Behavior Research Methods, Instruments, & Computers
2004, 36 (1), 41-51

Using logistic regression to estimate delay-discounting functions

E. PAUL WILEYTO and JANET AUDRAIN-McGOVERN
University of Pennsylvania, Philadelphia, Pennsylvania

LEONARD H. EPSTEIN
State University of New York, Buffalo, New York

and

CARYN LERMAN
University of Pennsylvania, Philadelphia, Pennsylvania

The monetary choice questionnaire (MCQ) and similar computer tasks ask preference questions in order to ascertain indifference, the perceived equivalence of immediate versus larger delayed rewards. Indifference data are then fitted with a hyperbolic function, summarizing the decline in perceived value with delay time. We present a fitting method that estimates the hyperbolic parameter k directly from survey responses. Binary preferences are modeled as a function of time (X_2) and a transformed reward ratio (X_1), yielding logistic regression coefficients β_2 and β_1. The hyperbolic parameter emerges as $\hat{k} = \beta_2/\beta_1$, where the logistic predicted $p = .5$ (the definition of indifference). The MCQ was administered to 1,073 adolescents and was scored using both standard and logistic methods. The means for $\ln(\hat{k})$ were similar (standard, -4.53; logistic, -4.51), and the results were highly correlated ($\rho = .973$). Simulated MCQ data showed that $\hat{k}$ was unbiased, except where $\beta_1 \geq -1$, indicating a vague survey response. Jackknife standard errors provided excellent coverage.

Delay discounting describes an individual's preference for a smaller immediate reward over a delayed but larger reward because the later reward is perceived as less valuable, due to the passage of time (Bickel, Madden, & Petry, 1998). Delay-discounting measures have been used as indices of impulsivity, on the basis of the premise that impulsive individuals prefer more immediate gratification (Crean, de Wit, & Richards, 2000). A higher level of discounting has been shown to be associated with a greater liability to substance use and abuse (Bickel, Odum, & Madden, 1999; Madden, Petry, Badger, & Bickel, 1997; Petry, 2001a, 2001b).

Traditional assessment of discounting involves fitting summary models to data concerning indifference with respect to the choice between immediate and delayed rewards (Crean et al., 2000). Indifference experiments offer choices of immediate rewards (value V_I) versus delayed rewards (value V_D) and ascertain the indifference point at which V_I is perceived to be equivalent to V_D (indifference is denoted V_I*, and the reward ratio $R = V_I/V_D$ is denoted R* at the indifference point). These preference questions are of the form, "Would you prefer to receive \$10 today or \$25 in 7 days?" (Here, $V_D = 25$, and $V_I = 10$.) A series of these indifference points is evaluated across a range of delay times, and nonlinear regression is used to fit a function describing the decline in perceived value as a function of time.

The summary model used most commonly is a hyperbolic decline in value with increased delay time (Johnson & Bickel, 2002; Kirby, 1997; Kirby & Maraković, 1996; Kirby, Petry, & Bickel, 1999; Madden, Bickel, & Jacobs, 1999), which has the form

$$R^* = \frac{V_I^*}{V_D} = \frac{1}{1+kt}. \quad (1)$$

This allows us to summarize a subject's tendency to discount as a single value, k, the scale parameter for the hyperbolic function of delay time. The possible range of k is 0 to ∞, but the practical range falls between 0 and 0.5. Small values of k (10^{-4}, 10^{-3}, . . .) indicate a lack of discounting and a general preference for delayed rewards, whereas larger values (0.1+) indicate heavy discounting and a preference for immediate rewards. The response does seem to vary with the magnitude of V_D (Green, Myerson, & McFadden, 1997; Johnson & Bickel, 2002; Kirby, 1997; Kirby & Maraković, 1995), so separate k values are often assessed for small, medium, and large V_D.

This work was supported by Transdisciplinary Tobacco Use Research Center Grant P50 CA/DA84718 to C.L. The authors extend their thanks to K. P. Burnham and J. L. Norris for discussions on assessing confidence intervals and to K. Kirby, W. Bilker, and three anonymous reviewers for reviews of an earlier version of this article. Correspondence concerning this article should be sent to E. P. Wileyto, Tobacco Use Research Center, University of Pennsylvania, 3535 Market Street, Suite 4100, Philadelphia, PA 19104 (e-mail: epw@mail.med.upenn.edu).

 Copyright 2004 Psychonomic Society, Inc.

Although fitting a model to indifference points is straightforward, determining the indifference points is cumbersome. To keep the number of questions down to a reasonable level, computer task assessments given in the laboratory adjust V_I depending on prior answers (Crean et al., 2000; Richards, Mitchell, de Wit, & Seiden, 1997), yet even this quasi-optimal sampling still leaves us with an indeterminate number of questions, usually well over 100 (Lerman, unpublished data;[1] Richards et al., 1997).

Moving from the laboratory venue to more routine assessment requires a shorter uniform series of questions. Kirby et al. (1999) introduced the 27-item monetary choice questionnaire (MCQ) as such an alternative to the computer task (requiring only nine questions devoted to each of small, medium, and large delayed rewards). The survey is scored by inspection, assigning a k value for each block of nine questions by looking for a single indifference point at which the subject's behavior switched from preferring the delayed reward to preferring the immediate reward. Each block is assigned the k value most consistent with the array of realized choices and the geometric mean of possible values, if more than one k value turned out to be most consistent with the data block. Kirby's (2000) method reduces the scoring of the MCQ to entering survey data into a spreadsheet and matching behavioral breakpoints in a discrete lookup table.

All of the approaches described thus far depend on treating indifference points as the random variables of interest, an approach that is ideal for determining what type of function to use in summarizing discounting, in addition to estimating the parameters. Although Kirby's (2000) method is crude, it does highlight the fact that it takes fewer data to estimate the parameters when it has been decided ahead of time that the function is hyperbolic than it does to determine what type of function fits best. This focus on assessing indifference also does not allow us to simulate survey data; thus, the approach cannot provide the usual measures of bias that accompany standard estimation procedures. In addition, the standard MCQ scoring procedure does not yield any measures of uncertainty.

