UNIVERSITY OF PIRAEUS
SCHOOL OF FINANCE AND STATISTICS
DEPARTMENT OF BANKING AND FINANCIAL MANAGEMENT
MASTER OF SCIENCE IN BANKING AND FINANCIAL MANAGEMENT

Master Thesis

DISTRIBUTIONAL UNCERTAINTY AND THE CROSS-SECTION OF EXPECTED STOCK RETURNS

by

CHAITAS NIKOLAOS CHRISTOS

MXRH 1832

Supervisor:	Professor Kourogenis Nikolaos
Evaluation Committee:	Professor Kourogenis Nikolaos
	Professor Tsiritakis Emmanouil
	Professor Englezos Nikolaos

Dedication

To my grandpa, for teaching me the value of knowledge.

Acknowledgments

I am grateful to my supervisor, prof. Nikolaos Kourogenis, for his kindness, support, and exceptional guidance, and to the rest of my dissertation committee – prof. Emmanouil Tsiritakis and Englezos Nikolaos – who accepted to be the examiners of this thesis.

I would also like to express my gratitude to all the Academic Staff of the Department of Banking & Financial Management for the knowledge and the support they offered me throughout the whole period of this master's degree.

Furthermore, I would like to thank the Administrative Staff of the Department of Banking & Financial Management for their exceptional assistance in offering me helpful advice and answering all my questions.

Last but not least, I would like to thank my family for their wise counsel and sympathetic ear. You are always there for me.

Abstract

My dissertation aims at understanding possible differences of the distribution of stock returns. With the help of the proxies for distribution uncertainty, I examine the difference of distribution uncertainty between each individual stock and the market return. The result is that those stocks with higher distribution uncertainty exhibit higher returns, therefore, the difference between portfolios with higher and lower distribution uncertainty is significantly positive. The empirical results for the impact of distribution uncertainty persist after taking into consideration individual firm characteristics.

Table of Contents

Introduction

This dissertation investigates questions within the framework of whether there is an explanation to the variation of returns across different stocks at any point in time. This is a legitimate question, which goes through the mind of every student at any university around the globe who has ever taken a course in finance instantaneously.

In one of the most creative papers in financial economics, Markowitz (1952a) distinguished two stages when it comes to select a portfolio: The first stage starts with the observation and experience and concluded with the future beliefs about asset returns, and the second stage, with the optimisation of the portfolio on the basis of these beliefs. So, a plethora of traditional asset pricing models like the CAPM neglect the first stage and are based on the optimisation of the second stage. The problem is that the investors who have picked these models had already beliefs about the form of the distribution of the asset returns. For instance, under CAPM, the investor assumed that asset returns from his portfolio following a multivariate normal distribution or they have quadratic utility, so the mean-variance is optimal but, as the empirical evidence confirms (see, e.g. Fama, 1965, and Rosenberg, 1974) the portfolio returns are not normally distributed and even (a fuzzy argument by Tsay, 2010) a specific stock return distribution does not exist. The above arguments imply that investors do not actually know the exact distribution of future stock returns. Particularly, a recent paper form Kacperczyk and Damien (2011), assumes that the form of the distribution of returns is unknown, and advance a novel method (the Bayesian model) to incorporate the uncertainty into the return distribution (distribution uncertainty) in order to attain an optimal mix between a risky and a riskless asset. While there are difficulties in understanding the form of the uncertain distribution which its problem is widely known, oddly very few is known about whether the uncertainty or the return distribution affects empirical phenomena in finance, such as the cross-sectional difference of asset returns.

Since, Treynor (1961-2), Sharpe (1964), Lintner (1965) and Black (1972), who have introduced the Capital Asset Pricing Model, there have been many studies conducted on the cross-section of stock returns. A plethora of studies have tested the empirical validity of CAPM using different approaches, but in this specific thesis I will not use CAPM or another traditional asset pricing model. Specifically, I will use proxies for distribution uncertainty of a stock return

such as the Kolmogorov-Smirnov (KS), the Cramer-von Mises (CM) and the Kuiper (K) statistics, which are non-parametrical, and the differences between empirical return distributions of an individual stock and a benchmark portfolio. As a result, I aim to investigate if there exists a significant relation between distribution uncertainty and expected stock returns.

