Stockholm School of Economics

Thesis in Finance

Tutor: Peter Englund

Presentation: 10:15, 20th December 2011

Discussant: Jasper ter Laak

Variance and Intuitive Risk Heuristics

David Enander (18801) & Jonas Wikfeldt (21085)

Abstract

This thesis attempts to further the understanding of what risk actually constitutes in a financial setting. More specifically, we hone in on the definition of risk used in Markowitz's portfolio selection framework. The existence of the human bias of risk aversion is established, and models purporting to manage that risk abound. Even so, few of those models have taken the time to justify the use of their chosen risk measure. The absence of discussion has led to variance being the oft-chosen unexamined stand-in for how people intuitively evaluate risk, and Markowitz's portfolio selection framework is no exception. This thesis examines the extent to which Markowitz's choice of variance was sound. We also devise a new risk measure that attempts to bring some ideas brought forth by Prospect Theory into the realm of finance, and conjointly test it with variance as a potential replacement.

In order to answer the questions posed by this thesis, a survey is conducted where the respondents are asked to rank hypothetical portfolios. The data is analyzed to reveal genuine preferences at group level. The findings suggest that variance leaves much to be desired in terms of being a suitable stand-in for how people intuitively evaluate risk. The risk measure of our own making did fare better, but the tests reveal that it is far from perfect.

***Keywords**: Risk Measure, Variance, Mean-Variance, Prospect Theory, Experimental Psychology*

***Acknowledgements:** We would like to express our gratitude to our tutor Peter Englund for his valuable comments and his patience with us finishing this thesis. Furthermore, we would like to thank FrontlineSolvers for their graciousness providing us with the, for this thesis, invaluable tool "Premium Solver for Education V 7.0". We also thank Per-Olof Edlund for invaluable input in statistical matters.*

Contents

1. Introduction

Finance and risk management are today almost synonymous. Yet the attitudes and science surrounding the risk that financial managers purport to manage have, during the ages, undergone significant changes. Even the concept of risk aversion being an almost universal human trait, once had its detractors. Paul A. Samuelson (1960) recounts that, as late as 1738, the existence of risk aversion needed to be demonstrated with dramatic effect by Daniel Bernoulli, who in his treatment of the St. Petersburg paradox *"dramatized the fact that men feel they value money losses and gains at something different from their expected or arithmetic-mean values."*

Even after the existence of risk aversion was beyond reproach, it was still long considered of little consequence within the field of finance by many, since it was thought that the practice of diversification could eliminate all risk (Rubinstein, 2002). Even the founder of fundamental analysis, John Burr Williams (1938), wrote in his book *The Theory of Investment Value*:

"The customary way to find the value of a risky security has been to add a "premium for risk" to the pure rate of interest. /.../ Given adequate diversification, gains on such purchases will offset losses, and a return at the pure interest rate will be obtained. Thus the net risk turns out to be nil."

In 1952, Markowitz changed it all after publishing his seminal paper *Portfolio Selection* where he formalized the method in which riskiness was minimized given each level of expected return: The mean-variance framework (MVF). The parts constituting the MVF might have been mentioned in earlier literature, but Markowitz put it all together and presented it to the world as a coherent investment decision rule under uncertainty (Rubinstein, 2002). For his followers, in order to maximize the utility of an investment strategy, one would need to consider that both the expected return and the expected volatility of one's total holdings will affect one's final utility. The risk-reducing properties of diversification were also made more modest, with the introduction of systematic risk. The most attractive feature though, was the ease with which any investor could find his or her utility maximizing portfolio. All that was required, was to choose the most preferred portfolio out of a manageable set of "efficient" portfolios, and as long as one's true preferences corresponded with the MVF assumptions one would have found one's utility maximizing investment strategy. The set of efficient portfolios is called the *efficient frontier* and will be prominently featured in the following pages, in various forms, as we investigate claims to the effect that: In the absence of a risk free asset, one of the portfolios on the efficient frontier must be the one that maximizes one's utility, depending on the risk measure used in its construction.

The MVF provided the foundation for new theories attempting to further the understanding of how financial markets work; premier among these was Sharpe's Capital Asset Pricing Model (CAPM)

(Sharpe, 1964). With the MVF as its cornerstone, the CAPM served to price assets in a competitive market by assuming all agents to be mean-variance (MV) optimizing and rational. Notwithstanding the accolades the CAPM has received from academics and practitioners alike, empirical research still found the theory lacking. Fama and French (1992), as well as Lakonishok and Shapiro (1986), found the relationship between beta and average return left much to be desired. We will only investigate one potential explanation to why that might be. Maybe it is the CAPM's foundation that has unsettled what otherwise could be as sturdy as it is elegant – maybe there are inherent flaws in the MVF. Our inquiry hones in on one part of the MVF in particular: The choice of using variance as the mathematical measure of dispersion used for measuring the volatility of assets and portfolios.

We do not dispute that variance, or its sibling standard deviation, has many strengths. Variance is mathematically elegant, its estimator is unbiased and it holds much information about the statistical dispersion; However, if it were to be used in the construction of an efficient frontier it must fulfill yet another requirement: In the world of the MVF, it is also tasked with being a predictor of utility.

The use of variance is how the loosely defined concept of risk aversion was incorporated in the MVF, but risk aversion is not by definition equivalent to an aversion against returns with a high variance. There are many types of measures of statistical dispersion and it is possible that variance is not the one risk measure that most closely corresponds to how people intuitively evaluate what constitutes a risky gamble, i.e. there might exist another risk measure that better reflects the way people perceive risk.

The question we ask is therefore: *Is variance well suited to the task?* Equivalently: *Must one of the portfolios on an MV efficient frontier really correspond to the most utility maximizing portfolio choice for a typical agent?*

This has, to our knowledge, not been directly investigated. We aim to explore this question with the help of empirical research and careful examination of existing literature.

Although this thesis should primarily be considered an investigation into the suitability of variance, the methodology chosen also provides us with the opportunity to test an additional risk measure without much additional effort. We will therefore devise a risk measure of our own making and examine it in conjunction with variance.

2. Ambitions

For clarity, we have summarized below the questions we want to have answered in this thesis.

- *Is variance a well suited risk measure to explain the way people assess risk?* Equivalently: *Must one of the portfolios on an efficient frontier based on variance really correspond to the most utility maximizing portfolio choice for a typical agent?*
- *Is it possible to create a new risk measure with better descriptive properties than variance, in terms of a typical agent's preferences for risky gambles?*

As the methodology used to examine these questions is novel and developed by the authors, we will exercise extra care examining the robustness of the results:

- *Are the investigative methods used suitable to answer the aforementioned questions?*

3. Delimitations

Although this thesis concerns the suitability of the examined risk measures in terms of being used to construct efficient frontiers, we judge their suitability squarely on how well human preferences are reflected in the final set of efficient portfolios.

The other important feature of a risk measure, if it is to be used in the construction of an efficient frontier, lays in its mathematical properties and how enabling they are in actually constructing the final set in a reasonable time. In other words, how well suited the risk measure is in finding the combinations of assets yielding the highest returns for given risk levels, taking into account that every asset up for consideration can potentially be correlated to all other assets.

The MVF is well suited to this task, since variance can be regarded as merely a special form of covariance – covariance being the standard measure of how two stochastic variables move in relation to each other. The calculations required to derive the set of portfolios that make up the efficient frontier is therefore fairly straightforward.

If another risk measure aims to replace variance, the same problem of determining the nature of co-movements of different assets needs to be solved. It is likely that the use of covariance need not be the best way to measure the relationship between two assets in this case. This is a part of the puzzle that needs to be solved when creating an alternative to variance. We do not provide an algorithm in which the risk measure we created can be used in this manner, so it not presented in this thesis as a full-fledged real world contender to variance. It serves in this paper only as an aid for examining variance and as a proof of concept that when it comes to human preferences, variance might fall short.

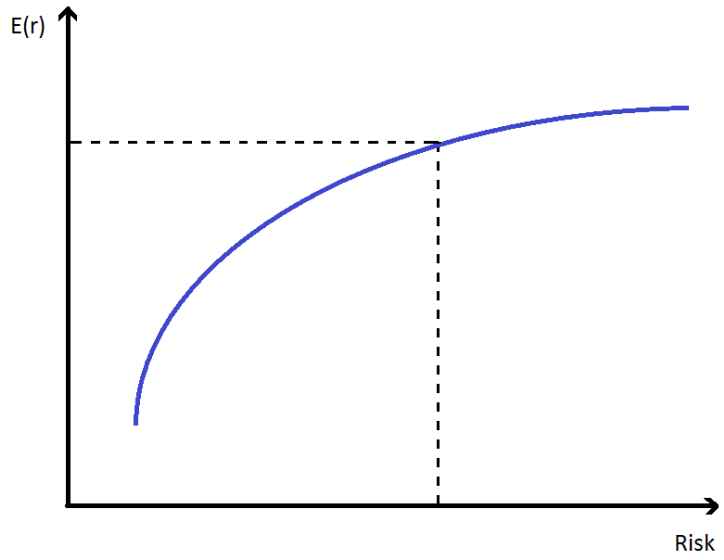
It bears mentioning though, in a world of evolutionary algorithms and ever increasing machine aided computation abilities, a straightforward way to compute co-relationships is not as important as it once was. Since no attempts were made calculating an efficient frontier using the risk measure we developed using real assets, we cannot argue with much vigor how feasible it would be in a practical setting, but we do not regard it as outside the realm of possibility that such an attempt would have been fruitful.

4. Theoretical Background

4.1. The Markowitz Legacy

The MVF, unsurprisingly, assumes that investors are risk averse, i.e. given two portfolios with equal expected returns the investor will chose the less risky one. The rational investor will then only be prepared to take on more risk if he is compensated by a higher expected return. Hence, a tradeoff between risk and expected return needs to be made. The MVF leaves that choice up to the individual investor, who in the case of no risk free asset, is left to his own devices in choosing one of many potential portfolios on the so called "efficient frontier of optimal portfolios", where all portfolios on display represent the minimal risk for a given expected return. The efficient frontier lends itself well to being graphed due to its two dimensional nature, and is shown in Figure 1.

Figure 1 - The Efficient Frontier



A generalized Efficient frontier. Two types of risk measures will be used in this paper in place of the "Risk" the x-axis denote here.

In the MVF, as the name implies, risk is evaluated in terms of variance.

When Markowitz introduced the MVF he did not argue why variance was chosen as the risk measure used in his model. However, in Markowitz's prize lecture, after being one of the three who were awarded the 1990 *Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel*, he provided some insight into his initial choice of variance as a measure of risk when he said:

"It seemed obvious that investors are concerned with risk and return, and that these should be measured for the portfolio as a whole. Variance /.../ came to mind as a measure of risk of the portfolio." (Markowitz H. M., 1990).

Markowitz gives the question of risk measures more thought in his article *Foundations of Portfolio Theory* (1991), where he argued that semi-variance makes a realistically superior choice as a measure of risk. Still, variance is the measure that is predominantly used today – perhaps for no other reason than that it was the first measure that *"came to mind"*.

4.2. Criticism of the MVF

Even though the MVF is considered the foundation of modern portfolio theory it has endured some extensive critique during the last decades. Our thesis is solely focused on one aspect of the MVF: The efficient frontier. The critique below should therefore not be seen as all-encompassing critique of the MVF.

4.2.1. Normal Distribution

The MVF models the volatility of assets and portfolios using variance. This must be the optimum in the case where returns are always normally distributed, i.e. jointly normally distributed, since one can wholly describe a normal distribution with only its mean and variance. However, it has been shown that asset returns are generally not normally distributed. Extreme swings, such as three to six standard deviations from the expected value, occur far more often than what is predicted in a normal distribution (Mandelbrot & Hudson, 2004). It is said that financial return series are permeated by skewness and/or "fat tails" (Mandelbrot & Taylor, 1967). A natural defense would be to argue that the *Central Limit Theorem* (CLT) stipulates that a stochastic variable created to be the average of a number of other non-correlated stochastic variables will be approximately normally distributed as long as the number is large enough. This would hold regardless of what probability distribution those stochastic variables could be described with (Bergh & van Rensburg, 2008). The use of this line of thinking is of limited value in practical applications, since assets are generally correlated to some degree (Statman, 1987), or as William F. Sharpe (1999) put it:

"*While the central limit theorem provides a powerful inducement to assume that investment returns and values are normally distributed, it is not sufficient in its own right. While most investment results depend on many events and most portfolios contain many securities, it is unlikely that the influences on overall results are unrelated."*

4.2.2. Total Holding and the Relationship to Utility

While the CLT may be the strongest ally of the MVF, it suffers from another problem not as obvious as the problem with systematic risk: Depending on one's frame of reference, it is not certain that the number of returns is always large enough for the CLT to be applicable.

The efficient frontier is supposed to be constructed of all potential holdings available to an investor. The MVF would have it so, that only by looking at the "big picture" an investor could maximize his or her utility. To a modern investor that "picture" is indeed big. Even in a small country such as Sweden the potential number of domestic assets satisfies the modest requirements of the CLT by a large margin. Under the assumption that all assets are jointly uncorrelated, the number of assets for the requirements of the CLT to be fulfilled is commonly thought to be only 20-50. A rational MV optimizing investor should therefore have no problem including a number of assets large enough in his portfolio, so at least that criterion of the central limit theorem is satisfied.

