Revealed Preference and Mean-Variance Investing: An Experiment

BENJAMIN L. BUSHONG

SENIOR THESIS · JUNE 2007

Revealed Preference and Mean-Variance Investing: An Experiment

BENJAMIN L. BUSHONG

ABSTRACT

THIS PAPER IS AN EXAMINATION OF PAST INFORMATION CONCERNING UTILITY FUNCTIONS AND MEAN-VARIANCE INVESTMENT THEORY. THE PURPOSE OF THIS INVESTIGATION IS TO ESTABLISH THE APPROPRIATE LIMITATIONS ON THE REALM OF MEAN-VARIANCE ANALYSIS. I SHOW, THROUGH EXPERIMENT USING REVEALED PREFERENCE TOOLS, THAT MEAN-VARIANCE IS EMPIRICALLY TESTABLE. THROUGH THIIS EXPERIMENT, I ALSO SHOW THAT STUDENTS HAVE CLEARLY DEFINED MEAN-VARIANCE PREFERENCES WHICH SATISFY THE AXIOMS OF REVEALED PREFERENCES. IN FACT, CHOICES IN MEAN-VARIANCE SPACE ARE HIGHLY EFFICIENT. USING NUMERACY SCORES AS CONTROL, I ALSO VERIFY THAT NUMERACY DOES NOT HAVE A SIGNIFICANT EFFECT ON THE AVERAGE NUMBER OF REVEALED PREFERENCE VIOLATIONS. FINALLY, I OFFER SUGGESTIONS INTO FURTHER RESEARCH.

DATE 6/8/2007

APPROVED: PROF. WILLIAM HARBAUGH

I. Introduction

There is a valuable reason for understanding the utility-theoretic underpinnings of mean-variance (MV) investing. It appeals to the "inner investor" as a logical, simplistic summation of a large set of complexities. While there is a school of thought that believes MV is a subset of a more-advanced or more-general choice model, it remains a bastion in the finance community. It's practical, simple, and useful for theory and practice. The von Neumann-Morgenstern (NM) axioms of rational behavior offer a specific and reliable foundation for analysis of this theory. This basis also allows for a definition of efficiency in MV space. However, the basic hypothesis of MV theory has not been subject to experimental review using revealed preference techniques; that hypothesis being that an investor displays consistent MV preferences.

In this paper I test, experimentally, whether people's choices in MV space are consistent with axioms of revealed preference. More specifically, do investor's decisions demonstrate consistency when their choices are presented strictly in terms of mean and variance? I will show that the design is consistent with literature which allows a corresponding NM utility function. Finally, the experiment design will allow to test the extent of any violations in revealed preference using an Efficiency Index.

First, I will briefly discuss the basis of the theory. The fundamentals of MV investing are straightforward. The theory assumes that portfolio a_1 is preferred to portfolio a_2 if and only if the expected utility of a_1—denoted $EU(a_1)$—exceeds that of a_2. This can be simplified given that the returns of a_1 and a_2, which we denote X_1 and X_2, are the result of some random function with probability distributions $F(X_1)$ and $F(X_2)$. Therefore we can simplify this into:

$$a_1 >_{nm} a_2 \;\; iff \;\; EU(a_1 \mid F(X_1)) > EU(a_2 \mid F(X_2))$$ [1]

The final assumption of MV theory states that the expected utility function above can be simplified into a function of only the mean and variance of the return on the portfolio.

$$a_1 >_{nm} a_2 \;\; iff \;\; EU(a_1 \mid F(X_1)) = V(\mu_1, \sigma_1^2) > V(\mu_2, \sigma_2^2) = EU(a_2 \mid F(X_2))$$

Applying the standard expected utility formula for continuous probability functions allows for simplification into:

$$a_1 >_{nm} a_2 \;\; iff \;\; \int U(x)F(X_1)dx = V(\mu_1, \sigma_1^2) > V(\mu_2, \sigma_2^2) = \int U(x)F(X_2)dx$$

where U(x) is a NM utility function over wealth. Therefore, by definition, an investor with mean-variance preferences must have some function *V* which is solely a function of mean and variance of returns, which reflects their expected utility.

Given that the purpose of the experiment is to establish whether MV decisions are consistent with axioms of revealed preference, the above is sufficient background for MV theory. However, in order to reject the conclusion that they do not due to a lack of corresponding NM utility function, I include a review of the literature and provide a solid theoretical basis for the test.

Having established this, the paper is broken into four sections which discuss the following:

1. Limitations placed on a NM utility function which corresponds to MV preferences.
2. Experimental tools utilized and their theoretical background, including a cursory discussion of GARP and other revealed-preference principles.
3. Experiment design.
4. Results and models.

The scope of this paper also includes possibilities for future research into the subject.

[1] Note that we use the $>_{nm}$ notation to designate that an alternative is strictly preferred in the NM sense. While we chose to examine strong preference, clearly indifference follows logically.

II. Acceptable forms of utility functions for optimal portfolio selection using mean-variance analysis

In order to fully develop a set of limitations of MV analysis, one must understand the goals they wish to pursue. For instance, if one is seeking to rank alternatives in order of preference using means and variance of returns, the limitations they face are in stark contrast to those of someone whom seeks to identify only the optimal portfolio for an investor. The purpose of this paper is limited to the latter. Namely, since the goal is to examine whether choices are consistent with GARP, we must only concern ourselves with rejecting the possibility that decisions are inconsistent due to the lack of a corresponding NM utility function.

Baron (1977) presented a series of propositions which express the relationship between MV preferences and NM expected utility models. These propositions will form the foundation of the literature review portion of this paper, and will be used to structure the arguments. Each of the issues will be addressed mathematically as well as intuitively. This means that the paper will include a discussion of the connection between risk aversion and decisions in MV space.

