Discounting Model Selection with Area-based Measures: A Case for Numerical Integration

Shawn P. Gilroy

Department of Psychology

National University of Ireland, Galway

Donald A. Hantula

Department of Psychology

Temple University, Philadelphia

This manuscript is not the copy of record and may not exactly replicate the final, authoritative version. The version of record is available at https://doi.org/10.1007/s40489-018-0136-6

Correspondence may be sent to:

Shawn Patrick Gilroy
shawnpgilroy@gmail.com
shawn.gilroy@nuigalway.ie

Grant Sponsor: The charity RESPECT and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013)
Grant Number: PCOFUND-GA-2013-608728'

ABSTRACT

A novel method for analyzing delay discounting data is proposed. This newer metric, a model-based Area Under Curve (AUC) combining approximate Bayesian model selection and numerical integration was compared to the point-based AUC method developed by Myerson, Green & Warusawitharana (2001). Using data from Monte Carlo simulation and published study data, comparisons of the two methods indicated that model-based AUC offered a more robust measurement of discounting than provided by the point-based method alone. Beyond a calculation of area, numerical integration methods permitted discounting model selection to proceed in cases when the Effective Delay 50 (ED50) measure could not be calculated. This allowed discounting model selection to proceed in conditions where data are traditionally more challenging to model and measure, a situation where area-based methods are often enlisted. Results from simulation and existing data indicated that numerical integration methods extended both the area-based interpretation of delay discounting as well as the discounting model selection approach. Limitations of point-based AUC as a first-line analysis of discounting and additional extensions of discounting model selection were also discussed.

INTRODUCTION

Temporal discounting, or delay discounting, is an account of how the value of items or events change as a function of their distance from the present (Chung & Herrnstein, 1967; Ainslie, 1975; Mazur, 1987; Green, Fry & Myerson, 1994). Established in both human and nonhuman animal decision-making research, delay discounting presents an account of how organisms adapt in a complex and uncertain world (Ainslie, 1975; Ainslie & Herrnstein, 1981; Mazur, 1987; Critchfield & Kollins, 2001). For both human and nonhuman subjects, discounting procedures permit analyses of decision-making as delays to rewards are manipulated and differences in choices are observed (Kirby & Herrnstein, 1995). While hundreds of studies have utilized delay discounting procedures to analyze human and nonhuman choice, the theories and frameworks for analyzing and interpreting these phenomena have varied considerably (Ainslie, 1975; Rachlin, 2000; Madden & Bickel, 2010; Doyle, 2013).

Researchers in economics, marketing, psychology, and decision analysis have produced related, but contrasting accounts of how individuals arrive at specific delayed and temporally extended choices (Laibson, 1997; Doyle, 2013; van den Bos, & McClure, 2013). While few would dispute that the subjective value of an item is influenced by its distance from the present, the specific mechanism and course of these changes remains a particularly contentious topic (Samuelson, 1937; Ainslie, 1975; Laibson, 1997; Green & Myerson, 2004, Rachlin, 2006; van den Bos, & McClure, 2013). Debate continues between researchers as to whether the hyperbolic (Ainslie, 1975), exponential (Samuelson, 1937), quasi-hyperbolic (Laibson, 1997), or some hyperboloid model variant (Green & Myerson, 2004; Rachlin, 2006) is the best suited model for analyzing intertemporal choice. Doyle (2013) presents an extensive review of these models, as well as others.

Despite decades of research and hundreds of studies in this area, no single model has been able to adequately characterize all instances of individual intertemporal choice (Laibson, 1997; van den Bos & McClure, 2013). Traditional methods have historically employed single parameter models (e.g., exponential, hyperbolic), though more recent models have introduced additional parameters, augmenting the shape and characteristics of either the exponential (e.g., Quasi-Hyperbolic) or hyperbolic model (e.g., Hyperboloid) models (Ainslie, 1975; Green & Myerson, 2004; Laibson, 1997; Rachlin, 2006). Beyond these modifications, some have argued that discounting models should broaden much further to better account for other biological, neurological and contextual factors (van den Bos & McClure, 2013). In a break from model fitting altogether, others have proposed interpreting intertemporal choice data more simply and directly as the *empirical discounting function* (Myerson et al., 2001).

Myerson et al (2001) proposed the use of AUC, a measure of individual decision-making constructed by calculating the area beneath pairs of data points, interpolated with straight lines, and dividing the sum of these segments by the total possible area. Myerson and colleagues referred to this interpolation between data points as the *empirical discounting function*, hereafter referred to as *point-based AUC*, a conceptually-neutral method for quantifying choice over time. Point-based AUC linearly interpolates between pairs of data points with the underlying assumption that straight-line approximations provide a conceptually-neutral estimation of data at unmeasured delays. That is, the empirical discounting function uses a series of straight lines to estimate a “curve” of sorts. While not a model in the traditional sense, the interpolation constructed using the empirical discounting function could be viewed as a sort of saturated, segmented model. Such a model would contain as many parameters as there are data points, connected by linear segments, to the effect of linearly interpolating the observed data. The *empirical discounting function* operates in this way, reconstructing the data "perfectly" by connecting data with straight lines (e.g., a makeshift curve).

The *empirical discounting function* (point-based AUC) shares several of the same limitations inherent in saturated models. First, this type of model cannot be compared to any other because there is no error variance available to analyze. By interpolating the data with straight lines, the function passes through each point perfectly (e.g., no residual error). The fit is "perfect", in that the model is essentially a re-construction of the observed data. Without a means to make comparisons with other models, it cannot be determined whether any other model could more simply and effectively account for that data. That is, the *empirical discounting function* can only ever be compared to itself. Second, this type of model is essentially a re-construction of the observed data. The model is an interpolation of the data and any projections beyond the data provided would be highly questionable and inappropriate. For example, the *empirical discounting function* cannot make predictions at unmeasured delays that are before the initial or after the final observed data point. Further, it assumes that any delays between the measured points in the dataset will perfectly follow that linear path between the measured points. As a result of these limitations, the *empirical discounting function* offers a basic *summary* of the available data but little in the way of *predictive* value--it can summarize data but not explain how delays affect choice behavior.

Discounting researchers who have traditionally used point-based AUC in lieu of model fitting likely elected to do so for one of several reasons. First, discounting data are inherently variable and it is often the case that instances of discounting can vary substantially from individual to individual. As a result, individual discounting patterns can and will be better characterized by different types of discounting models (Franck et al., 2015). Researchers planning to use some prescribed or assumed model (e.g., hyperbolic) for all participants in a study may find themselves foregoing model fitting altogether and instead applying point-based AUC methods when a model of interest does not fit well with one or more participants. For these participants, researchers may elect to instead use the point-based AUC method to summarize the observed data of each participant, rather than explore alternative models. Second, point-based AUC methods do not require specialized software to compute and preclude the need to apply any model fitting techniques (Myerson et al., 2001). It is computationally simple to perform, is readily calculated by hand or with basic spreadsheet software (e.g., Kaplan et al., 2016) and does not require the researcher to systematically compare other competing models.

