E c o n o m i c s B u l l e t i n

Comparison of Mean-Variance Theory and Expected-Utility Theory through a Laboratory Experiment

Andrea Morone
UniversitÃ degli studi di Bari

Abstract

In the 40's and early 50's two decision theories were proposed and have dominated the scene of the fascinating field of decision-making. Since 1944 - when von Neumann and Morgenstern showed that if preferences are consistent with a set of axioms then it is possible to represent these preferences by the expectation of some utility function - Expected Utility theory provides a natural way to establish "measurable utility". In the early 50's Markowitz introduced the Mean-Variance theory that is the basis of modern portfolio selection theory. Since then, both models were analyzed from virtually all possible points of view and were tested against several generalizations. However, these two models should be tested against each other. This paper tries to fill this gap, investigating (using experimental data) which of these two models better approximate subjects' preferences.

Citation: Morone, Andrea, (2008) "Comparison of Mean-Variance Theory and Expected-Utility Theory through a Laboratory Experiment." *Economics Bulletin,* Vol. 3, No. 40 pp. 1-7
Submitted: May 11, 2008. **Accepted:** August 8, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C90003A.pdf

1. Introduction

This paper is motivated by simple questions: why, in two different branches of Economics, do two different preference functionals dominate the scene? Does Expected Utility perform significantly better than Mean Variance? Do we have a loss of accuracy if we use Mean Variance instead of Expected Utility?

Expected Utility leads the field of decision making in Economics because, since 1944 – when von Neumann and Morgenstern showed that if preferences are consistent with a set of axioms then it is possible to represent these preferences by the expectation of some utility function – Expected Utility provides a natural way to establish "measurable utility": it is a simple and elegant way to derive utility cardinality.

Mean Variance leads the field of decision making in Financial Economics. It was developed in the 50's and 60's by Markowitz (1952), Tobin(1958), Sharpe (1964) among others. It is an important model of investment based on decision theory. Actually, it is the simplest model of investment that is sufficiently rich to be directly useful in applied problems.

It is clear that both models have nice desirable properties. It is, also, rather obvious that Expected Utility should perform better than Mean Variance. Indeed, it is a more general model (Levy and Markowitz (1979), Kroll et al. (1984)). And, finally, we should expect that using Mean Variance instead of Expected Utility, we have to accept a loss in accuracy. But what is rather striking is that neither the presumed superiority of Expected Utility nor the accuracy loss of Mean Variance has been systematically investigated. The aim of the present paper is to fill this gap. In a certain sense, we are addressing three questions:

1. Why, in two different branches of Economics, do two different preference functionals dominate the scene?
2. Does Expected Utility perform significantly better than Mean Variance?
3. Does the use of Mean Variance instead of Expected Utility produce a loss of accuracy?

In section 2, we briefly describe the data, which we used to estimate the three preference functionals. Section 3 illustrates the features of the preference functionals analyzed and presents the estimation procedures. Section 4 discusses the estimation results. Finally, results are presented and discussed in section 5.

2. The data

Much effort has been expended to produce a better theory of decision making under risk than that provided by EU. Therefore, there is now an abundant literature that compares EU with a number of its generalizations (e.g. Harless and Camerer (1994), Hey and Orme (1994), Hey (1995, 2001)). It seems fairly natural to follow one of these approaches to compare MV and EU. We decided to follow Hey and Orme's approach. Thus we need a set of pair-wise choice questions. Each pair-wise choice is composed of two lotteries, labeled "Left Gamble" and "Right Gamble". Each subject has to report his preference between the two lotteries. The incentive mechanism is that the preferred lottery is played for real.

The enormous activity of this branch of experimental economics makes it unnecessary to run our own pair-wise choice question experiment, since we can address our questions using a data set from a previous experiment[1].

The experiment took place in the EXEC laboratory at the University of York with 53 participants. Each participant had to attend five separate treatments, Set 1, Set 2, Set 3, Set 4 and Set 5. Each of the five treatments was composed of the same 100 pair-wise choice questions, with different chronological order, and randomized left/right position. The pair-wise choice questions

[1] I have to thank John Hey for letting me use one of his experiment's data set. A more detailed presentation of the experimental design can be found in Hey (2001).

were presented in the form of segmented circles, and subjects were asked to report, for each pair, their preferences.

The 100 questions included three of the following four outcomes: -£25, £25, £75, and £125.[2] One of these four outcomes involves a negative pay-off; to avoid that subjects can experience a real monetary loss a participation fee of £25 for attending all the 5 sessions of the experiment was paid.