Consider a set of pure strategies, A^*, with strategies a_i, $i = 1,2,\ldots,n$. Each strategy is mutually exclusive—therefore a strategy could be "invest all available money in security i" or another equivalent expression. For the purpose of experimentation, a portfolio and a single security are equivalent in this paper, as further explained in Section II. Given this set of pure strategies, the payoffs resulting from each choice are denoted by X_i which is a random variable that takes on values x_i in the set X^*. The random return X_i are distributed with function $F(X_i)$.

We may consider a pure strategy as indicated by a probability vector $p = (0,\ldots 0,1,\ldots 0)$ with *ith* entry equal to one corresponding to the a_i strategy. Therefore a mixed strategy corresponds to a probability vector in which some entries are between zero and 1 and the total

entries of the vector sum to one. Intuitively, this is equivalent to a probability mixture between some set of strategies.

Mixed strategies in portfolio optimization present a major roadblock for experimental testing, as the experiment would have to be done over much iteration in order to tease out the underlying mixed strategies. However, Baron provides a solution in his third proposition (see Appendix 1) and demonstrates that, in the case when both pure strategies and mixed are available, pure strategies dominate mixed for all increasing NM utility functions.

Our next concern is that initially presented in Markowitz (1959), in which he shows that for any set of distributions, there exists a $V(\mu , \sigma^2)$ if and only if the corresponding NM utility function takes a quadratic form. While the mathematical proof is not necessary, it is important to take note of this discovery. This conclusion, on its surface, would appear to preclude any analysis with revealed preference.

Borch's paradox offers a solution for experimental purposes. Borch (1969) states, in short, that if an individual has preferences that are both MV and satisfy NM axioms then they cannot be indifferent between all distributions that have the same mean and variance. Thus, turning this proposition around allows for a greater scope. If we limit ourselves to the class of normal distributions, the classes of NM utility functions we can apply expands, as discussed by Baron.

In order to reject the possibility that people do not have well-specified preferences (as represented by some function) in MV space because there does not exist some NM utility function which directly corresponds to it, the experiment utilizes normally distributed returns for all securities. With this simplification normally distributed returns, I largely refute such claims and validate the use of revealed preference testing.

This will allow for greater generalization when it comes to results. Having rejected the posibilities presented earlier, I can generalize to a larger subsection of people choosing over similar inversments.

III. Revealed preference axioms, theory, and practice in experiment

Revealed preference theory was pioneered in Samuelson's seminal 1938 work, which simply stated that a person's utility can best be derived from actual behavior, rather than some function in general. In short, a person's decisions are completely reflective of their preferences. This concept was expanded and best summarized by Varian (2005), which outlines the major contributions to the theory. Among those is Samuelson's weak axiom of revealed preference (WARP) which states a direct revealed preference relation, which applies pairwise. This was expanded by Houthakker (1950) to the strong axiom which can be applied to more than two budget constraints, and states:

***SARP definition**: If x is revealed preferred to y [directly or otherwise], then it is not the case that y is revealed preferred to x.*

This definition gives a relation which can be utilized mathematically as well. However, this form is not sufficient to develop experimentation. Afriat (1967) set up a different approach: he took a finite set of observed prices and choices and asked how to actually construct a utility function that would be consistent with these choices. This began the movement which is utilized here.

Varian (1982) himself developed a Generalized Axiom of Revealed Preference (GARP), a minor variant on the SARP. The purpose of GARP is to establish a necessary and sufficient condition for consistency with the maximization of a continuous, concave utility

function. This end is also consistent, as shown above, with the financial needs; namely a concave utility function over consumption. Thus Varian provided a way to analyze choices without a loss of economic generality.

However, the work of early authors focused on continuous choice sets. As will be later explained, participants in this experiment are not presented with continuous choice sets; rather a set of discrete bundles along a budget constraint. Harbaugh (2001) faced a similar problem with decision making and children. Rather than restate his conclusion, I will present it here.

> "[W]e need to restate GARP in terms of choices, as follows. First, we say that a person directly reveals that they prefer bundle x_i to bundle x when they choose x_i over x or over a bundle with at least as much of every good as in x, and more of at least one. We say that a person indirectly reveals that they prefer x_i to bundle x when some sequence of directly preferred three relations between bundles connects x_i to x. GARP then requires that if a person directly or indirectly reveals that they prefer x_i to x, they cannot choose bundle x_i when some alternative x_j with at least as much of every good as in x, and more of at least one, is available. We can then state a new version of Afriat's theorem: If choices satisfy this version of GARP, they are consistent with the maximization of a continuous, concave, strongly monotonic utility function."

Using this definition of the GARP, we can analyze the choices of a person within the framework of revealed preferences. What we seek to discover is violations of the consistency required in revealed preference axioms. A number of minor violations[2] is expected; however, if there are a great number of violations, that ultimately points to the inability of a person to make consistent decisions within the space of goods. Therefore, a number of revealed preference violations in this case would indicate the lack of a well-specified utility function $V(\mu, \sigma^2)$ in MV space. More to the point from a finance perspective, it would indicate that people do not have preferences which can be specified only by the mean and variance of a security's return.

[2] The degree of a violation has been established by Afriat (1972) through an efficiency index. We will use a similar measure.

