Optimizing experiment design for estimating parametric models in economic experiments*

James R. Bland†

2025-09-27

Abstract: Estimating parametric models of behavior from economic experiments is becoming increasingly common. This paper discusses methods for designing an experiment so that parameters in these models are estimated as precisely as possible. Using an example of designing a battery of pairwise lottery choices, I demonstrate how experiments can be designed for various inferential goals.

1 Introduction

When the goal is to estimate a structural model of decision-making, how can we best design our experiments to achieve this? This question has of course been at the forefront of some experimenters' minds when designing these experiments. For example, in an investigation of decision-making under risk, Hey and Di Cagno (1990) narrate their experiment design as follows:

> The choice of questions was not easy ... We tried to get a mixture so that the slope of the line joining the pair of gambles varied considerably ... The idea behind this was that we would then be able to distinguish between very risk-averse people and not-very-risk averse people, ...

As Moffatt (2007) notes, early experimental economists were largely theorists, and so while early experiments were probably designed with a lot of understanding of the *economic* models being used, perhaps there was less understanding of the *econometric* models being used. In fact, being more focused on inference from observational data where there is much less control over the data-generating process, even a traditional modern treatment of econometrics may not have left the experimenter with good tools to tune their experiment to best answer their research question.

In the meantime, we have of course developed a better understanding of experiment design. This is especially the case for the design of binary choice experiments of decision-making

*This project benefitted greatly from comments by Glenn Harrison, Brian Monroe, and attendees of the 2023 North American ESA meetings.

†The University of Toledo. james.bland@utoledo.edu https://orcid.org/0000-0002-7117-9998

under risk where decisions can be represented in the Marschak-Machina (MM) triangle, which I shall use as an example in this paper. That is, when all lottery pairs in an experiment can be described as probability distributions over three prizes. An example of a MM triangle is shown in Figure 1. The MM triangle can represent all lotteries over three prizes. The probability of winning the largest prize is shown in the vertical coordinate, and the probability of winning the lowest prize is shown in the horizontal coordinate. A pair of lotteries is shown in this Figure as a line, with the endpoints denoting the lotteries themselves. In this example "pair 1", denoted by the solid line, contains one lottery that is dominated by the other (in the first-order sense). We can see this because one end of this line is above and to the left of the other. This means that compared to its counterpart, this lottery has more probability on the highest prize and less probability on the lowest prize. "pair 2" on the other hand, denoted by the dashed line, does not contain a dominated lottery: the lottery with the greater probability of winning the largest prize also has a greater probability of winning the smallest prize.

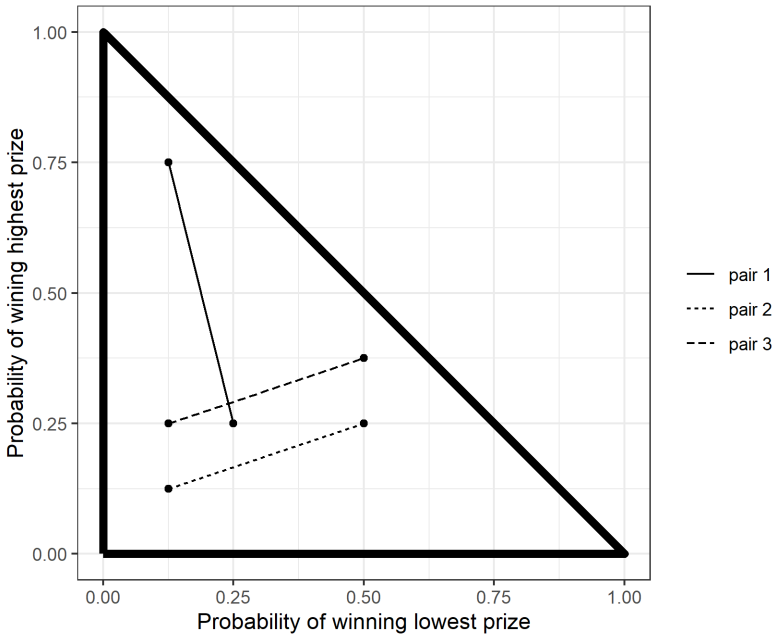

Figure 1: A Marschak-Machina (MM) triangle showing two lottery pairs. Each lottery is denoted by a point, and each pair of lotteries is connected by a line

As noted by Loomes and Sugden (1998), an important design feature to *test* expected utility theory is to have several lottery pairs with the same slope connecting the lotteries in the MM triangle. An example of this is shown in Figure 1, where pairs 2 and 3 are connected with parallel line segments. Since the indifference curves for expected utility theory are also parallel straight lines, it must be that both decisions are either at the left ends of the lines, or both decisions are at the right ends of the lines.[1] Therefore in their design Loomes and Sugden (1998) had many lottery pairs with the same slopes to *test* expected utility theory, and some variation in the slopes in order for their test to be appropriately powerful over a plausible range of risk preferences. This is also why Hey and Di Cagno (1990) correctly decided that

[1]This assumes that the decision-maker is not indifferent.

varying these slopes—not keeping them parallel—was an important design consideration for their experiment. By varying the slopes, we give expected utility participants an opportunity to reveal their risk preferences. That is, important features of an experiment designed for a *test* of a theory do not necessarily make for a good experiment design if we want to *estimate* the parameters in a model: by including two lottery pairs with parallel connecting line segments, we may be giving up some precision in our estimate of risk-aversion because we could have presented the participant with a less redundant choice.

Neither is it clear that an experiment designed for testing a theory in the sense of testing *implications* of a theory would be a good design for testing a theory in the sense of testing a *parameter restriction* in a parametric model. For example Loomes and Sugden (1998) tests an implication of expected utility theory compared to a generic alternative hypothesis, whereas Monroe (2020) focuses on testing whether a parameter restriction on a rank dependent utility model nesting expected utility theory holds. In the first case, a test of expected utility theory against "something else" would be more powerful if there were many pairs of parallel lines in the MM triangle. Here the participant would have many opportunities to violate the predictions of expected utility theory (in this case by making a "common ratio" pattern of choices). In the second case, the design goal would be to have as precise as possible estimates of the parameter(s) that nest expected utility theory. As noted by Monroe (2020), existing experimental designs (specifically Hey and Orme (1994) and Harrison and Ng (2016)) are inadequate by normal acceptable standards in distinguishing between parametric rank-dependent utility and expected utility models. Taken together, we have a good understanding of how to design an experiment if our goal is testing an implication of a theory. But it is not clear that the features of an experiment that make it a good test of a theory will also make it a good experiment for estimating the parameters in a model, and in some cases these design goals can provide conflicting recommendations.

The task of designing an experiment tuned to estimate parameters of a model of course becomes more complex as the model becomes more complex. Given that design considerations for testing versus estimating a model could diverge even for simpler models like expected utility theory, exploring the design space using computational methods becomes increasingly beneficial. For example, suppose that we have narrowed down the design considerations of our experiment to choosing a number of lottery pairs in the MM triangle. While the model may become more complicated, the design space for our experiment is exactly the same whether we want to estimate a simple 2-parameter expected utility model, or a generalization of this with more free parameters. Wilcox (2023), for example, uses simulation to explore the design space of an experiment with the goal of tuning it to non-parametrically estimate probability weighting functions in a rank-dependent utility model. In this case, the design space is relatively small,[2] but the number of parameters in the model is large because probability weighting functions are estimated non-parametrically.

I describe a framework for designing an economic experiment with the goal of estimating specific parameters, or transformations of parameters, in a model as precisely as possible. This method is based on $\mathcal{A}$-optimal experiment design (e.g. Nguyen and Miller (1992)),

[2]For experiment 2, Wilcox (2023) restricts the design space to lotteries that can be expressed in increments of 0.25.

which seeks to design experiments so that the trace of the Fisher information matrix is maximized. In doing so, the experiment design will maximize the expected precision of our estimator. However for many inferential goals, some parameters in a model are "nuisance parameters", which must be estimated alongside the parameters that answer our research question, but whose values we do not need to know to answer our research question. In these cases, designing an experiment to estimate *all* parameters as precisely possible may come at the expense of *specific* parameters' precision. This can be overcome by focusing on the section of the information matrix that corresponds just to the parameters that are directly relevant to our research question. Furthermore, it may be that the experimenter is not interested in the parameters themselves, but a *transformation* of them (e.g. a welfare measure or a prediction in another scenario). I extend $\mathcal{A}$-optimal experiment design in order to target precision of such transformations of parameters.

A substantial complication to this process is that in order to compute the information matrix in most nonlinear models, one needs to know the parameters of the model. As noted in Moffatt (2007), this poses a "chicken an egg" problem: how can one have estimates of the parameters of interest before running an experiment? I address this problem through using a prior over the model's parameters. Such a prior could be calibrated to data from existing similar economic experiments. Where such experiments are not informative enough, perhaps because similar enough experiments have not been run at the same location, this process is still applicable, but would rely more on the experimenter's expert knowledge of the environment. For example, Bland (2025a) discusses methods for calibrating priors to known properties of behavior in economic experiments.