There is an additional consideration however: Individuals usually look at their assets much differently. *Mental Accounting* introduced by Richard Thaler (1985) is an empirically well proven theory which states that people rather look at their portfolio as fragments where different parts belong to different mental accounts, each contributing in a positive or negative manner towards final utility. If the final utility derived then in reality comes from several smaller "mental portfolios" within the real portfolio, the role of CLT is further marginalized since the set of assets for each "mental portfolio" is smaller than the total set.

4.2.3. Human Risk Assessment Heuristics

If one cannot justify the use of variance due to the CLT, it has to stand up on its own merits as a stand in for how people intuitively assess riskiness in a gamble if it is to be used as a means to maximize utility. To investigate this claim we can examine variance in terms of it fulfilling the conditions of the four von Neumann-Morgenstern axioms of expected utility theory (Von Neumann, J & Morgenstern, O, 1953). If a utility function fulfills the four relatively modest axioms of "rationality", it is deemed a valid utility function.

Our starting point is that the MVF is not *"consistent with the Von Neumann-Morgenstern axioms of expected utility theory unless either (i) asset prices follow normal probability distributions, or (ii) utility functions representing investor preferences are quadratic"* (Kerstens, Mounir, & Van de Woestyne, 2008). We have examined the first condition *(i)*, in the two previous sections and found the evidence wanting, then the burden rest solely on *(ii)*: Is it reasonable to suspect that the utility function of investors is quadratic in nature? The short answer is *"no"* – a quadratic utility function has several unintuitive properties. Prime among those is that *"this simple utility function has the unrealistic implication that the investors absolute risk aversion is increasing in wealth, so that he at some point of wealth is worse off as he gets richer"* (Kerstens, Mounir, & Van de Woestyne, 2008). This bizarre feature is easy to see by studying the quadratic utility function below, where W is wealth and β is a positive constant.

$$u(W) = W - \frac{\beta W^2}{2} \qquad \textbf{(Equation 1)}$$

An investor who follows a quadratic utility function is also an investor who is indifferent to higher moments than variance (Hagströmer, Anderson, & Binner, 2007). The former property is so unrealistic it is hard to find any research examining it. The second property, the one suggesting an investor makes no consideration to higher moments, is also unwarranted. Large scale studies have demonstrated the existence of a preference for positive skewness among investors to an ample degree (Mitton & Vorkink, 2007).

Even if we abandon the implied quadratic utility function required by the von Neumann-Morgenstern axioms, the symmetric nature of variance has also amassed critique from other directions. It has been argued, investors are far more concerned with the probability of losing rather than the possibility of winning, i.e. investors are generally loss averse. Studies have shown that the power of a loss is psychologically approximately twice as strong as the power of a gain (Kahneman & Tversky, 1979). According to this view, people's intuitive concept of risk is fundamentally asymmetric in nature and variance seems like an unlikely candidate to approximate the way in which people think.

5. Contribution

5.1. Risk Measures and Efficient Frontiers

There exists research concerning alternative methods of constructing efficient frontiers as well as empirical research examining intuitive human risk perception. What we have not found, are instances where these two fields of research are married into a coherent framework which can produce conclusions about the desirability of new risk measures in financial contexts.

Research papers (Bergh & van Rensburg, 2008), (Mhiri & Prigent, 2010), (Stacey, 2008) examining alternative ways to construct efficient frontiers have many commonalities:

- They try to augment the MVF by including higher moments (skewness and kurtosis usually).
- They are predominately focused on the mechanics of the construction of the alternative efficient frontier.
- They do not contribute with any empirical psychological research examining if the inclusion of higher moments indeed affects investors' utility.

It is not surprising that these papers have been focused on the technical difficulties stemming from the inclusion of additional moments, since leaving the two parameter approach of the MVF and constructing a three dimensional efficient frontier gives rise to many technical challenges that need to be addressed. The downside of this focus is that the desirability of also taking higher moments into consideration is often just assumed based on perceived shortcomings of variance to approximate non-normal distributions (Bergh & van Rensburg, 2008).

Our research, in comparison, does not attempt to solve the problem of the actual construction of an alternative efficient frontier but is instead focused on what that efficient frontier should attempt to minimize. The other important difference is, for reasons we will go into in the next chapter, us choosing to use a two parameter approach instead of a four parameter approach as in the case when using higher moments. Our focus on the "why" instead of the "how" as it pertains to risk aversion and efficient frontiers remains the most salient feature of this paper.

For empirical research into the underpinnings of risk aversion, one will have to look elsewhere. This also means that the most fastidious experiments carried out, exploring risk aversion, are not directly geared towards being useful in the field of finance (Acerbi, 2002), (Kahneman & Tversky, 1979). Prospect Theory, for example, put forward utility maximizing functions of the general type where riskiness and expected return are conflated into one utility estimate. In contrast, our empirical research is of a more restricted nature where the expected return is held static and the risk measure we construct

bears a closer resemblance to a general volatility measure. We will address this issue in greater detail in section 6.1.

5.2. Methodology

The manner in which we examine the data generated by the questionnaire is also, in part, of our own creation. We quickly realized that the task of extracting the maximum amount of information from a small data set consisting of ordinal rankings was not part of the standard fare of statistical testing. Our scant data set therefore compelled us to devise our own testing procedure. It might be wise to postpone the reading of the elucidation of our contribution below until after reading section *10 - Analysis* to aid understanding.

We feel it prudent to mention that our contributions in this field could not have been achieved without the aid and guidance of Per-Olov Edlund.

Several of our tests use the practice of dividing an ordinal set into all possible permutations, then reconstituting them to groups large enough so that a general chi square test can be performed. This is either novel or, perhaps more likely, rediscovered by us. The idea of reconstructing an ordinal order into a set of six questions, which we then can subject to further tests, is likewise, a statistical test procedure that is to our knowledge unique to this paper. Consequently, our concept of degrees-of-fitness, which is based on the aforementioned six questions, is also an original addition in this thesis.

6. Raison D'être for a New Risk Measure

Variance is such a commonly used measure of statistical dispersion it is often taken to be synonymous with variability in stochastic variables. To do this would be a mistake. The only limitation to the number of ways one could measure the spread of a probability distribution is one's imagination. Other examples of risk measures are: Range, interquartile range, median absolute deviation, semi variance and value at risk. Since all these risk measures use different mathematical formulas in order to calculate their respective estimates, they may or may not agree with each other regarding which of several hypothetical probability distributions have a high or a low level of risk.

Unless one also stipulates what intended use one has for a proposed risk measure, there is no way to say that one risk measure is better than any other. In our case, we are interested in a suitable risk measure that can help us construct an efficient frontier. From this point on when we refer to the suitability of a risk measure, the metrics used to measure that suitability is how well suited the risk measure is in the creation of an efficient frontier; Or, with the delimitations of this paper in mind, how well the risk measure captures how humans natively and instinctively experience riskiness, which has a close relationship to its suitability in the creation of an efficient frontier.

If we recall that an efficient frontier is based on the assumption that people have two distinct preferences for their combined financial holdings, the highest possible return and the least amount of risk, there can be little doubt that as long as we state these two conditions in such general terms the only thing an investor needs to do in order to maximize his or her utility, is to select the most preferred portfolio from the curve of efficient portfolios. It is when we are forced to define how this risk is measured we run into problems. The preferences surrounding risk are like all preferences - wholly wrapped in the mysteries of the mind. In finding that intuitive risk assessment function, it therefore exists no substitute to empirical testing. We will subject variance as well as another risk measure of our own making to such empirical tests.

The reason why we want to create a new risk measure is twofold. Firstly, the new risk measure aids us in examining whether variance is a suitable risk measure (and vice versa). This is a consequence of our testing methodology which is presented in greater detail in the relevant section, but the gist of the idea is that in testing two risk measures at once, it allows us to test each risk measures in two ways instead of just one: If a given risk measure is descriptive in how people value a series of gambles having different risk levels and if the risk measure is descriptive in how people value a series of gambles sharing the same risk level. We could achieve the same outcome by creating random data that has no relation to a new risk measure, but since the tests are symmetric in nature, testing two risk measures instead of just one requires very little additional effort.

Secondly, a lot of new research has been amassed since the creation of the first efficient frontier in 1952, so we see an opportunity to try to best variance with the help of the advances that have been made in the last 62 years. Since our methodology is so allowing, we see no reason why we should not, at least attempt to, construct the best risk measure we can muster.

The question is: How do one create a new risk measure to replace variance?

There are two natural ways in one would go about creating a new risk measure. One could try to simply add *"higher moments"* to augment variance, such as skewness and kurtosis, but as Markowitz (1991) put it:

"/.../ when four moments are required, the investor must pick carefully from a three-dimensional surface in a four-dimensional space. This raises serious operational problems in itself, even if we overcome computational problems due to the non-convexity of sets of portfolios with given third moment or better."

Markowitz then goes on to speak an alternate approach to achieve the same end:

"But perhaps there is an alternative. Perhaps some other measure of portfolio risk will serve in a two parameter analysis for some of the utility functions which are a problem to variance."

For the reasons Markowitz himself laid out, we will attempt this two parameter approach in searching for an alternative measure of portfolio risk.

6.1. The Creation of a New Risk Measure

In examining previous research, we find that the most notable attempt to address questions concerning intuitive evaluation of risk is Prospect Theory (PT). From experimental economic research, Kahneman and Tversky have proven that the manner in which people make decisions under uncertainty systematically deviates from traditional economic theory, and consequently implicitly also from what would be predicted by the MVF. For our purposes, the most important aspect of PT is the existing theoretical framework explaining the asymmetrical nature of risk aversion with the empirical research to support it.

PT seems like a promising starting point for a new risk measure, but unfortunately there are confounding factors. One of the more striking findings by Kahneman and Smith (2002) for example, is people often being much more sensitive to the way an outcome differs from some non-constant reference level (such as the status quo) than to the outcome measured in absolute terms. As they put it:

"*This focus on changes rather than levels may be related to well established psychophysical laws of cognition, whereby humans are more sensitive to changes than to levels of outside conditions, such as temperature or light.*"

Interesting as it may be, findings such as the one mentioned above are incorporated into PT and makes it difficult, if not impossible, for us to apply the PT framework directly in creating a risk measure that is as suitable for use within the field of finance, especially when it comes to the calculations of efficient frontiers as is our requirement.

The source of this incompatibility is that a risk measure used in the creation of an efficient frontier must have certain properties. Namely, it must measure only the dispersion around an arbitrarily set mean, and the mean itself must not be taken directly into account in the creation of the measure. Prospect Theory does not provide such a risk measure since it is designed to evaluate a gamble in terms of its safe equivalent. In contrast, any risk measure that fulfills the properties we require would explain a gamble in terms of two separate properties: Its expected return and also its risk level. In order for us to use Prospect Theory for our purposes, we must therefore create a risk measure that can only be described as being loosely based on the same principles as prospect theory. One would be overstating if one were to refer to this new risk measure as a Prospect Theory risk measure. It would be a more accurate statement to say that the risk measure was inspired by Prospect Theory. We will thus refer to this risk measure as a *Prospect Theory Inspired Volatility Measure* (PTIVM) throughout this paper.

6.1.1. Prospect Theory – a Review

In order to understand our risk measure, the PTIVM, it is helpful to have a fuller understanding of how Prospect Theory and the refinement of the same, Cumulative Prospect Theory (CPT), evaluate a gamble (Tversky & Kahneman, 1992). We will use the following notation in order to describe a gamble:

x_i: Outcome i, $i = 1,\ldots,n$

p_i: Probability that the outcome of the gamble will be x_i

A gamble with n outcomes can thus be denoted:

$$(x_1, p_1; \ldots; x_n, p_n)$$

For example, a random throw of a normal six sided die would be described as:

$$(1, \frac{1}{6}; 2, \frac{1}{6}; 3, \frac{1}{6}; 4, \frac{1}{6}; 5, \frac{1}{6}; 6, \frac{1}{6})$$

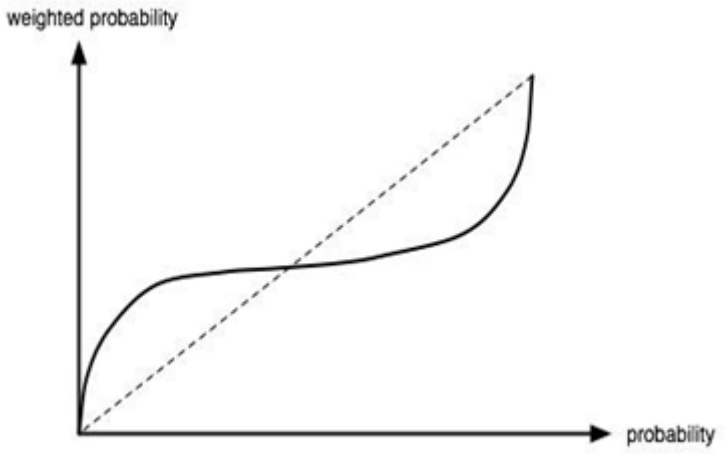
In CPT as well as in PT, the valuation of any gamble can be determined only after an editing phase where the probabilities and their respective outcomes are weighed by separate functions. Every distinctive outcome is transformed by these two functions: The probability weighing function and the value function.

The probability weighing function in PT is (Tversky & Kahneman, 1992):

$$\pi_i = \frac{p_i^y}{\left(p_i^y + \left(1 - p_i^y\right)\right)^{1/y}} \qquad \textbf{(Equation 2)}$$

where *y* is an estimate given from how people perceive different probabilities. CPT uses a different probability weighing function, which is also the only difference between CPT and CP. This difference is of little importance to us since, for reasons we will dwell into later, we will not use any probability weighing function in the calculation of the PTIVM. The probability weighing function is presented here, only so that our justifications for its exclusion will be easier to absorb in section 6.1.2. A graphical representation of the probability weighing function is shown in Figure 2.

