Empirical Test

Zark Wang

2023-04-13

1 Data

I select two datasets from previous studies: the first is from Ericson et al. (2015), containing 23,131 observations from 939 participants; the second is from Chávez et al. (2017), containing 34,515 choices from 1,284 participants. I term the first dataset as *Ericson* data, and the second dataset as *Chávez* data. Each dataset is used in more than one academic study.[1] Thus, readers interested in the method and the results of this paper can easily compare them with those of other papers.

The experiments corresponding to each dataset ask the participants to answer a series of intertemporal choice questions. In each question, the participants are required to select one option between an early small reward (denoted by SS) and late large reward (denoted by LL). I denote the magnitude and delay of reward by x_s and t_s for option SS, and by x_l and t_l for option LL, where $x_l > x_s$, $t_l > t_s$. I mainly focus on out-of-sample model performance. For each dataset, I draw the responses from 20% of the participants as the test sample, and set the rest as the train sample. To mitigate the overfitting issue, I implement a 10-fold cross-validation procedure on the train sample.

[1] For example, *Ericson* data is also used by Wulff and Bos (2018) for comparing different intermporal choice models. *Chávez* data is also used by Gershman and Bhui (2020) for testing their proposed attention-based theory.

2 Empirical Strategy

I test three types of intertemporal choice model: discounted utility model, trade-off model, and heuristic model.

The discounted utility model assumes that the decision maker tends to choose the option with greater discounted utility. Let the discounted utility for option j ($j \in \{l, s\}$) be $v_j = d(t_j)u(x_j)$, where $d(.)$ is the discounting function and $u(.)$ is the instantanoues utility function. Suppose the decision maker's perceived discounted utility for each option, denoted by is $\tilde{v_l}$ and $\tilde{v_s}$, is noisy. I set $\tilde{v_l} = v_l + \eta_l$, $\tilde{v_s} = v_s + \eta_s$. When η_l and η_s are independent and both follow $Gumble(0, \rho)$, where the scale parameter $\rho \in (0, \infty)$, then the probability that the decision maker chooses LL is

$$P\{\tilde{v_s} \leq \tilde{v_l}\} = \frac{1}{1 + \exp\{-\frac{1}{\rho}(v_l - v_s)\}}$$

The trade-off model (Scholten et al., 2014; Scholten and Read, 2010) assumes that when thinking of whether to choose LL, the decision maker makes a comparison between attributes (reward and time), rather than between options (LL and SS). If the benefit of receiving a larger reward exceeds the cost of waiting a longer time, then she will choose LL; otherwise, she will choose SS. I denote the benefit of receiving a larger reward by B, the cost of waiting longer by Q. The value of B can be simply represented by $u(x_l) - u(x_s)$. Following Scholten et al. (2014), I represent Q by

$$Q = \frac{\kappa}{\zeta_1} \ln \left(1 + \zeta_1 \left(\frac{w(t_l) - w(t_s)}{\zeta_2}\right)^{\zeta_2}\right)$$

where η_q is a noise term, and $w(t) = \ln(1 + \omega t)/\omega$. The parameter ω measures how much time is distorted in the decision maker's mind; κ measures the relative importance of reducing waiting time compared with increasing reward magnitude; ζ_1, ζ_2 jointly determine the curvature of changes in Q relative to $t_l - t_s$. Scholten et al. (2014) use ζ_1, ζ_2 to ensure that Q follow a S-shape curve in relation to $t_l - t_s$ and that the behavioral pattern can shift between sub-additivity and super-additivity.

I assume the decision maker's perception of B and Q, denoted by $\tilde{B}$ and $\tilde{Q}$, is noisy. Therefore, $\tilde{B} = B + \eta_B$, $\tilde{Q} = Q + \eta_Q$, where η_B and η_Q are independent noises. Again, assume both η_B and η_Q follow $Gumble(0, \rho)$, then the probability that the decision maker chooses

LL is

$$P\{\tilde{Q} \leq \tilde{B}\} = \frac{1}{1+\exp\{-\frac{1}{\rho}(B-Q)\}}$$

For the heuristic model, I employ a decision tree algorithm called XGBoost, which has been widely used in solving classfication problems (including predicting human risky choices). The intuition underlying XGBoost is that, the decision-maker uses a chain of if-then rules to make a choice, and repeats this process for several times, adding up the results of each iteration to make the final decision. To better fit the data, I extract features from each intertemporal choice question, following the methods in Read et al. (2013) and Ericson et al. (2015). Meanwhile, I tune the hyper-parameters of the algorithm via grid search. The features that I use to fit *Ericson* data are x_s, x_l, t_s, t_l, the absolute and relative differences between x_s and x_l, the absolute and relative differences between t_l and t_s, the interest rate of LL when SS is invested as principal. For *Chávez* data, given that $t_s = 0$, I omit t_s and the differences between t_s and t_l.