"We perceive the world before we react to it, and we react not to what we perceive, but always to what we infer"

-Frank Knight-

Chapter 1

Uncertainty

1.1 Introduction

The basic notion for many investors is that they prefer to take risks in situations where they know the probability rather than an alternative risk situation in which the probability is completely ambiguous. This fear for the uncertainty is, and always be, in the nature of man, because always we are always going to choose a known probability of winning over an unknown probability of winning even if the known probability is low and the unknown could be a guarantee of winning[1]. Because of this, many economics analysts wanted to describe this distinction between the known and unknown as **risk** vs. **uncertainty** (Knight, 1921), **unambiguous** vs. **ambiguous probability** (Ellsberg, 1961), **precise or sharp** vs. **vague probability** (Savage, 1954), and so forth.

1.2 Uncertainty and Risk

The main problem of uncertainty and risk in economics is, of course, not new. We live in a world of change and a world of uncertainty. As Knight (1921) has said; "We live only by knowing something about future while the problems of life, or of conduct at least, arise from the fact that we know so little" [p.199][2]. Indeed, one of the major issues of modern decision

[1] Ellsberg Daniel, 1961, Risk ambiguity and the Savage axioms, Quarterly Journal of Economics 75, 643-669

[2] Knight Frank H., 1921, Risk, Uncertainty and Profit, 1st end, Houston Mifflin, Boston, MA

theory is the analysis of decisions under ignorance or ambiguity, where the probabilities of potential outcomes are neither specified in advance nor readily assessed on the basis of the available evidence. This problem was examined in detail by many academic researchers from economic or even from statistical perspective; one of them, of course, was Knight (1921) who distinguished the risk and uncertainty as measurable and unmeasurable uncertainty, respectively. In other words, the proper definition that we can give in order to describe the word "risk" is that of referring to any kind of uncertainty viewed from the perspective of the unfavourable contingency and the term "uncertainty" similarly with reference to the favourable outcome. Nevertheless, if the above logic is correct, we can use the expressions' "objective" and "subjective" probability to describe the risk and uncertainty respectively, as these terms are already generally used for this purpose.

The foundation differences between those two terms are that the objective probability is more accurate for the description of a given outcome than subjective probability. That is because the subjective probability is referred to the judgement and experiences of the individual rather than objective probability which allows the observer to gain insight from historical data and then evaluate the likelihood of a given outcome. A striking example about uncertainty is in judging or shaping these views in the future course of events, which in fact guide most of our behaviour. Furthermore, if we are able to know the possible outcomes from a probability then we can discard any real uncertainty by the expedient of grouping instances. It seems possible but does not necessarily mean this will be done; and we should observe the outset of probability because when an individual instance only is at issue, the measurable risk and an unmeasurable uncertainty have no differences. The individual who has already observed via on his evaluation of the value of his opinion into the probability form of p (probability of success) and q (probability of failure) and the "feeling" about any other probability situation; besides this unique instance, it can be concluded that it is possible for the real probability to be calculated, if of course we not only know exactly how much success there will be in a number of games but also, the odds against us then it does not matter whether we place all our bets in one kind of game or in as many as possible different games as there are bets. Nevertheless, it is important an individual who enters in this world to remember that it will be against him, if he considered any single case as isolated and the only thing that comes to my mind to reverse this important logic is when his fortune is at stake.

Additionally, in the case that we do not have a quantitative probability in the process of grouping, still we have the change to tend to some results of cancelling out some fluctuations and approach constancy at some point. For Knight (1921) there appears to be two kinds of

elements as far as he is concerned about making judgements; the first one is **determinate factors** (or uncertainty of an opinion) and the second one **truly accidental factor** (or true probability). The difference between these two is that we have no means to scatter them and estimate them, neither by calculation a priori nor by empirical sorting. However, if we applied some magnitude within narrow limits then, the sorting method can be conducted.