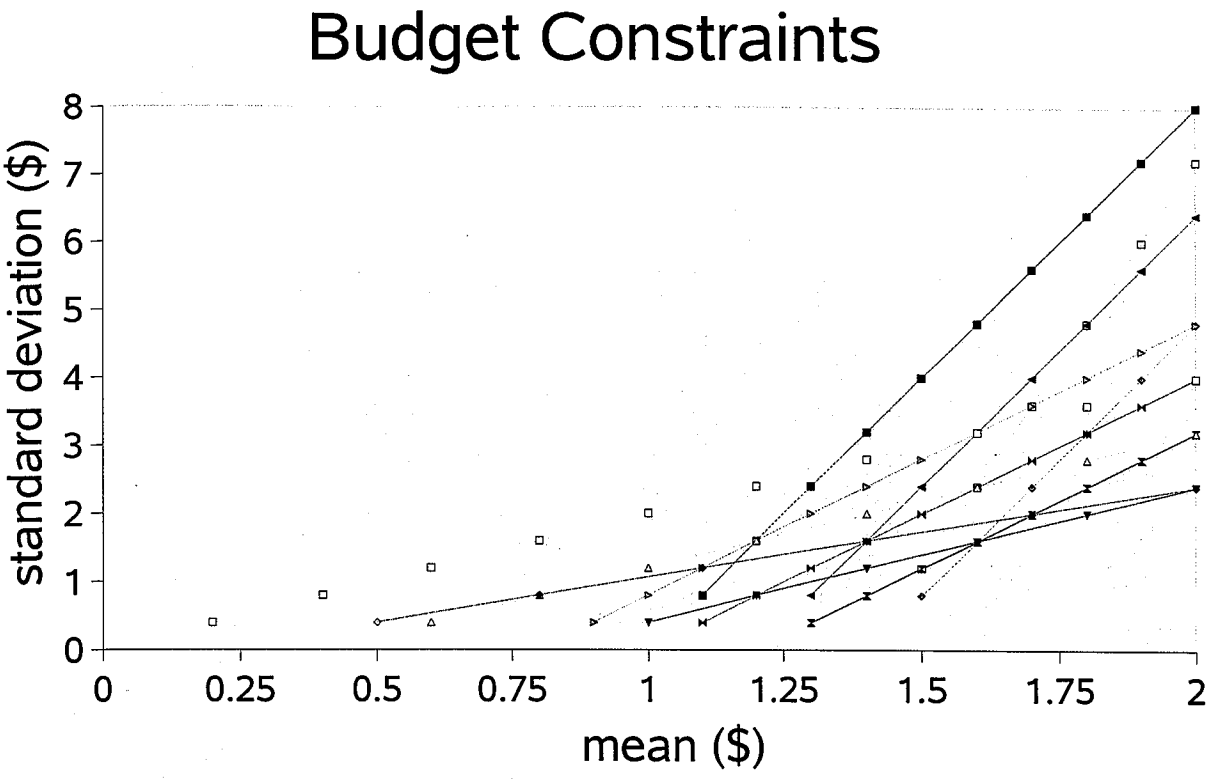


FIGURE 1

III. EXPERIMENTAL DESIGN

The experiment design is based on the work of Harbaugh and, before him, Andreoni and Miller (1998). Participants (college students, n=19, 6 male) were given choices over a series of 14 discreet budget constraints containing "prospects" meant to resemble securities in mean-variance space (see above). They were then asked to choose the prospect they most preferred for each constraint.

However, an important distinction is made from Harbaugh. Rather than downward sloping budget constraints, participants were presented upward sloping constraints (with mean on the y axis, variance[3] on the x). The purpose of this design is straightforward: additional

[3] Participants will actually choose over mean and standard deviation bundles. This simplification can be made without loss of generality, but with the added bonus of clarity and familiarity for participants.

variance in the return of a security is considered an economic "bad" according to finance theory. That is, people seek to maximize mean return subject to some maximal acceptable variance.

This variation, while possibly worrisome, is wholly acceptable under finance theory, and is easily reconciled with previous work. When the assumption of a continuous, concave utility function is made, variance is indeed a bad and decisions (over a normally distributed return) adhere to mean-variance preferences. Furthermore, the assumption will not ultimately lead participants down the path of violations of revealed preference; rather, it is an assumption and an experimental design with the purpose of capturing the maximal amount of violations.

Participants were given fifteen dollars to start the experiment[4], which they were forced to "invest" for real payoffs. I showed the participants sufficient money to make the promise credible. Participants then read the following (approximately) concerning their choices:

> "We're paying you $15 for participating in this experiment. We're now going to pass out $15 to each of you. During the experiment you can either add to this amount, or lose some of it. Whether you gain more money, or lose, will depend in part on your decisions and in part on chance. We will pay you your earnings in cash at the end. The experiment is set up so that you can't end up owing us money.
> In brief, the experiment works like this:
> - You start with $15.
> - You will see lists of investments.
> - Each chart represents a single investment and the payoffs corresponding to it.
> - For each list of investments, indicate the one investment you most prefer.

[4] This choice of design is made with a very specific purpose. Without the money given upfront, the variance of an investment's payoff cannot lead to a negative result. Namely, a person (without the 15 dollars upfront) could not face an investment which has a chance of losing money. Therefore, this "fifteen dollar" trick could be considered a device with the purpose of inducing risk adversity. Furthermore, this design more accurately reflect the reality of investment—namely the placing of one's own money into investment decisions which involve risk. Since the participants were presented with nearly mean zero/variance zero investments (i.e. keep all fifteen dollars with near certainty) there is still a method to detect a high degree of risk aversion.

A critic of this technique is certain to bring up prospect theory, which, in short, states that a person with a concave utility function over gains may have a convex utility function over losses. This seems to be pertinent here, since it would suggest that the variance of the investment may be an economic "good" to people displaying prospect theory properties. While this has been fully considered, it does not rule out the technique. If this holds, then people will simply maximize both mean and variance and such decisions will a) stick out like a sore thumb and b) not violate any revealed preference axioms. Therefore, while seemingly pertinent, the problems of prospect theory prove to be somewhat irrelevant.

- After you have marked every list, we'll pick one list at random.
- The investment you selected from that list will be played for real money. We will randomly select the final outcome of that investment.
- The payoff you receive will be $15 plus any money earned or minus any money lost."

Participants then progressed through a series of instructions designed to familiarize them with the normal curve and its implications. Throughout this section (which lasted approximately 20 minutes), there are quiz questions which will be used as proxy measures for confusion.