The purpose of this article is threefold. First, we present a new approach to assessing discounting and estimating parameters with logistic regression, to provide a direct estimate of the hyperbolic parameter k and obtain confidence intervals, using the individual choices made on each preference question as the random variables of interest. Second, we compare the new approach with standard scoring on MCQ data, to determine how well the new measure corresponds to the standard. Finally, we use the logistic model to simulate MCQ survey data and examine the performance of the estimation procedure against known values of k. The procedure we develop here has the potential to streamline analysis of MCQ data while giving a better account for people who do not have a perfectly defined indifference point, as well as to reduce the amount of data needed for the computer task assessment of discounting, all using standard statistical software.

A Logistic Probability Model of Discounting

We examined only a simple model that was independent of V_D. Logistic regression (Hosmer & Lemeshow, 1989) uses a transformation of a linear score variable ($\Sigma\beta X$) to predict a probability of a binary event:

$$p = \frac{e^{\Sigma\beta X}}{1 + e^{\Sigma\beta X}}. \tag{2}$$

Using logistic regression (with no intercept), we fit the choice responses (0 for immediate, 1 for delayed), with a logistic model using time (t) and a transformed ratio variable ($1 - 1/R$) as predictors, to give us

$$\Sigma\beta X = \beta_1\left(1 - \frac{1}{R}\right) + \beta_2 t. \tag{3}$$

We were interested in the value of the reward ratio ($R = R^*$) that sets the probability of taking the delayed reward to .5. From Equation 2, $p = .5$ where $\Sigma\beta X = 0$, making R^*

$$R^* = \frac{1}{1 + \frac{\beta_2}{\beta_1} t}, \tag{4}$$

and our estimate of k is

$$\hat{k} = \frac{\beta_2}{\beta_1}.$$

The approach is similar philosophically to probit analysis, except that we are using a logistic model of an event probability, rather than a cumulative normal distribution. Standard errors are readily estimated for $\hat{k}$ and $\ln(\hat{k})$, using the delta method, after obtaining the covariance matrix($\mathbf{V}$) for β (Hogg & Craig, 1995). We focused on $\ln(\hat{k})$ to prevent the lower confidence limit around $\hat{k}$ from dropping below 0. The variance for our estimate $\ln(\hat{k})$ is then $\mathrm{Var}[\ln(\hat{k})] = \mathbf{g}\mathbf{V}\mathbf{g}^T$, where $\mathbf{g} = \partial\ln(\hat{k})/\partial\beta = [-1/\beta_1, 1/\beta_2]$, and the standard error is $SE_\Delta = \sqrt{\mathrm{Var}[\ln(\hat{k})]}$. Standard errors for $\ln(\hat{k})$ may also be obtained using the jackknife procedure (SE_{JK}), by systematic deletion of one survey question (Efron & Tibshirani, 1993). (See the appendices for the code for estimating k, SE, and SE_{JK}.)

METHOD

Our purpose was to examine the performance of a new estimation procedure for rapidly fitting the hyperbolic delay-discounting function to smaller quantities of uniform survey data. The data came from two sources. First, we utilized data that were collected as part of a longitudinal study of health-related behavior among high school students in the Washington, D.C. metropolitan area. The MCQ was administered in the spring of the sophomore year, and 1,073 students completed the survey. Estimates of k were generated using both the standard procedure

and logistic regression. Finally, the estimates from the two methods were compared using correlation methods.

Although real data allow us to compare the output of the new technique with a standard approach, real data cannot tell us how well the procedure works in an absolute way. To address that question, we simulated random MCQ data, using Monte Carlo methods based on Equations 2 and 3, testing the procedure's ability to identify the original parameters used in generating the simulated data. We generated 1,200 surveys for each combination of parameters, where $k = [0.002\ 0.01\ 0.02\ 0.1\ 0.25]$ and $\beta_1 = [-0.5\ -1\ -3\ -5\ -8]$, for a total of 30,000 simulated surveys. We assumed that k was uniform across all V_D levels and estimated only a single value. We applied the logistic regression procedure to each survey and calculated an upper and a lower confidence limit for $\ln(\hat{k})$, using SE_Δ and SE_{JK}. We also recorded all related events, such as success or failure to converge on an estimate and inclusion of the true value of $\ln(k)$ inside each of the confidence intervals (CIs). Finally, we calculated summary statistics on our output data, including the bias for $\hat{k}$ and $\ln(\hat{k})$, the average length of the CI, the probability of successful convergence, and the coverage probability for each of the CIs.

Software

We used two software packages. First, we wrote a routine for standard scoring of the MCQ survey, using the Matlab programming language (Mathworks, Natick, MA). The routine duplicated the logic in a spreadsheet that had been written for scoring the survey (Kirby, 2000), but we included loop statements for filing through a large data set and recording output values.

We used Stata (Stata Corporation, College Station, TX) for estimation of the k parameter, including the logit and the generalized linear models (GLM) procedures (see Appendix A). We implemented the delta method for estimating CIs around k, utilizing the covariance matrix available from the system variables. We also used the jackknife procedure to obtain an alternate SE and CI.

Stata also provides a flexible facility for maximum-likelihood (ML) estimation, allowing us to tailor likelihood functions that fit our needs. We used the Stata ML to place boundaries on parameter values, using a penalty function that came into play any time parameters stepped beyond the boundaries during optimization (see Appendix B).

