Latent Process Heterogeneity in Discounting Behavior

by

Maribeth Coller, Glenn W. Harrison and E. Elisabet Rutström†

April 2010

Forthcoming, *Oxford Economic Papers*

Abstract. We show that observed choices in discounting experiments are consistent with roughly one-half of the subjects using exponential discounting and one-half using quasi-hyperbolic discounting. We characterize the latent data generating process using a mixture model which allows different subjects to behave consistently with each model. Our results have substantive implications for the assumptions made about discounting behavior, and also have significant methodological implications for the manner in which we evaluate alternative models when there may be complementary data generating processes.

JEL Classification Codes: D91, C91
Keywords: time preferences, discounting, experimental methods

† School of Accounting, Moore School of Business, University of South Carolina, USA (Coller); Department of Risk Management & Insurance and CEAR, Robinson College of Business, Georgia State University, USA (Harrison), and Department of Economics, College of Business Administration, University of Central Florida, USA (Rutström). E-mail: MBETH@MOORE.SC.EDU, GHARRISON@GSU.EDU and ERUTSTROM@BUS.UCF.EDU. Helpful comments have been received from Steffen Andersen, John Hey, Morten Lau, Melonie Sullivan, seminar participants, and two referees. We are grateful for funding from the Friends of the Accounting Department, School of Accounting, Moore School of Business, University of South Carolina. We thank the U.S. National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746.

Two major hypotheses about discounting behavior dominate economic analysis. The traditional exponential model assumes that agents have constant discount rates for different time horizons. The quasi-hyperbolic model holds that the individual has a "fixed premium" that applies to money deferred for some initial time delay, but has constant discount rates for any monies deferred for further horizons beyond the initial delay.[1] The empirical evidence is widely viewed as supporting the quasi-hyperbolic model over the exponential model.[2] We present evidence that suggests a richer characterization of the data: *that choices are divided roughly equally between those that follow the exponential model and those that follow the quasi-hyperbolic model.*

In retrospect, this finding should not be a surprise, since preferences are subjective. It is perfectly reasonable that some people make choices that are best characterized by an exponential model and that other people make choices that are best characterized by a quasi-hyperbolic model. None of the protagonsists in the debates over these alternative models would, we think, go so far as to claim that *every* choice by *every* subject was best characterized as generated by *just one* of the processes.

Given this uncontroversial proposition, how is one to evaluate if there is heterogeneity in the *type* of discounting function employed by different individuals? One natural approach is to collect enough responses from each individual, estimate different models for each individual (or one general model), and then tally up how many individuals are best characterized by one model or the other. For models of decision making under risk, this approach was employed by Hey and Orme [1994]. For models of discounting, this approach was employed by Benhabib, Bisin and Schotter [2009].

Unfortunately, this approach is not always attractive. It may not always be feasible to ask so many questions of each individual to be able to estimate flexible structural models, particularly if one

---

[1] So-called "quasi-hyperbolic discount functions" were introduced by Phelps and Pollack [1968], and have been adopted by Laibson [1997], O'Donoghue and Rabin [1999], Angeletos et al. [2001] and others. Such specifications have been given various names. For example, O'Donoghue and Rabin [1999] call them "present-biased preferences" to emphasize that they reflect differences in pure rates of time preference. Others call them simply "(β, δ) preferences."

[2] See Frederick, Shane; Loewenstein, George, and O'Donoghue [2002; §4.1, §5.1] for a review.

wants procedures that generalize from the lab to the field. Moreover, individual-level estimation can often lead to implausible estimates, given the random nature of behavior that is tightly correlated with the individual generating the entire estimation sample.[3] In fact, Benhabib, Bisin and Schotter [2009] admit that in "our data the annual discount rate is on the order of 472%." This is *wildly* different from the roughly 16% p.a. inferred by Coller and Williams [1999], the 28% p.a. inferred by Harrison, Lau and Williams [2002], and the 10% inferred by Andersen, Harrison, Lau and Rutström [2008]. Of course, many things differ across these studies, most notably the use of a front end delay on the earlier payment of money and the correction for non-linear utility functions in the final study. Related to this last correction, which causes a huge drop in the elicited discount rate, Benhabib, Bisin and Schotter [2009; Tables 7, 8] report virtually *no risk aversion* in their data, in direct conflict with *every* rigorous experimental study.[4]

We propose an approach that is complementary to estimating at the level of the individual, and that still recognizes that different choices and different individuals might employ different models. Our approach is to use data that is pooled across individuals, and to estimate a finite *mixture model* using maximum likelihood. We allow the observed choices to be generated by two latent data generating processes, corresponding to the exponential and quasi-hyperbolic models.[5] The empirical model allows each model to receive support, or allows one to dominate if that best characterizes observed behavior.

Our approach is complementary in another sense, even if one does have access to enough data to estimate at the level of the individual. The structural econometric models required to allow

[3] At a technical level, the problem arises from the imprecision of their estimates of the parameter controlling curvature of the discounting function in their specification, θ. For example, review the estimates of θ in Benhabib, Bisin and Schotter [2009; Tables 2 - 9], many of which are statistically insignificant, negative, or extremely high. There are also many instances where they honestly report that their estimation algorithm simply "failed to converge."