I demonstrate this framework by designing two experiments of decision-making under risk, with the goal of estimating a parametric model of decision-making as precisely as possible. The first is a toy example to highlight the process in a simple one-parameter environment. The second, reflecting a popular format for such experiments, is for a battery of 80 pairwise lottery choices. Here I calibrate a prior for the model's parameters using existing experimental data (Harrison and Swarthout 2023), and then explore through simulation the sampling properties of the estimator when data are generated from experiments designed to achieve various inferential goals.

While this computer-aided design can be a useful tool in designing economic experiments, there is still much room for the expert knowledge of the experimenter to improve experiment designs in areas where the econometric model is silent, or makes unrealistic simplifying assumptions about behavior. Presenting a participant with a choice over the same lottery pair multiple times, for example, may not be as informative as our model that assumes decisions are independent would like to believe. Other design considerations in these experiments for example include "boundary effects", which coming from observations by Camerer (1989) and Camerer (1992), note that observed decisions are generally more consistent with expected utility theory when both lotteries are strictly in the interior of the MM triangle.[3] If these effects cannot be produced by the model we are estimating, then it is likely that a computer-aided design could be improved by incorporating this knowledge. As such, this paper makes

[3]Due to these effects Harrison and Ng (2016) choose to have half of their lottery pairs be in the interior of the MM triangle, and the other half include boundaries.

a call for more computer-*aided* design of experiments, but not experiments designed *only* by computers.

2 Experiment design as an optimization problem

Consider the task of designing an experiment as an optimization problem. As the experimenter, we seek to maximize our utility given design constraints by choosing experimental conditions $d \in \mathbb{D}$. Our utility function $V(d)$ measures our *ex-ante* utility from running our experiment with conditions d. The concept of maximizing a utility function subject to constraints should not be unfamiliar to experimental economists, however the process of *determining what this utility function should look like* may not be so obvious. Broadly speaking, the choice of $V(d)$ should be determined by the research question: larger values of $V(d)$ would then indicate *ex-ante* "better" answers to the research question. For example, if the goal of an experiment is to estimate a certainty equivalent, then a reasonable choice of $V(d)$ could be the expected mean squared error of the estimator of this certainty equivalent. On the other hand if the goal of the experiment is to distinguish between two competing models (say, expected utility theory and rank dependent utility), then $V(d)$ could be the expected power of the test between these two theories (holding test size constant).

While this utility-maximization problem is conceptually simple, there are several practical hurdles that one will need to overcome in order to solve it. Firstly, we need to formally specify the function $V(d)$ so that we can maximize it. While there are some out-of-the-box suggestions for this based on functions of the Fisher information matrix, an approach more tailored to the research question is clearly more desirable. Secondly, for non-linear models even these out-of-the-box suggestions for $V(d)$ are a function of our model's parameters θ, and so we need to formalize our beliefs about these parameters in order to meaningfully incorporate them into $V(d)$. Furthermore, these parameters are likely heterogeneous across participants, and so we need to incorporate participant heterogeneity into our beliefs. I address this by using a prior for the model's parameters θ. Finally, since we typically have many different choices when designing an experiment, maximizing $V(d)$ will be a high-dimensional problem. I address this using an exchange algorithm. I elaborate on these three issues in the remainder of this section.

2.1 Constructing the utility function $V(d)$

I consider a class of problems where we can write down our utility of using a point estimate $\tilde{\theta}$ when the true value of the parameter is θ. I will denote this utility function as $v(\tilde{\theta} \mid \theta)$. I will assume that $v(\tilde{\theta} \mid \theta)$ has the following properties:

1. Differentiability: $v(\tilde{\theta} \mid \theta)$ is twice continuously differentiable with respect to both of its arguments.
2. Optimization: $\arg\max_{\tilde{\theta}} v(\tilde{\theta} \mid \theta) = \theta$, and $\frac{\partial v(\tilde{\theta}|\theta)}{\partial \tilde{\theta}}\Big|_{\tilde{\theta}=\theta} = 0$

These assumptions are required to make an approximation for $V(d)$, which is discussed below. The second assumption states that we (the experimenters) are happiest when our point

estimate $\tilde{\theta}$ is equal to the true value θ.

Since $\tilde{\theta}$ is not known before our experiment, we integrate out its uncertainty with respect to the sampling distribution of our estimator conditional on θ, $p(\tilde{\theta} \mid d, \theta)$:

$$E[v(\tilde{\theta} \mid \theta) \mid d, \theta] = \int_{\Theta} v(\tilde{\theta} \mid \theta) p(\tilde{\theta} \mid d, \theta) \mathrm{d}\tilde{\theta} \tag{1}$$

Importantly, we can influence the sampling distribution of our estimator $p(\tilde{\theta} \mid d, \theta)$ through choosing our experiment design d.

Since the true value of the parameter θ is also also not known before the experiment, we integrate out our uncertainty with respect to our prior over θ, $p(\theta)$. This yields our expected utility of using design d:

$$\begin{aligned} V(d) &\doteq E\left[E\left(v(\tilde{\theta} \mid \theta) \mid d, \theta\right) \mid d\right] = \int_{\Theta} \int_{\Theta} v(\tilde{\theta} \mid \theta) p(\tilde{\theta} \mid \theta, d) \mathrm{d}\tilde{\theta} p(\theta) \mathrm{d}\theta \\ &= E\left[v(\tilde{\theta} \mid \theta) \mid d\right] \end{aligned} \tag{2}$$

where the second line follows from the law of iterated expectations. However the expectation in the first line is useful because it allows us to approximate $V(d)$ in two steps.

2.2 Choosing and scruitinizing an appropriate prior

Constructing $V(d)$ requires specifying a prior $p(\theta)$ over participant-specific parameters. For parametric models of behavior for example, θ could include elements describing risk-aversion, choice precision, probability weighting, time preferences, and other-regarding preferences, to name but a few. In situations where previous experiments can speak to typical ranges of parameter values, these can be used to calibrate $p(\theta)$ to match these ranges. For example Bland (2025a) calibrates a prior for a coefficient of relative risk-aversion to data from Holt and Laury (2002). Such back-of-the-envelope calculations can provide useful, fast information that will hopefully ensure that the researcher is designing an experiment for participants in the right ballpark.

For applications where there already exists data from similar experiments, these data can be further used to *estimate* an appropriate prior. Bayesian hierarchical models such as (for example) Gao, Harrison, and Tchernis (2023) and Gao, Harrison, and Tchernis (2020) provide us with estimates of the distribution of θ_i in their subject pools. If the subject pool and experiment are similar enough to those of the experiment we are designing, then these estimates can be used as a starting point for constructing $p(\theta)$. This is the technique I use in my second example below.

There are, of course, situations in which existing data may not be available or appropriate to use. For example this could be because our participants will be drawn from an unusual subject pool, or because our experiment will be investigating something substantially novel. In these situations we must rely more on our own expert judgement about appropriate values for θ_i, but we can also use our model to help us choose $p(\theta)$. That is, we can explore our model's

predictions given different values of θ_i and ask whether these predictions seem plausible or nonsensical. For example, Bland (2025a) shows that some realizations of the logit choice precision parameter λ above a particular value imply almost deterministic choices, and that realizations below another smaller value imply behavior not substantially different to uniform randomization. The "interesting" part of the distribution is between these values, and we should therefore calibrate our prior to place most of its mass in this region.

Whether our prior is calibrated through a judgement call, estimated, or otherwise, we can also use our prior to scrutinize the performance of our experiment once it is designed (and before it is run). If, for example, we are worried that the prior we chose to arrive at optimal design d was, say, too "informative", in that the prior variance for some or all of the parameters seemed too small, we can use our model to investigate our experiment's performance for other distributions. That is, if we can evaluate $V(d)$ for our calibrated prior $p(\theta)$, then we can also evaluate it for any other prior $p'(\theta)$. If our design d does not suffer too much under these alternative priors, then we can be confident that our experiment, while not explicitly designed with $p'(\theta)$ in mind, will not perform too badly if this was the actual distribution of θ_i.

2.3 Computational considerations

Optimizing $V(d)$ poses (at least) three computational problems: (i) we need to evaluate the inner expectation in (2), (ii) we need to evaluate the outer expectation in (2), and finally (iii) we need to optimize $V(d)$. There are many ways of approaching each of these problems, which may be better or worse for different applications. I present one approach to each of these problems in the sections below.