RESULTS AND DISCUSSION

Whole Population Analysis

The standard scoring procedure was applied to the student survey data, yielding a full 1,073 k estimates for each of the three V_D levels. We averaged the $\ln(\hat{k})$ values across the V_D levels, to yield a mean $\ln(\hat{k})$ of -4.53 ($SE = 0.046$; CI, -4.623 to -4.443). This is approximately halfway across the range of the standard assessment (-8.740336 to -1.39938). An analysis of the whole population with the Stata logit procedure yielded $\beta_1 = 1.67$ and $\beta_2 = -0.0184$, for a population-averaged $\hat{k}$ of 0.011, or $\ln(\hat{k})$ of -4.51. Using the delta method, we obtained the variance of $\ln(\hat{k})$, $\mathrm{VAR}_{\ln(\hat{k})} = 0.00173$, for an SE of 0.042, yielding a 95% CI of -4.593 to -4.427. Overall, the whole population measures of $\ln(\hat{k})$ appear very similar.

Comparing Logistic With Standard Scoring

We used the Stata GLM procedure to score the student MCQ survey data, this time analyzing by subject. (Implementation of logistic regression is different for logit and GLM, and GLM works better on the small samples of the MCQ.) The GLM procedure determined estimates for 986 subjects, with a mean $\ln(k)$ of -4.51 and a range of -10.58 to -0.95. (The range for standard scoring is artificially restricted to -8.74 to -1.40. For extreme questionnaires, almost all one choice, the procedure can project an estimate outside of this range.) A number of k estimates (87) were out of bounds (negative) or failed to converge. For $\hat{k}$ to be positive, both β_1 and β_2 must be negative.

Logistic-regression based estimates are plotted against their corresponding standard estimates in Figure 1A. Most of the points lie along the 45° line. Some small exceptions occur at the extremes, where our point estimate is free to reach more extreme values than are allowed under Kirby's (2000) scoring, giving the relationship a slightly "sigmoid" appearance. The overall correlation between the measures is .944. However, if we restrict the analysis to those with a $\beta_1 < -1$, the correlation becomes .972. As we shall discuss later, $\beta_1 < -1$ indicates a better defined indifference response, and as is indicated by the hollow points in Figure 1A, most of the outliers were $\beta_1 \geq -1$.

It is useful now to discuss the (optimization) pathology that prevented 87 surveys (8% of the total) from being scored by Stata GLM. To examine whether the subjects were responding sensibly, we ordered the questions by ratio and recorded a regression slope (choice: delayed = 1 vs. order). We expected a negative trend across the order. All of the 986 surveys fitted by Stata showed a negative trend. Of the 87 not fitted, 13 showed a slight positive trend. This implies that the subject's choices were changing in the "wrong" direction, so that they were choosing the immediate reward more often when it was small and rejecting it more often when it was closer in value to the delayed reward. (This is like choosing the delayed reward when asked, "Would you like $100 right now or in 30 days?) We cannot think of any explanation for this, other than the fact that answering questions at random, in the sense of a coin toss, will occasionally lead to a positive trend. Another 48 of the surveys that failed to converge showed a zero trend, having been answered uniformly all the way across. Kirby's (2000) own method of scoring deals with those 48 uniform surveys in a sensible way, assigning them the high or the low extreme values from his available range. We cannot obtain a point estimate for these subjects without placing bounds on the parameter values, yielding an "at least" or "at most" k value. The remaining 26 showed a slight or vague neg-

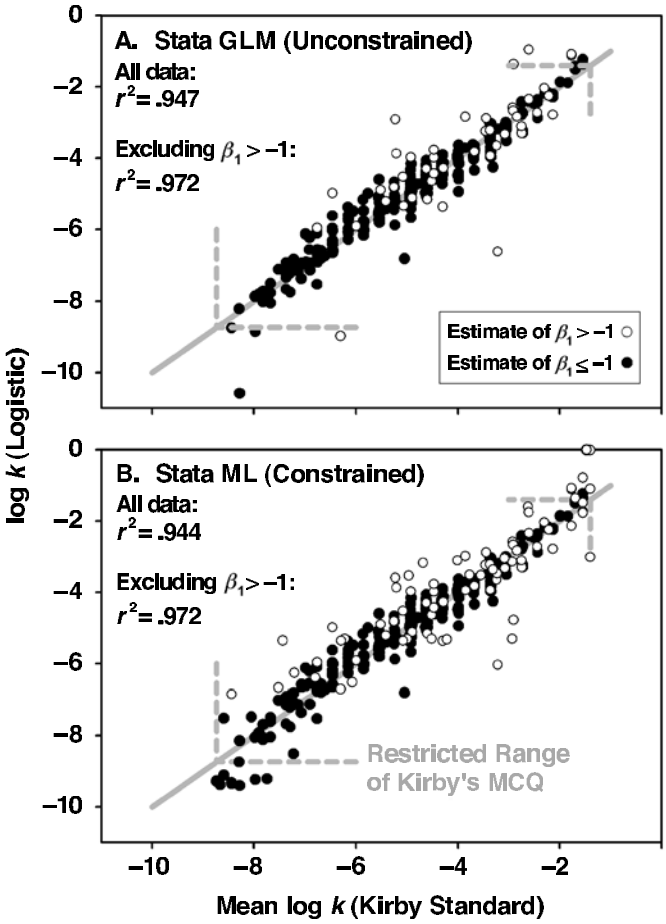

Figure 1. (A) Results from the Stata generalized linear models (GLM) procedure, comparing logistic and standard scoring. The *x*-axis represents the average of the three ln(*k*) values obtained through standard scoring, and the *y*-axis represents the result of the logistic procedure. Open circles indicate an estimate of $\beta_1 > -1$, indicating a vague or less consistent survey response on the part of the subject. Results lie generally along the 45° line (gray), indicating a one-to-one correspondence. The correlation is .947 if the open circles are included, and .972 if the open circles are discarded. The GLM procedure scored 986 of 1,073 surveys. (B) Similar results using Stata maximum likelihood (ML), with bounds placed on the parameters (β_1, β_2, $\hat{k}$), using a penalty function. Constraints allow all 1,073 surveys to be scored, and uniform surveys (all 0 or all 1) are correctly placed at the extremes. The correlation is .944 if the open circles are included and .972 if the open circles are discarded. Standard scoring for the survey produces results for log(*k*) between −8.74 and −1.40, and this restricted range is indicated by gray dashed lines in both parts of the figure. Our logistic scoring falls outside of those lines when a set of extreme survey results causes the model to project *p* = .5 past the standard range. This happened more often for the constrained procedure, since the constraints made it possible to fit all of the surveys, giving (B) a sigmoid appearance. Our constraints were wider than standard scoring, because our purpose in constraining the estimate was to keep *k* in the legal bounds for the hyperbolic function, rather than to restrict it to range of the monetary choice questionnaire (MCQ).

ative trend. When the actual patterns displayed by those last 26 surveys were examined, the subjects had made either random (coin toss) choices all the way across the order or uniform choices, with one or two exceptions.

