Experimental Investigation of Percieved Risk in Random Walk Processes

Gneezy, U.; Das, J.W.M.

Publication date:
1996

Citation for published version (APA):
Gneezy, U., & Das, J. W. M. (1996). *Experimental Investigation of Percieved Risk in Random Walk Processes*. (CentER Discussion Paper; Vol. 1996-85). Econometrics.

Experimental Investigation of Perceived Risk in Finite Random Walk Processes

Uri Gneezy and Marcel Das *

CentER for Economic Research and Dept. of Econometrics

Tilburg University

September 1996

Keywords: probability assessment, risk perception, heuristics

Abstract

The hypothesis that, on average, people accurately estimate probabilities in random walk processes is experimentally investigated. Individuals are confronted with a process that starts with \$$X$, and in every stage either goes up or down by \$1, with probabilities p and $1-p$ respectively. For different values of p, individuals were asked to estimate what is the chance that after 10 stages the system will be at a point higher than or equal to \$$X$. Systematic mistakes in estimations were observed. In particular, estimations were centered around the stage-by-stage probability (p) rather then around the actual probability. Implication of this result to random walk processes in finance is considered.

*We wish to thank Eric van Damme, Martin Dufwenberg, Bas Werker, and Jenke ter Horst for helpful comments.

Correspondence to: CentER for Economic Research, P.O. Box 90153, 5000 LE Tilburg, The Netherlands (e-mail: GNEEZY@KUB.NL)

1 Introduction

There is a large psychological literature concerning the way individuals estimate probabilities. In particular, many systematic biases are documented [see e.g. the book edited by Kahneman et al. (1982)]. The relevance of these studies to economic problems is not always clear. In this paper we consider probability estimation in random walk processes. These type of processes are of great importance in finance, since they are assumed to describe price changes in so-called efficient markets (see section 4.2). Clearly, systematic mistakes in probability estimation will lead to systematic mistakes in risk perception. This may be helpful in understanding observations from real markets, and in developing a better behavioral theory.

We experimentally investigate the following question:

How do people perceive risk in random walk processes?

In particular, we investigate whether people assess the probability of several outcomes after n periods correctly, and if not, are the mistakes systematically 'optimistic' (pessimistic), i.e. viewing the process as less (more) risky than it really is. At this point no explicit definition of risk is given, it will be clear in the context of the experiment.

We use the following setup:

An investor is given a stock that is worth \$X. Then a process of N stages begins. In every stage, the price of the stock either goes up or down by \$1, with probability p and $1-p$ *respectively. If the price of the stock reaches \$0 within N stages, then the stock will be worth \$0 for ever.*
Problem: What do you think is the probability that the stock will be worth at least \$X after N stages (p_X)*?*

We investigated this game with $p = 0.25, 0.33, 0.4, 0.6, 0.67$, and 0.75. In all treatments $N = 10$ and $X = 3$. 104 subjects participated, each presented with only one p. The subjects were awarded according to the accuracy of their answers (see the Appendix).

In the literature on probability assessment there is extensive discussion on problems connected with applications of Bayesian methods. Relatively little has been done about the practical problem of assessing prior distributions. Gneezy (1996) presented experimental evidence from a process with similar rules, that ends when the price of the stock reaches either \$0 or \$$n$ ($X < n$), i.e. there is no limit on the number of stages, and in principle the game may last forever. He tested this for $p > .5$, and found that individuals systematically under-estimated the probability of ending up with \$$n$ (p_n), rather then with \$0. In particular, 'anchoring' heuristics was found: estimations did not differ much from the stage-by-stage probability ('anchored' to p). The under-estimations seemed to be the result of the fact that when $p > .5$, $p_n > p$.

This paper deals with a finite horizon. We also extend the investigation to $p < .5$, in order to test whether, if estimation mistakes will be found, they will be in the same direction for $p < .5$ as for $p > .5$ (i.e. over-estimation or underestimations for all p's). This is important because the anchoring heuristics gives a different prediction, namely it suggests that for $p > .5$ individuals will under-estimate p_X (because in that range $p < p_X$), and for $p < .5$ individuals will over-estimate p_X (because in that range $p > p_X$).