Besides the above two kinds of elements, the main problem for human attitude concerning uncertainty is that an individual is surrounded by difficulties as the uncertainty itself. Of course, the reaction itself is varied from one individual to another but the common, as common can be, reaction of a human being is subject to well-known deviations from the conduct which sound logic would dictate. So, as is well detailed by Adam Smith, the man going to risk a small amount in the chance of winning a larger one when the adverse probability, which is probably known or estimated, against winning is much more than the ratio of two amounts, while he will deny bearing a small chance of losing a larger amount for a virtual certainty of winning a smaller one, even if the value of chance is in their favour. However, on this prejudice should be added the word "luck" for the part of the individual because his inveterate belief of his own luck is specifically powerful when the basis of uncertainty is the quality of his own judgment. But what is this "feeling" that makes a person think that he has taken the right choice? Is it a mere "hunch" or "I am feeling it in my bones"? The man in the street knows the real value of his opinions better than the knowledge of the "logic" on which they repose. Thus, the choice that an individual takes is not based only on logic but also on superstition. Any incident that strikes attention is probable to be elevated into a law of nature, creating a belief in an unerring "sign" which, without real or imaginary basis in the mind of the person itself, might be accepted as a valid base for action and treated as a beyond doubt verity.

Undoubtedly, an imperative needed for rationality judgment for the human's thoughts and needs such as whims and impulses are and always will be necessary and because of this the limitations on the below hypothesis to rational grounds of action seem justified. An individual wants to bear a sacrifice for the sake of a future benefit, of course, should this benefit be considered not only certain, but also greater than his sacrifice and his anticipation. Obviously, the subjective uncertainty, which is determined in the above case, therefore, what the individual believes the chances to be, whether if his degree of confidence is based upon an objective probability in the situation itself or in an estimate of his own powers of prediction. Additionally, both types, objective and subjective might be, involved at the same time; a man's forecasting or opinion may be as estimation of an objective probability and the estimation itself be acknowledged as a certain degree of validity to be the product of two probability ratios and felt

the degree of uncertainty. However, it should be emphasised that all the decisions which an individual makes are derived from his conduct in real life and undoubtedly most of these opinions, which first evaluate them through scrutiny, easily resolve themselves into an opinion of probability.

One of the major problems of the uncertainty in economics is the future character of the economic process itself. Generally, goods are produced to satisfy wants; through from goods two elements of uncertainty are introduced by Knight (1921) as two different kinds of foresight which should be exercised. Firstly, from the beginning the operations of productivity must be evaluated. Unfortunately, it is impossible for someone to know accurately what the results will be when entering the productive activity, in physical terms such as the **quantities** and **qualities** of goods which will be the outcome from the cost of given resources. Secondly, goods which are produced to satisfy needs are going to be fulfilling this satisfaction in the future so the producer must estimate (1) **the future demand** which he is endeavouring to satisfy and (2) **the future results of his operations** in his attempt to satisfy this demand.

Obviously, the rational conduct endeavour decreases to the minimum the uncertainties which are affected in adapting means to an end. This does not mean that uncertainty as such is loathsome for the individual. We should really not be proponents of the notion that everything in this world is defined and therefore we should not want our activity to be perfectly rational. Yet, because it is in our nature to try to foresee the next day and in some cases, even the next minute, we are attempting with "intelligent" moves to secure perfectly adaptation as much as possible. As noted before, this element of paradox in our conduct must not be ignored. We find ourselves compelled to struggle after things which in other conditions we acknowledge that we do not want, at least not this completeness and perfection. Maybe it is this apparent impossibility as a human imperfection of reaching the end which makes it interesting after. In other words, we are trying to decrease uncertainty although we should not want it eliminated from our lives.

In addition to the above two kinds of foresight, Knight (1921) introduces through two basic sets of conditions another two kinds respectively; (3) **control of the future** and (4) **increased power of prediction**. As far as the first fundamental condition is concerned, uncertainties are fewer in groups of cases than in a single instance. Yet, if a priori probability of uncertainty tends to disappear completely it is because of the increasing group of inclusiveness; with statistical probabilities, the same tendency is obvious to a lesser degree, being limited by defectiveness of classification. As far as the second condition, it concerns the reduction of uncertainty is different among individuals. Finally, Knight (1921) presents another two and last

kinds of foresight, the (5) **"diffusion"** of the consequences of untoward contingencies and through observation of consolidation[3] and specialisation which are intimately connected, there is the (6) **possibility of directing industrial activity** approximately along lines in which the minimum amount of uncertainty is included.