An important contribution to the language throughout this section is the link between "risk and reward" and "variance". These words are commonplace in the investment world, but the two phrases are not easily linked in the mind of the average student. Through repetition and graphical representation, it will hopefully become clear that the two are synonymous for the purpose of this experiment.

After the quiz, participants were given approximately 20 minutes to make their decisions. A sample of the sheets from which participants chose can be found in the appendix. At the end of this period, we randomly selected a list which was used to calculate payoffs for all participants. Their final payoffs were determined by chance according to the distribution they chose.

During the experiment, the instructions clearly stated that choice sheets were independent of one another. Therefore, there would be no reason to mix strategies (i.e. high risk on one sheet followed by low risk on another). While a test on diversification would have interesting and potentially useful conclusions, for this experiment it was critical to stress that each sheet should be treated as if it were the only one being played for real money.

Directly before students were paid, they will were given four minutes to fill out a numeracy test. Peters et al (2006) demonstrate and validate the link between numeracy and

decision making. Their conclusions relate to framing and other effects; given the relative nature of comparing securities across many choice sets, it seems pertinent to control for numeracy. The test Peters et al employed by Lipkus et al (2001) was used in this experiment.

There were two versions of the choice sheets that participants were given. One had choices presented with low mean/low variance choices at the top and high mean/high variance choices at the bottom. The second version had the opposite order; high risk on top, low risk on bottom.

IV. Results

A table of descriptive statistics follows. Men scored 1.5 points better on the numeracy test, on average, than did women. The numeracy test is scored by the number of questions answered incorrectly out of 11. Likewise, the quiz is scored by the number incorrect. Thus a score of zero is perfect for both the quiz and numeracy test.

TABLE I: DESCRIPTIVE STATISTICS

	Statistic			
Variable	Mean	Standard Dev.	Min.	Max.
Numeracy score	2.53	1.99	0.0	8.0
Quiz score	0.37	0.58	0.0	2.0
Mean of selected security	1.80	0.68	0.3	3.3
Std. deviation for selected security	1.93	1.93	0.2	7.2

Clearly, if the quiz was a viable measure for understanding, there was little confusion during the experiment. Most participants scored very high on the quiz for understanding. Likewise, the numeracy score results, while not as high, seem very reasonable, as they closely reflect the results found in Peters (2006). In that study, the mean score was 2.6. Similar to the results in that study, the distribution of numeracy scores here was highly skewed.

In addition to descriptive statistics, I ran regressions to test the effects of the numeracy test, quiz score, and the minimum mean. The minimum mean was the lowest mean on the sheet—it roughly approximates price effects, as steeper (and thus more costly for additional mean) budget constraints generally had additional mean to begin. In addition, I tested the significance of two dummy variables: one dummy variable for the second treatment of the choice sets in which high mean/high variance choices are presented at the top, and a second dummy variable for gender.

I chose to regress these variables against the choice number. This gives an approximation of the average effect of the variables on the average choice number. It does not correlate to GARP violations, as we will see. The results from these regressions are on the following pages. Note that the minimum mean is approximately the x-intercept of the budget constraints.

As we can see from Table II, the backward ordering was very significant in the average selection. This can be attributed to "top bias"; namely, the propensity for participants to choose options closer to the top of the page rather than those further down. This suggests that participants may not be maximizing utility due to some framing effects.

TABLE II: REGRESSIONS

	Dependent Variable					
Independent Variable	Choice number					
Backward ordering (dum)	0.99					1.22
	0.54					*0.74*
Numeracy score		-0.18				-0.12
		0.14				*0.22*
Quiz score			0.22			0.66
			0.51			*0.58*
Male (dum)				1.24		1.26
				0.56		*0.73*
Minimum Mean (proxy for price increases)					-3.95	-3.95
					0.37	*0.37*
Adjusted or Psuedo R-square	0.1639	0.0828	0.1598	0.2235	0.2649	0.3866

**Italicized values are standard errors*

Also of note, the numeracy score of a participant seems to have little effect on the average choice of the participants. This would rule out the conclusion that the innumerate were somehow at a natural disadvantage and either did not take enough risk or took too much. Rather, participants of lower numerical ability chose roughly the same, on average, as their counterparts. Likewise, the quiz of understanding seems to have little effect on the average choice.

Males were greatly more risk-taking than females, with an average selection 1.2 picks higher. However, given the small sample size, this was only marginally significant. Males also scored, on average, 1.5 points higher than females on the numeracy test.

Finally, the minimum mean on the page was highly significant. However, since this was entered as a proxy for price effects, it seems more efficient to directly compare prices to the choices. Results from such regressions can be found in Table III.

TABLE III: PRICE EFFECT REGRESSIONS

		Dependent Variable	
Independent Variable	Choice Mean	Standardized Mean	Log Standardized Mean
Price of $1 Increase in Mean	0.02 *0.01*	-0.08 *0.01*	
Log of Price			-0.92 *0.06*
Backward ordering (dum)	0.18 *0.08*	0.18 *0.08*	
Adjusted or Psuedo R-square	0.0817	0.3573	0.6483

In the above regressions, choice mean represents the mean of the choice selected. However, as one can see by the budget constraints and by the regression results, this is not a good measure. This is not a good measure because certain choice sets begin with a higher minimum mean. Therefore, the choice mean could be very high even if a participant chose the least risky option in a set. Thus, a better measure is a standardized mean: namely, the mean minus the minimum mean in that choice set. When this new meteric, which I call the standardized mean or excess mean demanded, is regressed against the relative price, we now get a much more meaningful result. In order to counter the effects of the two ordering sets, the backward dummy was used in the regressions. We find that there is a negative relation between demand for mean and its price (in terms of higher variance), as one would expect.

This result, in fact, could be generalized into a derivation of the demand curve for mean relative to variance in MV space choice sets. Furthermore, logging the standardized mean and the relative price offers a third meaningful tool: the price elasticity. Here, we find that the price elasticity of mean is almost perfectly unit elastic.

