MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Modeling Scale Attraction Effects: An Application to Charitable Donations and Optimal Laddering

Kee Yeun Lee
Polytechnic University
Hong Kong

Fred Feinberg
Stephen M. Ross School of Business
University of Michigan

UNIVERSITY OF MICHIGAN

Modeling Scale Attraction Effects:
An Application to Charitable Donations and Optimal Laddering

Kee Yeun Lee and Fred M. Feinberg*

Draft: June, 2012

(Invited for 2nd round review, Journal of Marketing Research)

Modeling Scale Attraction Effects:

An Application to Charitable Donations and Optimal Laddering

Abstract

Charities seeking donations usually employ an "appeals scale," a set of specific suggested amounts presented directly to potential donors. Choosing them well is crucial: if charities select overly high scale points, they risk their being ignored or even alienating donors and receiving nothing, while overly low scale points may encourage donors to give less than they'd have otherwise. Despite their widespread use, little is known about the degree to which the points on such scales affect both whether a donation is made and, if so, its size. Using panel data from a 3.5 year quasi-experiment, we develop a joint model accounting for both donation incidence and amount. The model incorporates heterogeneity across donors in both scale attraction effects and in donation patterns (e.g., seasonality), and allows tests of distinct operationalizations of internal and external reference price theories. Results suggest that scale points do exert substantial attraction effects; that these vary markedly across donors; that donors are more easily persuaded to give less than more; and that there are strong seasonal donation patterns in giving. A significantly negative correlation between error terms in (latent) donation propensity and (observed) donation amount highlights the importance of accounting for selectivity effects. We illustrate the framework with a speculative application to "laddering": how much charities should increase amounts subsequently requested of individual donors, based on their donation histories.

1 Introduction

Solicitations for charitable donations are a part of everyday life, with requests being made at physical locations (stores, workplaces), through the mail, various media (radio, television), and increasingly via the Internet (e-mail, websites, social networks). The Center on Philanthropy at Indiana University (2010) recently listed over 1.2 *million* charitable organizations in the United States alone, as of 2009.[1] The total amount of giving in the US has increased rapidly over the past decade, with $303B in 2009, an 80% inflation-adjusted increase over 20 years earlier. These donations account for 2.1% of U.S. GDP, a quantity ahead of all but 3 corporations in revenues and all in profits.

Private citizens have been generous to charities, even during the recent economic downturn, with 65% of US households donating per annum, $1,940 on average (including non-donors). Household-level donation, $227.4B in 2009, accounts for 75% of the U.S. total, followed by foundations ($38.4B; 13%), bequests ($23.8B; 8%), and corporations ($14.1B; 4%).

Potential donors are typically presented with an "appeals scale", a list of suggested amounts or scale points selected by fundraisers. Figure 1 presents three such scales, used for recent funding drives by Wikipedia, the United Way, and the U. N. Foundation. Each features the most common sort of appeals scale: a series of specific donation amounts, along with an "open" category. The appeals scale serves several functions, but its main role is to provide concrete anchors to help donors select an appropriate quantity; donors can, of course, also choose to give nothing, or some amount not listed on the scale, including amounts outside the range of listed values.

[**Figure 1 about here**]

Holding aside questions involving the design of an entire *scale*, an immediate practical concern for fundraisers is simply about how much to ask for: too little, and a donor may be more likely to give, but to give less; too much, and a donor may fail to be influenced by the request, or simply not donate at all. Charities wish to maximize donations, and so must attempt to tailor their requests to avoid asking for inappropriate or suboptimal donation amounts.

---

[1]2009 is the last year for which comprehensive statistics are presently available, so it is adopted consistently for comparison purposes.

Despite their ubiquity in charitable requests and fundraising, there is neither theory nor a body of empirical findings on whether and to what degree such requests, and the scales comprising them, affect individual donor behavior. As a result, fundraisers have little rigorous guidance in assessing and optimizing their appeal requests, instead falling back on prior experience, coupled with summary metrics arising from trial and error (which, as we shall see, can produce misleading or even null results). Part of the problem in providing such guidance is the need for household-level, longitudinal data on both charitable requests and outcomes - "whether" and "how much" - which charities typically possess, along with a (suitably heterogeneous) statistical model for scale attraction effects, which they typically do not.

Some of these issues have been addressed in prior literature, for example, Desmet and Feinberg (2003) and De Bruyn and Prokopec (2011), each of whom examined scale effects statistically via recourse to both internal (donors' latent, planned amounts) and external (how much one is asked for) reference points (Mayhew and Winer, 1992; Mazumdar and Papatla, 2000). Although both detected scale-based effects, neither was able to incorporate heterogeneity (the basis of individually-tailored appeals), seasonal variation in giving (which is pronounced in our empirical application), nor simultaneously account for whether and how much to give, which can lead to selection biases (Van Diepen, Donkers, and Franses, 2009; Wachtel and Otter, 2011). In this paper, we resolve these and several other issues via a novel model that measures *individual-level* scale attraction effects. The model, which builds upon a classic Type 2 Tobit formulation, is calibrated on donation history panel data from a French charity.

The remainder of the paper is organized as follows. We first provide a concise overview of prior literature on scale attraction, donation behavior, reference effects, and related areas. We then describe our empirical application, develop the model, and present both empirical results and model comparisons. An illustrative simulation exercise examining the effect of tailored appeals scales is followed by potential limitations and associated future research.

2   Literature Review

The contextual effects of scaling on responses have been intensively examined in social psychology over the past two decades. Schwarz (1999)'s comprehensive review suggests that features of research instruments - question wording, format, and scaling, among others - can substantially affect respondents' self-reported behaviors and attitudes. In particular, response scales presented to respondents are far more than a simple "measurement device," but can work as reference frames that directly influence respondents' judgments (Schwarz et al., 1991).

Researchers working in the area of social norms have found them to systematically influence human behaviors. Individuals seek out social norms to better understand or more effectively react to social situations they encounter, especially under high uncertainty (Cialdini and Goldstein, 2004). Fisher and Ackerman (1998) support this "normative" perspective in their studies on volunteerism, and several studies have examined the effects of social information on donation behavior specifically. It has long been observed that manipulating such information (i.e., what *other* donors gave) can strongly affect donation behaviors (Reingen, 1982); Shang, Reed, and Croson (2008) and Shang and Croson (2009) found exactly this in a field test for a national radio fundraising campaign. When other donors' behavior is not disclosed during a donation appeal (which is typical), respondents are more uncertain in deciding a donation amount, so a given set of response alternatives - an appeals scale - can provide contextually normative information via the location (i.e., distribution) of its scale points (Schwarz et al., 1991).

Many studies have addressed charitable donations directly, and examined the role of request size on donation behavior (amount and compliance) in laboratory and field data (Doob and McLaughlin, 1989; Fraser, Hite, and Sauer, 1988; Schibrowsky and Peltier, 1995; Weyant and Smith, 1987). Although contexts and methods vary across them, these studies largely confirm scale manipulation effects, yet differ as to whether they affect donation likelihood, donation amount, or both (see De Bruyn and Prokopec, 2011, for review). These differences may have originated from variations in compliance techniques, solicitation methods, and the suggested donation amounts. One particularly compelling potential source for inconsistencies across prior studies is lack of an account of *internal* referents. In the words of De Bruyn and Prokopec (2011), "... most fundraising

research to date has overlooked the crucial role of a donor's internal reference point in moderating the impact of appeals scales on behavior." In marketing specifically, reference price theory has been a cornerstone of consumer behavior research, and supported empirically in dozens of studies (Kalyanaram and Winer, 1995, provide an extensive review).

We make especial use of one of the key findings from this literature: that two distinct kinds of reference prices play a role in choice decisions. One is *internal* reference prices, consumer-specific, memory-based amalgams of actual, recent (and "fair") prices; the other is *external* reference prices, present at the time of purchase. It is well-known that both internal and external reference points play a role in consumer purchase decisions (Mayhew and Winer, 1992); in donation contexts, the former is characterized by what the donor typically gives and/or *plans* to give, the latter what the donor is *asked* to. Specifically, the internal referent is an unobservable construct that must be inferred from other observable information (e.g., past donation behavior), while external referents are those presented at the time of the request via the appeals scale. Prior work in donations was unable to employ both referents, since individual-specific donation histories were lacking. Thus, researchers were unable to avail of potential donors' internal referents when designing scales for experiments. This may have led to inconsistent scale manipulation results as reported by Weyant and Smith (1987) vs. those of Doob and McLaughlin (1989). Weyant and Smith (1987) found no significant difference in the average donation *amount* between the "smaller request" and "larger request" conditions, only in donation *rate*. Assimilation-contrast theory (Sherif, Taub, and Hovland, 1958) suggests that stimuli are evaluated with regard to a point of reference based on previous experience, and so depend on a "latitude of acceptance"; Doob and McLaughlin (1989) suggest that the listed amounts in the "larger request" condition fell outside this latitude of acceptance, and so had little effect on donors. When more plausible amounts (i.e., lower) were substituted in the larger request condition, they found a significant difference in the average donation *amount*, but none in donation *rate*. In short, taking account of appropriate *internal* referents literally reversed the pattern of substantive results.

Another potential source of inconsistencies involves not accounting for heterogeneity in internal referents. Most previous studies could avail only of aggregate data (e.g., control / experimental

group, or segment level; e.g., Desmet and Feinberg, 2003) to assess the mean scale manipulation effect across conditions. Because donor-specific internal referents were unavailable, group-level may dilute the effect of scale manipulation. In this regard, De Bruyn and Prokopec (2011) were unique in having obtained each donor's last donation before the field experiment, used it a proxy for a donor's internal referent. Despite this advance, the "one shot", before-after nature of their data precludes incorporating parametric, "unobserved" heterogeneity, which likewise plagues all prior studies relying on cross-sectional data. In a similar vein, no previous study of which we are aware reflects seasonal variation in donation patterns: donors are more likely to give, and/or give more, at certain times of year, such as Christmas in the U.S.; results may therefore be sensitive to *when* data are collected, especially so for field experiments. For these and other reasons, a panel of individual donors provides by far the best platform to detect and measure scale effects. Panel data further enables us to examine donors' internal referents evolve over time, as well as provide a fully heterogeneous account of scale attraction effects. This information is critical in designing optimal, dynamic appeals for each donor separately.

Lastly, none of the studies that employed scale manipulation provided a unified account of both donation incidence and donation amount. Models should not simply presume that *whether* to donate and *how much* to donate are behaviorally or econometrically unrelated. Doing so could introduce well-known measurement errors (Heckman, 1979). An especially appealing modeling framework is afforded by a Type 2 Tobit model, which comprises two components: one accounts for selection ("did they donate?"), the other the conditional output of interest ("if so, how much?"). In marketing, Type 2 Tobit models have been deployed to analyze disparate consumer decisions making processes (e.g., Donkers et al., 2006; Ying, Feinberg, and Wedel, 2006; Van Diepen, Donkers, and Franses, 2009; Zhao, Zhao, and Song, 2009), with the degree of selectivity between incidence and amount represented by a correlation parameter. Most relevant to our research, although not involving scale manipulation specifically, Donkers et al. (2006) and Van Diepen, Donkers, and Franses (2009) used such a model in donation contexts, but with somewhat different results: Donkers et al. (2006) found a small negative correlation, while Van Diepen, Donkers, and Franses (2009) found a very large positive correlation. We return to this point later

5

when discussing our own results.

3 Data Description

Our data were provided by a French charity that conducted a large-scale field experiment as part of a national fundraising campaign. The charity holds three fund-raising drives a year, at Easter, June, and Christmas. Data were collected for three and a half years, from Easter 2000 to Easter 2003, for 10 donation appeals in total. The database contains household-level records for the appeals scale presented to donors, whether a donation was made and, if so, the donation amount. Donation appeals were made by door-to-door canvassing to "regular" donors; the charity judged regularity based on each donor's frequency (the number of donations during past two years) and recency (the number of periods since last donation). Subjects were partitioned into two groups ("levels" 1 and 2) according to their average donation amounts over the two years prior to the start of the experiment. Household-averaged donations in the level 1 and 2 groups fall within 100 FF-199 FF and 200 FF-399 FF, respectively.[2]

[**Table 1 about here**]

The charity sought to better understand the role of appeals scales in donation behavior, so manipulated it by randomly assigning respondents to receive either a "standard" or a "test" scale. The standard scale had previously been used for all subjects prior to the experiment, and thereby helps establish a baseline. Scales all consisted of five suggested amounts (e.g., 100, 150, 250, 500, 1000 FF for the standard scale), as well as an "Other" category, which allowed donations below or above all five scale points, or between any adjacent pair. The test appeals scales manipulated these five suggested amounts; these all appear in Table 1.

The charity thereby implemented a 2×2 design: (prior donation) "level 1" or "level 2" $\times$ random assignment of either a "standard" or "test" appeals scale. It is important to note that the charity was collecting real donations, and therefore did not have the luxury of 'optimally' designing

---

[2]The currency unit in the data is French Francs (FF), trading during the collection window at approximately 7 to the US dollar.

the scale for the purposes of the experiment, such as orthogonalizing, including extreme values, and the like. Thus, the points comprising the "test" scale for the level 2 (higher prior) donation group were higher than those used in the test scale for the level 1 group. This 'endogeneity' is a data limitation over which we had no control, and our model will take care to treat scales as a collection of anchor points, in part to mitigate this concern.