2.3.1 Approximating the inner expectation

The inner expectation in (2) is the expected utility of the experimenter, conditional on the design d and true parameters θ. I approximate this with a second-order Taylor series expansion and an assumption that the variance of th estimator dominates the bias of the estimator. Specifically, I approximate the innermost expectation using a second-order Taylor series expansion at $\tilde{\theta} = \theta$ as follows:

$$
\begin{aligned}
E\left[v(\tilde{\theta} \mid \theta) \mid d, \theta\right] &\approx E\left[v(\theta \mid \theta) + (\tilde{\theta} - \theta)^\top \frac{\partial v(\theta \mid \theta)}{\partial \tilde{\theta}} + \frac{1}{2}(\tilde{\theta} - \theta)^\top \frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \tilde{\theta}^\top}(\tilde{\theta} - \theta) \mid d, \theta\right] \\
&= v(\theta \mid \theta) + \frac{1}{2} E\left[(\tilde{\theta} - \theta)^\top \frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \tilde{\theta}^\top}(\tilde{\theta} - \theta) \mid d, \theta\right] \\
&= v(\theta \mid \theta) + \frac{1}{2} E(\tilde{\theta} - \theta \mid \theta, d)^\top \frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \tilde{\theta}^\top} E(\tilde{\theta} - \theta \mid \theta, d) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \tilde{\theta}^\top}\mathcal{V}(\theta, d)\right) \\
&\approx v(\theta \mid \theta) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \tilde{\theta}^\top}\mathcal{V}(\theta, d)\right)
\end{aligned}
\tag{3}
$$

where the second line holds because (by assumption) the first-order condition identifies the maximizer, the last line holds with equality in the limit as the number decisions approaches infinity for a consistent estimator, and $\mathcal{V}(\theta, d)$ is the variance-covariance matrix of our estimator conditional on θ and d. In the last line, we are assuming that the variance of our estimator will dominate its bias.

2.3.2 Approximating the outer expectation

To address the second problem, the inner expectation is approximated by (3), all of which can be obtained by analytically taking the derivatives of v and the log-likelihood function. The outer expectation can then be approximated using Monte Carlo integration as follows:

$$E\left[E\left[v(\tilde{\theta} \mid \theta) \mid d, \theta\right] d\right] \approx \frac{1}{S}\sum_{s=1}^{S}\left[v(\theta^s \mid \theta^s) + \frac{1}{2}\text{tr}\left(\frac{\partial^2 v(\theta^s \mid \theta^s)}{\partial\tilde{\theta}\partial\tilde{\theta}^\top}\mathcal{V}(\theta^s, d)\right)\right] \tag{4}$$

where $\{\theta^s\}_{s=1}^S$ are draws from the prior distribution $p(\theta)$.

2.3.3 Maximizing the (approximated) utility function

The third problem–maximizing $V(d)$–arises because experiment design problems are typically high dimensional. For example in Harrison and Ng (2016) and the second example in this paper, there are 80 lottery pairs over three prizes, which means that there are $2 \times (3-1) \times 80 = 320$ different choice variables for the probabilities in these lotteries alone. In order to manage this high-dimensional problem, I use an exchange algorithm (e.g. Nguyen and Miller (1992)). The algorithm simplifies the problem by partitioning the design space and then focusing on the elements of the partition separately. Specifically, the algorithm I use is:

0. Start with an initial experiment design d, and partition it into T elements so it can be represented as $d = \{d_t\}_{t=1}^T$
1. For each subset d_t of the partition, optimize the experiment with respect to the choice variables in this subset. Replace subset d_t with the optimized subset d_t^*.
2. Go back to step 1 until negligible improvements to the experiments are made

Exchange algorithms typically do not find a global optimum on one attempt, so it is recommended that this technique is performed with several initializations (Nguyen and Miller 1992).

3 Examples

3.1 Example 1: Designing a 2-decision experiment

To fix ideas, consider designing a simple experiment with just two binary choices. Our goal is to estimate as precisely as possible the parameter r_i in the constant relative risk-aversion (CRRA) expected utility function for every participant i:

$$EU(L) = E\left[X^{r_i} \mid L\right] \tag{5}$$

where $X \mid L$ is the lottery over monetary prizes induced by choosing option L.

In order to further simplify the problem, I will assume that one of the lotteries in each pair is set in stone, and so they will not be design choices. I will further assume that the length of the line segment connecting each lottery pair is not a design choice. That is, we are left with choosing d_1 and d_2 in the following lottery pairs:

$$\begin{array}{lll} \text{Pair 1} & \text{Lottery } L_1: & p_1 = 0.250,\ p_3 = 0.125 \\ & \text{Lottery } R_1: & p_1 = 0.250 + \rho \sin d_1,\ p_3 = 0.125 + \rho \cos d_1 \\ \text{Pair 2} & \text{Lottery } L_2: & p_1 = 0.125,\ p_3 = 0.250 \\ & \text{Lottery } R_2: & p_1 = 0.250 + \rho \sin d_2,\ p_3 = 0.125 + \rho d_2 \end{array} \tag{6}$$

where $\rho = 0.4$ is the length of the line segment connecting each lottery pair, and d_1 and d_2 are the angles to the horizontal of each of these lines. This design space $\mathbb{D}$ is shown graphically in Figure 2.

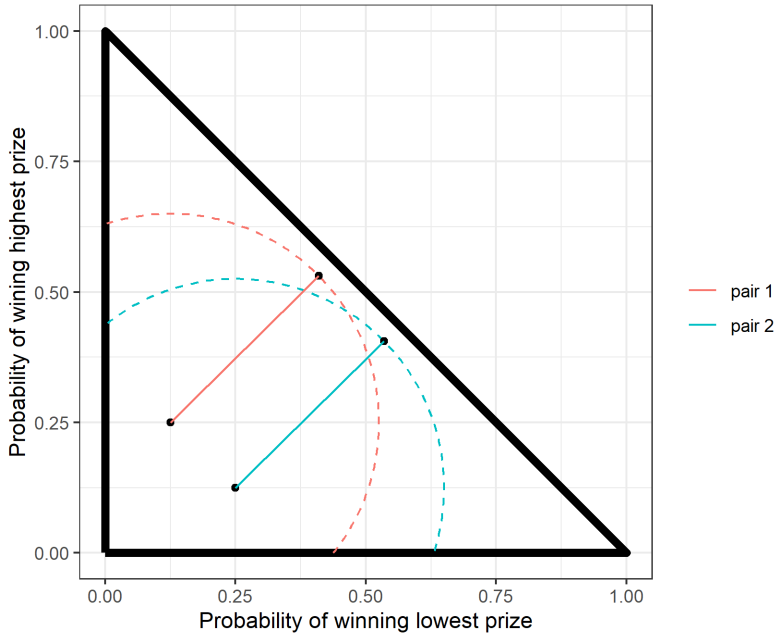

Figure 2: Design space for the simple 2-decision experiment

To further simplify things, we will assume that the three prizes are set to $\{1.0, 0.5, 0.0\}$, and we know that decision-makers will use logistic choice with choice precision parameter $\lambda = 30$. Hence, we can write the probability of choosing the left lottery as:

$$
\begin{aligned}
p(y_{i,t} = 1 \mid r_i) &= \Lambda\left[\lambda\left((p^L_{t,1} - p^R_{t,1}) + 0.5^{r_i}(p^L_{t,2} - p^R_{t,2}) + 0(p^L_{t,3} - p^R_{t,3})\right)\right] \\
&= \Lambda\left[\lambda\left((p^L_{t,1} - p^R_{t,1}) + 0.5^{r_i}(1 - p^L_{t,1} - p^L_{t,3} - 1 + p^R_{t,1} + p^R_{t,3})\right)\right] \\
&= \Lambda\left[\lambda\left((1 - 0.5^{r_i})(p^L_{t,1} - p^R_{t,1}) + 0.5^{r_i}(p^R_{t,3} - p^L_{t,3})\right)\right] \\
&= \Lambda\left[\lambda\rho\left(-(1 - 0.5^{r_i})\sin d_t + 0.5^{r_i}\cos d_t\right)\right]
\end{aligned} \tag{7}
$$

where $\Lambda(x) = (1 + \exp(-x))^{-1}$ is the inverse logit function.

Now we can write the log-likelihood function as:

$$
\mathcal{L}(r_i) = \sum_{t=1}^{2} y_{i,t} \log\left(p(y_{i,t} = 1 \mid r_i)\right) + (1 - y_i)\log(1 - p(y_{i,t} = 1 \mid r_i)) \tag{8}
$$

In order to find the information matrix, we need the derivative of the likelihood function:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(r_i)}{\partial r_i} &= \sum_{t=1}^{2}\left[\frac{y_{i,t}}{p(y_{i,t} = 1 \mid r_i)} - \frac{1 - y_{i,t}}{1 - p(y_{i,t} = 1 \mid r_i)}\right]\frac{\partial p(y_{i,t} = 1 \mid r_i)}{\partial r_i} \\
&= \sum_{t=1}^{2}\left[\frac{y_{i,t} - p(y_{i,t} = 1 \mid r_i)}{p(y_{i,t} = 1 \mid r_i)(1 - p(y_{i,t} = 1 \mid r_i))}\right]\frac{\partial p(y_{i,t} = 1 \mid r_i)}{\partial r_i}
\end{aligned} \tag{9}
$$

$$
\frac{\partial p(y_{i,t} = 1 \mid r_i)}{\partial r_i} = \lambda\rho(\sin\theta_t + \cos\theta_t)p(y_{i,t} = 1 \mid r_i)(1 - p(y_{i,t} = 1 \mid r_i))0.5^{r_i}\log(0.5) \tag{10}
$$

$$
\frac{\partial \mathcal{L}(r_i)}{\partial r_i} = \sum_{t=1}^{2}\left[y_{i,t} - p(y_{i,t} = 1 \mid r_i)\right]\lambda\rho(\sin\theta_t + \cos\theta_t)0.5^{r_i}\log(0.5) \tag{11}
$$