While point-based methods are considerably simpler to compute, this simplicity is not without caveats. Point-based AUC calculations remain conceptually neutral throughout analysis. It is always possible that some alternative model may more simply represent the data. To remain conceptually neutral, point-based AUC operates with certain assumptions. First, this method assumes that data are visually inspected to determine that some degree of *discounting* actually took place (Myerson et al., 2001). This assumption is paramount, as it is typically regarded that the point-based AUC measure generally represents a *downward* sloping function. Visual analysis is required for every data series in this approach because point-based AUC methods will provide a summary of the data regardless of whether discounting has taken place. For example, consider a case in which three participants in a typical delay discounting task yielded data in which the point-based AUC is identical. As illustrated in Figure 1, one participant may have discounted (downward slope) while the others may have shown either a noise pattern (no systematic slope) or accretion (upward slope). Regardless of the direction of the data, the point-based measurement of AUC would treat all cases as if they behaved in the same manner. Figure 1 shows that calculations of point-based AUC are not specific to *discounting* and that many very different arrangements of data can produce an identical measurement of point-based AUC. The conceptually neutral stance provided by point-based methods make it possible that AUC may be closer to 1.0 (e.g., nearer 100% of maximum) because data are *increasing* towards the maximum as delays increase. When interpreted as a saturated model, this may make some good sense because point-based AUC methods can provide a *summary* of the observed data, but cannot *predict* how future delays are influencing choice behavior (e.g., are the data increasing or decreasing). This is because the model here serves to reconstruct the data.

More recent attempts to address the difficulties encountered when fitting a single model across participants have explored a multi-model approach combined with generalized measures of discounting. Franck et al. (2015) described a method that used the Bayesian Information Criterion (BIC; Schwarz, 1978) to compare, and select from, multiple discounting models. Franck and colleagues utilized approximate Bayesian model selection to identify the best performing model for an instance of discounting and then measured discounting in a way that could be compared regardless of the model used. Using the Effective Delay 50 (ED50), the point at which a function arrived at 50% of its initial value, instances of discounting could be measured and interpreted similarly regardless of the discounting model it was derived from (Yoon & Higgins, 2008). While the discounting model selection procedure is robust and well-suited to discounting applications, calculations of the ED50 have certain assumptions that must be met. First, calculations of ED50 require that a discounting model candidate fits well and that this model demonstrates some type of systematic discounting. That is, the function must be moving in a downward sloping direction towards 50% of its starting value. Calculations of ED50 cannot be performed when data are either *increasing* or characterized by the Noise model. A series characterized by the Noise model is best represented by the intercept alone (e.g., the mean of the Y values), which neither ascends or descends; in other words, a straight line. In this instance, an ED50 cannot be calculated because that model is not moving in the direction towards 50% of its maximum value. Second, the ED50 is potentially limited when certain models rapidly reach 50%, or below, the initial value. For example, the quasi-hyperbolic (Laibson) model cannot have an ED50 calculated when the *beta* parameter is 0.5 or below. In such a case, where the present-moment bias (*beta*) is already 50% of a maximum value, the ED50 cannot be calculated because the 50% mark has occurred before any delays took place. While ED50 cannot be calculated in the above cases, an area-based measurement is still possible using the most probable model. While researchers always have the option to use point-based AUC, and forego model fitting entirely, the discounting model selection approach can still proceed in these cases by using numerical integration to calculate the area beneath a fitted model.

Similar to how the point-based AUC method calculates area, numerical integration entails the calculation and summation of area segments. While point-based AUC is calculated by summing the area under pairs of data points, numerical integration calculates the area under a fitted model between specific bounds (e.g., first and last delay). Performing numerical integration on a fitted discounting model permits the summation of theoretically infinite segments--numerical integration is not constrained by the amount of data collected. Through a process of calculating and summing a large number of very small and precise segments, the area under the model ultimately converges to a measure that captures the area beneath the unique shape of the most probable, fitted model. Once area from numerical integration is divided by the maximum possible area, the result is a measurement that ranges from 0.0 and 1.0, which is consistent with the traditional interpretation using point-based AUC while retaining its interpretive and statistical advantages. In addition to a measure of area, calculations of model-based AUC are accompanied by a known model and a set of fitted model parameters. A model-based AUC combines the *summary* value of an area-based measurement of discounting with the *predictive* power of a known model. The model and parameter(s) provided offer useful, statistically desirable information regarding a subject that traditionally would have had to been inferred from visually inspecting the data. Within this information, individual model parameters speak directly to whether a function is discounting or not--information not afforded when calculating area with data points outright.

Numerically integrating models of discounting offers new opportunities for discounting researchers. Among these, a model-based AUC can be calculated using most software already used for applying nonlinear curve fitting (R Core Team, 2017). In these cases, there is no additional software necessary to either purchase or learn. Second, calculating area from discounting models allows researchers to calculate a conceptually-driven AUC from models that can be systematically and statistically compared (i.e., through model selection). This is an improvement over applying the point-based AUC method outright, as it is very likely that instances of individual discounting could be represented by a more robust and parsimonious model than the empirical discounting function. Third, the conceptual and empirical bases of discounting models are preserved and thus available for further analysis in model-based calculations of AUC. Area calculations performed in this manner can still directly reference a conceptual model and its individual parameters, error variance, and other goodness-of-fit measures. No parameters are added, lost, or modified in this process. Among such benefits, these metrics provide the researcher with empirical measures of whether discounting was observed. This may be of benefit in instances where the visual analysis of every instance of each participant is particularly taxing (e.g., results from 1000+ participants). Fourth, numerical integration substantially extends the capabilities of the discounting model selection approach. Area-based calculations enable the discounting model selection approach to continue in cases where the ED50 cannot be calculated. This is a substantial extension of the multiple model approach to analyzing individual discounting. Numerical integration methods allow discounting model selection to capitalize on the flexibility and interpretability provided by area-based measurements. That is, this combination allows model selection to proceed in cases where data series are particularly challenging to model.

To evaluate a model-based calculation of AUC, and examine its strengths and limitations, analyses with simulated and real data were needed to compare both methods of calculating AUC across a range of conditions. This study sought to answer the following questions: (1) are model-based calculations of AUC, using numerical integration, consistent with point-based calculations of AUC when data emerge from various models of discounting, (2) are model-based calculations of AUC consistent with point-based calculations of AUC when data are characterized by the Noise model and (3) does the discounting model selection using model-based AUC produce similar results to point-based AUC when real-world data are used?

Simulation Study

METHOD

Simulated Participants

Simulation studies were conducted to compare the results of point- and model-based AUC. Individual simulations were designed and performed for each of the models included within the discounting model selection procedures described in Franck et al. (2015). Individual simulations were constructed for the exponential, hyperbolic, quasi-hyperbolic, Myerson hyperboloid and Rachlin hyperboloid models to produce data that emerged from a range of underlying models. Individual models and their structure are shown in Table 1. Each simulation was programmed to produce 10,000 randomly-generated discounting series. Identical to the simulations from Franck et al. (2015), delay points included 1 day, 1 week, 2 weeks, 1 month, 6 months, 1 year, 5 years, and 25 years. Simulation parameters were derived from parameter values and error variance from 111 college students completing a hypothetical monetary delay discounting task (Franck et al., 2015). A more complete description of data simulation procedures and parameters (e.g., starting parameters, error variance) can be found in Franck et al. (2015) and the resources necessary to replicate the present simulations are described in the appendix of this work.

