Empirical Test

Zijian Zark Wang

2023-04-14

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] Readers interested in the empirical 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 randomly 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.

The attention-adjusted discounting factor is dependent on the decision maker's initial attention allocation. I test the model under the assumptions that initial attention allocation is exponential and uniform (I term the former as "*attention*", the latter as "*attention_uni*"). Along with the attention-adjusted discounting, I also employ 8 other methods to draw the discounting factor, which are:

1. exponential, denoted by "*expo*"

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

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

2. double exponential, denoted by "*expo2*" (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, denoted by "*hb*"

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

where the parameter is k.

4. dual-parameter hyperbolic, denoted by "*hb2*" (Loewenstein and Prelec, 1992)

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

where the parameters are k, a.

5. magnitude-dependent hyperbolic, denoted by "*hbmd*" (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, denoted by "*quasihb*"(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, denoted by "*quasihb_fc*" (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, denoted by "*hce*" (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 shows the fitness of each model in cross-validation. The heuristic model has the highest accuracy rate, the lowest log loss, and the lowest MAE (mean absolute error). The trade-off model with power utility performs the lowest MSE (mean squared error). The magnitude-dependent hyperbolic (*hbmd*) model with power utility ranks third in all evaluation metrics. On the test sample, these three models also perform the best in MSE, MAE and accuracy rate (see Table 2). However, the heuristic model has the highest MSE in out-of-sample test, which is 10.8005, nearly 18 times as the *hbmd* and trade-off models with power utility. This may imply that the heuristic model can perform well in predicting whether the probability of choosing LL is greater than 0.5, but is not good at predicting the exact value of it.

Notably, the heuristic model has the most parameters that need to be fitted. Apart from the heuristic model, the trade-off model has the highest number of parameters, which is 6. The *hbmd* model has only 3 parameters to be fitted, and on the test sample, it performs very similarly with the trade-off model. It may be reasonable to conclude that magnitude-dependent hyperbolic discounting plus power utility can best describe the data, for it performs the best out-of-sample prediction accuracy (70.1%) with the lowest number of parameters. In addition, to test the correlation between the predictions of *hbmd* and trade-off models, I randomly draw 1,000 choice questions from *Ericson* data, and use choices predicted by trade-off model

[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.

with power utility as the labels, letting the other models to predict them. The accuracy rate for *hbmd* with power utility is the highest, which is 94.5%, even higher than the heuristic model (91.6%) and the trade-off model with exponential utility (85.4%).

Table 1: Cross-Validation Results on *Ericson* Data

model	**utility**	**MSE**	**MAE**	**log_loss**	**accuracy**
heuristic	–	0.2988	0.2988	0.5812	0.7012
tradeoff	power	0.2019	0.4035	0.5914	0.6949
hbmd	power	0.2042	0.4093	0.5976	0.6943
quasihb_fc	power	0.2074	0.4151	0.6036	0.6889
quasihb	power	0.2074	0.4148	0.6037	0.6889
expo2	power	0.2074	0.4157	0.6038	0.6858
attention_uni	power	0.2088	0.4184	0.6072	0.6835
hb2	power	0.2094	0.4186	0.6086	0.6846
hb	power	0.2097	0.4199	0.6092	0.6833
hce	power	0.2100	0.4203	0.6097	0.6833
expo	power	0.2100	0.4203	0.6097	0.6833
attention	power	0.2132	0.4278	0.6179	0.6782
tradeoff	cara	0.2160	0.4321	0.6232	0.6729
attention	cara	0.2276	0.4549	0.6473	0.6348
attention_uni	cara	0.2278	0.4553	0.6477	0.6338
hbmd	cara	0.2278	0.4554	0.6478	0.6341
expo2	cara	0.2280	0.4557	0.6480	0.6323
quasihb	cara	0.2280	0.4557	0.6480	0.6320

Table 1: Cross-Validation Results on *Ericson* Data

model	**utility**	**MSE**	**MAE**	**log_loss**	**accuracy**
quasihb_fc	cara	0.2280	0.4557	0.6480	0.6320
expo	cara	0.2280	0.4558	0.6481	0.6326
hce	cara	0.2280	0.4558	0.6481	0.6326
hb2	cara	0.2281	0.4558	0.6482	0.6325
hb	cara	0.2282	0.4561	0.6485	0.6323

Table 2: Out-of-Sample Test Results on *Ericson* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**	**pred_ll**
hbmd	power	0.2044	0.4074	0.5983	0.7008	0.2132
heuristic	–	0.2997	0.2997	10.8005	0.7003	0.2647
tradeoff	power	0.2020	0.4022	0.5919	0.6977	0.2106
hb	power	0.2084	0.4185	0.6067	0.6903	0.2451
quasihb_fc	power	0.2067	0.4167	0.6029	0.6899	0.2839
quasihb	power	0.2090	0.4001	0.6109	0.6897	0.2056
expo	power	0.2090	0.4158	0.6082	0.6814	0.2056
hce	power	0.2092	0.4161	0.6087	0.6814	0.2056
attention_uni	power	0.2108	0.4190	0.6114	0.6798	0.1491
attention	power	0.2159	0.4288	0.6242	0.6685	0.1246
tradeoff	cara	0.2192	0.4326	0.6298	0.6630	0.1078
attention	cara	0.2302	0.4570	0.6528	0.6318	0.0709