1.3 Risk, Ambiguity, and the Savage Axioms

In the above chapter we have discussed the Knightian theory which distinguishes the "measurable uncertainty" and "risk". Moreover, the risk can be represented by numerical probabilities, but the "unmeasurable uncertainty" cannot. In this chapter we are going to discuss another decision theory which was presented in 1954[4] by Savage and a set of axioms constraining preferences over a set of options that guarantee the existence of a pair or probability and utility functions relative to which the preferences can be represented as maximising expected utility.

Besides, the possible outcomes which do not come with probabilities rather than state of world the options or prospects in Savage theory are familiar to lotteries. Truly, the primitives in Savage's theory are consequences and states. As a result, the former are the good or bad states of affairs that affect and matter to a person, while the latter are the features of the world that the human has no control over, and which are the locus of her uncertainty about the world. The distinction between consequences and states serves to separate "desire" and "belief" respectively.

Consider this following example: In tabular form the act-state-consequence-outcome distinction from Savage's decision theory, can be represented with rows serving as acts that yield a given outcome for each state/event column.

Table 1.3. An example which illustrates acts, states, and outcomes

	No rain	Rain
Stroll without umbrella	Very comfortable stroll	Miserable wet stroll
Stroll with umbrella	Comfortable stroll	Comfortable stroll

[3] Irving Fisher's term (The Nature of Capital and Income, p.288 which introduce the five ways which risk may be reduced)

[4] Savage, L. J., 1954, *The Foundations of Statistics*, Wiley, New York

Constant act	Miserable wet stroll	Miserable wet stroll

Table 1.3 depicts the two acts mentioned above plus a third one that the decision maker might care about: the acts i) "go for stroll without umbrella", ii) "go for stroll with umbrella", and iii) the bizarre constant act. Of course, the set of acts required for Savage's theorem involve even more acts that account for all the possible combinations of states and outcomes.
Without further ado, let state the Savage's axioms. These are intended as constraints on an agent's preference relation, $\preceq$, over a set of acts, F, as described above. The first of Savage's axioms is the basic ordering axiom.

P1. (Ordering)
The relation $\preceq$ is complete and transitive.
The next axiom is reminiscent of vNM's Independence axiom. We say that alternative f "agrees with" g in event E if, for any state in event E, f and g yield the same outcome.

P2. (Sure Thing Principle)
If f, g, and f′, g′ are such that:

- f agrees with g and f′ agrees with g′ in event ¬E,
- f agrees with f′ and g agrees with g′ in event E,
- and $f \preceq g$,

then $f' \preceq g'$.

The idea behind the Sure Thing Principle (STP) is essentially the same as that behind Independence: since we should be able to evaluate each outcome independently of other possible outcomes, we can safely ignore states of the world where two acts that we are comparing results in the same outcome. Putting the principle in tabular form may make this more apparent. The setup involves four acts with the following form:

	E	¬E

f	X	Z
g	Y	Z
f’	X	W
g’	Y	W

The intuition behind the Sure Thing Principle is that if g is weakly preferred to f, then that must be because the consequence Y is considered at least as desirable as X, which, by the same reasoning, implies that g’ is weekly preferred to f’.

P3. (State Neutrality)

If f(si) = X and g(si) = Y whenever si $\in$ E and E is not null, then f $\preceq$ g given E just in case X $\preceq$ Y.

The next axiom is also necessary for it to be possible to determine a comparative belief relation from an agent’s preferences. Above it was suggested that by asking you to stake a prize on whether a coin comes up heads or tails, it can be determined which of these events you find more likely. But that suggestion is only plausible if the size of the prize does not affect your judgement of the relative likelihood of these two events. That assumption is captured by the next axioms and is illustrated in the tabular form below:

P4.