Kirby's (2000) own method determines a consistency score for each block of nine questions, and he rejects observations with consistency scores less than 0.77. We observed that over 2% of blocks have low consistency, and

3.7% (39) are inconsistent in at least one block. Excluding the uniform responses, our nonidentifiable rate is similar to his, and the very surveys that are difficult to evaluate using the standard method create optimization problems for the logistic method.

We repeated the logistic regression procedure with Stata ML, applying penalty function bounds to the parameters, in order to determine whether a constrained optimization routine would reduce our failure rate. Bounds were applied so that $-10 < \beta_1 < -0.2$, $\beta_2 < -0.00001$, and $0.00002 < \hat{k} < 0.5$. The ML logistic procedure generated a $\hat{k}$ for all surveys (Figure 1B). Most of the distribution looks the same. The exceptions are the observations with β_1 estimated to be > -1 and the addition of those at the extreme ends of the distribution. The bounded procedure correctly assigned the 48 uniform responses the most extreme values it could, subject to the constraints, and matched the behavior of the procedure outlined by Kirby (2000).

What β_1 and β_2 Mean in the Context of the MCQ

The MCQ was designed perfectly to estimate k with a small sample of questions. The layout of data points in the MCQ is shown in Figure 2A. The scatter points at the lower left correspond to an immediate reward that is very small versus a comparatively large reward after only a small delay. At the upper right extent of the scatter, the reward values are almost equal, but one must wait a very long time for the slightly larger one. Somewhere across the range of scatter points, we would expect to see a shift in behavior between choosing the delayed reward and choosing the immediate reward. Kirby et al. (1999) set up the array of questions so that the placement of the best curve would be apparent. This arrangement works for the logistic method, because the no-intercept (a forced 0 intercept) logistic model forces the hyperbolic y-intercept through 1. Like Kirby, we identify the hyperbolic function from two points: the y-intercept and the point at which it crosses the picket line.

On reviewing a large number of MCQ results, we found that some individuals showed very clean shifts in preference from delayed to immediate rewards, so that when the ratio was smaller than the indifference point, the delayed reward was always chosen, and when the ratio was higher, the immediate reward was always chosen. Others appeared to be "answering at random" or, more properly, tossing a hypothetical coin for each survey item in the mid-range, sometimes showing a clean response at the extremes, sometimes not. There were many surveys in between these extremes.

The parameter β_1 can be thought of as a vagueness parameter, related to the clarity of the indifference response. It represents the slope of the logistic probability for choosing delayed as the reward ratio increases as the probability crosses .5 (at R^*). Figure 2B shows a sample indifference model, indicated by the bold line with y-intercept of 1. Above and below the bold line we see contours for $p = .75$ and $p = .25$ that arise from two different values of β_1. The closer those contours are to the bold line, the steeper the probability gradient across the indifference level of .5. (The profile view is in Figure 2C.) The possible range of β_1 is $(-\infty < \beta_1 < 0)$, although a more realistic range is $(-20 < \beta_1 < -0.1)$. Large negative values of β_1 ($\beta_1 < -5$) are associated with clear, steep behavior changes in crossing the indifference curve (Figure 2C), and individuals who show clean breaks in behavior (the kind that one expects to see with standard scoring) are represented by those values. Small negative values are associated with vague, flat changes (Figures 2B and 2C), and individuals who appear to be "answering at random" are represented by those small negative values. The β_1 parameter allows us to simulate individual variation in survey behavior and to examine the impact of a vague or clean response on our ability to identify the model from the data.

The parameter β_2 is a time coefficient, or a scale parameter. It can be thought of as scaling the k value to match the units on the x-axis in Figure 2A or 2B, which are days but could as easily be hours, weeks, or months.

Performance of the Procedure

Thus far, we have demonstrated that our procedure based on logistic regression can match the estimates made with Kirby's own method on his discounting survey. Here, we present an examination of the performance of the procedure, using simulated survey data generated from known parameter values. As was expected, not all simulated surveys could be scored. Figure 3A reveals the probability of obtaining an estimate based on the logistic method. The method failed frequently where the true value of β_1 was -0.5. Our earlier discussion on the meaning of β_1 makes sense of this, because this is where the location of the $p = .5$ contour would be the most vague. Failures also occurred often where the true β_1 was -1, and the true k was 0.002 (small). The gradient is still vague where β_1 is -1, and the indifference curve is very close to the upper edge of the range of the data shown in Figure 2A. A smaller number of failures occurred at the other end of the k range (true $k = 0.25$).

The meaningful values of $\hat{k}$ span four orders of magnitude. Thus, we calculated bias $[E(\hat{\theta}) - \theta]$ for $\log10(\hat{k})$, using successful estimates, so that bias would be presented in orders of magnitude (Figure 3B). $\hat{k}$ appears to have been relatively unbiased, except for a positive bias where k was small (true $k = 0.002$) and β_1 was a small negative number (-0.5, -1). The bias in those cases may have been a direct result of the values omitted due to estimation failures; the positive bias reflects the surveys that are, by chance, more like the middle of the k range and easier to estimate. Bias was reduced (Figure 3B) using the constrained routine in Stata ML. The most likely explanation is that the estimates from the difficult data sets were included, rather than excluded, and were important in centering the distribution.