Finally, to determine the information matrix, we take the negative expectation of the derivative of this expression.

$$
\begin{aligned}
\mathcal{I}(r_i) &= E\left[\left(\frac{\partial \mathcal{L}(r_i)}{\partial r_i}\right)^2 \mid r_i\right] \\
&= E\left[\left(\sum_{t=1}^{2}\left[y_{i,t} - p(y_{i,t} = 1 \mid r_i)\right]\lambda\rho(\sin\theta_t + \cos\theta_t)0.5^{r_i}\log(0.5)\right)^2 \mid r_i\right] \\
&= \lambda^2\rho^2 0.5^{2r_i}\log^2(0.5)\sum_{t=1}^{2} p(y_{i,t} = 1 \mid r_i)(1 - p(y_{i,t} = 1 \mid r_i))(\sin\theta_t + \cos\theta_t)^2
\end{aligned} \tag{12}
$$

So the variance "matrix" of our estimator is:

$$\mathcal{V}(r_i \mid d) = \left[\lambda^2 \rho^2 0.5^{2r_i} \log^2(0.5) \sum_{t=1}^{2} p(y_{i,t} = 1 \mid r_i)(1 - p(y_{i,t} = 1 \mid r_i))(\sin\theta_t + \cos\theta_t)^2\right]^{-1} \tag{13}$$

Now suppose that we wish to estimate r_i as precisely as possible, and so choose a quadratic form for our utility function:

$$v(\tilde{r}_i \mid r_i) = -(\tilde{r}_i - r_i)^2 \tag{14}$$

The relevant derivatives of this function are:

$$\begin{aligned} \frac{\partial v(\tilde{r}_i \mid r_i)}{\partial \tilde{r}_i} &= -2(\tilde{r}_i - r_i) \\ \frac{\partial^2 v(\tilde{r}_i \mid r_i)}{\partial \tilde{r}_i \partial \tilde{r}_i} &= -2 \end{aligned} \tag{15}$$

and so:

$$E\left[v(\tilde{r}_i \mid r_i) \mid d, r_i\right] \approx \frac{-2}{\lambda^2 \rho^2 0.5^{2r_i} \log^2(0.5) \sum_{t=1}^{2} p(y_{i,t} = 1 \mid r_i)(1 - p(y_{i,t} = 1 \mid r_i))(\sin\theta_t + \cos\theta_t)^2} \tag{16}$$

which we are trying to maximize.

At this point, we need to assume a prior for r_i. For this, I will use $\log r_i \sim N(-0.87, 0.55^2)$. See Section 3.3.2 for more details about how this prior was calibrated. Figure 3 shows a contour plot of this approximated experimenter's utility function. The two global optima are marked with an "X". Here it should not be too surprising that there are *two* optima, because an expected utility-maximizer's utility is linear in probabilities. Hence, the left endpoints of the lottery pairs do not matter to such a participant. One of the global maximizers is $(d_1, d_2) = (0.464, 1.22)$, and the other simply re-labels the angles: $(d_1, d_2) = (1.22, 0.464)$.

Figure 4 shows the optimized experiment in the MM triangle. I show these lottery pairs (colored lines) alongside indifference curves for the median participant (dotted lines), which was $r = \exp(-0.87) \approx 0.42$. Note that the slope of the indifference curves is in between the slopes of the two lottery pairs. This is not a coincidence: these indifference curves represent the median risk preferences in our prior, and so having one pair steeper, and one pair flatter, than the indifference curves provides us with a switching point for the most likely part of the distribution of r_i.

3.2 Example 2: Desinging a battery of eighty binary lottery choices

Perhaps one of the most common applications of structural estimation using data from economic experiments is the estimation of risk preferences. A popular experiment design

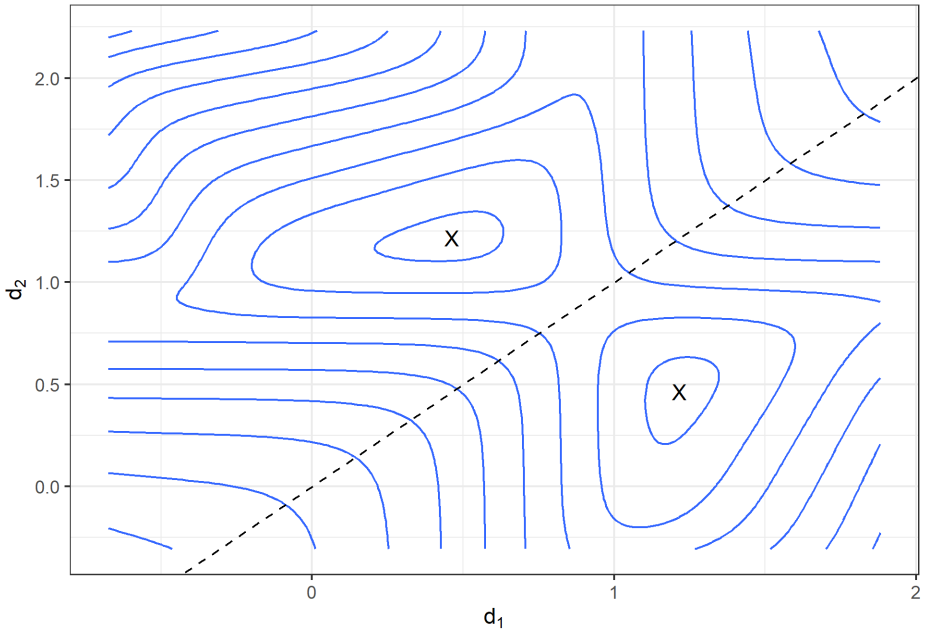

Figure 3: Utility function of the experiment design. Blue curves show contours. Black 'X's show global optima.

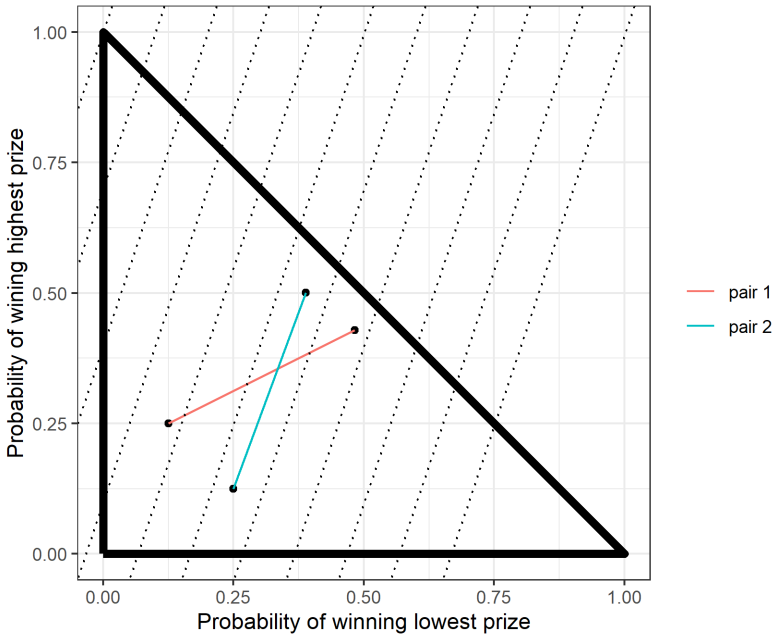

Figure 4: Optimized experiment with two decisions. Dots conneted with lines show lottery pairs. Dotted lines show indifference curves for an expected utility participant with the prior median level of risk aversion.

for this application is the battery of binary lottery choices. In this design, participants are presented with a number of pairwise choices over lotteries. These choices can then be used to estimate parameters in a model of decision-making under risk. In this section, I seek to design an experiment consisting of a battery of binary lottery choices in the MM triangle that estimates as precisely as possible the parameters in a parametric model of rank-dependent utility (RDU) maximization (Quiggin 1982) for individual participants. Here I choose the RDU model for two reasons. Firstly, it is a commonly-used extension of the expected utility model, and hence the resulting experiment design will be tuned to a common inferential goal. Secondly, the parametric model is sufficiently rich that, unlike the expected utility model, good rules of thumb for designing experiments with this inferential goal are not so obvious.[4]

I will restrict attention to designing an experiment with $T = 80$ decisions per participant, and the same three, equally-spaced, prizes for all lotteries $x = \{1.0, 0.5, 0.0\}$.[5] While these restrictions simplify the design problem considerably,[6] the benefit is that the experiment designs can be viewed in the same MM triangle without any loss of context. Furthermore, the experiment designs will be directly comparable to Harrison and Ng (2016), and so the simulation exercises below can provide us with a sense of scale for how much additional precision these experiments can bring when compared to an existing design. I further restrict the design space $\mathbb{D}$ to only include lottery pairs whose connecting line segment forms an angle to the horizontal of between 5° and 85°. This ensures that there is at least some minimal trade-off between the probability of winning the largest prize and the probability of winning the smallest prize.[7]

3.3 Model and likelihood function

The RDU model assumes that participants maximize "anticipated" utility. The anticipated utility of lottery X is:

$$U_i^X = \sum_{k=1}^{3} x_k^{r_i} \pi_{i,k}^X \tag{17}$$

where $\pi_{i,k}^X$ are "decision weights". If the decision weights were to equal the probabilities associated with eack prize x_k, then the model reduces to expected utility. Decision weights are calculated using a probability-weighting function $\omega_i(\cdot)$, which distorts cumulative probabilities

[4]For example, if we were designing an experiment in the MM triangle to estimate the parameters in an expected utility model, it is clear that we need a large variety of slopes of the line segments connecting lottery pairs. Furthermore, longer line segments should imply more precise choices, so these may also be desirable.