Figure 2 - The PT Probability Weighing Function



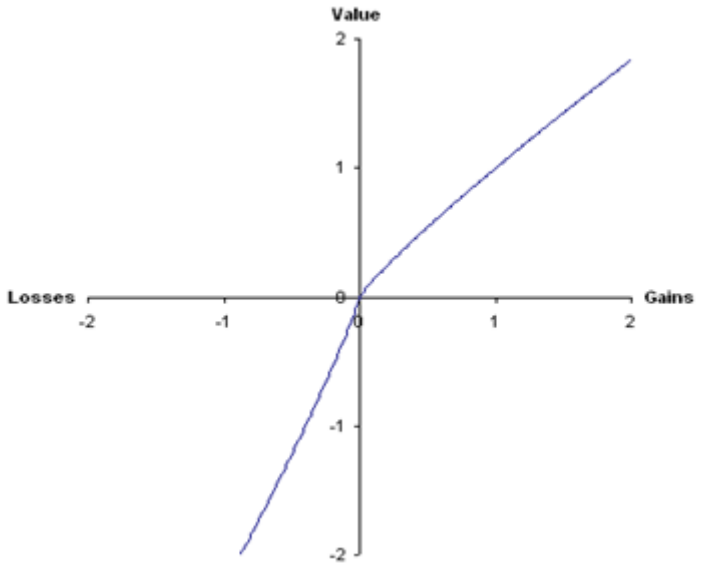
A stylized version of the probability weighing function used in Prospect Theory. Note that Cumulative Prospect Theory uses another type of probability weighing function.

The second function is the value function:

$$v(x_i) = \begin{cases} x_i^a, & x_i \geq 0 \\ -\lambda(-x_i)^a, & x_i < 0 \end{cases} \quad \textbf{(Equation 3)}$$

where Tversky and Kahneman (1992) estimated λ and a to be 2.25 and 0.88 respectively. A graphical representation of the value function can be seen in Figure 3.

Figure 3 - The PT Value Function



A stylized version of the value weighing function used in Prospect Theory and Cumulative Prospect Theory.

Used together, the PT certainty equivalent can then be calculated with:

$$\pi(p_1)v(x_1) + \ ... + \pi(p_n)v(x_n) \quad \textbf{(Equation 4)}$$

The PT certainty equivalent, like the name suggest, is the fixed amount one would regard as equally attractive to the gamble.

The different treatment of utility in PT from how it is treated in the MVF should be apparent by now: The PT certainty equivalent is a final utility estimate while an efficient frontier is a set of potentially utility maximizing choices, leaving the final choice up to the investor.

6.1.2. Exclusion of the Prospect Theory Probability Weighing Function

After an initial informal preliminary study into different ways to display a gamble, we noticed people having trouble perceiving the impact of heterogeneous probabilities across outcomes. This notion was given more credence when we found previous research having discovered much the same thing. Birnbaum (2004) have presented a very convincing case of people violating "stochastic dominance" when confronted with choices involving heterogeneous probabilities. To be more precise, he found that people were not able to "coalesce". In case these are unfamiliar terms, below is a short explanation:

- *Stochastic dominance* is a form of stochastic ordering where one lottery (i.e. a probability distribution) can be ranked as superior to another (Hadar & Russell, 1969). A lottery *A* has a stochastic dominance over lottery *B* if the probability of winning *x* or more in gamble *A* is greater than or equal to this same probability in gamble *B* for all *x*, and the probability of winning *x* or more in lottery *A* is higher in at least one of the outcomes. In simpler terms: Lottery *A* is unequivocally better than lottery *B*.

The study of systematic violations of stochastic dominance is therefore an especially interesting avenue of research since one can rule out that these violations are due to some form of genuine preference.

- *Coalescing* concerns the way in which the probability distribution of a gamble is presented and is easiest to demonstrate with an example.
  A coin toss is a two-branch split form:

  $(win, \frac{1}{2}; lose, \frac{1}{2})$

  Similarly, rolling a dice where even numbers generates a gain and odd numbers a loss, is a six-branch split form:

  $(win, \frac{1}{6}; win, \frac{1}{6}; win, \frac{1}{6}; lose, \frac{1}{6}; lose, \frac{1}{6}; lose, \frac{1}{6})$

  To regard these two gambles as equal would be an example of coalescing. It is the assumption that if a gamble has two or more outcomes that lead to the same consequence the participant will be able to sum the probabilities together before evaluating and comparing it to other lotteries (Birnbaum, 2004).

Birnbaum demonstrated that people have trouble coalescing, this is not noteworthy because PT assumes that people never violate any kind of stochastic dominance. It is important because PT

implicitly assumes that coalescing is specifically not a source of stochastic dominance (Tversky & Kahneman, 1986).

The "PT editing phase" includes no consideration for if a gamble is displayed in a coalesced form or not. Indeed, in the experiments Tversky and Kahneman carried out, all gambles were already "pre-coalesced" (Kahneman & Tversky, 1979).

Birnbaums research even points to people being tasked with coalescing being the main reason for violations against stochastic dominance. He calls this phenomenon the *splitting effect* (Birnbaum, 2004). The fear is: If PT neglects to include a very prominent reason why people violate stochastic dominance, yet attempts to provide a descriptive theory of how people value risky gambles, its findings will suffer. Birnbaum expressed it thusly:

"*The main problem for CPT is that it must satisfy the property of coalescing, so it cannot explain splitting effects /.../ nor can it explain paradoxes that can be deduced as implications of coalescing. Violations of CPT are observed whether probability is presented as text or accompanied by pie charts, whether by lists or by frequencies, whether by marbles or by tickets, and using either branch or decumulative probabilities."*

Since it is the PT probability weighing function that is responsible for capturing the "irrationality" people exhibit when faced with heterogeneous probabilities – and it has no awareness of the splitting-effect – the weighing function might suffer from serious inherent problems.

Birnbaum (2004) concludes that if the aim is to minimize violations against stochastic dominance the gamble should be presented in an appropriately split format with an equal number of branches in each gamble. Further studies found that violations of stochastic dominance can be nearly eliminated by also letting the probabilities across branches be homogeneous, which they referred to as *canonical split form* (Birnbaum, Johnson, & Longbottom, 2008). This is a central finding in our thesis and serves as our main justification why we have chosen to represent all our gambles in canonical split form when we constructed our survey. Our use of homogenous probabilities consequently also renders the need of a probability weighing function superfluous. This aspect will be explored more thoroughly in section 8.1.

The unawareness of the splitting-effect is not the only form of criticism that has been waged at the PT and CPT probability weighing functions. Indeed, the reason CPT was developed was due to the increasing criticism that was waged against the original weighing function (Tversky & Kahneman, 1992). We feel that, despite its lofty aspirations, CPT fails to correct the flaws of its predecessor. The flaws can be made more apparent by giving a couple of examples.

- In PT, due to its weighing function, a gamble of a series of outcomes all with small probabilities could result in a safe equivalent exceeding the expected return of the gamble.
- In CPT the probability weighing function does not take the magnitude of change between two outcomes into full consideration. Two gambles, where the only difference is that one outcome, (x, p), is replaced by two outcomes, $\left(x, \frac{p}{2}; x + \varepsilon, \frac{p}{2}\right)$ where ε is an arbitrarily small number will result in a significant difference in the two gambles' resulting safe equivalents.

The latter example is especially damning for us, since the gambles we have constructed can very well have separate outcomes with only miniscule differences. This is an unavoidable consequence of the method in which they are constructed.

The last question we ask is: What effect will retaining the value function have if we exclude the probability weighing function? The probability weighing function and value function are happily by design supposed to have an uncorrelated effect on people's evaluation of a gamble (Equation 1-3). I.e. the probability weighing function is agnostic to the type of outcome the probability is affixed to and, in the same fashion, the value function is unmoved by how probable an outcome is. Not using one would at most result in less accurate estimates, not estimates that would be biased in one way or the other.

6.1.3. Mathematic Attributes of the PTIVM

With the probability weighing function excluded, only the value function (Equation 2) remains, which as previously mentioned, is used in both CPT and PT. In order to reformulate the value function in such a way we can use it as a risk measure, with the properties required, we need to make further adjustments. That is, we must transform it in such a way that it enables us to measure the dispersion around the mean.

First, one needs to ensure that values that lay below the mean does not negate values that lay above the mean. When calculating variance this problem is addressed by squaring the deviation from the mean:

$$(x_i - \mu_x)^2$$

This in itself emphasizes extreme values and can be considered a primitive utility function. Since $v(x_i)$ already is a utility function, an emphasis of extreme values serves no purpose and could only add distortion to its estimate; so we have chosen to use the absolute value of the deviation from the mean instead. Our idea is thusly to use the utility estimates produced by $v(x_i)$ and then calculate the mean absolute deviation. To simplify comparisons between variance and the PTIVM, they are displayed side by side in similar mathematic forms for a gamble with *n* outcomes in Table I below.

Table I - Variance and the PTIVM, Mathematical Forms

Variance	PTIVM
$v(x_i) = \{x_i\}$	$v(x_i) = \begin{cases} {x_i}^a, & x_i \geq 0 \\ -\lambda(-x_i)^a, & x_i < 0 \end{cases}$
$R = \frac{1}{n}\sum_{i=1}^{n}\left(v(x_i) - \frac{1}{n}\sum_{i=1}^{n} v(x_i)\right)^2$	$R = \frac{1}{n}\sum_{i=1}^{n}\left\lvert v(x_i) - \frac{1}{n}\sum_{i=1}^{n} v(x_i)\right\rvert$

Mathematical forms of Variance and the PTIVM. Although variance is here formulated in an atypical way, it is the standard definition of variance that is on display. $v(x_i)$ *is the value function of outcome* x_i. R *denotes the final estimate of the risk measure.*

7. Methodology

7.1. General Outline

To see whether variance and/or the PTIVM succeed in capturing the typical investor's aversion toward risk, we have constructed two sets of four hypothetical portfolios. All eight portfolios, A-H, have the same expected return and all portfolios in each set have the same level of risk with the difference that the risk is measured with variance in one set and the PTIVM in the other. The attributes of the two sets can be seen in Table II and Table III below.

Table II – Descriptive Statistics Portfolio Set I

	PTIVM	Variance	Exp. Ret.	Skewness	Kurtosis
A	6,00	113,00	10,00	-0,53	2,17
B	6,00	144,00	10,00	1,97	5,04
C	6,00	82,00	10,00	-0,2	-0,43
D	6,00	109,00	10,00	0,85	-0,64

Statistical properties for the portfolios of the first set.

Table III – Descriptive Statistics Portfolio Set II

	PTIVM	Variance	Exp. Ret.	Skewness	Kurtosis
E	6,10	118,00	10,00	0,99	-0,17
F	4,80	118,00	10,00	2,55	8,72
G	6,70	118,00	10,00	-0,53	1,85
H	7,40	118,00	10,00	-0,22	-0,21

Statistical properties for the portfolios of the second set.

The salient features being that portfolios in set I got a homogenous PTIVM level, and a heterogeneous variance risk level, while the portfolios in set II got a homogenous variance level, and a heterogeneous PTIVM level. These attributes are crucial to our testing procedure, so well worth noting.

For readability purposes, we will henceforth refer to the set consisting of four hypothetical portfolios that have a homogenous level of risk according to the PTIVM, *the PTIVM static set*. In the same fashion, the set consisting of four hypothetical portfolios that have a homogenous level of risk according to variance, will be denoted *the variance static set*.

The two sets of portfolios were then distributed in the form of a questionnaire to our test subjects, who were instructed to order the hypothetical portfolios in each set according to their preferences, under the assumption that they were investing their entire net worth. They were also asked to state their self-experienced strength of preference for their selected orderings. The questionnaire can be seen in its entirety in Appendix 2 - *The Questionnaire*.

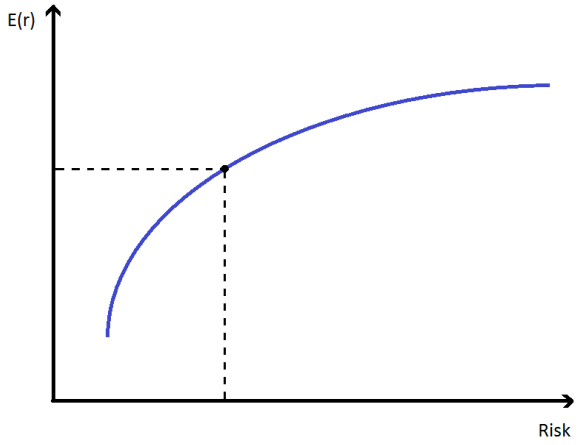
The data was subsequently analyzed in a number of ways in order to reach conclusions about the suitability of using variance and the PTIVM in the creation of efficient frontiers. The details of the tests can be found in section 10 - *Analysis*, but broadly speaking the tests can be categorized into two main groups.

- Tests of Absence of Preference

  These tests take advantage of the fact that each portfolio in a set has a set level of risk according to either variance or the PTIVM. If a risk measure is suitable, all portfolios in a set would therefore be positioned on the exact same place on the efficient frontier, if the efficient frontier is based on the same risk measure as the portfolios in that set have in common. This principle is graphically represented in Figure 4.