[4] For example, see Holt and Laury [2002][2005], Harrison, Johnson, McInnes and Rutström [2005], and Harrison and Rutström [2008] for a review.

[5] One could readily extend this approach to consider alternative parameterizations of hyperbolic preferences, noted below. We restrict attention to two major competing models in order to illustrate the basic methodological point.

for non-linear utility and alternative discounting functions are highly non-linear in the key parameters. Hence one would naturally expect numerical sensitivity of estimates. It is then invaluable to have a baseline set of estimates, from pooled data, to begin the analysis of each individual. The alternative is to allow "black box" algorithms to find starting values, and run the risk of reporting local optima that are *a priori* implausible. We stress that these are complementary approaches, and that neither is necessarily superior.

In section 1 we describe the experimental data. The procedures are standard and deliberately simple: different subjects are asked to make choices between a certain amount of money today and different amounts of money in the future, and between different lotteries to allow us to identify their utility function. We identify the discount function by varying the horizon for the delayed reward across subjects. In section 2 we present the formal mixture model specification. In section 3 we present results, discuss alternative specifications, and valuable extensions.

1. Experimental Data

The basic question used to elicit individual discount rates is simple, and is of this general form: do you prefer \$100 today or \$100+x in the future, where x is some positive amount? If the subject prefers the \$100 today then one can infer that the discount rate is higher than x% over the period of the delay; otherwise, we can infer that it is x% or less over the period of the delay.[6] The format of our experiment modifies this question in three ways.

First, we pose a number of such questions to each individual, each question varying x by some amount. Table 1 illustrates the typical set of questions presented to each subject. When x is zero we would expect the individual to reject the option of waiting for a zero rate of return. As we increase x we would expect more individuals to take the future income option. For any given

[6] We assume that the subject does not have access to perfect capital markets, as explained in Coller and Williams [1999; p.110] and Harrison, Lau and Williams [2002; p.1607ff.]. This assumption is plausible, and confirmed by responses to the financial questionnaire that we ask each subject to complete.

individual, the point at which he switches from choosing the current income option to taking the future income option provides a bound on his discount rate. That is, if an individual takes the current income option for all *x* from 0 to 10, then takes the future income option for all *x* from 11 to 100, we can infer that his discount rate lies between 10% and 11% for this time interval. The finer the increments in *x*, the finer is the discount rate of the individual that we will be able to bracket. We then select one question at random for actual payment after all responses have been completed by the individual. In this way the results from one question do not generate income effects which might influence the answers to other questions.

Second, we consider many different horizons, ranging from 1 day to 60 days. We assign one of these time horizons to each subject at random, and elicit a discount rate response for that subject pertaining only to that horizon. We are thus able to plot discount rates over a variety of horizons and explicitly test how rates are related to the time horizon.

Third, we provide respondents with information on the implied interest rates associated with the delayed payment option. There is an important field counterpart to this information: truth in lending and savings laws require disclosure of both the annual and effective interest rates associated with credit market instruments. Hence, the provision of the implied rates serves to provide information in the laboratory that is consistent with the information available to individuals in naturally occurring credit markets.

Table 1 shows a payoff table for the case of a subject facing a 60 day horizon. The subject circled which option, A or B, he would prefer in each row. For subjects facing a different time horizon, the annual and effective rates were the same as those in Table 1, but the payment amounts for option B were appropriately different.

We recruited 87 subjects from the general undergraduate population of the University of South Carolina to participate in two sessions. The subjects were initially presented with the written instructions. These instructions were read out loud by an experimenter, who then ran the subjects through a "hands on" trainer in which they made choices over receiving Hershey's Kisses immediately or more Hershey's Kisses at the end of the experiment. After the trainer was completed, and all subjects had an opportunity to ask any questions, the main experiment was implemented. When all subjects had made their choices on record sheets similar to the one appearing in Table 1, the responses were collected and subjects completed a questionnaire eliciting socio-demographic information and information about their personal finances. One payoff alternative was selected at random, and then one subject was selected at random for payment. All random choices were implemented with a visible bingo cage. Each experiment lasted an average of 45 minutes.

We too particular care to equalize the "transactions costs" of selecting money sooner or later. Because our subjects are students, we avoided payouts that fall outside of the regular academic year during which the experiments were conducted, during recognized holidays, or on weekends. We also made sure that the instructions stressed the credibility of the payment. For instance, each subject was told that

> An experimenter will immediately take the Assignee to the office of XXX, a Notary Public, in the Moore School of Business. There the Assignee will receive a certificate which is redeemable under the conditions dictated by his or her chosen payment option under the selected payoff alternative. This certificate is guaranteed by Professor Elisabet Rutström, Beamlab Director, Moore School of Business, University of South Carolina. Professor Rutström's signature is notarized by XXX and thus the certificate is a binding legal contract between the Assignee and Professor Rutström. A sample certificate is shown below. At any time on or after the payment date the certificate can be redeemed for a University of South Carolina check in the office of XXX, Administrative Assistant, Department of Economics, room 412, Business Building. Alternatively, the assignee may mail the certificate to Professor Rutström at the address provided, and Professor Rutström will then mail a university check to the assignee.