[5]As I am using the constant *relative* risk-aversion utility function coupled with the contextual utility normalization, payoff scale is irrelevant.

[6]This, for example, rules out the multiple-context designs of Harrison and Swarthout (2023) and Hey and Orme (1994), and rules out lotteries with more than three prizes.

[7]Earlier iterations of this design process, without this restriction, were resulting in experiment designs with some lottery pairs forming either vertical or horizontal line segments. Such pairs would likely provide much information about the choice precision parameter λ if the data-generating process were correctly specified, however it is unlikely that human participants would respond to these lottery pairs in the way that the model would predict, as there would be a first-order stochastically dominated lottery in these pairs.

relative to expected utility. Specifically, if the probabilities for lottery X are $\{p_1^X, p_2^X, p_3^X\}$, then the decision weights become:

$$\begin{aligned}\pi_{i,1}^X &= \omega_i(p_1^X) \\ \pi_{i,2}^X &= \omega_i(p_1^X + p_2^X) - \omega_i(p_1^X) \\ \pi_{i,3}^X &= \omega_i(p_1^X + p_2^X + p_3^X) - \omega_i(p_2^X + p_1^X)\end{aligned} \tag{18}$$

For the probability weighting function $\omega_i(\cdot)$, I use the flexible functional form proposed by Prelec (1998):

$$\omega_i(q) = \exp(-\eta_i(-\log q)^{\psi_i}) \tag{19}$$

where η_i and ψ_i are parameters to be estimated. Note that this model nests expected utility theory when $\eta_i = \psi_i = 1$. As with the previous example, I adopt a constant relative risk-aversion utility function $u_i(x) = x^{r_i}$ and a logit choice rule with precision parameter λ_i.

Letting $\theta_i = (r_i, \psi_i, \eta_i, \lambda_i)^\top$, and assuming that choices are independent conditional on θ, we can now write down the log-likelihood function as:

$$\mathcal{L}(\theta) = \sum_{t=1}^{80} \left[y_{i,t} \log\left(\Lambda(\Delta_{i,t})\right) + (1 - y_{i,t}) \log\left(1 - \Lambda(\Delta_{i,t})\right) \right] \tag{20}$$

where $y_{i,t} = 1$ (0) if participant i choice lottery L (R) for decision t, and:

$$\Delta_{i,t} = \lambda_i \sum_{k=1}^{3} x_k^{r_i} \left(\pi_{i,t,k}^L - \pi_{i,t,k}^R \right) \tag{21}$$

is the difference in anticipated utility between the two lotteries, inflated by choice precision λ_i.

From here we can compute the derivative of the log-likelihood function as follows:

$$\begin{aligned}\frac{\partial \mathcal{L}(\theta)}{\partial \theta} &= \sum_{t=1}^{80} \left[\frac{y_{i,t}}{\Lambda(\Delta_{i,t})} - \frac{1 - y_{i,t}}{1 - \Lambda(\Delta_{i,t})} \right] \Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t})) \frac{\partial \Delta_{i,t}}{\partial \theta} \\ &= \sum_{t=1}^{80} \left[y_{i,t} - \Lambda(\Delta_{i,t}) \right] \frac{\partial \Delta_{i,t}}{\partial \theta}\end{aligned} \tag{22}$$

And then we can determine the information matrix:

$$
\begin{aligned}
\mathcal{I}(\theta) &= E\left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta}\frac{\partial \mathcal{L}(\theta)}{\partial \theta^\top} \mid \theta\right] \\
&= E\left(\left[\sum_{t=1}^{80}\left[y_{i,t} - \Lambda(\Delta_{i,t})\right]\frac{\partial \Delta_{i,t}}{\partial \theta}\right]\left[\sum_{s=1}^{80}\left[y_{i,s} - \Lambda(\Delta_{i,s})\right]\frac{\partial \Delta_{i,s}}{\partial \theta^\top}\right] \mid \theta\right) \\
&= \sum_{t=1}^{80}\sum_{s=1}^{80} E\left[(y_{i,t} - \Lambda(\Delta_{i,t}))(y_{i,s} - \Lambda(\Delta_{i,s})) \mid \theta\right]\frac{\partial \Delta_{i,t}}{\partial \theta}\frac{\partial \Delta_{i,t}}{\partial \theta^\top} \\
&= \sum_{t=1}^{80}\Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t}))\frac{\partial \Delta_{i,t}}{\partial \theta}\frac{\partial \Delta_{i,t}}{\partial \theta^\top}
\end{aligned}
\tag{23}
$$

where the final line follows by noting that $E\left[(y_{i,t} - \Lambda(\Delta_{i,t}))(y_{i,s} - \Lambda(\Delta_{i,s})) \mid \theta\right] = \text{cov}(y_{i,t}, y_{i,s} \mid \theta)$, which is zero except for when $s = t$, because choices are assumed to be independent conditional on θ. Evaluating the derivatives $\frac{\partial \Delta_{i,t}}{\partial \theta}$ is straightforward but tedious, so their derivations are relegated to Appendix A.

3.3.1 Design goals

In total, I design four experiments. Three of these are designs tuned for specific inferential goals, and the fourth is a randomly-designed experiment to provide a sense of scale for the worst possible implementation. Other than the random design, the designs differ in the specification of the experimenter's utility function $v(\tilde{\theta} \mid \theta)$.

First, I design an experiment to estimate all four parameters as precisely as possible. To do this, I use a quadratic loss function for the errors between actual and estimated parameters, that is:[8]

$$
v^{\text{FULL}}(\tilde{\theta} \mid \theta) = -(\tilde{\theta} - \theta)^\top(\tilde{\theta} - \theta) \tag{24}
$$

Next, I design an experiment to estimate a certainty equivalent as precisely as possible. Specifically, I aim to estimate the certainty equivalent of a lottery that pays out \$30 with probability 50% and \$0 otherwise. This certainty equivalent is equal to:

$$
CE(\theta) = 30\omega_i(0.5)^{\frac{1}{r_i}} \tag{25}
$$

Note in particular that the certainty equivalent only includes three of the four model parameters: it does not include choice precision λ_i. For this design, I use a quadratic loss function over the certainty equivalent:

$$
v^{\text{CE}}(\tilde{\theta} \mid \theta) = -\Big(CE(\tilde{\theta}) - CE(\theta)\Big)^2 \tag{26}
$$

I also design an experiment to estimate the Prelec probability weighting parameters as precisely as possible. This design may be useful if one is wishing to distinguish between

[8] This design goal is equivalent to designing an $\mathcal{A}$-optimal experiment.

RDU and expected utility types. For this, I define a quadratic loss function over just these parameters:

$$v^{\text{PRELEC}}(\tilde{\theta} \mid \theta) = -(\tilde{\eta} - \eta)^2 - (\tilde{\psi} - \psi)^2 \tag{27}$$

Finally, I "design" an experiment by randomly selecting 80 lottery pairs in the MM triangle. To do this, I draw each lottery's probabilities independently from the Dirichlet$(1, 1, 1)$ distribution.

3.3.2 Calibrating a prior to an existing experiment

Evaluating the expectation in $E[v(\tilde{\theta} \mid \theta)]$ requires the prior information $p(\theta)$, which describes our belief about our model's parameters before designing and running our experiment. One option here is to elicit such a prior based on our understanding of typical parameter values, and the behavior they imply. For example Bland (2025a) describes a process for assigning appropriate priors for estimating a parametric model of expected utility for a single participant. In this process, one investigates the implications of a candidate prior for the distribution of quantities of interest to the researcher. This could be particularly useful if there do not exist data that are appropriately informative of typical parameter values for the model one wishes to estimate.

A second option is to use data from a previously-run experiment (or experiments) to *estimate* an appropriate prior. Bayesian hierarchical models lend themselves especially well to this exercise, and this is the route I choose for this example. In particular, an appropriately-specified Bayesian hierarchical model will estimate parameters governing the population-level distribution of θ. From here, it is straightforward to simulate draws from this population distribution that can be used for the Monte Carlo integration described above.

