CESifo Working Papers

www.cesifo.org/wp

Hidden Skewness: On the Difficulty of Multiplicative Compounding under Random Shocks

Ludwig Ensthaler
Olga Nottmeyer
Georg Weizsäcker
Christian Zankiewicz

CESIFO WORKING PAPER NO. 4760
CATEGORY 13: BEHAVIOURAL ECONOMICS
APRIL 2014

CESifo *Working Paper No. 4760*

Hidden Skewness: On the Difficulty of Multiplicative Compounding under Random Shocks

Abstract

Multiplicative growth processes that are subject to random shocks often have a skewed distribution of outcomes. In a number of incentivized laboratory experiments we show that a large majority of participants either strongly underestimate skewness or ignore it completely. Participants misperceive the outcome distribution's spread to be far too narrow-band and they estimate the median to lie too close to the distribution's center. The observed bias in expectations is irrespective to risk preferences and fairly robust to feedback. It is consistent with a behavioral model in which geometric growth is confused with linear growth. The misperception is a possible explanation of investors' difficulties with real-world financial products like leveraged ETFs.

JEL-Code: C910, D030, D140, G020.

Keywords: behavioral economics, irrational expectations, binomial tree.

Ludwig Ensthaler
Humboldt University & WZB
Berlin / Germany
ludwig@ensthaler.com

Olga Nottmeyer
IZA Bonn
Germany
nottmeyer@iza.org

*Georg Weizsäcker**
Humboldt University & DIW
Berlin / Germany
weizsaecker@hu-berlin.de

Christian Zankiewicz
DIW Berlin
Germany
czankiewicz@diw.de

*corresponding author

April 7, 2014
We thank Erik Eyster, Dorothea Kübler, Erik Mohlin, Peter Mörters, Tobias Schmidt, Adam Szeidl and Heinrich Weizsäcker and audiences at DICE, DIW Berlin and ESMT for helpful comments and we thank colleagues at the decision laboratories of Technical University Berlin and University College London, especially Brian Wallace and Mark Henninger, for their excellent contributions in the preparation and conduct of the experiments. Financial support by the ERC (Starting Grant 263412) is gratefully acknowledged.

1 Introduction

Many household investors face a particular mismatch in the time frames of asset return evaluations. They acquire their most important financial assets with the intention to liquidate them in the relatively distant future but the available return information concerns a much shorter investment interval such as a year. Real estate investments, retirement savings plans or investments in college funds typically share this feature. In order to evaluate the odds of the investment yielding the desired return on the planned (or any plausible) distant selling date, an investor needs to extract the price distribution at the selling date by compounding the available short-term return distributions–a formidable task for the average person. The discrepancy between the two relevant time horizons can be very large and the importance of compounding therefore substantial: for many retirement savings, the 30-year performance is relevant but only information about the 1-year retirement fund is easily available. But mismatches of this flavour occur also for most shorter-term households investments.

It is well documented that decision-makers both in the laboratory and in the field have difficulties when compounding deterministic growth processes–the so-called exponential growth bias. When asked to estimate the final value of an account that accumulates 7% interest for ten years a substantial fraction of respondents give an answer that is closer to 70% than to the actual 97%. The analysis of Stango and Zinman (2009) indicates that the bias is empirically relevant as it affects households' borrowing and saving decisions.

But the cognitive errors in compounding may have another, even uglier face. There is strong anecdotal evidence that cognitive problems matter especially in settings where growth is not deterministic but random. In random processes the decision maker needs to generate a probability distribution over all possible random paths, which can lead to new misunderstandings. As an example consider a relatively new class of retail financial products, so-called leveraged exchange-traded funds (leveraged ETFs). These assets move by a given multiple relative to an underlying asset, compounded at the end of each trading day. A triple leveraged ETF on the Dow Jones Industrial Average increases by three per cent on a trading day if the DJIA increases by one per cent on that day and it falls by three per cent if the DJIA falls by one per cent. Leveraged ETFs have come under severe scrutiny shortly after their introduction as many investors were perplexed when the products made a loss in a period where the underlying index made a gain. This can occur even after only a few days and is a frequent event. Regulatory units and the financial media have responded by extensive warnings that involve explanations of these counter-intuitive possibilities.

This paper presents a series of incentivized laboratory experiments that extend the research on the multiplicative growth bias to the stochastic domain and test the decision makers' ability to compound a sequence of multiplicative random shocks. More precisely, we provide our participants with the one period return distribution of an asset and give incentives to compute the distribution of its selling price after a specified number of periods. We find that participants' estimates are biased and deviate systematically from the rational prediction. Generally, participants perceive a distribution that is too symmetric, i.e., they ignore or significantly underestimate the distribution's skew. This effect is stronger for longer periods and for more volatile assets.

Our experimental results are largely in line with the predictions of a simple model of misperception of compounding of shocks. This model, which we label "linearity bias model", extends the hypothesis that the agent fails to do deterministic compounding to stochastic settings. It stipulates that a biased decision maker perceives a linear evolution in the sense that she perceives the distributions of absolute changes as constant over time, instead of the relative changes being constant over time. In effect, all multiplicative growth is mistaken as additive growth with a constant distribution of increments. The model thereby straightforwardly produces several testable hypotheses. In a deterministic setting, the agent expects a linear growth path. In stochastic settings, the agent ignores all skewness that arises from multiplicative compounding. If the distribution of an asset's per-period returns is symmetric, then the agent perceives all longer-period returns as symmetric, too. Moreover, the linearity bias model predicts larger deviations from the rational benchmark if the investment horizon is long relative to the length of a single period and if the volatility of returns is high. The deviations are predicted to occur in terms of misperceptions of the median, the spread, and the skewness.

We start our analysis with a stylized experiment. There, participants have to estimate the median of an asset that increases in value by 70% or decreases in value by 60% in every period, each with a chance of one half. The asset must be held for twelve periods and is then sold. We find that, as is predicted by the linearity bias model, participants overestimate the median of the resulting distribution of prices significantly and strongly: the large majority of participants overestimate the median by an order of magnitude.

We then test the predictions of the model in more realistic settings. We use both artificial asset descriptions as well as actual historical data on the German DAX index to test for a participant's perception of the median as well as the 10th and 90th percentile of the distribution of long-run investments.

The findings confirm that skewness is underestimated by most participants. In all relevant treatments, the majority of participants reveal a perceived skew of the outcome distribution that is below the rational prediction. These findings are highly significant in treatments where the considered asset is more volatile or the investment horizon is longer. About ninety percent of participants underestimate the skew, about ninety percent underestimate the spread and about eighty percent estimate the median to lie too close to the distribution's center. All findings are irrespective of risk preferences: the experimental design gives participants a list of decisions between two investments, each involving a probability of winning a fixed bonus. This renders risk preferences irrelevant.

The rest of this paper is organized as follows. Section 2 briefly discusses related literatures. Section 3 introduces the setting and experimental design by way of reporting the shorter, stylized experiment. Section 4 describes the behavioral model and relevant implications. Section 5 explains the main experiment in detail and reports the experimental findings and their statistical analysis. Section 6 concludes.

2 Review of related literature

Classic studies in cognitive psychology discuss quite extensively whether or not the human cognitive apparatus is able to correctly account for the distinction of linear versus nonlinear relations between relevant variables. Wagenaar and Sagaria (1975) ask participants to predict an exponential data series representing an index for pollution. They find that participants strongly underestimate exponential growth. Wagenaar and Sagaria (1978) show that underestimation of exponential growth is robust to the amount of information available to the participants and Wagenaar and Sagaria (1979) show that the effect is robust to the framing of the information. Kemp (1984) surveys perceptions of changes in the cost of living. Respondents systematically underestimate the increase in cost, which is also in line with a misperception of exponential growth. Much of the early data analysis uses responses to quiz-type questions, but a subsequent specialization of this literature more and more focuses on economic contexts, like the perception of of compound growth from interest or loan payments. Eisenstein and Hoch (2005) and Stango and Zinman (2009), among others, document that participants underappreciate the effects of compound interest and thereby predictably underestimate the compound effect of growth. Chen and Rao (2007) show that retailers can strategically use this bias by posting double dip price discounts (a discount of 20% followed by another 25% discount is perceived to be a 45% reduction, not the actual 40%). As described in the introduction, our paper can be viewed as an extension of this literature to non-deterministic growth processes.

An important predecessor of our paper is the study by Benartzi and Thaler (1999) who, among other things, also study biases in the compounding of the long term distributions from a given short term distribution. Their experimental participants choose different hypothetical retirement plans depending on whether they are given the historical return distribution of retirement plans for a one year period or a 30 year period. Benartzi and Thaler (1999) relate this bias to the effects of myopic loss aversion (see also Samuelson, 1963, Redelmeier and Tversky, 1992, Gneezy and Potters, 1997, Klos, Weber, and Weber, 2005). While we agree that myopic loss aversion likely plays a role in households' long term investment decisions, our experiments suggest that household decisions can also be misguided by a biased perception of the underlying growth processes.[1] This is also consistent with the hypothetical investment choice experiments in Stutzer and Jung Grant (2010) who find an inflated investment rate in those treatments where their participants have to calculate the compound return by themselves.[2]

Another related literature studies whether experimental participants have a correct understanding of financial options. We refer the reader to Gneezy (1996) and Abbink and Rockenbach (2006) for previous results in this–surprisingly small–literature. We note that the assets that we construct have the same structure as the underlying asset in the well-known model of Cox, Ross, and Rubinstein (1979) of European call options. A consistent finding of misperceptions of such assets may therefore indicate

[1] A relevant distinction between our study and the existing experimental work on myopic loss aversion is that the existing papers largely make use of additive growth processes.

[2] The experiment by Stutzer and Jung Grant (2010) uses a quite similar experimental wording as the experiment described in Section 3 and in our paper's first version (Ensthaler et al, 2010) despite having been developed and written independently.

a potential mispricing. This is not further studied in our paper, which focusses on investments in the underlying asset itself.

3 A Stylized Experiment to Start With

The experiment follows the binomial tree model of Cox, Ross and Rubinstein (1979). Participants consider a security that has a seemingly simple price transition. They are told that they can buy the security at a price of 10,000 Euros. During each month, the security's price either increases by 70% or decreases by 60%. The two possible price changes in each month occur with equal probabilities. All random draws are independent. The security, if bought, has to be held for exactly 12 months and is then to be sold.

A decrease by 60% cannot be undone by a single increase by 70% and therefore the typical price path tends downward. Already with a fixed maturity as short as 12 months, the median selling price is as low as 989 Euros. The mean selling price after 12 months is much higher than the median, at 17,959 Euros. Thus, the outcome distribution is heavily skewed.

Our stylized experiment tests whether participants correctly locate the median. Through a sequence of simple choice problems we identify bounds on the median of each participant's subjectively expected distribution. The instructions differ by treatment condition only in the degree to which they show the values of any compound price changes that accumulate over time. Participants may thus misperceive the random price process to a greater extent in one treatment condition than in in the other.

3.1 Experimental Design

Choice problems: As indicated above, the experiment elicits the participants' expectations irrespective of their risk preferences. The monetary incentives achieve this property by using only two possible payments in each choice problem–"receive a bonus" versus not–making it optimal for any participant with monotonic preferences to maximize the subjectively perceived probability of receiving the bonus.

The choice problems come with a financial investment context: in each round of the experiments, two risky securities are on offer and the selling price of the chosen security determines whether or not the participant receives the bonus.[3] Security A is the security described above, with a 70% / -60% price change in each of 12 periods. A participant who chooses this security receives the bonus if the selling price at maturity exceeds a given threshold t_A. The alternative choice is Security B, which yields the bonus with probability one half. One can immediately see that it is subjectively optimal for a participant to choose Security A if and only if she believes that Security A yields the bonus with

[3] The descriptions begins with the wording: "You are a manager and have to make a decision between two risky investments."

probability more than one half. A choice for Security A thus reveals that the median of her subjective probability distribution of Security A's selling price is above t_A.