Model Selection Procedures

Model selection entails a comparison of potential models based on some index of model performance (Franck et al., 2015). Following simulations of hypothetical discounting data, model selection procedures were performed for each series using the BIC (Schwarz, 1978) and subsequent model probabilities (Kass & Raftery, 1995). As indicated in Table 1, five commonly used discounting models were included in the model selection procedure and these were selected based on the present state of discounting model selection (Franck et al., 2015). A Noise model was included as a candidate to model instances where discounting was best characterized by a single, straight line (i.e., intercept only). Consistent with model selection procedures in Franck et al. (2015), models that could not be optimized were not included in model comparisons and selection procedures continued with the remaining models. Models were compared using the BIC and the model with the highest model probability was selected for use in subsequent analyses.

Nonlinear Model Fitting and Numerical Integration

The R statistical program was used to fit all statistical models and to perform numerical integration. The *nls* package within R was used to fit all discounting models (Bates & Watts, 1988; R Core Team, 2017). Using the default Gauss-Newton optimizer, maximum iterations were fixed to 50, error sum of squares to elicit convergence was set to 10^{-5}, and step sizes were set to no less than 2^{-10}. Consistent with Franck et al., (2015), the Laibson β and δ parameters were constrained within the range of 0-1 and all other parameters were fitted freely. The Rachlin model *s* parameter was not constrained by default, which was necessary to meet conditions for regularity necessary for the model selection approach. The *integrate* package within R was used to perform all numerical integration (Piessens, deDoncker–Kapenga, Uberhuber & Kahaner, 1983; R Core Team, 2017). The upper and lower bounds were set to the highest and lowest delay points, respectively, and the most probable discounting model was set to be the function integrated upon. All other settings remained at defaults.

Point-based AUC Calculation using the Trapezoid Rule

Point-based AUC calculations were performed as described in Myerson et al. (2001). This form of AUC was calculated using the Trapezoid Rule, where individual area segments were calculated by averaging adjacent Y-values and multiplying them by the respective increment, see Equation 1. The result of this calculation was a rectangular segment approximating the area underneath pairs of points. The area of the collective rectangular segments was divided by the maximum area possible. The maximum area possible was the product of the full domain, or between the first and final delays, and the maximum value of the discounted item. The final calculation yielded a ratio of area under data points to maximum area, which ranged from 0.0 to 1.0.

Equation 1. $\frac{Y_n+Y_{n+1}}{2}(X_{n+1}-X_n)$

Model-based AUC Calculation using Numerical Integration

Numerical integration is a methodology that allows for calculating the area beneath an arbitrary function. In the case of discounting functions, numerical integration allows one to determine the area beneath a fitted model over a fixed domain as defined by the first and final delay point. The calculation of maximum possible area was identical to that of point-based AUC. Model-based AUC was calculated using a modified version of the Discounting Model Selector (DMS; Gilroy, Franck & Hantula, in press). This software conducted discounting model selection and was modified to apply numerical integration methods rather than calculate the ED50. This modified version of the software, and resources to recreate it, are listed in the Appendix of this work. Once the area beneath a fitted discounting function was solved for, model area was divided by the maximum possible area to calculate a ratio of area under curve to maximum area, which ranged from 0.0 to 1.0.

Model-based AUC Calculation for data characterized by the Noise Model

In data series characterized by the Noise model, discounting model selection would have found that the intercept alone characterized the data better than other, more complicated models (e.g., exponential, hyperbolic). As such, the model was simply the intercept extending from the first to the final delay point, see Figure 2. In these cases, AUC was calculated as the area beneath the intercept between the first and final delay point. While numerical integration could be performed upon this line, it was computationally more efficient to simply divide the Y value of the intercept by the maximum value of the commodity. Calculated in this way, this measure represented a ratio of area beneath the function to total area, yielding a figure that ranged between 0.0 and 1.0.

RESULTS

Discounting Model Selection for Simulated Data

Discounting model selection procedures were applied to all simulated discounting series. Consistent with earlier studies on discounting model selection, a range of discounting models were found to characterize simulated participants (n=50,000). The exponential (n=8,776; 17.55% of series), hyperbolic (n=11,003; 22.01% of series), quasi-hyperbolic (n=11,073; 22.15% of series), Myerson-Green Hyperboloid (n=6,158; 12.32% of series) and Rachlin Hyperboloid model (n=11,867; 23.74% of series) were characterized by similar proportions of all data series overall. Noise models characterized a small percentage of simulated series (n=1,123; 2.25% of series).

Comparing Point-based AUC to Model-based AUC for Discounting Models in Simulation Data

Point- and model-based AUC calculations were compared within and across the most probable discounting model for simulated data. A strong relationship was observed between both forms of AUC overall, as illustrated in Figure 3. A Spearman rank order correlation between these two measures, across all underlying models, revealed a strong relationship overall (S = 8.8444e+11, rho = .9575, p < .0001). Correlations between the two measures were strong at the individual model level as well. Spearman rank order correlations for individual discounting model comparisons were as follows: Hyperbolic (S = 1.0244e+10, rho = .9539, p < .0001), Exponential (S = 9.0772e+9, rho = .9194, p < .0001), Beta-Delta (S = 2.5314e+10, rho = .8881, p < .0001), Myerson Hyperboloid (S = 1.6759e+8, rho = .9957, p < .0001), and Rachlin Hyperboloid models (S = 5.9538e+9, rho = .9786, p < .0001).

Comparing Point-based AUC to Model-based AUC for Noise Models in Simulations

Roughly 2% (n=1,123) of simulated discounting data were characterized by the Noise model. These data were analyzed to examine how model-based AUC utilizing a Noise model (e.g., straight, horizontal line) related to the point-based AUC method (e.g., linear interpolation of data points). A Spearman rank order correlation between both forms of AUC calculations indicated a very strong relation in these situations (S = 1.1624e+7, rho = .9507, p < .0001). The relationship between these metrics is illustrated in Figure 4. As indicated within marginal histograms, series best characterized by the Noise model tended to be series with either little-to-no (closer to 1.0) or very-high (closer to 0.0) discounting. Instances of discounting between these two extremes were more often characterized by one of the more traditional discounting models.

Comparing Point-based AUC to Model-based AUC in Published Data

To demonstrate the utility of a discounting model selection approach using numerical integration, data were extracted from a published paper and re-analyzed. This served to both apply a model-based calculation of AUC to real data as well as review the process of applying discounting model selection with numerical integration. All model selection, numerical integration, and calculations of point-based AUC were as described earlier in the methods section.

METHOD

The specific data, methods, analyses and results were as listed and described in Takahashi, Masataka, Malaivijinond and Wongsiri (2008). Takashi et al (2008) examined potential differences in the rates of discounting between reinforcer types and magnitude. More specifically, they evaluated the impact of delays on the subjective value of food (preservable rice) and currency (baht) across high- and low-magnitude conditions. The baht was a currency used in Thailand, where the study was conducted.