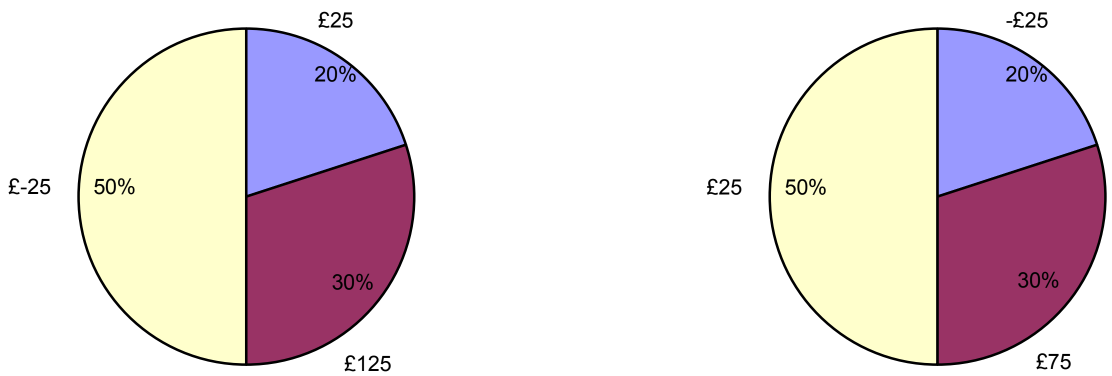

Figure 1: The Presentation of Lotteries

The lotteries were presented as segmented circles on the computer screen. Figure 1 presents an example. If a subject received a particular lottery as reward he or she had to spin a wheel on the corresponding circle. The amount won was then determined by the segment of the circle in which the arrow on the wheel stopped.

3. Some notes on estimation techniques

The estimation of the parameters of the utility function from pairwise choice data follows Hey and Orme (1994). Lets indicate the two lotteries in the pairwise choice by L and R; On the one hand, if Eu(L) > Eu(R) then the subject chooses L with probability 1 - ε; on the other hand if Eu(L) < Eu(R) he/she choose L with probability ε. Where ε denotes the noise or measurement error. We assume that ε is normally distributed with mean 0 and variance s^2. Given the actually reported preferences we will proceed to the estimation of the parameters using maximum likelihood methods.

Note that when estimating a utility function from an experiment, there are two usual approaches: (a) to assume a particular functional form and estimate the parameters of that form; or (b) to estimate the utility at the various outcome values used in the experiment. In the experiment there were four outcome values (-£25, £25, £75, and £125) which we denote by x1, x2, x3 and x4. If we adopt the usual normalisation, we put u1 = 0 and u4 = 1, where we denote u(xi) by ui. This means that, following approach (b), we simply estimate u2 and u3.

[2] See Hey, 2001.

4. Estimation procedure and preference functionals

Our estimation procedure is similar to the one used by Hey and Orme (1994) which is motivated by two fundamental observations. First, there is not necessarily one best preference functional for all subjects but the behavior of different subjects may be explained best by different functionals. Second, subjects make from time to time errors in their responses which demand a stochastic specification of preference functionals for our empirical test. To take into account the first observation we have estimated the models subject by subject. To take into account the second observation we have added an error term to each preference functional. We assume that errors are identically and independently distributed among subjects and questions.

In our analysis, we will consider three preference functionals:

- Risk Neutral (RN)[3];
- Mean Variance (MV);
- Expected Utility (EU).

First some notation, let $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ be the vector of outcomes; $\mathbf{p} = \{p_1, p_2, ..., p_n\}$ is the probability vector of the Left Gamble and $\mathbf{q} = \{q_1, q_2, ..., q_n\}$ the probability vector of the Right Gamble. W denotes the subject's preference function. Therefore, if $W(\mathbf{p}) > W(\mathbf{q})$ Left will be preferred to Right and if $W(\mathbf{p}) < W(\mathbf{q})$ then Right will be preferred to Left.

Altogether subjects' derived preferences are determined by $W(\mathbf{p}) - W(\mathbf{q}) + \varepsilon$, where ε is an error term. We assume that ε is symmetric and has a mean of zero.

The first model, we have estimated, is RN given by

$$RN\!: W(\mathbf{p}) - W(\mathbf{q}) + \varepsilon = k\sum_{i=1}^{n} p_i x_i - k\sum_{i=1}^{n} q_i x_i + \varepsilon .$$

For RN, we have to estimate only the parameter k which is the relative magnitude of subjects' errors.

