unif$(-4, 4)$	Σ_1	10	5	10
B	0.8	iid Unif$(-4, 4)$	Σ_2	10	5	10
C	0.95	iid Unif$(-4, 4)$	Σ_1	10	5	10
D	0.95	iid Unif$(-4, 4)$	Σ_2	10	5	10

In order to evaluate the quality of fit for all cases, we compare the degree of similarity between the true process parameters and the fitted ones. Figure 4 shows the estimated values of w for each simulation. The averages for all simulations are also shown (in red). It is clear that our algorithm does reasonably well at determining the anchoring effect, with all estimates within 0.02 of the true value. There is no clear evidence of a difference between the distribution of estimates under the different survey paradigms. Although there is some sampling error, the true values of w fall within 95 percent confidence intervals based on sampling statistics in all 12 cases.

Under the latent Gaussian model, the distribution of rating sequences is indirectly defined by the multivariate Gaussian distribution with parameters μ and Σ. Thus, an assessment of the parameter estimates based on a measure of distance between the true and fitted distributions seems to be an appropriate evaluation. One such measure of the distance between distributions is given by the symmetrized Kullback-Leibler divergence (Kullback 1959). If (μ, Σ) define the true multivariate normal distribution and $(\hat{\mu}, \hat{\Sigma})$ as data-based estimates, then the symmetrized Kullback-Leibler divergences between the fitted and true values will be

$$\mathrm{KL}(\mu, \Sigma, \hat{\mu}, \hat{\Sigma}) \quad = \quad \mathrm{tr}\left[\Sigma\hat{\Sigma}^{-1} + \hat{\Sigma}\Sigma^{-1} - 2I_J\right] + (\mu - \hat{\mu})^T\Sigma^{-1}(\mu - \hat{\mu}) + (\hat{\mu} - \mu)^T\hat{\Sigma}^{-1}(\hat{\mu} - \mu),$$

where I_J is a $J \times J$ identity matrix. The computed Kullback-Leibler divergences are shown in Figure 5 for all samples. The most noticeable aspect of the plot is that for each of the four simulations, the average divergence is much smaller when a larger sample size is used, whereas there seems to be no gain from

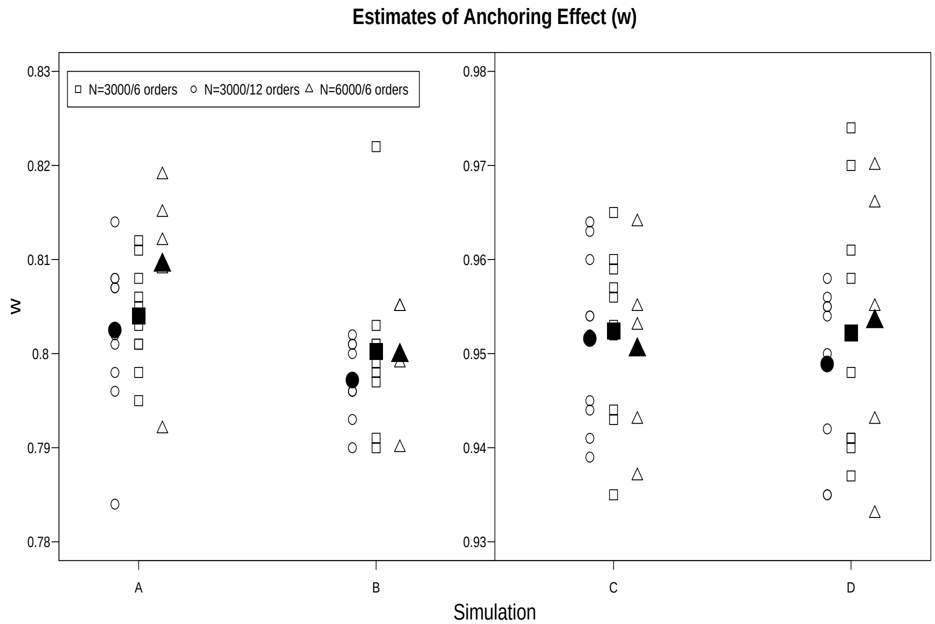


Figure 4: Estimated values of w for each simulation and paradigm type. Filled points represent the averages across all simulations.

increasing the number of orderings.

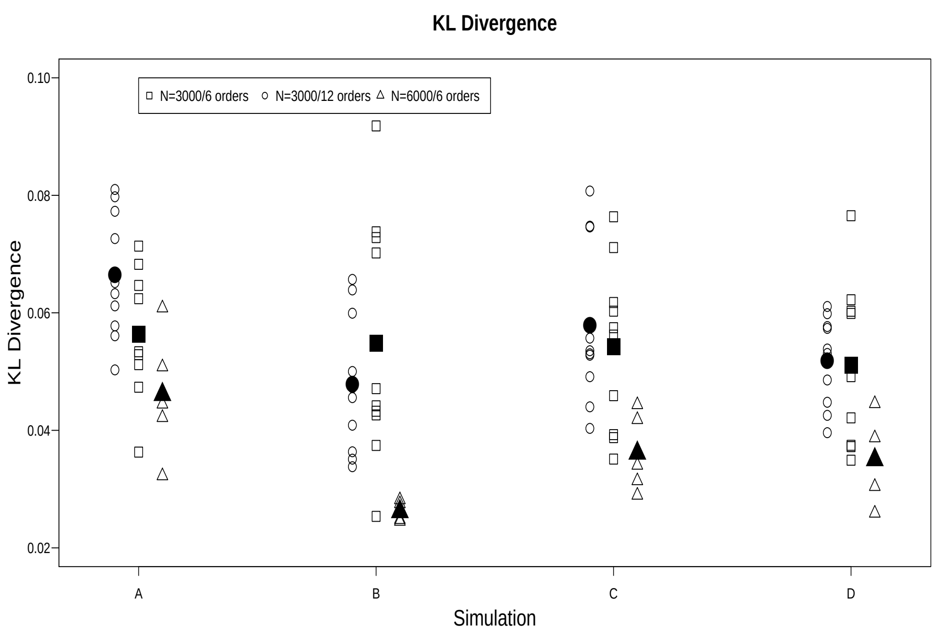


Figure 5: Kullback-Leibler divergence of estimates of (μ, Σ) for each simulation and paradigm type. Filled points represent the averages across all simulations.

Overall, our algorithm seems to perform well for all four processes regardless of the survey paradigm.

The algorithm is robust, as running it several times on the same datasets led to very similar log-likelihoods and parameter estimates. Even with strong anchoring effects, estimated parameters match the true values quite well. An interesting revelation is that, at least for sample sizes of $3{,}000$ (and presumably more), increasing the number of orderings does not aid much in the quality of the estimates. Evidently, there is enough information about the anchoring effect in the six original orderings that additional ones provide little extra information. It seems possible that as few as two orderings might provide enough insight into the bias introduced by the anchoring. While adding additional orderings does not seem to help, increasing the sample size does, especially when it comes to improving estimates of the underlying Gaussian parameters. This is a welcome result, as consistency of an estimator is desirable.

4.2.2 Simulations with No Sequential Anchoring

A second set of simulations was devoted to verifying that the algorithm did not recognize a sequential anchoring effect when it was not present. For the purposes of this exercise, $\#o$ was kept at six, though the assignment of ratings was done independently of the orderings. We constructed 10 datasets from three different generating models, none of which included a sequential anchoring effect:

(I) Latent Gaussian models defined by $w = 1$, $\mu_j \sim \mathrm{Unif}(-4, 4)$, and Σ_1 or Σ_2.

(II) A multinomial model with independence across items.

(III) A multinomial model with strong correlation between items.

For sequences corresponding to item (III), the correlation between item ratings was generated by a Markovian procedure in which the rating probabilities for one item depend on the rating of the previous item in the standard ordering. In the latter two cases, the marginal rating probabilities were chosen to be such that the latent Gaussian model cannot precisely match even the marginal rating probabilities for each item. As a result, in (II), the optimal latent Gaussian model provides a worse fit in terms of likelihood than the multinomial model with assumed independence. For each dataset, we fit the latent Gaussian model twice, once with the anchoring parameter w left to be estimated and once with w fixed to one. The estimates of w ranged from 0.985 to 1.00, with the latter being the optimal value in six out of 10 simulations. Perhaps more importantly, deviances (twice the log-likelihood differences) between the two fits were small, with a maximum of 1.65, which under a Chi-square distribution with one degree of freedom corresponds to a p-value of 0.19. Therefore, in each case, a comparison of the fits affirms the hypothesis that there is no evidence of sequential anchoring. This aspect of our model is vital, since we want to minimize the probability of falsely identifying a sequential anchoring effect.