Consider the following acts:

	E	¬E
f	X	X’
g	Y	Y’

	F	¬F
f’	X	X’
g’	Y	Y’

Now suppose:

$$X' \preceq X,$$

$$Y' \preceq Y,$$

$$f' \preceq f$$

Then

$$g' \preceq g$$

Less formally (and stated in terms of strict preference), the idea is that if you prefer to stake the prize X on f rather than f', you must consider E more probable than F. Therefore, you should prefer to stake the prize Y on g rather than g' since the prize itself does not affect the probability of the events.

The next axiom is arguably not a rationality requirement, but one of Savage's "structural axioms" (Suppes 2002). An agent needs to have some variation in preference for it to be possible to read off her comparative beliefs from her preferences; and, more generally, for it to be possible to represent her as maximising expected utility. To this end, the next axiom simply requires that there be some alternatives between which the agent is not indifferent:

P5.

There are some $f,g \in F f,g \in F$ such that $f \prec g f \prec g$.

When these five axioms are satisfied, the agent's preferences give rise to a comparative belief relation, $.\precsim..\precsim.$, which has the property of being a *qualitative probability* relation, which is necessary for it to be possible to represent $.\precsim..\precsim.$ by a probability function. In other words, $.\precsim..\precsim.$ satisfies the following three conditions, for any events EE, FF and GG:

1. $.\precsim..\precsim.$ is transitive and complete,
2. if $E \cap G = \emptyset = F \cap G E \cap G = \emptyset = F \cap G$, then $E.\precsim.F \Leftrightarrow E \cup G.\precsim.F \cup G E.\precsim.F \Leftrightarrow E \cup G.\precsim.F \cup G$,
3. $\emptyset.\precsim.E, \emptyset.\precsim.E, \quad \emptyset.\prec.S \emptyset.\prec.S$

Being a qualitative probability relation is, however, not sufficient to ensure the possibility of probabilistic representation. To ensure this possibility, Savage added the following structural axiom:

P6. (Non-atomicity)

Suppose $f \prec g f \prec g$. Then for *any* $X \in O X \in O$, there is a finite partition, $\{E1,E2,\ldots Em\}\{E1,E2,\ldots Em\}$, of SSsuch that:

- $f'(si)=Xf'(si)=X$ for any $si \in Ejsi \in Ej$, but $f'(si)=f(si)f'(si)=f(si)$ for any $si \notin Ejsi \notin Ej$,
- $g'(si)=Xg'(si)=X$ for any $si \in Ejsi \in Ej$, but $g'(si)=g(si)g'(si)=g(si)$ for any $si \notin Ejsi \notin Ej$,
- $f' \prec g f' \prec g$ and $f \prec g' f \prec g'$.

Similar to the Continuity axiom of vNM, Non-Atomicity implies that no matter how bad an outcome X is, if gg is already preferred to ff, then if we add X as one of the possible outcomes of ff—thereby constructing a new alternative f′—g will still be preferred to the modified alternative as long as the probability of X is sufficiently small. In effect, Non-Atomicity implies that S contains events of arbitrarily small probability. It is not too difficult to imagine how that could be satisfied. For instance, any event F can be partitioned into two equiprobable sub-events according to whether some coin would come up heads or tails if it were tossed. Each sub-event could be similarly partitioned according to the outcome of the second toss of the same coin, and so on.

Savage showed that whenever these six axioms are satisfied, the comparative belief relation can be represented by a *unique* probability function. Having done so, he could rely on the vNM representation theorem to show that an agent who satisfies all six axioms[7]can be represented as maximising expected utility, relative to a unique probability function that plausibly represents the agent's beliefs over the states and a cardinal utility function that plausibly represents the agent's desires for ultimate outcomes (recall the statement of Savage's theorem above).