Let us now turn to MV where we have:

$$MV\!: W(\mathbf{p})-W(\mathbf{q})+\varepsilon = v\sum_{i=1}^{n} p_i x_i + w\sum_{i=1}^{n}\left(p_i\left(x_i - \sum_{j=1}^{n} p_j x_j\right)\right)^2 - \left[v\sum_{i=1}^{n} q_i x_i + w\sum_{i=1}^{n}\left(q_i\left(x_i - \sum_{j=1}^{n} q_j x_j\right)\right)^2\right] + \varepsilon$$

Concerning MV, we have to estimate v and w, which represent, respectively, the weight that each subject gave to the mean of the lottery and to its variance.

Finally EU:

$$EU\!: W(\mathbf{p}) - W(\mathbf{q}) + \varepsilon = \sum_{i=1}^{n} p_i u(x_i) - \sum_{i=1}^{n} q_i u(x_i) + \varepsilon .$$

For EU, we estimated $u(x_i)$, we normalised $u(x_1)$ to zero, and the variance of the error term to one.

4. The estimation results

The first question we are trying to address is which – RN, MV, and EU – of the various preference functionals best explain subjects' behaviour. A very natural way to compare the performances of our three preference functionals is ranking them according to the Akaike Information Criterion (AIC). This is a measure of goodness of fit, which takes into account the

[3] RN will be a kind of low benchmark

model parsimony. In table 1, it is reported the frequency of ranking first, second or third by the three models according the AIC[4].

	RN			MV			EU		
	1	2	3	1	2	3	1	2	3
Set 1	0	0	53	3	50	0	50	3	0
Set 2	0	0	53	5	48	0	48	5	0
Set 3	0	0	53	6	47	0	47	6	0
Set 4	0	0	53	4	49	0	49	4	0
Set 5	0	1	52	4	49	0	52	0	1

Table 1: frequency of ranking first, second or third according the AIC

Looking at table 1 we have a very clear picture: EU performs better then its challengers. At this stage, we can conclude that according to the AIC, we have to prefer EU to MV. The strength of this kind of analyses is that it gives us a complete ranking of the preference functionals, but it does not help us to answer our second question: how much one preference functional is better than the other one. To investigate this particular aspect, we can analyse the log- likelihood value. This value gives us the probability that a preference functional fits correctly the subject actual preferences, but it is not correct for the degree of freedom (that is, it does not penalize for the number of parameters).

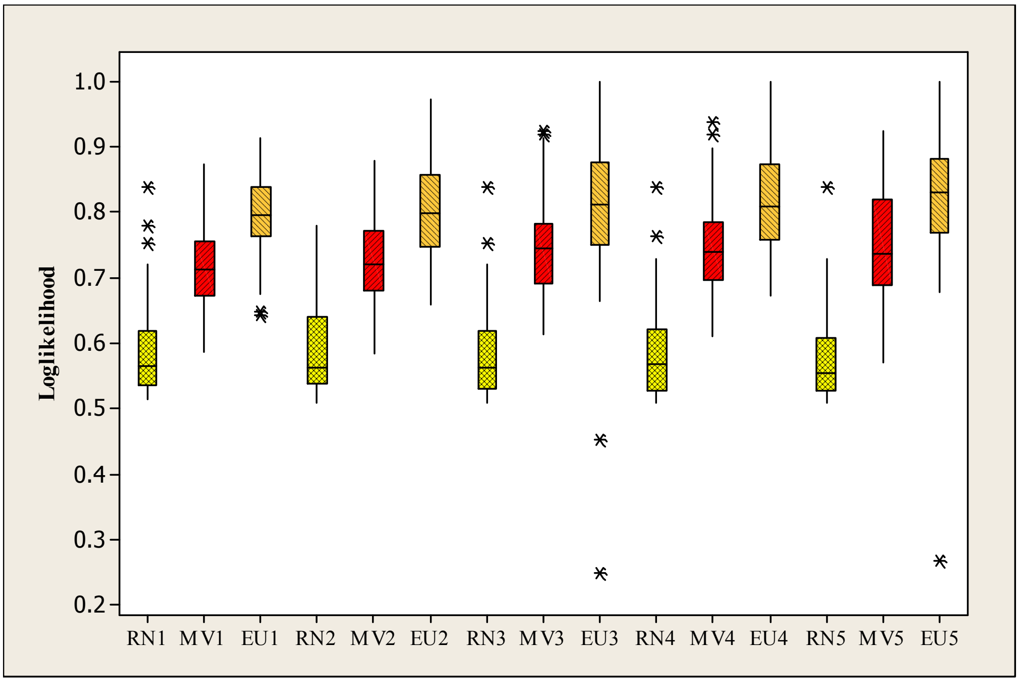

Figure 2: Box plot of the estimated log-likelihood

In figure 2 we report the box plot of the estimated log-likelihood. As we can see EU performs roughly the 10% better then MV. Since RN performs particularly poorly from now on we will concentrate our attention only on MV and EU.