Security B's description is analogous to that of Security A, with the difference that only a single price change of +70% or -60% (equiprobably) occurs during the 12 months. A participant who chooses Security B receives the bonus if the selling price of B exceeds a separate threshold t_B. This threshold is fixed at the initial price of 10,000 Euros throughout the experiment whereas the threshold t_A varies between 10 different values. Each experimental participant makes a choice between A and B for each of the 10 possible values of t_A, allowing us to infer bounds on her subjective median of the selling price of Security A. Table 1 lists the 10 choice problems (Task 1, Task 2, etc.) as seen by the participants. Given that the true median of Security A's selling price is 989, the rational prediction is for the participants to chose A in Tasks 1 and 2 and to choose B in all subsequent tasks.

	Thresholds for Security A	Thresholds for Security B	Your decision (A or B)
Task 1	100	10,000	_
Task 2	500	10,000	_
Task 3	2,000	10,000	_
Task 4	6,000	10,000	_
Task 5	9,000	10,000	_
Task 6	12,000	10,000	_
Task 7	20,000	10,000	_
Task 8	35,000	10,000	_
Task 9	90,000	10,000	_
Task 10	250,000	10,000	_

Table 1: The 10 binary choices.

Treatment conditions: Participants are randomly assigned to one of two conditions that differ in the extent to which the experimental instructions explain the implied distributions. The NO_HELP condition presents the basic explanation. To introduce Security A, the instructions first describe the price transition rule using simple language. This is followed by a statement about the independence of random draws and by the paraphrase that after month 1, the security's price is either at 17,000 Euros or at 4,000 Euros. The instructions then repeat the random price transition, but without calculating compound effects explicitly: "At the end of month 2, the price is either 70% higher or 60% lower than at the end of month 1. At the end of month 3, the price is either 70% higher or 60% lower than at the end of month 2. And so on, ..." Security B is described with identical wording to that of Security A, where applicable. Next, the thresholds t_A and t_B are explained and two examples are given. Finally, participants face an understanding test of four questions which they have to answer correctly before they may proceed. The examples and understanding test are carefully chosen to not suggest any responses to the participants.

A possible concern is that data patterns in the NO_HELP condition might not be due to cognitive limitations but instead are driven by the choice format, the context frame or other cues. In particular, the set of 10 threshold values can conceivably influence the responses.[4] We address this concern by including the HELP condition where we provide the participants with an additional explanation, leaving the remainder of the instructions unchanged. The additional text (about one written page) gives an explicit calculation of the distribution of compound price changes after two periods. It also points out the asymmetry in the selling price distribution and lists the implicit probabilities of receiving the bonus from choosing Security A for each value of t_A. None of the explanations in HELP adds any substantive information relative to the descriptions in NO_HELP. The only difference is that the relevant distributions are explicit in HELP and implicit in NO_HELP. Any difference in responses under the two conditions must stem from differences in the understanding of these implied truths.

Feedback and repetitions: After the participants make their 10 choices, they receive individual feedback in the form of a sample pair of selling prices of Securities A and B. This concludes the first round of the experiment. The experiment is then repeated for four additional rounds of the same nature, each including 10 choices and individual feedback. The feedback procedure and the choice format are identical for both treatment conditions.[5]

Procedures and payments: All 128 participants (68 in NO_HELP and 60 in HELP) are students at Technical University Berlin. Six sessions, three in each treatment condition, are conducted in a paper-and-pencil format. The protocol is fixed across all sessions. The instructions are read aloud to the participants up to the beginning of the understanding test. Participants receive a participation fee of 5 Euros and a possible bonus of 5 Euros per round. That is, participants can earn up to five bonuses of 5 Euros each, one per round of the experiment. After completing all choices, each participant receives five random draws of integers between 1 and 10 to determine which of the 10 choice problems in each round is payoff relevant for her. She receives the bonus for a given round if the selling price of the chosen security in the payoff-relevant problem exceeds its threshold.

3.2 Results

The data analysis is simplified by the observation that a participant with any subjective belief about selling prices maximizes her preference by choosing Security A for low values of t_A and switching to B for all values higher than her subjective median, i.e. she switches between the securities no more than once. We observe such unique switching points in the large majority of responses (93%) and restrict attention to these data.[6]

[4]We deliberately fixed the 10 values of t_A so that half of them exceed Security A's starting price of 10,000 Euros, in order to not suggest a direction of price change. However, this property may conceivably induce a midpoint effect, leading the participants to switch from A to B towards the middle of the list.

[5]Each additional round comes with the chance to earn a new bonus (see the next paragraph in the main text), but this does not affect the simple optimality conditions for choice. Independent of other choices it remains optimal to choose A iff the subjective median is above t_A, under a wide set of preferences for choice under uncertainty.

[6]If a participant has multiple switching points in one round, her answers in the remaining rounds are still considered. None of our conclusions would change if we dropped all responses by participants who switch strictly more than once in at least one round (12% of participants), or if we included all data and considered each of the 10 tasks separately.

Range of subjective median for Security A			Share of participants switching from A to B NO_HELP	HELP
[0	–	100)	0.018	0.000
[100	–	500)	0.000	0.000
[500	–	2,000)	0.000	0.703
[2,000	–	6,000)	0.036	0.109
[6,000	–	9,000)	0.107	0.063
[9,000	–	12,000)	0.411	0.063
[12,000	–	20,000)	0.196	0.031
[20,000	–	35,000)	0.179	0.016
[35,000	–	90,000)	0.054	0.000
[90,000	–	250,000)	0.000	0.000
[250,000	–	∞)	0.000	0.016

Table 2: Subjective medians in round 1.

Table 2 lists the implied ranges for the medians of the participants' subjective distributions of Security A's selling price in round 1. Not a single NO_HELP participant reveals a subjective median between 500 and 2,000 Euros (i.e. rational switching at Task 3). Instead, 98% of NO_HELP participants in round 1 reveal that their subjective medians are above 2,000 Euros, and 86% do so still in round 5.[7] The modal choice in round 1 indicates a subjective median between 9,000 and 12,000 Euros in the NO_HELP condition. Under the HELP condition, 70% of responses are at the optimal switching point of Task 3 already in round 1. Altogether, the data show strong differences between the two conditions, and parametric t-tests as well as non-parametric Wilcoxon rank-sum tests confirm that all round-by-round comparisons between the two conditions are statistically significant at $p < 0.001$. In particular, the treatment effects are still highly significant in the last round of the experiment and thereby exhibit a consistent pattern that the performance is poor under the NO_HELP condition and much better in HELP. This result is confirmed by random effects regressions (not included in the paper) that exploit the panel structure of the data. They show a significant treatment effect that is robust to changes in the regression model specification.[8]

To summarize, almost all NO_HELP participants exhibit a sizable misperception of the median when confronted with a seemingly simple stochastic growth process. The effect largely vanishes in the HELP condition and we thus conjecture that the underlying bias is the participant's failure to correctly compound the multiplicative growth process. The deviation from optimality is so strong that much of the evidence is consistent with the simple hypothesis that participants falsely perceive a linear growth process, which is spelled out in the next section. Under this hypothesis, they would miss out entirely on the skewness of the relevant distributions.

[7]See the Appendix A.1 for results of rounds 2 to 5.

[8]The appropriateness of random effect regressions is confirmed by applying a Hausman test. See Ensthaler et al (2010) for regression results.

4 Theoretical Considerations

Following the data patterns of the stylized experiment above, we model a decision maker who suffers from linearization bias (LB decision maker) as follows. The LB decision maker correctly computes the distribution of all relevant random changes that occur in the first period. However, she is biased in the sense that she uses linear extrapolation to extend the distribution of absolute changes to all remaining periods. Formally, let Y_0 denote the initial price of an asset and let μ_t be the random variable describing the relative price changes occurring in t, e.g. $Y_1 = Y_0\mu_1$. The random variable describing the absolute price movements in t is captured by η_t, e.g. $Y_1 = Y_0 + \eta_1$. As in the basic model of Cox, Ross and Rubinstein (1979) we consider a price movement where μ_t has a distribution that is constant across t. The LB decision maker instead views the distribution of absolute changes η_t as constant across t. That is, where an unbiased decision maker perceives the true distribution as $Y_T = Y_0 \prod_{t=1}^{T} \mu_t$, with $\{\mu_t\}$ i.i.d., the LB decision maker perceives the final price as $\tilde{Y}_T = Y_0 + \sum_{t=1}^{T} \eta_t$, with $\{\eta_t\}$ i.i.d. and its distribution equal to that of η_1.

The LB decision maker misses out on all effects of multiplicative compounding.[9] In particular, the model is consistent with the failure to engage in compound interest accumulation that is documented in the literature (e.g. Christandl and Fetchenhauer, 2009, Stango and Zinman, 2009). Moreover, the LB decision maker ignores all skewness arising from compounding: if η_1 is symmetrically distributed, $\tilde{Y}_T$ is symmetrically distributed, too, because symmetry is preserved under addition of random variables.

Just as Cox, Ross and Rubinstein (1979), we assume that μ_t is a binary random variable with a constant 50-50 chance of moving up or down, i.e. $\mu_t \in \{\mu^h, \mu^l\}$ where the percental uptick $\mu^h \geq 0$ and the percental downtick tick $\mu^l \geq 0$ are equiprobable in each t. For the propositions that follow, we focus on the case that the processes' mean and median lie on different sides of the starting value. (Only Proposition 2. (ii) considers a different case, as point of comparison.) We therefore assume that $\mu^h\mu^l \leq 1$ (median lies below starting value) and $\mu^h + \mu^l > 2$ (mean lies above starting value).[10] For analytical convenience we consider an even-numbered length of the investment period, $T = 2n$ and $n \in \mathbb{N}$. We couch the model's implications in the propositions below, which serve as foundations for later hypothesizing. The propositions focus in turn on the LB decision maker's perception of the following distribution characteristics: the median (Proposition 1), spread (Proposition 2) and skewness (Proposition 3). All proofs are in Appendix B.1.

Proposition 1. *Let $q_{0.5}$ denote the true median of Y_T, and let $\tilde{q}_{0.5}$ denote the median of $\tilde{Y}_T$ as perceived by the LB decision maker. We find that:*

(i) The LB decision maker overestimates the median, $\tilde{q}_{0.5} > q_{0.5}$. Also, $lim_{T\to\infty} \frac{q_{0.5}}{\tilde{q}_{0.5}} = 0$ and $\frac{q_{0.5}}{\tilde{q}_{0.5}}$

[9] To see this formally, let e.g. $Y_T = Y_0(1+r)^T$, where r is the annual rate of interest. The LB decision maker perceives a constant additive increase instead, $\overline{Y}_T = Y_0 + T\overline{r}$ where $\overline{r} = Y_1 - Y_0$. For $r > 0$ and $T > 1$, we have $\overline{Y}_T < Y_T$, consistent with the available lab and field evidence.

[10] Mean and median generically differ in stochastic multiplicative growth processes like those described by Cox, Ross and Rubinstein (1979). We restrict attention to the case that mean and median lie on different sides of the starting value because it naturally arises in many growth processes and illustrates the possible misperceptions well.

strictly decreases in T.

(ii) Reducing the per-period volatility reverses the first statement of (i): There exists a $z^* \in (0, \frac{\mu^h-\mu^l}{2})$ such that for all $z \in (z^*, \frac{\mu^h-\mu^l}{2})$, changing the uptick from μ^h to $\mu^h - z$ and the downtick from μ^l to $\mu^h + z$ induces the LB decision maker to underestimate the median.[11]

Proposition 2. *Analogous to the notation in Proposition 1, let q_p denote the true p-percentile of the process, and let $\tilde{q}_p$ denote the p-percentile as perceived by the LB decision maker. For every $\delta \in (0.25, 0.5)$, the LB decision maker underestimates the spread of size 2δ for small T and overestimates the same spread for large T:*

(i) $\tilde{q}_{0.5+\delta} - \tilde{q}_{0.5-\delta} < q_{0.5+\delta} - q_{0.5-\delta}$ *for $T = 2$.*

(ii) $\tilde{q}_{0.5+\delta} - \tilde{q}_{0.5-\delta} > q_{0.5+\delta} - q_{0.5-\delta}$ *for T sufficiently large.*

Proposition 3. *The LB biased decision maker ignores the skew, i.e., perceives a perfectly symmetric distribution.*

5 The Experiment

In a sequence of experimental treatments we test the theoretical predictions of Section 4. Besides literal testing of the main propositions, we also ask whether the predicted effects disappear in an environment that is approximately deterministic. Moreover, we test the main predictions in a more realistic context frame.