Study Participants and Data from Takashi et al (2008)

Forty-eight students at a local university participated in group-based, hypothetical decision-making tasks. Individual participants were assigned to either a high- (*n*=20) or low-magnitude (*n*=28) condition. In each of these conditions, all participants completed hypothetical decision-making tasks for food (10 or 100 kg of rice) and currency (200 or 2000 baht). Both the high- and low-magnitude tasks were designed to be offer roughly equivalent worth across commodities (e.g., 10 kg of rice cost roughly 200 baht). Individual participant responding was extracted and analyzed as recorded in Tables 1, a-d of the source material (Takashi et al 2008).

Study Methods from Takashi et al (2008)

Participants in all groups completed a checklist-format questionnaire after being read a script outlining the discounting task. All instructions were presented in English. Hypothetical monetary choice tasks were presented both vocally by an experimenter as well as on written media (i.e., paper checklist, blackboard, etc.). The delays included in the assessment were six months, one year, five years, ten years, and twenty years. The possible immediate reward values were 1%, 2.5%, 5%, 7.4%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92.5%, 95%, 97.5%, and 99% of the maximum value of the commodity. One five-trial practice session was conducted in English prior to experimental conditions to confirm sufficient understanding of the English language.

Study Results from Takashi et al (2008)

Takashi et al (2008) presented analyses using both the hyperbolic discounting model (Mazur, 1987) and point-based AUC. They had foregone model-fitting for all participants following observations that a single model was insufficient to adequately characterize all instances of participant responding. For the sake of this example, only point-based AUC was reviewed and re-analyzed. Consistent with the source paper's analyses, a repeated measures ANOVA was conducted for the point-based AUC observed in the four conditions (reinforcer magnitude vs. reinforcer type). Analyses of point-based AUC with extracted data revealed a significant main effect for reinforcer type, $F = 8.016$ (1, 34), p < .01, but not for reinforcer magnitude, $F = 1.007$ (1, 34), p = .33. A significant interaction was observed between reinforcer type and magnitude, $F = 7.56$ (1, 34), $p = < .01$. All results were identical to that of Takashi et al (2008).

The same statistical comparison was repeated using the results of model-based AUC. Consistent with the results of Takashi et al (2008), a significant main effect was observed for reinforcer type, $F = 7.772$ (1, 34), p <.01, but not for reinforcer magnitude, $F = 1.014$ (1, 34), p = .32, using model-based AUC. A significant interaction was also observed between type and magnitude, $F = 8.686$ (1, 34), $p = < .01$, using this new metric. These results indicate that a model-based form of AUC provided nearly identical results to that of point-based AUC.

While both forms of AUC offered consistent measurements and comparisons, using a model-based approach to calculate AUC provided information that would have been unavailable using either a single-model approach or applying point-based AUC outright. As indicated in Table 2, re-analyses of the Takahashi et al. (2008) participant data revealed that individual decision-making was characterized by a range of different models overall. Interestingly, discounting model selection revealed that individual patterns of decision-making often varied at the individual-level when the type of reinforcers changed. For example, 80.95% (n = 17) and 60% (n = 9) of participants demonstrated different patterns in discounting between food and currency in the low- and high-magnitude conditions, respectively. Data gleaned from model selection suggested that individual discounting processes varied considerably between reinforcer types in both lower and higher magnitude conditions, even though a significant effect was not observed in the overall AUC between higher and lower magnitude conditions. Additionally, model-based AUC and discounting model selection was able to accommodate the two cases where discounting data were characterized by the Noise model. If discounting model selection would have used the ED50 as the metric of discounting instead, no data could have been calculated and these cases would have had to been dropped from any further analysis and comparison.

Beyond knowledge of the model that likely underlies AUC, a model-based AUC provided statistical information not provided when calculating point-based AUC outright. Principal among these benefits, fitted model parameters spoke to the general direction of the data over time. More specifically, these parameters guide the interpretation of AUC by indicating whether or not discounting actually took place. Using this approach, calculations of model-based AUC for participant number nineteen in the baht condition section of Takahashi et al. (2008) revealed that this participant did not actually discount over time. Rather, this individual's data increased as delays grew and this pattern resulted in an AUC deceptively close to 1.0. As depicted in Figure 5, AUC in this instance produced a figure that misleadingly suggested slight discounting (AUC: 0.766), but represented a savings function. While point-based AUC alone offered little to distinguish discounting from savings without visual analysis, the fitting of a probable discounting model (Rachlin hyperboloid) and interpretation of model parameters did indicate this statistically. The fact that the *s* parameter was signed negative suggested that the function operated in the opposing direction as typically assumed in a discounting function. Using discounting model selection in this manner, a wider range of statistical measures were provided to the researcher and assisted in the analysis and interpretation of individual choice behavior.

CONCLUSIONS

While numerical integration is a commonly used method in many areas of mathematics, its use in the behavioral sciences and in the measurement of decision-making represents a novel departure from established area-based methods. This type of calculation significantly extends the capabilities of the discounting model selection approach and the present study sought to answer the following questions: (1) does model-based AUC derived from a range of discounting models correspond with point-based AUC, (2) does model-based AUC derived from the Noise model correspond with point-based AUC and (3) does the discounting model selection using model-based AUC produce similar results to point-based AUC when real-world data are used? Analyses of these methods using both simulated and real-world data indicated that a model-based AUC using numerical integration produced an area metric that was highly consistent with traditional point-based AUC. Additionally, these comparisons suggested that model-based methods were equally applicable to both systematic and more variable data series (e.g., Noise model), which are often the case for using point-based AUC measure of discounting.

With respect to the first question, simulation results indicated that model-based AUC, derived from numerical application upon a probable discounting function, correlated very strongly with the traditional point-based AUC method. This type of relationship is not altogether surprising. When viewed as a saturated model of sorts, the empirical discounting function is based upon interpolated data that, to some unmeasured degree, resembles a curve. However, lacking any residual error variance, it impossible to determine whether a simpler, more parsimonious model could simply represent the approximated curve. As such, it is entirely possible that one or more traditional discounting models could fit that data. In contrast, a model-based AUC using model selection systematically analyzes error variance to determine which model most simply, best fits the data. As a result of this, the area under these two very different methods ultimately arrives at very similar measurements. While ultimately arriving at a relatively similar figure, the model-based approach to AUC provides both the *summary* value afforded by the area measurement and the *predictive* capabilities of an established model of discounting.

Per the second question, this study examined whether or not numerical integration applied to a Noise model (e.g., intercept alone) would relate to point-based AUC in cases where a data series was best characterized by a Noise model. This was a question of considerable interest, as calculations of point-based AUC are traditionally enlisted in situations where data are not easily modeled. In instances when the discounting model selection identified the Noise model as the most probable, point-based AUC was calculated as typical and model-based AUC was calculated as the area beneath a single, horizontal line of fit. While remarkably distinct, both AUC methods produced similar results overall. Statistically, this makes good sense from the perspective of discounting model selection. To have found the Noise model to be most probable, the intercept alone would have to have been found to represent that data better than more complicated alternatives. This is consistent with the notion that the empirical discounting function could be represented by some alternative, simpler model. This was observed in both simulated and extracted study data. In the case of Takahashi et al. (2008) specifically, participants numbered twenty-four (point-based AUC: 0.11; model-based AUC: 0.11; Lower magnitude food condition) and fifty (point-based AUC: 0.14; model-based AUC: 0.15; Higher magnitude currency condition) demonstrated responding that was best characterized by the Noise model. In these cases, traditionally challenging arrangements of data arrived at a similar figure and both individual- and group-level analyses yielded the same results with both forms of AUC.