Table 2: Out-of-Sample Test Results on *Ericson* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**	**pred_ll**
attention_uni	cara	0.2303	0.4574	0.6531	0.6307	0.0764
hbmd	cara	0.2304	0.4575	0.6533	0.6301	0.0762
quasihb_fc	cara	0.2309	0.4582	0.6544	0.6281	0.0546
quasihb	cara	0.2309	0.4582	0.6544	0.6281	0.0546
hb	cara	0.2310	0.4585	0.6546	0.6275	0.0657
hb2	cara	0.2317	0.4554	0.6562	0.6248	0.0399
hce	cara	0.2310	0.4583	0.6545	0.6242	0.0410
expo	cara	0.2310	0.4582	0.6545	0.6242	0.0410
expo2	cara	0.2406	0.4484	0.6819	0.6203	0.0083
hb2	power	0.3832	0.3832	8.8541	0.6168	0.0000
expo2	power	0.3832	0.3832	9.3569	0.6168	0.0000

I now focus on attention-adjusted models. First, it is notable that *hbmd* can be viewed as a special case of the attention-adjusted models. [3] Magnitude-dependent hyperbolic discounting plus power utility is identical to attention-adjusted discounting under uniform initial attention allocation, plus logarithmic utility. Second, to explain why attention-adjusted models with exponential or power utility underperform some other models, see the last column in Table 2, which reports the ratio of LL choices in the predictions. The two attention-adjusted models underestimate the decision makers' tendency to choose option LL in *Ericson* data. Third, for why using logarithmic utility in attention-adjusted models outperforms power utility in the given data, the experimental design may explains.

25.3% of the small early rewards is less than \$3, 19.3% of the large late rewards is less than

[3]Note that under the assumption that the initial attention is uniformly allocated, the discounting factor in attention-adjusted model is $1/(1+kt)$, where $k = e^{-u(x_t)/\lambda}$. Suppose $u(x)/\lambda = \ln \beta + \gamma \ln x$, then we can get the magnitude-dependent hyperbolic model with power utility.

\$3.

In Chavez data, the smallest x_s is 11, the smallest x_l is 25.

3.2 Results for *Chávez* data

Table 3shows the fitness of each model in cross-validation.

Table 3: Cross-Validation Results on *Chávez* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**
heuristic	–	0.2211	0.2211	0.4818	0.7789
tradeoff	power	0.1571	0.3140	0.4840	0.7826
expo2	power	0.1574	0.3146	0.4844	0.7826
quasihb	power	0.1574	0.3144	0.4845	0.7826
quasihb_fc	power	0.1576	0.3146	0.4850	0.7816
hb2	power	0.1577	0.3149	0.4854	0.7804
attention	power	0.1579	0.3170	0.4860	0.7826
hbmd	power	0.1583	0.3174	0.4868	0.7731
attention_uni	power	0.1584	0.3178	0.4881	0.7826
hbmd	cara	0.1592	0.3191	0.4900	0.7731
hb	power	0.1597	0.3209	0.4903	0.7731
attention_uni	cara	0.1601	0.3203	0.4905	0.7641
expo	power	0.1604	0.3229	0.4922	0.7731
hce	power	0.1604	0.3229	0.4922	0.7731
attention	cara	0.1625	0.3263	0.4956	0.7465
tradeoff	cara	0.1633	0.3268	0.4971	0.7430

Table 3: Cross-Validation Results on *Chávez* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**
hb2	cara	0.1659	0.3326	0.5042	0.7439
hb	cara	0.1685	0.3372	0.5093	0.7287
quasihb	cara	0.1679	0.3411	0.5099	0.7450
expo	cara	0.1692	0.3421	0.5107	0.7253
hce	cara	0.1700	0.3412	0.5124	0.7168
quasihb_fc	cara	0.1701	0.3407	0.5128	0.7203
expo2	cara	0.1747	0.3487	0.5213	0.6825

Table 4 shows the out-of-sample performance of each model.

Table 4: Out-of-Sample Test Results on *Chávez* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**	**pred_ll**
attention	power	0.1628	0.3230	0.4982	0.7702	0.3299
attention_uni	power	0.1635	0.3233	0.5008	0.7702	0.3299
tradeoff	power	0.1633	0.3206	0.4986	0.7674	0.2926
heuristic	–	0.2326	0.2326	8.3825	0.7674	0.2926
hb	power	0.1641	0.3315	0.5014	0.7603	0.3296
quasihb_fc	power	0.1876	0.4173	0.5640	0.7603	0.3296
expo	power	0.1646	0.3302	0.5032	0.7603	0.3296
hbmd	cara	0.1645	0.3283	0.5033	0.7603	0.3296
hce	power	0.1651	0.3359	0.5039	0.7603	0.3296

Table 4: Out-of-Sample Test Results on *Chávez* Data

model	**utility**	**mse**	**mae**	**log_loss**	**accuracy**	**pred_ll**
quasihb	power	0.1779	0.3969	0.5413	0.7603	0.3296
hbmd	power	0.1630	0.3252	0.4987	0.7603	0.3296
hb2	power	0.1650	0.3217	0.5024	0.7570	0.2184
attention_uni	cara	0.1821	0.3844	0.5425	0.6710	0.1118
tradeoff	cara	0.2500	0.5000	0.6931	0.6678	0.4780
attention	cara	0.1802	0.3552	0.5360	0.6483	0.0000
quasihb	cara	0.2060	0.3433	0.5958	0.6483	0.0000
expo	cara	0.1996	0.3290	0.5981	0.6483	0.0000
hce	cara	0.2085	0.3226	0.6464	0.6483	0.0000
hb	cara	0.2198	0.3268	0.6642	0.6483	0.0000
expo2	cara	0.2245	0.3201	0.7294	0.6483	0.0000
hb2	cara	0.2504	0.3283	0.7914	0.6483	0.0000
expo2	power	0.3517	0.3517	12.6783	0.6483	0.0000
quasihb_fc	cara	0.2019	0.3423	0.5866	0.6483	0.0000

3.3 Parametrication

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

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