5 Application to SCPC

In this section, we describe the application of our latent variable model to the payment characteristic data from the 2012 SCPC. To assess the variation in the sequential anchoring effect, we treat the six characteristics separately, fitting the model for each independently. In order to avoid imputation, we consider only those individuals who provided a rating for every instrument. For each characteristic, the percentage of individuals who met this criterion was upward of 98 percent. Below, we discuss the model fits as well as the implications of any measurement error on sample-based inference.

5.1 Results

As noted, the best test for the presence of sequential anchoring effects involves comparing the fits of the latent Gaussian model with $w = 1$ fixed and with w as a free parameter to be estimated. As a simple means of comparison, we also fit a multinomial model that treats item ratings as independent. The three models are:

Model 0: Latent-Gaussian model with anchoring.

Model 1: Latent-Gaussian model with no anchoring component ($w = 1$).

Model 2: Independent, multinomial model: $\mathcal{L}(R_{ij} = k) = q_{jk}$ for $\sum_{k=1}^{5} q_{jk} = 1$ and $R_{ij}, R_{i'j'}$ independent if either $i \neq i'$ or $j \neq j'$.

The first two models allow for dependence between an individual's response for one payment instrument and that for a different payment instrument, though only *Model 0* incorporates the sequential anchoring effect. By treating the given rating for each payment instrument as independent, *Model 2* not only ignores any sequential anchoring but also does not allow for any inherent dependencies between ratings for payment instruments. With a smaller number of instruments and a larger sample, one could consider estimating a more general multinomial distribution on all sequence of ratings. Unfortunately, with $J = 8$ instruments and five possible ratings for each instrument, there are 390,625 possible rating sequences. With a sample size as small as $N = 3{,}000$, a robust estimate of the distribution is unlikely.

We focus on comparing the differences in the negative log-likelihoods of the three models at their optimal fits. Thus, let nll_m represent the negative log-likelihood under the estimated parameters that maximizes the likelihood of the observed SCPC data under *Model m*, $m = 0, \ldots, 2$. For *Model 0*, the negative log-likelihood will be given by the log of (3) under the fitted parameters. Fits and log-likelihoods for the latent Gaussian model with no anchoring are determined by adjusting the procedure to force $w = 1$. For the independent multinomial model, *Model 2*, it is straightforward to determine the negative log-likelihood. If N_{jk} represents the number of individuals who rate payment instrument j with rating k for $j = 1, \ldots, 8$ and $k = 1, \ldots, 5$, then

$$\mathrm{nll}_2 \quad = \quad -\sum_{j=1}^{8}\sum_{k=1}^{5} N_{jk} \log \frac{N_{jk}}{N}.$$

As we are primarily interested in differences in the negative log-likelihoods between the anchoring-inclusive model (*Model 0*) and the rest, we define

$$\Delta_m = \mathrm{nll}_m - \mathrm{nll}_0$$

for $m = 1, 2$. These differences in log-likelihoods are shown in the left half of Table 3 for all six payment characteristics.

Table 3: Sample sizes, estimates of w, and improvements in negative log-likelihood over the independent-multinomial model for each payment characteristic. "All Data" includes everyone who rated every payment instrument, while "Nonvariants Removed" excludes all individuals who gave the same rating for each instrument.

	All Data				**Nonvariants Removed**			
Characteristic	N	$\hat{w}$	Δ_1	Δ_2	N	$\hat{w}$	Δ_1	Δ_2
Acceptance	3138	0.926	18	1899	3003	0.917	17	1371
Cost	3136	0.880	35	2191	2993	0.881	34	1713
Convenience	3136	0.957	8	1695	3035	0.958	7	1313
Security	3140	0.892	34	3238	3009	0.893	33	2652
Set-up Ease	3145	0.927	18	2937	2912	0.931	16	2094
Records	3140	0.937	14	2944	2994	0.940	13	1838

The log-likelihood differences between fits corresponding to the independent, multinomial model and the two latent Gaussian models indicate that the assumption of independence in responses for different payment instruments is highly unlikely, with this model carrying an increase of anywhere from a half to a whole negative log-likelihood unit per observation from the Latent-Gaussian model. The fitted marginal rating frequencies under *Model 1* are not exact matches of the empirical frequencies, although the degree of mismatch varies across the six payment characteristics and in some cases is quite small. Yet, in terms of likelihood, the dependence captured through the underlying multivariate normal greatly outweighs the imperfect matching of marginal rating frequencies. Incidentally, the latent Gaussian model also performed much better than a multinomial regression model in which the probability distribution of ratings for each item depended on age, gender, and income.

The differences in likelihood between the fit with $w = 1$ fixed and w estimated are much smaller. Nevertheless, as the two models are nested with only one degree of freedom difference, the deviances are significant under a Chi-square distribution for all six payment characteristics. Overall, there is strong evidence of a sequential anchoring effect. Table 3 shows the six estimates of w, which scatter around their average of $\bar{w} = 0.92$.

To test the hypothesis that the sequential anchoring effect is the same for all six payment characteristics, we refit the latent Gaussian model for all six payment characteristics with w fixed at 0.92. Summed over the six payment characteristics, the overall difference in log-likelihoods between this model and one in which w is estimated for each payment characteristic was 11.5. Again taking advantage of the nested nature, we make use of the likelihood ratio test with six degrees of freedom to find a p-value of 0.001, thus rejecting the hypothesis that the anchoring effects are the same across all characteristics. Based on the estimates of w, it seems that cost and security show the largest amounts of anchoring. We note that in our experience these two characteristics are the ones with the most ambiguity in their definition, perhaps making sequential anchoring a more important factor in the rating assignment process. Nevertheless, the estimates of w are fairly similar for all six payment characteristics.

5.2 Quality of Model Fit

In order to determine how well the estimated magnitudes of the sequential anchoring effects correspond to the true values, it is necessary to know how closely the latent Gaussian model conforms to the observed data. As noted, the class of latent Gaussian models cannot capture all valid probability distributions, and the larger the mismatch the more likely the estimated anchoring effect is incorrect. To study the quality of fit, we look at the predictive ability of our model, specifically by comparing observed data statistics to those expected under the fitted model.

For a given sequence of ratings $\mathbf{r} = (r_1, \ldots, r_8)$, we let $N^o(\mathbf{r})$ represent the number of individuals to give rating $\mathbf{r}$ under ordering o and let $N(\mathbf{r}) = \sum_{o \in \mathcal{O}} N^o(\mathbf{r})$ be the total number of individuals in our sample with such an ordering. In addition, for any ordering o, the probability of observing r under that ordering, $p^o(\mathbf{r})$, can be easily determined by integrating the multivariate normal distribution given in (2). Therefore, $N^o(\mathbf{r}) \sim \text{Binomial}(N^o, p^o(\mathbf{r}))$ meaning $N(\mathbf{r})$ will follow the distribution of the sum of six independent binomial distributions, one for each ordering. Exact confidence intervals for $N(\mathbf{r})$ are computationally difficult to determine, but we can approximate them by taking advantage of the fact that $p^o(\mathbf{r})$ are fairly similar to one another for different o. Taking a weighted average of the probabilities

$$p(\mathbf{r}) = \sum_{o=1}^{6} \frac{N^o}{N} p^o(\mathbf{r})$$

and assuming that $N(\mathbf{r}) \sim \text{Binomial}(N, p(\mathbf{r}))$ should provide a good approximation to the true distribution. Doing so makes it simple to determine a parameter-dependent 95 percent confidence interval for $N(\mathbf{r})$, which we we call $(L(\mathbf{r} \mid \hat{\theta}), U(\mathbf{r} \mid \hat{\theta}))$. Due to the fact that there are so many possible ratings relative to the sample size, more often than not it is the case that $U(\mathbf{r} \mid \hat{\theta}) = 0$. If this occurs, we set $U(\mathbf{r} \mid \hat{\theta}) = 1$. For a particular rating r we define

$$f(\mathbf{r} \mid \hat{\theta}) = \begin{cases} 1, N(\mathbf{r}) > U(\mathbf{r} \mid \hat{\theta}) \text{ or } N(\mathbf{r}) < L(\mathbf{r} \mid \hat{\theta}) \\ 0, N(\mathbf{r}) \in [L(\mathbf{r} \mid \hat{\theta}), U(\mathbf{r} \mid \hat{\theta})] \end{cases}$$

to be a measure of consistency between our observed data and the expected data. Thus, $f(\mathbf{r} \mid \hat{\theta}) = 1$ if the number of individuals with rating sequence $\mathbf{r}$ falls drastically above or below the number expected under the model fits and the sampling design.