There are 6 treatments. Participants consider either an artificially constructed asset (4 treatments) or an asset that we construct using real-world data from (leveraged) Exchange Traded Funds (2 treatments). We elicit the locations of the participants' 10th percentiles, the medians and the 90th percentiles for the relevant asset prices. Just as in the introductory experiment of Section 3, we identify bounds on these quantiles of the subjectively expected distribution of each participant through a sequence of investment problems. This is detailed in Section 5.1. Moreover, where applicable, we elicit the participants' perceived probabilities of the assets returning a profit. Section 5.2 reformulates the theoretical predictions as behavioral hypotheses that relate directly to the available data. Section 5.3 shows a summary of the data. Section 5.4 formally tests the hypotheses using Bayesian inferences and describes the results of the elicitation profit probabilities.

5.1 Experimental Design

Variables of interest: Following the theoretical considerations we examine three properties of a participant's subjective distribution of an asset return: the median, the spread and the skewness. A

[11] The assumption that the median lies below the starting value does not hold for $z > z^*$. All other assumptions are maintained also in this case.

participant's perception of all three can be approximated by studying the participant's 10th, 50th and 90th percentiles. We measure, as described above, merely intervals that contain the percentiles. This leaves some room for variation in the data analysis. (Yet, the intervals are sufficiently tight to potentially reject precise hypotheses.) In the main text, we only discuss results that use the elicited lower bounds of the relevant percentiles:[12]

$\underline{q}_{0.1,i}^{a,k}$: the revealed lower bound of participant i's 10th percentile of asset a in treatment k

$\underline{q}_{0.5,i}^{a,k}$: the revealed lower bound of participant i's median of asset a in treatment k

$\underline{q}_{0.9,i}^{a,k}$: the revealed lower bound of participant i's 90th percentile of asset a in treatment k

The following paragraphs explain how we generate these variable. As measures of a perceived distribution's spread and skew, we construct:

$$spread_i^{a,k} = \underline{q}_{0.9,i}^{a,k} - \underline{q}_{0.1,i}^{a,k}$$

$$skew_i^{a,k} = \frac{1}{3}\sum_{j\in\{0.1,0.5,0.9\}}(\underline{q}_{j,i}^{a,k} - \bar{q}_i/sd_i)^3,$$

with $\bar{q}_i$ and sd_i denoting the mean and the standard deviation of participant i's three responses $\underline{q}_{0.1,i}^{a,k}$, $\underline{q}_{0.5,i}^{a,k}$ and $\underline{q}_{0.9,i}^{a,k}$, respectively. Finally, we elicit the subjectively perceived probability of the asset making a profit, as detailed below.

Treatment conditions: The instructions in all treatments give the same basic explanation of an investment situation, analogous to the NO_HELP condition in the experiment of Section 3. We employ a between-subject design and randomly assign the participants to one of six treatments that differ in the properties of "Security A". In four treatments Security A is an artificially created asset that mimics the theory of Section 4. In the remaining two treatments it is a (more realistic) DAX-based ETF. In all treatments participants can buy the security at a price of $£100$. If they buy it they have to sell after T^k periods, where k indexes the treatment.

The first two treatments specify a high level of volatility such that the price moves by about 20 percent in each period: In treatments *High Volatility_Short* (HVS) and *High Volatility_Long* (HVL), the parameters specifying upticks and downticks are $\mu^{h,HVS} = \mu^{h,HVL} = 1.212$ and $\mu^{l,HVS} = \mu^{l,HVL} = 0.811$. The sole difference between these two treatments is in the lengths of time periods until maturity: $T^{HVS} = 14$ and $T^{HVL} = 140$.

[12]Almost all results are also replicated using upper bounds of the relevant intervals and are reported in Appendix A.3.

The other two treatments that involve artificially constructed assets are *Low Volatility_Short* (LVS) and *Low Volatility_Long* (LVL). Their price volatility is much lower at $\mu^{h,LVS} = \mu^{h,LVL} = 1.012$ and $\mu^{l,LVS} = \mu^{l,LVL} = 1.011$. Hence, the asset price motion is approximately deterministic in these treatments. The number of time periods until maturity is again set at $T^{LVS} = 14$ and $T^{LVL} = 140$, respectively.

The two remaining treatments study ETFs that are based on the German stock index DAX30. The two treatments differ only in that their respective versions of Security A differ in per-period volatility. In treatment *ETF_3*, the relevant security is a triple-leveraged DAX ETF. Its price changes, on each trading day, by three times the daily percentage changes of the underlying index DAX30. In treatment *ETF_1*, in contrast, Security A is simply the DAX ETF itself. The number of time periods until maturity of the ETF is specified at $T^k = 2000$ trading days for both $k = ETF\_3$ and $k = ETF\_1$. To generate realized price paths, we use for each participant a separate historical realization of 2000 consecutive DAX30 values. These are drawn randomly from the time period 1964 to 2012.[13] As in the other four treatments, participants can buy the ETF at a price of $\pounds 100$ and have to hold it until maturity.

Elicitation of subjective quantiles: As in the introductory experiment presented in Section 3, in all six treatments a participant who chooses Security A receives a fixed bonus if the selling price at maturity exceeds a given threshold t_A. These thresholds differ between treatments for the artificial assets and are held constant between ETF treatments. Table 3 lists the respective thresholds of Security A.

The alternative choice option is Security B which yields the bonus with a certain probability. To elicit three different quantiles, Security B has three different specifications. Each participant faces each specification once. Security B1 yields the bonus with 10% probability, B2 with 50% and B3 with 90%. Accordingly, each participant faces three choice lists. First, she chooses between Securities A and B1 for the different thresholds of Security A. This allows us to infer bounds on her subjective 10th percentile of the Security A's selling price. For example, suppose that participant i in treatment HVS chooses Security A over Security B1 in Tasks 1 and 2 and chooses Security B1 over Security A in Tasks 3 through 10. Inspecting Table 3 we see that this is subjectively optimal iff participant i's subjective 10th percentile for Security A's selling price is between 30 and 45. In this example, we would thus assign the lower bound as $\underline{q}_{0.1,i}^{A,HVS} = 30$. As her second set of tasks the participant faces the analogous choices between Securities A and B2 (with the same list of thresholds for Security A). This allows us to infer a lower bound on her subjective median, $\underline{q}_{0.5,i}^{A,HVS}$. Finally, she faces the analogous list of choices between Security A and B3, allowing us to infer $\underline{q}_{0.9,i}^{A,HVS}$.

[13]The instruction in the ETF treatments are analogous to the other four treatments. Additionally, they supply participants with general information about the DAX30 and a data summary of daily DAX30 movements in the relevant time period. The information is given in the form of a histogram as well as statements specifying the 90% confidence interval ([-1.8%,1.8%]) and the overall average of daily percentage changes (0.03%).

	Values of t_A *LVS*	Values of t_A *LVL*	Values of t_A *HVS*	Values of t_A *HVL*	Values of t_A *ETF_1 / ETF_3*
Task 1	104.0	185	15	2	30
Task 2	104.5	210	30	5	60
Task 3	105.5	240	45	15	90
Task 4	107.0	290	65	60	140
Task 5	109.0	340	95	140	200
Task 6	111.5	400	125	230	260
Task 7	114.5	460	155	350	330
Task 8	118.0	520	190	550	450
Task 9	122.0	625	225	700	650
Task 10	126.5	850	265	1,000	1,000
Task 11	-	-	-	-	1,600

Table 3: The thresholds t_A by treatment condition.

Elicitation of subjective profit probabilities: Having elicited the subjective quantiles we ask for the participants' beliefs of Security A making a profit. This profit probability is trivially equal to 1 for treatments LVS and LVL as all price changes are increases. We therefore do not include the elicitation task in these treatments.

The employed mechanism was proposed by Grether (1981) and later by Holt (2007) and Karni (2009). Participants are told that they can invest in either Security A and receive the bonus if Security A yields a profit at maturity or invest in an alternative security whose probability to yield the bonus is drawn from a uniform distribution on [0,1]. They submit a threshold probability for the alternative security which states the highest probability value at which they would prefer to invest in Security A instead of an alternative security with that stated probability to yield the bonus. Regardless of participants' risk preferences it is payoff-maximizing to provide a threshold that equals the subjectively believed probability of Security A making a profit.

Feedback and repetitions: The computer terminals report feedback to the participants in the form of a sample selling price of Security A. For treatments HVS, HVL, LVS and LVL, the computer carries out the relevant random draws. For the ETF conditions the computer randomly samples a sequence of 2000 trading days from the set of all such available sequences (i.e. all 2000 day histories of the DAX30) and uses its respective sequence of daily returns to simulate the asset price at maturity.

Participants in treatments LVS and LVL receive their individual feedback after they complete the quantile elicitation tasks. Participants in the other four treatments receive the feedback after completion of the quantile elicitation tasks and the profit probability elicitation tasks. In each treatment, this concludes the first round of the experiment. The experiment is then repeated for four additional rounds. Hence, the experiment comprises five identical rounds for each participant.

Procedures and payments: All 175 participants are undergraduate students at University College London. 48% of them are currently taking or have taken at least one university mathematics course in their curriculum. Nine sessions are conducted in a computer-based format programmed using the software z-Tree (Fischbacher, 2007). Each treatment condition is faced by a random subset of participants in each session. Participants are supplied with calculators that they can use throughout the experiment.

The protocol is fixed across all sessions. First, printed instructions are distributed and participants have to pass an understanding test.[14] Then the computer-based experiment commences and guides the participants through the five rounds in immediate succession. Participants receive a participation fee of $\pounds 5$ and a possible bonus of $\pounds 5$ per round. That is, participants can earn up to five bonuses of $\pounds 5$ each, one per round of the experiment. In each round, a participant can earn the bonus either through the quantile elicitation task or through the profit probability elicitation task. For each round and each participant, the relevant task type (quantile or profit probability) is determined by a simulated coin flip at the end of the experiment. This gives an ex-ante incentive to act optimally in each task. In treatments LVS and LVL, the quantile elicitation is always payoff relevant as these treatments skip the profit probability elicitation. Within each task type of each round the computer makes the relevant random draws and adds the round's bonus to the participant's account if appropriate, i.e. iff the security chosen by the participant yields the bonus. If the coin flip determines that the quantile elicitation is payoff relevant in a given round, the computer randomly and equiprobably selects a single choice task between Security A and one of Securities B1, B2, or B3.

5.2 Hypotheses

This subsection collects predictions about the behavior of an LB decision maker in our experiment. For the treatments that involve artificially constructed assets (HVS, HVL, LVS and LVL), the hypotheses directly take the propositions to the experimental setting. They are adjusted to account for the complication that we only observe bounds on the quantiles and not the precise quantiles themselves. For the treatments involving ETFs we base the hypotheses on computer simulations where we generate behaviors for the LB decision maker under the assumption that she correctly appreciates the distribution of daily returns but connects them in an additive way instead of doing the correct multiplicative compounding. The results of the simulations are analogous to those of the propositions. We present the hypotheses for different treatments jointly, where appropriate.

All hypotheses are stated using the variables defined at the beginning of Section 5.1. In what follows, we denote the rational response in the experiment as $\underline{q}_p^{A,k}$ (without participant index i), which is the best possible lower bound of the true p-percentile of Security A's price in treatment k. More precisely, $\underline{q}_p^{A,k}$ is the threshold $t^{A,k}$ such that rational behavior would prescribe switching to Security B for all strictly higher thresholds in the respective treatment. Additional optimal benchmarks $spread_p^{A,k}$ and $skew_p^{A,k}$ are constructed analogously to Subsection 5.1, using $\underline{q}_p^{A,k}$ for different values of p. The hypotheses are phrased in terms of the deviations of a participant i from the optimal benchmarks.