The delta method 95% CI did not perform well over much of the parameter range tested (Figure 4A). The probability of including the true value of k (coverage probabil-

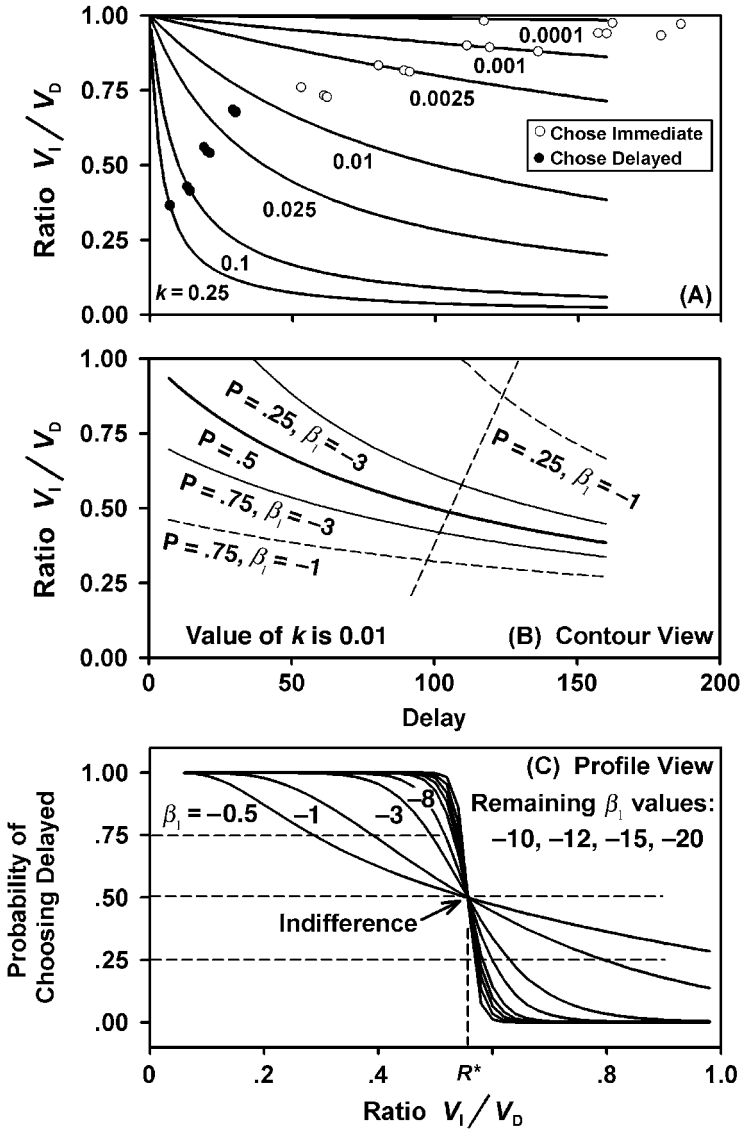

Figure 2. (A) How standard scoring of the monetary choice questionnaire (MCQ) identifies *k*. Points indicate the reward ratio values in the survey versus delay times, and they correspond to hyperbolic functions that have specific values of *k*. The estimate of *k* is identified by looking for a shift in behavior. At the lower left, the immediate reward is worth very little, as compared with the delayed reward, and subjects choose the delayed reward almost exclusively, whereas at the upper right, the immediate reward is almost as valuable as the delayed reward, and they choose the immediate reward almost exclusively. The shift in behavior indicated here corresponds to a *k* of approximately 0.01. Standard scoring has a more difficult time with responses that do not break cleanly, which are quite likely because the choices are given out of order and respondents "shoot from the hip." (B) Our approach does not require a clean break, because we model the probability of choosing the delayed reward and identify where the probability is .5. The contours reflect a gradual change in the probability that the participant will choose delayed rewards, and the slope of this gradual shift in behavior is reflected by parameter β_1. If the change in behavior is sharper, β_1 is a large negative number. If the change is more gradual, β_1 is a smaller negative number. The bold line represents a probability contour of $p = .5$ for choosing the delayed reward. Additional lines show the .75 and .25 probability contours under two different values of β_1. (C) Here, we display the same model of a gradual shift in behavior in profile, indicating elevations along the dashed line in contour view. It shows gradients crossing the $p = .5$ contour for various values of β_1. Large negative values may not be separated easily with small samples. Stata's generalized linear models procedure seems to work better than the logit procedure for these extreme values.