  If one ordering of portfolios in a set presents itself as a more prominent choice among our test subjects, it would cast doubt on the risk measure being homogenous in that particular set. For these tests, the variance static set will therefore be used to test the suitability of variance and the PTIVM static set will be used to test the suitability of the PTIVM.

Figure 4 – Tests of Absence of Preference



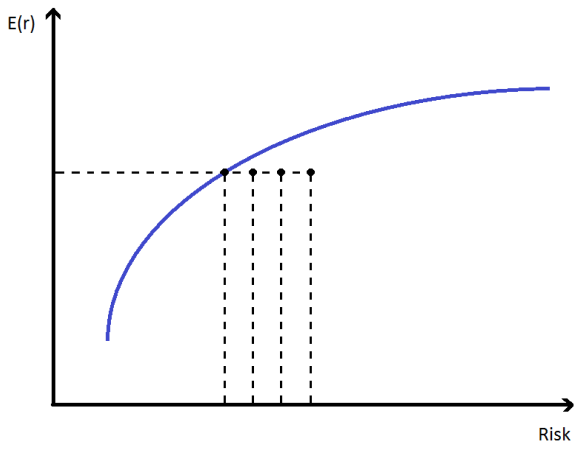
If a risk measure is suitable, and all portfolios in a set have the same expected outcome and risk level, all portfolios would be positioned on the exact same place on the efficient frontier. We categorize all tests taking advantage of this relationship as "Tests of Absence of Preference".

- Tests of Absolute Preference

The other type of tests we use exploit that each portfolio in a set has heterogeneous levels of risk according to either variance or the PTIVM. If a risk measure is suitable, only one of the portfolios could feasibly be positioned on the efficient frontier since they all share the same expected outcome, yet have heterogeneous risk levels. All portfolios should also be preferred in the reverse order of their riskiness. This principle is graphically presented in Figure 5.

If the order of portfolios which is ordered from the least risky to the most risky, according to the risk measure that is heterogeneous in that particular set, does not present itself as the most prominent choice among our test subjects, it would cast doubt on that risk measure. For these tests, the variance static set will therefore be used to test the suitability of the PTIVM and the PTIVM static set will be used to test the suitability of variance.

Figure 5 - Tests of Absolute Preference



If a risk measure is suitable, and all portfolios have the same expected outcome, yet different levels of risk, only one could feasibly be positioned on the efficient frontier and all portfolios have a clear order of preference. We categorize all tests taking advantage of this relationship as "Tests of Absolute Preference".

8. The Construction of the Questionnaire

8.1. Presentation – Form, Format and Framing

When conducting any kind of survey it is important to bear in mind how the presentation of the alternatives might influence the results. Many recent studies have found paradoxes in decision-making, and consequently, that fundamental assumptions of widely accepted descriptive theories of choice are often violated (Birnbaum & Navarrete (1998), Birnbaum (2004)).

As our survey aims to examine how people would determine riskiness in real life investment environments, it is not only central to minimize unwanted forms of bias in the results, we would also, to the extent possible, like to maintain the influence of the biases that affect typical investment decisions.

Birnbaum (2004) examined what impact the presentation of probabilities and outcomes have in making people violate stochastic dominance or not. In his empricial experimentation, he manipulates three factors concerning the way a gamble is displayed in order to examine the influence the factors have on the results. Our choices concerning how we display our gambles will be explained below on the basis of Birnbaum's chosen categorization: *Form, format and framing.*

8.1.1. Form

The *form* of a gamble concerns the manner in which its probability distribution is conveyed to the gambler. Birnbaum found that among the three categories of representation he studied, form had the largest influence over whether people violated stochastic dominance or not. We have already talked about the influence of form at length in section 6.1.2., since the construction of our risk measure was highly influenced by Birnbaums findings. We will not go over the justification here again, but we can recall the conclusion we arrived at: We took heed of Birnbauim et. al. (2008) findings and let all our outcomes in each gamble have equal probabilities.

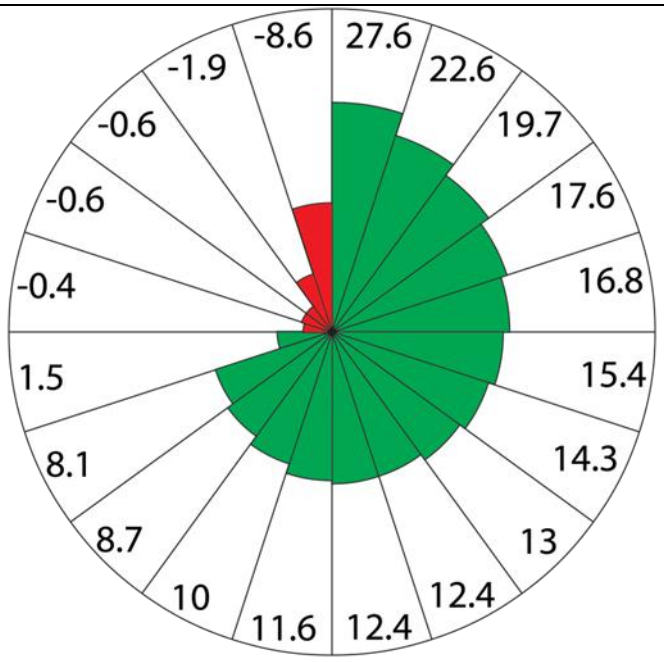
The choice of using twenty outcomes and no other number however is not a choice as much as it is dictated by necessity. Twenty branches was the minimal number we could use given the constraints imposed on us by the computational difficulties in creating gambles with fewer outcomes, as will be discussed in greater detail in section 8.2.

8.1.2. Format

The *format* refers to how the probability mechanism, consequences, and probabilities are presented to the participants (e.g. in a figure, table, plain text, etc.). Since we could not rely on our subjects taking an extraordinary time to understand the premises of our questions we chose to utilize the fact that people already have a mental framework concerning gambles where all outcomes are equally probable, yet the outcomes could differ: *The wheel of fortune* (WoF).

Unfortunately, this is a form of gamble associated more with fun and games than with serious investment decisions, but we regarded that as a small price to pay for such a powerful explanatory vehicle. Each gamble is therefore represented by a wheel of fortune consisting of twenty splinters of proportional size to the probability of each outcome, which in our case results in all splinters being of the same size. A representative example of the type of wheel of fortune we use is presented in Figure 6.

Figure 6 - Wheel of Fortune



Our final design for our display of a gamble: A stylized Wheel of Fortune.

Birnbaums (2004) found no specific format that completely eliminated violations of stochastic dominance but he found a positive effect when he facilitated comparisons by arranging the outcomes of competing gambles in sorted tables next to each other. We therefore, in addition to the WoFs, include sorted tables showing the outcomes of the four gambles next to each other.

8.1.3. Framing

Framing is the way events and consequences are described (or manipulated). Framing the outcomes of a gamble is to display it in such a way that unconscious psychological responses might influence the decision making without actually changing the objective situation. Tversky and Kahneman (1986) argue that event framing can be used either to "mask" a dominance relation or to make it more transparent. For example, colors can be used to enhance the effect of what is experienced as a gain or a loss. In economic news, increases in stock indices are usually displayed in green numbers while decreases are displayed in red. In fact, almost all information that we perceive in real life investment situations is framed one way or another. In our experiment, our objective is to display the gamble as transparent as possible and also to frame it in line with what is commonly known and accepted. In our WoFs, outcomes that result in gains therefore have their corresponding splinters partly colored green. Similarly, outcomes that result in a loss are emphasized with the color red. The sizes of these colored areas are also proportional to the amount that is gained or lost. To facilitate comparisons, the splinters in each wheel are ordered according to their return. (clockwise descending order, with the top right splinter of each WoF having the highest rate of return).

The instructions given to our respondents were designed to be as easy to comprehend as possible while being as neutrally phrased as possible. The questionnaire in its entirely is shown in Appendix 2 - *The Questionnaire.*

8.2. Generating Data

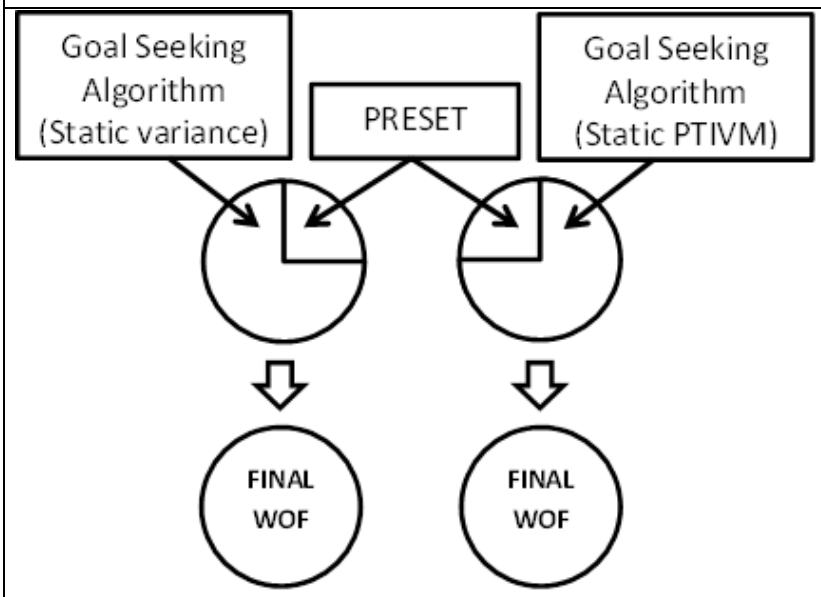
Before we explain how we generated the actual outcomes for each WoF, a side note needs to be made concerning the role the creator of this data has in influencing the final test results. It might seem all scientific endeavors are subject to this risk and any special mentioning thereof would be superfluous. The novelty of our testing procedure however, gives rise to an additional avenue to misrepresentation and source of erroneous conclusions, not entirely similar to the usual well known problems of data mining and outright data falsification. The problem arises from the creation of the WoFs in conjunction with the types of test we eventually perform. It might therefore be wise to revisit this section after the reader has developed a fuller understanding of the types of tests we carried out.

We require two sets of four WoFs. All WoFs in each set were created with two unwavering restrictions in order for our tests to be applicable: The twenty outcomes must have an expected outcome of 10% and they must have a set risk level. Since one can conceivingly create a very large number of WoFs that satisfy these two requirements, the question is then: How do one pick the WoFs one actually use in the testing procedure? Ideally one would want to pick the four WoFs that look the most dissimilar since our testing procedure is only capable of finding faults in a risk measure. All approval of a risk measure must therefore be implicit in nature. It would therefore be trivial for a

hypothetical nefarious scientist to select four WoFs in such a way our tests would find no faults in a given risk measure, even though a more prudent selection of WoFs would deem the same risk measure inappropriate.

Even if the hypothetical scientist in question was not particular nefarious, the problem of him introducing undue bias into the selection would still remain. The solution therefore seemed to be to remove the human element from the task of picking the WoFs as much as possible. If we had no influence and only let randomness dictate what WoFs were ultimately chosen, we would probably not end up with four WoFs that look as dissimilar as possible, so a compromise had to be made. Our solution to this dilemma was to preset one or several outcomes in the four WoFs in each WoF set in a predetermined fashion, while using the same presets for both WoF sets. The remaining unspecified outcomes were then calculated with the help of a computer program in order to satisfy the two constraints. Due to the nature of this optimization problem, there is no "one" solution, so we accepted the first proposed solution as the final solution for each WoF. A schematic illustration of this procedure can be seen in Figure 7.

Figure 7 – Creation of Outcome Distribution



Schematic illustration showing how the outcomes for the WoFs were generated. Note that each WoF share the same preset with another WoFs in the other WoF set.

The computer program in question was graciously provided to us by Frontline Systems Inc. and goes by the name *Premium Solver for Education V 7.0*. We used its evolutionary algorithm and even with a modern computer, with a manufacturing date circa 2007, just one proposed solution could take up to 48 hours to calculate. The more restrictions one imposes on the computer program the harder it will be for it to find a solution which is the main reason why we have twenty outcomes and not fewer, and why we allowed outcomes with one decimal instead of using only whole numbers.

As we previously mentioned the point of using presets were to increase the chance of creating WoFs that to the human eye would look different – an ambition which with the perfect risk measure would be a futile pursuit. The presets used were therefore chosen in such a way that we would capture different factors that previous literature suggest could have disproportional influence over how attractive a gamble is. The four WoFs in each WoF set therefore have the following presets:

- One WoF in each set (WoF B and F) has one predetermined outcome of 50% and one WoF in each set (WoF A and G) has one predetermined outcome of -20%. These presets were chosen since large positive or negative outcomes of low probability are overweighed according to Kahneman & Tversky (1992).
- We constructed one WoF in each set with five predetermined outcomes of 0% (WoF D and E). People are often more influenced by shifts from a given reference level than by levels in absolute terms, and in our survey this reference level – the "mental status quo" – corresponds to the respondents' current level of wealth before participating in the gamble. This corresponds to outcomes of zero, i.e the participant is neither better nor worse off.
- One WoF in each set has one predetermined outcome of 10% (WoF C and H). This is the unskewed WoF. Since our expected value is 10%, the preset contains what can be considered a typical outcome and the only thing it achieves is to make the rest of the outcomes slightly more volatile. Since the other WoFs are consciously skewed in one way or another, this WoF distinguishes itself on account of it being *typical.*

The outcome distributions that make up the final WoFs can be seen in Appendix 1 - TablesTable XIX and Table XX.