The first pair of hypotheses regards the median and follows Proposition 1. LB decision makers overestimate the median in all treatment with non-negligible volatility, and the result reverses in low-

[14] All participants passed the understanding test, in a few cases after asking for some additional explanations.

volatility settings.[15] Moreover, the bias becomes stronger if the length of time until maturity increases.

Hypothesis 1 (Directed Bias in Median Perception).

(a) $\underline{q}_{0.5,i}^{A,k} > \underline{q}_{0.5}^{A,k}$*, for* $k \in \{HVS, HVL, ETF\_3, ETF\_1\}$

(b) $\underline{q}_{0.5,i}^{A,k} < \underline{q}_{0.5}^{A,k}$*, for* $k = LVL$*.*

Hypothesis 2 (Bias in Median Perception increases in Time Horizon and Volatility).

$\underline{q}_{0.5,i}^{A,HVL} - \underline{q}_{0.5}^{A,HVL} > \underline{q}_{0.5,i}^{A,HVS} - \underline{q}_{0.5}^{A,HVS}$,

$\underline{q}_{0.5,i}^{A,HVL} - \underline{q}_{0.5}^{A,HVL} > \underline{q}_{0.5,i}^{A,LVL} - \underline{q}_{0.5}^{A,LVL}$,

$\underline{q}_{0.5,i}^{A,HVS} - \underline{q}_{0.5}^{A,HVS} > \underline{q}_{0.5,i}^{A,LVS} - \underline{q}_{0.5}^{A,LVS}$,

$\underline{q}_{0.5,i}^{A,ETF\_3} - \underline{q}_{0.5}^{A,ETF\_3} > \underline{q}_{0.5,i}^{A,ETF\_1} - \underline{q}_{0.5}^{A,ETF\_1}$ *and*

$\underline{q}_{0.5,i}^{A,LVL} - \underline{q}_{0.5}^{A,LVL} < \underline{q}_{0.5,i}^{A,LVS} - \underline{q}_{0.5}^{A,LVS}$.

The next pair of hypotheses follows Proposition 2 and considers the perceived spread, which is too small if volatility is non-neglible. This bias, too, increases in time horizon and volatility:[16]

Hypothesis 3 (Directed Bias in Spread Perception).

$spread_i^{A,k} < spread^{A,k}$ *for* $k \in \{HVS, HVL, ETF\_3, ETF\_1\}$.

Hypothesis 4 (Bias in Spread Perception increases in Time Horizon and Volatility).

[15]For treatment LVS, we do not state a hypothesis regarding the median, as the LB decision maker has the same prediction as a rational decision maker.

[16]Participants cannot underestimate the spread in LVS and LVL. Since our statistical tests will use proportions of participants underestimating the rational benchmarks, we do not state hypotheses regarding spread in these two treatments. The same observation applies to skewness perceptions.

$$spread_i^{A,HVL} - spread^{A,HVL} < spread_i^{A,HVS} - spread^{A,HVS} \textit{ and}$$

$$spread_i^{A,ETF\_3} - spread^{A,ETF\_3} < spread_i^{A,ETF\_1} - spread^{A,ETF\_1}.$$

Skewness perceptions are discussed analogously in Hypotheses 5 and 6, which follow Proposition 3.[17]

Hypothesis 5 (Directed Bias in Skewness Perception).

$$skew_i^{A,k} < skew^{A,k} \textit{ for } k \in \{HVS, HVL, ETF\_3, ETF\_1\}.$$

Hypothesis 6 (Bias in Skewness Perception increases in Time Horizon and Volatility).

$$skew_i^{A,HVL} - skew^{A,HVL} < skew_i^{A,HVS} - skew^{A,HVS} \textit{ and}$$

$$skew_i^{A,ETF\_3} - skew^{A,ETF\_3} < skew_i^{A,ETF\_1} - skew^{A,ETF\_1}.$$

5.3 Descriptive Statistics of the Experimental Data

We start the data analysis with a descriptive overview of the findings. The next subsection has a series of detailed results including statistical tests of all behavioral hypotheses.

For a simpler data analysis, the participants' computer interfaces restrict responses to satisfy two constraints. First, responses must exhibit at most one switching point on a choice list between Security A and a single B-type security. A participant must not switch back and forth between Security A and the respective B-type security. Second, participants must not switch from Security A to B1 (revealing bounds on her subjective 10th percentile) at a threshold exceeding the one at which she switches from Security A to B2 (revealing bounds on her subjective median), which in turn must not exceed the threshold at which she switches from Security A to B3 (revealing bounds on her subjective 90th percentile).[18]

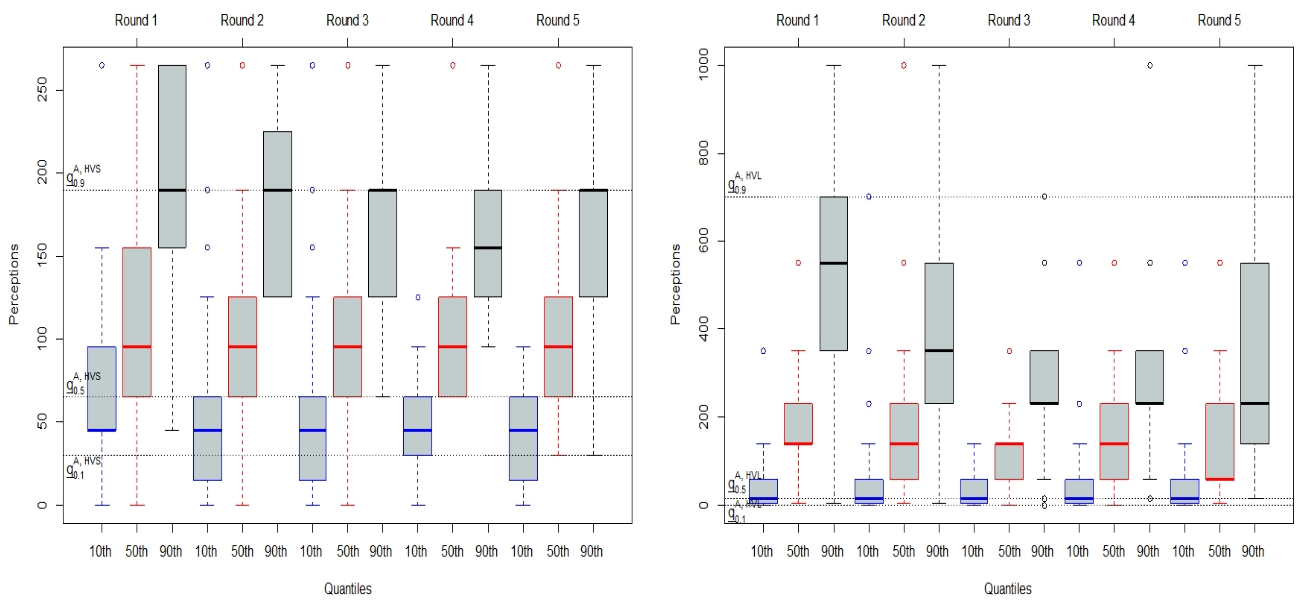
The boxplots in Figures 1, 2 and 3 summarize the distributions of participants' switching points for

[17] The two hypotheses make use of the skewness properties of the true distributions. This is not spelled out here but can be checked straightforwardly.

[18] The instructions explain that violations of these constraints are subjectively suboptimal. Additionally, the experimental software shows an error message if a participant violates either of the two constraints. Only 2% of the participants' inputs receive one or more error messages.

the three quantiles of interest, separately for each of the five rounds. The horizontal dashed lines depict the benchmark rational predictions for the respective treatment-specific quantiles, $\underline{q}_p^{A,k}$. As detailed above, we use the lower bounds of the rational and the participants' switching intervals for our statistical analysis. That is, the depicted distributions are distributions of $\underline{q}_{p,i}^{A,k}$ across different i and for the three relevant values of p.[19]

Figure 1(a) describes the switching points in treatment HVS (high volatility and short time horizon). It shows that in all five rounds the median observation of $\underline{q}_{0.5,i}^{A,k}$ (the solid median line in the second boxplot, within each triplet of boxplots) lies strictly above the rational benchmark $\underline{q}_{0.5}^{A,HVS}$. Thus, within each round, at least half of the participants strictly overestimate the median of the stochastic process in treatment HVS. The perceived spread can be gauged from the location of the two boxplots depicting the subjective 10th (first boxplot within each triplet) and 90th percentiles (third boxplot within each triplet). Comparing them to the respective rational levels, it shows that these perceived quantiles exhibit a too small spread. For example, measuring the distance between the median responses of the two quantiles shows that the typical spread is too narrow in each round. With regard to the skewness, the arrangement of boxplots within one round tends to be more symmetric than the rational benchmarks. This pattern, too, is fairly robust over the rounds.



(a) Perceived quantiles in *Treatment HVS*.

(b) Perceived quantiles in *Treatment HVL*.

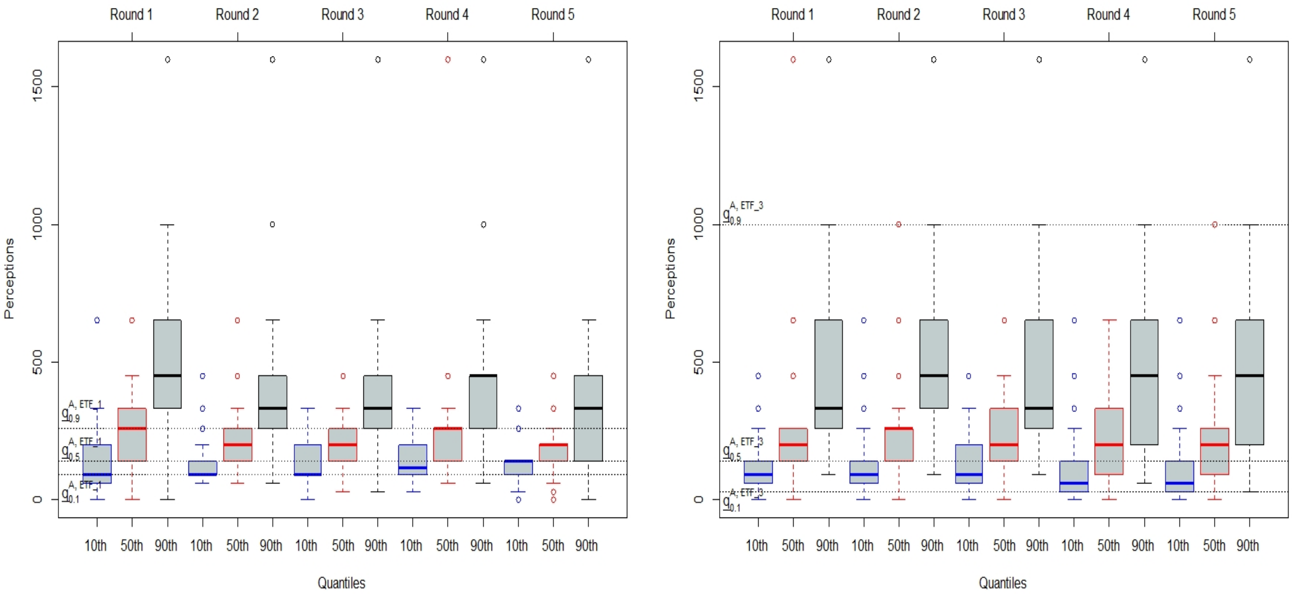
Figure 1: Elicited distributions of the participants' subjective quantiles over the five rounds, with subjective 10th percentiles (blue bordered boxplots), subjective medians (red bordered boxplots), and subjective 90th percentiles (black bordered boxplots).