The experiments were conducted on the same floor as the Department of Economics, and the bulk of the subjects had classes in the School of Business. No procedures are perfect, but we believe that these serve to make the task and future payment credible to subjects.[7]

[7] Andreoni and Sprenger [2009a] pay particular attention to these issues, and employ several procedural innovations that likely mitigate these "transactions cost" problems significantly.

2. Statistical Specification

Most of the debates over discounting have to do with the validity of the discounted utility model proposed by Samuelson [1937] and axiomatized by Koopmans [1960], Fishburn and Rubinstein [1982] and others. As explained by Andersen, Harrison, Lau and Rutström [2008], to correctly infer discount rates in this model one must jointly estimate utility functions and time preferences.[8] We first explain how one estimates the statistical model assuming exponential discounting, then we explain how that model is modified for quasi-hyperbolic discounting, and finally we allow for each type of data generation process in a mixture model.

General Statement

Consider the identification of risk and time preferences in the canonical case of mainstream economic theory. Specifically, assume EUT holds for the choices over risky alternatives, that subjects employ a CRRA utility function defined over the prizes they made choices over, and that discounting is exponential. Then a plausible axiomatic structure on preferences is known from Fishburn and Rubinstein [1982; Theorem 2] to imply the stationary structure

$$U(M_t) = 1/(1+\delta)^{\tau}\, U(M_{t+\tau}) \tag{1}$$

where $U(M_t)$ is the utility of monetary outcome M_t for delivery at time t, δ is the discount rate, τ is the horizon for later delivery of a monetary outcome, and U is a utility function for money that is stationary over time. Thus δ is the discount rate that makes the present value of the *utility* of the two monetary outcomes M_t and $M_{t+\tau}$ equal. Most analyses of discounting models implicitly assume that the individual is risk neutral, so that (1) is instead written in the more familiar form

$$M_t = 1/(1+\delta)^{\tau}\, M_{t+\tau} \tag{2}$$

in which δ is the discount rate that makes the present value of the two *monetary outcomes* M_t and $M_{t+\tau}$ equal.

[8] As a practical matter it is easiest to identify the shape of the utility function by evaluating the risk preferences of the subject, but it is the curvature of the utility function that is theoretically needed.

To state the obvious, (1) and (2) are not the same. This observation has an immediate implication for the identification of discount rates from observed choices over M_t and $M_{t+\tau}$. As one relaxes the assumption that the decision maker is risk neutral, it is apparent that the implied discount rate decreases since U(M) is concave in M. Thus one cannot infer the individual discount rate without knowing or assuming something about their risk attitudes.

Parametric Structure

We can quickly put some familiar parametric structure on this statement of the identification problem. Let the utility function be the CRRA specification

$$U(m) = m^{1-r}/(1-r) \tag{3}$$

where r is the CRRA coefficient and $r \neq 1$. With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. The experimental evidence is that the CRRA parameter is roughly 0.5 in laboratory experiments in the United States conducted by Holt and Laury [2002][2005] and Harrison, Johnson, McInnes and Rutström [2005]. The general conclusion from these studies is that *the utility of money function is concave* over the domain of prizes relevant for these experiments.

Of course, the other two parametric components of the specification include the assumption of EUT over risky lotteries, and the assumption of constant, exponential discounting.

Likelihood Specification

If one assumes that subjects made choices over uncertain lotteries, as well as over the timing of different prizes,[9] it is easy to write out the likelihood function and jointly estimate the parameter r of the utility function and the discount rate δ. Our subjects made choices over uncertain lotteries

[9] Experiments in which the same subjects were asked for both types of responses are reported in Harrison et al. [2005] and used in Andersen et al. [2008]. But we can also pool responses from distinct experimental samples drawn from the same population, as explained later.

using the “multiple price list” design of Holt and Laury [2002].[10]

Consider first the contribution to the overall likelihood from the risk aversion responses. Probabilities for each lottery outcome k_n, $p(k_n)$, are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery. If there were 2 outcomes in each lottery, as common in risk elicitation tasks of the type developed by Holt and Laury [2002], then the EU for lottery i is

$$EU_i = \sum_n [\ p(k_n) \times U(k_n)\ ] \tag{4}$$

for n = 1, 2.

A simple stochastic specification from Holt and Laury [2002] is used to specify likelihoods conditional on the model. The EU for each lottery pair is calculated for a candidate estimate of r, and the ratio

$$\nabla EU = EU_R^{1/\mu} \ / \ (EU_R^{1/\mu} + EU_L^{1/\mu}) \tag{5}$$

calculated, where EU_L is the left lottery in the display and EU_R is the right lottery, and μ is a structural “noise parameter” used to allow some errors from the perspective of the deterministic EUT model. The index ∇EU is in the form of a cumulative probability distribution function defined over differences in the EU of the two lotteries and the noise parameter μ.[11] Thus, as $\mu \rightarrow 0$, this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries; but as μ gets larger and larger the choice essentially becomes random. This is one of several different types of error story that could be used. The index in (5) is linked to the observed choices by specifying that the right lottery is chosen when $\nabla EU > 0.5$.