9. Data

Our data consist of the responses from the aforementioned questionnaire. The respondents, 94 in total, are all students at Stockholm School of Economics, attending classes in the third year or later. The data collected by the questionnaire consist of:

- The sex of the respondent.

There are two WoF sets in each questionnaire, The PTIVM static set and the variance static set. For each set, the respondents are asked to record their preference in two ways:

- An ordinal ranking of the four different WoFs in terms of preference. E.g. (C-D-A-B) where *C* is the most preferred WoF.
- A statement reflecting how strongly the respondent feel the preferences reflected in the respondent's ordinal ranking are to them, denoted in a number between zero and four, zero being "not important at all".

Additionally we have data concerning which of two different questionnaires a particular respondent answered, the only difference between the questionnaires being that the four WoFs for each risk measure is presented in a different order. This data will be used for a robustness check of our results. In Table IV and Table V the descriptive statistics are given showing the number of times and at what positions the different WoFs were ranked.

Table IV – Descriptive Statistics, All Observations, PTIVM Static Set

	1st choice	*2nd choice*	*3rd choice*	*4th choice*
A	*15*	*20*	*17*	*41*
B	*32*	*34*	*15*	*12*
C	*4*	*10*	*41*	*38*
D	*42*	*29*	*20*	*2*

The number of respondents that chose a certain WoF as their first, second, third or fourth choice in the PTIVM static set.

Table V - Descriptive Statistics, All Observations, Variance Static Set

	1st choice	*2nd choice*	*3rd choice*	*4th choice*
E	*33*	*37*	*16*	*7*
F	*41*	*35*	*6*	*11*
G	*10*	*16*	*40*	*27*
H	*9*	*5*	*31*	*48*

The number of respondents that chose a certain WoF as their first, second, third or fourth choice in the variance static set.

10. Analysis

We have the same type of data for both risk measures, and they will undergo an identical testing procedure. For the sake of brevity, the following elucidation on the tests performed will therefore be given in general terms. In contrast, the test results for the two risk measures are presented side by side, in order to facilitate comparisons.

We will first examine the quality and characteristics of the data collected. Depending on the results we will then proceed with the main analysis which has the structure outlined in chapter 7 - *Methodology*.

In aid of understanding the following sections it is prudent to recall that all WoFs have the same expected return and each wheel set is designed to have the same level of risk according to one of the two risk measures under examination.

10.1. Tests of Admissibility of Collected Data

Before we can start to press our data for answers to the questions of our hypotheses, we need to know whether we can use our collected data as a homogenous data set, more specifically we need to know:

- Does it matter in which order the WoFs are presented? If it does, we cannot assume that our test subjects' stated preferences rely only on the outcomes in the WoFs.
- Do men and women have different preferences? If we can see such differences we might want to treat the groups separately in order to be able to explain more of the diversity in the collected data.

10.1.1. Tests of Admissibility of Collected Data I – Test for Serial Position Effect

General principle: If the data collected through a questionnaire can be considered to have measured the preferences of a set of choices and nothing else, the ordering of the choices in the questionnaire should not affect the answers of the respondents.

In order to determine whether the preferences we can detect are "true" preferences or partly or wholly a result of a serial position effect, we created two versions of the same questionnaire which differ only in terms of what order the WoFs are presented in. The two questionnaire versions were randomly distributed to the respondents. If we can detect that the two versions of the questionnaire yielded different results we would therefore suspect that we are observing a serial positioning effect. The test we will use for this purpose is a test of association in contingency tables:

H_0: No association exists between the two types of questionnaires in the population.

H_1: Association may exist between the two types of questionnaires in the population.

The decision rule is:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} > \chi^2_{(r-1)(c-1),\alpha} \qquad \textbf{(Equation 5)}$$

Where O_{ij} is the observed value and E_{ij} is the expected value in a contingency table with r roes and c columns. The results are displayed in Table VI and Table VII.

Table VI - Test for Serial Position Effect, Variance Static Set

Chi-obs	*1,126*
Chi-krit	*12,840*
DF	*3*
p-value	*0,771*

Chi-test characteristics for the test examining whether there is a "serial position effect" among the WoFs that share the same variance.

Table VII - Test for Serial Position Effect, PTIVM Static Set

Chi-obs	*2,849*
Chi-krit	*12,840*
DF	*3*
p-value	*0,416*

Chi-test characteristics for the test examining whether there is a "serial position effect" among the WoFs that share the same PTIVM level.

The p-values are 0.771 and 0.416 respectively, which means none of the tests were significant or even close to significant at the 5% level. We therefore conclude that we have no reason to suspect that our data suffers from a serial positioning effect.

10.1.2. Tests of Admissibility of Collected Data II – Test for Gender Effect

General principle: If the data collected through a questionnaire can be considered a homogenous data set, the gender of the respondents should not contain information about the answers of the respondents.

As in the previous section we will perform a test of association in contingency tables in order to spot if men and women differ in the way they answer. The decision rule is as before shown in (Equation 5) and the results are shown in Table VIII and Table IX.

Table VIII - Test for Gender Effect, Variance Static Set

Chi-obs	*0,943*
Chi-krit	*12,840*
DF	*3*
p-value	*0,815*

Test characteristics for chi-test concerning whether one can detect gender differences in the rankings of WoFs that share the same variance.

Table IX - Test for Gender Effect, PTIVM Static Set

Chi-obs	*1,274*
Chi-krit	*12,840*
DF	*3*
p-value	*0,735*

Test characteristics for chi-test concerning whether one can detect gender differences in the rankings of WoFs that share the same PTIVM level.

The p-values are 0.815 and 0.735 respectively which means that neither of the tests were significant at the 5% level. We therefore conclude that we cannot detect any differences in how men and women respond to the questionnaire.

10.1.3. Tests of Admissibility of Collected Data – Concluding Remarks

Due to the novelty of our testing procedure and the importance of how a questionnaire is constructed in order not to bias the results in one way or another, robustness checks of the data collected is essential. The results shown in Table VI through Table IX do much to allay such fears. The p-values stretch from 0.416 to 0.815 which means none of the tests were significant or even close to significant at the 5% level. We therefore conclude that we have no reason to suspect that there is a serial positioning effect or any difference in how men and women respond to our questionnaire. We can therefore treat all our observations as one homogenous data set, and consequently all tests that follow do so.

10.2. Tests of Absence of Preference

General principle: If a risk measure is suitable, it should be impossible to form a preference for one gamble among other gambles which have the same risk level and expected return.

If all gambles have the same risk level and the same expected return all those gambles would be located on the same point on the efficient frontier. If that efficient frontier was constructed using a

suitable risk measure all those gambles would be considered equally attractive/unattractive. For our purposes, since the four WoFs in each wheel set share the same risk level (the PTIVM for the PTIVM static set and variance for the variance static set) and the same expected return, we can test if all WoFs are considered equally attractive by looking for randomness in the preferences of our test subjects. If the risk measure under examination is suitable we would not expect to find any preferences for any one WoF over another in the wheel set where all WoFs share the same risk level. A graphical representation of this principle was provided in section 7 - *Methodology*.

We are interested in an intuitive understanding of risk, so we will put more emphasis on the test subjects' actions rather than on their expressed opinions. In doing so, we hope to be able to pick up on implicit preferences even if the subjects are unaware of having any. Our method therefore dictates that the test subjects are not permitted to indicate that they consider two or more WoFs as equally desirable or neglect to include one or more WoFs from the ranking altogether. By forcing them to make a decision we cannot conclude anything from just one person, but on a group level, if the null hypothesis is true we expect to find nothing but randomness in the sum of their collective expressed preferences.

10.2.1. Test of Absence of Preference I – Aggregated Level

The first test of randomness will consist of a Pearson's chi-square test. This will test on an aggregate level if randomness can be seen in our test subjects' preferences. The test looks at how often the 24 possible combinations (e.g. A-B-C-D) occur in each WoF set. The test statistic in (Equation 6) was used where O_i is the number of observations in each group and E_i is the expected number of observations in each group given that people chose portfolios randomly. The tests have (*r-1*)(*c-1*) degrees of freedom.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \textbf{(Equation 6)}$$

The test stipulates the expected number in each group, E_i, to be at least five in order to not be biased. The combinations were therefore grouped in sets of three in order to satisfy the requirements of an unbiased Pearson's chi-square test. The null and alternative hypothesis is:

H_0: People choose portfolios randomly, i.e. all combinations occur with equal probability

H_1: People do not choose portfolios randomly, i.e. not all combinations occur with equal probability

Decision rule: Reject H_0 if:

$$\chi^2_{\text{Obs}} > \chi^2_{\text{Crit}}$$

The results of the tests can be seen in Table X and Table XI.

Table X – Test of Absence of Preference I, Variance Static Set

χ^2_{Obs}	*117.064*
$\chi^2_{Crit.5\%}$	*14.07*
Significance	*Sig.*
p-value	*0.000*

Test characteristics for chi test examining whether WoFs that share the same variance have been selected randomly.

Table XI – Test of Absence of Preference I, PTIVM Static Set

χ^2_{Obs}	*55.447*
$\chi^2_{Crit.5\%}$	*14.07*
Significance	*Sig.*
p-value	*0.000*

Test characteristics for chi test examining whether WoFs that share the same PTIVM level have been selected randomly.

The two Pearson's chi-square tests generated χ^2_{Obs} values of 117.064 and 55.447 respectively. Both tests are significant since both χ^2_{Obs} exceed the $\chi^2_{Crit.5\%}$ of 14.07 by a large margin. Our data therefore clearly suggests that on the aggregate levels our test subjects showed some form of preference when choosing between the 24 permutations. A perfect risk measure would have our test subjects choose between four WoFs that were equally attractive and therefore no preferences should be able to be detected. We therefore conclude that neither variance, nor the PTIVM is a *perfect* risk measure. However since the χ^2_{Obs} for the variance static WoF set was more than twice the size of the χ^2_{Obs} for the PTIVM static set, we can say that the PTIVM came closer to the platonic idea of a perfect risk measure in this test – subjected to the caveats mentioned in 8.2.

10.2.2. Test of Absence of Preference II - Pairwise

Since the Pearson chi-square test is a test on an aggregate level, it has the weakness that it will not give us any fine grained information. If our test subjects regard only some of the WoFs as equally attractive, the tests in the previous section would not show it. This type of data is of interest to us since such results could give us clues about what a suitable risk measure should look like. That is, we could use the finding of two WoFs that are regarded as equally attractive and use those WoFs as a litmus test of what a risk measure, not yet created, must regard as equally risky in order to be acceptable. We could also use the same findings to find traits among the WoFs that are not regarded as equally attractive and try to incorporate an awareness of such traits in a new risk measure. In order to be able to detect differences and commonalities among the WoFs, we have therefore devised a method in which we can examine the preferences embedded in our data with more detail.

With the purpose of extract the maximum amount of information from our data we must first transform it. The general principle of this transformation is that an ordinal ranking of four alternatives (A-D) can be completely described if one have the answers to at most six questions. The questions are:

- Is A preferred to B or vice versa?
- Is A preferred to C or vice versa?
- Is A preferred to D or vice versa?
- Is B preferred to C or vice versa?
- Is B preferred to D or vice versa?
- Is C preferred to D or vice versa?

The questions are not necessarily independent. If A is preferred to B and B is in turn preferred to C, then A must by necessity also be preferred to C. This is not a problem as long as we do not try to create a statistical test that group the individual questions together, but something one need to be aware of.

We can now test for randomness by reframing the six questions into proportions, e.g. the answer to the question: "Is A preferred to B or vice versa?" is contained in the ratio where A is preferred to B in $\hat{p}_{AB}$ of all cases. We can then use a simple proportion test for large samples in order to see if two WoFs are equally attractive. The null hypothesis being that the proportion of one WoF being preferred to another is 0.5, i.e. we can see only randomness.

$H_0: p = 0.5$, *The two WoFs are equally attractive.*

$H_1: p \neq 0.5$, *The two WoFs are not equally attractive.*

Decision rule: Reject H_0 if:

$$Z_{Obs} = \frac{\hat{p}_{xy} - p_0}{\sqrt{p_0(1-p_0)/n}} > Z_{Crit.} \quad \textbf{(Equation 7)}$$

or

$$Z_{Obs} = \frac{\hat{p}_{xy} - p_0}{\sqrt{p_0(1-p_0)/n}} < -Z_{Crit.} \quad \textbf{(Equation 8)}$$

The double sided test yields the following results for the different comparisons.

Table XII – Test of Absence of Preference II – Variance Static Set

	U(E)>U(F)	U(E)>U(G)	U(E)>U(H)	U(F)>U(G)	U(F)>U(H)	U(G)>U(H)
$\hat{p}_x$	0,330	0,479	0,287	0,851	0,404	0,138
p_0	0,500	0,500	0,500	0,500	0,500	0,500
Z_{Obs}	-3,301	-0,413	-4,126	6,807	-1,857	-7,014
$Z_{Crit.2.5\%}$	1,960	1,960	1,960	1,960	1,960	1,960
Significance	Sig.	Not sig.	Sig.	Sig.	Not sig.	Sig.
p-value	0,001	0,680	0,000	0,000	0,063	0,000

Test characteristics for proportion tests, concerning pairwise ordinal comparisons of the utility of WoFs sharing the same variance.