[4] When we calculated the average rankings two models got the same rank if they performed identically. If, for example, two models have the highest Akaike criterion, they both get the first rank and the next model gets rank three. For this reason the average of the average ranks may differ from the rank average.

In table 2 is reported the frequency of the difference between the likelihood value of EU and the likelihood value of MV.

	up to 1%	1%-5%	5%-10%	10%-15%	more than15%
Set 1	28.30	32.08	28.30	7.55	3.77
Set 2	24.53	43.40	20.75	7.55	3.77
Set 3	30.19	35.85	24.53	9.43	0.00
Set 4	24.53	45.28	28.30	0.00	1.89
Set 5	35.85	28.30	30.19	3.77	1.89

Table 2: likelihood of EU – likelihood of MV

From this table, we have again a clear picture of the superiority of EU, but more important, it gives us an indication on the loss of accuracy we have to be ready to accept if we use MV instead of EU.

This kind of analysis is only a statistical one, and even then we reach some important conclusion on the superiority of EU with respect to MV and the loss of accuracy. But we are interested also in some economics analysis to measure the accuracy loss. One way of answering this is the following. We can evaluate the distance between the real subjects' preferences and the estimated one. But unfortunately, it is not obvious how to define a distance function. Should we consider only the number of times that the estimated preference does not match with the actual preference or should we consider also the magnitude of the errors. It seems that the harmless mechanism should be counting how many mistakes are produced by a particular preference functional in the prediction of actual behaviour. In figure 3 we reported the difference between the mistakes produced by MV and the mistakes produced by EU. For less then the 20% of our subject pool MV performs better or equally to EU.

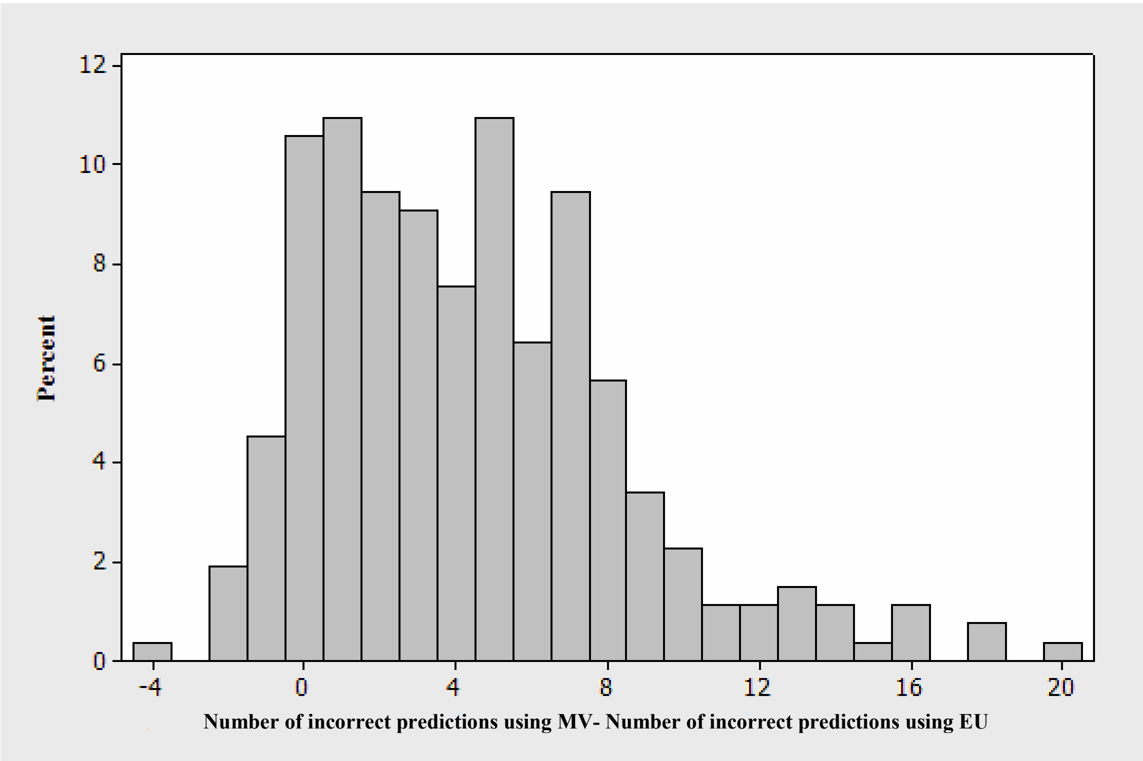

Figure 3: Difference between the mistakes produced by MV and the mistakes produced by EU

