Comments to the Editor on "The Willingness to Pay for Diversification"*

Aleksi Pitkäjärvi[a,b] Matteo Vacca[c,d]

March 2024

Abstract

Mahmoud (2022) presents experimental evidence showing that people have an intrinsic preference for diversification that is driven by risk aversion and loss aversion, with risk averse and loss averse individuals being more willing to pay for diversification. We document numerous discrepancies between the results reported in the paper and the results calculated from the experimental data. In particular, the data show that risk averse individuals are actually less willing to pay for diversification, while the effect of loss aversion is non-monotonic and not statistically significant. We thus show that the data do not support the paper's claim of a preference for diversification driven by risk aversion and loss aversion.

Keywords: experimental economics, diversification, risk aversion, loss aversion

DOI: https://doi.org/10.1287/mnsc.2023.01581

*This paper grew out of an assignment in the *Management Science* Reproducibility Project (Fišar et al., 2024) and would not have been possible without the *Management Science* Policy for Data and Code Disclosure. The authors gratefully acknowledge financial support from the OP Group Research Foundation (grant nos. 20220057 and 20220089).

[a]Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
[b]Tinbergen Institute, 1082 MS Amsterdam, The Netherlands
[c]Aalto University School of Business, 02150 Espoo, Finland
[d]Hanken School of Economics, 65101 Vaasa, Finland
Contact: a.pitkajarvi@vu.nl, https://orcid.org/0009-0002-5598-7685 (AP); matteo.vacca@aalto.fi, https://orcid.org/0000-0002-1002-6680 (MV)

1 Introduction

Mahmoud (2022) presents the results of an experiment designed to determine whether people have an intrinsic preference for diversification. The experiment consisted of 13 questions where participants were asked to allocate an initial endowment across two identical fair dice with three sides colored in red and three sides colored in blue. The participants could choose to allocate the entire endowment to one of the two dice, in which case there was no cost involved. Alternatively, the participants could choose to diversify their endowment across the two dice, in which case they had to pay a cost that increased from zero to 10% of the initial endowment. The dice were then rolled, and if a die landed on the color red the participant won the amount of money they had allocated to that die. Mahmoud (2022) finds that participants were willing to pay for diversification, and shows that this willingness to pay is driven by risk aversion and loss aversion, with risk averse and loss averse participants being more willing to pay for diversification.

In this comment, we document numerous discrepancies between the results reported in Mahmoud (2022) and the results calculated from the experimental data included in the replication package provided on the *Management Science* website. By applying the empirical methods from the original paper to the experimental data, we replicate the analyses reported in Figure 1 and Tables 2–8 of Mahmoud (2022), and show that the analyses contain multiple errors. We then follow a growing literature in replication (e.g., Simonsohn, 2013; März, 2019; Guest, 2021) and demonstrate the materiality of the errors by rectifying them.

In particular, we focus on two errors that significantly affect the central results of the paper. First, we show that the paper incorrectly classifies "strongly risk averse" participants as "less risk averse," and "less risk averse" participants as "strongly risk averse." As a result, the risk aversion results in the paper are incorrect, with the data actually suggesting the counter-intuitive result that risk averse individuals are less willing to pay for diversification. Second, we show that the classification of participants as "not loss averse," "loss averse," or "strongly loss averse" in the paper

is inconsistent with the data. As a result, the loss aversion results in the paper are also incorrect, with the data actually showing that the relation between loss aversion and the willingness to pay for diversification is non-monotonic and not statistically significant. The experimental data thus provide no evidence of a clear relation between loss aversion and the willingness to pay for diversification, while the effect of risk aversion is the opposite of what is claimed in the paper.

The rest of the comment is structured as follows. We discuss the risk aversion results in Section 2, the loss aversion results in Section 3, and other results in Section 4. Section 5 concludes. We provide additional details and present the corrected versions of Figure 1 and Tables 2–8 of Mahmoud (2022) in the Internet Appendix.

2 Risk Aversion Results

Mahmoud (2022) elicits participants' risk preferences with a task where participants are given an initial endowment of 10 Swiss francs, and then have the choice of either keeping the entire amount, or allocating some of it to a 50/50 gamble with an expected value of zero. Participants who choose to keep the entire amount are classified as "strongly risk averse," while participants who choose to allocate some of it to the gamble are classified as "less risk averse." According to Mahmoud (2022), 199 participants are classified as strongly risk averse and 42 participants are classified as less risk averse.