A less stylized set-up is discussed in Staël von Holstein (1970, Ch.10). He conducted an experiment which concerned forecasts of the changes in buying prices over fourteen day periods for twelve shares quoted on the Stockholm stock exchange. The experiment was run for 10 sessions to enable a study of the effect of feedback. Subjects were asked to state their probabilities for the following five events: 1) the buying price decreases more than 3%, 2) the buying price decreases more than 1%, but at most 3%, 3) the buying price changes at most 1%, 4) the buying price increases more than 1%, but at most 3%, 5) the buying price increases more than 3%. In total 72 subjects participated, some of which were stock market experts, trained statisticians, and bankers. It turned out that a "naive" assessment strategy, based on historical data from the two years preceding the experiment, had an average successes which was better than 92% of the subjects. We refer to this study although it is basically different from ours. It shows that even experts that were given time to learn (the experiment repeated 10 times with the same subjects for different periods) do not make good prediction of outcomes in real life situations. Other studies about biases in the assessment of compound probabilities, are Bar Hillel (1973) and Wagenaar and Sagaria (1975).

The paper proceeds as follows. Section 2 states the hypotheses to be tested, illustrates the computation of p_X, and ends with describing the method used. Section 3 gives the results, and section 4 contains some discussion of the results.

2 Hypotheses, computation of p_X, and method

2.1 Hypotheses

A traditional assumption in finance is that even if, for bounded rationality reasons, not all individuals estimate probabilities accurately, there are no systematic mistakes in the estimations. So the benchmark hypothesis we use is:

H1: The median of the individuals' estimations is p_X.

Against this benchmark hypothesis, we test the hypotheses:

H2: For a given p, the majority of individuals over-estimate p_X.

H3: For a given p, the majority of individuals under-estimate p_X.

Finally, we test the anchoring hypothesis:

H4: The median of the answers is the stage-by-stage probability p.

2.2 Computation of p_X

To calculate p_X we define the *transition probability matrix* P with elements p_{ij}:

$$P = [p_{ij}], \quad i, j = 1, 2, \ldots, X + N + 1,$$

where the elements p_{ij} are given by

$$p_{ij} = \begin{cases} 1 & \text{if } i = j = 1, \\ p & \text{if } i = 2, \ldots, X + N, \text{ and } j = i + 1, \\ 1 - p & \text{if } i = 2, \ldots, X + N + 1, \text{ and } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

If we define $q \equiv 1 - p$, then P can be written as

$$P = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ q & 0 & p & & & \\ & q & 0 & p & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \\ & & & q & 0 & p \\ 0 & & \cdots & & q & 0 \end{bmatrix}$$

Note that it is enough to have $X + N + 1$ rows, since after N stages the stock can be worth at most $\$(X + N)$. The probabilities of reaching a value of $\$j$ $(j = 0, \ldots, X + N)$ after N stages, when the starting amount is equal to $\$X$, can be found in the $(X + 1)$-th row of $Q \equiv [q_{ij}] = P^N$. Finally, the p_X is then calculated by

$$p_X = \sum_{j=X+1}^{X+N+1} q_{X+1,j}.$$

2.3 Method

We fixed $X = 3$ and had 6 treatments, with the stage-by-stage probability (p) of 0.25, 0.33, 0.40, 0.60, 0.67, 0.75 in treatment $1, \ldots, 6$ respectively. Altogether 104 subjects participated: 18, 15, 18, 17, 18, 18 in treatment $1, \ldots, 6$ respectively. Subjects were first-year students in economics at the second semester of their studies. Most of them had participated successfully in a basic statistic course. After a short introduction, subjects received the instruction (see the Appendix for the instructions for treatment 1). The answers of subjects were evaluated and rewarded as described in the instructions. The experiment took 20 minutes (excluding the paying time).

3 Results

The answers given by the subjects are presented in Table 1.

Table 1 : Data for the six different stage-by-stage probabilities; p = stage-by-stage probability and p_X = compound probability.

p	0.25	0.33	0.40	0.60	0.67	0.75
p_X	0.07	0.19	0.33	0.78	0.89	0.96
nr. obs.	18	15	18	17	18	18
	.02	.04	.03	.35	.04	.04
	.05	.05	.04	.53	.33	.50
	.05	.08	.05	.55	.47	.50
	.06	.15	.05	.60	.50	.50
	.07	.15	.13	.60	.52	.62
	.10	.15	.20	.60	.60	.65
	.14	.20	.24	.66	.65	.65
	.15	.21	.25	.70	.65	.70
	.17	.22	.35	.75	.66	.73
	.19	.30	.38	.77	.67	.75
	.20	.49	.40	.80	.67	.80
	.25	.63	.48	.80	.68	.80
	.34	.65	.53	.80	.75	.83
	.85	.67	.65	.80	.80	.84
	.89	.90	.65	.80	.87	.90
	.90		.70	.85	.95	.90
	.93		.90	.85	.97	.90
	.99		.95		1.00	.99