For the third and final research question, results showed that model-based AUC can be applied to real-world data. Re-analyses of an existing data set indicated that model-based AUC provides area estimates of discounting that were consistent with point-based AUC methods, but offered information not available through a conceptually-neutral approach. Adopting a conceptually-driven approach to AUC offers significant potential benefits to applied delay discounting researchers. First, having parameters from a probable discounting model offers an empirical means to enhance visual analysis of individual responding. That is, parameters of a fitted model indicate whether discounting took place. This may be of particular interest to researchers with larger sample sizes, when the number of individual series to analyze visually becomes particularly burdensome. For example, Jarmolowicz et al (2012) reported a delay discounting study conducted on MTurk with over 900 participants. Such larger sample sizes may become more common as an oft-invoked solution to the replication and reproducibility crisis in behavioral sciences (Open Science Collaboration, 2015), economics (Chang & Li, 2015) and medicine (Ioannidis, 2005) is to conduct studies with much larger sample sizes (Etz & Vandekerckhove, 2016; Francis, 2013).

Second, a conceptually-driven approach to AUC speaks to both *how* and *how much* discounting took place. While point-based AUC offers a means to summarize *how much* discounting was observed (if visually confirmed to show discounting), the conceptually-neutral basis for it prevents any further analysis of *how* discounting took place over time. This is consistent with the interpretation that point-based AUC as a saturated model having good *summary* value but little *predictive* power. Takahashi et al. (2008) provides a good example of this limitation. Using point-based AUC, they found a main effect for reinforcer type for the overall amount of discounting observed (e.g., *how much*) but were not able to describe *how* individual discounting individual might have been different for participants across conditions. It is entirely plausible (and likely) that individuals might have varied in both their *degree* of discounting and their *process* of discounting when reinforcers types and magnitudes changed. Point-based methods do not permit further analysis, or prediction, of how shorter and more delayed conditions affected an individual's choices. In contrast, model-based AUC methods provide information regarding an individual's rate of change over time and can account for the individual slope and path of the individuals fitted discounting function. In this manner, an individual's steeper discounting nearer the present, and more gradual discounting in the future, is retained for further analysis and interpretation. Maintaining a conceptually-driven approach throughout assessments of individual discounting could highlight additional factors that underlie differences in individual discounting.

The results of this study overall support the use of numerical integration within the discounting model selection approach because it can be adapted to offer both the strong conceptual support for models of discounting and the flexibility offered by an area-based interpretation of discounting. While these findings indicated that area-based calculations of discounting expand upon the present capabilities of discounting model selection, we do not mean to discourage the use of the ED50 metric. Rather, numerical integration should be seen as an additional, alternative means of measuring individual instances of discounting in cases where responding may be more challenging to model. This novel approach grants new benefits to researchers analyzing and comparing individual discounting. In a field marred by debate over the specific models and metrics used, discounting model selection may serve to provide a statistically-informed (but neutral) process for modeling discounting data with the opportunity to interpret responding in several ways (e.g., individual models and parameters, ED50, model-based AUC). A shared set of methods and procedures may assist in integrating lines of research that have historically reported only a single type of metric. Also, given that numerical integration is largely a transformation of some model into a proportion of area, this approach supports research at the individual model level as well.

In summation, numerical integration serves as a novel and robust methodology for measuring and analyzing individual discounting. All together, these methods combine approximate Bayesian model selection, robust mathematical procedures, the general interpretability of area metrics, and the rich conceptual support afforded by established models of discounting. The combination of these methods simultaneously permits research and analysis of individual discounting on multiple levels, providing researchers the option to evaluate responding at the model-level and beyond, together rather than apart.

LIMITATIONS

While results from computer simulation and re-analyses of existing data were encouraging, numerical integration methods are a substantial departure from the traditional methods for analyzing discounting phenomena using area. Additional experimental and clinical evaluation is necessary to better understand how a model-based form of AUC, which references the discounting model itself, may behave when different patterns of discounting are observed. This approach could potentially yield even more robust measures of discounting in a discounting model selection approach, especially in cases when models are unable to produce a calculation of ED50. However, additional research is necessary to real-world research of these methods would be necessary to do so.

APPENDIX

The source code necessary to compile and run all simulations are publicly shared on the corresponding author's public Git, located at https://github.com/miyamot0. All simulations were conducted using the R programming language. The R scripts necessary to conduct and re-create all simulations can be found on Github in the repository named "Discounting Model Selector Simulations." The raw data used in the study is publicly shared in the same location.

The software used to conduct all analyses in this work is freely available at https://github.com/miyamot0/ModelSelectorQt in the source code form. It is a cross-platform, open-sourced project (General Public License-Version 3.0), built on earlier behavioral decision-making tools. It will conduct discounting model selection and apply either the ED50 or model-based AUC, at the user's discretion. It can be run as a script for the R programming language or as a cross-platform Graphical User Interface (GUI). Runnable binaries can be compiled for the Windows, Linux, and Apple computing environments.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Christopher T. Franck for his statistical expertise as well as his open-sourced R scripts for simulating delay discounting data.

REFERENCES

Ainslie, G. W. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*, 463-496.

Ainslie, G. W. & Herrnstein, R. J. (1981). Preference reversal and delayed reinforcement. *Animal Learning and Behavior*, *9*, 476-482.

Bates, D. M. & Watts, D. G. (1988). Nonlinear Regression Analysis and Its Applications, John. Wiley & Sons, New York.

Chang, Andrew C., and Phillip Li (2015). "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not'," *Finance and Economics Discussion Series 2015-083*. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.083.

Chung, S. H. & Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, *16*, 67-74.

Critchfield, T. S. & Kollins, S. H. (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis*, *34*, 101-122.

Doyle, J. R. (2013). Survey of time preference, delay discounting models. *Judgment and Decision Making*, *8*, 116-135.

Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *Plos ONE*, *11*(2): e0149794. doi:10.1371/journal.pone.0149794.

Francis, G. (2013). We don't need replication, but we do need more data. *European Journal Of Personality*, *27*(2), 125-126.

Franck, C. T., Koffarnus, M. N., House, L. L., & Bickel, W. K. (2015). Accurate characterization of delay discounting: A multiple model approach using approximate Bayesian model selection and a unified discounting measure. *Journal of the Experimental Analysis of Behavior*, *103*(1), 218-233.

Gilroy, S. P., Franck, C. T., & Hantula, D. A. (in press). The Discounting Model Selector: Statistical software for delay discounting applications. *Journal of the Experimental Analysis of Behavior.* https://doi.org/10.1002/jeab.257.