[**Table 2 about here**]

Two hundred households from each of the four groups were randomly selected for analysis. Table 2 presents descriptive statistics for each, average donation amount (per household and per occasion), and yield rate. Level 1 and 2 differ substantially in per-household and in per-occasion average donation amounts; this is unsurprising, as the baseline donation amount was used by the charity to partition donors into different levels. However, yield rates are remarkably similar across the four groups, with all between 32% and 34%. Moreover, each of the descriptive statistics - yield rate and both per-occasion and per-household amount - fails to differ across the standard and test scales, within a donation level (1 or 2). One might therefore conclude that there were *no* effects attributable to the use of the test scale. As our analysis will show, such a conclusion based on aggregated metrics is not only premature, but highly misleading.

Table 3 suggests a clear (aggregate) seasonal pattern in both yield rate and average donation amount: people give more, and more often, at Easter than during June or Christmas. The difference in yield rates is striking - approximately $\frac{3}{4}$ of respondents donate at Easter (an important holiday in France), while under $\frac{1}{4}$ do at the other times of year - and these proportions are nearly identical in the level 1 and 2 donation groups (the latter, by construction, has higher donation *amounts* across the board). Holding aside any aggregate patterns, there is nonetheless sizable variation in *household*-level donation profiles. Table 4 presents donation histories for five households from the level 1/standard scale group, for illustrative purposes; considerable heterogeneity in timing (and some in amounts) is apparent. For example, households #3, 66, and 118 seem to be a "100FF in Easter, only", a "not in June", and a "never at Christmas" giver, respectively. By contrast, household #148 has no obvious seasonal or amount pattern. It is these variations

in donation patterns - both incidence, and amount - that we will model, in order to estimate the degree of "pull" of the appeals scale, which itself will vary across donors.

[**Table 3 about here**]

4 Model Development

4.1 Internal and External Reference Points

The model hinges on two assumptions, as discussed previously: that, for a particular request, each donor has some (latent) quantity, which serves as an *internal* referent (r^I); and that the request itself provides a set of alternatives, in the form of the appeals scale, that serve as *external* referents (r^E). If an appeals scale contains multiple points, we denote the kth as $r^{E,k}$.

The internal referent admits different operationalizations; because it is unobserved, it must be inferred based on data and the model. The reference pricing literature offers several contenders; among the most common are last price paid (Krishnamurthi, Mazumdar, and Raj, 1992; Mayhew and Winer, 1992) and a (perhaps weighted) average of past prices (Lattin and Bucklin, 1989; Kalyanaram and Little, 1994; Rajendran and Tellis, 1994; Mazumdar and Papatla, 2000), and we will empirically compare them. We include two additional specifications that can account for seasonal donation variations; so, the four (donor-specific) internal reference point models estimated are: the average of all prior observed donation amounts (IR-1); the last observed donation amount (IR-2); the average observed donation amount at the same time of year (IR-3); and the last observed donation amount at the same time of year (IR-4).[3]

That the external reference points are observable might make them appear simple to model. This might be so were there only a single requested amount. But, in practice, there are usually many, and so it is unclear how they exert their "joint pull": perhaps only the extremes are noticed; or only those nearest the internal reference have any influence; or some summary measure of all points (like the average or median); or something else entirely. We therefore empirically examine

---

[3]Instead of exponential or geometric time discounting, we used simple averaging, i.e., equal weights. Given the small number of observations per donor (2.99, on average), the difference between the formulations is minor.

five such formulations, where influence is exerted: by all scale points (ER-1); by the two scale points closest (above and below) the internal referent (ER-2); by the largest and the smallest scale point (ER-3); by the median (i.e., middle) of the scale points (ER-4); or by the mean of all scale points, which itself is typically *not* a point on the scale (ER-5). We consider such a wide range of possibilities because there is no prior theory to suggest how a *group* of referents exert collective influence. In fact, we view this as among the most intriguing open questions that our data and model can help address. Note that, when multiple points are presumed to exert influence (as in ER-1, ER-2, and ER-3), we must also specify the *weight* associated with each point; we address this in detail subsequently.

4.2 Modeling Scale Attraction Effects

In the absence of any appeals scale - for example, if a potential donor is simply asked how much s/he would like to give - whether and how much is donated would be influenced by the internal referent, not any external ones. However, when presented with (the external referents of) the appeals scale, observed behavior may be affected by both the internal and external referents. One way to visualize this is that the internal referent is "pulled" by the external ones, and that these separate pulls (if indeed more than one external referent is "noticed", as in ER-1-3) can cumulate in their effects. A simple metric for how influential a scale point is its "compliance degree," which we describe next.

4.2.1 Compliance Degree

We define CD^k, "compliance degree" of the kth external reference point as the proportional increase (or decrease) in donation amount from a donor's *internal* reference point (r^I) to an *external* one ($r^{E,k}$). More formally (with DA = Donation Amount received):

$$CD^k = \frac{DA - r^I}{r^{E,k} - r^I} \tag{1}$$

For example, if a donor is "planning" to give (i.e., has an internal referent of) $100, but is asked for $101, he will be very likely to comply, in which case both the numerator and denominator are

$1 and $CD^1 = 100\%$ (the superscript "1" indicates there was just one external reference point). However, if the same donor is asked to give $200 more (i.e., $300), the donor is less likely to fully comply; if the resulting donation is instead $140, $CD^1 = (\$140 - \$100)/(\$300 - \$100) = .2$, or 20%. In simple language, the donor "came up 20%" from a $100 baseline. An analogous calculation pertains to external referents below the internal one.

It is convenient to define the *distance*, d^k, between the kth external and the internal referent as an incremental/decremental ratio.

$$distance(d^k) = \frac{\left\|r^{E,k} - r^I\right\|}{r^I} \tag{2}$$

This allows both compliance degree as well and the pulling amount (described later) to be expressed as a dimensionless quantity for each donor. This in turn helps to unify the model; for example, it can treat the response of a donor planning to give $10, but asked to donate $20, similarly to that of one planning to donate $100, but asked for $200.

We will model both upward and downward "compliance degree curves", which satisfy three properties:

1) $CD^k \approx 1$ for $d^k \approx 0$: "Maximal compliance occurs near donors' internal referents."

2) CD^k decreases monotonically in d^k: "Compliance is worse for requests further from the internal referent."

3) $CD^k \geq 0$: "Compliance can't be worse than zero."

Properties 1 and 2 suggest donation is highly responsive to asking for amounts close to what was 'planned' (the internal referent), but increasingly less so for distant amounts. This is consistent with "latitude of acceptance" in Assimilation-Contrast Theory (Sherif, Taub, and Hovland, 1958), which has found prior support in a donation context (Doob and McLaughlin, 1989). Property 3 simply suggests that requests can be ignored, but do not literally repel donors from a scale point.

There are many ways to specify compliance degree curves satisfying these three properties, including using fully parametric (e.g., polynomial), semi-parametric, or non-parametric formulations. We select a translated gamma kernel function, for two reasons. First, it provides a parsimonious, yet flexible, functional form that naturally satisfies properties 1-3; this parsimony

is important for a *heterogeneous* account to be identified, given the small number of responses per donor during the data window. Second, the gamma kernel enables the pulling *amount* curves (described later) to follow a unimodal, yet flexibly-shaped, distribution, which in turn facilitates eventual optimization. Thus, we arrive at an especially simple form:

$$CD^k = \frac{\exp\left(-\frac{d^k+1}{\theta}\right)}{\exp\left(-\frac{1}{\theta}\right)} = \exp\left(-\frac{d^k}{\theta}\right); \qquad \theta = \begin{cases} \exp(\beta^U), & r^{E,k} \geq r^I \\ \exp(\beta^D), & r^{E,k} < r^I \end{cases} \tag{3}$$

where $\theta > 0$ is the gamma kernel scale parameter.

The compliance degree curve follows from a gamma kernel with "shape parameter" 1 and "scale parameter" θ. [4] This is then both translated and normalized - first horizontally translated by -1 so that it crosses the y-axis, then normalized to have a value of 1 at the origin - after which it follows a translated gamma kernel, anchored at (0,1) with curvature determined by the scale parameter. Note that there are actually two different compliance degree curves, depending on the relative location of the internal and the external referents. When $r^{E,k} \geq r^I$, we have an "upward" compliance degree curve, and otherwise a "downward" one.

Since the scale parameter (θ) must be positive, we specify β^U or $\beta^D = \ln(\theta)$, where β^U and β^D are the "upward" and "downward" parameters in (3). Figure 2 depicts both curves, which can have a variety of shapes, for different values of β^U and β^D. However, $\beta^U = \beta^D$ does not imply identical upward and downward curves, because the domain of the downward curve is bounded by 100%, since one cannot give less than zero (i.e., a 100% decrement).

[**Figure 2 about here**]

4.2.2 Pulling Amount

The pulling amount (PA^k) represents the *size* of effect exerted by a scale point, a simple matter of multiplying compliance by the (Euclidean) distance between the internal (r^I) and the *kth*

---

[4]Fixing the shape parameter at 1 yields a non-negative, monotonically decreasing, convex curve (with regard to the origin), satisfying properties 1-3. Numerous simulations showed recovery of *two* parameters (both scale and shape) was very poor, suggesting weak identification in data generated to resemble ours.

external reference point ($r^{E,k}$):

$$PA^k = CD^k \times \left\| r^{E,k} - r^I \right\| \tag{4}$$

The pulling amount suggests a trade-off between asking for too little and asking for too much: If a charity asks too little - that is, just a bit more than the internal referent - compliance (CD^k) may be high, but the potential surplus ($\left\| r^{E,k} - r^I \right\|$) is small. On the other hand, if a charity asks too much, the compliance degree may be low, while the surplus is large. In light of this trade-off (where the extremes are literally zero), optimizing donation drives requires considering both elements, that is, asking for a judiciously chosen amount from each donor.

Each of the two compliance degree curves therefore gives rise to a "pulling amount" curve: $r^{E,k} \geq r^I$ corresponds to "upward" pulling, $r^{E,k} < r^I$ to "downward". The simple nature of (4) implies that these curves also follow a gamma kernel, with shape parameter 2 and scale parameters $\exp(\beta^U)$ and $\exp(\beta^D)$. As depicted in Figure 3, these curves can have many shapes: the upward pulling curve has domain $[0, \infty)$, is unimodal (and thus has a unique maximum), with zero at the origin and asymptoting to zero for large d (for any β^U). The domain of the downward pulling amount curve is $[0, 1]$; it is unimodal (with unique maximum) if $\beta^D < 0$, and is monotonically increasing otherwise (with maximum at 1). These internal maxima allow us to derive a closed-form expression for optimal, donor-specific scale points, discussed in the section on the effect of individually tailored appeals scales.

[**Figure 3 about here**]

4.2.3   Accumulating Scale Attraction Effects

Because real appeals scales almost always comprise multiple amounts, their effects need to be somehow combined. Figure 4 illustrates the "accumulated pulling amount" accruing from multiple external referents; to match our empirical application, five external referents are depicted, with two distinct upward and downward curves on either side of the graph. Here, scale points 1, 2, and 3 are greater than the internal referent (set by convention to $d = 0$), so each induces an upward

pull on donation amount, tending to increase it. By contrast, scale points 4 and 5 are lower than the internal referent, tending to pull the observed donation downward.

The discussion thus far concerns the pulling amount for individual scale points, not how to combine them. Just as we considered a number of specifications for the effects of the internal and external referents, we will do so for this combination. Before detailing these, we highlight one simplifying assumption: that the effect of a particular scale point can be modeled separately from the existence or the location of the others. This is dictated by a data limitation: the charity did not change scales (over the course of the experiment, nor within each of the four donation groups), so that identifying *interactions* between scale points is not possible. Even were this not the case, such interactions would greatly weaken gamma kernel parameter (β^U and β^D) identification, owing to the small number of observations per donor (and again to the lack of within-donor scale variation during the experiment).

While at first blush such independence assumptions may appear unrealistic, they are mitigated by the weighted-averaging schemes explored for the "accumulated pulling amount", or APA. We examine three: i) the sum; ii) the mean; and iii) the weighted mean of the pulling amounts. Each is described as follows, along with potential caveats. In general:

$$APA = \sum_{k=1}^{K} w^k \times I^k \times PA^k; \qquad I^k = \begin{cases} 1, & r^{E,k} \geq r^I \\ -1, & r^{E,k} < r^I \end{cases} \qquad (5)$$

$$Sum : w^k = 1; \qquad Mean : w^k = \frac{1}{k}; \qquad Weighted\ Mean : w^k = \frac{PA^k}{\sum_{k=1}^{K} PA^k} \qquad (6)$$

Sum. Simple summation appears to be the most direct way to accumulate the separate pulling amounts. However, this specification has two inherent problems. First, the predicted donation amount can lie above the largest, or below the smallest, scale point. Although this is not impossible, our data contains very few instances in which the donation exceeds the largest scale point. Second, the effect of including *additional* scale points can be overstated (something that, in our data, will not be testable, since the charity fixed this at 5). For example, given an internal referent of 50, the APA of an appeals scale with four points of 9, 11, 99, and 101 is about

twice as large as for one with two scale points of 10 and 100, which seems decidedly unrealistic.