This relationship is best seen visually, with relative price compared to the excess demanded mean. Figure 2 shows the average results; Figure 3 shows the results for the two

treatments. Note that the average selected mean is the excess mean demanded. Also note that Figure 3 has the axis inverted for graphical ease.

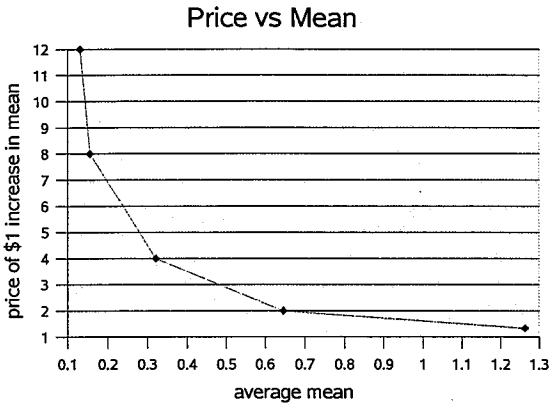


FIGURE 2

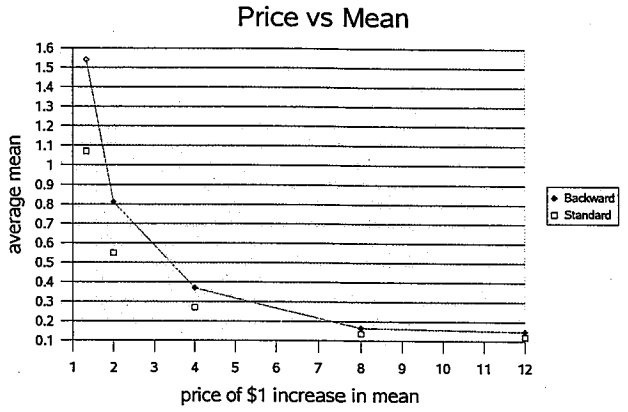


FIGURE 3

Even visually, it seems clear that demand is highly responsive to price. The exponential shape of the curves above also visually verify that the price elasticity of excess mean is nearly unit elastic.

This conclusion offers a startling result when put into a financial framework. If the aggregate demand for increased mean return on a security were unit elastic, it would suggest that investment levels would differ from those seen today. Using the typical mean-variance efficient sets introduced by Markowitz, we find that the relative price of additional mean increases with additional variance. This would correspond to the derivative of the efficient frontier on the following page.

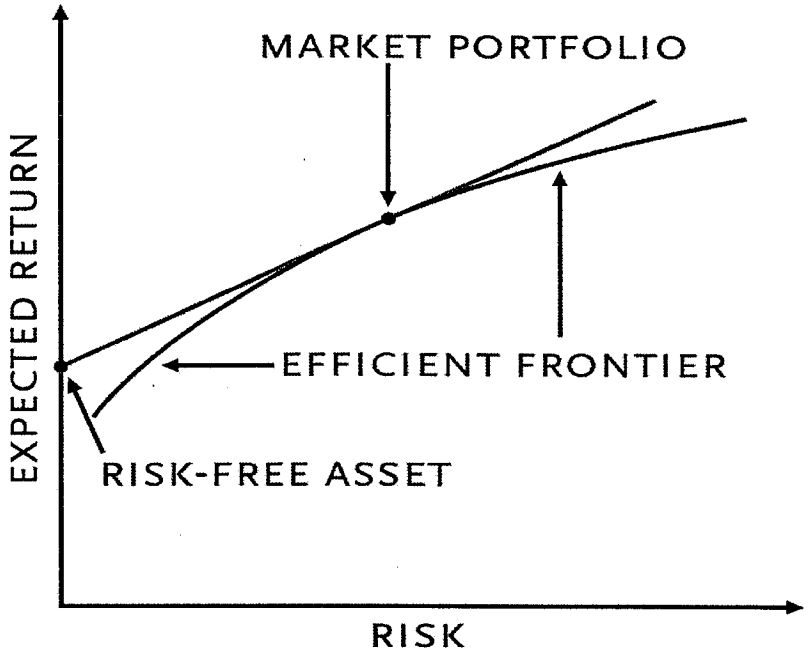


FIGURE 4

Note that the slope is increasing in expected return (or mean return). When standard deviation is used as the measure for risk as is expected in MV theory, we would find that a 1% price increase would lead to a 1% decrease in investment levels from the regression results discussed earlier. However, this would suggest that changing the shape of the efficient frontier would lead to volatile investment levels, as it would change relative prices. Given the history with the efficient frontier[5], this is an interesting conclusion. While this conclusion is certainly worth further investigation, there is a notable lack of covariance in my experiment (choice sets do not affect one another). As a result, this is left to later research.

In order to test participant's choices for GARP violations, I used software developed by Harbaugh and Varian. This software assumes downward sloping budget constraints and different relative prices. In order to fit the data to their program, I did a linear transformation of the choice sets. Theory states that variance will be considered as an "economics bad" in the MV

[5] The slope and shape of the efficient frontier has been frequently debated in the past. See Merton (1972); Elton, Gruber, Padberg (1978); and work of Eugene Fama and Kenneth French. There certainly are others as well.

space. However, in order to run GARP analysis, I needed to use a linear transformation and put choices in a space with two goods. This was necessary in order to scale the choice sets to the software. As this was a one-to-one transformation, no loss of power resulted.

TABLE IV: DESCRIPTIVE STATISTICS

Variable	Statistic			
	Mean	Standard Dev.	Min.	Max.
GARP violations	1.74	2.94	0.0	10.0
Afriat's efficiency score	0.97	0.06	0.8	1.0

**Afriat's is scored from 0-1, with 1 corresponding to efficiency*

Descriptive statistics for the 19 participants' GARP tests are above in Table IV. Quickly it becomes clear that there are a very smal number of violations and the average efficiency is extremely high. In fact, these results are striking in that they are amazingly efficient results. Harbaugh (2001) showed that, using 11 choice sets with fewer discrete choices[6], a randomly selected set would consist of ~8 GARP violations. In that study, children in second grade had an average of 4.3 violations; undergraduates had an average of 2.0. Here, we find that choices are even more in line with GARP axioms.