This classification is inconsistent with the experimental data. The data show that 42 participants chose to keep the entire amount, and should thus be classified as strongly risk averse, while 198 participants allocated at least some of it to the gamble, and should thus be classified as less risk averse.[1] The data thus show that Mahmoud (2022) incorrectly classifies strongly risk averse participants as less risk averse, and less risk averse participants as strongly risk averse.[2]

[1] See Internet Appendix A for an explanation of the discrepancy in the number of participants.

[2] To confirm that the classification reported in Mahmoud (2022) is incorrect, in Internet Appendix A we verify that the classification is inconsistent with (1) the oTree code used to generate the experiment, (2) the monetary payoffs rewarded to the participants, (3) the information reported in the data analysis file included in the replication package, and (4) the results from earlier papers that have elicited risk preferences with similar experiments.

In Table B1, we replicate Tables 3–5 of Mahmoud (2022) using the experimental data. Panel B shows that the percentage of strongly risk averse participants who diversify is consistently smaller than the corresponding percentage for less risk averse participants. Panel A confirms this result by showing that the mean and median Herfindahl-Hirschman Index values of strongly risk averse participants are consistently higher than the values of less risk averse participants, which implies that strongly risk averse participants are less diversified than less risk averse participants. However, the p-values in Panels A and C show that the differences in diversification across the risk aversion groups are rarely statistically significant at the 5% level. In contrast with what is reported in Mahmoud (2022), the experimental data thus provide suggestive, but statistically weak, evidence in favor of the counter-intuitive result that strongly risk averse participants are actually less willing to pay for diversification than less risk averse participants.

3 Loss Aversion Results

Mahmoud (2022) elicits participants' loss aversion parameters using the same method as in Imas, Sadoff, and Samek (2017). To start, a risk aversion parameter for gains is identified from a series of gambles with positive outcomes. A risk aversion parameter for losses is then identified from a series of gambles with negative outcomes. Once these parameters have been pinned down, the loss aversion parameter λ can then be identified from a series of gambles with both positive and negative outcomes. In each of these stages, the generally observed pattern is that participants start by choosing a risky option, and then later switch to a sure outcome. Following Imas, Sadoff, and Samek (2017), participants who exhibit extreme risk attitudes by never switching, and participants who exhibit inconsistent risk attitudes by switching back and forth, are excluded from the analysis. According to Mahmoud (2022), this leaves a total of 207 participants: 30 participants with $\lambda \leq 1$ that are classified as "not loss averse," 115 participants with $\lambda \in (1, 2]$ that are classified as "loss averse," and 62 participants with

$\lambda \in (2, 4]$ that are classified as "strongly loss averse."

This classification is inconsistent with the experimental data. After applying the exclusion criteria described above, the data contain a total of 163 participants: 69 participants classified as not loss averse, 60 participants classified as loss averse, and 34 participants classified as strongly loss averse.[3]

In Table B2, we replicate Tables 6–8 of Mahmoud (2022) using the experimental data. Panel A shows that there is no consistent pattern in the mean Herfindahl-Hirschman Index values across the different loss aversion groups. Similarly, Panel C shows that there is no consistent pattern in the percentage of participants who diversify across the different loss aversion groups. Moreover, the p-values in Panels A and B show that the differences in diversification across the loss aversion groups are almost never statistically significant at the 5% level. In contrast with what is reported in Mahmoud (2022), the experimental data thus provide no evidence of a clear relation between loss aversion and the willingness to pay for diversification.

4 Other Results

In Table B3, we replicate Table 2 of Mahmoud (2022) using the experimental data. We are mostly able to replicate the original table, though some of the values change slightly because of three reasons we discuss in Internet Appendix A. The most notable change is that the 1% variable cost condition is no longer statistically significant at the 5% level.

In Figure B1, we replicate Figure 1 of Mahmoud (2022) using the experimental data. We are unable to replicate the original figure. For example, the original figure shows that column 0.1 has the smallest number of participants. By contrast, the experimental data show that column 0.1 has the second highest number of participants.