Along with the attention-adjusted discounting (under exponential and uniform initial attention allocations), I employ 8 other methods to draw the discounting factor in the discounted utility model, which are

1. exponential

$$d(t) = \delta^t$$

where the parameter is δ and $\delta \in (0, 1]$.

2. double exponential (Bos and McClure, 2013)

$$d(t) = \omega\delta_1^t + (1-\omega)\delta_2^t$$

where the parameters are δ_1, δ_2, ω, and $\delta_1, \delta_2 \in (0, 1]$.

3. hyperbolic

$$d(t) = \frac{1}{1+kt}$$

where the parameter is k.

4. dual-parameter hyperbolic (Loewenstein and Prelec, 1992)

$$d(t) = \frac{1}{(1 + kt)^a}$$

where the parameters are k, a.

5. magnitude-dependent hyperbolic (Gershman and Bhui, 2020)

$$d(t) = \frac{1}{1 + kt}, \quad k = \frac{1}{bu(x_t)}$$

where the parameter is b.

6. quasi-hyperbolic (Laibson, 1997)

$$d(t) = \mathbf{1}\{t = 0\} + \beta\delta^t \cdot \mathbf{1}\{t > 0\}$$

where the parameters are β, δ, and $\beta, \delta \in (0, 1]$.

7. quasi-hyperbolic plus fixed delay cost (Benhabib et al., 2010)

$$d(t) = \mathbf{1}\{t = 0\} + (\beta\delta^t - \frac{c}{u(x_t)}) \cdot \mathbf{1}\{t > 0\}$$

where the parameters are β, δ, c, and $\beta, \delta \in (0, 1]$.

8. homogeneous costly empathy (Noor and Takeoka, 2022)

$$d_t = \kappa_t u(x_t)^{\frac{1}{m}}$$

where κ_t is decreasing in t. I set $\kappa_t = \delta^t$, where the parameters are m, δ and $\delta \in (0, 1]$.

For the parameters in discounting functions, except for those explicitly marked as having a domain between 0 and 1, the domain of all other parameters is $(0, \infty)$. Besides, I employed 2 types of utility functions: (1) exponential utility (CARA), $u(x) = 1 - e^{-\gamma x}$; (2) power utility (CRRA), $u(x) = x^\gamma$. In both functions, the parameter is γ and $\gamma \in (0, \infty)$. Therefore, for the discounted utility model, there are 20 specific model settings to fit; for the trade-off model, there are 2 specific model settings to fit. In model fitting, if a parameter has a lower bound of 0, I set its lower bound to 0.001; if a parameter has a upper bound of infinity, I set its upper bound to 100. I use the maximum likelihood method to estimate the parameters, and apply L-BFGS-B method for optimization. As the solutions of L-BFGS-B are sensitive to initial points and often converge to local optima, I use the basin-hopping algorithm to achieve global optimization. [2]

3 Result

3.1 Results for *Ericson* data

Table 1

3.2 Results for *Chávez* data

3.3 Discussion

Reference

Benhabib, J., Bisin, A., and Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*(2), 205–223. https://doi.org/10.1016/j.geb.2009.11.003

Bos, W. van den, and McClure, S. M. (2013). Towards a General Model of Temporal Discounting: General Model of Time Discounting. *Journal of the Experimental Analysis of Behavior*, *99*(1), 58–73. https://doi.org/10.1002/jeab.6

Chávez, M. E., Villalobos, E., Baroja, J. L., and Bouzas, A. (2017). Hierarchical Bayesian modeling of intertemporal choice. *Judgment and Decision Making*, *12*(1), 19–28. https://doi.org/10.1017/S1930297500005210

Ericson, K. M., White, J. M., Laibson, D., and Cohen, J. D. (2015). Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better Than Delay Discounting Does. *Psychological Science*, *26*(6), 826–833. https://doi.org/10.1177/0956797615572232

[2] The basin-hopping algorithm runs the L-BFGS-B method for several times, and after each iteration, the solution will randomly drift to a new point. We set this new point as the initial point for the next iteration, and compare the new solution with the solution of the last iteration. The algorithm tends to accept the better solution of them, but there is still some probability of accepting an inferior solution. The magnitude of drifting is dependent on a stepwise parameter, which I set as 0.5; the probability of accepting the inferior solution is dependent on a temper parameter, which I set as 1.0. I also set the maximum number of iterations as 500.