Table XIII - Test of Absence of Preference II – PTIVM Static Set

	U(A)>U(B)	U(A)>U(C)	U(A)>U(D)	U(B)>U(C)	U(B)>U(D)	U(C)>U(D)
$\hat{p}_x$	0,457	0,745	0,830	0,787	0,819	0,617
p_0	0,500	0,500	0,500	0,500	0,500	0,500
Z_{Obs}	-0,825	4,745	6,395	5,570	6,189	2,269
$Z_{Crit.2.5\%}$	1,960	1,960	1,960	1,960	1,960	1,960
Significance	Not sig.	Sig.	Sig.	Sig.	Sig.	Sig
p-value	0.409	0.000	0.000	0.000	0.000	0.023

Test characteristics for proportion tests, concerning pairwise ordinal comparisons of the utility of WoFs sharing the same PTIVM level.

In Table XII we can see that only two double sided tests were not significant, i.e. the tests did not show that one WoF were preferred to the other WoF. The two tests were U(E)>U(G) and U(F)>U(H) with p-values of 0.680 and 0.063. A naive conclusion would therefore be that among the WoFs that share the same variance, only two real levels of attractiveness were represented in the sample. However, when these tests are interpreted in conjunction it introduces the problem of them being more or less correlated. A more prudent approach might be to observe the largest p-value, consider the possibility of data mining, then try to derive some common traits from the two WoFs that cannot statistically be shown to be different. Since we do not have the means nor the inclination to go down this iterative road to further refinement of a perfect risk measure, we will perform no further testing along those lines.

In Table XIII we can see that in the WoF set where all WoFs share the same PTIVM level, only one comparative test is not significant: U(A)>U(B) with a p-value of 0.409. This suggests that WoF A and B were regarded as equally attractive by our test subjects. This notion is further reinforced when looking at the other comparisons where we can see that the sample proportions are roughly the same for A and B in all our tests; U(A)>U(C) and U(B)>U(C) are 0.745 and 0.787 and U(A)>U(D) and U(B)>U(D) are 0.830 and 0.819. As before we will not further investigate whether this is indeed the case since it is outside the scope of our thesis.

10.2.3. Test of Absence of Preference III – Reported Preference

As we have mentioned, we suspected that the test subjects' actions could tell another story then their stated preferences, but our focus is trying to deduce if any real preferences existed in the preferences that were embedded in the rankings the respondents were compelled to provide us with. We do, however, have data of self-reported strength of preference.

Our test subjects were asked to provide us with an answer to the question:

How strongly did you prefer your most preferred wheel of fortune to your least preferred wheel of fortune?

The answers were given as a rating between four to zero where four is "Very strongly" and zero is "Not at all". The question we ask ourselves is: When we examine the answers to this question, how should we interpret results that are in discordance with our results from our main form of analysis?

The field of experimental psychology is filled with examples where test subjects' stated opinions were in conflict with their actions and conventional wisdom seems to be that actions speak louder than words. If our aim is to provide a descriptive theory of how people behave, stated preferences are only important as long as they are in correspondence with observed behavior. Stated preferences should therefore, perhaps, be the last resort if one wishes to uncover true preferences but they, nevertheless, tell an interesting story as a measure of self-awareness.

We examined data of self-reported strength of preference, using a Student's t-test to see if the mean of the ranking differ from zero:

H_0: The mean ranking of strength of preference is zero, i.e. our test subjects reported no preferences.

H_1: The mean ranking of strength of preference larger than zero, i.e. our test subjects did report preferences.

We also examined if the two risk measures had differing levels of self-reported strength of preference with another Student's t-test with the hypotheses:

H_0: The mean ranking of strength of preference is the same for both WoF sets.

H_1: The mean ranking of strength of preference differ between the two WoF sets.

The results for tests examining the mere presence of stated preferences were both significant. The null hypothesis of no preferences can be rejected for both risk measures; i.e. people did have clear preferences for ordering both sets. This result is in agreement with the previous tests performed. Both the test for the variance static set and the PTIVM static set were highly significant and the test characteristics can be seen in Table XV and Table XVI.

The results for the comparative test can be found in Table XIV. A bit surprisingly, we can see that we cannot reject the null hypothesis and conclude that one risk measure prompted a different level of self-reported preference than the other. This is a disagreeing finding when compared to the findings in 10.2.1 which suggested that the orderings of gambles in the PTIVM static set were more random than the orderings of gambles in the variance static set.

Table XIV – Test for Difference of Reported Preferences between the PTIVM and Variance

t_{Obs}	0.2464
$t_{Crit.5\%}$	1.645
Significance	Not sig.
p-value	0.4028

Test characteristics for t-test examining whether the means of stated preferences is the same for both WoF sets

Table XV – Test for Difference of Reported Preferences and Zero, PTIVM Static Set

t_{Obs}	28.7155
$t_{Crit.5\%}$	1.645
Significance	Sig.
p-value	0.0000

Test characteristics for t-test examining whether the mean of stated preferences for the PTIVM static set differ from zero.

Table XVI – Test for Difference of Reported Preferences and Zero, Variance Static Set

t_{Obs}	29.5013
$t_{Crit.5\%}$	1.645
Significance	Sig.
p-value	0.0000

Test characteristics for t-test examining whether the mean of stated preferences for the variance static set differ from zero

10.3. Tests of Absolute Preference

General principle: If a risk measure is suitable, among a group of gambles, the gamble with the lowest risk should be preferred as long as all gambles have the same expected return.

If all gambles have the same expected return but different risk levels, only the one with the least amount of risk can be positioned on the efficient frontier. As long as all gambles have heterogeneous risk levels we should also be able to deduce a clear order of preference among them as long as the risk measure is well suited to the task. Each of our two WoF sets have a heterogeneous risk level according to one risk measure and a homogenous risk level according to the other.

The methodology used is based on the fact that that there are twenty-four permutations in a set of four members, e.g. {A,B,C,D}. Since we have four distinct risk levels in each WoF set, the risk measure under examination only corresponds to one of these permutations. For example, under the unrealistic assumptions of homogenous preferences and that variance is a perfect risk measure we would expect all our subjects to order the PTIVM static set {A,B,C,D} as C-D-A-B (least risky-to riskiest according to variance). We could therefore examine what proportion of our test subjects has chosen this

particular permutation and compare it to the expected proportion we would see under the null hypothesis that all permutations are equally likely. This however would fail to take into consideration that even though only one permutation is the exact order we would expect if the risk measure we are examining is a good predictor, the other permutation are not all equally incorrect. If C-D-A-B is the correct order we would still regard D-C-A-B as more correct than B-A-D-C, and if the risk measure under examination is suitable we would also expect the former permutation to be chosen more often than the latter. We will therefore include even non-perfect permutations in our testing procedure. We can determine how well a particular order fits an ideal order with the six questions introduced earlier. Only one permutation will correspond to the ideal order completely. Four permutations will fit the ideal order to 5/6 or higher. Nine permutations will fit the ideal order to 4/6 or higher. We will test all these three degrees of fitness for each risk measure.

Formally, the test utilized is a proportion test for large samples. The hypotheses are:

$H_0: p = p_0$, *The permutations predicted by the risk measure occur only at the same frequency as randomness would suggest.*

$H_1: p > p_0$, *The permutations predicted by the risk measure occur more frequently than randomness would suggest.*

where p_0 is determined by the number of permutations examined depending on the degree of fitness being examined.

Decision rule: Reject H_0 if:

$$Z_{Obs} = \frac{\hat{p}_x - p_0}{\sqrt{p_0(1-p_0)/n}} > Z_{Crit.} \qquad \textbf{(Equation 9)}$$

From Table XVII we can see that variance failed to predict the comparative popularity of the WoFs in the variance static WoF set. None of the fitness levels tested were significant (lowest p-value is 0.78), and not one test subject chose the exact order variance would have predicted.

In Table XVIII we can see that PTIVM fared much better. Almost 20% responded exactly as PTIVM would have predicted (p-value 0.191) in the PTIVM static WoF set. We can further see that all fitness levels were highly significant with p-values of 0.000.

Examining the underlying data we can even see that no other predicted permutation would have yielded a better result on all three fitness levels. PTIVM predicted the order F-E-G-H, which yielded

the observed test statics of 8.256, 11.716 and 7.269 for the three fitness levels. The closest other permutation is E-F-G-H which produced observed test statics of 7.830, 8.672 and 7.785.

Table XVII - Tests of Absolute Preference, PTIVM Static Set

% Fit	*4/6*	*5/6*	*6/6*
$\hat{p}_x$	*0.085*	*0.064*	*0.000*
p_0	*9/24*	*4/24*	*1/24*
Z_{Obs}	*-5.806*	*-2.675*	*-2.022*
$Z_{Crit.5\%}$	*1.65*	*1.65*	*1.65*
Significance	*Not sig.*	*Not sig.*	*Not sig.*
p-value	*1.000*	*0.996*	*0.780*

Characteristics for three proportion tests for three fitness levels, where 6/6 is a perfect fit. The tests examine how well variance can explain the choices in the WoF set where the WoFs have heterogeneous variance levels.

Table XVIII- Tests of Absolute Preference, Variance Static Set

% Fit	*4/6*	*5/6*	*6/6*
$\hat{p}_x$	*0.787*	*0.617*	*0.191*
p_0	*9/24*	*4/24*	*1/24*
Z_{Obs}	*8.256*	*11.716*	*7.269*
$Z_{Crit.5\%}$	*1.65*	*1.65*	*1.65*
Significance	*Sig.*	*Sig.*	*Sig.*
p-value	*0.000*	*0.000*	*0.000*

Characteristics for three proportion tests for three fitness levels, where 6/6 is a perfect fit. The tests examine how well PTIVM can explain the choices in the WoF set where the WoFs have heterogeneous PTIVM levels.

In conclusion, PTIVM performed as well as could be in this test and variance performed very poorly. It is still possible that this is just due to a friendly data set and a more challenging set of WoFs would have shown the inability of PTIVM in distinguishing smaller nuances.

11. Suggestion for Further Research

11.1. Suspension of Disbelief

The use of real monetary gain instead of just hypothetical gains and losses would most likely have benefitted the validity of the data collected. It is hard to quantify how much so.

Since the hypothetical scenario involved the gamble of the entirety of one's personal wealth, it would also be impossible; although one could conceivingly simulate it to closer degree if the study was repeated in a poverty stricken third world country.

11.2. Inclusion of Higher Moments

The inclusion of higher moments is as we have mentioned earlier an avenue of research that is not entirely unexplored. Empirical research into the question if higher moments capture the way people intuitively reason about risk is however not the typical focal point of the research that has been conducted.

Our data is not really suited to any in-depth exploration of whether the inclusion of higher moments can better the performance of variance. In our study of previous literature we found that when an incorporation of higher moments was attempted, both skewness and kurtosis were added as variables in conjunction, in order to create a four dimensional surface representing the efficient frontier. Since our survey data only has (in the case of the variance static set) variance and expected return fixed, this would leave us with two unconstrained variables which precludes the use of our methodology. That is, the four WoFs are no longer guaranteed to be positioned on the same spot on the efficient frontier or a clear order of preference cannot be established.

It is possible to utilize part of our method even in the examination of the higher moments, but considerably more computational power would be required since more variables are involved.

11.3. Co-relationship

We, as we stated at the offset, did not formalize a method in which we could construct a real efficient frontier based on real assets. This would however be the natural next step once one has found a risk measure that is sufficiently attractive to warrant the effort.

12. In conclusion

12.1. Method

As our testing methodology is novel we feel that our tests of admissibility of collected data fill an important role. We were pleased to find that no significant problems of serial position effect or gender effects were found. The lack of serial position effect is especially pleasing since its absence indicates that our respondent indeed took the time to report their true preferences even though we had no influence over their effort level. We are also pleased that our method made no assumptions about our respondents being aware of their preferences, since our test of stated preferences do suggest that a certain lack of insight might exist.

Any study that is based on data collected through a questionnaire is subjected to the risk of biased results due the respondents being unduly influenced or lacking a clear understanding of the tasks they were asked to perform. Although we have made considerable efforts ensuring this would not be the case for our study, ultimately this is hard to qualify. We do feel that our efforts to create a questionnaire that could be readily understood by our respondents were successful. This is of course a conclusion that must be somewhat speculative, as we have already mentioned, but based on the respondents' lack of questions and good return rate of the questionnaire, we have no reason to suspect otherwise.

The testing procedure, due to multiple tests being performed, provided means to ensure that the results were consistent both with each other and with one's intuition. The tests performed were very efficient for a small data set, yet they would also be suitable for a large data set with very few changes. All in all, we are very pleased with the quality of the data collected and the methods used to analyze it.

12.2. Results

Since we have performed multiple tests, we need to take the individual tests' strengths and weaknesses into account in order to arrive at conclusions of the more general kind. Of the two types of tests we performed, intuition would tell us that the tests measuring the absence of preference must be considered more sensitive tests than tests measuring the existence of preference. This follows, since in the former case, the absence of spreads (since the tests are of homogeneous risk levels) would enable the tests to detect even minute differences in suitability. In the case of tests examining the existence of preference, a substandard risk measure could appear flawless if the spreads are sufficiently large so that the coarseness of the risk measure is obscured.

This intuition is also what our findings reflect. The tests of absence of preference do find variance as well as PTIVM wanting, but when we test for absolute preferences PTIVM reigns supreme. If we were to perform the test of absolute preferences over smaller spreads we would therefore expect to see the PTIVM fail to make accurate predictions at some point as the spreads becomes smaller.