In order to reject the possibility that violations which did occur resulted from a lack of understanding of the instructions, I regressed the quiz score of participants against the number of violations they had. I found that quiz score had no significance. Likewise, regressing the numeracy test score against the number of GARP violations showed no effect on the part of the numeracy test. However, it should be noted that both coefficients were the anticipated sign.

[6]Increasing the number of discrete choices per set would almost certainly increase the number of GARP violations and decrease efficiency.

That is, greater numeric skills did correlate with lower GARP violations. Likewise, better performance on the quiz did correlate somewhat to GARP violations.

However, looking closely revealed an interesting side note: three participants scored 100% correct on the numeracy test. All three of those participants also had an Afriat Efficiency Index of 1, which corresponds to efficiency or near efficiency. This seems to suggest there may be a connection with numeracy and Afriat Efficiency Index. The efficiency index also reveals more about the incredible results from this experiment. Harbaugh found that that randomizing would lead to an index of .648; undergraduates had an average efficiency of .94. Here, participants' average index was .96, bettering the performance of any of Harbaugh's subjects.

A few conclusions immediately jump out from this table. First, it seems that participants' preferences are completely and amply described by only mean and variance. This is a significant conclusion, and I will attempt to validate it further. Secondly, the efficiency of choices is very high, again suggesting that preferences areproperly modeled. However, one must consider whether the number of violations is driven by a small number making many violations, or a large number of people making fewer violations. Table V shows the distribution of GARP violations. In addition, it also shows the Afriat's Efficiency Index score and the sets which included violations.

TABLE V: GARP VIOLATIONS

Participant #	Sets with GARP violations	# Violations	Afriat's Index
1	{2, 5, 6, 8, 9, 10}	6	0.89
2	n/a	0	1
3	n/a	0	1
4	n/a	0	1
5	{2, 7, 7, 9}	4	0.9
6	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}	10	0.78
7	n/a	0	1
8	n/a	0	1
9	{5, 9}	2	1
10	n/a	0	1
11	{5, 9}	2	1
12	n/a	0	1
13	n/a	0	1
14	n/a	0	1
15	n/a	0	1
16	{2, 5, 6, 7, 8, 9, 10}	7	0.89
17	n/a	0	1
18	n/a	0	1
19	{7, 9}	2	0.94

Clearly, the majority of participants had a small number of violations. This, along with the high efficiency indices, continues to suggest that there exists a fundamental set of preferences in MV space. However, there are a few things which could account for the range of GARP violations. In particular, the regressors in our early regressions may have had an effect on the number of GARP violations or the efficiency index of a participant. Perhaps the highly nummerate are more likely to have fewer violations. Likewise, perhaps if a participant had a better score on the quiz for understanding it would correspond to fewer violations. Therefore, I regressed the same variables against the number of violations and the efficiency index. The results follow in Tables VI and VII.

From Table VI, we see a number of interesting things. Numeracy appears to decrease the number of violations, but only a small bit and without any statistical significance. The second regressor, the quiz score, also has the expected sign, but with little significance.

Finally, males had, on average, fewer violations than females, but with only little statistical significance. The small sample size was deterimental to the power of regression techniques.

Table VII offers a similar result. Numeracy was correlated with greater efficiency, the quiz score had no significance, and males were far more efficient than females, but again with little statistical significance for any of the variables.

TABLE VI: REGRESSIONS

			Dependent Variable	
Independent Variable			Number of GARP violations	
Numeracy score	0.26			0.27
	0.34			*0.45*
Quiz score		-0.18		-0.69
		1.19		*1.44*
Male (dum)			-1.56	-1.20
			1.44	*1.65*
Adjusted or Psuedo R-square	0.0331	0.0013	0.0645	0.0884

TABLE VII: REGRESSIONS

			Dependent Variable	
Independent Variable			Afriat's Efficiency Index	
Numeracy score	-0.01			-0.01
	0.01			*0.01*
Quiz score		0.00		0.02
		0.03		*0.03*
Male (dum)			0.05	0.03
			0.03	*0.03*
Adjusted or Psuedo R-square	0.0882	0.0004	0.1305	0.1876

These tables offer no direct reason to counter the hypothesis that there is a fundamental degree of preference participants had in MV space. None of the variables is even close to statistically significant. While the sample number is very small, it still seems clear that there is

a clear pattern in the results. It seems safe to say that participants in this experiment have preferences which can be modeled using only mean and variance of return.

If people, both individually and on an aggregate level, have preferences which satisfy GARP axioms, then investment should be aggregately well-placed in terms of MV. That is, we should see investment which is aggregately on the MV efficient frontier. Such a conclusion may seem like a stretch given this experiment does not deal with sets which have covariance. However, this study does show that aggregately and individually, people's preferences can be properly modeled using only the mean and the variance of return of an investment.

IV. Areas for further research

My primary concern is whether the design used both proper language and proper visualizations which enabled participants to suitably make their decisions. Future research could benifit from cooperation from educational tools which may instruct better. However, my opinion is that this experiment satisfactorily enabled participants. In the future, using a group of students or adults whom are well-versed in MV efficient sets would allow a control group. A control group could answer such questions as to whether MV preferences are learned and not intuitive.

However, it still remains possible that participants may be only relying on some simple rule to make decisions in MV space. This result, as an end within itself, is very interesting. If the instructions sufficiently informed participants as I discussed above, then a group which relies solely on rules is very interesting. This result would suggest that most individuals should not control their investments. Such a resutls could be applied to the retirement population or other groups which engage in heavy securities trading.