[3]The data also contain an additional 18 participants with $\lambda > 4$, and who thus do not belong to any of the loss aversion groups defined in Mahmoud (2022). For consistency with the group definitions reported in the paper, we also exclude these participants. Including them in the strongly loss averse group does not change the conclusions of Table B2.

5 Conclusion

We document numerous discrepancies between the results reported in Mahmoud (2022) and the results calculated from the experimental data. In particular, the data show that risk averse individuals are actually less willing to pay for diversification, while the effect of loss aversion is non-monotonic and not statistically significant. We thus show that the data do not support the paper's claim of a preference for diversification driven by risk aversion and loss aversion.

References

Fišar, Miloš, Ben Greiner, Christoph Huber, Elena Katok, Ali Ozkes, and the *Management Science* Reproducibility Collaboration (2024). "Reproducibility in *Management Science*." *Management Science* 70.3, pp. 1343–1356.

Guest, Paul (2021). "Risk Management in Financial Institutions: A Replication." *Journal of Finance* 76.5, pp. 2689–2707.

Imas, Alex, Sally Sadoff, and Anya Samek (2017). "Do People Anticipate Loss Aversion?" *Management Science* 63.5, pp. 1271–1284.

Mahmoud, Ola (2022). "The Willingness to Pay for Diversification." *Management Science* 68.8, pp. 6235–6249.

März, Oliver (2019). "Comment on "Naiveté, Projection Bias, and Habit Formation in Gym Attendance"." *Management Science* 65.5, pp. 2442–2443.

Simonsohn, Uri (2013). "Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone." *Psychological Science* 24.10, pp. 1875–1888.

Internet Appendix to Comments to the Editor on "The Willingness to Pay for Diversification"

Aleksi Pitkäjärvi Matteo Vacca

March 2024

A Discrepancies in the Data Analysis

The Replication Package. We utilize three files from the replication package provided on the *Management Science* website: the raw data file, which contains the experimental data; the data analysis file, which generates the tables and figure in Mahmoud (2022); and the models.py file, which is part of the oTree code used to generate the experiment.

Number of Participants. Mahmoud (2022) states that 241 participants completed the experiment. However, the experimental data show that one participant did not complete the risk aversion and loss aversion tasks, and another participant did not complete the loss aversion task. Mahmoud (2022) thus incorrectly allocates one participant to a risk aversion group even though they did not complete the risk aversion task.

Risk Aversion Classification. The risk aversion classification reported in Mahmoud (2022) is inconsistent with (1) the oTree code used to generate the experiment, (2) the monetary payoffs rewarded to the participants, (3) the information reported in the data analysis file, and (4) the results from earlier papers that have elicited risk preferences with similar experiments.

(1) In the experimental data, the part of the initial endowment that a participant chooses to keep is stored in the column "player.question2a," while the part that they allocate to the gamble is stored in the column "player.question2b." By inspecting lines 82–83 of the models.py file, we confirm that the value stored in the "player.question2a" column is indeed the amount the participant chooses to "Keep (in CHF)," while the value stored in the "player.question2b" column is the amount the participant chooses to "Play (in CHF)." Classifying participants as "strongly risk averse" or "less risk averse" based on the values stored in these columns confirms that the classification reported in Mahmoud (2022) is incorrect.

(2) Line 117 of the models.py file shows that the monetary payoff a participant receives is given by the amount of the endowment they choose to keep ("self.question2a") plus twice the amount they invest in the gamble ("self.question2b") if the outcome of

the gamble is favorable. This reconfirms the fact that the amount a participant chooses to keep is stored in the column "player.question2a," while the part that they allocate to the gamble is stored in the column "player.question2b." Classifying participants as "strongly risk averse" or "less risk averse" based on the values stored in these columns confirms that the classification reported in Mahmoud (2022) is incorrect.

(3) The data analysis file includes "Keep" and "Play" columns in sheets "Table_3_(strongly_risk_averse)" and "Table_3_(less_risk_averse)." Participants in the "less risk averse" sheet all have "Keep" values of 10 and "Play" values of 0, which indicates that they should be classified as strongly risk averse. Participants in the "strongly risk averse" sheet all have "Keep" values smaller than 10 and "Play" values greater than 0, which indicates that they should be classified as less risk averse. The data analysis file thus confirms that the classification reported in Mahmoud (2022) is incorrect.