Green, L. & Myerson, J. (2004). A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological Bulletin, 30,* 769-792.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. *PLoS Med 2*(8): e124. doi:10.1371/journal.pmed.0020124

Jarmolowicz, D. P., Bickel, W. K., Carter, A. E., Franck, C. T., & Mueller, E. T. (2012). Using crowdsourcing to examine relations between delay and probability discounting. *Behavioural Processes*, *91*(3), 308-312. doi:10.1016/j.beproc.2012.09.001

Kaplan, B. A., Amlung, M., Reed, D. D., Jarmolowicz, D. P., McKerchar, T. L., & Lemley, S. M. (2016). Automating scoring of delay discounting for the 21- and 27-item Monetary Choice Questionnaires. *The Behavior Analyst*, *39*(2), 293-304. doi:10.1007/s40614-016-0070-9

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Kirby, K. N., & Herrnstein, R. J. (1995). Preference reversals due to myopic discounting of delayed reward. *Psychological Science*, *6*, 83-89.

Laibson, D. (1997). “Golden Eggs and Hyperbolic Discounting” *Quarterly Journal of Economics, 112*(2), 443-477.

Madden, G. J., & Bickel, W. K. (2010). *Impulsivity: The behavioral and neurological science of discounting*. Washington, DC US: American Psychological Association.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analysis of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value* (pp. 55-73). Hillsdale, NJ: Erlbaum.

Myerson, J., Green, L., & Warusawitharana M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, *76*, 235-243.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Piessens, R., deDoncker–Kapenga, E., Uberhuber, C., & Kahaner, D. (1983). Quadpack: A Subroutine Package for Automatic Integration; Springer Verlag.

Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior*, 85, 425-435.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Samuelson, P. A. (1937). A note on measurement of utility. *Review of Economic Studies*, *4*, 155-161.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, *6*(2), 461-464.

Takahashi, M., Masataka, N., Malaivijitnond, S., & Wongsiri, S. (2008). Future rice is discounted less steeply than future money in Thailand. *The Psychological Record, 58*(2), 175-190.

van den Bos, W., & McClure, S. M. (2013). Towards a general model of temporal discounting. *Journal Of The Experimental Analysis Of Behavior*, *99*(1), 58-73.

Yoon, J. H., & Higgins, S. T. (2008). Turning k on its head: Comments on use of an ED50 in delay discounting research. *Drug and Alcohol Dependence*, *95*(1), 169-172.

Table 1

Discounting Model Selection Candidates

Model	Source	Model Structure
Exponential	Samuelson (1937)	Ae^{-kx}
Hyperbolic	Mazur (1987)	$\frac{A}{1 + kx}$
Quasi-Hyperbolic	Laibson (1997)	bde^{-kx}
Myerson & Green Hyperboloid	Myerson & Green (1995)	$\frac{A}{(1 + kx)^s}$
Rachlin Hyperboloid	Rachlin (2006)	$\frac{A}{(1 + kx^s)}$

Table 2

Model Characterizations in Takahashi, Masataka, Malaivijinond and Wongsiri (2008)

	Low Magnitude		High Magnitude	
Model	Food	Currency	Food	Currency
Noise	0% (*n*=0)	4.76% (*n*=1)	0% (*n*=0)	6.67% (*n*=1)
Mazur	9.52% (*n*=2)	9.52% (*n*=2)	6.67% (*n*=1)	0% (*n*=0)
Exponential	0% (*n*=0)	4.76% (*n*=1)	0% (*n*=0)	6.67% (*n*=1)
BD	42.85% (*n*=9)	28.57% (*n*=6)	40% (*n*=6)	40% (*n*=6)
Green Myerson	19.04% (*n*=4)	38.09% (*n*=8)	26.67% (*n*=4)	20% (*n*=3)
Rachlin	28.57% (*n*=6)	14.28% (*n*=3)	26.67% (*n*=4)	26.67% (*n*=4)

Figure 1. Equivalent point-based AUC for discounting, noise and growth functions

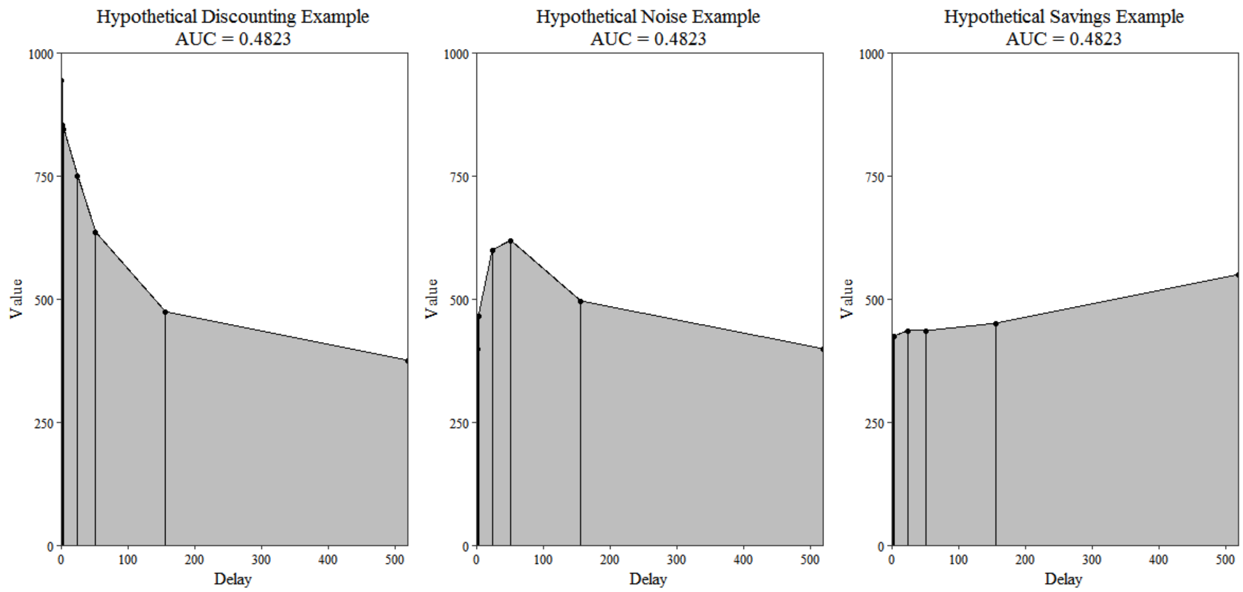

Figure 2. Comparison of point- and model-based AUC for data characterized with Noise Models