Mean. The mean specification retains additivity and resolves the two problems with the sum, but is not without problems of its own, owing to equal-weighting. For example, if a donor is asked to give \$2000 when the planned amount is \$100, the effect of such a large scale point on APA might be small or negligible. However, equal weighting forces a large scale point like \$2000 to have tremendous effect on the APA by substantially lowering the accumulated pulling amount. A straightforward fix involves the use of a weighted mean, as follows.

Weighted Mean. This makes use of weights, w^k, which one might imagine were estimable. Two data limitations prevent this, (once again) the lack of within-donor scale variation, and that only three different appeals scales (one standard and two test) appeared in the experiment. For this reason, and because we include heterogeneity β^U and β^D, even homogeneous w^k proved impossible to estimate.[5] Therefore, the weight is set in proportion to the size of pulling amount, based both on conceptual appeal and trial of multiple alternate schemes (which we do not report here). The key point is that the weighted mean allows a scale point with a larger pulling amount to contribute more to the total pull, unlike for either of the previous two specifications.[6]

4.3 General Model(*Type 2 Tobit*)

We begin by outlining the general model structure. The model has been set up to allow a "dimensionless" account of pulling effects, so that heterogeneity can be specified across the log-scale for donation amount. As discussed earlier, we use a Type 2 Tobit model (Amemiya, 1985) to jointly account for donation incidence and amount, as follows:

---

[5]De Bruyn and Prokopec (2011) tried to estimate each scale point's weight, which they term "absolute attraction weight". However, they could estimate only the weight of the smallest of four scale points, while fixing those of the other three to 1. They attribute this identification problem to the inherent correlation across suggested donation amounts on the appeals scale (i.e., suggested amounts increase monotonically and are highly correlated).

[6]In fact, we found model with weighed mean specification (explained next) fits better than that with equal weight mean specification, keeping all other model components the same: The RMSE and MAD of the former are 0.257 and 0.188 respectively; for the latter, are 0.271 and 0.195.

$$y^{s*} = X^s \beta^s + \epsilon^s \tag{7}$$

$$y^{a*} = (r^I + APA) + X^a \beta^a + \epsilon^a, \ \ where:$$

$$y^s = 1, \ \ if \ y^{s*} \geq 0; \ \ 0 \ \text{otherwise}$$

$$y^a = y^{a*}, \ \ if \ y^s = 1; \ \ \text{unobserved otherwise}$$

$$(\epsilon^s, \epsilon^a) \sim BVN(0, \Sigma_\epsilon); \quad \Sigma_\epsilon = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$$

The subscripts i and t (for donor and time) are suppressed, and X^s and X^a are covariates in the selection (s) and amount (a) equations, respectively, which we detail below.

In the amount equation, let y^{a*} denote the log of the latent donation amount, which is observed only when a donation is made, that is, when y^s is 1, which occurs when the latent variable $y^{s*} \geq 0$. The error terms of the selection and amount equations (ϵ^s and ϵ^a) follow a bivariate normal distribution; the variance of ϵ^s is fixed to 1 for identification. It is important to note that we model the logarithm of donation amount, for several reasons: first, it allows ϵ^a to be plausibly homoscedastic; second, it allows all effects in the amount equation to enter multiplicatively; and third, it allows for coefficient heterogeneity to act on a dimensionless quantity, which we address in detail shortly.

The amount equation (for y^{a*}) contains two deterministic components. The first is the sum of a donor's internal referent (r^I) and the accumulated pulling amount (APA). The second is all factors ($X^a\beta^a$) that affect the donation, other than those stemming from the appeals scale. Scale-based effects do not appear directly in the selection equation, because in our data all scales used were set in "reasonable" ranges for every donor (recall that these were real donors, and the charity was understandably reluctant to alienate them with unrealistically high requests, or lose funds with low ones). Hypothetically, were all or many of the suggested amounts exceedingly large, it is possible that the donor would become annoyed and give nothing. Therefore, although we cannot preclude this possibility for all data settings, for ours the appeals scale can exercise

influence on donation incidence only indirectly, via the correlation, ρ. [We did estimate a model allowing for scale effects in selection; the APA coefficient in selection was ns.]

[**Table 4 about here**]

4.4 Explanatory variables and Heterogeneity

4.4.1 Explanatory variables

Selection Equation

The selection equation contains three types of explanatory variable (X^s), which we detail subsequently: seasonal indicators, (log of) prior donation, and "level" fixed effects. Table 3 reveals strong aggregate seasonal variation in donation likelihood, by far highest at Easter; Table 4 suggests household-level variation as well. Previous studies, which were mostly one-shot, could not account for such seasonal variations, which are critical in our data. Three dummies - Easter (X_{it}^E), June (X_{it}^J), and Christmas (X_{it}^C) - represent when the donation request occurred.

The log of (1+ amount the donor gave on the last request), donated X_{it}^{lag}, is included to examine carryover effects; as discussed previously, logs help retain error homoscedasticity, among other benefits. The donation amount itself is not censored, but truncated, at 0; so, X_{it}^{lag} is 0 when we observe no donation taking place. The directly prior donation (0 or otherwise) may affect the decision to donate for several reasons: these are regular donors, so are likely aware of their donation patterns, and the interval between solicitations is relatively short. [7]

Although Table 3 shows only modest differences in yield rate between the "larger" (level 2) and "smaller" (level 1) donation groups, we still include an appropriate dummy (X_i^{level}) among the selection covariates, to control for potential differences in baseline donation likelihood *after* accounting for seasonal patterns. Coefficients for the three seasonal dummies, the log-donation lag, and the level dummy, are denoted β^E, β_i^J, β_i^C, β^{lag}, and $\beta^{level,s}$, respectively.

In the experiment, donors were randomly assigned to receive either a standard or a test appeals scale, so no dummies were entered for this difference (in either selection, or amount). Doing

---

[7]We estimated a similar model by replacing the log-donation lag with an indicator variable for *whether* one in the previous period, finding poorer in-sample fit. This may be because the continuous variable (log-donation) carries additional information, compared with a simple indicator variable.

so failed to improve fit, in any case, so we do not discuss these again.

[**Table 5 about here**]

Amount Equation

Based on examination of the data and unimproved fit of models including them, seasonal dummies are not included in the amount equation; the somewhat higher amounts indicated at Easter in Table 3, for example, will be well-explained by other covariates, like lags in setting "internal" referents (such as in IR-3 and IR-4). Table 4 shows far greater household variation in when to give, not how much; and both Figure 5 and Table 5 suggest that household-level seasonal variation in amount is very small for most donors. Similarly, we do not include a lag for prior donation amount. This may seem paradoxical, but recall that there is little *within-donor* variation in observed donation amount, suggesting that that people do not say, in effect, "I give more than usual last time, so will give less this time," or vice versa. [8] Lastly, although donation amount is mainly predicted by a donor's internal referent and scale effects, a level dummy (X_i^{level}) is included to account for the difference in baseline donation amount between the two groups, denoted $\beta^{level,a}$.

[**Figure 4 and Figure 5 about here**]

4.4.2 Heterogeneity

It is critical, in a model for household-level behavior, to incorporate "unobserved" heterogeneity, which we do in several ways. Given the large household-level seasonal donation variation, we model heterogeneity in the seasonal dummies for the June and Christmas coefficients (β_i^J and β_i^C).[9] Our empirical results suggested that household-level seasonal donation patterns were well

---

[8]To verify this choice, we estimated a series of models with $\ln(r^I + APA + \beta * PriorDonation)$ instead of the analogous term in the amount equation. For all reference point models (IR1-4), AIC, BIC, and in-sample fit does not show any improvement.

[9]Extensive simulations for data matching ours in marginal (summary) statistics failed to recover the true parameters - mean vector and covariance matrix for β_i^E, β_i^J, β_i^C - when the Easter, June and Christmas coefficients were all heterogeneous. Restricting the most common donation period (Easter, with a 74.1% yield rate) to be homogeneous led to nearly perfect parameter recovery.

reflected in heterogeneity for β_i^J and β_i^C, owing perhaps to much larger variation in giving in June and at Christmas.

Importantly, since our model is primarily meant to capture scale attraction effects, the two gamma kernel parameters (β_i^U and β_i^D) in the amount equation are heterogeneous. Imposing heterogeneity on the gamma parameters - especially "upwards", β_i^U - is crucial for formulating tailored appeals scales, which require identifying the request amount with maximum effect in "pulling" up a donor's internal referent. If β_i^U were homogeneous, each donor's optimum would be the *same* percentage above his/her internal referent. This might still provide a helpful guideline for fundraisers, but presumes all donors are equally 'elastic' in being cajoled upwards. Our results, in fact, will strongly weigh against this presumption. We similarly account for heterogeneity in the "downward" parameter, β_i^D, though it will play a lesser role in optimization.

Our formulation therefore specifies four heterogeneous parameters, to be recovered from the relatively short data window of 7 occasions; the 42.7% aggregate yield rate suggests that about 3 of these 7 requests resulted in donations, on average. Although it may appear ambitious to account for 4 household-level parameters based on relatively little data, simulations showed good recovery for all four heterogeneous parameters, and excellent recovery of the others.

5 Estimation

The full model (see Appendix A) is estimated using Markov chain Monte Carlo methods. Data augmentation (Tanner and Wong, 1987) converts the model to a Bayesian Hierarchical Seemingly Unrelated Regression. We obtain posterior draws via Metropolis-within-Gibbs algorithms: Gibbs sampling (Geman and Geman, 1984) if the full conditional of a parameter block is of known form, and Metropolis-Hastings, with a random walk proposal (Chib and Greenberg, 1995), otherwise. We set diffuse priors for all parameters of interest; detailed procedures appear in Appendix B. All estimates are based on 100,000 draws. We discard the first 50,000 draws for burn-in, and use the last 50,000 (thinned to every tenth) to calculate posterior densities. Gelman-Rubin scale reduction factors, using 5 chains with different stating points, are below 1.1 for almost all parameters, suggesting good convergence (Brooks and Gelman, 1998).

6 Results

For brevity, we only present full estimation results for the model with IR-1 (average of all observed donation amounts) and ER-1 (all scale points), as these provided the best fit compared with all possible combinations of the other internal and external references point formulations (IR 2-4 and ER 2-5). Table 6 summarizes posterior means and standard errors for all parameters, and detailed model comparison statistics appear in the following section.

6.1 Error Correlation in Selection and Amount equations

The mean of the marginal posterior for the correlation (ρ) between the selection and amount equation errors is negative (-0.387), and the 95% highest density region does not include zero. This suggests that unmeasured factors influencing selection are correlated with those influencing amount, and operate in opposite directions. The size of the correlation is moderate: neither close to 0 nor to 1. This differs from findings in previous research using related model formulations; for example, Donkers et al. (2006) found the correlation to be negligible and negative (-0.033), while Van Diepen, Donkers, and Franses (2009) found it to be very large and positive (0.958). A very small correlation fails to help correct for potential selection biases, and could reflect large, independent sources of error in each equation. Conversely, a large correlation might suggest important variables omitted in both equations.

It is difficult to generalize such results, since our model accounts for scale attraction effects, while prior ones do not. We did, however, find significant, moderate, negative values of ρ across a very wide range of candidate models, indicating that selectivity needs to be accounted for in our data. One interpretation of this finding, which is apparently robust, is that, knowing one has donated, the *conditional* expectation of the donation is smaller. Thus, models that account for "whether" and "how much" *separately* may overestimate total expected yield.

6.2 Selection: Seasonality

Comparing the Easter coefficient (0.782) to the means of the (heterogeneous) June and Christmas coefficients (-0.554, -1.019 respectively) accords with the aggregate benchmark, that giving is much more likely for Easter than June or at Christmas, on average (a finding that should not be extrapolated beyond French donors to a nationwide "general purpose" charity.). There is a substantial seasonal heterogeneity: the SD of individual-level parameters for June and Christmas are 0.406 and 0.831, respectively. The high (positive) correlation between these individual-level parameters (0.742) largely reflects the fact the yield rates in June and Christmas are both low (18.1%, 20.3%) and a high proportion of donors (65.6%) gave at neither time.

6.3 Level Dummies and Lagged Log-Amount

The level dummy is only marginally significant (mean 0.073, SE 0.039) in selection, but significantly positive in amount (mean 0.273, SE 0.019). So, as aggregate statistics suggest, level 2 donors give more than those in level 1, but with large difference in yield rates. The coefficient of the log-donation lag in selection is significantly negative (-0.131), indicating that a larger donation *amount* last time leads to being less likely to give *at all* this time.

[**Figure 6 and Figure 7 about here**]

6.4 "Pulling Effects": Gamma Kernel Parameters in Donation Amount

The values of β_i^U and β_i^D determine each donor's degree of compliance ("pull") to the scale points above and below the internal referent. Because the domains of the two compliance curves differ, we should not compare β_i^U directly to β_i^D. Instead (and ignoring for the moment the considerable variation in these across donors), Figure 6 shows both compliance curves at the posterior means of β_i^U and β_i^D. The downward compliance curve is far less 'pitched' than the upward. This makes intuitive sense: asking for much more than one is willing to give will eventually result in almost zero compliance, unlike asking for much *less*.