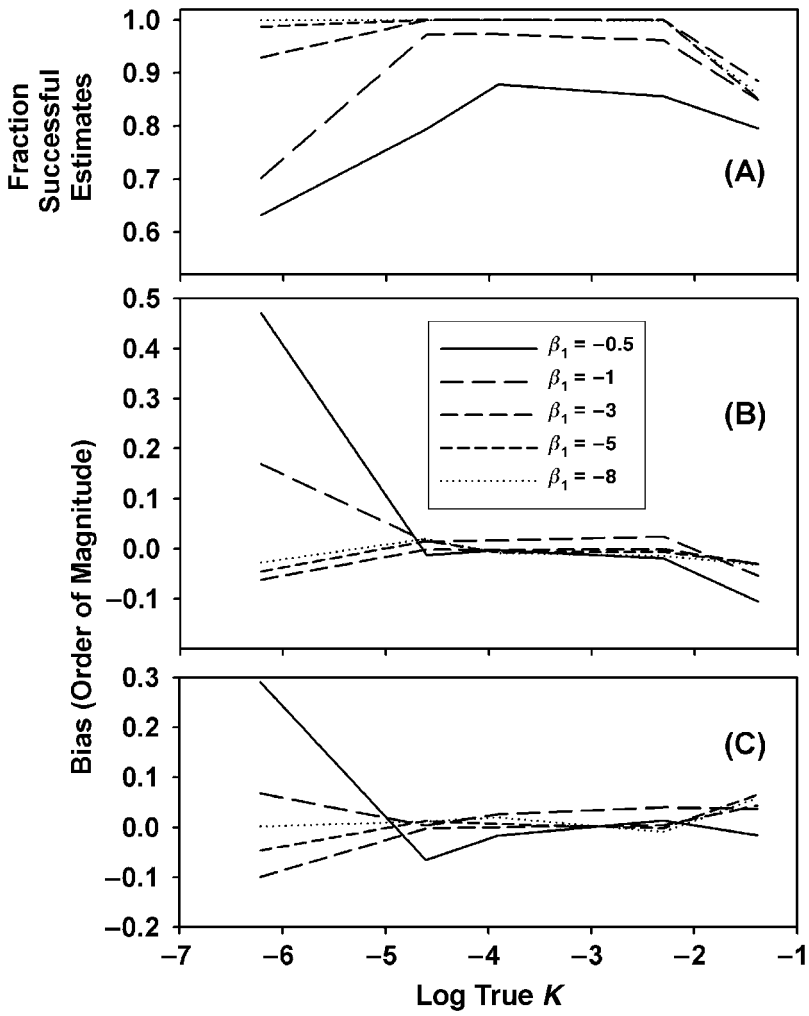

Figure 3. Simulation results. The x-axis shows the true value of $\ln(k)$, with separate lines showing the response for different values of β_1. (A) Proportion of simulated surveys that could be scored using Stata generalized linear models. Two things contributed to failures: k_{true} values at either extreme were near the edge of the survey's ability to detect them, and small negative numbers for β_1 (solid line) produced a vague response. (B) Bias for a parameter θ is $[E(\hat{\theta}) - \theta]$. A bias of 0 is perfect. Because the range of k can be several orders of magnitude, we calculated bias as $\{E[\log_{10}(\hat{k})] - \log_{10}(k)\}$. It was negligible except where β_1 was a small negative number and the probability of deleting one of the more extreme observations was high. (C) Bias for the bounded estimate (from Stata maximum likelihood) was reduced because extreme observations were included in the average.

ity) was closest to .95 for $\beta_1 = -0.5$ and, generally, .4 .5 for lower values of β_1. This may be related to the small sample obtained by a single survey and the excessively small standard errors generated by the delta method under some parameter combinations. However, the jackknife *SE* (SE_{JK}), used in conjunction with the GLM procedure, produced CIs with consistently good coverage (Figure 4B), generally falling in the 89% 97% range. We also applied the jackknife procedure to the bounded ML procedure. The initial assessment of coverage included values as low as 70%, higher than the worst given by the delta method but not close to the coverage provided when the jackknife procedure was applied to Stata GLM. This was remedied by removing from consideration CIs calculated for point estimates made with constraints active, which left the remaining coverage close to 95% across the board (Figure 4C).

SE_{JK} can be exceedingly large, and the distribution of SE_{JK} can be highly skewed. For that reason, Figure 4D displays the median length of the 95% CI resulting from SE_{JK}, revealing two trends. First, the median size of the CI is about the same magnitude as true k, and correspondingly, the CI generally increases as true k increases. The

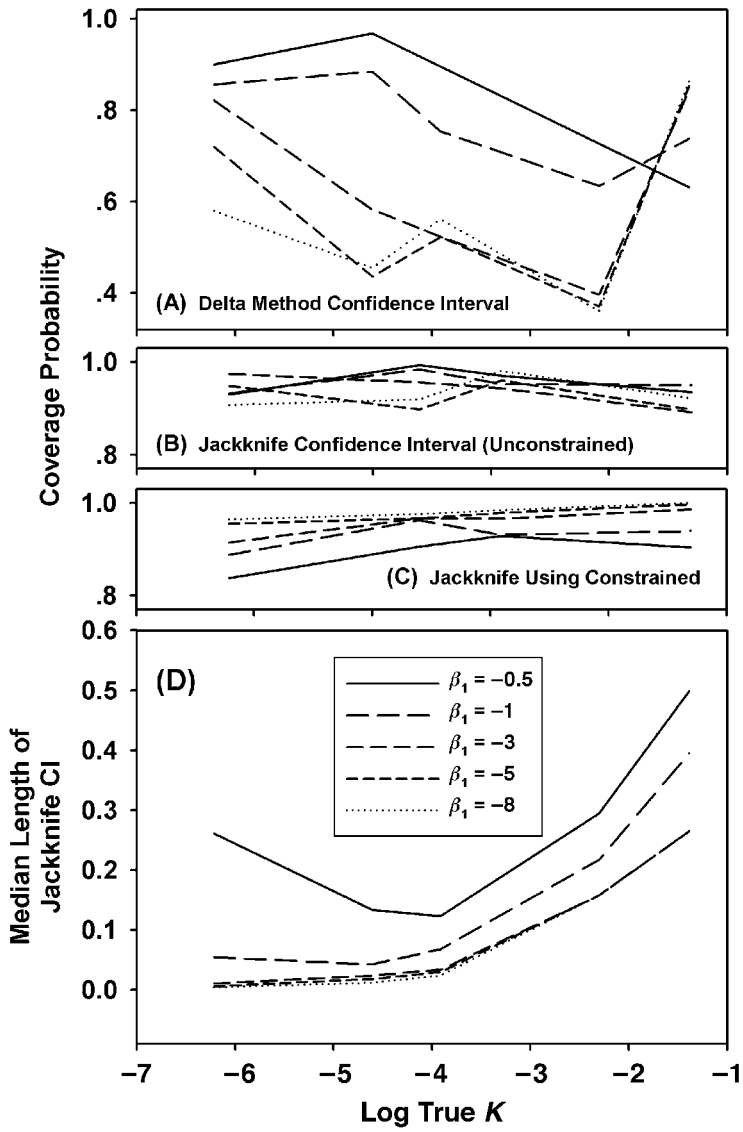

Figure 4. Performance of the confidence interval. The *x*-axis shows the true value of ln(*k*), with separate lines showing the response for different values of β_1. (A) Coverage of the confidence interval derived via the delta method falls far short of the purported 95%. The worst performance comes from what should be the best data (clean behavioral shifts and k_{true} in the mid-range). (B) Coverage of the confidence interval derived via the jackknife remains close to 95% across all parameter values tested. We recommend the jackknife procedure. (C) Coverage of the confidence interval derived via the jackknife when bounds are applied to the parameters remains close to 95% across all parameter values tested, as long as standard errors are discarded when the constraint is active. Coverage reached as low as 70% when observations with active constraints were included. We recommend ignoring the standard error estimate when any constraint is active for the point estimate. (D) The median width of jackknife confidence intervals for the constrained procedure (corresponding to that in panel C). Intervals are widest for large values of k_{true}, and for a small negative β_1.