(4) Following Simonsohn (2013), we use the results from earlier experiments to evaluate the classification reported in Mahmoud (2022). In particular, Gal (2006) and Ertac and Gurdal (2012) elicit participants' risk preferences with experiments that are similar to the one used by Mahmoud (2022).[1] Gal (2006) finds that 23% of participants keep the entire endowment, which is similar to the 18% of participants who keep the entire endowment in the experimental data, but inconsistent with the 82% of participants reported in Mahmoud (2022). Similarly, Ertac and Gurdal (2012) find that participants allocate an average of 51% of the endowment to the gamble. This is similar to the average allocation of 55% in the experimental data, but inconsistent with the maximum of 18% reported in Mahmoud (2022).[2] These earlier papers are thus consistent with the experimental data, but inconsistent with the classification reported in Mahmoud (2022).

[1]Specifically, we refer to the allocation task in Gal (2006) and to the experiment with $p = 2$ in Ertac and Gurdal (2012).

[2]Mahmoud (2022) does not report this value directly, but if 82% of participants keep the entire endowment, then the average allocation to the gamble across all participants cannot be larger than 18%.

Loss Aversion Classification. Following Imas, Sadoff, and Samek (2017), we calculate the risk aversion parameters used to identify the loss aversion parameter by setting the indifference point between the risky option and the sure outcome as the midpoint between the payoff at which a participant switches and the preceding payoff. For more details, see the Supplemental Materials of Imas, Sadoff, and Samek (2017).

Table 2. The data analysis file provides three reasons why Table B3 is slightly different from Table 2 in Mahmoud (2022). First, even though 241 participants completed this part of the experiment, most of the p-values are incorrectly calculated assuming a sample size of 242 or 243. Second, the calculation of the p-values incorrectly omits one of the participants. Third, some of the participants' responses in the data analysis file do not match their responses in the experimental data.

Table 5. The chi-squared test described in Table 5 of Mahmoud (2022) is inconsistent with the chi-squared test performed in the data analysis file. The null hypothesis stated in the paper is that "each risk attitude has the same proportion of diversifiers." We would thus expect to see a 2×2 test of the proportions of diversifiers and non-diversifiers across the risk aversion groups. Instead, the test performed in the data analysis file is a 3×2 test of the proportions of 50/50 diversifiers, other diversifiers, and non-diversifiers across the risk aversion groups. The p-values in the data analysis file are consistent with the p-values reported in the paper, implying that the p-values in the paper are indeed from this 3×2 test. In Panel C of Table B1, we report p-values from a 3×2 test of 50/50 diversifiers, other diversifiers, and non-diversifiers across the risk aversion groups.

Table 7. The chi-squared test described in Table 7 of Mahmoud (2022) is inconsistent with the chi-squared test performed in the data analysis file. The null hypothesis stated in the paper is that "each loss aversion group (loss averse, not loss averse, strongly loss averse) has the same proportion of observations." We would thus expect to see a 2×3 test of the proportions of diversifiers and non-diversifiers across the three loss

aversion groups. Instead, the test performed in the data analysis file is a 3×2 test of the proportions of 50/50 diversifiers, other diversifiers, and non-diversifiers across only two loss aversion groups: loss averse and not loss averse. The strongly loss averse participants are thus not treated as a separate group. The p-values in the data analysis file are consistent with the p-values reported in the paper, implying that the p-values in the paper are indeed from this incorrectly specified test. In Panel B of Table B2, we report p-values from a 3×3 test of 50/50 diversifiers, other diversifiers, and non-diversifiers across all three of the loss aversion groups.[3]

[3]The chi-squared tests in the data analysis file include 241 participants. This is difficult to reconcile with the claim from Mahmoud (2022) that only 207 participants are included in the loss aversion groups.