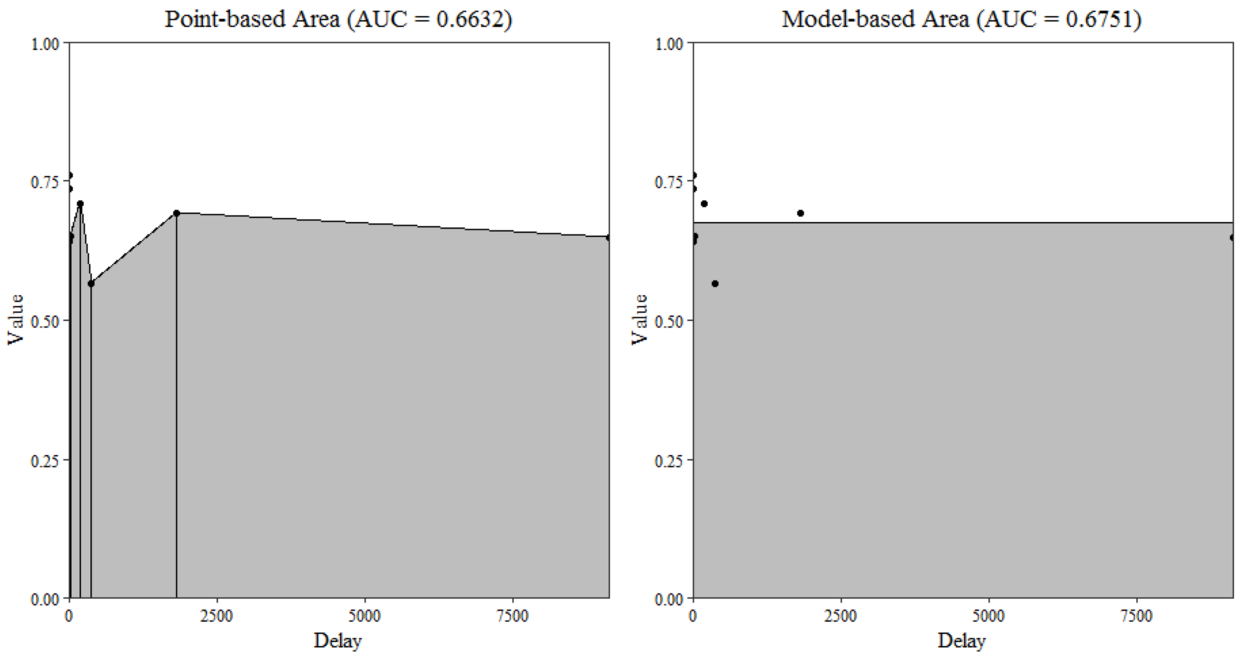

Figure 3. Area of Point AUC versus Most Probable Model AUC

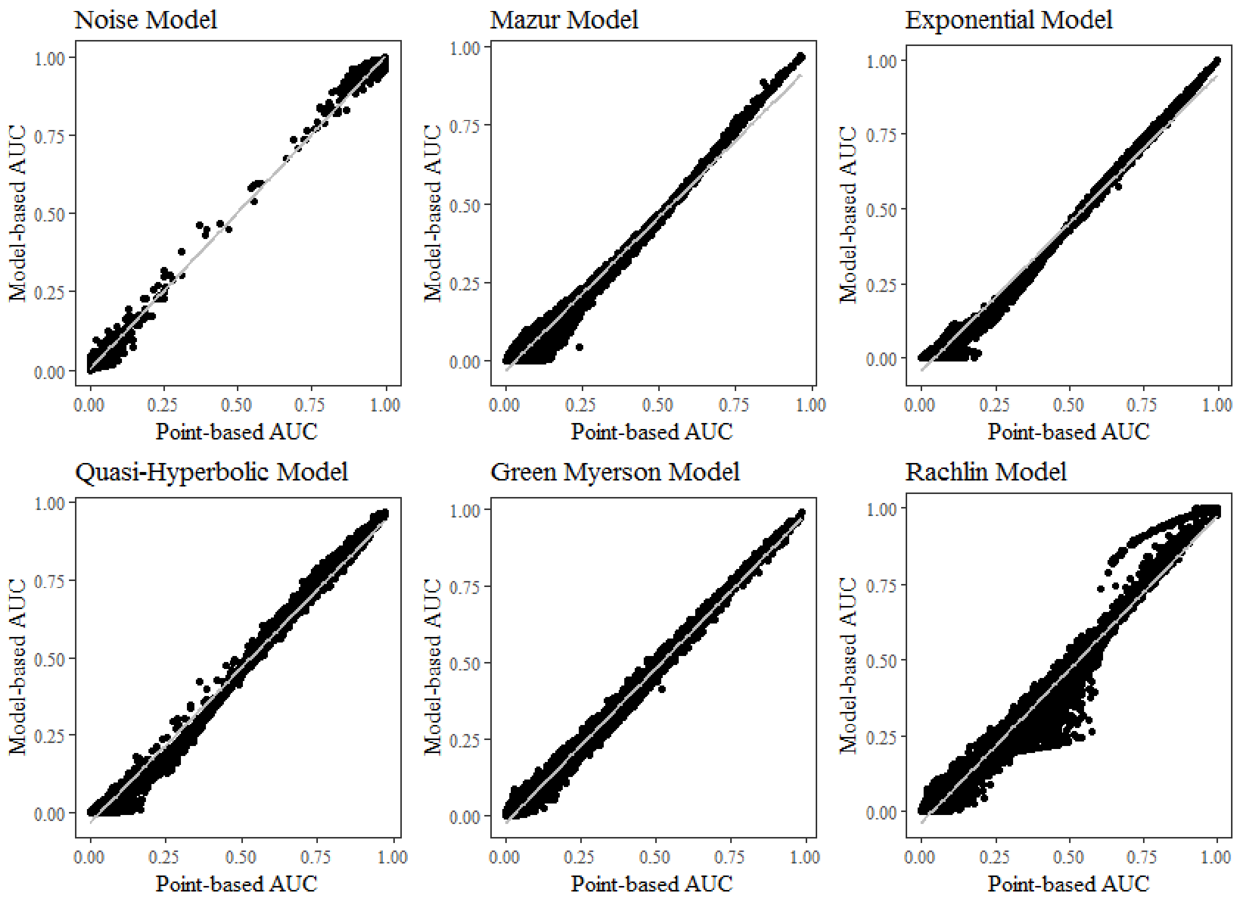

Figure 4. Comparison of Model-based AUC and Point-based AUC for Noise Models

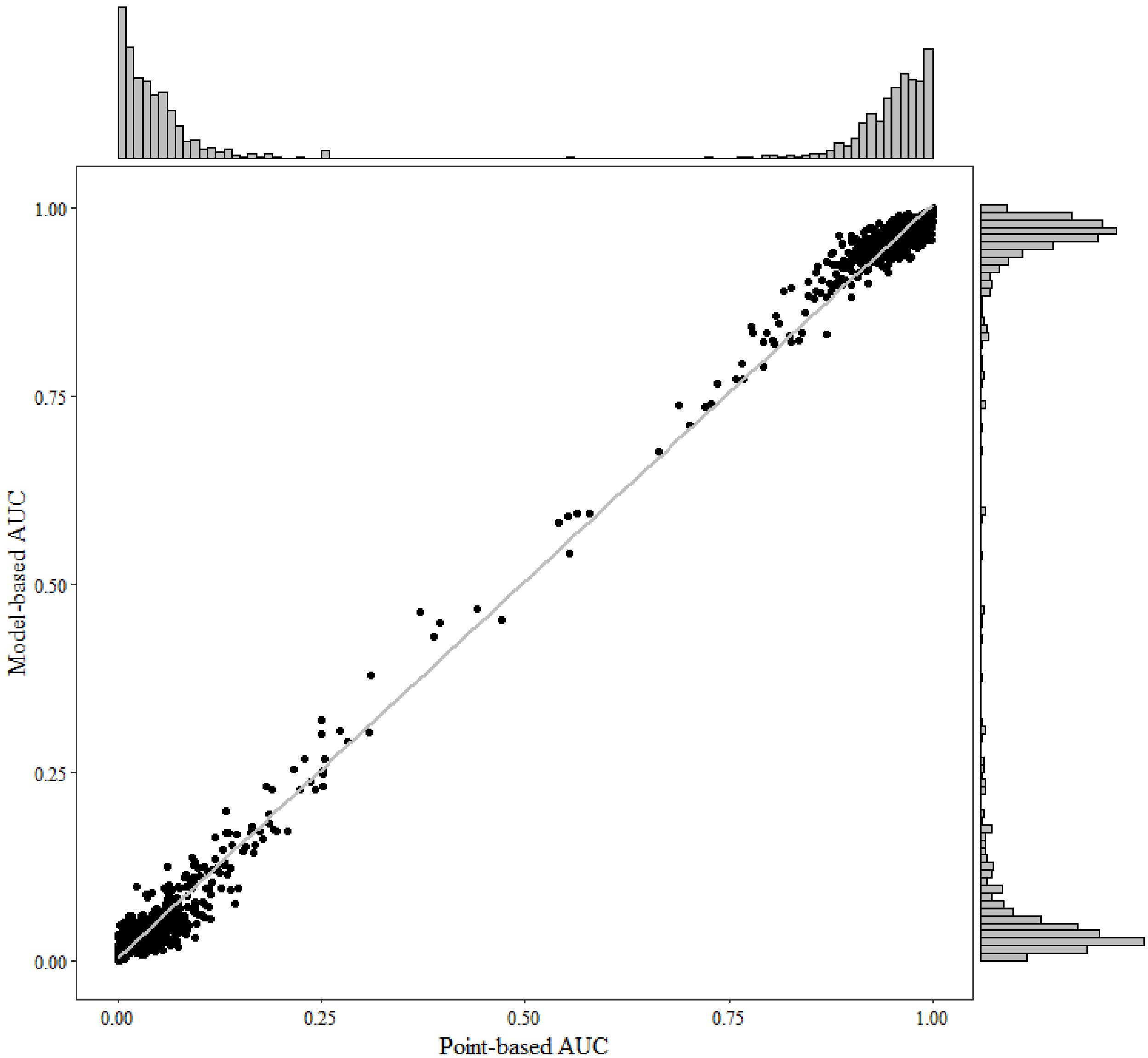

Figure 5. Participant number nineteen from Takahashi et al. (2008)

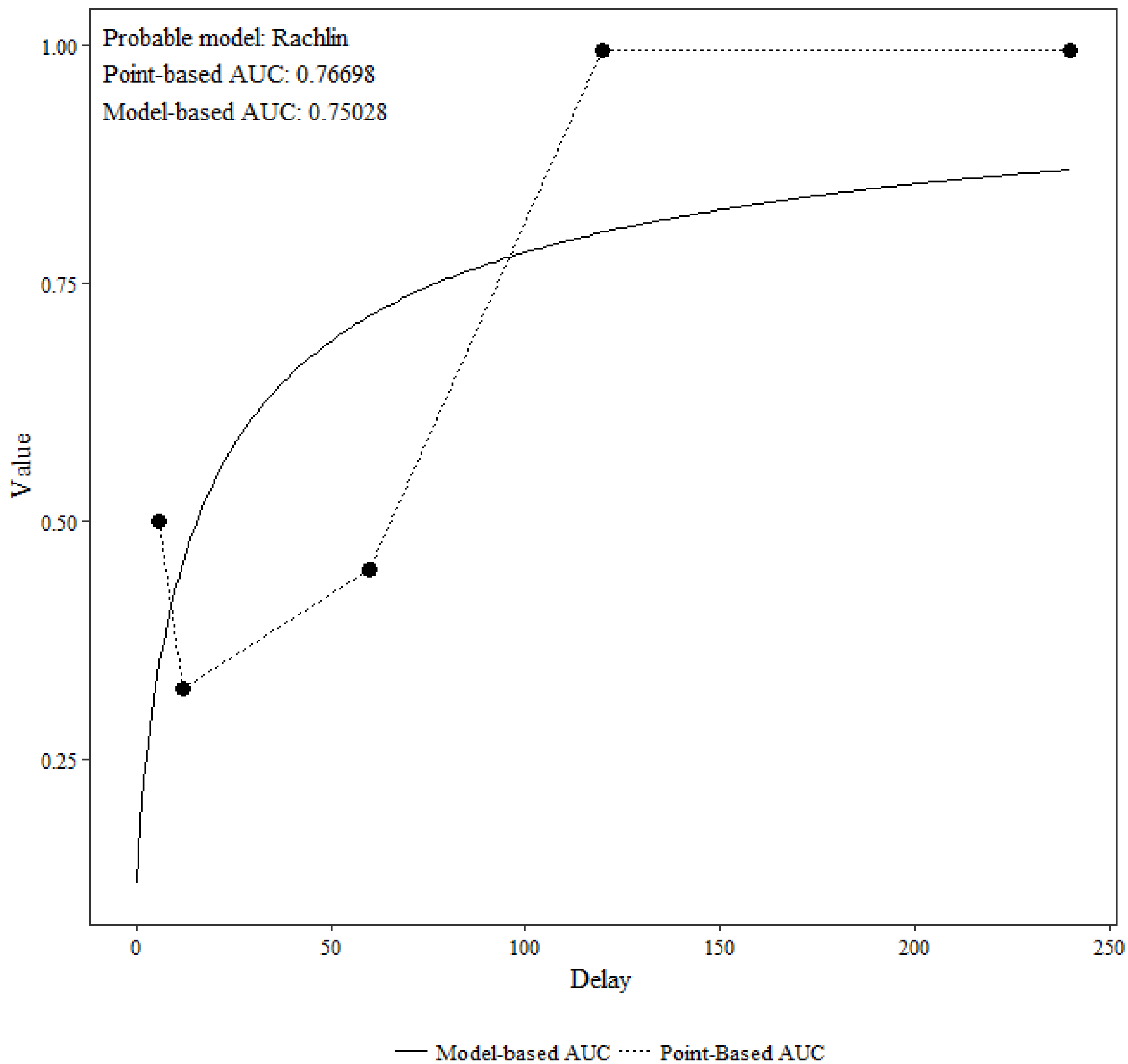

Figure Captions

Figure 1. The three plots displayed have the exact area beneath the interpolated data points, despite one being a decay function, one being a noise function and the final an interest function.

Figure 2. The upper plot in this figure illustrates the calculation of area for the empirical discounting function and the lower plot illustrates how area is calculated when a data series is best characterized by the Noise model.

Figure 3. The six plots within this figure illustrate the correlations between both types of AUC methods, with individual plots representing the results of both methods when a certain model was found to best characterize the data.

Figure 4. This plot is an enlarged and enhanced version of the Noise model plot in Figure 3. The Noise model is displayed here along with marginal histograms, which indicate the distribution of results for each metrics on the opposing axes.

Figure 5. This figure illustrates a positively increasing function from Takahashi, Masataka, Malaivijinond and Wongsiri (2008). This series yields an area-based metric, despite not characterizing discounting.