Essentially the same results, but to a greater extent, appear in Figure 1(b) for HVL. Here, the longer

[19] Choosing the intervals' upper bound does not alter the main conclusions but introduces some censoring pitfalls.

time horizon exacerbates each of the described biases. The interquartile ranges of all rounds median-perception boxplots are located strictly above the rational level. Hence, at least 75% of the participants overestimated the median in each round of this treatment. The perceived price spread in this condition is underestimated by almost all participants over the rounds. The arrangement of boxplots within each round is much more symmetric than the rational levels. Despite some changes over time, all of these findings are robust within the five rounds.

Figure 2(a) illustrates the distributions of participants' switching points for condition ETF_1. Recall that in the ETF treatments there is non-negligible volatility (and much more volatility in ETF_3 than in ETF_1). The figure shows that all medians of the median-perception distributions are above the rational level indicating a substantial degree of overestimation in terms of participants's assessment of the typical (median) price growth: participants in our experiment are overly optimistic about the simple index ETF. The arrangement of boxplots also shows that the perceived distributions are quite symmetric. But with a simple ETF, the true distribution is also relatively symmetric and thus, at first glance, no inferences about relative skewness misperception can be drawn. We will return to the issue in the next subsection.

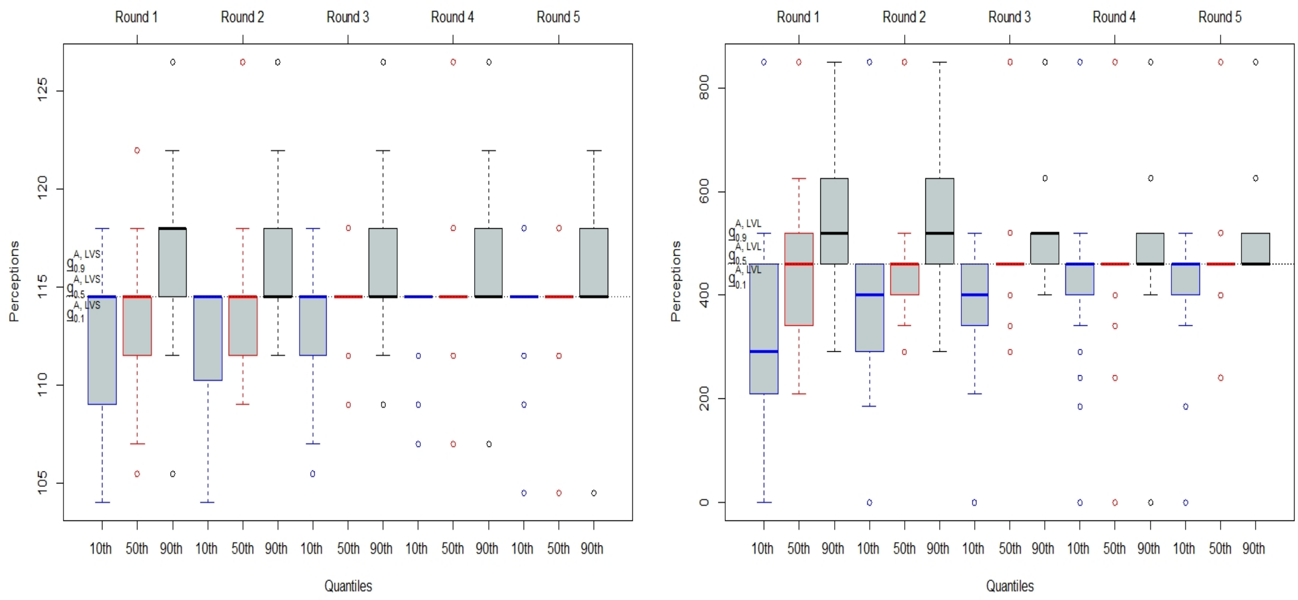


(a) Perceived quantiles in *Treatment ETF_1*.

(b) Perceived quantiles in *Treatment ETF_3*.

Figure 2: Elicited distributions of the participants' subjective quantiles over the five rounds, with subjective 10th percentiles (blue bordered boxplots), subjective medians (red bordered boxplots), and subjective 90th percentiles (black bordered boxplots).

Figure 2(b) captures the participants' switching behavior in treatment ETF_3. Again, perceived medians show a notable level of overestimation: the median of the perceived medians lies strictly above the optimal level. The arrangement of boxplots within a round also shows that the perceived spread is too small and the perceived skewness is also too small (in absolute terms). The participants do show a tendency to report skewed distributions but they far underappreciate the actual level of skewness.



(a) Perceived quantiles in *Treatment LVS*.

(b) Perceived quantiles in *Treatment LVL*.

Figure 3: Elicited distributions of the participants' subjective quantiles over the five rounds, with subjective 10th percentiles (blue bordered boxplots), subjective medians (red bordered boxplots), and subjective 90th percentiles (black bordered boxplots).

Figure 3(a) depicts perceptions in condition LVS, which differs from conditions in HVS only in that LVS considers a growth process that it essentially deterministic. (The differences in the three quantiles should therefore vanish.) The results are very different under these conditions. We observe perceptions of the median that are exactly at, or very close to, the optimal level already in round 1 of the experiment. Over the rounds, the distributions of price median perceptions quickly collapse towards a single point, which is located at the optimal level. Inspecting the arrangement of boxplots within one round, one can also see that the participants perceive skewness and price spread close to optimally: there is little dispersion in the subjective quantiles, and most of it vanishes in the course of five rounds.

Figure 3(b) shows a very similar picture of condition LVL. Again, elicited medians show low levels of misperception and become quite accurate from the third round on. In the first two rounds, spread and skewness perceptions are slightly too large but they converge to zero fast.

The result that deterministic processes are almost perfectly understood by participants is surprising. It implies that there is no exponential growth bias in our experiments as long as the process is deterministic. But especially in treatment LVL the effect of compounding is substantial. Nevetheless, the participants are able to predict it near perfectly. In part this may have to do with the feedback which is highly informative in this treatment if one understands the growth process well enough. However, there is no feedback in round 1 and yet we do not observe any underestimation of the cumulated growth.[20]

In sum, we find that the results of treatments HVS and HVL suggest a strong underestimation of skewness, which qualitatively carries over to the more realistic leveraged ETFs. This finding is consistent with the warnings that are issued against leveraged ETFs in the popular investor press and by regulators. Overall, the above observations support the hypotheses of Subsection 5.2 well for treatments HVS, HVL and ETF_3. That is, the medians are overestimated while skewness and price spread are underestimated whenever the true growth exhibits substantial volatility. For the treatments with negligible volatility, LVL and LVS, a surprising accuracy in perceptions appears in the data. The following subsection reports on appropriate tests for these and related findings.[21]

5.4 Hypothesis Testing

We employ Bayesian inference to test our hypotheses.[22] For each of the hypotheses of Section 5.2 we test whether a majority of responses concurs with the respective hypothesis. That is, the null hypothesis of each test is that at most 50% of participants show the respective pattern in their decisions.

5.4.1 The Bayesian Inferential Approach

Parametric tests are questionable in our context since we deal with interval data where the interval bounds are irregular. We thus focus on the frequencies with which the hypotheses are supported, ignoring the extent to which they are supported or not. In order to restrict our analyses to our sample of experimental participants, **y**, and not draw inferences for a hypothetical population, we use Bayesian inference (Rossi, Allenby, and McCulloch, 2005). Furthermore, the Bayesian features of the calculated p-values allow for a more straightforward interpretation of our results (Jeffreys, 1961/1939, Section 7.2). Our Bayesian probabilistic model yields a posterior distribution for the parameter of interest, θ, which is the proportion of participants who obey the respective hypothesis. The relevant random variable is whether a data observation supports the hypothesis or not, i.e., whether the observed switching point lies on the respective set of switching points that support the hypothesis. This variable is assumed to be binomially distributed with parameter θ. For θ's prior distribution we take an uninformative Jeffreys prior (Jeffreys, 1946),

$$p(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}}, \text{ for all } \theta \in (0,1).$$ [23]

[20]Two differences to the set-ups in Stango and Zinman (2009) and related studies are evident, but we would nevertheless find it surprising if they should explain all the difference. First, we use a different participant pool and different instructions. Second, we provide the participants with a calculator.

[21]We do not report the analysis of our participants' individual feedback reaction. These results are available upon request.

[22]We also employ frequentist testing as robustness checks (see Appendix A.2).

[23]A prior sensitivity analysis does not show any changes in our results. It is available upon request.

Our statistical tests assess whether the posterior distribution of θ, given the data, has significant mass lying strictly above 0.5.[24]

5.4.2 Perceptions of the Median

To test **Hypothesis 1 (Directed Bias in Median Perception)** and **Hypothesis 2 (Bias in Median Perception increases in Time Horizon and Volatility)**, we initially focus on the proportions of participants who show a bias in their median perception, consistent with the treatment-specific hypotheses. That is, we first examine round by round and treatment by treatment whether the respective proportions are significantly larger than 50%. Subsequently, we test for treatment differences in these proportions.

Hypothesis 1(a): In treatments HVS and HVL, the true medians of the two stochastic processes are as low as $\pounds 89$ and $\pounds 30$, respectively. A rational decision maker switches in Task 5 in treatment HVS and in Task 4 in treatment HVL. All participants switching at higher tasks are classified as strictly overestimating the median. Their proportions are reported in Table 4. Differences between entries and 1 are the remaining proportions, of participants who correctly perceive the median or underestimate it.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.551	(0.09)	0.655**	(0.09)	0.551	(0.09)	0.689**	(0.09)	0.620*	(0.09)
HVL	0.931***	(0.05)	0.793***	(0.08)	0.793***	(0.08)	0.793***	(0.08)	0.793***	(0.08)
ETF_1	0.733***	(0.08)	0.666**	(0.09)	0.566	(0.09)	0.733***	(0.08)	0.666**	(0.09)
ETF_3	0.655**	(0.09)	0.655**	(0.09)	0.655**	(0.09)	0.551	(0.09)	0.517	(0.09)

Bayesian significance levels of P($\theta \leq 0.5 \mid \mathbf{y}$): $*$: <10% $**$: < 5% $***$: < 1%

Table 4: Proportions of participants strictly overestimating the median over the rounds.

Table 4 lists the relevant relative frequencies by treatment condition and round. In each version of the high-volatility conditions, all shares are greater than 0.5 and range up to 0.9. In order to assign the relative frequencies a significance level, we use the posterior distributions obtained from our Bayesian probabilistic model to determine the probability of these shares being less or equal to 0.5 conditional on our experimental data, P($\theta \leq 0.5 \mid \mathbf{y}$). For treatment HVS, relative frequencies in three of the total five rounds are significantly greater than 0.5, while in condition HVL the same is true for all shares. In the two ETF treatments, the true medians are both located at around $\pounds 150$ such that switching in Task 5 would be rational. All relative frequencies of participants strictly overestimating the median are greater than 0.5. In ETF_1, four of the five rounds exhibit shares significantly greater than 0.5 while the same is true in ETF_3 in three rounds. Altogether, the above findings support Hypothesis 1(a).

[24]For each round within each treatment condition, we obtain a posterior distribution for θ by simulating samples of respectively 1000 independent values from a correspondingly adjusted *Beta* distribution. Appendix B.2 clarifies how Bayes' rule is used to derive a *Beta* distribution as posterior based on the above specifications.

Hypothesis 1(b): In LVL, the true median of the outcome distribution is $£496$. A rational decision maker would switch in Task 8. In order to test whether the participants' perceptions in LVL concur with Hypothesis 1(b), we test whether the proportion of participants who strictly underestimate the median is significantly greater than 50%. Table 5 lists the results for each of the five rounds.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
LVL	0.413	(0.09)	0.413	(0.09)	0.241	(0.08)	0.172	(0.07)	0.103	(0.06)

Bayesian significance levels of P($\theta \leq 0.5$ | **y**) : $*$: <10% $**$: < 5% $***$: < 1%

Table 5: Proportions of participants strictly underestimating the median over the rounds.

The table shows that not a single relevant proportion in the five rounds is greater than 0.5. Hypothesis 1 (b) therefore finds no support. The participants are well able to correctly assess deterministic growth in our experiment.

In sum, Hypothesis 1 finds support insofar as stochastic processes are concerned.