Table 1: Cross-validation result for Ericson data

dstyle	**ustyle**	**params**	**mse**	**mae**	**log_loss**	**accuracy**
gbdt	gbdt	–	0.2988	0.2988	0.5812	0.7012
trade	power	[4.783 1.275 39.186 2.194 0.159 0.926]	0.2036	0.4088	0.5947	0.6937
hbmd	power	[6.751 0.149 0.182]	0.2042	0.4094	0.5978	0.6942
attention_uni	power	[1.298 0.176 0.326]	0.2088	0.4185	0.6071	0.6840
expo2	power	[0.939 0.786 0.599 0.016 0.014]	0.2094	0.4165	0.6087	0.6845
expo	power	[0.992 0.018 0.015]	0.2099	0.4205	0.6093	0.6827
hb	power	[0.012 0.027 0.024]	0.2099	0.4204	0.6097	0.6820
hce	power	[0.993 4.22 0.01 0.013]	0.2102	0.4211	0.6098	0.6805
hb2	power	[0.003 3.939 0.02 0.019]	0.2105	0.4194	0.6113	0.6824
quasihb_fc	power	[0.922 0.861 9.834 0.086 1.253]	0.2160	0.4362	0.6222	0.6725
attention	power	[0.996 7.781 0.341 1.582]	0.2157	0.4358	0.6232	0.6787
trade	cara	[3.615 1.63 36.216 5.63 0.4 1.205]	0.2248	0.4499	0.6415	0.6485
quasihb	power	[0.997 0.834 0.121 1.875]	0.2251	0.4438	0.6440	0.6571
attention_uni	cara	[0.793 2.886 0.395]	0.2277	0.4554	0.6476	0.6339
expo	cara	[0.725 2.516 0.544]	0.2279	0.4558	0.6480	0.6326
hb2	cara	[0.079 4.857 2.677 0.502]	0.2281	0.4558	0.6481	0.6324
expo2	cara	[0.445 0.604 0.419 2.853 0.508]	0.2280	0.4556	0.6481	0.6325
hce	cara	[0.725 98.968 2.537 0.545]	0.2279	0.4558	0.6481	0.6326
hb	cara	[0.317 3.258 0.419]	0.2282	0.4560	0.6484	0.6324
attention	cara	[0.885 1.159 3.091 0.481]	0.2281	0.4560	0.6487	0.6318
hbmd	cara	[3.75 2.583 0.383]	0.2281	0.4559	0.6489	0.6363
quasihb_fc	cara	[0.995 0.491 19.741 18.769 2.305]	0.2366	0.4780	0.6658	0.6325
quasihb	cara	[0.971 0.488 13.833 2.652]	0.2372	0.4790	0.6670	0.6330

Gershman, S. J., and Bhui, R. (2020). Rationally inattentive intertemporal choice. *Nature Communications*, *11*(1), 3365. https://doi.org/10.1038/s41467-020-16852-y

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478. https://doi.org/10.1162/003355397555253

Loewenstein, G., and Prelec, D. (1992). Anomalies in Intertemporal Choice: Evidence and an Interpretation. *The Quarterly Journal of Economics*, *107*(2), 573–597. https://doi.org/10.2307/2118482

Noor, J., and Takeoka, N. (2022). Optimal Discounting. *Econometrica*, *90*(2), 585–623. https://doi.org/10.3982/ECTA16050

Read, D., Frederick, S., and Scholten, M. (2013). DRIFT: An analysis of outcome framing in intertemporal choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 573–588. https://doi.org/10.1037/a0029177

Scholten, M., and Read, D. (2010). The Psychology of Intertemporal Tradeoffs. *Psychological Review*, *117*(3), 925–944. https://doi.org/10.1037/a0019619

Scholten, M., Read, D., and Sanborn, A. (2014). Weighing Outcomes *by* Time or *against* Time? Evaluation Rules in Intertemporal Choice. *Cognitive Science*, *38*(3), 399–438. https://doi.org/10.1111/cogs.12104

Wulff, D. U., and Bos, W. van den. (2018). Modeling Choices in Delay Discounting. *Psychological Science*, *29*(11), 1890–1894. https://doi.org/10.1177/0956797616664342