We first test the benchmark hypothesis (H1). It is not wise to use a test based upon the mean of the observations, since this test will be very sensitive to outliers (in particular with this kind of number of observations, see e.g. Hampel et al., 1986). A sign test is an alternative. [1] Under H1, the number of individuals that under-estimates should equal the number of individuals that over-estimates the compound probability p_X. From Table 2 we can see that hypothesis H1 is rejected for $p = 0.25, 0.67$, and 0.75, and is not rejected for $p = 0.33, 0.40$, and 0.60 (we reject H1 when one of the two probabilities is lower than 2.5%). [2] We should note that for $p = 0.25$ there are some (unreasonable) high answers. This might be caused by a misunderstanding of the question. If we drop these observations from the analysis, the number of underestimations is still larger than the number of overestimation, but the result is not significant anymore.

When H1 is rejected, we then test whether H2 or H3 will be rejected. In the case of $p = 0.25$, we cannot reject hypothesis H2, but hypothesis H3 is rejected. For $p = 0.67$ and $p = 0.75$, H2 is rejected, but H3 is not rejected (see Table 2).

Table 2 : Significance probabilities corresponding to H1 (and H2 and H3).

X: number of respondents that underestimates p_X
r: realization

p	0.25	0.33	0.40	0.60	0.67	0.75
p_X	0.07	0.19	0.33	0.78	0.89	0.96
r	4	6	8	10	15	17
$P\{ X \leq r \}$	0.0245	0.304	0.407	0.834	0.999	1.000
$P\{ X \geq r \}$	0.994	0.849	0.760	0.315	3.77E-3	7.25E-5

Let's now consider the hypothesis H4. Again, we use a sign test: under H4, the number of individuals that give an answer below p equals the number of individuals that give an answer above p. From Table 3 we see that, for all treatments, we cannot reject hypothesis H4, i.e. estimations are centered around p (significance level is $2 \times 2.5\%$).

[1] Under the null hypothesis, the number of respondents that underestimates p_X follows a binomial distribution $B(n, q)$ with parameters n equal to the number of observations and q equal to $\frac{1}{2}$.

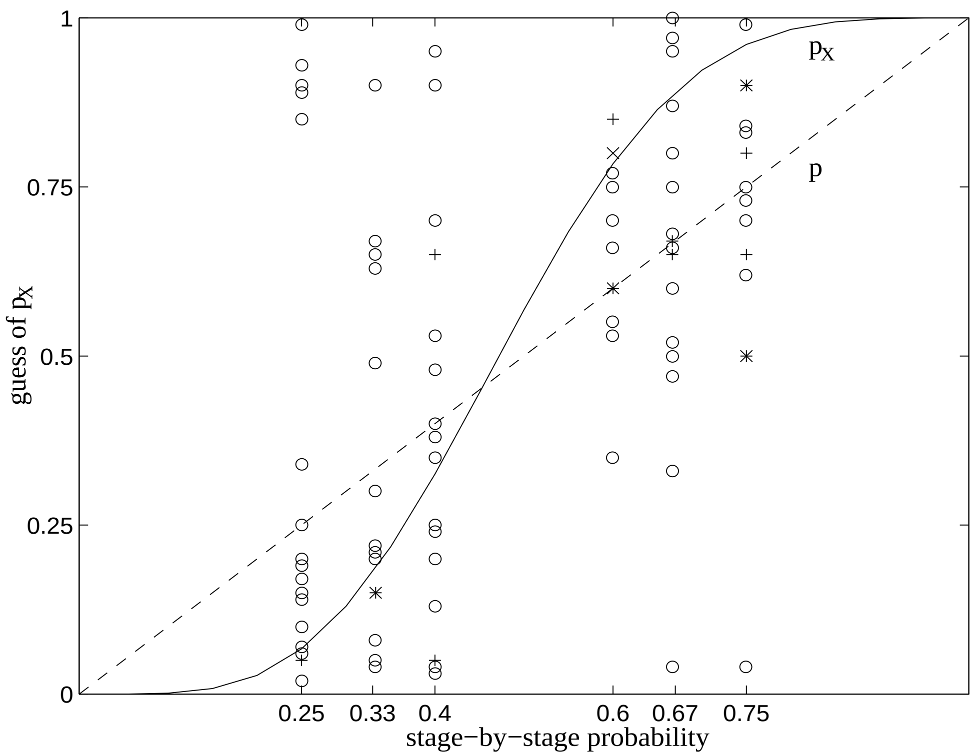
[2] For $p = 0.25$ one respondent gave the exact answer. In that case the test is based upon the answers different from p_X (conditional sign test).