Thus the likelihood of the risk aversion responses, conditional on the EUT and CRRA

[10] This popular design asks subjects to make 10 binary choices between a “safe” lottery and a “risky” lottery, in which the probability of the higher outcome in each lottery is incremented by 0.1 in each row of the decision sheet. It is discussed in detail by Harrison and Rutström [2008], and compared to alternative elicitation methods.

[11] An alternative approach might be to define an index as the difference of the EU’s of the two lotteries, and then specify some cumulative distribution function to link it to the observed choices. For example, the cumulative standard normal distribution leads to the probit specification.

specifications being true, depend on the estimates of r and μ given the above statistical specification and the observed choices. The conditional log-likelihood is

$$\ln L^{RA}(r, \mu; y) = \sum_i [(\ln (\nabla EU)\times \mathbf{I}(y_i = 1)) + (\ln (\nabla EU)\times \mathbf{I}(y_i = 0)] \tag{6}$$

where $\mathbf{I}(\cdot)$ is the indicator function, and y_i =1(0) denotes the choice of the right (left) lottery in risk aversion task i.

Turning to the discount rate choices, a similar specification is employed. Equation (4) is replaced by the present value of the utility of the two outcomes, conditional on some assumed discount rate, and equation (5) is defined in terms of those present values instead of the expected utilities. The present value of the utility of M_t at t is just

$$PV_L = U(M_t) \tag{7}$$

and the present value of the utility of $M_{t+\tau}$ at $t+\tau$ is

$$PV_R = 1/(1+\delta)^\tau\, U(M_{t+\tau}) \tag{8}$$

where the subscripts L and R refer to the left and right options in the choice tasks presented to subjects, illustrated in Table 1. The parametric form for the utility function in (7) and (8) is the CRRA form given in (3), so we can rewrite these as

$$PV_L = M_t^{1-r}/(1-r) \tag{7$'$}$$

$$PV_R = [1/(1+\delta)^\tau] \times [M_{t+\tau}^{1-r}/(1-r)] \tag{8$'$}$$

An index of the difference between these present values, conditional on δ and r, can then be defined as

$$\nabla PV = PV_R^{1/\nu} / (PV_R^{1/\nu} + PV_L^{1/\nu}) \tag{9}$$

where ν is a noise parameter for the discount rate choices, just as μ was a noise parameter for the risk aversion choices. It is not obvious that $\mu = \nu$, since these are cognitively different tasks. Our own priors are that the risk aversion tasks are harder, since they involve four outcomes compared to two outcomes in the discount rate tasks, so we would expect $\mu > \nu$. Error structures are things one should always be agnostic about since they capture one's modeling ignorance, and we allow the error terms to differ between the risk and discount rate tasks.

Thus the likelihood of the discount rate responses, conditional on the EUT, CRRA and exponential discounting specifications being true, depend on the estimates of r, δ and ν given the above statistical specification and the observed choices. The conditional log-likelihood is

$$\ln L^{DR}(r, \delta, \nu; y, \tau) = \sum_i [(\ln(\nabla PV) \times \mathbf{I}(y_i = 1)) + (\ln(1-\nabla PV) \times \mathbf{I}(y_i = 0))]. \tag{10}$$

The joint likelihood of the risk aversion and *exponential* discount rate responses can then be written as

$$\ln L^{E}(r, \delta, \mu, \nu; y, \tau) = \ln L^{RA} + \ln L^{DR} \tag{11}$$

and maximized using standard numerical methods explained in Harrison and Rutström [2008]. The statistical specification allows for the possibility of correlation between responses by the same subject.[12]

We did not undertake any risk aversion tasks with our subjects, but Harrison, Johnson, McInnes and Rutström [2005] did undertake such experiments with subjects drawn from the same population. If we assume that the subjects from the two sets of experiments are comparable, we can pool them and use the risk aversion estimates from one sample to condition the implied discount rates of the other sample.

This useful feature of the "joint estimation" approach is worth noting explicitly. Assume for the moment that we have one representative agent in the sample. If the same subjects had given the responses to the risk aversion and discounting tasks, the formal maximum likelihood problem would solve for an r and δ (and μ) that jointly maximizes the probability of observing *all* of the choices over both types of tasks. In effect, and literally in terms of the matrix calculations involved, the risk aversion choices are "stacked" on top of the discounting choices, a likelihood for each choice evaluated conditional on the trial values of r and δ, and then a search made for the maximum

[12] The use of clustering to allow for "panel effects" from unobserved individual effects is common in the statistical survey literature. Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. The procedures for allowing for clustering allow heteroskedasticity between and within clusters, as well as autocorrelation within clusters. Wooldridge [2003] reviews some issues in the use of clustering for panel effects, in particular noting that significant inferential problems may arise with small numbers of panels.

likelihood values of r and δ. But there is nothing in the numerical setup that "tells" the likelihood function that the top stack of choices were made by the same people who made the bottom stack.[13] All that is needed for the numerical evaluation of the likelihood of observing the bottom, discounting stack of choices is that we know some value of r; in fact, that value of r could have been directly assumed *a priori* or from other studies.