Future research with a larger sample size and a control group would greatly extend the conclusions of this paper and would allows for generalizations to the larger population. Numeracy remains an unsolved question of this paper. While it appeared to have little statistical significance in regressions, it remains that the only three people who got 100% on the numberacy test also ended with Afriat's Indices of 1. Further investigation may be able to resolve that issue.

Works Cited

Andreoni, James and Miller, John. "Giving According to GARP. An Experiment of Rationality and Altruism." Social Systems Research Institute Working Paper #9902, University of Wisconsin-Madison, 1998.

Baron, David P. "On the Utility Theoretic Foundations of Mean-Variance Analysis." *Journal of Finance*, 1977, pp. 1683-1697.

Borch, K. "A Note on Uncertainty and Indifference Curves." Review of Economic Studies, 1969, pp. 1-4.

Elton, E.J., Gruber, M.J., and Padberg, M.W. "Simple Criteria for Optimal Portfolio Selection: Tracing Out the Efficient Frontier." *Journal of Finance,* March 1978, pp. 296-302.

Harbaugh, William T., Krause, Kate and Berry, Tim. "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review*, 2001, pp. 1539-1545.

Houthakker, Hendrick S. "Revealed Preference and the Utility Function." *Economica,* May 1950, pp. 159-174.

Levy, H. and Markowitz, H.M. "Approximating Expected Utility by a Function of Mean and Variance." *American Economic Review*, 1979, pp. 308-317.

Lipkus, Isaac M., Samsa, Greg, and Rimer, Barbara K. "General Performance on a Numeracy Scale among Highly Educated Samples." *Medical Decision Making*, 2001, No. 1; 37-44.

Markowitz, H.M. "Portfolio Selection." *Crowles Foundation*, 1959.

Merton, Robert C. "An Analytic Derivation of the Efficient Portfolio Frontier." *Journal of Financial and Quantitative Analysis*, September 1972, pp. 1851-1872.

Peters, Ellen, Västfjäll, Daniel, Slovic, Paul, Mertz, C.K., Mazzocco, Ketti, and Dickert, Stephan. "Numeracy and Decision Making." *Psychological Science*, 2006, pp. 407–413.

Samuelson, P.A. "A Note on the Pure Theory of Comsumer's Behavior." *Economica,* 1938, pp. 61-71.

Varian, Hal R. "The Non-Parametric Approach to Demand Analysis." *Econometrica*, 1982, pp. 945-973.

Varian, Hal R. "Revealed Preference." Prepared for S*amuelsonian Economics and the 21st Century*, 2005. Available at http://www.sims.berkeley.edu/~hal/Papers/2005/revpref.pdf

Appendix I: Protocol

Instructions:

This experiment is an investment simulation. You will be given a sum of money and be asked to make decisions on how to invest that money.

We're paying you $15 for participating in this experiment. We're now going to pass out $15 to each of you. During the experiment you can either add to this amount, or lose some of it. Whether you gain more money, or lose, will depend in part on your decisions and in part on chance. We will pay you your earnings in cash at the end. The experiment is set up so that you can't end up owing us money.

In brief, the experiment works like this:

- You start with $15.
- We will go through a discussion of the investments you get to choose.
- You will see lists of investments.
- Each chart represents a single investment and the payoffs corresponding to it.
- For each list of investments, indicate the one investment you most prefer. Pick only one investment per page.
- After you have marked every list, we'll pick one list at random.
- The investment you selected from that list will be played for real money. We will randomly select the final outcome of that investment.
- The payoff you receive will be $15 plus any money earned or minus any money lost. You will be paid in cash at the end of the experiment.

Details:

In this experiment, you will be asked to make decisions which involve risk. Investment in general involves weighing risk against reward. To begin, we will help clarify the choices you will be given.

You will be asked to make decisions on investments based on the *mean*, or average, return of the investment and the *variance* of the return. Don't worry, we will explain those terms.

First, the return on an investment is the final amount of money minus the initial amount which was invested. Therefore, your final payoff will be $15 plus the return you earn. If you earn a negative return, then you will be paid $15 minus the return you earn on your investment decision.

The mean is the average return of the investment. Suppose an investment yields the following returns in previous years:

Year	Return
2000	$40
2001	$20
2002	$30
2003	$50
2004	$10

In the above case, the *mean* return is the average of the five numbers, or $30.

Measures of the mean (or average) don't tell you about how much the data values differ from each other. Take the following example

<u>Professor Ages</u>
50 50 50 50 50 (Mean = 50)
20 30 50 80 90 (Mean = 50)

Clearly you can't call these groups of professors the same! You need a way to describe the variability in ages in the second group.

Measures of dispersion or variability attempt to measure numerically the spread of the distribution.

We will use the standard deviation as the measure. The standard deviations measures how different the values in a distribution are from the mean. The greater the standard deviation, the less typical the mean.

As the standard deviation increases, so does the variability.

Simply put, the higher the standard deviation, the more spread out the values.

But how do you use the standard deviation along with the mean to reach some conclusion? In this experiment, it can be used to compare possible outcomes which are decided by chance.

First, though, let's run through a familiar example.

EXAMPLE: Test Scores

Is a 65 on an exam a bad grade? Not if it's the best in the class. Is an 85 a good grade? Not if it's in the bottom 25%.

Suppose the instructor tells you that the average score was a 70. Now, if you scored a 65 you would be below average and if you scored an 85 you would be above average.

But if all you know is your grade and the mean you can only tell if your score is less than, equal to, or greater than the mean. You can't say how far it is from the average because you don't know how other people scored. There could be a lot of people at the extreme ends, or there could be a lot of people who scored exactly 65.

You need to know the distribution of the test scores to know how to interpret your own score.

In this experiment, as is frequently the case with exam scores, the distribution is a bell-shaped curve.

So, for a normal distribution, almost all values lie within 3 standard deviations of the mean.