B Replication of Tables and Figure

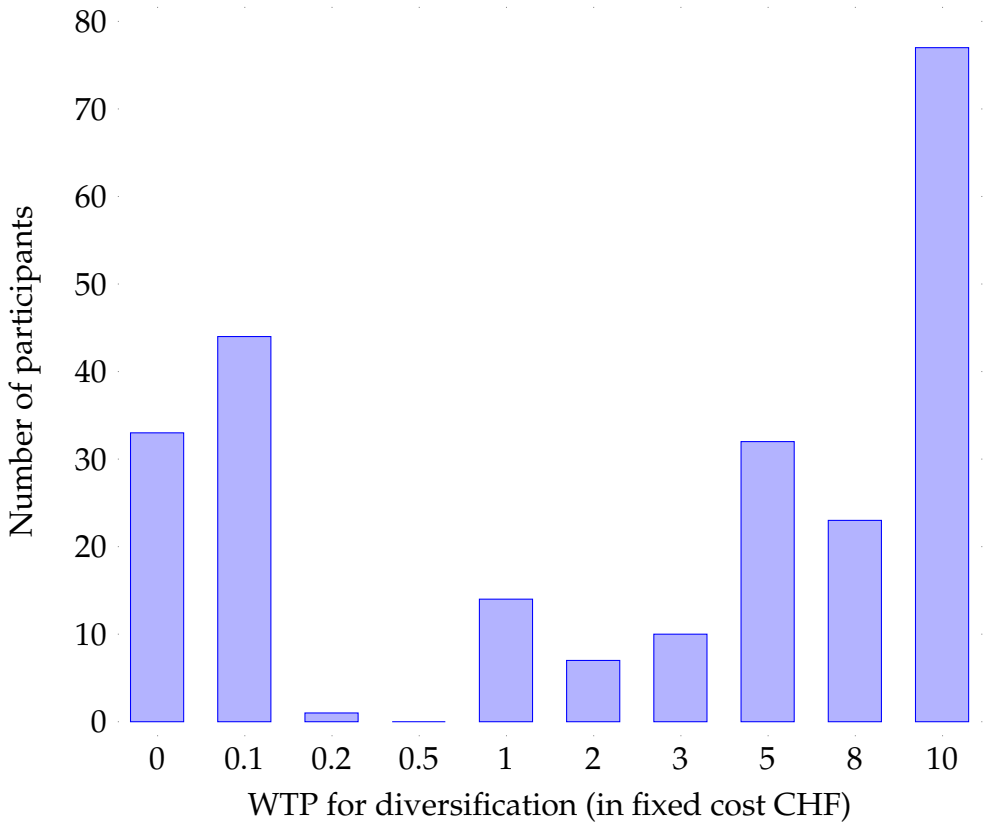

Figure B1: Distribution of Individual Willingness to Pay for Diversification.
This figure replicates Figure 1 of Mahmoud (2022) using the experimental data. For more details on the contents of the figure, see Mahmoud (2022).

Table B1: Risk Aversion Results.

This table replicates Tables 3–5 of Mahmoud (2022) using the experimental data. Panel A replicates Table 3, Panel B replicates Table 4, and Panel C replicates Table 5. For more details on the contents of the tables, see Mahmoud (2022).

Panel A: Replication of Table 3 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
Strongly risk averse													
Mean	0.17	0.48	0.50	0.47	0.55	0.55	0.54	0.63	0.72	0.74	0.65	0.70	0.75
Median	0.00	0.02	0.46	0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Less risk averse													
Mean	0.15	0.31	0.30	0.31	0.36	0.39	0.44	0.58	0.67	0.73	0.46	0.56	0.71
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.04	1.00	1.00	1.00	0.20	0.90	1.00
p-value	(0.64)	(0.10)	(0.08)	(0.11)	(0.06)	(0.14)	(0.44)	(0.58)	(0.42)	(0.79)	(0.04)	(0.12)	(0.71)
Panel B: Replication of Table 4 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
Strongly risk averse	85.71	54.76	52.38	54.76	47.62	47.62	47.62	38.10	28.57	26.19	38.10	33.33	30.95
	(0.00)	(0.32)	(0.44)	(0.32)	(0.68)	(0.68)	(0.68)	(0.96)	(1.00)	(1.00)	(0.96)	(0.99)	(1.00)
Less risk averse	86.36	71.72	72.22	72.22	66.67	63.13	58.08	43.94	34.85	28.79	58.08	50.00	34.34
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.96)	(1.00)	(1.00)	(0.01)	(0.53)	(1.00)
Panel C: Replication of Table 5 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
p-value	(0.52)	(0.07)	(0.01)	(0.05)	(0.05)	(0.12)	(0.16)	(0.68)	(0.69)	(0.81)	(0.06)	(0.07)	(0.71)

Table B2: Loss Aversion Results.