Now extend this example to allow the parameter r to be a linear function of a binary dummy indicating the sex of the individual. In this case we would have one r for men, and one r for women, in effect. The r for men is used in the bottom stack to condition the calculation of the likelihood of the observed discounting choices made by men, and the r for women is used to condition the choices made by women. Similarly, when we have a richer set of demographics, the r for an individual is matched with the discounting choices of everyone else that shares exactly the same set of demographic characteristics.

It is then a small step to see how one can use responses from a distinct sample[14] to generate estimates of r that can be used to condition inferences about the discount rate δ for another sample. Obviously the conditioning is tighter if the samples are drawn from the same population, and it is ideal that they be the same individuals. But this is not necessary. In this manner, one can imagine "re-conditioning" inferences in experimental designs that did not anticipate the need for a control for non-linear utility, rather than having to run a completely new experiment in which the same subjects did all of the tasks needed by theory to infer discount rates.

3. Results

Results with Exponential Discounting

The results are reported in panel A of Table 2. The estimated CRRA coefficient is consistent

[13] There is a connection in standard errors when one allows for clustering, but that is not essential, and the joint estimation approach can allow for clustering by individual and simply recognize that the individuals in one stack are different from the individuals in the other stack.

[14] To see the logic, consider that it could have been the same sample just responding to experimental tasks in two different sessions.

with previous evidence from a wide range laboratory subjects, such as those reported in Holt and Laury [2002][2005] and Harrison, Johnson, McInnes and Rutström [2005]. These estimates, implying statistically significant concavity of the utility function, are sharply different from the individual-level estimates of risk neutrality reported by Benhabib, Bisin and Schotter [2009].

The estimated discount rate is 28.6%, which is high and comparable to the worst credit card interest rates. On the other hand, it is nowhere near the 472% reported from individual-level estimates by Benhabib, Bisin and Schotter [2009].

Quasi-Hyperbolic Discounting

The quasi-hyperbolic (QH) specification assumes that individuals have some "passion for the present" that leads to *any* future payment being discounted relative to current payments. The standard QH specification proposed by Laibson [1997] then assumes that individuals have constant discount rates for all future time delays. This means that we would replace

$$PV_R = 1/(1+\delta)^\tau \, U(M_{t+\tau}) \tag{8}$$

with

$$PV_R = \beta/(1+\delta)^\tau \, U(M_{t+\tau}) \tag{8*}$$

where $\beta < 1$ is the parameter capturing a "passion for the present." This implies that the subject has some preference for the M_t outcome in comparison to all of the $M_{t+\tau}$ outcomes, irrespective of $\tau>0$. The decision maker still discounts later payoffs for $\tau>0$, but in the standard manner. Thus, as $\tau \rightarrow \infty$ the discounting function looks more and more like the standard exponential model. In a similar manner to equation (11), the joint likelihood of the risk aversion and *quasi-hyperbolic* discount rate responses can be written as $\ln L^{QH}$ $(r, \beta, \delta, \mu, \nu; y, \tau)$.

Panel B of Table 2 contain estimates for the quasi-hyperbolic specification. The parameter β is estimated to be 0.987, and there is clear evidence for β being significantly less than 1. The estimated per-period discount rate δ is only 9.8% per annum.

Weighing the Evidence

To allow the evidence for each of the discounting models to be weighed, we use statistical mixture models in which the "grand likelihood" of the data is specified as a probability-weighted average of the likelihood of the data from each of K>1 models.[15] In other words, with the two modeling alternatives we have, which we can call models E and QH, the overall likelihood of a given observation is the probability of model E being true times the likelihood conditional on model E being true, plus the probability of model QH being true times the likelihood conditional on model QH being true.[16]

If we let π^E denote the probability that the exponential discounting model is correct, and $\pi^{QH} = (1-\pi^E)$ denote the probability that the quasi-hyperbolic model is correct, the grand likelihood can be written as the probability weighted average of the conditional likelihoods. Thus the likelihood for the overall model estimated is defined by

$$\ln L(\omega, r, \beta, \delta, \mu, \nu; y, \tau) = \textstyle\sum_i \ln [(\pi^E \times L_i^E) + (\pi^{QH} \times L_i^{QH})], \tag{13}$$

where L^E and L^{QH} were defined earlier, and ω is a parameter defining the log odds of the probabilities of each model as $\pi^E = 1/(1+\exp(\omega))$. This log-likelihood can be maximized to find estimates of the parameters.[17]

The estimates from this specification indicate that both models have roughly equal weight. The estimate of π^E is 0.59, with a standard error of 0.072 and a 95% confidence interval between

[15] Mixture models have an astonishing pedigree in statistics: Pearson [1894] examined data on the ratio of forehead to body length of 1000 crabs to illustrate "the dissection of abnormal frequency curves into normal curves,...". In modern parlance he was allowing the observed data to be generated by two distinct Gaussian processes, and estimated the two means and two standard deviations. Modern surveys of the evolution of mixture models are provided by Titterington, Smith and Makov [1985], Everitt [1996] and McLachlan and Peel [2000].