Hypothesis 2 (Bias in Median Perception increases in Time Horizon and Volatility): We carry out pairwise treatment comparisons of the relative frequencies of participants who falsely estimate the median, in accordance to the stated hypothesis. The results are in Table 6 and Table 7. Positive values for the differences in proportions are consistent with the hypothesis. In Table 6, we examine treatment comparisons where larger time horizon or larger volatility leads to a larger bias, according to the hypothesis. The Bayesian significance levels (indicated by asterisks) are the posterior probabilities for the null hypothesis that the proportions relating to the shorter or less volatile assets, θ_1, are greater or equal than the proportions relating to longer or more volatile version, θ_2, given our experimental data: P($\theta_2 \leq \theta_1$ | **y**).

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.380***	(0.10)	0.138*	(0.12)	0.242**	(0.12)	0.104	(0.11)	0.173*	(0.12)
HVL vs. *LVL*	0.620***	(0.09)	0.620***	(0.10)	0.655***	(0.09)	0.724***	(0.08)	0.689***	(0.09)
HVS vs. *LVS*	0.379***	(0.11)	0.482***	(0.11)	0.413***	(0.11)	0.586***	(0.10)	0.551***	(0.10)
ETF_3 vs. *ETF_1*	-0.078	(0.12)	-0.011	(0.08)	-0.089	(0.13)	-0.182†	(0.13)	-0.149	(0.13)

Bayesian significance levels of P($\theta_2 \leq \theta_1$ | **y**): $*$: <10% $**$: < 5% $***$: < 1%

Table 6: Differences for the proportions of participants overestimating the median over the rounds.

Table 6 shows that in the comparisons that involve the artificially created assets (HLS, HVL, LVS, LVL) all difference values are positive, indicating that the differences in proportions support Hypothesis 2. In each case, the median is overestimated more strongly in cases where time horizon and volatility are larger. Almost all cases exhibit significant differences between the two respective treatments. Concerning the ETF conditions, all differences show negative signs indicating ETF_1's proportions of overestimation to be greater than ETF_3's, which is inconsistent with Hypothesis 2. However, most of these differences are small and only one is statistically significant.[25]

In sum, the above results deliver substantial support for Hypothesis 2, with the exception of the comparisons of ETF perceptions.

Table 7 reports the differences in shares of participants underestimating the median over the rounds in the low-volatility condition. The listed differences in the first four rounds are positive, consistent with Hypothesis 2. However, none of the differences are statistically significant. A longer time horizon has no significant influence on participants' perceptions of deterministic growth.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
LVL vs. *LVS*	0.069	(0.12)	0.103	(0.12)	0.035	(0.11)	0.035	(0.09)	-0.034	(0.09)

Bayesian significance levels of P($\theta_2 \leq \theta_1$ | **y**) : $*$: <10% $**$: < 5% $***$: < 1%

Table 7: Differences for the proportions of participants underestimating the median over the rounds.

5.4.3 Perceptions of the Spread

If a participant's value of $spread_i^{a,k}$ lies below the treatment specific benchmark for a rational decision maker, $spread^{a,k}$, she is classified as underestimating the price spread, otherwise not. We test **Hypothesis 3 (Directed Bias in Spread Perception)** and **Hypothesis 4 (Bias in Spread Perception increases in Time Horizon and Volatility)** using this classification.

Hypothesis 3: Table 8 shows the treatment specific proportions of participants underestimating the price spread over the rounds. For HVS and HVL, all shares are significantly greater than 0.5, supporting the hypothesis. In HVL, most of them are even above 0.9. Likewise, all of the shares in ETF_3 are significantly greater than 0.5. However, in ETF_1, all range below 0.5. In sum, we find support of Hypothesis 3, as most participants in the relevant treatment conditions significantly underestimated the price spread–with the sole exception of condition ETF_1.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.655**	(0.09)	0.655**	(0.09)	0.724***	(0.08)	0.724***	(0.08)	0.655**	(0.09)
HVL	0.896***	(0.06)	0.965***	(0.03)	1.000***	(0.00)	0.965***	(0.03)	0.931***	(0.05)
ETF_1	0.266	(0.08)	0.233	(0.08)	0.333	(0.09)	0.300	(0.08)	0.400	(0.09)
ETF_3	0.862***	(0.06)	0.931***	(0.05)	0.931***	(0.05)	0.965***	(0.04)	0.931***	(0.05)

Bayesian significance levels of P($\theta \leq 0.5$ | **y**): $*$: <10% $**$: < 5% $***$: < 1%

Table 8: Proportions of participants underestimating the price spread over the rounds.

Hypothesis 4: Table 9 lists the differences in proportions of participants who strictly underestimate spreads of the price distributions, in pairwise comparisons between the two high-volatility treatments

[25] In the 4th round, the measure P($\theta_1 \leq \theta_2$ | **y**) lies below 10%, as indicated by †.

and the ETF treatments. (In the low volatility treatments, the optimal spreads are zero and hence cannot be strictly underestimated. We therefore skip all comparisons involving these treatments.) As before, positive entries indicate support of the hypothesis. The table shows that in both comparisons, all values are positive and significant. That is, we find strong evidence for Hypothesis 4.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.241***	(0.10)	0.310***	(0.10)	0.276***	(0.08)	0.241***	(0.09)	0.276***	(0.10)
ETF_3 vs. *ETF_1*	0.596***	(0.10)	0.698***	(0.09)	0.598***	(0.10)	0.665***	(0.09)	0.531***	(0.10)

Bayesian significance levels of P($\theta_2 \leq \theta_1$ | **y**): $*$: <10% $**$: < 5% $***$: < 1%

Table 9: Differences for the proportions of participants underestimating the price spread over the rounds.

5.4.4 Perceptions of the Skewness

To assess skewness perceptions, we compare $skew_i^{A,k}$ with its rational prediction $skew^{A,k}$ and classify the participants according to whether or not they underestimate skew. We then use the respective proportions in the different treatments to test **Hypothesis 5 (Directed Bias in Skewness Perception)** and **Hypothesis 6 (Bias in Skewness Perception increases in Time Horizon and Volatility)**.

Hypothesis 5: The relative frequencies of participants underestimating skewness during the experiment appear in Table 10. For HVS and HVL, all proportions are greater than 0.5 and most of them are above 0.8. Furthermore, for condition HVS, the relevant proportions in four of the total five rounds are significantly larger than 0.5 while in HVL this is true for all five rounds. For ETF_1 and ETF_3, too, all relevant proportions of participants underestimating skewness are greater than 0.5. In ETF_1, three of the total five rounds exhibit entries that are significantly larger than 0.5, whereas ETF_3 shows statistical significance without exception. Altogether, the above results strongly support Hypothesis 5: participants in our experiment underestimated skewness in HVS, HVL, ETF_1 and ETF_3. The bias is very robust to feedback information.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.551	(0.09)	0.758***	(0.07)	0.827***	(0.07)	0.896***	(0.06)	0.827***	(0.07)
HVL	0.931***	(0.05)	0.827***	(0.07)	0.827***	(0.07)	0.896***	(0.06)	0.931***	(0.05)
ETF_1	0.566	(0.10)	0.633*	(0.09)	0.566	(0.09)	0.700**	(0.08)	0.633*	(0.09)
ETF_3	0.896***	(0.06)	0.793***	(0.08)	0.896***	(0.06)	0.862***	(0.07)	0.758***	(0.08)

Bayesian significance levels of P($\theta \leq 0.5$ | **y**): $*$: <10% $**$: < 5% $***$: < 1%

Table 10: Proportions of participants underestimating skewness over the rounds.

Hypothesis 6: Pairwise comparisons of the relative proportions of skew underestimation between HVS and HVL as well as between ETF_1 and ETF_3 are reported in Table 11. The table lists the relevant treatment differences in proportions of skew-underestimating participants. In the high-volatility condition all differences are positive, indicating that the proportions underestimating skewness are greater in HVL compared to HVS. The differences are statistically significant in two out of five rounds. Concerning the ETF condition, all differences show positive signs and all are statistically significant. Taken together, we find support for Hypothesis 6.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.380***	(0.10)	0.069	(0.10)	0.000	(0.10)	0.000	(0.08)	0.104*	(0.09)
ETF_3 vs. *ETF_1*	0.330***	(0.10)	0.160*	(0.11)	0.330***	(0.10)	0.162*	(0.10)	0.125*	(0.11)

Bayesian significance levels of P($\theta_2 \leq \theta_1$ | $\mathbf{y}$): $*$: <10% $**$: < 5% $***$: < 1%

Table 11: Differences for the proportions of participants underestimating skewness over the rounds.

5.4.5 Profit Probabilities

While the elicited profit probabilities are only mildly responsive to the treatment variation (with means 0.47, 0.41, 0.55 and 0.51 in treatments HVS, HVL, ETF_1 and ETF_3, respectively) there is considerable variation across participants within treatments.

We conjecture that participants prone to the linearity bias also misperceive the profit probabilities of volatile stochastic processes. That is, they mistakenly see the growth processes in a more positive light than participants with a more accurate understanding of multiplicative compounding. In order to test this, we run treatment specific random effects regressions.[26] We use the elicited profit probability in a given round as the dependent variable and $Round$, $Feedback$, and $Biased$ as regressors. By including the number of centered rounds ($Round$) and the individually received selling price of Security A ($Feedback$) in our model, we control for round fixed effects and feedback influences, respectively. This allows us to isolate the partial difference between those participants who are linearity biased ($Biased$) versus the others. $Biased$ is a dummy variable with value 1 if the participant overestimated the median in the respective round (see entries of Table 4) and zero otherwise. Model (1) includes the coefficient of being linearity biased only. Model (2) contains additional feedback and round fixed effects and model (3) also accounts for regressor interactions.

Table 12 lists the results. In three out of four treatment conditions, participants who are linearity biased state on average significantly higher profit probabilities than those in the same treatment who are not biased, irrespective of the econometric model specification. While no evidence for a feedback effect can be inferred from the estimates, significant round fixed effects are present.[27]

[26] A Hausman test supports the appropriateness of random effect regressions.

[27] Round fixed effects show in 3 out of 4 conditions a negative sign. This points to the fact that, except for *ETF_1*, participants exhibit a downward adjustment behavior net of feedback information. The result that this is not the case for *ETF_1* is also reasonable as this condition's stochastic process is more likely to end up with a profit than the other processes.

	Dep. Var.: elicited profit probability											
Regressors	*HVS*			*HVL*			*ETF_1*			*ETF_3*		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$Biased$	0.84***	0.063*	0.083**	0.059*	0.064**	0.061**	0.039	0.031	0.035	0.091***	0.073***	0.072***
	(0.032)	(0.037)	(0.035)	(0.034)	(0.029)	(0.030)	(0.040)	(0.039)	(0.038)	(0.027)	(0.028)	(0.027)
$Feedback$		3.15E-4	-2.10E-4		5.37E-6	-1.37E-6		8.66E-5	-2.89E-5		2.29E-5	2.54E-5**
		(1.96E-4)	(2.48E-4)		(5.28E-6)	(1.03E-6)		(1.45E-4)	(3.32E-4)		(1.61E-5)	(1.20E-5)
$Round$		-0.015	0.008		-0.027***	-0.046**		0.023**	0.018		-0.022**	-0.024**
		(0.009)	(0.011)		(0.009)	(0.018)		(0.010)	(0.018)		(0.008)	(0.010)
$Feedback \times Biased$			6.77E-4**			2.64E-5**			1.51E-4			-4.53E-6
			(2.97E-4)			(1.34E-5)			(3.57E-4)			(2.63E-5)
$Round \times Biased$			-0.035**			0.023			0.007			0.004
			(0.015)			(0.018)			(0.022)			(0.017)
Intercept	0.418***	0.394***	0.436***	0.365***	0.358***	0.368***	0.524***	0.513***	0.530***	0.462***	0.463***	0.463***
	(0.029)	(0.033)	(0.030)	(0.032)	(0.027)	(0.026)	(0.040)	(0.043)	(0.065)	(0.025)	(0.024)	(0.026)
N	116	116	116	116	116	116	120	120	120	116	116	116

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 12: Results from random effects regressions with cluster-robust standard errors.