Table 3 : Significance probabilities corresponding to H4.

X:	number of respondents that underestimates p					
r:	realization					
p	0.25	0.33	0.40	0.60	0.67	0.75
p_X	0.07	0.19	0.33	0.78	0.89	0.96
r	11	10	10	3	9	9
P$\{ X \leq r \}$	0.928	0.941	0.834	0.0287	0.773	0.685
P$\{ X \geq r \}$	0.166	0.151	0.315	0.994	0.402	0.500

In Figure 1 we graphically illustrate what we just found. The the guesses of p_X are plotted for each treatment, together with the actual p_X as a function of p.

Figure 1 : Graphical representation of the guesses of p_X for each treatment together with the stage-by-stage probability (p, dashed line) and the compound probability (p_X, solid line). The 'o' corresponds with one observation, '+' with two, '*' with three and 'x' corresponds with five observations.



As can be seen from this figure, for $p = 0.25, 0.33$, and 0.40, we have $p_X < p$, and for $p = 0.60, 0.67$, and 0.75 we have $p_X > p$. Since, as shown above, estimations are 'centered' around p and not around p_X, it is not surprising that in cases where $p_X < p$ the fraction of under-estimations is smaller than 0.5, and when $p_X > p$ it is larger than 0.5.

4 Discussion

It is difficult to extrapolate the experience from laboratory experiments to real life situations. For example, it may be that the deviation of estimations from the bench mark compound probability is due to the difficulty of the assessment task. It could also be that subjects were inexperienced in the task of quantifying judgment in probabilistic terms. However, the task in the experiment is certainly less difficult than in real life problems, such as risk assessment in stock markets. The fact that the stage-by-stage probability was given reduces one degree of difficulty compared with real markets. The fact that a short period was used ($N = 10$) reduces another. We showed that even though subjects were familiar with basic statistics, it seems that their knowledge was not quite useful to the problem presented in the experiment.

It is shown in the experiment that the fraction of under- estimations increases in the stage-by-stage probability. The median of the estimations for low p's lies above p_X, and the median of the estimations for high p's lies below p_X. Moreover, it is shown that for all tested p's, estimations are centered around p. This is argued to cause the systematic mistakes in the assessment of p_X.

Finally we will consider two points of view about the results presented in this paper, namely a psychological point of view and an economic point of view.

4.1 Psychological point of view

Many tasks in life demand the use of heuristics, rather than explicit calculations. As shown e.g. by Kahneman et al. (1982), people often use heuristics that lead to biases in assessments. In this paper we consider the assessment of probabilities in random walk processes, which are claimed to mimic some features of real financial markets. We finger out a heuristics that is used by people in assessing compound probabilities in these kind of set-ups, namely anchoring to the given stage-by-stage probability. This heuristics improved the estimations compared with having estimations that are uniformly distributed in the interval $(0, 1)$. An important observation is that although people use 'wrong' heuristics, it seems to us that in the process of teaching students, virtually all the attention is given to developing their skills in solving problems analytically. We believe that not enough attention is given to the development of 'heuristic skill', e.g. the marginal contribution for an MBA student of a course which tries to develop heuristic skills may be very high. The first step in developing such heuristic skills is to understand the heuristics people use, and when they should be improved. In this paper we indicate one important situation in which an improvement is needed.

4.2 Economical point of view

One of the most important hypotheses which has evolved from the research of financial markets, and has been empirically investigated, is the efficient markets hypothesis. According to this hypothesis, financial markets are 'efficient', and prices should reflect a rational forecast of the present value of future dividend payment. The efficient market hypothesis has also been traditionally associated with the assertion that future price changes are unpredictable, although stock prices have a positive drift, see De Bondt and Thaler (1989, p.189). The arbitrage forces are supposed to guarantee that prices adjust, and then move again, randomly, in response to unpredictable events. In a classical example, Fama (1965, p.98) writes: "It seems safe to say that this paper has presented strong and voluminous evidence in favor of the random walk hypothesis." Of course, this does not imply that all stocks follow the same random walk. The future prospects of stocks may still differ. These future prospects may create different random walk processes for different stocks. Looking n periods into the future, different processes typically imply different probability distribution over prices. Investors are assumed to have a preference relation over these probability distributions. A common assumption about this preference relation is that investors are risk averse, i.e. when comparing two processes with the same expected return, they prefer the stock that is "less risky". The state of the art today is the belief that only to a first approximation, financial markets follow a random walk. For an elaborate discussion of the efficient market and random walk hypothesis, see De Bondt and Thaler (1989), Fama (1991).