[16] The earliest applications of mixture models in the evaluation of experimental data are El-Gamal and Grether [1995] and Stahl and Wilson [1995]. More recent applications by Harrison and Rutström [2009] and Harrison, Humphrey and Verschoor [2009] consider mixtures of EUT and non-EUT models of choice under risk, and provide references to the larger literature. Andersen et al. [2008] consider mixtures of E and Hyperbolic discounting models when there is a positive FED.

[17] One could alternatively define a grand likelihood in which observations or subjects are completely classified as following one model or the other on the basis of the latent probabilities π^E and π^{QH}. El-Gamal and Grether [1995] illustrate this approach in the context of identifying behavioral strategies in Bayesian updating experiments.

0.45 and 0.74. The point estimates (standard errors) of r, β and δ in the mixture model are 0.51 (0.047), 0.94 (0.011) and 0.12 (0.027), respectively. We can easily reject the null hypothesis that β=1, with a *p*-value less than 0.001.

Thus the sample in this setting is *roughly split between those that use exponential discounting and those that use quasi-hyperbolic discounting*. We can formally test the hypothesis that the exponential model and quasi-hyperbolic model have equal support from these data, and the *p*-value for this two-sided test is 0.19. Thus the weight of the evidence is slightly in favor of the exponential model, but the two models are substantively entitled to roughly "equal billing."

The mixture model generates slightly different estimates for the core parameters of each discounting model. This makes sense. The earlier analysis, and results in Table 2, assumed that one model accounted for every data point, and estimates were found that did that conditional on that model being true *for every subject*. When we relax the conditional, we would expect the estimates to change since the exponential model does not need to explain data generated by quasi-hyperbolic subjects, and *vice versa*. In fact, we find that β is estimated to be 0.94 in the mixture model, which is much lower than before, signifying a greater "passion for the present." The parameter δ is estimated to be 11.6% for both the QH model and E models, with a standard error of 3% and a 95% confidence interval between 6% and 17%. These are sensible ranges for these parameters.

3. Limitations and Extensions

Our specification uses common parametric assumptions. Relaxing the specification of the utility function would not be expected to change the results dramatically, since the essential qualitative property is concavity of income and not the specific functional form. Similarly, moving from an EUT specification to a rank-dependent or prospect theory specification would likely not alter the main results, since one again typically finds concave utility functions in the gain domain employed here (e.g., see Andersen et al. [2008] and Harrison and Rutström [2008]).

On the other hand, there are some that argue that there could be an interaction between violations of EUT and departures from exponential discounting (e.g., Halevy [2008] and Andreoni and Sprenger [2009b]). For this reason it would be useful to incorporate mixture specification of alternative models of decision-making under risk, such as in Harrison and Rutström [2009], with mixture models of alternative models of discounting. In terms of the present modeling alternatives, this might imply four types in an extended mixture specification: EUT and exponential, EUT and quasi-hyperbolic, non-EUT and exponential, and non-EUT and quasi-hyperbolic. This type of structure would be feasible to evaluate with larger samples.

There are also alternative specifications of hyperbolic preferences, which might be worth examining. The most common functional form of the older literature is due to Herrnstein [1981], Ainslie [1992] and Mazur [1987], and would replace

$$PV_R = 1/(1+\delta)^\tau \, U(M_{t+\tau}) \tag{8}$$

with

$$PV_R = 1/(1+\alpha\tau) \, U(M_{t+\tau}) \tag{8'}$$

for $\alpha > 0$.[18] One could also evaluate the effect of using a more general hyperbolic specification[19] proposed by Prelec [2004; p.526]. This specification replaces (8) with

$$PV_R = \exp\{-\beta\tau^\alpha\} \, U(M_{t+\tau}) \tag{8''}$$

The exponential discounting model emerges as a limiting case as α tends to 1. One can think of the parameter α as characterizing the "decreasing impatience" of the decision-maker, a smoother and inverse counterpart of the notion of a "passion for the present" in quasi-hyperbolic characterizations. As α takes values below 1, the discount function takes on the familiar shape of

[18] An even more general version suggested by Ainslie [1992; p.71] [2001; p. 35] would add a parameter ϕ to influence valuation at time zero, so that $PV_R = 1/(\phi+\alpha\tau) \, U(M_{t+\tau})$. Values of $\phi>1$ would capture some of the discontinuous effects of the quasi-hyperbolic model as one considers any delay at all. Applied to our data, and assuming $\pi^E=0$, we estimate $\phi=1.013$ with a *p*-value of 0.003 the hypothesis test that $\phi=1$, and $\alpha=0.0955$ (*p*-value=0.088). The log-likelihood value of this hyperbolic specification is virtually the same as the quasi-hyperbolic specification. However, when we constrain $\phi=1$ the log-likelihood is significantly worse. This suggests that one definitely needs to accommodate the jump-discontinuity in value for *any* time delay in order to account for these data, but that there are various ways to model that phenomenon.