The properties of $f(\mathbf{r} \mid \hat{\theta})$ in our sample provide some detail about the fit, although distributions of functionals of $f(\mathbf{r} \mid \hat{\theta})$ are difficult to determine because of our adaption of $U(\mathbf{r} \mid \hat{\theta})$ and because $f(\mathbf{r} \mid \hat{\theta})$ and $f(\mathbf{r}' \mid \hat{\theta})$ are not independent. Nevertheless, the overall prediction rate for the ratings observed in the sample, defined as the average of $f(\mathbf{r} \mid \hat{\theta})$ with respect to the ratings observed in the sample, ranged from 89 percent to 92 percent across payment characteristics. In order to learn more about where our model diverges from what is observed, we consider the failure rate as a function of the number of unique rating values given in a sequence $\mathbf{r}$, which we define as $v(\mathbf{r})$. Thus, the sequence $\mathbf{r} = (3, 4, 3, 3, 5, 4, 4, 1)$ has $v(\mathbf{r}) = 4$, since there are four unique ratings in the sequence $(1, 3, 4, 5)$. Naturally, $v(\mathbf{r})$ can take integer values from one to five, with smaller values indicating less distinguishing of payment instruments. We can then define $\mathcal{V}_k = \{\mathbf{r} \in \mathbf{R} \mid v(\mathbf{r}) = k\}$ to be the collection of observed ratings with k unique rating values and

$$\bar{f}_k(\mathbf{R}, \hat{\theta}) \quad = \quad 1 - \frac{\sum_{r \in \mathcal{V}_k} f(\mathbf{r} \mid \hat{\theta})}{\sum_{r \in \mathcal{V}_k} 1} \tag{8}$$

to be the average prediction rate by the number of rating values given. These rates for all values of k are shown in Figure 6. It is clear that our model does not provide accurate predictions for the number of individuals who give low-variation sequences. As the number of rating values increases, the prediction rate increases. It is important to remember that the number of possible ratings in each group increases with the number of ratings given (there are only five possible ratings for which $v(\mathbf{r}) = 1$). This raises the concern that our model does not accommodate individuals who have very high anchoring effects.

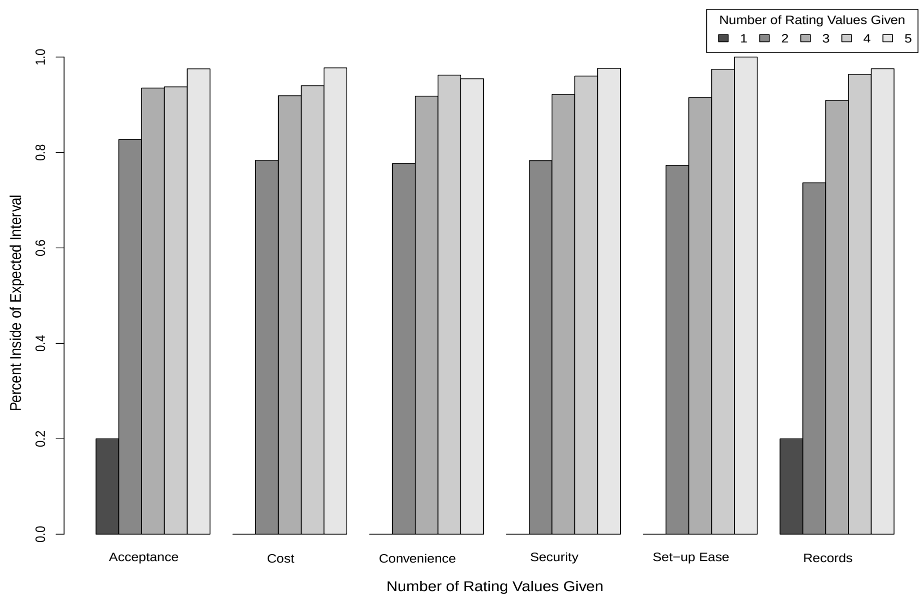


Figure 6: Prediction rate by the number of unique ratings given for each characteristic.

To explore this further, we plot the observed counts $N(\mathbf{r})$ with respect to the confidence bounds defined in $(L(\mathbf{r} \mid \hat{\theta}), U(\mathbf{r} \mid \hat{\theta}))$. This is done in Figure 7 with different symbols differentiating low-variation rating sequences in which $v(\mathbf{r}) = 1$ or $v(\mathbf{r}) = 2$. These plots suggest that the latent Gaussian model fits better for certain characteristics, such as cost or set-up ease, than for others, such as convenience or security. Figure 7 also indicates that the most egregious deviations from model-based expectations occur in cases in which $v(\mathbf{r})$ is low. For example, there are significantly more individuals who give the same rating to each instrument than would be expected. For the purposes of this work, we call these individuals "nonvariants." There is a very high number of nonvariants relative to what our model would predict (about $125 - 150$ for each payment characteristic), and it seems reasonable to question the value of their responses. Even ignoring the plausible explanation that the same rating was given to decrease the time spent in taking the survey, rating sequences in which each instrument has the same rating, and where that rating is not the most central rating (3), are difficult to interpret in the context of other rating sequences.

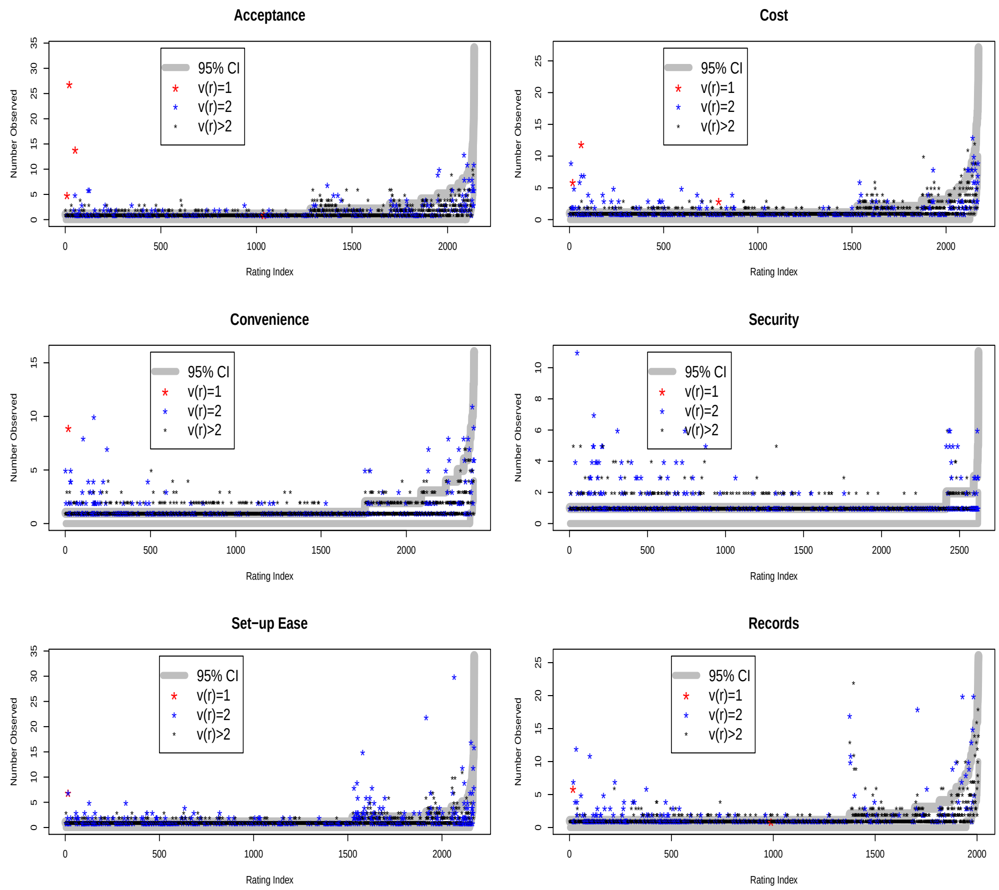


Figure 7: Observed counts for each rating along with approximate 95 percent confidence intervals based on parameter estimates. In some cases, the value of $N(\mathbf{r})$ in which $v(\mathbf{r}) = 1$ is large enough that it extends past the shown axes.