This table replicates Tables 6–8 of Mahmoud (2022) using the experimental data. Panel A replicates Table 6, Panel B replicates Table 7, and Panel C replicates Table 8. For more details on the contents of the tables, see Mahmoud (2022).

Panel A: Replication of Table 6 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
Not loss averse													
Mean	0.15	0.36	0.37	0.37	0.49	0.54	0.56	0.69	0.79	0.84	0.54	0.69	0.80
Loss averse													
Mean	0.19	0.41	0.41	0.41	0.41	0.44	0.46	0.60	0.72	0.80	0.50	0.55	0.71
p-value	(0.60)	(0.72)	(0.67)	(0.68)	(0.22)	(0.12)	(0.15)	(0.11)	(0.11)	(0.18)	(0.33)	(0.06)	(0.07)
Strongly loss averse													
Mean	0.21	0.40	0.40	0.40	0.36	0.37	0.43	0.56	0.59	0.64	0.45	0.56	0.66
p-value	(0.72)	(0.38)	(0.41)	(0.39)	(0.35)	(0.35)	(0.46)	(0.40)	(0.11)	(0.04)	(0.30)	(0.45)	(0.32)
Panel B: Replication of Table 7 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
p-value	(0.54)	(0.97)	(0.68)	(0.85)	(0.44)	(0.12)	(0.48)	(0.42)	(0.11)	(0.04)	(0.87)	(0.50)	(0.36)
Panel C: Replication of Table 8 of Mahmoud (2022)													
Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0	1%	5%	10%
Not loss averse	85.51	65.22	63.77	65.22	52.17	47.83	44.93	31.88	21.74	15.94	49.28	36.23	23.19
	(0.00)	(0.01)	(0.01)	(0.01)	(0.40)	(0.68)	(0.83)	(1.00)	(1.00)	(1.00)	(0.60)	(0.99)	(1.00)
Loss averse	81.67	60.00	60.00	60.00	60.00	56.67	55.00	41.67	30.00	21.67	51.67	48.33	35.00
	(0.00)	(0.08)	(0.08)	(0.08)	(0.08)	(0.18)	(0.26)	(0.92)	(1.00)	(1.00)	(0.45)	(0.65)	(0.99)
Strongly loss averse	82.35	64.71	67.65	67.65	67.65	67.65	58.82	47.06	44.12	38.24	55.88	47.06	38.24
	(0.00)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.20)	(0.70)	(0.80)	(0.94)	(0.30)	(0.70)	(0.94)

Table B3: Willingness to Pay for Diversification.

This table replicates Table 2 of Mahmoud (2022) using the experimental data. For more details on the contents of the table, see Mahmoud (2022).

Cost	0.0	0.1	0.2	0.5	1.0	2.0	3.0	5.0	8.0	10.0
Div.	86.31	68.88	68.88	69.29	63.49	60.58	56.43	43.15	34.02	28.63
Conc.	13.69	31.12	31.12	30.71	36.51	39.42	43.57	56.85	65.98	71.37
p-val.	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.99)	(1.00)	(1.00)
Cost	1%	5%	10%							
Div.	54.77	47.30	34.02							
Conc.	45.23	52.70	65.98							
p-val.	(0.08)	(0.82)	(1.00)							

References

Ertac, Seda and Mehmet Gurdal (2012). "Deciding to Decide: Gender, Leadership and Risk-Taking in Groups." *Journal of Economic Behavior & Organization* 83.1, pp. 24–30.

Gal, David (2006). "A Psychological Law of Inertia and the Illusion of Loss Aversion." *Judgment and Decision Making* 1.1, pp. 23–32.

Imas, Alex, Sally Sadoff, and Anya Samek (2017). "Do People Anticipate Loss Aversion?" *Management Science* 63.5, pp. 1271–1284.

Mahmoud, Ola (2022). "The Willingness to Pay for Diversification." *Management Science* 68.8, pp. 6235–6249.

Simonsohn, Uri (2013). "Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone." *Psychological Science* 24.10, pp. 1875–1888.