[19] Loewenstein and Prelec [1992] and Prelec [2004; p.515] proposed a similar generalized hyperbolic function which has fragile numerical properties when used for estimation.

earlier hyperbolic specifications. One can also think of the parameter β as characterizing time preferences in the usual sense (Prelec [2004; p.524]). The instantaneous discount rate implied by this discount function is $\alpha\beta t^{\alpha-1}$, which collapses to β as $\alpha \rightarrow 1$. In general one could extend the mixture specification to allow for exponential, quasi-hyperbolic or generalized hyperbolic data generating processes, since there is no clear *a priori* reason to prefer one hyperbolic specification over another when one is accounting for subjective time preferences.

Two particularly attractive variants on the Quasi-Hyperbolic model are worth considering, but would entail extensions of our experimental design to ensure econometric identification. One is the fixed cost model of Benhabib, Bisin and Schotter [2009]. Simplifying slightly, to see their main insight, we would replace

$$PV_R = \beta/(1+\delta)^{\tau}\ U(M_{t+\tau}) \tag{8*}$$

with

$$PV_R = \beta/(1+\delta)^{\tau}\ U(M_{t+\tau}) - \alpha\ M_t \tag{8$'$}$$

or with

$$PV_R = \beta/(1+\delta)^{\tau}\ U(M_{t+\tau}) - \alpha\ U(M_t) \tag{8$''$}$$

to capture the idea that there is some "fixed cost" α of foregoing money today. If $\alpha=0$ these two specifications collapse to the usual QH specification (8*), but if $\alpha>0$ then it is as if the individual needs some monetary premium or utility premium to forego money today.[20] This is quite plausible: if someone is staring at a \$100 principal, then they might reasonably behave as if no increment less than \$1, \$5 or \$10 would be enough to get them to consider delaying receipt. In order to reliably identify α it is apparent that one simply needs to have discounting tasks that vary the principal.[21] Our design did not have that feature, and that would be a valuable extension.

Yet another interesting variant on the mainstream QH specification, proposed by Jamison

[20] In their evaluation of the effects of allowing for concave utility, Benhabib, Bisin and Schotter [2009] implicitly assume that the premium is defined in terms of money. But it need not be, of course.

[21] Or one could impose the severe identifying assumption implied by (8′), that the individual discounts utility streams in terms of time preferences but has a fixed premium for delay defined in terms of money.

and Jamison [2007], is to allow for the discontinuity in the discount factor to be some finite time in the future, rather than *any* time in the future. Thus one would simply define the τ in

$$PV_R = \beta/(1+\delta)^\tau \, U(M_{t+\tau}) \tag{8*}$$

as being greater than some value $\tau_0 \ggg 0$. Of course, one has to estimate τ_0, and for that one would need an experimental design that varied the FED.[22] Our design did not have that feature, which would also be a valuable extension.

One could also relax the parametric assumption that the core parameters are estimated as scalars, and let them be linear functions of observable characteristics of individuals. Our sample is not sufficiently large for such exercises to be numerically reliable. However, such an extension would provide obvious insight into the characteristics of subjects better characterized by one model or the other.

4. Conclusion

We find that the data indicates roughly equal support for the exponential and quasi-hyperbolic models. Our experimental design and statistical model therefore reconcile the major competing theories, simply by recognizing that different subjects might behave according to different discounting models. Our results are conditional on parametric assumptions about functional forms, and we identify a range of immediate extensions that should be evaluated using our approach.

[22] An alternative would be to assume various values of τ_0 and evaluate the maximum likelihood values of remaining parameters conditional on that value. This approach defines what is known as a "profile likelihood" in econometrics. It is possible to then find the value of τ_0 that is associated with the highest conditional maximum likelihood, but one cannot view that value of τ_0 as an estimate.

Table 1: Illustrative Payoff Matrix

Payoff Alternative	**Payment Option A** (pays amount below today)	**Payment Option B** (pays amount below in 60 days)	**Annual Interest Rate** (AR)	**Annual Effective Interest Rate** (AER)	**Preferred Payment Option** (Circle A or B)
1	$500	$501.67	2.00%	2.02%	A B
2	$500	$502.51	3.00%	3.05%	A B
3	$500	$503.34	4.00%	4.08%	A B
4	$500	$504.18	5.00%	5.13%	A B
5	$500	$506.29	7.50%	7.79%	A B
6	$500	$508.40	10.00%	10.52%	A B
7	$500	$510.52	12.50%	13.31%	A B
8	$500	$512.65	15.00%	16.18%	A B
9	$500	$514.79	17.50%	19.12%	A B
10	$500	$516.94	20.00%	22.13%	A B
11	$500	$521.27	25.00%	28.39%	A B
12	$500	$530.02	35.00%	41.88%	A B
13	$500	$543.42	50.00%	64.81%	A B
14	$500	$566.50	75.00%	111.53%	A B
15	$500	$590.54	100.00%	171.45%	A B

Table 2: Initial Estimates of Time Preferences

Parameter	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
A. Assuming that the Exponential discounting model is the only data generating process					
r	0.523	0.050	0.000	0.424	0.622
δ	0.286	0.060	0.000	0.167	0.405
μ (for RA)	0.093	0.014	0.000	0.064	0.121
ν (for DR)	0.011	0.002	0.000	0.007	0.015
B. Assuming that the Quasi-Hyperbolic discounting model is the only data generating process					
r	0.523	0.050	0.000	0.424	0.622
β	0.987	0.004	0.000	0.979	0.995
			0.002 (H_0: β=1)		
δ	0.098	0.063	0.119	-0.026	0.222
μ (for RA)	0.093	0.014	0.000	0.064	0.121
ν (for DR)	0.011	0.002	0.000	0.007	0.015

References

Ainslie, George, *Picoeconomics* (New York: Cambridge University Press, 1992).