The compliance curves are, by construction, monotonic. By contrast, the pulling *amount* curves need not be. These are depicted, at the posterior means for β_i^U and β_i^D, in Figure 7.

20

The upward pulling amount curve is inverted-U (i.e., unimodal), indicating a single "best request" value, to which we return later. By contrast, the downward curve decreases monotonically, suggesting that donors tend to give less as the suggested amount decreases (with lower bound 0).

[**Figure 8 about here**]

Figure 8 in some sense encapsulates our main results: the upward and downward pulling parameters (β_i^U and β_i^D) for each donor. There is clearly a good deal of heterogeneity, indicating differing degrees of susceptibility to the appeals scale, *despite only modest differences in prior donation behavior*. The upward pulling parameter (β_i^U) displays larger variation (SD 0.93) than the downward (SD 0.48). This might be expected: most everyone can go along with being asked for less, but people react to being asked for more very differently.

By allowing a bivariate density for (β_i^U,β_i^D), the model helps assess *overall* scale compliance. Specifically, we find a substantial correlation (0.380) in these values, suggesting that donors who are "upward compliant" tend to be "downward compliant" as well. There is no reason to expect these should be correlated at all, let alone positively, and we believe this finding to be the first of its kind. This bivariate density for (β_i^U,β_i^D) leads immediately to the joint distribution of *maximal pulling amounts*, those scale points that lead to the greatest overall effects; we do not call these "optimal", since a large *downward* pull is to be avoided.

[**Figure 9 about here**]

Heterogeneity in (β_i^U,β_i^D) leads to substantial variation in maximally effective potential scale point locations, depicted for the "upward" pull in Figure 9 (we omit the analogous "downward" distribution, as for most respondents these are zero). The model suggests that the scale point with maximal upward pull, which varies across donors, ranges from 27.0% to 198.5%, with a mean of 71.7%, above one's internal referent, which seems reasonable.[10] This non-trivial variation has an important implication: that it may be possible to substantially increase donations by *personalizing* an appeals request, based on each donor's history. We discuss this possibility later, along with associated calculations.

---

[10]Discussions with a large university's fundraising team suggested that the success of "laddering" dropped nearly to zero when appeals hit 200% above a donor's typical or last donation amount.

21

Figure 10 in some sense integrates the key elements of the model, and presents its main substantive findings in the context of the original data, specifically: How much can a maximally-effective appeal (either up or down) pull from one's reference donation? It depicts, across donors, the maximal percentage increase and decrease (see Appendix E for derivation). This also allows a direct comparison of the "strength" of upward and downward scale attraction effects, heterogeneously, which was not sensible using (β_i^U, β_i^D), given their different domains of operation. The maximum percentage increase ranges from 9.9% to 73.0% (mean = 26.3%; SD = 7.2%); maximum percentage decrease ranges from 20.5% to 89.9% (mean = 74.3%; SD = 4.8%). The correlation in these values is 0.298 (echoing the 0.380 value for β_i^U and β_i^D). Figure 10 suggests that the maximum percentage decrease is greater than the analogous increase for most donors: 81.6% of the donors lie above the diagonal (dotted) line. This is reminiscent of the asymmetric effects in Desmet and Feinberg (2003), whose lack of individual-level data precluded any distributions across donors, and De Bruyn and Prokopec (2011), who only had one-shot (i.e., "before" and "after") data unsuited to modeling heterogeneity or carryover effects.

[**Figure 10 about here**]

7   Model comparison

The data give clear indication of the existence of scale attraction effects. But one might reasonably question whether these were strongly dependent on the particular form of the model, four of its elements in particular: 1) internal reference point specification; 2) external reference point specification; 3) the importance of including correlation (Type 2 Tobit), seasonality, and scale effects; and 4) incorporating response heterogeneity. We examine each of these in some detail, to assess relative "contribution" to overall model fit.

With respect to internal reference formulation, we compare four, as described in the model development section, each donor-specific: the average of all prior donation amounts (IR-1); the last amount (IR-2); the average of all amounts at the same time of year (IR-3); and the last amount at the same time of year (IR-4).[11] We similarly examine the five external reference

---

[11]For IR-3, IR-4, if we don't observe donation at a certain time of year in the initialization period (first full year,

formulations explained earlier: all scale points (ER-1); the two scale points closest (above and below) the internal referent (ER-2); the largest and the smallest point in an appeals scale (ER-3); the median (i.e., middle) of all scale points (ER-4); and the mean of all scale points (ER-5).

We call the model with all the aforementioned components - internal and external referents; error correlation; seasonality; heterogeneity - the "full model". Alternative models include those lacking: error correlation ("no correlation"), scale effects ("no scale effect"), and both ("simple regression"). We similarly examine the effects of homogenous seasonality, homogenous scale effects, and both of these.

[**Table 7 and Table 8 about here**]

For model comparisons, owing to short donation histories (which preclude 'squandering' an entire year for prediction purposes), we compare fit in-sample. "Fit" is assessed via mean absolute deviation (MAD) and root mean square error (RMSE) for donation amount predictions, which appear in Tables 7 and 8. Table 7 shows that the proposed model (the "full" model with IR-1, ER-1 and both seasonal and scale effect heterogeneity) provides a better fit than all alternatives; moreover, including error correlation and scale effects improves fit regardless of internal reference formulation (IR1-5) and the inclusion of heterogeneity. Table 7 also allows us to judge relative contribution to overall model fit: scale effects easily best both correlation and seasonality. For example, in subtable 7-1, which incorporates heterogeneity in both seasonality and scale effects, failing to account for scaling effects inflates RMSE and MAD approximately 30%; the corresponding figure for removing correlation alone is $\approx 5\%$; and for dropping heterogeneity entirely, $\approx 20\%$. Even in the final subtable (7-4), for homogeneous seasonality and scale effects, discarding scale attraction inflates RMSE and MAD by approximately 10%; dropping correlation, about 1%. These comparisons suggest that scale attraction effects appear to explain more variation in giving than those typically modeled in prior donation research combined, although it will require additional data applications and distinct settings to verify whether this holds generally.

In terms of internal reference point specification, IR-1, the average of all prior donation amounts appeared to dominate across the board. The degree of dominance was not trivial,

---

or three data points), we initialize using the mean of the all observed amounts in each group.

sometimes hovering near 10%. To our knowledge, such a test of 'internal' referents is unique in donation contexts, and we know of no prior theory that would have anticipated it. We found as well that allowing for heterogeneous scale effects (subtables 1 and 3) provides superior fits compared with analogous homogeneous scale effects models (subtables 2 and 4). However, allowing heterogeneity in seasonality parameters offered rather small increases in fit. So, although heterogeneity is itself important, overall, it is much more so in terms of scale effects.

Based on the results of Table 7, we restrict our attention to the "full" model with IR-1, and Table 8 summarizes fits of five distinct *external* reference specifications (ER 1-5) for this model. ER-1, with all five scale points included, clearly dominates. We hesitate to term this a general finding, as the charity did deliberately choose all scale points to be "reasonable", since they were understandably more focused on revenue than testing reference point theories. Regardless, the "full" model with IR-1 and ER-1 was verified, using all discussed metrics, to provide the best fit to the data among the $2 \times 2 \times 2 \times 2 \times 4 \times 5$ (scale effects?; scale effect heterogeneity?; seasonality heterogeneity?; error correlation?; IR1-4; ER1-5) design.

To guard against concluding in favor of a potentially overparameterized model based on in-sample fit, we report log-likelihood (LL), AIC, and BIC.[12] LL is based on the Tobit 2 likelihood, calculated at parameters' posterior means. The proposed model has the largest LL and smallest AIC and BIC, consistent to the in-sample fit test result.

8 Illustrative Application: Effect of individually tailored appeals scales

We conclude with a preliminary examination of setting a superior appeal. While a laudable goal might be to devise optimal donor-specific appeals scales or even a single optimal scale applicable to the entire donor pool, the nature of the available data do not permit this, for a number of

---

[12]We do not report two common Bayesian model comparison measures, the Bayes factor and DIC, for the following reasons. The marginal likelihood (on which the Bayes factor is based) is well known to be sensitive to prior specification and calculation method. We in fact computed marginal likelihoods under several different diffuse priors and common calculation methods, finding it highly variable across them. DIC is also not appropriate here, as it has multiple plausible definitions in missing data models (Celeux et al., 2006). Parameters for AIC and BIC are based on Level-I for homogeneous parameters and Level-II for heterogeneous ones. AIC and BIC are included for comparison purposes only, not as definitive bases for Bayesian model selection.

reasons. Foremost among these was that the scales used by the charity changed neither during the course of the experiment nor across the four donation groups ([level 1, 2] × [test, standard]). Although the *existence* (and nature) of scale attraction effects was strongly verified, this lack of within-donor variation made identifying *interactions* between scale points impossible, since they were perfectly confounded with the experimental conditions themselves. Such interactions are necessary to avoid placing *all* points in a k-point scale at the same "optimal" spot, an obvious absurdity never seen in real appeals scales. Detailed simulations (available from the authors) verified these claims, which we view simply as a data limitation stemming from constraints put on the charity provider.

We instead focus on choosing a single optimum request - often referred to as an attempt to "ladder" an individual donor - and the effect of tailoring these to each donor, as opposed to using a common best ladder (increase percentage over the internal referent) for each, the ubiquitous practice among real fundraisers, even those with donation histories at their disposal. Finding a single optimal request may appear paradoxical, but we must remember that the charities that use appeals scales do *not* vary them across individuals - the very issue that provoked the modeling effort here - and moreover would never have reason to ask for *less* than an internal referent, given its small pulling effect.[13]

Given these data limitations, we conducted a simulation study involving a single request amount, and a three-period (i.e., one full year) look-ahead. We immediately discovered that seasonal patterns, which were exceedingly strong in our data, vastly swamped any benefit of potential strategies of "let's ask for less now, so they'll give somewhat less, and use that carryover to make them more likely to give again next time." It simply did not matter. Thus, the "three-period look-ahead" optimization, which involved discrete dynamic programming, gave results identical to the three "myopic one-period look-ahead" optimizations, on which we report. Our sole focus, as stated earlier, is quantifying the results of the appeal itself when it is generated

---

[13]Although neither our data nor model allows us to rectify this issue, it's unclear whether using *multiple* scale points is useful in charitable appeals, let alone how many. An appeals scale with multiple points might be beneficial when the charity cannot estimate donors' internal referents, as when soliciting from first-time donors. If there is wide variation in internal referents, a suitably-spaced appeals scale may help donors find a reasonable anchoring point. Regardless, whether it is beneficial to present multiple scale points to regular donors is an empirical question awaiting suitable data, in which the *number* of scale points is systematically varied.

heterogeneously, based on donation history, as opposed to a "one ladder fits all" strategy.

To run the simulation, the optimal suggested amount and the expected value of donation amount are calculated, based on β_U. The optimal external referent, r^{E*}, can be calculated by solving a first-order condition for the upward pulling amount, PA (see Appendix C for derivation).

$$r^{E*} = (\exp(\beta_U) + 1)r^I \tag{8}$$

The expected value of donation amount can also be calculated analytically. We use the generic symbols ν^s and ν^a for the deterministic parts in the selection and donation amount equations in (7). Because y^{a*} is a logged quantity, it needs to be exponentiated to calculate the expected donation amount, which is as follows (see Appendix D for derivation):

$$\mathbb{E}[\exp(y^a)] = \exp\left(\nu^a + \frac{\sigma^2}{2}\right) \times \Phi(\nu^s + \sigma\rho) \tag{9}$$

Three different types of appeals were used in the simulation: i) a "pseudo" appeal, with no suggested amount; ii) a *group*-level optimized suggested amount; and iii) an *individually* customized suggested amount. In the first case, where no external referent is provided, we presume donors hew to their internal referents; this thus serves as a convenient benchmark. The group-level request is the single common value that would optimize overall donations, given the heterogeneous distribution of β_i^U. Individual-level amounts are calculated from each donor's β_i^U separately. Simply put, the first scenario does not account for scale effects at all; the second type does, but presumes a common "best request" for everyone (as per current practice in the fundraising industry); and the third accounts for individual-level scale effects.

In each scenario, we compare the expected donation amount of 1,000 donors in each of two hypothetical groups (level 1 and 2). These hypothetical donors are generated via the homogeneous parameters (e.g., group dummies) and draws from the joint multivariate normal density obtained from the field experiment (Table 6). One challenge in conducting a simulation study is that we cannot "observe" internal reference points and lag-amounts for hypothetical donors. We thus use the group-wise mean of all observed amounts in the real data as a proxy for the group-wise

internal reference points for hypothetical donors. As a proxy for the last donation amount (the lag amount in each period), we use the group-wise mean of all observed amounts for Easter, in June, and at Christmas. Then, for each period, the proportion of hypothetical donors giving in each period was set to reflect the response rate in the real data.