The assumption that returns on stocks follow a random walk with positive drift, can be compared with $p > .5$ in our study. If, in real markets, investors under-estimate the compound probability of a stock to come up ahead after a few periods (like we show in our stylized example), they will regard the stock as riskier than it really is. That may be a partial explanation of the equity premium puzzle. The puzzle refers to the fact that over the last century the risk-return relationship has been so much more favorable for stocks than for bonds, that unreasonably high level of risk aversion would be needed to explain why investors are willing to hold bonds at all (Mehra and Prescott, 1985). Could it be the case that it is not the level of risk aversion that is 'wrong', rather it is a case of misjudgment of risk? [3] It may be, as Arrow (1982) argues, that evidence of bounded rationality which is found in many cases in stylized experimental work, may teach us about risk perception in complex financial markets. It could also be the case that, like is commonly argued by economists, that even when individuals make systematic mistakes markets are not biased (see e.g. Camerer 1987). This we would like to investigate in future research.

[3] Note that this kind of irrationality does not allow for arbitrage against the irrational individual.

Appendix

Your name: ....

Welcome to our experiment in decision theory. In the experiment you will be presented with a problem, and asked to estimate the chance of a certain outcome. The more accurate your estimation will be, the more money you will earn. After you will finish answering we will pay you according to the following rule:

You will start with f 20, and for every 1% of mistake, f 1 will be deduced from your payoff. The mistake is the absolute value of [your guess (in percentages) minus the actual chance].

For example, if you will guess accurately, you will get f 20. If you will make a 10% mistake (either over-estimate or under-estimate), you will get f 10. If your mistake will be bigger or equal to 20% you will not be paid at all.

The problem is based on the following: Mr. X is given a stock that is worth f 3 today. He will hold the stock for 10 years. In each year, the price of the stock either goes up or down by f 1. The chance of it going up is 2/3 (i.e. 67%), and the chance of it going down is 1/3 (i.e. 33%). So, after the first year, the stock will be worth either f 2 (with 33% chance) or f 4 (with 67% chance). In the second year, again, the price of the stock will either go up or down by f 1, with the same chances, and so on. If the price of the stock will reach f 0 within the 10 years, then the stock will be worthless for all future periods. Otherwise, Mr. X will sell the stock after 10 years.

The problem :

What do you think is the chance that the price of the stock after 10 years will be at least f 3? ....%

After all of you will finish answering, we will collect and check the answers, and pay you as described above.

Do you have any questions?

References

Arrow, K.J. (1982), Risk Perception in Psychology and Economics, *Economic Inquiry*, 20, pp. 1-9.

Bar-Hillel, M. (1973), On the Subjective Probability of Compound Events, *Organizational Behavior and Human Performance*, 9, pp. 396-406.

Camerer, C. (1987), Do Biases in Probability Judgment Matter in Markets? Experimental Evidence, *American Economic Review*, 77 pp. 981-997.

De Bondt, W, and R.H. Thaler (1989), Further Evidence on Investor Overreaction and Stock Market Seasonality, *Journal of Finance*, 42, pp. 557-581.

Fama, E.F. (1965), The Behavior of Stock Markets Prices, *Journal of Business*, 38, pp. 34-105.

Fama, E.F. (1991), Efficient Capital Markets: II, *Journal of Finance*, 46, pp. 1575-1617.

Gneezy, U. (1996), Probability Judgements in Multi-Stage Problems: Experimental Evidence of Systematic Biases, forthcoming in *Acta Psychologica*.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986), *Robust Statistics*, Wiley, New York.

Kahneman, D., P. Slovic, and A. Tversky (1982), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Mehra, R., and E.C. Prescott (1985), The Equity Premium Puzzle, *Journal of Monetary Economics*, 15, pp. 145-61.

Staël von Holstein, C.-A. (1970), *Assessment and Evaluation of Subjective Probability Distributions*, The Economic Research Institute at the Stockholm School of Economics.

Wagenaar, W.A., and S.D. Sagaria (1975), Misperception of Experimental Growth, *Perception and Psychophysics*, 18, pp. 422-426.