1.4 Distribution Uncertainty

While a plethora of authors how tried to describe the uncertainty in their unique theories and scepticism, a recent paper of Kacperczyk and Damien (2011) has suggested a novel method to incorporate "distribution uncertainty" therefore, an uncertainty about the type of return distribution to obtain an optimal portfolio. Specifically, this paper approaches Feller's (1971) notion about **scale mixture representation** and therefore, the formula of the conditional distribution with which you can capture wide ranges of kurtosis in the data. Before exploring the mathematics further, a brief intuitive understanding of the idea is as follows. Considering the equation

$$r_{t+1} = \mu_{(t)} + e_{t+1}\, t = 1,\ldots,T \qquad (1)$$

where r is market return, $\mu_{(t)}$ means that μ either be constant or could depend on a set of predictor variables up to time t. The important aspect of this equation is that it does not impose any distributional form on the error term. Considering, *e* which may come from any distribution

with mean 0 and variance σ^2. The subsequent sections show how this component of return process can be used to generate different forms of distribution uncertainty.

Specifically, the key aspect to modeling distribution uncertainty via the above equation (1) is that the conditional distribution of the asset return is Normal, leading to a parametric form for the predictive distribution. Under distribution uncertainty, the above assumption is relaxed by stating that the conditional distribution of the returns itself is uncertain. In other words, this means that you must place a prior distribution on a wide class of distribution functions, which any member of this could possibly be the conditional distribution of returns.

To model distribution uncertainty, we use family models from the Kacperczyk et al. (2011) paper and adapt the Semiparametric Scale Mixture of Betas (SSMB) whose approach is based on the belief of "scale mixture representation" which is an idea dating back to Feller (1971) and the references therein. Consider the equation (1). Typically, one writes the sampling distribution of r, say as a Normal $(0, \sigma^2)$, or some other distribution f(r). Feller generalized this by first introducing an auxiliary variable U. The distribution of U, f(U), could be any probability density. Supposing that we have a random draw of u from f(U). With this *u* Feller states the conditional distribution of r as a uniform distribution in addition to the Uniform (a,b), where b is now a product of u and the standard deviation of the sampling distribution f(r). This idea from Feller (1971) is remarkable if you think that now one could induce much more flexibility in modelling the higher-order moments of the sampling distribution f(r); particularly, skewness and high levels of kurtosis are readily handled by appropriately choosing f(U). In the paragraph below we are going to obtain the Normal distribution as a special case for the form of f(r).

For the moment, what if investors believe that the conditional distribution of the excess returns is unimodal. Since the unimodal density for the returns is the desirable one, we use uniform and beta distributions in the scale mixture model. Furthermore, we simplify the intuition underlying our model by first addressing symmetric distributions for the sampling distribution of the data. This simplification is practically relevant as our empirical application considers monthly excess market returns. Campbell et al. (1997) mentions that the observed deviations from normality observation in the monthly returns are more pronounced as a result of excess kurtosis than skewness. In principle, note that with symmetric distributions still appeals to the idea of distribution uncertainty in that one does not have to assume a particular form for the underlying sampling distribution of the data; with r as observed data and U as a latent mixing random variable, Feller's (1971) formulation of the conditional distribution of r is given by:

$$F(r|U = u) \sim \text{Uniform}(\mu-\sigma\sqrt{u}, \mu+\sigma\sqrt{u}), \qquad (2)$$

$$u \sim F$$

for some distribution function F with support on (0, ∞). As F ranges over all such distribution functions then the density of r ranges over all unimodal and symmetric density functions. As a result, with flexible F we can capture wide ranges of kurtosis in the data. In order to ensure maximum flexibility F must be nonparametrically. In Kacperczyk et al. (2011) paper demonstrate that √u rather than u in the formulation above is helpful since one can express higher moments for r in terms of lower moments for U. So, they rewrite the model as

$$f(r|U) = \sigma\sqrt{U}\ (1\text{-}2\ \text{beta}\ (1,1)) \qquad (3)$$

which will suggest the form of generalizations to asymmetric or skewed densities. An interesting fact is that if F is distributed Gamma (3/2, ½), then the distribution of r is normal (0, σ^2). Similarly, by changing the specifications of the parameters in F, one can obtain other commonly used distributions, such as t, generalized exponentially and so on.