To deal with this fact, we fit the three models again, this time having removed all nonvariants from the sample. The new estimates of the anchoring effect w and differences in the negative log-likelihoods between our model and the independent-multinomial model are in Table 3. It is not surprising that the relative improvement gained with our model, as indicated by difference in log-likelihoods per number of observations, decreases. Removing the nonvariants allows for a stronger case for independence between payment instruments, a key aspect of the independent-multinomial model. Nevertheless, excluding this subset of people seems to improve the fit of our model substantially, with the new prediction rates, defined in (8), now ranging from 92 percent to 93 percent. The improvement is predominantly due to better alignment of the observed and expected results for ratings with higher values of $v(\mathbf{r})$.

5.3 Implications for the Data

While the model fits are not perfect, they seem to do a reasonably good job of capturing the data trends, especially with the nonvariants removed. In the following section, we study the implications of the observed sequential anchoring on the sample results. The analysis is based on the subsample with the nonvariants removed. Doing so changes the raw sample rating averages only marginally (at most by 0.02), and most conclusions are the same as with the full sample.

Table 4 shows the sample average rating for each payment instrument and payment characteristic as well as the deviations in the averages based on the fitted values for both latent Gaussian models. In the case where $w = 1$ (second row), the differences are minor and can be fully explained by mismatches in marginal probabilities for each item due to the limited flexibility of the latent Gaussian models. However, in the case where w is estimated, the differences are more substantial. As might be expected, empirical and fitted means are similar for the three instruments that are featured first in some ordering (C, CC, BA), since the model assumes that the responses for items that come first in a sequence (in this case, about one-third of all responses for each of the three instruments) are unaffected by sequential anchoring. For the remaining five instruments, the degree of change in the mean depends partly on the marginal distributions of the items considered. If marginal distributions of consecutive instruments are similar, it is even possible for a strong anchoring effect have little effect on the overall mean rating. The anchoring-adjusted ratings differ the most for instruments that routinely follow instruments with average ratings on the far sides of the spectrum. For example, the largest drops occur for acceptance (from 3.43 to 3.24) and cost (3.86 to 3.65) of check, the instrument that comes after cash, which has high scores for both characteristics. The largest increase occurs for record-keeping of checks (4.16 to 4.24), as cash has a very low average rating for this characteristic. Even for smaller changes, the adjustment for the sequential anchoring effects always involves a change in mean rating away from the mean rating of the previous instrument.

Because mean rating estimates correspond to sample averages of approximately $3{,}000$ individuals, many of the differences uncovered by the latent Gaussian model with sequential anchoring are statistically

Table 4: The average ratings in the 2012 SCPC for all eight instruments and all six characteristics. The averages were calculated ignoring any ordering or potential anchoring effects. The deviations from the average ratings as predicted by the fit are also shown.

Characteristic	C	Ch	MO	CC	DC	PC	BA	OB
Acceptance	4.63	3.43	3.19	4.42	4.54	3.89	2.65	3.65
Model 1	-0.01	–	–	-0.02	-0.03	-0.01	-0.01	-0.01
Model 0	-0.01	-0.19	-0.02	–	-0.03	-0.09	-0.02	+0.06
Cost	4.49	3.86	3.12	4.00	2.88	3.32	3.92	4.11
Model 1	+0.01	-0.01	–	–	+0.02	+0.01	-0.01	-0.01
Model 0	+0.02	-0.21	-0.11	–	-0.10	+0.05	+0.02	–
Convenience	4.15	3.23	2.37	4.39	4.37	3.43	3.09	3.98
Model 1	-0.01	–	+0.01	-0.02	-0.03	–	+0.01	-0.02
Model 0	-0.01	-0.05	-0.02	-0.01	-0.05	-0.06	+0.01	–
Security	2.76	3.05	3.21	3.20	3.31	2.85	2.76	3.28
Model 1	–	-0.01	-0.01	-0.01	-0.02	-0.01	–	-0.01
Model 0	–	+0.05	–	-0.02	-0.02	-0.05	-0.01	+0.04
Set-up Ease	4.35	3.77	3.08	4.02	3.72	3.48	3.42	3.57
Model 1	–	-0.01	–	–	–	–	–	-0.01
Model 0	+0.01	-0.09	-0.05	+0.02	-0.04	-0.04	–	–
Records	2.33	4.16	3.03	4.20	4.36	2.76	4.05	4.26
Model 1	+0.01	-0.01	–	-0.02	-0.02	+0.01	-0.02	-0.02
Model 0	-0.02	+0.08	-0.07	-0.01	-0.02	-0.09	+0.01	-0.02

significant (estimates of the margin of error range from 0.015 to 0.03, depending on the characteristic and instrument). As a result, inference about the relative population-wide rankings for different instruments with respect to a particular payment characteristic may differ if anchoring is accounted for. For example, with no anchoring, the estimated average ratings for cost of check and bank account number were 3.86 and 3.92, respectively, which yields little evidence that population attitudes differ for these instruments. However, once anchoring was taken into account, the estimated average ratings were 3.65 and 3.94, which now show strong evidence of a difference in perception of cost. In some cases, such as security for online banking and debit card, the relative ranking among instruments changes depending on whether anchoring is taken into account or not.

Comparisons of instrument ratings also occur at the individual level, perhaps most notably as predictors in revealed preference models (Samuelson 1938). In the case of the SCPC, Schuh and Stavins (Schuh and Stavins 2010; Stavins 2013) use a measure of the relative ratings of payment instruments in studying how the perceived security of payment instruments corresponds to adoption of those instruments. Specifically, a logistic model is used in which some of the covariates include the relative ratings of different instruments in the form $\log\left(\frac{R_{ij}}{R_{ij'}}\right)$ for $j \neq j'$. By its very nature, sequential anchoring affects the relative ratings by an individual, by masking some of the variability across items. A systematic bias that shrinks the diversity in item ratings at the individual level will lead to regressions based on a skewed view of relative item ratings.

Perhaps the simplest way to see how the anchoring observed in the SCPC might influence the individual level results is to consider its implications on the conditional distribution of ratings for one instrument given a particular rating for another. We select cash and check ratings ($j = 1$ and $j = 2$ under the standard ordering) with respect to the ease of set-up. Set-up ease is the payment characteristic in which the empirical and fitted rating frequencies under the latent Gaussian model with no sequential anchoring are most similar and has a fairly representative value of $\hat{w}$. Using the results of the model, we compare the probability distribution of ratings for check given a rating of five for cash, $\mathcal{L}(R_{i2} = r \mid R_{i1} = 5)$, as well as the distribution given a rating of one for cash, $\mathcal{L}(R_{i2} = r \mid R_{i1} = 1)$. We are interested in comparing the true conditional distributions, as reported with no anchoring, and the conditional distributions based on the sequential anchoring effect.

The results are shown in Table 5. A study of the unbiased results suggests a positive correlation between ratings, with individuals who reported higher ratings for cash also more likely to give higher values for check, and a slight shift away from higher ratings for a low cash rating. Nevertheless, the anchoring effect is made quite clear as the mass is shifted even more toward the preceding rating, often resulting in misclassification. For example, with no measurement error, 32.4 percent of individuals who gave a rating of five to cash would also give a rating of five to check. However, the sequential anchoring effect of magnitude $w = 0.927$ means that in a survey in which check always follows cash 36.2 percent would report a rating of five for check conditional on a rating of five for cash. Overall, studying the differences in Table 5, on average 13 percent of individuals with a five rating for cash and 12.7 percent of individuals with a rating of one for cash are misclassified. This is a sizeable percentage of individuals with measurement error, which could lead to erroneous inferences in the revealed preference models.