6 Conclusion

This paper investigates, and finds support for, the hypothesis that people underestimate the level of skewness in growth processes where nothing in the description of the environment explicitly raises the prevalence of skewness. A wide set of stochastic multiplicative growth processes have this feature–skewness is "hidden". The paper adds to the list of facts on cognitive biases in financial decision problems. In realistic (stochastic) settings, a failure to correctly compound can have more consequences than just an underestimation of growth.

Questions about compound interest are, by now, standard procedure in surveys about financial literacy–see e.g. the relevant module in the Health and Retirement Survey documented in Lusardi and Mitchell (2011). The typical evidence is that calculations of multiplicative growth effects show a strong downward bias, often to the extent that all compounding is ignored. The bias seems robust and economically important. However, very few studies include an incentivised experiment to corroborate the evidence (a notable exception from psychology is Christandl and Fetchenhauer, 2009). Our experiments arguably give the respondents a very good shot at correctly detecting the speed of exponential growth, especially since we use highly trained students and provide them with calculators. Indeed we do not find evidence of the exponential growth bias in settings where growth is deterministic. It is perhaps all the more notable that we find a strong bias in stochastic settings. There, skewness is indeed strongly under-appreciated.

References

[1] Abbink, K., and Rockenbach, B. (2006). *Option pricing by students and professional traders: a behavioral investigation.* Managerial and Decision Economics 27 (6): 497-510.

[2] Benartzi, S., and Thaler, R.H. (1999). *Risk aversion or myopia? Choices in repeated gambles and retirement investments.* Management Science 45: 364-381.

[3] Berger, J. O., Bernardo, J. M., and Sun, D. (2009). *The formal definition of reference priors.* Annals of Statistics, 37: 905-938.

[4] Chen, H., and Rao, A.R. (2007). *When two plus two is not equal to four: Errors in processing multiple percentage changes.* Journal of Consumer Research 34: 327-340.

[5] Christandl, F., and Fetchenhauer, D. (2009). *How laypeople and experts misperceive the effect of economic growth.* Journal of Economic Psychology 30: 381-92.

[6] Cox, J.C., Ross, S.A., and Rubinstein, M. (1979). *Option pricing: a simplified approach.* Journal of Financial Economics 7: 229-263.

[7] Eisenstein, E.M., and Hoch, S.J. (2005). *Intuitive compounding: Framing, temporal perspective, and expertise.* Working paper, Johnson Graduate School of Management, Cornell University.

[8] Ensthaler, L., Nottmeyer, O., and Weizsäcker, G. (2010) *Hidden skewness*, Working paper, Humboldt University Berlin.

[9] Fischbacher, U. (2007). *z-Tree: Zurich Toolbox for Ready-made Economic Experiments*, Experimental Economics 10(2): 171-178.

[10] Gneezy, U. (1996). *Probability judgments in multi-stage problems: experimental evidence of systematic biases.* Acta Psychologica 93: 59-68.

[11] Grether, D. (1981). *Financial Incentive Effects and Individual Decision Making* Working Paper 401, California Institute of Technology.

[12] Holt, C.A. (2007). *Markets, Games & Strategic Behavior.*

[13] Jeffreys, H. (1961/1939). *Theory of Probability* (3rd edition; 1st edition: 1939). Oxford: Clarendon.

[14] Jeffreys, H. (1946). *An invariant form for the prior probability in estimation problems.* Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 186(1007): 453-461

[15] Karni, E. (2009). *A mechanism for eliciting probabilities.* Econometrica, 77(2): 603-606.

[16] Kass, R. E., and Wasserman, L. (1996). *The selection of prior distributions by formal rules.* Journal of the American Statistical Association 91: 1343-1370.

[17] Kemp, S. (1984). *Perception of changes in the cost of living.* Journal of Economic Psychology 5(4): 313-323.

[18] Klos, A., Weber, E.U., and Weber, M. (2005). *Investment decisions and time horizon: risk perception and risk behavior in repeated gambles.* Management Science 51: 1777-1790.

[19] Lusardi, A., and Mitchell, O.S. (2011). *Financial literacy and planning: Implications for retirement wellbeing.* NBER Working Paper 17078.

[20] Mörters, P., and v. Weizsäcker, H. (2009). *Stochastische Methoden.* (3rd edition) Universität Kaiserlautern.

[21] Redelmeier, D.A., and Tversky, A. (1992). *On the framing of multiple prospects.* Psychological Science 3(3): 191-93.

[22] Rossi, P., Allenby, G., and McCulloch, R. (2005). *Bayesian Statistics and Marketing.*

[23] Samuelson, P. A. (1963). *Risk and uncertainty: a fallacy of large numbers.* Scientia 98: 108-113.

[24] Stango, V., and Zinman, J. (2009). *Exponential growth bias and household finance.* Journal of Finance 64(6): 2807-2849.

[25] Stutzer, M., and Jung Grant, S. (2010). *Expected return or growth rate? Choices in repeated gambles that model investments.* Working paper, Leeds School of Business, University of Colorado.

[26] Wagenaar, W.A., and Sagaria, S.D. (1975). *Misperception of exponential growth.*, Perception & Psychophysics 18(6): 416-422.

[27] Wagenaar, W.A., and Timmers, H. (1978). *Extrapolation of exponential time series is not enhanced by having more data points.*, Perception & Psychophysics 24(2): 182-184.

[28] Wagenaar, W.A., and Timmers, H. (1979). *The pond-and-duckweed problem: Three experiments on the misperception of exponential growth.*, Acta Psychologica 43(3): 239-251.

A Tables

A.1 Outcomes of the Stylized Experiment

Range of subjective median for Security A	Share of participants switching from A to B: Round 2 NO_HELP	HELP	Round 3 NO_HELP	HELP	Round 4 NO_HELP	HELP	Round 5 NO_HELP	HELP
$[0 - 100)$	0.000	0.032	0.000	0.047	0.018	0.046	0.000	0.092
$[100 - 500)$	0.000	0.016	0.000	0.000	0.000	0.046	0.035	0.046
$[500 - 2,000)$	0.054	0.612	0.072	0.666	0.072	0.676	0.107	0.661
$[2,000 - 6,000)$	0.145	0.145	0.127	0.095	0.200	0.138	0.303	0.046
$[6,000 - 9,000)$	0.090	0.048	0.254	0.063	0.309	0.046	0.142	0.061
$[9,000 - 12,000)$	0.381	0.064	0.309	0.063	0.236	0.000	0.196	0.030
$[12,000 - 20,000)$	0.181	0.064	0.109	0.031	0.127	0.462	0.142	0.046
$[20,000 - 35,000)$	0.090	0.000	0.109	0.015	0.036	0.000	0.053	0.015
$[35,000 - 90,000)$	0.054	0.000	0.000	0.015	0.000	0.000	0.017	0.000
$[90,000 - 250,000)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$[250,000 - \infty)$	0.000	0.016	0.018	0.000	0.000	0.000	0.000	0.000

Table 13: Subjective medians for rounds 2-5.

A.2 Binomial Testing of the Pure Experimental Data

The following tables report proportions as in Tables 4 to 11 and indicate significance levels based on binomial tests of the respective null hypotheses.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.551	(0.09)	0.655*	(0.09)	0.551	(0.09)	0.689**	(0.09)	0.620	(0.09)
HVL	0.931***	(0.05)	0.793***	(0.08)	0.793***	(0.08)	0.793***	(0.08)	0.793***	(0.08)
ETF_1	0.733***	(0.08)	0.666**	(0.09)	0.566	(0.09)	0.733***	(0.08)	0.666**	(0.09)
ETF_3	0.655*	(0.09)	0.655*	(0.09)	0.655*	(0.09)	0.551	(0.09)	0.517	(0.09)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 14: Proportions of participants overestimating the median over the rounds.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
LVL	0.413	(0.09)	0.413	(0.09)	0.241	(0.08)	0.172	(0.07)	0.103	(0.06)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 15: Proportions of participants strictly underestimating the median over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.380***	(0.10)	0.138	(0.12)	0.242**	(0.12)	0.104	(0.11)	0.173*	(0.12)
HVL vs. *LVL*	0.620***	(0.09)	0.620***	(0.10)	0.655***	(0.09)	0.724***	(0.08)	0.689***	(0.09)
HVS vs. *LVS*	0.379***	(0.11)	0.482***	(0.11)	0.413***	(0.11)	0.586***	(0.10)	0.551***	(0.10)
ETF_3 vs. *ETF_1*	-0.078	(0.12)	-0.011	(0.08)	-0.089	(0.13)	-0.182*	(0.13)	-0.149	(0.13)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 16: Differences for the proportions of participants overestimating the median over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
LVL vs. *LVS*	0.069	(0.12)	0.103	(0.12)	0.035	(0.11)	0.035	(0.09)	-0.034	(0.09)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 17: Differences for the proportions of participants underestimating the median over the rounds.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.655*	(0.09)	0.655*	(0.09)	0.724**	(0.08)	0.724**	(0.08)	0.655*	(0.09)
HVL	0.896***	(0.06)	0.965***	(0.03)	1.000***	(0.00)	0.965***	(0.03)	0.931***	(0.05)
ETF_1	0.266	(0.08)	0.233	(0.08)	0.333	(0.09)	0.300	(0.08)	0.400	(0.09)
ETF_3	0.862***	(0.06)	0.931***	(0.05)	0.931***	(0.05)	0.965***	(0.04)	0.931***	(0.05)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 18: Proportions of participants underestimating the price spread over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.241**	(0.10)	0.310***	(0.10)	0.276***	(0.08)	0.241***	(0.09)	0.276***	(0.10)
ETF_3 vs. *ETF_1*	0.596***	(0.10)	0.698***	(0.09)	0.598***	(0.10)	0.665***	(0.09)	0.531***	(0.10)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 19: Differences for the proportions of participants underestimating the price spread over the rounds.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.552	(0.09)	0.758***	(0.07)	0.827***	(0.07)	0.896***	(0.06)	0.827***	(0.07)
HVL	0.931***	(0.05)	0.827***	(0.07)	0.827***	(0.07)	0.896***	(0.06)	0.931***	(0.05)
ETF_1	0.566	(0.10)	0.633*	(0.09)	0.566	(0.09)	0.700**	(0.08)	0.633*	(0.09)
ETF_3	0.896***	(0.06)	0.793***	(0.08)	0.896***	(0.06)	0.862***	(0.07)	0.758**	(0.08)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 20: Proportions of participants underestimating skewness over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.380***	(0.10)	0.069	(0.10)	0.000	(0.10)	0.000	(0.08)	0.104	(0.09)
ETF_3 vs. *ETF_1*	0.330**	(0.10)	0.160*	(0.11)	0.330**	(0.10)	0.162*	(0.10)	0.125	(0.11)

Significance levels: $*$: <10% $**$: < 5% $***$: < 1%

Table 21: Differences for the proportions of participants underestimating skewness over the rounds.

A.3 Test Results for *Upper Bounds*

The following tables report analogous results to those in Section 5.4 using the upper bound of each subjective p-quantile instead of the the lower bound $\underline{q}_{p,i}^{a,k}$. A differentiation between upper and lower bounds does not matter for the median-perception analysis as it considers the respective switching intervals.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.655*	(0.09)	0.655*	(0.09)	0.724**	(0.08)	0.724**	(0.08)	0.655*	(0.09)
HVL	0.896***	(0.06)	0.965***	(0.03)	1.000***	(0.00)	0.965***	(0.03)	0.931***	(0.05)
ETF_1	0.233	(0.08)	0.200	(0.07)	0.233	(0.08)	0.266	(0.08)	0.366	(0.09)
ETF_3	0.862***	(0.06)	0.931***	(0.05)	0.931***	(0.05)	0.965***	(0.04)	0.931***	(0.05)

Bayesian significance levels of P($\theta \leq 0.5 \mid \mathbf{y}$): $*$: <10% $**$: < 5% $***$: < 1%

Table 22: Proportions of participants underestimating the price spread over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.241***	(0.10)	0.310***	(0.10)	0.276***	(0.08)	0.241***	(0.09)	0.276***	(0.10)
ETF_3 vs. *ETF_1*	0.629***	(0.10)	0.731***	(0.09)	0.698***	(0.10)	0.699***	(0.09)	0.565***	(0.10)

Bayesian significance levels of P($\theta_2 \leq \theta_1 \mid \mathbf{y}$): $*$: <10% $**$: < 5% $***$: < 1%

Table 23: Differences for the proportions of participants underestimating the price spread over the rounds.