Together our tests indicate that variance is not suitable as a descriptive risk measure of human risk heuristics and in a real world scenario (one where one cannot assume that returns are normally distributed) one of the portfolios on an efficient frontier based on variance is not necessarily the most utility maximizing portfolio choice for a typical agent.

The PTIVM share the same verdict, judged with stern enough eyes. While it is clearly more descriptive than variance as a measure of human risk heuristics, one of the portfolios on an efficient frontier based on the PTIVM is not necessarily the most utility maximizing portfolio choice for a typical agent either.

What we can say is that we did manage to create a risk measure with better descriptive properties while still using the two-parameter approach favored by Markowitz. The PTIVM's failure to pass our most stringent tests were of little surprise, it is after all only a first effort. Its performance in comparison to variance however, do give credence to the notion that a much better risk measure can be found. We would like to think that if such a risk measure were discovered the tests outlined in this paper could be used to demonstrate its properties.

13. References

Acerbi, C. (2002). Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion. *Journal of Banking & Finance 26*, 1505-1518.

Bergh, G., & van Rensburg, P. (2008). Hedge Funds and Higher Moment Portfolio Selection. *Journal of Derivatives & Hedge Funds, Vol. 14, No. 2*, 102-126.

Bernoulli, D. (1738). *Commentaries of the Imperial Academy of Science of Saint Petersburg*.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes 95*, 40-65.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of Stochastic Dominance and Cumulative Independence. *Journal of Risk and Uncertainty 17*, 49-78.

Birnbaum, M. H., Johnson, K., & Longbottom, J.-L. (2008). Tests of Cumulative Prospect Theory with graphical displays of probability. *Judgment and Decision Making, Vol 3, No 7*, 528-546.

Burr Williams, J. (1938). *The Theory of Investment Value.* Fraser Pub Co.

Chen, Z., & Yi Wang, Y. (2008). Two-sided Coherent Risk Measures and their Application in Realistic Portfolio Optimization. *Journal of Banking & Finance 32*, 2667.

Coombs, C. H., Bezembinder, T. G., & Goode, F. M. (1967). Testing Expectation Theories of Decision Making without Measuring Utility or Subjective Probability. *Journal of Mathematical Psychology 4*, 72-103.

Einhorn, D. (2008). Private Profits and Socialized Risk. *Global Association of Risk Professionals 8*, 12.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 427-465.

FrontlineSolvers - Developers of Excel Solvers. (n.d.). Retrieved March 15, 2010, from www.solver.com

Hadar, J., & Russell, W. R. (1969). Rules for Ordering Uncertain Prospects. *American Economical Review, Vol 59*, 25-34.

Hagströmer, B., Anderson, R. G., & Binner, J. M. (2007). Expanding the Merit of Utility Maximisation for Portfolio Choice. 1-33.

Hagströmer, B., Anderson, R. G., & Binner, J. M. (2007). Mean-Variance vs. Full-Scale Optimization: Broad Evidence for the UK. 1-25.

Kahneman, D., & Smith, V. (2002). *Foundations of Behavioral and Experimental Economics 16.* Stockholm: The Royal Swedish Academy of Sciences.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decisioin Under Risk. *Econometrica Vol 47*, 263-291.

Kerstens, K., Mounir, A., & Van de Woestyne, I. (2008). Geometric Representation of the Mean-Variance-Skewness Portfolio Frontier Based upon the Shortage Function. *HUB Research Paper*, 1-32.

Lakonishok, J., & Shapiro, A. C. (1986). Systematic Risk, Total Risk and Size as Determinants of Stock Market Returns. *Journal of Banking and Finance 10*, 115-132.

Levy, M., & Levy, H. (2002). Prospect theory: Much Ado with Nothing? *Management Science, 48*, 1334-1349.

Mandelbrot, B., & Hudson, R. L. (2004). *The (Mis)Behavior of Markets: A fractal View of Risk, Ruin and Reward.* London: Profile Books.

Mandelbrot, B., & Taylor, H. M. (1967). On the Distribution of Stock Price Differences. *Operations Research 15*, 1057-1062.

Maringer, D., & Parpas, P. (2009). Global Optimization of Higher Order Moments in Portfolio Selection. *Journal of Global Optimization 43*, 219-230.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 77-92.

Markowitz, H. (1991). Foundations of Portfolio Theory. *The Journal of Finance*, 469-477.

Markowitz, H. M. (1990). Foundations of Portfolio Theory. *Nobel Lecture* (pp. 280-287). New York: The City University of New York.

Mhiri, M., & Prigent, J.-L. (2010). International Portfolio Optimization with Higher Moments. *International Journal of Economics and Finance, Vol 2*, 157-169.

Mitton, T., & Vorkink, K. (2007). Equilibrium Underdiversification and the Preference for Skewness. *Review of Financial Studies 20*, 1255-1288.

Perrings, C., Folke, C., & Mäler, K.-G. (1992). The Ecology and Economics of Biodiversity Loss: The Research Agenda. *Ambio 21*, 201-211.

Premium Solver for Education, version 7.0. (n.d.). FrontlineSolvers.

Rubinstein, M. (2002). Markowitz's "Portfolio Selection": A Fifty-Year Retrospective. *The Journal of Finance*, 1041-1045.

Ruszczynski, A., & Vanderbei, R. J. (2003). Frontiers of Stochastically Nondominated Portfolios. *Econometrica Vol 71*, 1287-1297.

Samuelson, P. A. (1960). The St. Petersburg Paradox as a Divergent Double Limit. *International Ecomomic Review 1*.

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

Sharpe, W. F. (1999, 10 08). Retrieved 11 06, 2011, from Mean, Variance and Distributions: http://www.stanford.edu/~wfsharpe/mia/rr/mia_rr1.htm

Sharpe, W. F. (n.d.). *Mean, Variance and Distributions*. Retrieved 06 06, 2010, from http://www.stanford.edu/~wfsharpe/mia/rr/mia_rr1.htm

Stacey, J. (2008). Multi-Dimensional Risk and Mean-Kurtosis Portfolio Optimization. *Journal of Financial Management and Analysis, 21*, 47-56.

Statman, M. (1987). How Many Stocks Make a Diversified Portfolio? *Journal of Financial and Quantitative Analysis 22*.

Thaler, R. H. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, 199.

Tversky, A., & Kahneman, D. (1986). Rational Choice and Framing of Decisions. *The Journal of Business 59 (4)*, 251-278.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*, 297-323.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*, 297-323.

Von Neumann, J, & Morgenstern, O. (1953). *Theory of Games and Economic* (3rd ed.). Princeton University Press.

14. Appendix 1 - Tables

Table XIX – WoF Set I

A	B	C	D
31.4%	50.0%	27.6%	29.9%
23.1%	25.9%	22.6%	29.0%
22.3%	20.8%	19.7%	28.5%
21.6%	18.2%	17.6%	25.2%
21.8%	18.6%	16.8%	14.7%
14.2%	11.3%	15.4%	14.6%
12.3%	10.8%	14.3%	14.4%
10.5%	9.7%	13.0%	13.1%
10.5%	9.4%	12.4%	9.1%
8.9%	7.0%	12.4%	5.6%
8.3%	6.0%	11.6%	5.4%
7.6%	5.5%	10.0%	5.1%
6.5%	4.5%	8.7%	3.6%
5.4%	4.0%	8.1%	1.0%
4.5%	2.6%	1.5%	0.8%
4.3%	0.8%	-0.4%	0.0%
4.1%	0.4%	-0.6%	0.0%
1.6%	-1.1%	-0.6%	0.0%
1.2%	-1.7%	-1.9%	0.0%
-20.0%	-2.9%	-8.6%	0.0%

The outcomes that make up the four WoFs in WoF set I.

Table XX – WoF Set II

E	F	G	H
35.3%	50.0%	32.8%	28.3%
27.6%	17.7%	20.7%	28.0%
26.5%	17.6%	20.5%	24.8%
25.4%	16.2%	20.0%	19.3%
19.8%	15.0%	19.7%	18.4%
12.3%	13.4%	19.4%	15.5%
11.7%	11.0%	15.7%	15.6%
10.5%	9.8%	13.3%	13.6%
9.7%	9.8%	12.8%	12.5%
6.8%	9.3%	11.2%	10.0%
6.6%	7.3%	7.8%	9.6%
3.4%	6.8%	6.3%	9.1%
3.2%	5.3%	3.8%	8.1%
1.5%	4.5%	3.7%	4.9%
0.0%	2.9%	2.8%	5.1%
0.0%	2.7%	2.7%	1.2%
0.0%	1.1%	2.6%	-0.4%
0.0%	0.4%	2.3%	-2.7%
0.0%	0.1%	1.9%	-7.3%
-0.5%	-1.1%	-20.0%	-13.4%

The outcomes that make up the four WoFs in WoF set II.

Table XXI – Ordinal Permutations

PTIVM	
A-B-C-D	0
A-B-D-C	4
A-C-B-D	1
A-C-D-B	0
A-D-B-C	6
A-D-C-B	4
B-A-C-D	0
B-A-D-C	6
B-C-A-D	1
B-C-D-A	6
B-D-A-C	8
B-D-C-A	11
C-A-B-D	0
C-A-D-B	4
C-B-A-D	1
C-B-D-A	0
C-D-A-B	0
C-D-B-A	0
D-A-B-C	8
D-A-C-B	2
D-B-A-C	6
D-B-C-A	24
D-C-A-B	2
D-C-B-A	0
	94

Variance	
E-F-G-H	19
E-F-H-G	10
E-G-F-H	0
E-G-H-F	3
E-H-F-G	0
E-H-G-F	1
F-E-G-H	18
F-E-H-G	15
F-G-E-H	6
F-G-H-E	2
F-H-E-G	0
F-H-G-E	1
G-E-F-H	1
G-E-H-F	1
G-F-E-H	4
G-F-H-E	1
G-H-E-F	2
G-H-F-E	1
H-E-F-G	2
H-E-G-F	1
H-F-E-G	1
H-F-G-E	0
H-G-E-F	3
H-G-F-E	2
	94

Number of respondents who have chosen each possible permutation of ordinal order.

15. Appendix 2 - The Questionnaire

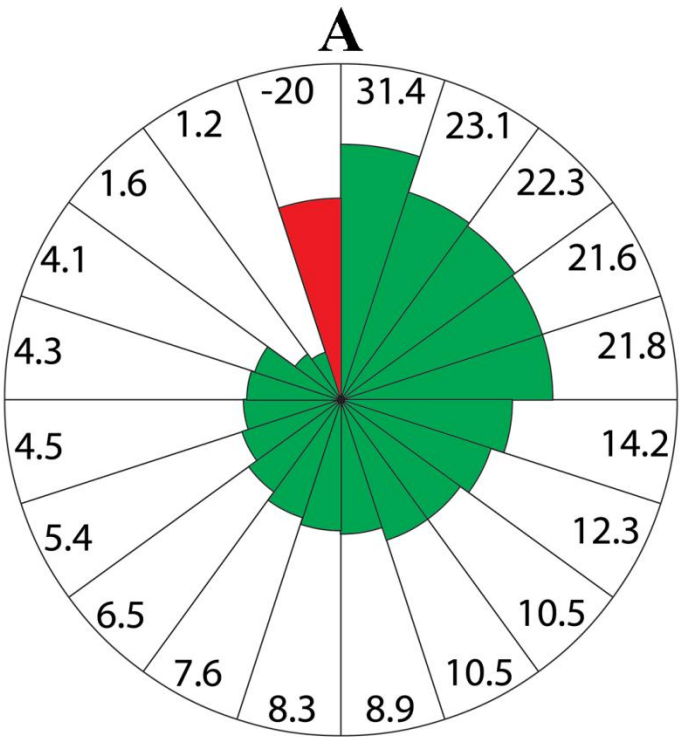
Survey of human risk assessment

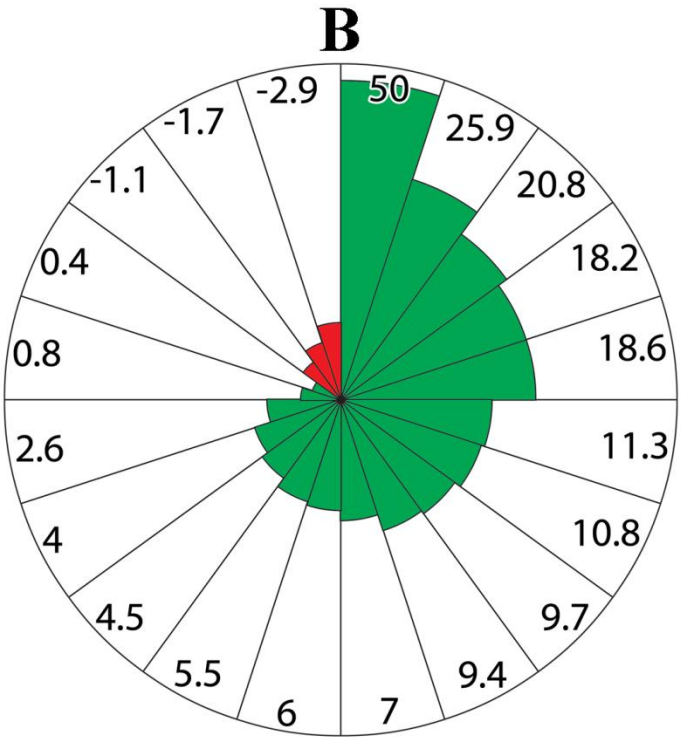
Assume that you are forced to invest everything you own in one out of four potential types of investments. Each type of investment has twenty potential outcomes. All outcomes are equally likely to occur. Each investment can therefore be represented as a gamble on a "wheel of fortune", where the wheel is divided into twenty slices and each slice is assigned one outcome. Each potential outcome is in the form of a percentage yield (the yield can be negative). The potential investments will all be represented by these kinds of wheels of fortune.