3.3.2.1 Data I use data from Harrison and Swarthout (2023) to estimate this model. This dataset is ideal for estimating a hierarchical model for (at least) two reasons. Firstly, with 175 undergraduate participants,[9] to my knowledge this represents one of the largest participant groups studied in such an experiment. As such, it provides us with much information about the population-level parameters that we need in order to evaluate $E[v(\tilde{\theta} \mid \theta)]$. Secondly, since each participant faced a battery of 100 pairwise lottery choices, there should be sufficient information about individual-level parameters to inform our estimates of the population-level quantities. Some lotteries in this experiment were framed as losses or mixed gains and losses, with an endowment added to ensure earnings were not negative. For these lotteries, I add this endowment so that all prizes are weakly positive.

3.3.2.2 Modeling the data-generating process At the individual level, I assume that participants' choices conditional on their individual-level parameters are generated from the RDU model with CRRA utility and Prelec (1998) probability weighting function described above. In order to estimate the choice precision parameter λ_i on a scale that will

[9] I exclude the MBA students and the "earned endowment" undergraduates in this study from my analysis.

be appropriate to take through the experiment design process, I use the contextual utility normalization proposed by Wilcox (2011). This ensures that anticipated utility differences are scaled to be between negative one and one, which also matches the normalization made for the experiment design above.

I then make the following transformation of this model's parameters in order to permit a multivariate normal hierarchical specification:

$$\begin{aligned} \theta_i &= \begin{pmatrix} \log(r_i) \\ \log(\lambda_i) \\ \log(\eta_i) \\ \log(\psi_i) \end{pmatrix} \sim N\left(\mu, \Sigma\right) \\ \Sigma &= \text{diag\_matrix}(\tau)\Omega\text{diag\_matrix}(\tau) \end{aligned} \tag{28}$$

where μ is the population mean vector of the transformed parameters, τ is the population standard deviation vector, and Ω is the correlation matrix. Here the natural logarithm transform $\log(\cdot)$ ensures that the individual-level parameters are strictly positive. Decomposing the variance-covariance matrix Σ into τ, a vector of (marginal) standard deviations, and Ω, a correlation matrix, will permit thinking about assigning priors to these parameters independently.[10]

3.3.2.3 Assigning hyper-priors to the hyper-parameters The distribution of the model's individual-level parameters θ is governed by the model's "hyper-parameters" μ, τ, and Ω. These are the parameters that we must assign "hyper-priors" to in order to estimate the model. Following the recommendation in Section 1.13 of the *Stan User's Guide* (Stan Development Team 2022), I adopt normal, half Cauchy, and LKJ priors for these hyperparameters, respectively:

$$\begin{aligned} \mu_r &\sim N(\log(0.5), 10^2) \\ \mu_\lambda &\sim N(0, 10^2) \\ \mu_\eta &\sim N(0, 10^2) \\ \mu_\psi &\sim N(0, 10^2) \end{aligned} \tag{29}$$

For the vector of standard deviations, I assign the following prior:

$$\tau_k \sim \text{Cauchy}^+(1), \quad k \in \{r, \lambda, \eta, \rho\} \tag{30}$$

For the correlation matrix Ω, I assign a prior of

$$\Omega \sim \text{LKJ}(4) \tag{31}$$

[10] See Section 1.13 of the *Stan User's Guide* (Stan Development Team 2022) for further discussion of this specification.

Table 1: Posterior estimates from the hierarchical model. Posterior means with posterior standard deviations in parentheses.

	$\log(r)$	$\log(\lambda)$	$\log(\eta)$	$\log(\psi)$
μ	-0.875 (0.051)	2.56 (0.039)	-0.304 (0.035)	-0.145 (0.034)
τ	0.555 (0.043)	0.436 (0.035)	0.348 (0.033)	0.376 (0.029)
corr: r		-0.129 (0.103)	-0.018 (0.12)	-0.085 (0.098)
corr: λ			-0.028 (0.117)	-0.062 (0.104)
corr: η				-0.125 (0.108)

This achieves a prior marginal distribution of the correlation between any two (transformed) individual-level parameters with a mean, median, and mode of zero, but with a good amount of density assigned to correlations substantially different to zero.[11]

3.3.2.4 Estimates I simulate the posterior distribution of the model described above using *Stan* (Carpenter et al. 2017). A summary of the population-level estimates is shown in Table 1. Here the first two rows show estimates of the parameters governing the marginal distribution of the individual-level parameters. Specifically, μ is the population mean of the transformed variables, and τ is the population standard deviation of the transformed variables. We can interpret μ as an estimate of the parameters for the median, or "representative" participant. For simplicity, using just the posterior mean point estimate of this parameter, this median participant is (utility) risk-averse with $r = \exp(-0.875) \approx 0.41$.

In order to evaluate $E[v(\tilde{\theta}, \theta)]$, I also simulate draws of the individual-level parameters θ from this posterior distribution. These draws are *not* the posterior "shrinkage" estimates for the participants in Harrison and Swarthout (2023), which could be denoted $\theta_i \mid y, i$. Instead, they are new draws from the estimated population distribution $\theta_j \mid y$. That is, the former is a posterior draw for a participant i who participated in Harrison and Swarthout (2023). Hence, we can condition their draw not only on the data y in general, but also their specific decisions y_i. On the other hand, $\theta_j \mid y$ is a posterior draw for somebody in the same subject pool as Harrison and Swarthout (2023), but who did not participate in this experiment. Hence, if the subject pool for the experiments I design here is not too different from the subject pool of Harrison and Swarthout (2023), then $E[v(\tilde{\theta}, \theta)]$ will be well-approximated and the experiment will be well-tuned to identifying the model's parameters.

3.3.3 Procedures and experiment designs

I implement the exchange algorithm described above in *R* (R Core Team 2024). Specifically, I partition the experiment design into the 80 lottery pairs presented to participants. In each

[11] See Section 6.4.1 of Bland (2025b) for an explanation of the LKJ prior, and in particular Figure 6.1, which shows the marginal distribution of an individual correlation for different parameterizations.

step of the exchange algorithm, the procedure chooses probabilities for this lottery pair to maximize the approximated utility function. The algorithm then repeats this procedure for ten iterations over each of the 80 decisions. This was enough for the objective functions to converge to reasonable tolerances. I use this algorithm with five different randomly-generated starting values for each experiment design, and choose the design with the greatest objective function. In practice, a re-parameterization from four probabilities (p_1^L, p_3^L, p_1^R, p_3^R) to the polar-like coordinates used in the first example proved helpful. In particular, the re-parameterization used the coordinate of the Left lottery (p_1^L, p_3^L), and defined the coordinate of the Right lottery with an angle to the horizontal and the length of the line segment connecting the two lotteries.

Figure 5 shows the resulting experiment designs, alongside Harrison and Ng (2016), in the MM triangle.

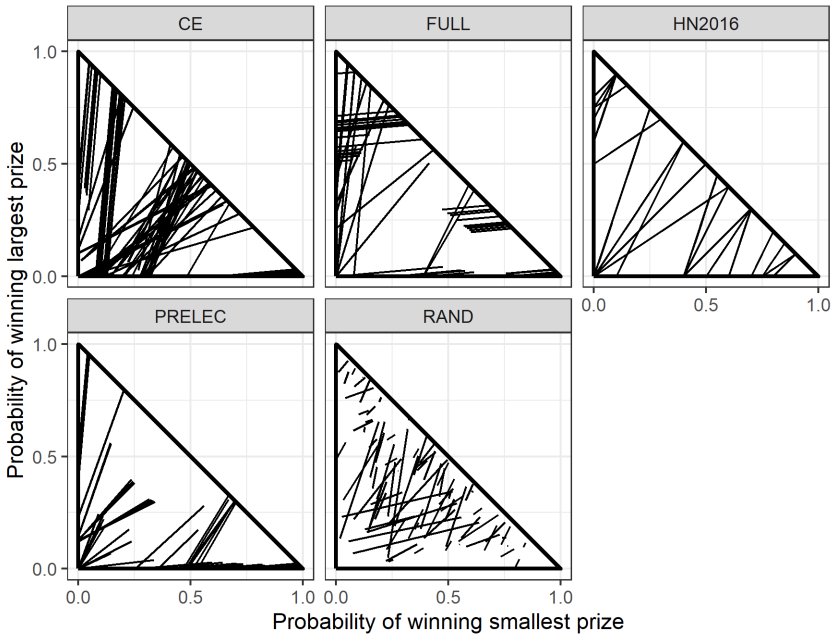

Figure 5: Experiment designs depicted in the MM triangle. Note that some lottery pairs are obscured in Harrison and Ng (2016) due to overlapping connecting line segments.

3.3.4 Evaluating the experiment designs

In order to evaluate the experiment designs, I explore the performance of the maximum likelihood estimator of the RDU model in achieving three inferential goals. That is, I assess how well the estimator:

1. Recovers the true parameters that generated the data,
2. Recovers the true certainty equivalent implied by the parameters that generated the data, and
3. Distinguishes between the RDU model and the expected utility model using a Wald test.