[**Table 9 about here**]

Table 9 shows the effect of an individually tailored appeals scale compared to a group-wise appeals scale. Improvements (from the group-wise to the individual-level appeals scale) are larger in the level 2 group (than in level 1), because their baseline donation amounts are higher. Likewise, Easter entails the largest increase due to its much higher response rate. The full-year expected donations, for an average level 1 and level 2 donor combined, are 519.8FF, 646.8FF, and 660.7FF, for the "no appeal", "group appeal", and "individual appeal" cases, respectively. The latter two both show marked improvement over "no appeal", 24.4% and 27.1%, respectively, offering a "scale-free" (if we may use that term) assessment of the power of making a donation request. By contrast, making an *individually-tailored* request offers a more modest improvement of 2.2% over a group-wise request. While this may seem comparatively small, it is certainly not so for charities. Moreover, charities are *already* making common group-wise requests (albeit, probably not close to optimal ones), so the question is how to leverage the individual donation histories they already have to boost total yield more. Here, we have done so by simultaneously modeling the effects of such a request on both yield rate and donation amount, as well as the intercorrelation of their unobserved influences (errors), and donor-level response heterogeneity.

Our simulation study should be taken as an illustrative exercise awaiting more detailed field data from a properly orthogonalized, custom design (although one wonders which charity will risk potential losses from such an experiment). Most notably, the nature of our data made it impossible to optimize an entire *scale*. However, the results clearly suggest both the importance of making an appropriate request, and of incorporating heterogeneity. We believe the platform developed here would allow for optimal "laddering," which is the dominant practice in real fundraising, for individual donors

9   Conclusion

Charities have long relied on appeals scales as cornerstones of their donation requests, though with little theory or measurement to guide them. In that light, the model developed here offers a heterogeneous, joint account of donation incidence and amount, addressing potential selectivity in modeling these separately. Importantly, different specifications for internal and external reference point theories can be assessed via model comparison.

Results suggest that variation across donors in scale attraction effects and seasonal donation patterns can be substantial. Such a finding depends critically on the availability of donation *histories*, explaining its absence from prior studies. A moderate, significantly negative, correlation between donation incidence and amount indicates the selectivity-based pitfalls of separately modeling incidence and amount. In terms of internal and external referents, we found that the mean of the previous donation amounts (internal referents) and including all points in an appeals scale (external referents) offered the best fit with our data. Importantly, the model allowed for a preliminary exploration of *laddering*, the common practice of asking for successively greater donation amounts.

Our study has some notable limitations, stemming mainly from the data. The charity that designed and carried out the study was not interested in optimal experimental practice, but in gaining some degree of insight within the context of a live donation drive. So, there is a potential issue with appeals scale amounts roughly tracked prior donation level in each segment. Second, because of the lack of group- and time-wise variation in appeals scale, the weight of each scale point - let alone any potential *interactions* among them - simply cannot be estimated; nor could the optimal *number* of points be ascertained. Third, because the appeals scales in the experiment contained only 'reasonable' amounts, effects of extreme scale points, such as ignoring them or even of alienating donors, could not be measured. Despite these data limitations, the model showed clear and strong evidence for scale attraction, in both upward and downward directions, and that the degree of attraction varied greatly across donors.

Some of the data limitations suggest clear directions for future experimental and field research. First and foremost would be some scheme for orthogonalizing appeals scale amounts across vari-

ous donor groups, and even the number of points on the scale, so that a truly "optimal" scale for each donor can be devised. Key to doing this is accounting for interactions among scale points, which is necessary to avoid "bunching up" in optimization. Future research might also identifying subtleties of weighting: do some consumers ignore endpoints, while others anchor on them? Experiments could similarly include extreme scale points, to see whether they are ignored entirely, lead respondents not to donate at all, or something more subtle. Any such data could be analyzed through variants of the basic framework employed here, and would help validate cross-study norms about scale point attraction effects, as well as fashion individually-tailored, multi-period laddering plans.

References

Amemiya, T. 1985. *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Brooks, S.P. and A. Gelman. 1998. "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics* 7 (4):434–455.

Celeux, G., F. Forbes, C. Robert, and M. Titterington. 2006. "Deviance Information Criteria for Missing Data Models." *Bayesian Analysis* 1 (4):651–674.

Chib, S. and E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *American Statistician* 49 (4):327–335.

Cialdini, R.B. and N.J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55:591–621.

De Bruyn, A. and S. Prokopec. 2011. "Maximizing Donations with Individually Tailored Appeals Scales: Bridging the Remaining Gaps between Theory and Practice." *working paper* .

Desmet, P. and F.M. Feinberg. 2003. "Ask and ye shall receive: The effect of the appeals scale on consumers' donation behavior." *Journal of Economic Psychology* 24 (3):349–376.

Donkers, B., R. Paap, J.J. Jonker, and P.H. Franses. 2006. "Deriving Target Selection Rules from Endogeneously Selected Samples." *Journal of Applied Econometrics* 21 (5):549–562.

Doob, A.N. and D.S. McLaughlin. 1989. "Ask and You Shall be Given: Request Size and Donations to a Good Cause." *Journal of Applied Social Psychology* 19 (12):1049–1056.

Edwards, Y.D. and G.M. Allenby. 2003. "Multivariate Analysis of Multiple Response Data." *Journal of Marketing Research* 40 (3):321–334.

Fisher, R.J. and D. Ackerman. 1998. "The Effects of Recognition and Group Need on Volunteerism: A Social Norm Perspective." *Journal of Consumer Research* 25 (3):262–275.

Fraser, C., R.E. Hite, and P.L. Sauer. 1988. "Increasing Contributions in Solicitation Campaigns: The Use of Large and Small Anchor Points." *Journal of Consumer Research* 15 (3):284–287.

Geman, S. and D. Geman. 1984. "Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (6):721–741.

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

Kalyanaram, G. and J.D.C. Little. 1994. "An Empirical Analysis of Latitude of Price Acceptance in Consumer Package Goods." *Journal of Consumer Research* 21 (3):408–418.

Kalyanaram, G. and R.S. Winer. 1995. "Empirical Generalization from Reference Price Research." *Marketing Science* 14 (3):161–169.

Krishnamurthi, L., T. Mazumdar, and S.P. Raj. 1992. "Asymmetric Response to Price in Consumer Brand Choice and Purchase Quantity Decisions." *Journal of Consumer Research* 19 (3):387–400.

Lattin, J.M. and R.E. Bucklin. 1989. "Reference Effects of Price and Promotion on Brand Choice Behavior." *Journal of Marketing Research* 26 (3):299–310.

Mayhew, G.E. and R.S. Winer. 1992. "An Empirical Analysis of Internal and External Reference Prices Using Scanner Data." *Journal of Consumer Research* 19 (1):62–70.

Mazumdar, T. and P. Papatla. 2000. "An Investigation of Reference Price Segments." *Journal of Marketing Research* 37 (2):246–258.

Rajendran, K.N. and G.J. Tellis. 1994. "Contextual and Temporal Components of Reference Price." *Journal of Marketing* 58 (1):22–34.

Reingen, P.H. 1982. "Test of a List Procedure for Inducing Compliance with a Request to Donate Money." *Journal of Applied Psychology* 67 (1):110–118.

Schibrowsky, J.A. and J.W. Peltier. 1995. "Decision Frames and Direct Marketing Offers: A Field Study in a Fundraising Context." *Journal of Direct Marketing* 9 (1):8–16.

Schwarz, N. 1999. "Self-reports: How the Questions Shape the Answers." *American Psychologist* 54 (2):93–105.

Schwarz, N., H. Bless, G. Bohner, U. Harlacher, and M. Kellenbenz. 1991. "Response Scales as Frames of Reference: The Impact of Frequency Range on Diagnostics Judgements." *Applied Cognitive Psychology* 5 (1):37–49.

Shang, J. and R. Croson. 2009. "A Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods." *The Economic Journal* 119 (540):1422–1439.

Shang, J., A. Reed, and R. Croson. 2008. "Identity Congruence Effects on Donations." *Journal of Marketing Research* 45 (3):351–361.

Sherif, M., D. Taub, and C.I. Hovland. 1958. "Assimilation and Contrast Effects of Anchoring Stimuli on Judgments." *Journal of Experimental Psychology* 55 (2):150–156.

Tanner, M.A. and W.H. Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82 (398):528–540.

The Center on Philanthropy at Indiana University. 2010. "Giving USA 2010." *Giving USA Foundation/Giving Institute* .

Van Diepen, M., B. Donkers, and P.H. Franses. 2009. "Dynamic and Competitive Effects of Direct Mailings: A Charitable Giving Application." *Journal of Marketing Research* 46 (1):120–133.

Wachtel, S. and T. Otter. 2011. "Successive Sample Selection and Its Relevance for Management Decisions." *working paper* .

Weyant, J.M. and S.L. Smith. 1987. "Getting More by Asking for Less: The Effect of Request Size on Donations of Charity." *Journal of Applied Social Psychology* 17 (4):392–400.

Ying, Y., F. Feinberg, and M. Wedel. 2006. "Leveraging Missing Ratings to Improve Online Recommendation Systems." *Journal of Marketing Research* 43 (3):355–365.

Zhao, Y., Y. Zhao, and I. Song. 2009. "Predicting New Customers' Risk Type in the Credit Card Market." *Journal of Marketing Research* 46 (4):506–517.

Table 1: Appeals Scales used in the Field Experiment

	Standard Scale					
	100 FF	150 FF	250 FF	500 FF	1000 FF	Other
Prior Donation level	*Test Scales*					
1	120 FF	180 FF	250 FF	350 FF	500 FF	Other
2	120 FF	200 FF	350 FF	500 FF	750 FF	Other

Table 2: Average Donation Amounts and Frequencies

Prior Donation Level	*Scales*	*Average Donation Amount*		*Yield Rate*
		per Household	*per Occasion*	
1	*Standard*	430.2	136.5	31.73%
	Test	434.3	137.3	31.61%
2	*Standard*	844.7	286.2	33.88%
	Test	839.5	283.8	33.81%

Table 3: Yield Rate and Average Amount of Observed Donations across Seasons

	Level 1			Level 2		
	Easter	June	Christmas	Easter	June	Christmas
Yield Rate	72.8%	18.6%	19.6%	75.3%	17.6%	20.9%
Average Donation	140.1	126.7	129.7	265.0	221.7	215.3

Table 4: Examples of Donation Histories for Several Randomly Selected Households

id #	2000			2001			2002			2003
	Easter	June	Xmas	Easter	June	Xmas	Easter	June	Xmas	Easter
3	100	0	0	100	0	0	100	0	0	100
20	0	150	0	0	150	0	150	0	0	250
66	200	0	150	200	0	150	150	0	150	250
118	100	100	0	100	100	0	100	0	0	150
148	0	90	0	100	0	100	150	150	100	150

Table 5: Mean and SE of observed amounts (FF) for each donor

	Observed Donation Amounts	
Prior Donation level	Mean	SE
1	137.71	22.60
2	259.05	47.43

Table 6: Parameter Estimates for Full Model

	Coefficient	mean	SE	95% HDR
Homogeneous	correlation (ρ)	-0.387	0.049	(-0.479, -0.288)
	sd of log amount (σ)	0.296	0.005	(0.286, 0.307)
	Easter dummy (β^E)	0.782	0.035	(0.714, 0.852)
	level dummy in selection ($\beta^{level,s}$)	0.073	0.039	(0.002, 0.156)
	level dummy in amount ($\beta^{level,a}$)	0.273	0.019	(0.237, 0.311)
	log amount lag in selection (β^{lag})	-0.131	0.009	(-0.149, -0.113)
Heterogeneous	June dummy (β_i^J)	-0.554	0.058	(-0.668, -0.441)
	Christmas dummy (β_i^C)	-1.019	0.070	(-1.163, -0.889)
	"gamma up" (β_i^U)	-0.418	0.070	(-0.563, -0.297)
	"gamma down" (β_i^D)	1.278	0.224	(0.858, 1.731)
	sd(June)	0.406	0.071	(0.277, 0.554)
	sd(Christmas)	0.831	0.088	(0.667, 1.010)
	sd(gamma up)	0.478	0.040	(0.404, 0.565)
	sd(gamma down)	0.932	0.154	(0.672, 1.260)
	corr(June, Christmas)	0.742	0.108	(0.506, 0.906)
	corr(June, gamma up)	-0.004	0.050	(-0.101, 0.094)
	corr(June, gamma down)	0.003	0.050	(-0.096, 0.103)
	corr(Christmas, gamma up)	0.002	0.049	(-0.097, 0.096)
	corr(Christmas, gamma down)	0.006	0.050	(-0.093, 0.102)
	corr(gamma up, gamma down)	0.380	0.145	(0.071, 0.626)

Table 7: In-sample fit of observed donation amounts

1. Heterogeneous seasonality and scale effects

	Full model				No correlation				No scale effect				Simple regression			
	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4
RMSE	**0.264**	0.287	0.294	0.290	0.276	0.296	0.298	0.301	0.334	0.351	0.320	0.344	0.319	0.331	0.335	0.320
MAD	**0.194**	0.207	0.215	0.213	0.204	0.215	0.218	0.223	0.229	0.227	0.233	0.228	0.233	0.240	0.238	0.232
LL	**-2745**	-2950	-2910	-2915	-2762	-2923	-2917	-2935	-3201	-3345	-3093	-3270				
AIC	**5532**	5943	5862	5873	5563	5885	5874	5910	6427	6715	6210	6563				
BIC	**5671**	6082	6001	6012	5696	6018	6006	6043	6506	6794	6290	6643				