6 Discussion

The latent-variable model developed in this paper, along with the Expectation-Maximization/ Monte Carlo simulated annealing hybrid procedure to estimate the parameters, serves primarily as a tool for identifying the presence of sequential anchoring and, if present, estimating its magnitude. The model framework mirrors the logic of a hypothesis test, with a single parameter responsible for recognizing

Table 5: The likelihood of each rating for check set-up ease conditional on rating cash set-up ease as five and one, respectively.

Cash Rating $= 5$					
	Check Rating				
Reporting Type	1	2	3	4	5
Unbiased	1.3	8.0	24.2	34.1	32.4
With Anchoring	0.5	5.6	20.9	36.8	36.2

Cash Rating $= 1$					
	Check Rating				
Reporting Type	1	2	3	4	5
Unbiased	12.1	28.3	33.9	19.8	5.9
With Anchoring	12.8	30.3	37.0	16.6	3.2

the presence of asymmetries in the joint distribution of item ratings under different orderings. Although limited in scope, our simulation results suggest that our approach does well in rejecting the notion of sequential anchoring when it is not present. In addition, with data generated through the latent Gaussian model, the algorithm does well in identifying all parameters. However, the latent Gaussian model cannot correspond to all rating distributions. In the cases in which it does not, the accuracy of the estimated sequential anchoring effect, w, will likely depend on how closely the model corresponds to the data.

We fit our model to the data for six payment characteristics from the 2012 SCPC and found evidence of sequential anchoring in all six cases. The quality of fit of the latent Gaussian model varied across payment characteristics as did the estimates of the sequential anchoring effect. We expect the magnitude of the effect to depend on the topic, so it is important to be careful in generalizing our results to a broader class of surveys. Nevertheless, our results suggest that sequential anchoring is generally present and that its effects on the sample data can be significant.

It is our opinion that the potential for sequential anchoring bias is an aspect every researcher should be aware of when designing and analyzing a questionnaire. To this effect, we highly recommend the randomization of the item ordering in Likert-scale sequences. Doing so allows the researcher to test for sequential anchoring and possibly adjust certain sample statistics for its effects. If no evidence is found of sequential anchoring, the orderings can be ignored, and there is no inherent harm in the randomization. Our simulations suggest that not many orderings are necessary to determine the presence of sequential anchoring, although larger sample sizes inevitably help with parameter estimation.

Ideally, survey techniques that reduce or eliminate sequential anchoring could be developed. One option, for example, is to pose each question on a separate page or screen. However, sequential anchoring is only one of many potential context effects, and any change in the questionnaire could introduce discrepancies in the results. Experiments have shown that responses to a series of psychometric questions in web surveys, in which individuals declare the level of agreement with several similar statements on a five-point scale, tend to be more internally consistent, as measured by Cronbach's alpha (Cronbach 1951), when all questions are presented on one screen than when each question is presented on a different screen (Couper, Traugott, and Lamias 2001; Tourangeau, Couper, and Conrad 2004). At the same, Tourangeau, Couper, and Conrad (2004) found that in the one-screen survey design, respondents were more likely to ignore the reverse wording of a question, in which agreement with the statement indicates the opposite general attitude than agreement with the other statements. The desirability of either design is likely to depend on the particular topic of interest and the goals of the researcher. Analysis and interpretability of results may also be easier if the number of nonvariants is minimized. This can be attempted by including explicit instructions and live checks for nonvariant sequences in online surveys. In general, there are mixed findings on the efficacy of forewarning in reducing anchoring, with some studies finding a significant effect (Tversky and Kahneman 1974; Wilson et al. 1996) and others not (Epley and Gilovich 2005; LeBoeuf and Shafir 2009).

A natural extension of this work is to consider more complicated model structures for the latent process and the anchoring effect. Perhaps the most obvious step involves dropping the assumption that the sequential anchoring effect, w, is fixed across the individuals in the population. Allowing variation in the anchoring effect, either across classes of respondents or at the individual level, would presumably help to identify the low-variation individuals. Of course, this makes parameter estimation much more difficult and it is likely that strong assumptions about the distribution of the anchoring effects would be needed.

A Estimating w, μ

In this section, we describe the procedure for estimating the parameters w and μ conditional on Σ in the optimization procedure. For simplicity of notation, we drop the superscript (k) to indicate the estimates during the k^{th} iteration and simply denote the most recent estimates with $\hat{w}, \hat{\mu}$, and $\hat{\Sigma}$ and the expectations based on those estimates as $\hat{\mathbf{M}}_i$.

By fixing the value of Σ to the most recent estimate, $\hat{\Sigma}$, it becomes conceptually straightforward to update estimates of w and μ. The reason for this is that for a given value of w the corresponding value of μ that optimizes the expected value of the full data log-likelihood (7) is easy to calculate. It is therefore helpful to view the maximum likelihood estimate of μ as a function of w: $\hat{\mu}(w)$. From (7) it is clear that for a given value of w, the estimate of μ will be

$$\hat{\mu}(w) \quad = \quad \frac{1}{N} \sum_{i=1}^{N} O_i^T W^{-1} \hat{\mathbf{M}}_i. \tag{9}$$

The only difficulty in evaluating (9) lies in calculating $\hat{\mathbf{M}}_i$. While it is easy to determine $\mathrm{E}\left[Y_{ij} \mid R_{ij}, o_i, \theta\right]$, it is considerably less so to determine $\mathrm{E}\left[Y_{ij} \mid \mathbf{R}_i, o_i, \theta\right]$ for all $j = 1, \ldots, J$. A conceptually simple way to calculate these expectations is to rely on a Gibbs sampler to draw from the distribution of $\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \theta$ by repeatedly sampling Y_{ij} conditional on the $Y_{ij'}$ for $j' \neq j$. Taking the most recent estimates of the parameters, $\hat{\theta}$, we write $\mathbf{Y}_i \sim \mathcal{MVN}(\hat{\mu}_i, \hat{\Sigma}_i)$, where $\hat{\mu}_i = O_i^T \hat{W} O_i \hat{\mu}$ and $\hat{\Sigma}_i = O_i^T \hat{W} O_i \hat{\Sigma}\, O_i^T \hat{W}^T O_i$. In addition, let $\mathbf{Y}_{i,-j}$ represent the collection $\left\{Y_{ij'} \mid j' \neq j\right\}$. Then, in order to draw from the target distribution of $\mathcal{L}(\mathbf{Y}_i \mid \mathbf{R}_i)$, we can sequentially draw from $\mathcal{L}(Y_{ij} \mid \mathbf{Y}_{i,-j}, \mathbf{R}_i)$ for $j = 1, \ldots, J$. Because knowledge of the value of Y_{ij} supplants the information contained in the value of R_{ij}, this latter conditional distribution reduces to $\mathcal{L}(Y_{ij} \mid \mathbf{Y}_{i,-j}, R_{ij})$. Now, since $\mathbf{Y}_i$ follows a multivariate distribution, it is known that $Y_{ij} \mid \mathbf{Y}_{i,-j}$ follows a normal distribution as well with mean and variance easily determined from $\hat{\mu}_i$ and $\hat{\Sigma}_i$. Sampling from $Y_{ij} \mid \mathbf{Y}_{i,-j}, R_{ij}$, then, involves sampling from a truncated normal distribution. By proceeding in this way for all j, conditioning on the current draws of $\mathbf{Y}_{i,-j}$, convergence of the Markov Chain assures draws from $\mathbf{Y}_i \mid \mathbf{R}_i$. Taking the sample averages produces estimates of $\mathrm{E}\left[\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \theta\right]$.

However, running this type of procedure for each individual is relatively time consuming, so we estimate the expectations with a variant of the above Gibbs sampler. We find that our simplification produces good results, while speeding up the optimization procedure considerably. To begin, define

$$T(r, m, s) \quad = \quad \left[\int\limits_{\ell(r)}^{u(r)} \frac{1}{\sqrt{2\pi} s} \exp\{-\frac{(x-m)^2}{2s^2}\} \mathrm{d}x \right]^{-1} \int\limits_{\ell(r)}^{u(r)} x \frac{1}{\sqrt{2\pi} s} \exp\{-\frac{(x-m)^2}{2s^2}\} \mathrm{d}x$$

to be the expectation of a Gaussian random variable with mean m and variance s^2 conditional on taking a value in $[\ell(r), u(r)]$. The algorithm we adopt for calculating $\hat{\mathbf{M}}_i$ is as follows.