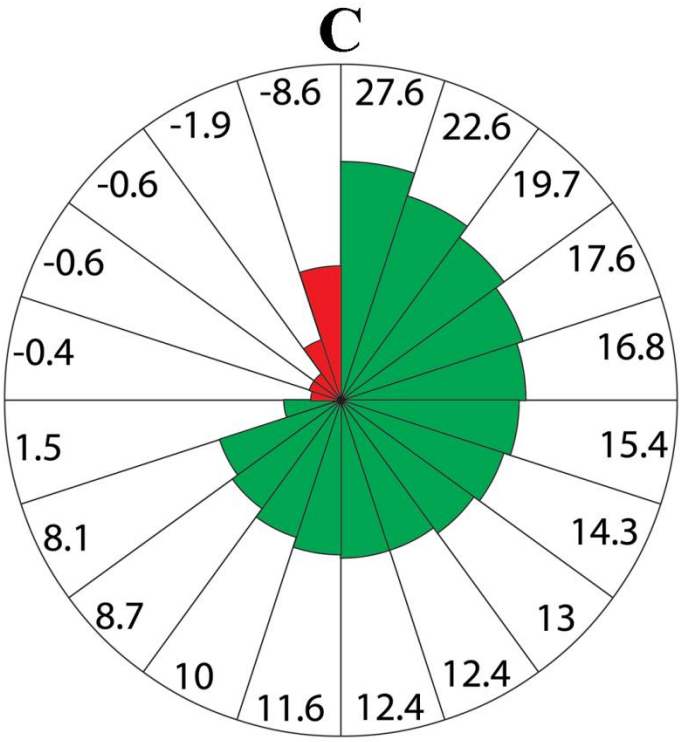
You will find two sets of wheels on the following pages. The outcomes of both sets of wheels are also represented in tabular form for your convenience. Please carefully examine the two sets of wheels in turn and answer all the questions. Both sets should be considered independent of each other. As a closing note, it is important to remember that the **average returns are the same for all wheels of fortune**.

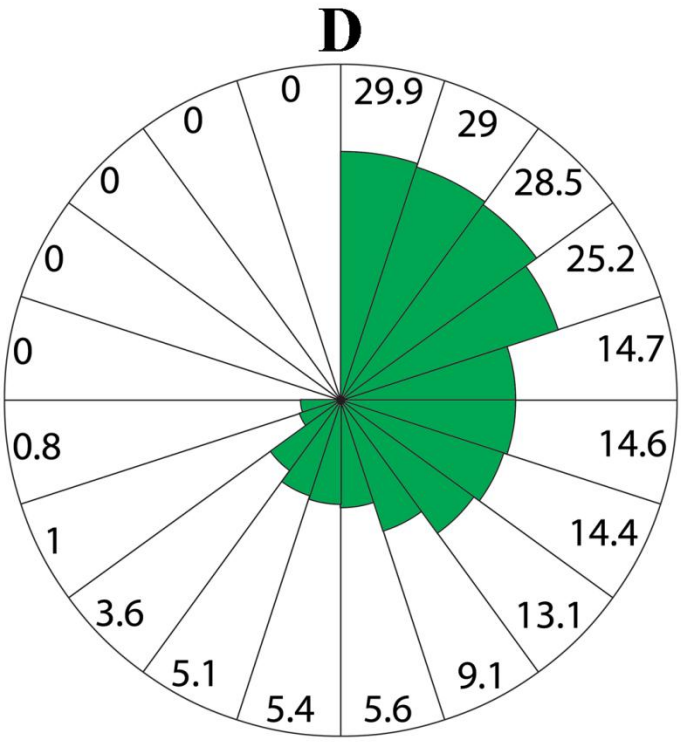
Gender :

- ☐ Male
- ☐ Female









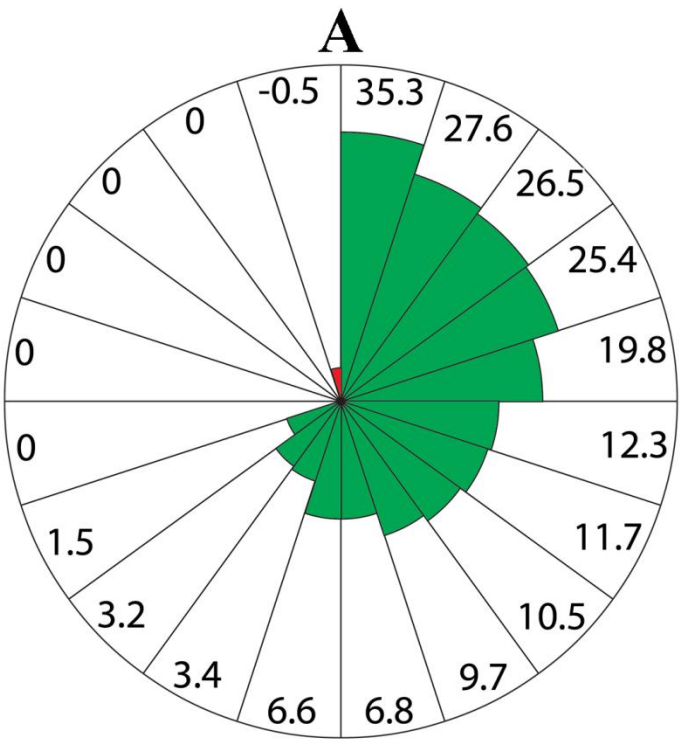
A	B	C	D
31.4%	50.0%	27.6%	29.9%
23.1%	25.9%	22.6%	29.0%
22.3%	20.8%	19.7%	28.5%
21.6%	18.2%	17.6%	25.2%
21.8%	18.6%	16.8%	14.7%
14.2%	11.3%	15.4%	14.6%
12.3%	10.8%	14.3%	14.4%
10.5%	9.7%	13.0%	13.1%
10.5%	9.4%	12.4%	9.1%
8.9%	7.0%	12.4%	5.6%
8.3%	6.0%	11.6%	5.4%
7.6%	5.5%	10.0%	5.1%
6.5%	4.5%	8.7%	3.6%
5.4%	4.0%	8.1%	1.0%
4.5%	2.6%	1.5%	0.8%
4.3%	0.8%	-0.4%	0.0%
4.1%	0.4%	-0.6%	0.0%
1.6%	-1.1%	-0.6%	0.0%
1.2%	-1.7%	-1.9%	0.0%
-20%	-2.9%	-8.6%	0.0%
Average:	10.0%	10.0%	10.0%	10.0%

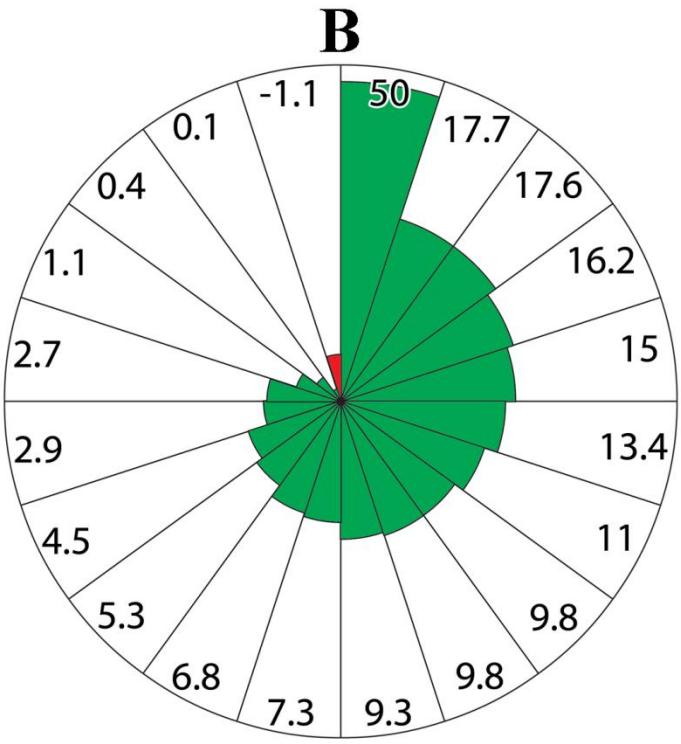
Rank the wheels of fortune according to your preference (A,B,C and D):

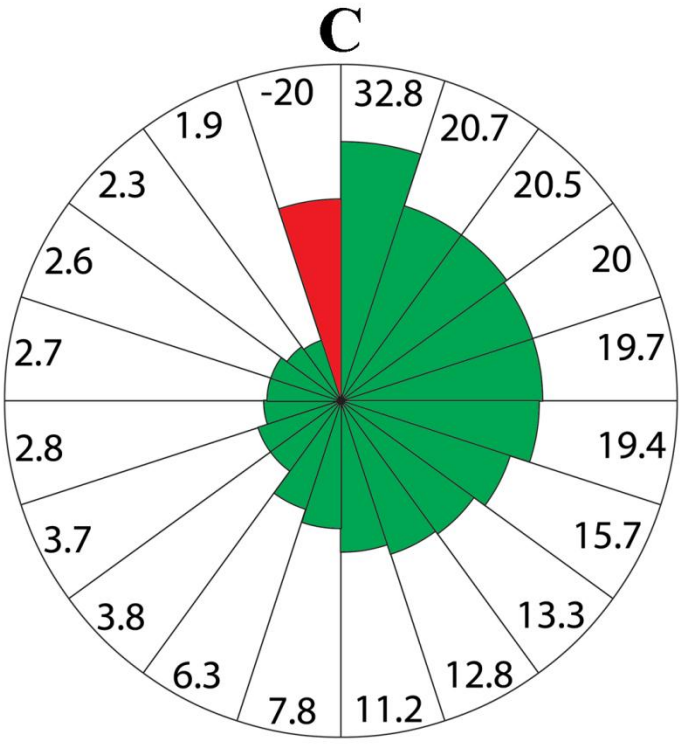
Most preferred ____ - ____ - ____ - ____ Least Preferred

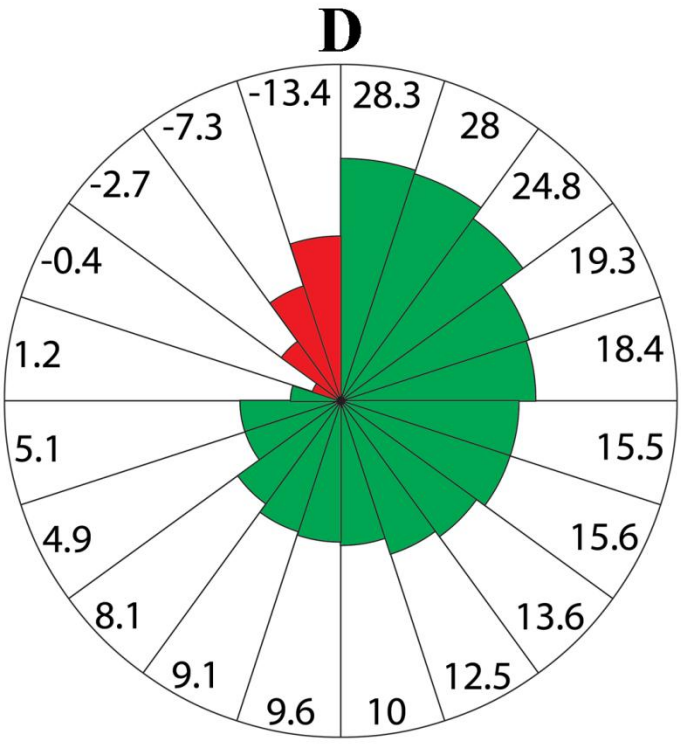
How strongly did you prefer your most preferred wheel of fortune to your least preferred wheel of fortune (circle your choice)?

Very strongly 4 - 3 - 2 - 1 - 0 Not at all









A	B	C	D
35.3%	50.0%	32.8%	28.3%
27.6%	17.7%	20.7%	28.0%
26.5%	17.6%	20.5%	24.8%
25.4%	16.2%	20.0%	19.3%
19.8%	15.0%	19.7%	18.4%
12.3%	13.4%	19.4%	15.5%
11.7%	11.0%	15.7%	15.6%
10.5%	9.8%	13.3%	13.6%
9.7%	9.8%	12.8%	12.5%
6.8%	9.3%	11.2%	10.0%
6.6%	7.3%	7.8%	9.6%
3.4%	6.8%	6.3%	9.1%
3.2%	5.3%	3.8%	8.1%
1.5%	4.5%	3.7%	4.9%
0.0%	2.9%	2.8%	5.1%
0.0%	2.7%	2.7%	1.2%
0.0%	1.1%	2.6%	-0.4%
0.0%	0.4%	2.3%	-2.7%
0.0%	0.1%	1.9%	-7.3%
-0.5%	-1.1%	-20.0%	-13.4%
Average: 10.0%	10.0%	10.0%	10.0%

Rank the wheels of fortune according to your preference (A,B,C and D):

Most preferred ____ - ____ - ____ - ____ Least Preferred

How strongly did you prefer your most preferred wheel of fortune to your least preferred wheel of fortune (circle your choice)?

Very strongly 4 - 3 - 2 - 1 - 0 Not at all