second trend is that high values of β_1 tend to have larger CIs than do low values of β_1. The high values of β_1 are more vague responses to the series of survey questions (Figure 2A). One exception, where the CI is unusually large, occurs where β_1 is -0.5 and k is 0.002. This combination implies a vague response on the part of the subject, and a true k value near the edge of the survey's ability to detect.

Limitations of Our Analysis

Although our method of analysis can be applied to computer task data, as well as to any fixed format survey designed to assess discounting, we restricted our analysis to data that would be found in the MCQ (Kirby et al., 1999), because it is in current use and amenable to simulation. Our restriction to MCQ data was also appropriate because this survey is growing in popularity; it is the only standard survey for assessing delay discounting and is being placed in a large number of current protocols because it is compact (Audrain-McGovern et al., 2004; K. N. Kirby, personal communication; N. M. Petry, personal communication; M. Seligman, personal communication). What we report is probably the most extreme test of the fitting technique, and better results would be expected with larger formats.

For simplicity in this initial analysis, we treated discounting as if it were uniform, regardless of the magnitude of the delayed reward. Nevertheless, we know that k is smaller if the delayed reward is more valuable (Green et al., 1997; Johnson & Bickel, 2002; Kirby, 1997; Kirby & Maraković, 1995) and that future efforts will need to account for this. Kirby's (2000) own method accounts for this by determining an estimate for k over each of three blocks of nine questions. Our failure rate when attempting to estimate k for each block of questions with the simpler unconstrained routines was too high to support that approach, and for that reason, we focused on averages. However, the constrained routine was reliable in providing an estimate by block, with coefficients of determination for regressions of our results on those from Kirby's (2000) own method at .88, .89, and .90 for small, medium, and large reward blocks, respectively. With an expanded format for this survey, we expect that this approach will enable us to fit more complex models that incorporate V_D as a covariate and to model the effect of the magnitude of the delayed reward. (The multiplicative formulation described by Ho, Mobini, Chiang, Bradshaw, & Szabadi, 1999, appears to work with the same approach. Another suggestion would be to replace the hyperbolic parameter k with a linear model of k: $k = k_0 + \beta V_D$.)

Finally, we recommend using the constrained procedure when possible and the jackknife procedure for obtaining CIs. The interval estimates are based on *SE*s that are inaccurate if the constraints are active. The user should examine estimates of β_1, β_2, and k, determine whether any of them are at the constrained values, and ignore *SE*s and CIs calculated for those observations.

Further Research

The technique described here will probably see its greatest use in scoring larger fixed discounting surveys and reducing the size of computer task assessments of discounting. Many of those larger surveys and tasks were constructed to ascertain indifference, to be used in determining what function best fits the data, as well as in estimating the parameters. Most authors now assume that the best function to fit is hyperbolic, so gathering data sets large enough to determine the best-fitting function is not necessary.

We have conducted some preliminary trials (by simulation) of slightly expanded formats for the survey, including simple range extensions of the current layout (including smaller ratios at the left end and longer times at the right end of the scheme in Figure 2C and different layouts). We have seen some initial improvements in the ability to identify the model at the edges of the usual k range, but a full analysis is outside the scope of this article.

REFERENCES

Audrain-McGovern, J., Rodriguez, D., Tercyak, K. P., Epstein, L. H., Goldman, P., & Wileyto, E. P. (2004). Applying a behavioral economic framework to understanding adolescent smoking. *Psychology of Addictive Behaviors*, **18**, 64-73.

Bickel, W. K., Madden, G. J., & Petry, N. M. (1998). The price of change: The behavioral economics of drug dependence. *Behavior Therapy*, **29**, 545-565.

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology*, **146**, 447-454.

Crean, J. P., de Wit, H., & Richards, J. B. (2000). Reward discounting as a measure of impulsive behavior in a psychiatric outpatient population. *Experimental & Clinical Psychopharmacology*, **8**, 155-162.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, **25**, 715-723.

Ho, M. Y., Mobini, S., Chiang, T. J., Bradshaw, C. M., & Szabadi, E. (1999). Theory and method in the quantitative analysis of "impulsive choice" behaviour: Implications for psychopharmacology. *Psychopharmacology*, **146**, 362-372.

Hogg, R. V., & Craig, A. T. (1995). *Introduction to mathematical statistics*. Englewood Cliffs, NJ: Prentice-Hall.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression* (1st ed.). New York: Wiley.

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, **77**, 129-146.

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, **126**, 54-70.

Kirby, K. N. (2000). *Instructions for inferring discount rates from choices between immediate and delayed rewards*. Unpublished manuscript.

Kirby, K. N., & Marakovic, N. N. (1995). Modeling myopic decisions: Evidence for hyperbolic delay-discounting within subjects and amounts. *Organizational Behavior & Human Decision Processes*, **64**, 22-30.

Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review*, **3**, 100-104.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than nondrug-using controls. *Journal of Experimental Psychology: General*, **128**, 78-87.