(i.) For $j = 1, \ldots, 8$, let $\hat{M}_{ij} = T(r_{ij}, \hat{\mu}_{ij}, \hat{\sigma}_{ij})$ where $\hat{\mu}_{ij}$ is the j^{th} element of $\hat{\mu}_i$ and $\hat{\sigma}_{ij}$ is the square root of the j^{th} element in the diagonal of $\hat{\Sigma}_i$.

(ii.) For $j = 1, \ldots, 8$ do:

a. Calculate $m_{ij} = \mathrm{E}\left[Y_{ij} \mid Y_{ij'} = \hat{M}_{ij'}\ \forall j' \neq j\right]$ and $s_{ij}^2 = \mathrm{Var}\left[Y_{ij} \mid Y_{ij'} = \hat{M}_{ij'}\ \forall j' \neq j\right]$.

b. Let $\hat{M}_{ij} = T(r_{ij}, m_{ij}, s_{ij})$.

(iii) Repeat step (ii.) until the $\hat{\mathbf{M}}_i$ converge.

Essentially, we continue to update the expected value of Y_{ij} conditional on the most recent estimates of the other $Y_{ij'}, j' \neq j$ and the given range of Y_{ij} prescribed by r_{ij}. The equilibrium point will correspond to $\mathrm{E}[\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \hat{\theta}]$.

For any pair $(w, \hat{\mu}(w))$ and for our assumed covariance matrix $\hat{\Sigma}$, we can compare the quality of fit by evaluating the observed data likelihood function $\mathrm{lik}(w, \hat{\mu}(w), \hat{\Sigma} \mid \mathbf{R}, \mathbf{o})$ as given by (3) and (4). For N around $3{,}000$ evaluating this likelihood takes around 40 seconds in R when done sequentially and can be sped up through parallelization. Most importantly, doing so allows us to avoid calculating $\hat{\mathbf{S}}_i$. Because with fixed Σ, the likelihood function is effectively determined by the choice of w, we perform a Golden Section search algorithm over $w \in [0, 1]$ and update $\hat{w}, \hat{\mu}$ to the pair $(w, \hat{\mu}(w))$ that has the lowest negative log-likelihood for the most recent estimate $\hat{\Sigma}$. Once we have updated our estimates of w and μ, we proceed to updating the estimate of Σ.

B Estimating Σ

In this section, we describe the adopted procedure for updating the estimate of the covariance matrix Σ in a given iteration of the optimization procedure. Again, we drop the superscript (k) to indicate the estimates during the k^{th} iteration and simply denote the most recent estimates with $\hat{w}, \hat{\mu}$, and $\hat{\Sigma}$ and the expectations based on those estimates as $\hat{\mathbf{M}}_i$. To avoid the calculation of $\hat{\mathbf{S}}_i$, we proceed by a Monte Carlo-based methodology in which we simulate possible vectors $\mathbf{Y}_i$ conditional on the observed $\mathbf{R}_i$ and the most recent parameter estimates, $\hat{\theta}$. Based on the sampled vectors, we can estimate Σ directly from the full-data negative log-likelihood (6). If this candidate for Σ, along with $\hat{w}, \hat{\mu}$, proves a better fit to the observed-data likelihood (3) than the current estimate, we update our estimate and continue with the algorithm. If not, we simply draw a new set of potential $\mathbf{Y}_i$ and generate a new estimate of Σ.

Within each iteration, we continue to sample $\mathbf{Y}_i$ conditionally on $\mathbf{R}_i$ until we find an improved estimate of Σ or until we have generated some threshold number of replicates without having improved the likelihood, in which case we simply keep our current estimate $\hat{\Sigma}$. Similar to simulated annealing procedures, in iteration k, we can choose to draw $n_k \geq 1$ independent samples of $\mathbf{Y}_i$ for each individual i. By having n_k increase with k, we decrease the variability in the sample covariance matrix, thus narrowing the space over which we are effectively searching. Below we provide details of the estimation of Σ, but for simplicity of notation we assume $n_k = 1$.

We refer to the randomly drawn values of $\mathbf{Y}_i \mid \mathbf{R}_i, o_i, \hat{\theta}$ as $\mathbf{Y}_i^*$ and consider the conditional negative log-likelihood of Σ given $\hat{\mu}$ and $\hat{w}$. This function takes the form

$$\mathrm{nll}(\Sigma \mid \hat{w}, \hat{\mu}, \mathbf{Y}^*, \mathbf{o}) \quad \propto \quad \sum_{i=1}^{N} \left\{ \log|\Sigma| + \left(O_i \mathbf{Y}_i^* - \hat{W} O_i \hat{\mu}\right)^T \hat{W}^{-T} O_i \Sigma^{-1} O_i^T \hat{W}^{-1} \left(O_i \mathbf{Y}_i^* - \hat{W} O_i \hat{\mu}\right) \right\}.$$

By letting $\mathbf{Z}_i^* = O_i \mathbf{Y}_i^* - \hat{W} O_i \mu$ we can simplify this expression to

$$\mathrm{nll}(\Sigma \mid \hat{w}, \hat{\mu}, \mathbf{Y}^*, \mathbf{o}) \quad \propto \quad N \log|\Sigma| + \mathrm{tr}\left[\Sigma^{-1} \sum_{i=1}^{N} O_i^T \hat{W}^{-1} \mathbf{Z}_i^* \mathbf{Z}_i^{*T} \hat{W}^{-T} O_i\right]. \tag{10}$$

The expression in (10) is simply the negative log-likelihood from a multivariate normal distribution with mean zero and variance Σ of a sample of N iid vectors whose sample covariance matrix is given by

$$\hat{\mathbf{C}} \quad = \quad \frac{1}{N} \sum_{i=1}^{N} O_i^T \hat{W}^{-1} \mathbf{Z}_i^* \mathbf{Z}_i^{*T} \hat{W}^{-T} O_i.$$

Therefore, the maximum likelihood estimate will be given by $\hat{\mathbf{C}}$.

C Algorithm Details

In this section we provide some details about several aspects of the optimization procedure. Perhaps the most important aspect not already discussed is the generation of the starting parameters, especially those of $\hat{\mu}^{(0)}$ and $\hat{\Sigma}^{(0)}$. We determine our starting values by sampling $\mathbf{Y}$ via a simulation. To do so, we take advantage of the fact that the number of item orderings relative to the number of respondents is small. For each $o \in \mathcal{O}$, we use the subset of the sample assigned to this ordering to estimate

$$\nu_o = \mathrm{E}\left[\mathbf{Y}_i \mid o_i = o\right] \quad \text{and} \quad \Omega_o = \mathrm{Var}\left[\mathbf{Y}_i \mid o_i = o\right].$$

The parameters ν_o, Ω_o do not depend on w, which means they can be estimated from the observed ratings without considering the anchoring effect.

The procedure begins by estimating ν_{oj} and Ω_{ojj} for $j = 1, \ldots, 8$, representing the mean and variance of $Y_{ij} \mid o_i = o$. This can be done by optimizing the marginal likelihood of the ratings. Thus, let N^o_{jk} represent the number of individuals with ordering o to rate question j with $k = 1, \ldots, 5$. Then, the likelihood of $N^o_j = \{N^o_{j1}, \ldots, N^o_{j5}\}$ for a given mean and standard deviation is given by

$$\mathrm{ll}(m, s \mid N^o_j) \quad = \quad \sum_{k=1}^{5} N^o_{jk} \log p_{ojk}, \tag{11}$$

where $p_{ojk} = \int_{\ell(k)}^{u(k)} \frac{1}{\sqrt{2\pi}s} \exp\{-\frac{1}{2s^2}(x-m)^2\}dx$. It is relatively straightforward to find the values of (m, s) that maximize (11) through numerical optimization techniques. We call these estimates $\hat{\nu}_{oj}$ and $\hat{\Omega}_{ojj}$.