2. Heterogeneous seasonality and Homogeneous scale effects

	Full model				No correlation				No scale effect				Simple regression			
	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4
RMSE	0.302	0.312	0.320	0.326	0.308	0.320	0.320	0.330	0.334	0.351	0.320	0.344	0.319	0.331	0.335	0.320
MAD	0.224	0.227	0.233	0.240	0.231	0.235	0.233	0.242	0.229	0.227	0.233	0.228	0.233	0.240	0.238	0.232
LL	-3067	-3174	-3102	-3200	-3002	-3090	-3093	-3163	-3201	-3345	-3093	-3270				
AIC	6163	6377	6232	6427	6031	6205	6212	6351	6427	6715	6210	6563				
BIC	6255	6470	6325	6520	6117	6291	6298	6438	6506	6794	6290	6643				

3. Homogeneous seasonality and Heterogeneous scale effects

	Full model				No correlation				No scale effect				Simple regression			
	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4
RMSE	0.266	0.286	0.296	0.291	0.276	0.296	0.298	0.301	0.334	0.351	0.320	0.345	0.319	0.331	0.335	0.320
MAD	0.196	0.208	0.217	0.214	0.204	0.215	0.218	0.223	0.229	0.228	0.233	0.228	0.233	0.240	0.238	0.232
LL	-3062	-3242	-3314	-3270	-3169	-3330	-3348	-3366	-3608	-3718	-3508	-3690				
AIC	6152	6512	6656	6568	6364	6686	6721	6758	7234	7454	7035	7399				
BIC	6245	6605	6749	6661	6450	6773	6807	6844	7294	7514	7094	7459				

4. Homogeneous seasonality and scale effects

	Full model				No correlation				No scale effect				Simple regression			
	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4	IR-1	IR-2	IR-3	IR-4
RMSE	0.304	0.313	0.320	0.328	0.308	0.320	0.320	0.330	0.334	0.351	0.320	0.345	0.319	0.331	0.335	0.320
MAD	0.226	0.228	0.233	0.241	0.231	0.235	0.233	0.241	0.229	0.228	0.233	0.228	0.233	0.240	0.238	0.232
LL	-3380	-3462	-3508	-3551	-3428	-3515	-3518	-3589	-3608	-3718	-3508	-3690				
AIC	6781	6945	7037	7125	6875	7049	7056	7197	7234	7454	7035	7399				
BIC	6854	7018	7110	7198	6941	7116	7122	7263	7294	7514	7094	7459				

Table 8: In-sample fit of observed donation amounts (Full model with IR-1)

	RMSE	MAD	LL	AIC	BIC
ER-1: All five scale points	**0.264**	**0.194**	**-2745**	**5532**	**5671**
ER-2: Two closest scale points from the internal referent	0.280	0.195	-2871	5784	5923
ER-3: Largest and smallest scale points	0.293	0.208	-2899	5840	5980
ER-4: Middle scale point	0.285	0.209	-2932	5906	6045
ER-5: Mean of all five scale points	0.312	0.217	-3078	6198	6337

Table 9: Expected value of average donation amount in simulation

	Level 1				*Level 2*			
	Easter	*June*	*Xmas*	***Level 1 Total***	*Easter*	*June*	*Xmas*	***Level 2 Total***
Individually Tailored Request	127.6	28.6	32.2	**188.4**	321.3	70.0	81.1	**472.4**
Group-wise Request	124.8	28.0	31.4	**184.2**	314.7	68.5	79.4	**462.6**
No suggested amount	100.4	22.5	25.3	**148.3**	252.8	55.0	63.8	**371.5**

A. Wikipedia: wikipedia.org

B. United Way: liveunited.org

C. United Nations Foundation: unfoundation.org

Figure 1: Examples of Appeals Scales used by Charities
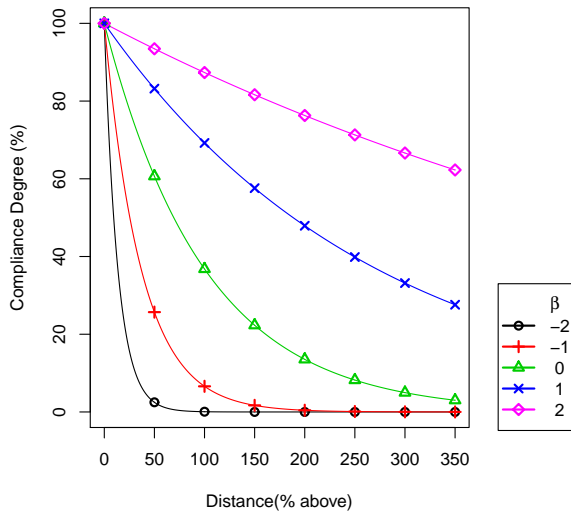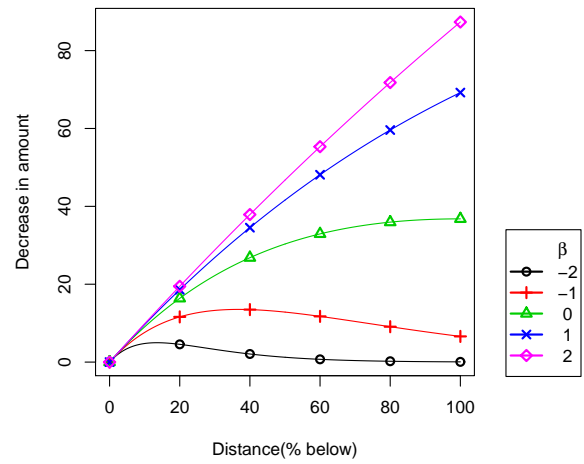
A. Upward Compliance Degree
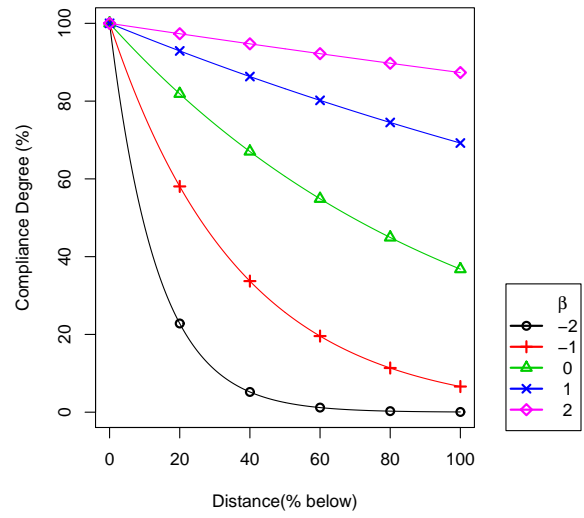
B. Downward Compliance Degree

Figure 2: Compliance Degree Curves
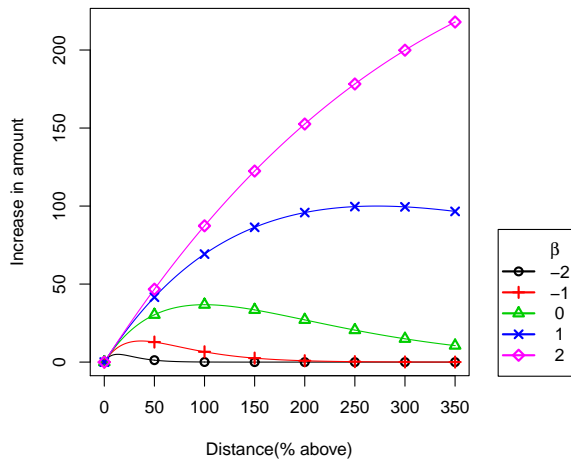
A. Upward Pulling Amount
(Internal reference point=100)

B. Downward Pulling Amount
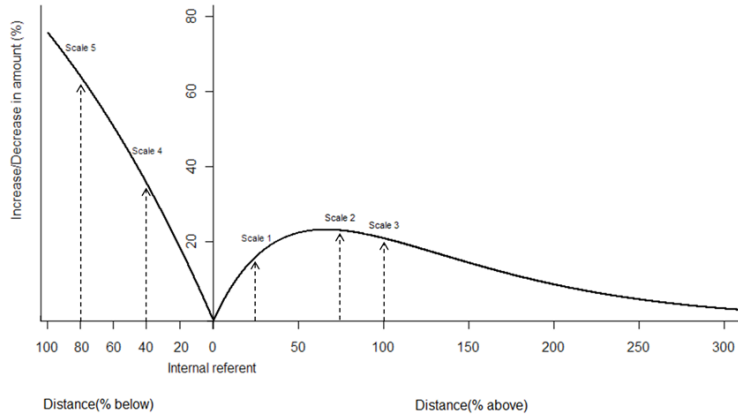(Internal reference point=100)

Figure 3: Pulling Amount Curves
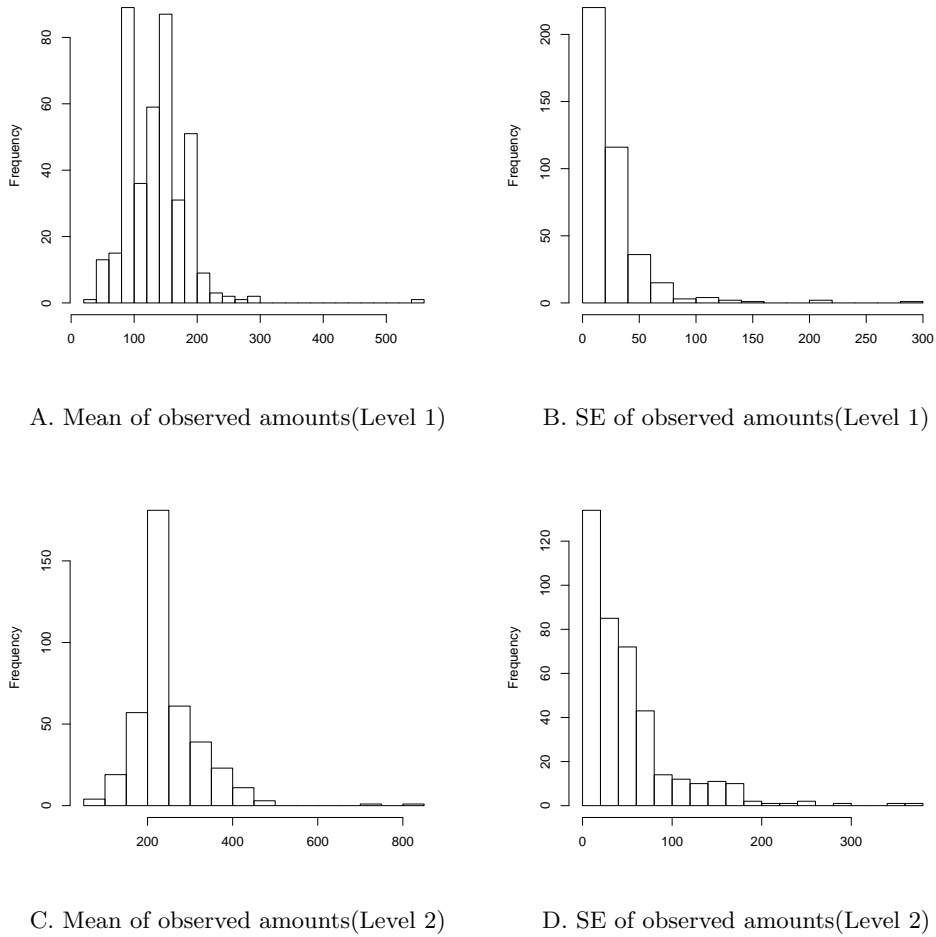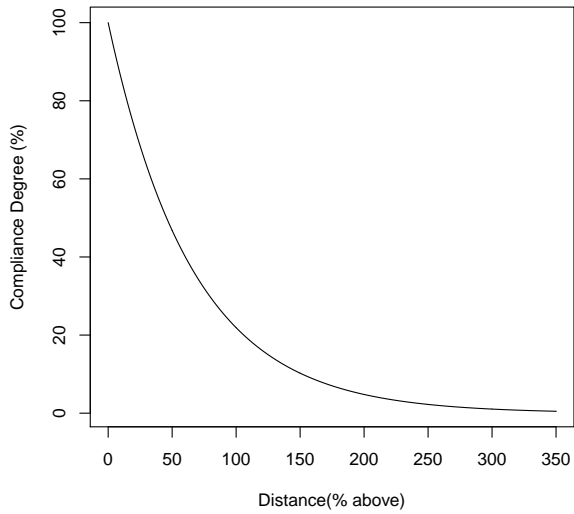
Figure 4: Pulling amounts owing to multiple scale (external reference) points

A. Mean of observed amounts(Level 1)

B. SE of observed amounts(Level 1)

C. Mean of observed amounts(Level 2)

D. SE of observed amounts(Level 2)

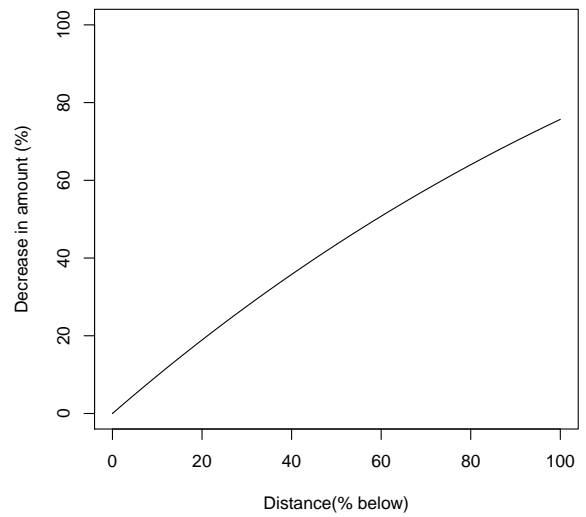Figure 5: Distribution of Mean and SE of observed amounts for each donor