Table 2: Median prediction error of the maximum likelihood estimator

	Quantity				
experiment	r	λ	η	ψ	CE
CE	0.118	3.739	0.245	0.178	1.558
FULL	0.121	3.174	0.199	0.162	2.526
HN2016	0.198	6.503	0.328	0.188	1.860
PRELEC	0.121	3.331	0.157	0.124	2.075
RAND	0.252	13.049	0.714	0.436	5.349

I investigate how well the five experiment designs in Figure 5 achieve these inferential goals through Monte Carlo simulation. Specifically, for each experiment design I draw 10,000 parameter vectors from the calibrated prior. For each of these vectors, I simulate one dataset and then estimate the RDU model using maximum likelihood estimation.[12]

As maximum likelihood estimation in experiments like these is prone to outliers, I focus on the *median* rather than *mean* prediction errors. These are shown in Table 2. Each row of this Table is an experiment design, and each column reports the median absolute error in estimating a parameter or the certainty equivalent. Here smaller numbers indicate better estimates. To begin with the first inferential goal, the FULL experiment mostly does better than other experiments at estimating the model's fundamental parameters r, λ, η, and ψ. The exception to this is for parameters η and ψ, which are best estimated by the PRELEC experiment. This should be unsurprising, as the PRELEC experiment was tuned to estimate these two parameters as precisely as possible.

Now to turn to the second inferential goal: recovering a certinaty equivalent. The right-most column of Table 2 shows the median prediction error for the estimated certainty equivalent. This is best estimated using the CE design. Coming in third by this measure, and closely behind the "PRELEC" design, is the Harrison and Ng (2016) design (HN2016). That this human-designed experiment does relatively well here should not be too surprising, as the goal of this battery was in fact to estimate certainty equivalents (although not necessarily the one studied here).

Finally, turning to the third inferential goal, I perform a Wald test that the parameter restriction on the Prelec probability weighting function implies expected utility maximization ($\eta = \psi = 1$). The rejection probability as a function of the Euclidean distance the true parameters are from this restriction (i.e. $\sqrt{(\eta-1)^2 + (\psi-1)^2}$) is shown in Figure 6. As *none* of the simulated participants were expected utility, we hope to reject this null, and so larger numbers are better. Here the PRELEC experiment (blue line) performs the best, which is comforting as this was the experiment design tuned to estimating these parameters as precisely as possible. Of particular interest here is the comparison to Harrison and Ng (2016) (green line). The properties of this test using the Harrison and Ng (2016) design

[12]As is common with maximum likelihood estimation of these models, some did not converge. These estimations were thrown out.

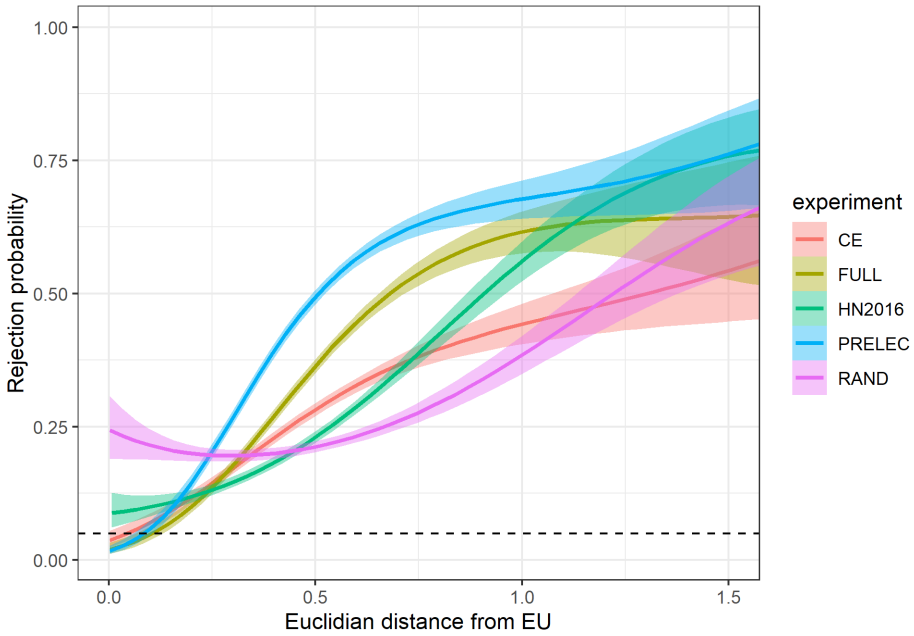

Figure 6: Rejection probability for a Wald test that $\eta = \psi = 1$ (quartic logit fit). Dashed line shows the nonimal test size (5%).

were simulated in Monroe (2020), who found that rejection rates were well below generally accepted levels. The PRELEC design here perhaps provides an upper bound to these rejection rates that can be achieved by re-designing the lotteries in this experiment.

4 Conclusion

A good experiment design for estimating a parametric model is not necessarily the same as a good experiment design for other purposes. This paper describes methods for achieving this first goal. In particular, we can use our prior knowledge of reasonable parameter values to tune our experiment design toward various inferential goals. These goals could include (i) estimating all parameters in a model as precisely as possible, (ii) estimating a subset of parameters in the face of other "nuisance" parameters, (iii) testing a parameter restriction, or (iv) estimating a transformation of the parameters. However these goals are not necessarily aligned. The lottery experiment designed to have overall the most precisely-estimated parameters did not distinguish between models as well as an experiment designed specifically to do this, but this second experiment sacrificed precision in other areas. For estimating parametric models, at least, we need to be very specific about the research question(s) we wish to answer *before* we design our experiments.

While I have focused exclusively here on estimating separate models for each participant in an experiment, these techniques can also be applied to experiments where the inferential goal is to learn about the population of participants. Although the task would be computationally more intensive, it is not conceptually difficult to optimize an experiment for estimating

(say) a hierarchical model (e.g. Gao, Harrison, and Tchernis (2023)) or a mixture model (e.g. Harrison and Rutström (2009)). Even if one (or one's computer) cannot directly optimize an experiment for a hierarchical model, optimizing an experiment to estimate parameters precisely at the individual level could prove to be an adequate second-best solution. Specifically, the information matrix of the individual-level model measures how informative an experiment is of the individual-level parameters. Loosely speaking, if there is less uncertainty in these individual-level parameters, then there will be less uncertainty carried through to our population-level parameters.

In this paper, I focus the examples on experiments involving a battery of binary lottery choices. I do this due to this experiment type's popularity in estimating structural models. However the techniques outlined here need not be confined to this class of experiment, or the models that are typically estimated from them. For example, these techniques could easily be applied to improve on the choices presented in Bruhin, Fehr, and Schunk (2019) to elicit other-regarding preferences. Other improvements to our estimates' precision could also be made in convex budget set experiments like Andreoni and Miller (2002) and Halevy, Persitz, and Zrill (2018). In these experiments there are typically fewer decisions made by participants, and so every decision presented to a participant comes at a premium.

Appendices

A – Properties of the likelihood function

Letting $\Delta_{i,t}$ be the utility difference between the two options scaled by λ_i, the likelihood function is:

$$\log p(y_i \mid \theta) = \sum_{t=1}^{T} \left[y_{i,t} \log \Lambda(\Delta_{i,t}) + (1 - y_{i,t}) \log(1 - \Lambda(\Delta_{i,t})) \right]$$

$$\begin{aligned}
\frac{\partial \log p(y_i \mid \theta)}{\partial \theta} &= \sum_{t=1}^{T} \left[\frac{y_{i,t}}{\Lambda(\Delta_{i,t})} - \frac{1 - y_{i,t}}{1 - \Lambda(\Delta_{i,t})} \right] \Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t})) \frac{\partial \Delta_{i,t}}{\partial \theta} \\
&= \sum_{t=1}^{T} \frac{y_{i,t} - y_{i,t}\Lambda(\Delta_{i,t}) - \Lambda(\Delta_{i,t}) + y_{i,t}\Lambda(\Delta_{i,t})}{\Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t}))} \Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t})) \frac{\partial \Delta_{i,t}}{\partial \theta} \\
&= \sum_{t=1}^{T} \left[y_{i,t} - \Lambda(\Delta_{i,t}) \right] \frac{\partial \Delta_{i,t}}{\partial \theta}
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}(\theta) &= E \left[\frac{\partial \log p(y_i \mid \theta)}{\partial \theta} \frac{\partial \log p(y_i \mid \theta)}{\partial \theta^\top} \mid \theta \right] \\
&= \sum_{t=1}^{T} E \left[\left(y_{i,t} - \Lambda(\Delta_{i,t}) \right)^2 \frac{\partial \Delta_{i,t}}{\partial \theta} \frac{\partial \Delta_{i,t}}{\partial \theta^\top} \mid \theta \right] \\
&= \sum_{t=1}^{T} \Lambda(\Delta_{i,t})(1 - \Lambda(\Delta_{i,t})) \frac{\partial \Delta_{i,t}}{\partial \theta} \frac{\partial \Delta_{i,t}}{\partial \theta^\top}
\end{aligned}$$

where the second line follows by assuming that choices are independent conditional on θ. Now we can focus on the derivatives $\frac{\partial \Delta_{i,t}}{\partial \theta}$, where:

$$\Delta_{i,t} = \lambda \sum_{j=1}^{3} x_j^r (\pi_j^L(p_t;\eta,\psi) - \pi_j^R(p_t;\eta,\psi))$$

$$\frac{\partial \Delta_{i,t}}{\partial \lambda} = \sum_{j=1}^{3} x_j^r (\pi_j^L(p_t;\eta,\psi) - \pi_j^R(p_t;\eta,\psi))$$

$$\frac{\partial \Delta_{i,t}}{\partial r} = \lambda \sum_{j=1}^{3} x_j^r \log(x_j)(\pi_j^L(p_t;\eta,\psi) - \pi_j^R(p_t;\eta,\psi))$$