Once we have $\hat{\nu}_{oj}$ and $\hat{\Omega}_{ojj}$ for all $o = 1, \ldots, 6$ and $j = 1, \ldots, 8$, we consider each pair of instruments in order to estimate the covariances conditional on these estimated means and variances. Thus, let $\Omega_{ojj'}$ represent $\mathrm{Cov}\left(Y_{ij}, Y_{ij'} \mid o_i = o\right)$. Again, we can evaluate the likelihood by considering $N^o_{jj'kk'}$ to be the number of individuals with order o who rated question j with rating k and question j' with rating k'. The collection of all pairs of ratings for a pair of questions is called $N^o_{jj'}$. The likelihood then can be written as

$$\mathrm{ll}(\rho \mid \hat{\nu}_{oj}, \hat{\nu}_{oj'}, \hat{\Omega}_{ojj}, \hat{\Omega}_{oj'j'}, N^o_{jj'}) \quad = \quad \sum_{k=1}^{5}\sum_{k'=1}^{5} N^o_{jj'kk'} \log\left[p_{ojj'kk'}(\rho \mid \hat{\nu}_{oj}, \hat{\nu}_{oj'}, \hat{\Omega}_{ojj}, \hat{\Omega}_{oj'j'})\right],$$

where

$$p_{ojj'kk'}(\rho \mid m_j, m_{j'}, s_j, s_{j'}) \quad = \quad \int\limits_{\ell(k)}^{u(k)} \int\limits_{\ell(k')}^{u(k')} \frac{1}{2\pi|S|^{-\frac{1}{2}}} \exp\left(x-m\right)^T S^{-1}(x-m)\mathrm{d}x$$

for

$$m = \begin{bmatrix} m_j \\ m_{j'} \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} s_j^2 & s_j s_{j'} \rho \\ s_j s_{j'} \rho & s_{j'}^2 \end{bmatrix}.$$

Again, we rely on numerical techniques to estimate the optimal correlation, $\hat{\rho}_{ojj'}$, and use this to estimate $\Omega_{ojj'} = \hat{\rho}_{ojj'}\sqrt{\Omega_{ojj}\Omega_{oj'j'}}$. We thus have a ready estimate of ν_o and Ω_o (Ω_o is not guaranteed to be positive-definite, but if necessary one can impose this condition by manipulating the eigenvalues). As a result, for each individual with ordering o, we draw $\mathbf{Y}_i^*$, the anchor-effected latent variables conditional on that individual's observed ratings, $\mathbf{R}_i$, and the estimated moments ν_o, Ω_o. Repeating this exercise for every ordering, we have a simulated version of $\mathbf{Y}_i$ for every individual. Given this supposed sample of the underlying variables, it is simple to find the optimal values of w, μ, and Σ without having to rely on the EM algorithm or Monte Carlo sampling. Instead, for a given choice of w, maximum likelihood estimates of the mean and covariance function will be given by

$$\hat{\mu}(w) = \frac{1}{N}\sum_{i=1}^{N} O_i^T W^{-1}\mathbf{Y}_i^* \quad \text{and } \hat{\Sigma}(w) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} O_i^T \hat{W}^{-1}\mathbf{Z}_i^*\mathbf{Z}_i^{*T}\hat{W}^{-T}O_i$$

for $\mathbf{Z}_i^* = O_i\mathbf{Y}_i^* - WO_i\hat{\mu}(w)$. We evaluate the $\left(w, \hat{\mu}(w), \hat{\Sigma}(w)\right)$ for a series of different values of w, and the triplet that maximizes the likelihood $\mathcal{L}\left(\mathbf{Y}_1^*, \ldots, \mathbf{Y}_N^* \mid w, \hat{\mu}(w), \hat{\Sigma}(w)\right)$ is chosen as the starting value.

Once the algorithm is running, there are many ways to declare convergence to a minimum. Our stopping time is a function of n_k, or the number of independent samples of $\mathbf{Y}_i$ drawn in the Monte Carlo-based search for Σ. We begin with $n_k = 1$ and increase to $n_k = 2$ only if an improved estimate of Σ was found or if in 300 consecutive draws of $\{\mathbf{Y}_i\}$, no better estimate was found. Afterwards, an increase in n_k occurs if an improved estimate of Σ was found or if 100 consecutive draws failed to produce a better fit to the likelihood. The increase is such that $n_{k+1} = \mathrm{int}(1.5n_k)$, where the function $\mathrm{int}(\cdot)$ represents the integer part of any number. Once n_k becomes greater than 50, we stop the entire algorithm if three consecutive values of n_k have failed to produce an improvement. Overall, we found this decision process to be robust.

The algorithm itself is run in the software package R. Random samples from the truncated multivariate Normal distribution were made through calls to the rtmvnorm function in the tmvtnorm package, while the graphical lasso algorithm was conducted via the glasso library (Friedman, Hastie, and Tibshirani 2008). In order to speed up the optimization process, we relied on the snowfall and snow libraries in order to parallelize the evaluation of the observed-data negative log-likelihood function (3) and the calculation of $\hat{\mathbf{M}}_i$.

D Estimating Σ with LASSO Penalty

We briefly consider the effects of imposing a LASSO penalty on the elements of Σ^{-1} when estimating the covariance matrix Σ. It is well documented that including a penalty proportional to the L_1 norm of Σ^{-1}, or equivalently proportional to the sum of the absolute value of the elements in Σ^{-1}, in the objective function will have the effect of driving certain elements of Σ^{-1} to zero (Friedman, Hastie, and Tibshirani 2008). Sparse inverse covariance matrices, often called precision matrices, in turn represent a fundamental change in the dependence structure of the variables in question. Specifically if the $(j, j')^{th}$ element of Σ^{-1} is zero then this means that conditional on all other X_{ik}, $k \neq j, j'$, X_{ij} and $X_{ij'}$ are independent.

Such a covariance structure, while not desirable for all cases, certainly seems plausible for some Likert-scale sequences. Such a sequence often involves items that are inherently related, and it is often the case that the nature of these relations are reflections of general attitudes towards broader classes of the items. Homogeneity within the broader classes but independence across them would lead to a sparse precision matrix. For example, in the SCPC data, it is possible that an individual's attitudes toward the convenience of payment instruments can be deconstructed into attitudes about the convenience of the three general groups of instruments. In addition, penalties for sparsity in the precision matrix have been imposed in cases where inference about Σ is limited by the number of observed data (N) relative to the dimension of the covariance matrix (J) (Huang et al. 2006). This suggests that our model is useful for identifying associations in ratings of different items even when there are many items and the sample is relatively small. In the case of the SCPC data, N is significantly greater than J, so it is unlikely that LASSO penalties will be necessary.

To invoke the LASSO penalty in the optimization procedure, we write the conditional negative log-likelihood function for Σ as

$$\text{nll}(\Sigma \mid \hat{w}, \hat{\mu}, \mathbf{Y}^*, \mathbf{o}) \quad \propto \quad N \log|\Sigma| + \text{tr}\left[\Sigma^{-1}\hat{\mathbf{C}}\right] + \lambda\|\Sigma\|_1,$$

where $\|\cdot\|_1$ represents the L_1 norm of Σ and $\lambda \geq 0$. As λ increases, the degree of shrinkage increases and a value of $\lambda = 0$ corresponds to the estimate $\hat{\Sigma} = \hat{\mathbf{C}}$. For a given value of λ, determining the optimal estimate $\hat{\Sigma}$ is a well-studied and can be determined by the graphical lasso algorithm for one (Friedman, Hastie, and Tibshirani 2008).

To test the effect of the LASSO penalty on parameter estimations, we compare the results of the

LASSO-based estimates of Σ with $\lambda = 0.05$ to those with no LASSO penalty, but only for the Simulations A and D with paradigm $N = 3{,}000, \#_o = 6$. This choice of λ is somewhat arbitrary, but our goal is not to find the optimal value of λ, but simply to get a sense of how the sparsity constraint on the precision matrix influences the results. Because the underlying process and the paradigm are the same, we gain extra power by being able to compare the fit with and without the sparsity constraint for the same samples. Therefore, for each simulation, we estimate μ and Σ twice, once with and once without the LASSO penalty.

For each estimate we can compare $\mathrm{KL}(\mu, \Sigma, \hat{\mu}, \hat{\Sigma})$, shown in Figure 8. It is fairly clear that there does seem to be a gain in the accuracy of the covariance estimate for Simulation A, with the average divergence being 0.052 without the penalty and 0.047 with the penalty. In addition, the LASSO penalty decreased the Kullback-Leibler divergence in nine out of 10 simulations. There was no such pattern for Simulation D. A look at the different covariance matrices in Simulations A and D, as shown in Figure 3, suggests a reason for this result. The covariance matrix in Simulation A is sparse, with three independent blocks, while that in Simulation D is not so. As the LASSO algorithm was specifically designed for sparse covariance matrices, it is no surprise that it performs better in our algorithm.