A. Upward Compliance Degree
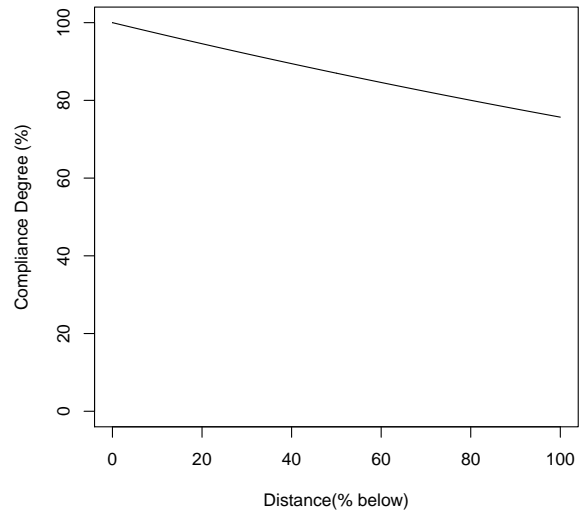
B. Downward Compliance Degree

Figure 6: Upward and downward compliance curves at gamma parameter posterior mean
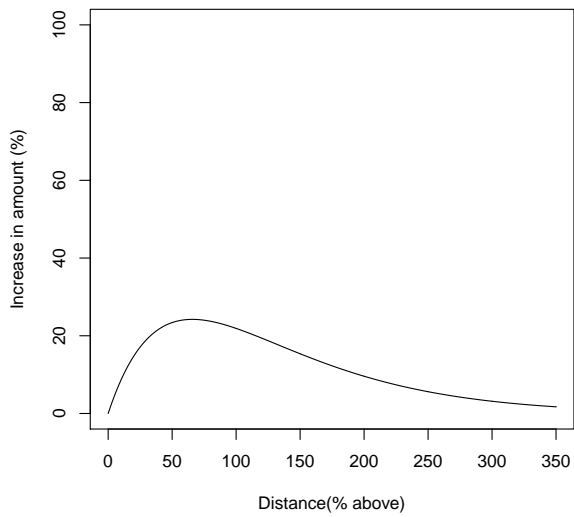
A. Upward Pulling Amount

B. Downward Pulling Amount

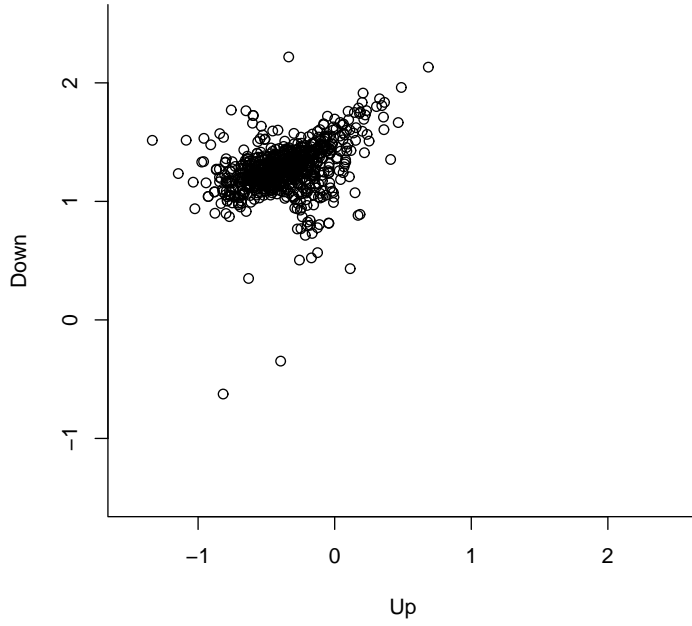Figure 7: Upward and downward pulling amount curves at gamma parameter posterior mean

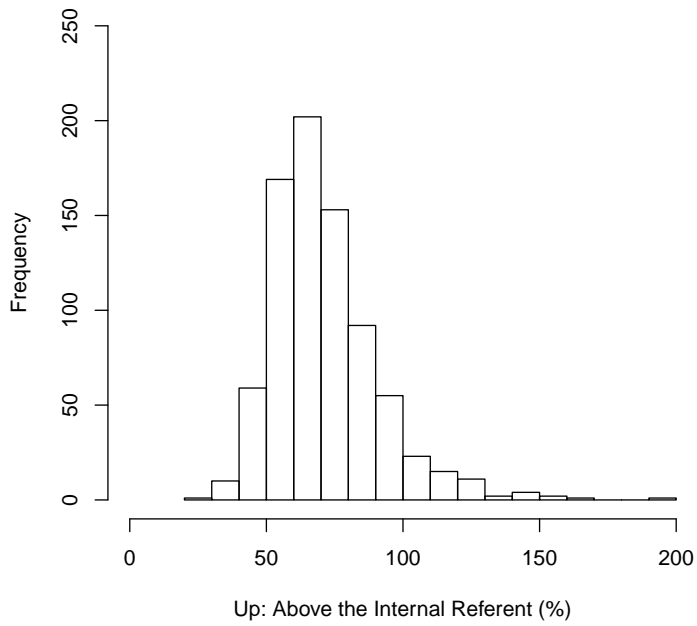Figure 8: Gamma parameters (up and down) for each donor

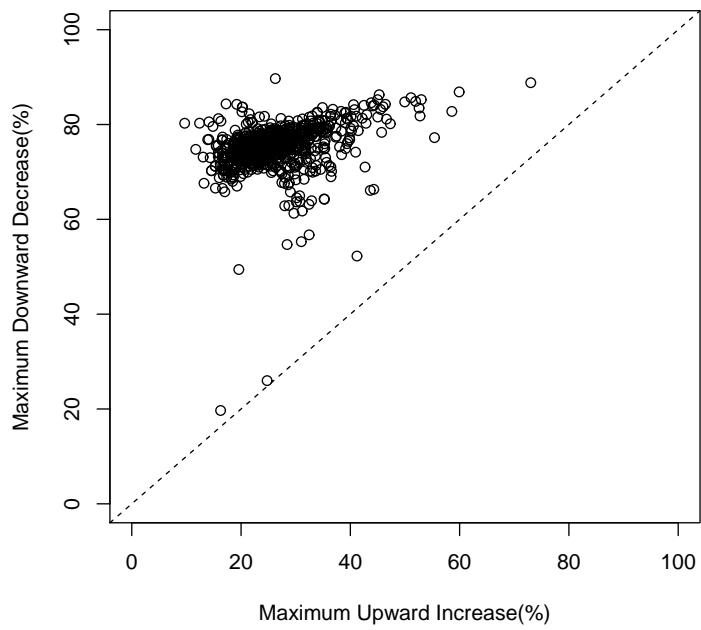Figure 9: Scale point with maximum upward pull

Figure 10: Maximum pulling up and down amounts for each donor

43

Appendix A. Full Model Specification

As discussed above, we can write the entire model as follows ($i =$ donor; $t =$ time):

$$y_{it}^{s*} = \beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level} + \epsilon_{it}^s$$

$$y_{it}^{a*} = \ln(r_{it}^I + APA_{it}) + \beta^{level,a} X_i^{level} + \epsilon_{it}^a, \ \ where:$$

$$y_{it}^s = 1, \ \ if \ y_{it}^{s*} \geq 0; \ \ 0 \ otherwise$$

$$y_{it}^a = y^{a*}, \ \ if \ y_{it}^s = 1; \ \ unobserved \ otherwise$$

$$APA_{it} = \sum_{k=1}^K w_{it}^k \times I_{it}^k \times PA_{it}^k; \quad w_{it}^k = \frac{PA_{it}^k}{\sum_{k=1}^K PA_{it}^k}, \quad I_{it}^k = \begin{cases} 1, & r_i^{E,k} \geq r_{it}^I \\ -1, & r_i^{E,k} < r_{it}^I \end{cases}$$

$$PA_{it}^k = CD_{it}^k \times \left\| r_i^{E,k} - r_{it}^I \right\|$$

$$CD_{it}^k = \exp\left(-\frac{d_{it}^k}{\theta_i}\right); \quad \theta_i = \begin{cases} \exp(\beta_i^U), & r_i^{E,k} \geq r_{it}^I \\ \exp(\beta_i^D), & r_i^{E,k} < r_{it}^I \end{cases}, \quad d_{it}^k = \frac{\left\| r_i^{E,k} - r_{it}^I \right\|}{r_{it}^I}$$

$$(\epsilon_{it}^s, \epsilon_{it}^a) \sim BVN(0, \Sigma_\epsilon); \quad \Sigma_\epsilon = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$$

$$\beta_i \sim MVN(\Delta, \Sigma_\beta), \ \ where \ \beta_i = (\beta_i^J, \beta_i^C, \beta_i^U, \beta_i^D)$$

Note that the internal reference point, r_{it}^I, for donor i can change over the course of the experiment, and is subscripted accordingly. However, since the appeals scale for donor i does *not* change over time, the kth external reference point for a donor i, $r_i^{E,k}$, lacks a t subscript. Again, the variance of ϵ^s is fixed to 1 for identification. Finally, the vector of heterogeneous parameters (β_i) follows a multivariate normal distribution with mean μ_β and full-rank covariance matrix Σ_β.

Appendix B. MCMC Algorithm and Priors

Here we present the prior distributions and sampling algorithm used in estimation. Because the requirement that setting error variance of the binary probit model (for donation incidence) be set to one ruins useful conjugacy properties, we instead make random draws from the unidentified space, as suggested by Edwards and Allenby (2003), and report post-processed estimates. Below, we specify Σ_ϵ in the unidentified space as $\Sigma_\epsilon = \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_a \\ \rho\sigma_s\sigma_a & \sigma_a^2 \end{bmatrix}$.

1. Data Augmented Likelihood

$$\prod_{i=1}^n \prod_{t=1}^T [(y_{it}^{s*}, y_{it}^{a*})|\beta_h, \beta_i, \Sigma_\epsilon] \times \prod_{i=1}^n [\beta_i|\mu_\beta, \Sigma_\beta]$$

where $\beta_h = (\beta^E, \beta^{lag}, \beta^{level,s}, \beta^{level,a})$ is a vector of homogeneous parameters and $\beta_i = (\beta_i^J, \beta_i^C, \beta_i^U, \beta_i^D)$ is a vector of heterogeneous parameters.

2. Prior Distribution

We use proper but diffuse priors.

(1) $\beta_h \sim MVN(M, V)$, where $M = \mathbf{0}$, $V = 10^4 I$

(2) $\Sigma_\epsilon \sim IW(\nu_{\Sigma_\epsilon}, V_{\Sigma_\epsilon})$, where $\nu_{\Sigma_\epsilon} = 5$, $V_{\Sigma_\epsilon} = 5I$

(3) $\Delta \sim MVN(\bar{\Delta}, A)$, where $\bar{\Delta} = \mathbf{0}$, $A = 10^4 I$

(4) $\Sigma_\beta \sim IW(\nu_{\Sigma_\beta}, V_{\Sigma_\beta})$, where $\nu_{\Sigma_\beta} = 7$, $V_{\Sigma_\beta} = 7I$

3. Posterior Distribution

$$\prod_{i=1}^{n}\prod_{t=1}^{T}[(y_{it}^{s*}, y_{it}^{a*})|\beta_h, \beta_i, \Sigma_\epsilon] \times \prod_{i=1}^{n}[\beta_i|\mu_\beta, \Sigma_\beta] \times [\beta_h|M, V] \times [\Sigma_\epsilon|\nu_{\Sigma_\epsilon}, V_{\Sigma_\epsilon}] \times [\Delta|\bar{\Delta}, A] \times [\Sigma_\beta|\nu_{\Sigma_\beta}, V_{\Sigma_\beta}]$$

4. Sampling Algorithm

Step 1. Draw y_{it}^{s*} and y_{it}^{a*} (Data augmentation step)

$$[(y_{it}^{s*}, y_{it}^{a*})|y_{it}^{s}, y_{it}^{a}, \beta_h, \beta_i, \Sigma_\epsilon]$$

1. If $y_{it}^{s} = 1$ then y_{it}^{a} is observed. We set $y_{it}^{a*} = y_{it}^{a}$ and draw y_{it}^{s*} from the truncated normal distribution below:

$$TN(\beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level} +$$
$$\frac{\rho\sigma_s}{\sigma_a}[y_{it}^a - (\ln(r_{it}^I + APA_{it}) + \beta^{level,a} X^{level})], (1-\rho^2)\sigma_s^2), \ y_{it}^{s*} \geq 0$$

2. If $y_{it}^{s} = 0$ then y_{it}^{a} is not observed. We draw $(y_{it}^{s*}, y_{it}^{a*})$ by following steps

a. Draw y_{it}^{s*} from $TN(\beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level}, \sigma_s^2), \ y_{it}^{s*} < 0$

b. Draw y_{it}^{a*} conditional on y_{it}^{s*} from normal distribution below:

$$N(\ln(r_{it}^I + APA_{it}) + \beta^{level,a} X_i^{level} +$$
$$\frac{\rho\sigma_a}{\sigma_s}[y_{it}^{s*} - (\beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level})], (1-\rho^2)\sigma_a^2)$$

Step 2. Draw β_i

$$[\beta_i|\beta_h, \Sigma_\epsilon, \Delta, \Sigma_\beta] \propto \prod_{t=1}^{T} [(y_{it}^{s*}, y_{it}^{a*})|\beta_h, \beta_i, \Sigma_\epsilon] \times [\beta_i|\Delta, \Sigma_\beta]$$

The full conditional distribution is also of unknown form. Therefore, we use a Metropolis-Hastings algorithm with a normal random walk proposal to make draws.