For $k \in \{\eta, \psi\}$

$$\frac{\partial \Delta_{i,t}}{\partial k} = \lambda \sum_{j=1}^{3} x_j^r \left(\frac{\partial \pi_j^L(p_t;\eta,\psi)}{\partial k} - \frac{\partial \pi_j^R(p_t;\eta,\psi)}{\partial k} \right)$$

where:

$$\pi_j(p;\eta,\psi) = \omega(\overline{p}_j;\eta,\psi) - \omega(\overline{p}_{j-1};\eta,\psi)$$

where $\overline{p}$ are the lottery's *cumulative* probabilities.

$$\begin{aligned}
\omega(\overline{p};\eta,\psi) &= \exp(-\eta(-\log \overline{p})^{\psi}) \\
\frac{\partial \omega(\overline{p};\eta,\psi)}{\partial \eta} &= -(-\log \overline{p})^{\psi} \omega(\overline{p};\eta,\psi) \\
\frac{\partial \omega(\overline{p};\eta,\psi)}{\partial \psi} &= \eta \log(-\log \overline{p})(-\log \overline{p})^{\psi} \omega(\overline{p};\eta,\psi)
\end{aligned}$$

Alternatively:

$$\begin{aligned}
\omega &= \exp(-\eta(-\log p)^{\psi}) \\
-\log \omega &= \eta(-\log p)^{\psi} \\
\log(-\log \omega) &= \log \eta + \psi \log(-\log p) \\
\frac{\partial \log(-\log \omega)}{\partial x} &= \frac{1}{\omega \log \omega} \frac{\partial \omega}{\partial x} \\
\frac{\partial \omega}{\partial \eta} &= \frac{\omega \log \omega}{\eta} \\
\frac{\partial \omega}{\partial \psi} &= \omega \log \omega \log(-\log p)
\end{aligned}$$

Or for the Kumaraswamy weighting function:

$$\begin{aligned}\omega(\overline{p};a,b) &= 1-(1-\overline{p}^a)^b \\ \frac{\partial\omega(\overline{p};a,b)}{\partial a} &= b\overline{p}^a\log(\overline{p})(1-\overline{p}^a)^{b-1} \\ \frac{\partial\omega(\overline{p};a,b)}{\partial b} &= -(1-\overline{p}^a)^b\log(1-\overline{p}^a)\end{aligned}$$

B – Experimenter utility functions

B.1 – Certainty equivalent

Suppose we consider the goal of estimating a certainty equivalent for a lottery which pays out $\overline{x}>0$ with probability p, and 0 with probability $1-p$.

$$\begin{aligned}CE(r,\eta,\psi) &= [\omega(p;\eta,\psi)\overline{x}^r]^{\frac{1}{r}} \\ &= \overline{x}\omega(p;\eta,\psi)^{\frac{1}{r}}\end{aligned}$$

$$\frac{\partial CE}{\partial r} = -\overline{x}\frac{\omega(p;\eta,\psi)^{\frac{1}{r}}\log\omega(p;\eta,\psi)}{r^2}$$

for $k\in\{\eta,\psi\}$:

$$\frac{\partial CE}{\partial k} = \frac{\overline{x}\omega(p;\eta,\psi)^{\frac{1}{r}-1}}{r}\frac{\partial\omega(p;\eta,\psi)}{\partial k}$$

and:

$$\frac{\partial CE}{\partial\lambda} = 0$$

Now, define the experimenter's utility function as the negative squared difference between the estimated CE and the true CE:

$$v(\tilde{\theta}\mid\theta) = -\left(CE(\tilde{r},\tilde{\eta},\tilde{\psi})-CE(r,\eta,\psi)\right)^2$$

$$\begin{aligned}\frac{\partial v(\tilde{\theta}\mid\theta)}{\partial\tilde{\theta}} &= -2\left(CE(\tilde{r},\tilde{\eta},\tilde{\psi})-CE(r,\eta,\psi)\right)\frac{\partial CE(\tilde{r},\tilde{\eta},\tilde{\psi})}{\partial\tilde{\theta}} \\ \frac{\partial^2 v(\tilde{\theta}\mid\theta)}{\partial\tilde{\theta}\partial\tilde{\theta}^\top} &= -2\frac{\partial CE(\tilde{r},\tilde{\eta},\tilde{\psi})}{\partial\tilde{\theta}}\frac{\partial CE(\tilde{r},\tilde{\eta},\tilde{\psi})}{\partial\tilde{\theta}^\top} - 2\left(CE(\tilde{r},\tilde{\eta},\tilde{\psi})-CE(r,\eta,\psi)\right)\frac{\partial^2 CE(\tilde{r},\tilde{\eta},\tilde{\psi})}{\partial\tilde{\theta}\partial\tilde{\theta}^\top}\end{aligned}$$

and we need to evaluate this at $\tilde{\theta}=\theta$, so :

$$\frac{\partial^2 v(\theta \mid \theta)}{\partial \tilde{\theta} \partial \theta^\top} = -2 \frac{\partial CE(r, \eta, \psi)}{\partial \theta} \frac{\partial CE(r, \eta, \psi)}{\partial \theta^\top}$$

References

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–53.

Bland, James R. 2025a. "Some Guidance for the Choice of Priors for Bayesian Structural Models in Economic Experiments." *Journal of the Economic Science Association*, 1–10.

———. 2025b. "Structural Bayesian Techniques for Experimental and Behavioral Economics." *Available at SSRN 5241268*. https://jamesblandecon.github.io/StructuralBayesianTechniques/section.html.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. "The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences." *Journal of the European Economic Association* 17 (4): 1025–69.

Camerer, Colin F. 1989. "An Experimental Test of Several Generalized Utility Theories." *Journal of Risk and Uncertainty* 2: 61–104.

———. 1992. "Recent Tests of Generalizations of Expected Utility Theory." *Utility Theories: Measurements and Applications*, 207–51.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Gao, Xiaoxue Sherry, Glenn W Harrison, and Rusty Tchernis. 2020. "Estimating Risk Preferences for Individuals: A Bayesian Analysis." Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

———. 2023. "Behavioral Welfare Economics and Risk Preferences: A Bayesian Approach." *Experimental Economics* 26: 273–303. https://doi.org/10.1007/s10683-022-09751-0.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. "Parametric Recoverability of Preferences." *Journal of Political Economy* 126 (4): 1558–93.

Harrison, Glenn W, and Jia Min Ng. 2016. "Evaluating the Expected Welfare Gain from Insurance." *Journal of Risk and Insurance* 83 (1): 91–120.

Harrison, Glenn W, and E Elisabet Rutström. 2009. "Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral." *Experimental Economics* 12 (2): 133–58.

Harrison, Glenn W, and Todd Swarthout. 2023. "Cumulative Prospect Theory in the Laboratory: A Reconsideration." In *Research in Experimental Economics: Models of Risk Preferences: Descriptive and Normative Challenges*, edited by G. W. Harrison and D. Ross. Bingley, UK: Emerald.

Hey, John D, and Daniela Di Cagno. 1990. "Circles and Triangles an Experimental Estimation of Indifference Lines in the Marschak-Machina Triangle." *Journal of Behavioral Decision Making* 3 (4): 279–305.

Hey, John D, and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica: Journal of the Econometric Society*, 1291–1326.

Holt, Charles A, and Susan K Laury. 2002. "Risk Aversion and Incentive Effects." *American*

Economic Review 92 (5): 1644–55.

Loomes, Graham, and Robert Sugden. 1998. "Testing Different Stochastic Specificationsof Risky Choice." *Economica* 65 (260): 581–98.

Moffatt, Peter G. 2007. "Optimal Experimental Design in Models of Decision and Choice." *MEASUREMENT IN ECONOMICS*, 357.

Monroe, Brian Albert. 2020. "The Statistical Power of Individual-Level Risk Preference Estimation." *Journal of the Economic Science Association* 6 (2): 168–88.

Nguyen, Nam-Ky, and Alan J Miller. 1992. "A Review of Some Exchange Algorithms for Constructing Discrete d-Optimal Designs." *Computational Statistics & Data Analysis* 14 (4): 489–98.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica*, 497–527.

Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior & Organization* 3 (4): 323–43.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stan Development Team. 2022. "Stan User's Guide, Version 2.31."

Wilcox, Nathaniel T. 2011. "'Stochastically More Risk Averse:'a Contextual Theory of Stochastic Discrete Choice Under Risk." *Journal of Econometrics* 162 (1): 89–104.

———. 2023. "Unusual Estimates of Probability Weighting Functions." In *Models of Risk Preferences: Descriptive and Normative Challenges*, 69–106. Emerald Publishing Limited.