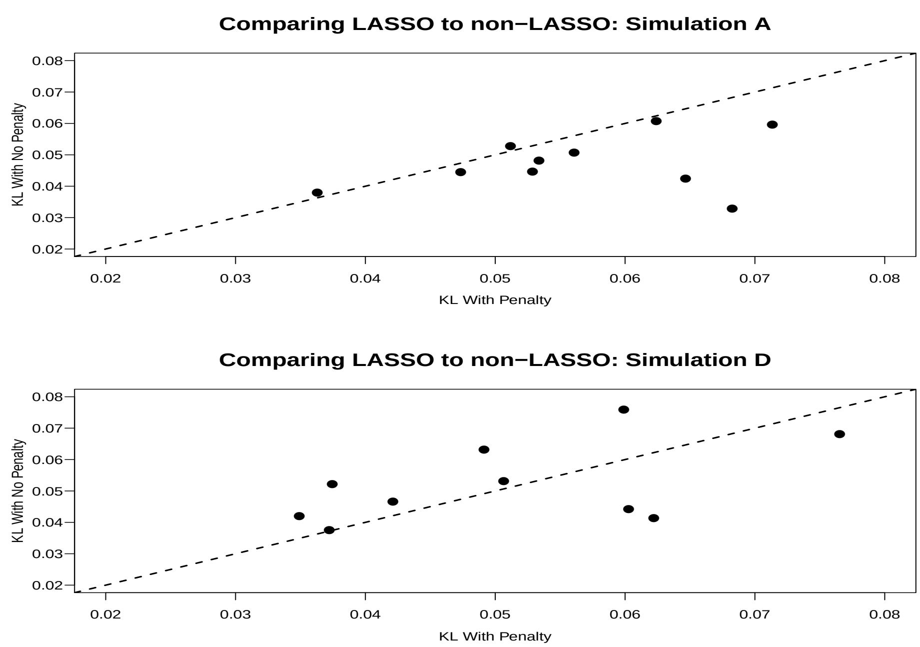


Figure 8: Kullback-Leibler divergences of parameter estimates with and without the LASSO penalty for each of 10 simulated datasets for Simulation A and Simulation D under paradigm ($N = 3{,}000, \#_o = 6$).

References

Agresti, Alan. 2002. *Categorical Data Analysis*. New York, NY: Wiley, 2 ed.

Booth, James T. and James P. Hobert. 1999. "Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm." *Journal of the Royal Statistical Society. Series B* 61(1):265–285.

Chapman, G. B. and E. J. Johnson. 1994. "The Limits of Anchoring." *Journal of Behavioral Decision Making* 7:223–242.

Chapman, Gretchen B. and Eric J. Johnson. 1999. "Anchoring, Activation and the Construction of Values." *Organizational Behavior and Human Decision Processes* 79:1–39.

Clogg, Clifford C. 1979. "Some Latent Structure Models for the Analysis of Likert-type Data." *Social Science Research* 8:287–301.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65(2):230–253.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16(3):297–334.

Daamen, Dancker D. L. and Steven E. de Bie. 1992. "Serial Context Effects in Survey Interviews." In *Context Effects in social and psychological research*, edited by Norbert Schwarz and Seymour Sudman. Springer-Verlag, 97–113.

Dawes, John. 2008. "Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales." *International Journal of Market Research* 50(1):61–77.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39(1):1–38.

Enders, Craig K. 2010. *Applied Missing Data Analysis*. New York, New York: Guilford Press.

Epley, Nicholas and Thomas Gilovich. 2005. "When Effortful Thinking Influences Judgmental Anchoring: Differential Effects of Forewarning and Incentives on Self-Generated and Externally Provided Anchors." *Journal of Behavioral Decision Making* 18:199–212.

Friedman, Hershey, Paul Herskovitz, and Simcha Pollack. 1994. "Biasing Effects of Scale-Checking Style in Response to a Likert Scale." *Proceedings of the American Statistical Association Annual Conference: Survey Research Methods* :792–795.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. "Sparse Inverse Covariance Estimation with the Grahical Lasso." *Biostatistics* 9(3):432–441.

Furnham, Adiran and Hua Chu Boo. 2011. "A Literature Review of the Anchoring Effect." *The Journal of Socio-Economics* 40:35–42.

Galinsky, Adam D. and Thomas Mussweiler. 2001. "First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus." *Journal of Personality and Social Psychology* 81:657–669.

Huang, Jianhua Z., Naiping Liu, Mohsen Pourahmadi, and Linxu Liu. 2006. "Covariance Matrix Selection and Estimation via Penalised Normal Likelihood." *Biometrika* 98:85–98.

Kirkpatrick, Scott, C. Daniel Gelatt, and Mario P. Vechhi. 1983. "Optimization by Simulated Annealing." *Science* 220:671–680.

Kullback, Solomon. 1959. *Information Theory and Statistics*. New York, NY: Wiley.

LeBoeuf, Robyn A. and Eldar Shafir. 2009. "Anchoring on the "Here" and "Now" in Time and Distance Judgments." *Journal of Experimental Psychology* 35:81–93.

Likert, Rensis. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 140:1–55.

Mason, Robert, John Carlson, and Roger Tourangeau. 1994. "Contrast Effects and Subtraction in Part-Whole Questions." *The Public Opinion Quarterly* 58:569–578.

Masters, Geofferey N. 1985. "A Comparison of Latent Trait and Latent Class Analyses of Likert-Type Data." *Psychometrika* 50(1):69–82.

McCullagh, Peter. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society, Series B* 42:109–142.

McCullagh, Peter and John A. Nelder. 1989. *Generalized Linear Models*. London, UK: Chapman and Hall.

McCulloch, Charles E. 1994. "Maximum Likelihood Variance Components Estimation for Binary Data." *Journal of the American Statistical Association* 89:330–335.

Press, William H., Saul A. Teukolsky, William T. Vetterlin, and Brian P. Flannery. 2007. *Numerical Recipes: The Art of Scientific Computing*. New York, New York: Cambridge Univeristy Press, 3 ed.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York, New York: Wiley and Sons.

Samuelson, Paul. 1938. "A Note on the Pure Theory of Consumers' Behaviour." *Economica* 5:61–71.

Schuh, Scott and Joanna Stavins. 2010. "Why Are (Some) Consumers (Finally) Writing Fewer Checks? The Role of Payment Characteristics." *Journal of Banking and Finance* 34(8):1745–1758.

Schuman, Howard. 1992. "Context Effects: State of the Past/State of the Art." In *Context Effects in Social and Psychological Research*, edited by Seymour Sudman Norbert Schwarz. Springer-Verlag, 5–20.

Schuman, Howard and Stanley Presser. 1996. *Question and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage Publications.

Schwarz, Norbert and Hans-J. Hippler. 1995. "Subsequent Questions May Influence Answers to Preceding Questions in Mail Surveys." *The Public Opinion Quarterly* 59(1):93–97.

Schwarz, Norbert, Barbel Kanuper, Hans-J. Hipler, Elisabeth Noelle-Neumann, and Leslie Clark. 1991. "Numeric Values May Change the Meaning of Scale Labels." *The Public Opinion Quarterly* 55(4):570–582.

Stavins, Joanna. 2013. "Security of Retail Payments: Do We Meet the New Strategic Objective?" Tech. rep., Consumer Payment Research Center, Federal Reserve Bank of Boston.

Strack, Fritz and T. Mussweiler. 1997. "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility." *Journal of Personality and Social Psychology* 73:437–446.

Tourangeau, Roger, Mick P. Couper, and Federick Conrad. 2004. "Spacing, Position, and Order: Interpretive Hueristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68(3):368–398.

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

Tversky, Amos and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131.

Wilson, Timothy D., Christopher E. Houston, Kathryn M. Etling, and Nancy Brekke. 1996. "A New Look at Anchoring Effects: Basic Anchoring and its Antecedents." *Journal of Experimental Psychology: General* 125:387–402.