Madden, G. J., Bickel, W. K., & Jacobs, E. A. (1999). Discounting of delayed rewards in opioid-dependent outpatients: Exponential or hyperbolic discounting functions? *Experimental & Clinical Psychopharmacology*, **7**, 284-293.

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control participants: Drug and monetary rewards. *Experimental & Clinical Psychopharmacology*, **5**, 256-262.

PETRY, N. M. (2001a). Delay discounting of money and alcohol in actively using alcoholics, currently abstinent alcoholics, and controls. *Psychopharmacology*, **154**, 243-250.

PETRY, N. M. (2001b). Pathological gamblers, with and without substance use disorders, discount delayed rewards at high rates. *Journal of Abnormal Psychology*, **110**, 482-487.

RICHARDS, J. B., MITCHELL, S. H., DE WIT, H., & SEIDEN, L. S. (1997). Determination of discount functions in rats with an adjusting-amount procedure. *Journal of the Experimental Analysis of Behavior*, **67**, 353-366.

NOTE

1. Our own use of this software has generally required approximately 120 questions to identify all indifference points. The actual number is indeterminate and may be different for each subject.

APPENDIX A
Stata Code for Estimating *k* Using Logit and GLM

Data must be in long format, with each question item on a separate line, 27 lines per person. We use integer variables "idn" for subject, "choice" for outcome (0,1), and (continuous) variables "ratio" as reward ratio V_I/V_D, "time" as delay time, and "ip1" as the constructed variable ($1 - 1/Ratio$). The best routine in Stata for our purposes is glm, which fails to converge less often than logit. The entry for estimating the model across the whole sample is:

```
glm choice ip1 time, nocons f(binomial) l(logit)
```

To collect individual estimates from many subjects, each identified by "idn," we use the "STATSBY" command ("/* " and "*/" continue a command across lines):

```
statsby "glm choice ip1 time, nocons f(binomial) l(logit)" /*
   */ _b lkhat = (log(_b[time]/_b[ip1])), by(idn) clear
```

So far, our code generates only estimates of *k* for each subject. In order to use the jackknife procedure to estimate standard errors, more coding is required. First, install the "levels" command (from the ssc Web archive) if you have not already done so.

```
ssc install levels
```

We then "GENERATE" empty variables to receive our results, create a local list of the idn values, use a loop (FOREACH) to apply the "JACKKNIFE" procedure to each subject, saving results for each successful run, and "COLLAPSE" to one line per subject.

```
gen b1 = .
gen b2 = .
gen lkhat = .
gen lkerr = .

levels idn, local(idnlevs)

foreach i of local idnlevs {

capture jknife "glm choice ip1 time, nocons f(binomial) l(logit)" /*
   */ _b lkjk = (log(_b[time]/_b[ip1])) if idn = = `i',e
scalar rc3 = _rc
replace b1 = r(stat1) if idn = = `i' & rc3 = = 0
replace b2 = r(stat2) if idn = = `i' & rc3 = = 0
replace lkhat = r(stat3) if idn = = `i' & rc3 = = 0
replace lkerr = r(se3) if idn = = `i' & rc3 = = 0
}

collapse b1 b2 lk*, by(idn)
```

The confidence interval is then the estimate *lkhat*, ± 1.96 *lkerr*.

APPENDIX B
Routines for Logistic Regression Using Stata ML, Constraining Parameter Values

First specify the likelihood function, including a penalty for violating bounds (copy exactly).

```
program define hyplogmlec
        version 7
        args todo b lnf
        tempvar theta1 tmp1 myb2 d1 d2 d3 d4 d5
        capture matrix `myb2' = `b'
        mleval `theta1' = `b',eq(1)
        quietly gen double `tmp1' = exp(`theta1')/(1 + (exp(`theta1')))
        mlsum `lnf' = $ML_y1*ln(`tmp1') + (1-$ML_y1)*ln(1-`tmp1')

note: Calculation of penalties.

        scalar `d1' = ((`myb2'[1,2]/`myb2'[1,1]) /*
                */ < 0.0001)*((`myb2'[1,2]/`myb2'[1,1])- 0.0001)^2
        scalar `d2' = ((`myb2'[1,2]/`myb2'[1,1]) /*
                */ > 1)*((`myb2'[1,2]/`myb2'[1,1])-1)^2
        scalar `d3' = ((`myb2'[1,1]) < -10)*((`myb2'[1,1]) + 10)^2
        scalar `d4' = ((`myb2'[1,1]) > -.2)*((`myb2'[1,1]) + .2)^2
        scalar `d5' = ((`myb2'[1,2]) > -.001)*((`myb2'[1,2]) + .001)^2
        scalar `lnf' = `lnf'-(`d2' + `d4' + `d1' + `d3' + `d5')*5000
        if `lnf' = = . { exit}
end
```

Again, we record the estimates, and the jackknife standard error of $\log(\hat{k})$, using a routine that calls the likelihood function. The ID variable "idn" may represent subject or subject by delayed reward size block (small, medium, large). The constrained procedure also provides estimates for the smaller nine-item sample. Complete routines for STATA may be found in the Statistical Software Components Archive (maintained by Boston College): http://ideas.repec.org/s/boc/bocode.html.

```
gen b1 = .
gen b2 = .
gen lkhat = .
gen lkerr = .
levels idn, local(idnlevs)

foreach i of local idnlevs {

capture jknife /*
*/ "ml model d0 hyplogmlec (choicevar = ip1 time, nocons),max" /*
*/ _b bias = (log(_b[time]/_b[ip1])) if idn = = `i',e

scalar rc3 = _rc
replace b1 = r(stat1) if idn = = `i' & rc3 = = 0
replace b2 = r(stat2) if idn = = `i' & rc3 = = 0
replace lkhat = r(stat3) if idn = = `i' & rc3 = = 0
replace lkerr = r(se3) if idn = = `i' & rc3 = = 0
}

collapse b1 b2 lk*, by(idn)
```

(Manuscript received October 1, 2002;
revision accepted for publication September 9, 2003.)