Ainslie, George, *Breakdown of Will* (New York: Cambridge University Press, 2001).

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Eliciting Risk and Time Preferences," *Econometrica*, 76(3), May 2008, 583-618.

Andreoni, James, and Sprenger, Charles, "Estimating Time Preferences from Convex Budgets," *Working Paper*, Department of Economics, University of California San Diego, November 2009a.

Andreoni, James, and Sprenger, Charles, "Risk Preferences are Not Time Preferences," *Working Paper*, Department of Economics, University of California San Diego, November 2009b.

Angeletos, George-Marios; Laibson, David; Repetto, Andrea; Tobacman, Jeremy, and Weinberg, Stephen, "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 15(3), Summer 2001, 47-68.

Benhabib, Jess; Bisin, Alberto, and Schotter, Andrew, "Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs," *Games & Economic Behavior*, forthcoming 2009.

Coller, Maribeth, and Williams, Melonie B., "Eliciting Individual Discount Rates," *Experimental Economics*, 2, 1999, 107-127.

El-Gamal, Mahmoud A., and Grether, David M., "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 432, December 1995, 1137-1145.

Everitt, Brian S., "An Introduction to Finite Mixture Distributions," *Statistical Methods in Medical Research*, 5, 1996, 107-127.

Fishburn, Peter C., and Rubinstein, Ariel, "Time Preference," *International Economic Review*, 23(5), October 1982, 677-694.

Frederick, Shane; Loewenstein, George, and O'Donoghue, Ted, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, XL, June 2002, 351-401.

Halevy, Yoram, "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect," *American Economic Review*, 2008, 98(3), 1145-1162.

Harrison, Glenn W.; Humphrey, Steven J., and Verschoor, Arjan, "Choice Under Uncertainty: Evidence from Ethiopia, India and Uganda," *The Economic Journal*, 119, 2009 forthcoming.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005, 897-901.

Harrison, Glenn W.; Lau, Morten Igel, and Williams, Melonie B., "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review*, 92(5), December 2002, 1606-1617.

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility *And* Prospect Theory: One Wedding and A Decent Funeral," *Experimental Economics*, 12(2), June 2009, 133-158.

Herrnstein, Richard J., "Self-Control as Response Strength," in C. M. Bradshaw, E. Szabadi and C.F. Lowe (eds.), *Quantification of Steady-State Operant Behavior* (Amsterdam: North-Holland, 1981).

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, 902-912.

Jamison, Dean T., and Jamison, Julian, "The Amount and Speed of Discounting," *Working Paper*, School of Public Health, University of Washington, November 2007

Koopmans, Tjalling C., "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 1960, 287-309.

Laibson, David I., "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62(2), 1997, 443-478.

Loewenstein, George, and Prelec, Drazen, "Anomalies in Intertemporal Choice: Evidence and Interpretation," *Quarterly Journal of Economics*, 107, 1992, 573-592.

Mazur, James E., "An Adjustment Procedure for Studying Delayed Reinforcement," in M.L. Commons, J.E. Mazur, J.A. Nevin and H. Rachlin (eds.), *The Effect of Delay and Intervening Events on Reinforcement Value* (Hillsdale, NJ: Erlbaum, 1987).

McLachlan, Geoffrey, and Peel, David, *Finite Mixture Models* (New York: Wiley, 2000).

O'Donoghue, Ted, and Rabin, Matthew, "Doing It Now or Later," *American Economic Review*, 89(1), 1999, 103-124.

Pearson, Karl, "Contribution to the mathematical theory of evolution," *Philosophical Transactions A*, 185, 1894, 71-110.

Phelps, Edmund S., and Pollak, Robert A., "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 1968, 185-199.

Prelec, Drazen, "Decreasing Impatience: A Criterion for Non-Stationary Time Preference and 'Hyperbolic' Discounting," *Scandinavian Journal of Economics*, 106(3), 2004, 511-532.

Samuelson, Paul, "A Note on Measurement of Utility," *Review of Economic Studies*, 4, 1937, 155-161.

Stahl, Dale O., and Wilson, Paul W., "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 1995, 218-254.

Titterington, D.M., Smith, A.F.M., and Makov, U.E., *Statistical Analysis of Finite Mixture Distributions* (New York: Wiley, 1985).

Wooldridge, Jeffrey, "Cluster-Sample Methods in Applied Econometrics," *American Economic Review (Papers & Proceedings)*, 93, May 1993, 133-138.