Journal of Behavioral and Experimental Economics 67 (2017) 14–19

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

# To choose or not to choose: An experiment in hedging strategies and risk preferences

Matthew J. Butler[a], James H. Cardon[b,1,*], Mark H. Showalter[b]

[a] Center for Population Science & Discovery, University of Arizona, Tucson, Arizona 85721, USA
[b] Department of Economics, Brigham Young University, Provo, UT, 84602, USA

ARTICLE INFO

*Article history:*
Received 5 July 2016
Revised 30 January 2017
Accepted 31 January 2017
Available online 3 February 2017

*Keywords:*
Risk preference
Risk aversion
Insurance
Testing
Hedging
Fatigue

ABSTRACT

We conduct a small field experiment on hedging and risk-taking in a relatively high-stakes environment. Students in an economics principles class at a large private university are allowed to choose multiple answers on multiple choice exams with a corresponding reduction in the maximum attainable points. In addition to the usual grade pressure for such classes, this class is also a key component for determining entrance into a coveted limited enrollment business major. We control for question difficulty using a second section of the same course taught by the same instructor. We also build a simple model to explain the findings. We find hedging propensities increase with the question number, suggesting that fatigue plays a role in students' decisions. We also find that student quality (as measured by ACT scores) significantly reduces hedging on the Final exam but not on the Midterm. This may be evidence of learning among more able students.



## 1. Introduction

Risk averse agents can often reduce risk by buying insurance against potential losses, choosing costly risk- or loss-reducing expenditures, or by hedging to offset risky assets. An interesting venue to explore choice under uncertainty is a school setting. Students become accustomed to making risky choices while taking multiple choice examinations. We examine the risk-taking behavior of students in taking multiple choice exams by offering them a chance to hedge by selecting more than one answer. We control for student quality (American College Test (ACT), GPA, class), and for question difficulty by using as a control information from another section of the same course. In addition, we examine the effect of wealth on hedging behavior by including results of prior exams in the regressions.

Students first experience with hedging may come from national standardized exams. The Scholastic Aptitude Test (SAT) until 2016 included a 0.25 point penalty for each incorrect answer. Each correct answer is worth one point, each question has 5 possible answers, so a correct guess raises the index by one point, while an incorrect guess lowers it by 0.25. It is therefore possible to do better, on average, by skipping the question than by guessing. The typical advice given was that "guessing" is better than skipping if you can eliminate at least one of the possible answers. Otherwise, skipping the question–a form of hedging–is the safe option, allowing the student to avoid a potential loss.[1] Such a rule of thumb ignores risk aversion and treats each question in isolation without regard to the wealth effects from past exams or background risk from other questions or from future exams.

There are several factors that might influence hedging behavior on exams. Foremost would be the difficulty of the question. If students are certain of the right answer on a particular question, hedging would lower the expected score and therefore would not be a good strategy. But if a student were unsure, then a hedging strategy could increase the expected utility of a given choice.

Another factor would be the background risk associated with the choice. Harrison et al. (2007) show that risk-aversion increases in a field experiment when there is additional risk associated with a given choice beyond just the immediate decision. In our context, the course grade is ultimately what is likely being optimized by the student, and the strategy on how to answer a particular question is embedded in this larger optimization problem. But this background risk concerning the course grade is likely to affect the level of risk aversion and therefore the decision behavior

* Corresponding author.
*E-mail addresses:* matthewbutler@email.arizona.edu (M.J. Butler), cardon@byu.edu (J.H. Cardon), showalter@byu.edu (M.H. Showalter).

1 We thank Scott Condie, Rick Evans and two anonymous referees for helpful suggestions.

1 Graduate Record Examination (GRE) subject exams also have a penalty. ACT and the GRE general exam do not.

http://dx.doi.org/10.1016/j.socec.2017.01.010
2214-8043/

on a given question. Additionally, this background risk might change over the semester as more information about the students performance is revealed.[2]

We might also expect demographics to influence hedging. For example, there is a growing body of evidence that suggests women are more risk averse than men (eg. Byrnes et al., 1999; Cadsby and Maynes, 2005; Eckel et al., 2008). If so, we would expect women to hedge more often than men, although there is some contrary evidence on this point (Harrison et al., 2007)). We might also expect native ability and previous performance to have an effect. In the context of the literature, the latter point could be considered a wealth effect; we might expect students with higher scores on previous exams to be less risk averse if preferences exhibit decreasing absolute risk aversion (DARA).

There is relatively little evidence of how individuals behave in the face of risk in an educational setting. Walker and Thompson (2001) presented a testing strategy similar to the one proposed here. The authors allowed their students in one section to select two choices from a set of four possible answers to hedge against uncertainty. Walker and Thompson (2001) argued that risk aversion would lead the "treatment" class to hedge to the point that their average score would be lower than the "control". The authors found a non-statistically significant negative impact of "treatment" – failing to reject the null hypothesis that students are risk neutral. Espinosa and Gardeazabal (2013) presented two alternative scoring methods for multiple-choice tests; the first method penalizes students for incorrect answers and the second rewards students for omitting answers. Both scoring methods are meant to create an incentive for students not to guess. The authors suggested a normalization for both scoring methods that equalizes their average expected average outcome. They then conducted a field experiment in which they varied the type of scoring method (with and without normalization) applied to students within the same section. They found that students' scores do not differ with the normalized scoring methods but students scores do differ without the normalization; the authors concluded that students, especially women, are risk averse when taking multiple-choice exams.