→You can estimate this range of highly likely outcomes by adding and subtracting three times the standard deviation to the mean. For instance, with a mean of 10 and a standard deviation of 2, the range of likely outcomes with 99.7% certainty is (10 - (3 x 2)) to (10 + (3 x 2)) or simply 4 to 16.

This is an important rule which makes it easier to find out likely outcomes.

***At this time, *tear this page out* of the packet so that you can reference it later.**

EXAMPLE: Test Scores (continued)

Let's revisit the test score example.

If the average score for an exam is 70 and the standard deviation is 5, a score of 80 is quite a bit better than the rest. It is two standard deviations above the mean.

But if the standard deviation is 15, the score of 80 isn't very terrific. It's less than one standard deviation above the mean.
You can see how the combination of information tells you a lot!

[Quiz]
Answer the following questions by circling the correct answer. These questions have no effect on your payoffs in the experiment and are only used to make certain that you understand the experiment.

The experiment:

During the experiment we will show you various investments. Each investment is represented by a chart which shows the *probability* of certain outcomes and the corresponding return, should that outcome occur.

Remember, you will be paid $15 plus any return you earn. If you earn a negative return, we will take that out of your initial $15.

For example, suppose the following:

EXAMPLE

<u>Stock A</u>

Mean: $2.00

Standard Deviation: $0.25

Consider another investment:

<u>Stock B</u>

Mean: $2.00

Standard Deviation: $2.00

EXAMPLE: Suppose you are given this list and asked to choose the investment you most prefer.

<u>Stock A</u>

Mean: $1.00
Standard Deviation: $0.25

With 99.7% certainty, you know the following:

Stock A would give a return between $0.25 and $1.75. Therefore, your final payout would be between $15.25 and $16.75

Stock B would give a return between -$1.50 and $4.50. Therefore, your final payout would be between $13.50 and $19.50

Stock C would give a return between -$3.50 and $7.00. Therefore, your final payout would be between $11.50 and $23.00

Ultimately it's a choice over how much chance you're willing to take.

[QUIZ]

Using the previous page, answer the following questions. These questions have no effect on your payoffs in the experiment and are only used to make certain that you understand the experiment.

1. Which of the investments represents the investment which is most likely to give a final payoff above $15.00

_____

2. Which of the investments represents the investment which is most likely to give a final payoff below $15.00?

_____

3. Which of the investments has the greatest certainty over the final payoff?

_____

4. Which of the investments has the least amount of certainty over the final payoff?

_____

5. Which investment do you prefer? [There is no "right" answer here]

_____

6. Suppose a fourth investment, Stock D, was an option. It has a mean of $2.00 and a standard deviation of $2.00. What is the range of final payouts which will occur with 99.7% certainty? HINT: Use torn out sheet.

_____

Stop, wait for us to review the answers with you.

Please don't turn the page until we tell you to.

This ends the instruction portion of the experiment. Let's review briefly.

You will see lists of investments, much like the lists you have seen in the previous pages.

For each list of investments you see, pick the investment which you most prefer. You may indicate the investment you most prefer by clearly circling the stock name.

After you've made all your choices, we will randomly pick one of the lists to be the "list that counts," by drawing a card randomly from a deck that has one card representing each list.

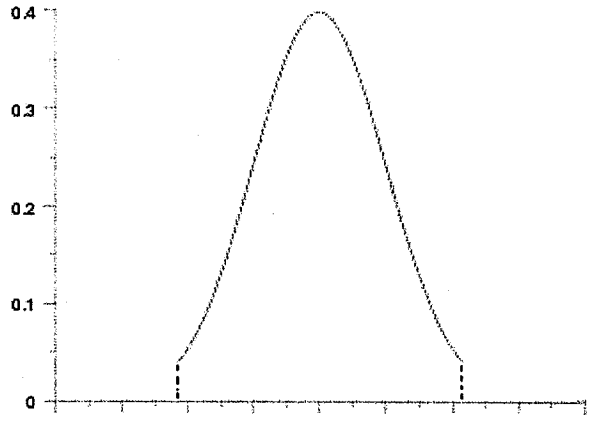
Since you don't know which list of investments we will pick, you should treat each list as if it is the one that counts, and mark your choice as if you are choosing only from that list.

Once we've picked a list, your investment pick for that list will be played for real money. We will decide the return on your investment using a randomly generated number from a computer.

You will then be paid $15.00 plus or minus the return on the investment you chose.

In the real world, there would be a chance that you could lose a great deal of money. However, we don't want you to end up owing any money. So we have "cut" ends off of the distribution (see diagram below).

Note that this is an exaggeration. The real diagrams will look like those presented earlier.



There are no payoffs which are greater than $15.00 and no returns smaller than -$15.00.

Therefore, your final payoff will range between $0.00 and $30.00

Outcomes in which you owe money were very, very unlikely, but to be safe we removed the possibility.

Finally, nobody likes getting pennies, so your final payoff will be rounded to the nearest quarter. There is no need to worry about getting a strange payout with lots of pennies and dimes!

Again, in brief the experiment works like this:

You start with $15.

- You will see lists of investments
- For each list, select the one you most prefer.
- After you marked every list, we'll pick one list at random.
- The investment you selected from that list will be used to determine the final payoff.
- We will add or subtract the earnings you receive from the investment to your initial $15 to determine your final payoff.

You will have ample time to go through the following pages and choose one investment per page.

Take your time and read carefully.

For simplicity, the mean of the investments will always decrease from the top of the page down. That is, high mean/high risk investments are at the top; low mean/low risk investments are at the bottom.

If you would like a calculator, you are welcome to use one. If you do not have one, please raise your hand and we will provide one for you.

DON'T TURN THE PAGE UNTIL WE TELL YOU TO DO SO.

[END APPENDIX I]

Appendix II: Sample Choice Sheet

Appendix II: Sample Choice Sheet (Backwards Treatment)