In table 3 is reported the percentage of ratio between the number of times EU's prediction is different from the actual subject preference and the number of times MV's prediction is different from the subject actual preference.

	up to 1	1-1.5	1.5-2	2-2.5	more than 2.5
Set 1	18.87%	35.85%	30.19%	5.66%	9.43%
Set 2	11.32%	39.62%	26.42%	3.77%	18.87%
Set 3	24.53%	30.19%	20.75%	15.09%	9.43%
Set 4	16.98%	37.74%	32.08%	3.77%	9.43%
Set 5	16.98%	41.51%	13.21%	15.09%	13.21%

Table 3: ratio between the number of times EU's prediction is different from the actual subject preference and the number of times MV's prediction is different from the subject actual preference

From this table it is clear that MV performances are not particularly good. In fact, only in 18-25% of the cases, its performance is better than EU. It is particularly surprising that 10-19% of the subjects using MV instead of EU will produce an error more than 2.5 times bigger.

A formal comparison of EU and MV is not straightforward, since these two models are non-nested. For the purpose of this comparison, we make use of Vuong's (1989) non-nested likelihood ratio test (Z)[5]. As proved by Vong Z is distributed as a standard normal distribution, and a significantly positive value of Z indicates that EU is closer to the true data generating process than MV (while a significantly negative value of Z indicates that MV is closer to the true data generating process than EU). In table 4 are reported the Z statistics for the five repetitions.

	Set1	Set2	Set3	Set4	Set5
Z	5.039103	4.913833	4.336087	4.821314	4.596744

Table 4: Vuong's non-nested likelihood ratio test (Z)

According to the Vuong's non-nested likelihood ratio test (Z) we can accept the hypothesis that EU performs better than MV.

5. Conclusion

This article produces two important results, one in the experimental field and the other in the financial one. On one hand, it covers the gap in the literature of decision under risk comparing the Expected Utility Theory with Mean-Variance Theory.

In terms of best-fitting preference functional EU emerges to perform better than its challenger. On the other hand, it suggests that the loss of accuracy using MV instead of EU in terms of fitting is generally low (for more than 50% of the subjects it is less than 5%). But from a non statistical analysis, we learned that it is dangerous to use MV instead of EU because with data for 10-19% of the subjects using MV instead of EU will produce an error more than 2.5 times bigger.

[5] For a more detailed explanation of the Voung's test see Loomes et al. (2002)

References

Harless and Camerer (1994), The Predictive Utility of Generalized Expected Utility Theories, *Econometrica,* 62: 1251-1289.

Hey, J. D. (1995). “Experimental investigations of error in decision making under risk.” *European Economic Review,* 39:633-640.

Hey, J. D. (2001). “Does repetition improve consistency?” *Experimental Economics,* 4(1)5-54.

Hey, J. D. and Orme, C. D. (1994). “Investigating generalizations of expected utility theory using experimental data.” *Econometrica,* 62:1291-1326.

Kroll, Y., Levy, H and Markowitz, H. (1984). “Mean-Variance versus direct utility maximization.” *Journal of Finance,* 39:47-62.

Levy, H. and Markowitz, H. (1979). “Approximating expected utility by function of mean and variance.” *American Economic Review,* 69:308-319.

Loomes, G. C., Moffatt, P. G., and Sugden, R. (2002) “ A Microeconometric Test of Alternative Stochastic Theory of Risky Choice” *The Journal of Risk and Uncertainty* 24:2, 103-130.

Markowitz, H.M. (1952), "Portfolio Selection," Journal of Finance, 77-91.

Sharpe, W.F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," Journal of Finance, 19, 425-442.

Tobin, J. (1958), "Liquidity Preferences as Behavior Towards Risk," Review of Economic Studies, 25, 65-87.

von Neumann, J., and Morgenstern O., (1944), Theory of games and economic behavior, Princeton University Press.

Vuong Q., (1989) “Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses”, *Econometrica* 57 307-333.