Step 3. Draw β_h

$$[\beta_h|\{\beta_i\}, \Sigma_\epsilon] \propto \prod_{i=1}^{n} \prod_{t=1}^{T} [(y_{it}^{s*}, y_{it}^{a*})|\beta_h, \beta_i, \Sigma_\epsilon] \times [\beta_h|M, V]$$

Again, we use a Metropolis-Hastings algorithm with a normal random walk proposal to make draws.

Step 4. Draw Σ_ϵ

$$[\Sigma_\epsilon|\beta_h, \{\beta_i\}] \propto \prod_{i=1}^{n} \prod_{t=1}^{T} [(y_{it}^{s*}, y_{it}^{a*})|\beta_h, \beta_i, \Sigma_\epsilon] \times [\Sigma_\epsilon|\nu_{\Sigma_\epsilon}, V_{\Sigma_\epsilon}]$$

$$\propto \prod_{i=1}^{n} \prod_{t=1}^{T} BVN\left(\begin{pmatrix} y_{it}^{s*} \\ y_{it}^{a*} \end{pmatrix} \Bigg| \begin{pmatrix} \beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level}) \\ \ln(r_{it}^I + APA_{it}) + \beta^{level,a} X_i^{level} \end{pmatrix}, \Sigma_\epsilon \right) \times$$

$$IW(\nu_{\Sigma_\epsilon}, V_{\Sigma_\epsilon})$$

$$[\Sigma_\epsilon|\beta_h, \{\beta_i\}] \sim IW(\tilde{\nu}_{\Sigma_\epsilon}, \tilde{V}_{\Sigma_\epsilon}), \;\; where$$

$$\tilde{\nu}_{\Sigma_\epsilon} = \nu_{\Sigma_\epsilon} + nT,$$

$$\tilde{V}_{\Sigma_\epsilon} = V_{\Sigma_\epsilon} + \sum_{i=1}^{n} \sum_{t=1}^{T} \begin{pmatrix} y_{it}^{s*} - (\beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level}) \\ y_{it}^{a*} - (\ln(r_{it}^I + APA_{it}) + \beta^{level,a} X_i^{level}) \end{pmatrix} \times$$

$$\begin{pmatrix} y_{it}^{s*} - (\beta^E X_{it}^E + \beta_i^J X_{it}^J + \beta_i^C X_{it}^C + \beta^{lag} X_{it}^{lag} + \beta^{level,s} X_i^{level}) \\ y_{it}^{a*} - (\ln(r_{it}^I + APA_{it}) + \beta^{level,a} X_i^{level}) \end{pmatrix}^{T}$$

Step 5. Draw Δ

$$[\Delta|\{\beta_i\}, \Sigma_\beta] \propto \prod_{i=1}^{n} [\beta_i|\Delta, \Sigma_\beta] \times [\Delta|\bar{\Delta}, A] \propto MVN_{nk}(B^*|[Z \otimes I_k]\Delta^*, I_n \otimes \Sigma_\beta) \times MVN_{nk}(\Delta^*|\bar{\Delta}, A)$$

where β_i is a vector of length k,

$$B = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_n^T \end{bmatrix}, \quad B^*_{nk \times 1} = vec(B^T), \quad Z = \begin{bmatrix} \iota_1^T \\ \iota_2^T \\ \vdots \\ \iota_n^T \end{bmatrix}, \quad \Delta^* = vec(\Delta^T)$$

$$[\Delta^* | \{\beta_i\}, \Sigma_\beta] \sim MVN_{nk}(\Delta^* | \tilde{\Delta}, \tilde{A})$$

where $\tilde{\Delta} = \tilde{A}([Z \otimes \Sigma_\beta^{-1}]B^*_{nk \times 1} + A^{-1}\bar{\Delta}), \quad \tilde{A} = [(Z^T Z) \otimes \Sigma_\beta^{-1} + A^{-1}]^{-1}$

Step 6. *Draw* Σ_β

$$[\Sigma_\beta | \{\beta_i\}, \Delta] \propto \prod_{i=1}^{n} [\beta_i | \Delta, \Sigma_\beta] \times [\Sigma_\beta | \nu_{\Sigma_\beta}, V_{\Sigma_\beta}] \propto MVN_{nk}(B | Z\Delta, I_n, \Sigma_\beta) \times IW(\nu_{\Sigma_\beta}, V_{\Sigma_\beta})$$

$$[\Sigma_\beta | \{\beta_i\}, \Delta] \sim IW(\tilde{\nu}_{\Sigma_\beta}, \tilde{V}_{\Sigma_\beta})$$

where $\tilde{\nu}_{\Sigma_\beta} = \nu_{\Sigma_\beta} + n, \quad \tilde{V}_{\Sigma_\beta} = V_{\Sigma_\beta} + (B - Z\Delta)^T(B - Z\Delta)$

Appendix C. Single Optimal Appeal

The single optimal appeal is that (scale) point (r^{E*}) that maximizes the log of the latent donation amount (y^{a*}). Clearly, r^E should be greater than r^I and maximizes the upward pulling amount (PA), so that r^{E*} solves a first order condition (PA with respect to r^E). Subscripts for donors and times (i, t) are suppressed for simplicity.

$$y^{a*} = \ln(r^I + PA) + X^a \beta^a + \epsilon^a$$

$$PA = \exp\left(-\frac{d}{\theta}\right)(r^E - r^I), \quad d = \frac{r^E}{r^I} - 1, \quad \theta = \exp(\beta_U)$$

$$FOC : \quad \frac{\partial PA}{\partial r^E} = 0$$

$$r^{E*} = (\theta + 1)r^I = (\exp(\beta_U) + 1)r^I$$

Appendix D. Expected Donation Amount

Following the Type 2 Tobit model specification, let y^{s*} and y^{a*} be the latent dependent variables in the selection and the amount equations, and y^s and y^a be their observed counterparts in (9). We set ν^s, ν^a to be deterministic parts of each equation and $\epsilon_s, \sigma\epsilon_a$ to be stochastic parts. Now we have the system of equations:

$$
\begin{aligned}
y^{s*} &= \nu^s + \epsilon_s \\
y^{a*} &= \nu^a + \sigma\epsilon_a
\end{aligned}
$$

Because the errors are correlated, we can re-write the system as (writing ϵ_s as ϵ):

$$
\begin{aligned}
y^{s*} &= \nu^s + \epsilon \\
y^{a*} &= \nu^a + \sigma(\rho\epsilon + \bar\rho z)
\end{aligned}
$$

where z is a standard normal draw uncorrelated with ϵ, and $\bar\rho = \sqrt{1-\rho^2}$. We want to calculate $\mathbb{E}[\exp(y^a)]$, the expected donation amount, and so merely integrate over the two *uncorrelated* errors, z and ϵ:

$$
\begin{aligned}
\mathbb{E}[\exp(y^a)] &= \int_{\epsilon=-\nu^s}^{\epsilon=\infty}\int_{z=-\infty}^{z=\infty} \exp[\nu^a + \sigma(\rho\epsilon + \bar\rho z)]\phi(z)\phi(\epsilon)dzd\epsilon \\
&= \exp(\nu^a) \times \int_{\epsilon=-\nu^s}^{\epsilon=\infty}\exp(\sigma\rho\epsilon)\phi(\epsilon)d\epsilon \times \int_{z=-\infty}^{z=\infty}\exp(\sigma\bar\rho z)\phi(z)dz
\end{aligned}
$$

The second term can be calculated by completing the square:

$$
\begin{aligned}
\int_{\epsilon=-\nu^s}^{\epsilon=\infty}\exp(\sigma\rho\epsilon)\phi(\epsilon)d\epsilon &= \int_{\epsilon=-\nu^s}^{\epsilon=\infty}\exp(\sigma\rho\epsilon)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\epsilon^2}{2}\right)d\epsilon \\
&= \int_{\epsilon=-\nu^s}^{\epsilon=\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\epsilon-\sigma\rho)^2}{2}\right)d\epsilon \times \exp\left(\frac{(\sigma\rho)^2}{2}\right) \qquad Let \ \ u = \epsilon - \sigma\rho \\
&= \int_{u=-\nu^s-\sigma\rho}^{u=\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du \times \exp\left(\frac{(\sigma\rho)^2}{2}\right) = \Phi(\nu^2+\sigma\rho) \times \exp\left(\frac{(\sigma\rho)^2}{2}\right)
\end{aligned}
$$

The third term can also be calculated by completing the square:

$$
\begin{aligned}
\int_{z=-\infty}^{z=\infty}\exp(\sigma\bar\rho z)\phi(z)dz &= \int_{z=-\infty}^{z=\infty}\exp(\sigma\bar\rho z)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz \\
&= \int_{z=-\infty}^{z=\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z-\sigma\bar\rho)^2}{2}\right)dz \times \exp\left(\frac{(\sigma\bar\rho)^2}{2}\right) = \exp\left(\frac{(\sigma\bar\rho)^2}{2}\right)
\end{aligned}
$$

Putting this all together, the expected value of donation amount is:

$$
\begin{aligned}
\mathbb{E}[\exp(y^a)] &= \exp(\nu^a) \times \Phi(\nu^2 + \sigma\rho) \times \exp\left(\frac{(\sigma\rho)^2}{2}\right) \times \exp\left(\frac{(\sigma\bar{\rho})^2}{2}\right) \\
&= \exp(\nu^a) \times \Phi(\nu^2 + \sigma\rho) \times \exp\left(\frac{\sigma^2\rho^2}{2}\right) \times \exp\left(\frac{\sigma^2(1-\rho)^2}{2}\right) \qquad \because \bar{\rho} = \sqrt{1-\rho^2} \\
&= \exp\left(\nu^a + \frac{\sigma^2}{2}\right) \times \Phi(\nu^2 + \sigma\rho)
\end{aligned}
$$

Appendix E. Distance ratio and the incremental/decremental amount, calculated at the scale point with the maximum pulling amount

1.Upward pulling amount

When $r^E \geq r^I$, the upward pulling amount PA_U follows:

$$
PA_U = \exp\left(-\frac{d_U}{\theta}\right)(r^E - r^I), \quad d_U = \frac{r^E}{r^I} - 1, \quad \theta_U = \exp(\beta_U)
$$

The scale point with maximum upward pulling amount (r^{E*}) can be calculated by solving first order condition of PA_U with respect to r^E.

$$
r^{E*} = (\theta_U + 1)r^I = (\exp(\beta_U) + 1)r^I
$$

At the scale point of r^{E*}, the incremental ratio in distance (d_U^*) is determined to be $\exp(\beta_U)$ and the maximum incremental ratio in the amount to be as follows:

$$
\frac{\exp\left(-\frac{d_U^*}{\exp(\beta_U)}\right)(r^{E*} - r^I)}{r^I} = \exp(-1)\exp(\beta_U)
$$

2. Downward pulling amount

When $0 \leq r^E \leq r^I$, the downward pulling amount PA_D follows:

$$
PA_D = \exp\left(-\frac{d_D}{\theta}\right)(r^I - r^E), \quad d_D = 1 - \frac{r^E}{r^I}, \quad \theta_D = \exp(\beta_d)
$$

The scale point with maximum downward pulling amount (r^{E*}) can be calculated by solving first order condition of PA_D with respect to r^E. We should note that there is a corner solution if $\beta_D > 0$.

$$r^{E*} = \begin{cases} (1 - \theta_D)r^I = (1 - \exp(\beta_D))r^I, & \beta_D \leq 0 \\ 0, & \beta_D > 0 \end{cases}$$

At the scale point of r^{E*}, the decremental ratio in distance (d_D^*) is determined to be

$$\begin{cases} \exp(\beta_D), & \beta_D \leq 0 \\ 1, & \beta_D > 0 \end{cases}$$

And the maximum decremental ratio in the amount is determined to be as follows:

$$\begin{cases} \dfrac{\exp\left(-\frac{d_U^*}{\exp(\beta_U)}\right)(r^I - r^{E*})}{r^I} = \exp(-1) - \exp(\beta_U), & \beta_D \leq 0 \\ \dfrac{\exp\left(-\frac{d_U^*}{\exp(\beta_U)}\right)(r^I - r^{E*})}{r^I} = \exp\left(-\frac{1}{\exp(\beta_D)}\right), & \beta_D > 0 \end{cases}$$