Treatment	Proportion Round 1		Proportion Round 2		Proportion Round 3		Proportion Round 4		Proportion Round 5	
HVS	0.517	(0.09)	0.724***	(0.08)	0.758***	(0.08)	0.862***	(0.06)	0.793***	(0.08)
HVL	0.931***	(0.05)	0.827***	(0.07)	0.793***	(0.08)	0.896***	(0.06)	0.931***	(0.05)
ETF_1	0.400	(0.09)	0.533	(0.09)	0.466	(0.09)	0.533	(0.09)	0.566	(0.09)
ETF_3	0.896***	(0.06)	0.827***	(0.07)	0.862***	(0.06)	0.827***	(0.07)	0.758**	(0.08)

Bayesian significance levels of P($\theta \leq 0.5 \mid \mathbf{y}$): $*$: <10% $**$: < 5% $***$: < 1%

Table 24: Proportions of participants underestimating skewness over the rounds.

Treatment	Difference Round 1		Difference Round 2		Difference Round 3		Difference Round 4		Difference Round 5	
HVL vs. *HVS*	0.414***	(0.10)	0.103	(0.09)	0.035	(0.10)	0.034	(0.10)	0.138*	(0.09)
ETF_3 vs. *ETF_1*	0.496***	(0.10)	0.294**	(0.11)	0.396***	(0.10)	0.294***	(0.10)	0.192**	(0.11)

Bayesian significance levels of $P(\theta_2 \leq \theta_1 \mid \mathbf{y})$: $*$: <10% $**$: < 5% $***$: < 1%

Table 25: Differences for the proportions of participants underestimating skewness over the rounds.

B Mathematical Appendix

B.1 Proofs

Recall that $T = 2n$ and that we assume that $\mu^h\mu^l \leq 1$ and $\mu^h + \mu^l > 2$.

Proof of Proposition 1. *(i) The true median is* $q_{0.5} = Y_0(\mu^h\mu^l)^n$ *and the LB decision maker's perceived median is* $\tilde{q}_{0.5} = Y_0 + nY_0(\mu^h + \mu^l - 2)$. *It holds that* $Y_0 + nY_0(\mu^h + \mu^l - 2) > Y_0 \geq Y_0(\mu^h\mu^l)^n$*, which proves the first part of the claim. The true median* $(\mu^h\mu^l)^n$ *is decreasing in* n. *As* $1 + n(\mu^h + \mu^l - 2)$ *strictly increases in* n*, the claim that* $\frac{q_{0.5}}{\tilde{q}_{0.5}}$ *strictly decreases in* n *follows. In order to show that the fraction vanishes in the limit, observe that by l'Hôpital's rule,* $lim_{n\to\infty}\frac{q_{0.5}}{\tilde{q}_{0.5}} = \lim_{n\to\infty}\frac{(\mu^h\mu^l)^n}{1+n(\mu^h+\mu^l-2)} = lim_{n\to\infty}\frac{d\,(\mu^h\mu^l)^n/dn}{d\,(1+n(\mu^h+\mu^l-2))/dn} = lim_{n\to\infty}\frac{ln(\mu^h\mu^l)(\mu^h\mu^l)^n}{\mu^h+\mu^l-2} = 0$. □

(ii) The assumption $\mu^h + \mu^l > 2$ *implies that for* z^* *close enough to* $\frac{\mu^h-\mu^l}{2}$*, we have* $(\mu^h - z^*)(\mu^l + z^*) > 1$ *and* $(\mu^h - z^*) + (\mu^l + z^*) > 2$*, and the analogous inequalities hold if* z^* *is replaced by* $z \in (z^*, \frac{\mu^h-\mu^l}{2})$. *The true and perceived median of a binomial-tree growth process with factors* $\mu^h - z$ *and* $\mu^l + z$ *are, respectively,* $q_{0.5} = Y_0((\mu^h - z)(\mu^l + z))^n$ *and* $\tilde{q}_{0.5} = Y_0 + nY_0(\mu^h + \mu^l - 2)$. *For* z^* *arbitrarily close to* $\frac{\mu^h-\mu^l}{2}$*, the difference* $q_{0.5} - \tilde{q}_{0.5}$ *becomes arbitrarily close to* $x^{2n} - (1 + n(2x - 2))$ *for* $x = \frac{\mu^h+\mu^l}{2} > 1$. *For* $n = 1$*, this difference is strictly positive. Moreover, the term* x^{2n} *increases faster in* n *than the term* $(1 + n(2x - 2))$ *increases in* n *and the difference therefore increases in* n*, ensuring that it is positive for all* n. □

Proof of Proposition 2. *(i) For* $T = 2$*, the* $0.5 + \delta$ *quantile coincides with the max of the distribution and the* $0.5 - \delta$ *quantile coincides with the min, for the true distribution as well as for the LB decision maker's perceived distribution. The true spread is thus* $q_{0.5+\delta} - q_{0.5-\delta} = Y_0(\mu^h)^2 - Y_0(\mu^l)^2$ *and the LB decision maker's perceived spread is* $\tilde{q}_{0.5+\delta} - \tilde{q}_{0.5-\delta} = 2Y_0(\mu^h - \mu^l)$. *Due to the assumption* $\mu^h + \mu^l > 2$ *we find that* $Y_0(\mu^h)^2 - Y_0(\mu^l)^2 = Y_0(\mu^h + \mu^l)(\mu^h - \mu^l) > 2Y_0(\mu^h - \mu^l)$. □

(ii) In order to prove the claim we show that the probability that the final price of the asset is close to 0 *approaches* 1 *as* T *tends to* ∞ *(see e.g. Mörters and v. Weizsäcker, 2009, section 3.5). First, observe that* $\mu^h\mu^l < 1$ *implies that* $\mathbb{E}(log\mu_1) < 0$. *Then,* $Pr(\mu_1 \cdots \mu_T \leq e^{\frac{T}{2}\mathbb{E}(log\mu_1)}) = Pr(log(\mu_1 \cdots \mu_T) \leq T/2\mathbb{E}(log\mu_1)) = Pr(\frac{1}{T}log(\mu_1 \cdots \mu_T) \leq \frac{1}{2}\mathbb{E}(log\mu_1)) \geq Pr(|\frac{1}{T}log(\mu_1 \cdots \mu_T) - \mathbb{E}(log\mu_1)| \leq -\frac{1}{2}\mathbb{E}(log\mu_1)) = Pr(|\frac{1}{T}\sum_{t=1}^{T} log\mu_t - \mathbb{E}(log\mu_1)| \leq -\frac{1}{2}\mathbb{E}(log\ \mu_1)) \to 1$ *as* $T \to \infty$*, by the weak law of large numbers, which proves the claim as* $e^{\frac{T}{2}\mathbb{E}(log\mu_1)} \to 0$ *as* $T \to \infty$. □

Proof of Proposition 3. *Let* $f(\cdot)$ *be the respective probability density function and define* $\tilde{Y} := Y_0 + nY_0(\mu^h + \mu^l - 2)$ *as the perceived median by a LB decision maker. Suppose that* $\tau = \kappa Y_0(\mu^h - \mu^l)$ *with* $\kappa \in \{1, ..., n\}$. *Then* $f(\tilde{Y} + \tau) = f(Y_0 + Y_0(n + \kappa)(\mu^h - 1) + (n - \kappa)(\mu^l - 1) = \binom{2n}{n+\kappa}/2^{2n} = f(Y_0 + Y_0(n - \kappa)(\mu^h - 1) + (n + \kappa)(\mu^l - 1) = f(\tilde{Y} - \tau)$. *Otherwise,*

$f(\tilde{Y}+\tau) = f(\tilde{Y}-\tau) = 0$. *This proves that a LB decision maker perceives a perfectly symmetric distribution and hence a skewness equal to zero.* □

B.2 Applying Bayes' Rule

We want to employ an *uninformative* prior distribution. Using an uniform prior in that instance, however, does not ensure invariance under reparametrization.[28] A widely applied alternative to the uniform distribution is the Jefferys prior which was proposed to tackle the invariance issue and can be considered as (virtually) *uninformative* at the same time.[29] For the one-parameter case, it is defined in terms of the Fisher information as

$$p_J(\theta) \propto \sqrt{I(\theta)}.$$

Thus, for the case of a binomially distributed random variable, $p_J(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$ which is the form of a *Beta* ($\alpha_1 = 0.5$,$\alpha_2 = 0.5$):

$$p(\theta|\alpha) \sim Beta(\alpha_1, \alpha_2) = \frac{\theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1}}{\mathcal{B}(\alpha_1, \alpha_2)},$$

with $\mathcal{B}(\alpha_1, \alpha_2)$ depicting the Euler integral of the first kind. The binomial likelihood can be written as

$$p(N_1, N_2|\theta) = \binom{N_1 + N_2}{N_1} \theta^{N_1}(1-\theta)^{N_2}$$

where N_1 denotes the number of participants who obey the respective hypothesis, and N_2 labeling all other counts. Combining the likelihood and the prior according to Bayes' rule yields a posterior of the form:

$$\begin{aligned} p(\theta|N_1, N_2, \alpha_1, \alpha_2) &= \frac{\binom{N_1+N_2}{N_1}\theta^{N_1}(1-\theta)^{N_2}\frac{\theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1}}{\mathcal{B}(\alpha_1,\alpha_2)}}{\int_{x=0}^{1}\binom{N_1+N_2}{N_1}x^{N_1}(1-x)^{N_2}\frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{\mathcal{B}(\alpha_1,\alpha_2)}dx} \\ &= \frac{\binom{N_1+N_2}{N_1}\theta^{N_1+\alpha_1-1}(1-\theta)^{N_2+\alpha_2-1}/\mathcal{B}(\alpha_1,\alpha_2)}{\binom{N_1+N_2}{N_1}\frac{\mathcal{B}(N_1+\alpha_1,N_2+\alpha_2)}{\mathcal{B}(\alpha_1,\alpha_2)}\int_{x=0}^{1}\frac{x^{N_1+\alpha_1-1}(1-x)^{N_2+\alpha_2-1}}{\mathcal{B}(N_1+\alpha_1,N_2+\alpha_2)}dx} \\ &= Beta(N_1+\alpha_1, N_2+\alpha_2). \end{aligned}$$

[28] Consider for example a binomial distributed variable $X \sim \text{Bin}(n, \theta)$. Further, you want to put a prior distribution on θ. Without any information about the location of the actual value of θ in the parameter space, using an uniform prior seems reasonable. However, to take a single example, it can be shown by the density transformation rule that a reparameterization of θ as log-odds ratio ($\rho = \log\frac{\theta}{1-\theta}$) would alter the uniform prior on the new variable ρ to not be flat (and uninformative) anymore. An appropriate *uninformative* prior, however, should represent ignorance about parameter values in an absolute sense, and not relative to a specific parameterization of the probabilistic model. Thus, it must be the case that, irrespective of which parameterization is chosen, the resulting posterior distribution should match (after transformation) with all those posteriors obtained under different valid reparameterizations (for further discussions see Kass and Wasserman, 1996).

[29] The Jeffreys prior can be called (virtually) *uninformative* because it coincides with the Berger-Bernardo-Sun reference prior for a one-dimensional parameter space (as it is the case for our statistical analysis). That is, the Kullback-Leibler divergence between the prior and the posterior is maximal (see Berger et al., 2009). This measure represents the amount of information brought by the data and therefore by employing a Jeffereys prior the data brings the maximal amount of information.