Both Walker and Thompson (2001) and Espinosa and Gardeazabal (2013) treat the test score as the dependent variable in their analysis. This ignores variation present in each multiple choice question on any given exam. Suppose an exam has 50 multiple choice questions. Using the test score as the unit of analysis provides one experiment to analyze risk aversion in the educational setting, when in fact there are 50; that is, each question has uncertainty that is captured by the subjective probability generated by the control section. In our empirical work, we use the negative binomial regression to model the count of the total number of hedged questions for each student. However, we also model the probability of hedging on *each* question using logit and fixed effects logit models. In some sense, the fixed effect logit and count data models are complementary: fixed effects models difference out all variables that do not vary across questions, including demographics and past exam performances, while count data models ignore all question-specific information. The logit model lies in between and allows us to measure the effects of student- and question-specific variables, while controlling for fixed effects.

In both logit specifications, we take individual questions as the unit of observation and control for average question difficulty using data from the control section, in which students could not hedge. Therefore, for the two exams on which hedging was allowed we have either 60 or 90 observations for each student and can use panel data estimators to control for unobserved individual effects. For both examinations we find evidence that students are more likely to hedge on harder questions. In addition, hedging propensity increases in the discrimination index, a commonly-used measure of question quality. We find some evidence of a negative wealth effect on the propensity to hedge, so that students have performed well on the most recent exam are less likely to hedge. Men are less likely to hedge than women, though the estimated coefficients are not statistically significant. Student quality, as measured by ACT score, has a significant, negative effect on hedging propensities, but only for the Final exam. We also find a significant, positive effect of question number on the propensity to hedge, suggesting that fatigue plays a role in the decision to hedge.

[2] There is also evidence that individuals tend to become more risk averse as the importance of the transaction increases (See Kachelmeier and Shehata, 1992 and Holt and Laury, 2002). For an extensive review of laboratory experiments testing for risk aversion see Harrison and Rutström (2008).

## 2. A simple model of hedging on a true/false exam

To develop the theory we simplify to the case in which the test consists of $N$ True/False questions. The score from each questions, $s_i$, then, is an independent draw from a Bernoulli distribution with parameter $p_i = Pr(s_i = 1)$. Without hedging, the optimal strategy is to choose the option that seems more likely to be correct. The probability is formed as the question is considered and while doing any computations necessary to arrive at a best guess of the correct answer. Each $p_i$, then, is the student's unique subjective probability assessment that her selected answer is correct. If the student has literally no information about the question, then $p_i = 0.5$. The total score for the exam is therefore $S = \sum_{i=1}^{N} s_i$, with $E(S) = \sum_{i=1}^{N} p_i$ and $Var(S) = \sum_{i=1}^{N} p_i(1 - p_i)$.

We assume the student has preferences over performance for the class, not just the current exam. Let $y$ represent fixed wealth from previous exams.[3] The score on each exam question for the current exam is $s_i$, and the score for the exam is $S = \sum_{i=1}^{N} s_i$. Let $f_i(s_i)$ be the marginal pdf and $f(s_1, \ldots, s_N)$ be the joint pdf. If the questions are identical ($p_i = p$ for all $i$), then $S \sim Bin(N, p)$. Define "wealth" $W = y + S$ and assume strictly concave utility function $U(W)$. If we allow the student to hedge on some of the questions, then the student will choose the number of hedged questions $k$ to maximize expected utility

$$E(U(W|k)) = \sum_{x=0}^{N-k} \binom{N-k}{x} p^x (1-p)^{N-k-x} U\left(y + x + \frac{k}{2}\right),$$

where $x$ is the number of correct answers. The student receives $\frac{1}{2}$ for each hedged answer and one for each correctly-guessed answer.

*Special Case:* $N = 1$

Suppose the test is a single T/F question. Expected utility for Guessing is $EU = pU(y+1) + (1-p)U(y)$ and for Hedging is $EU^H = U(y + \frac{1}{2})$. The student will Hedge if and only if

$$\frac{U(y+\frac{1}{2}) - U(y)}{U(y+1) - U(y)} \geq p^*. \tag{1}$$

Strict concavity of $U(W)$ implies $U(y + \frac{1}{2}) > \frac{U(y+1)+U(y)}{2}$, which implies that $p^* > \frac{1}{2}$ if the student is risk averse. Therefore, there exists a $p > \frac{1}{2}$ such that the student will Hedge.

[3] Andersen et al. (2012) gives a nice discussion of the issues surrounding the treatment of wealth and income in an expected utility framework. We choose to model the process as additive for simplicity and clarity.