1.5 Subjective Expected Utility Theory

First and foremost, utility theory can be described as the beliefs of individuals' preferences. For the economical world this theory can explain the behaviour of a particular person based on the assumption that people can consciously order their choices depending on their preferences. Each individual is going to show different preferences, which will appear to be hard-wired within them. As a result, we can state that individuals' preferences are intrinsic. Therefore, the above theory can be further divided into **expected utility theory (EU)** of von Neumann and Morgenstern (1947) and **subjective expected utility theory (SEU)** of Savage (1954)**.**

On the one hand, the **expected utility theory (EU)** is an economic term which summarises the utility as an entity or as an aggregate economy and is expected to be reached under any number of circumstances. Furthermore, in order to calculate the expected utility, we should take the weighted average of all possible outcomes under specific circumstances, with the

weights being assigned by the likelihood or probability which any particular event will occur. In other words, the probabilities of outcomes are known. On the other hand, in **subjective expected utility theory (SEU)**, probabilities are not necessarily objectively known, so the SEU applies more widely than EU and, since this thesis shall discuss this uncertainty, I am going to debate further the latter rather than the former.

Decisions made in the face of uncertainty spread through the life of every individual and organisation. Even animals, generally speaking, may be faced with uncertainty when continuously making decisions (most commonly for their survival) and the psychological mechanisms behind these decisions, by which animals decide, may have much in common with that of men. Yet, formal reasoning presumably makes no difference in the decisions of animals, little in those of children and even less than be wished in those of man.

Reasoning is commonly associated with logic, which, obviously, as many have pointed out, in the face of uncertainty these implications of what commonly is called logic are meager. Therefore, it has constantly been asked whether logic cannot be extended by principles which are acceptable as those of logic itself to bear more fully on uncertainty. First and foremost, as far as logic is concerned, with implications among propositions, a plethora of individuals have been thinking that to extend logic criteria must be set up for the extent to which one proposition tends to imply or provide evidence for another. However, it is obvious that the main problem was not if we need the criteria but what will be these criteria so that criteria being ultimately wanted for deciding among possible courses of action. Therefore, the generalisation of the notion of the related implications seems at best a roundabout method of attack. It must be acknowledged that the logic itself should guide our decision criteria because when it is implied by a proposition known to be true, it in turn is not only true, but also relevant (sometimes) for decision making. Furthermore, if some idea of partial implication is in a way proven and articulated with decision than is implication itself, that would be great; yet another question emerged of how such an idea be sought except by explicitly studying the decision meaning process? Secondly, it is appealing to assume that if two individuals who are in the same situation and act reasonably, have the same taste and are provided with equal information, then they will act in the same way. Such a belief in which view of probability is apparent (besides the personal opposite point of view), is certainly worth looking for.

But what are the consequences of our acts and decisions? To say that a decision must be taken is to say that one or two of the actions must be chosen or decided. When deciding to act, one must take into consideration the potential situations of the world, as well as the consequences involved in each action for every possible situation in the world. One

consequence is that everything can happen to an individual. Yet, so as to be more to enlightened to the above I believe that an example from Savage (1954)[5] will be suitable.

Consider this example. Your spouse has just broken five good eggs into a bowl, when you come in and volunteer to finish the omelette. A sixth egg is beside the bowl, which for some reason must either be used for the omelette or be wasted altogether. Hence, you must decide what you are going to do with this unbroken egg. The sixth egg might or might not be rotten. As a result, you must decide among three acts only; to break it into the bowl which contains the other five, to break it into a saucer for inspection, or to throw it away without inspection. Depending on the state of the egg, the above three acts have some consequences of concern to you and these three acts are illustrated on the Table 1.5.

Table 1.5. An example which illustrates acts, states, and consequences

Act	State	
	Good	Rotten
Break in bowl	six-egg omelette	no omelette and five good eggs destroyed
Break into saucer	six-egg omelette and a saucer to wash	five-egg omelette and a saucer to wash
Throw away	five-egg omelette and one good egg destroyed	five-egg omelette

With this example it is easy for someone to perceive the variety of things or experiences which are addressed as consequences can be. They could generally be about money, life, health, the well-being of others, or anything else that could possibly concern an individual. Consequences could be appropriately called situations of the person, as opposed to situations of the world. They could also be mentioned, with some extension of the economic concept of income as possible by the individual. Therefore, in any problem, the set of consequences will be donated by F, and the consequences of individual will be donated by f, g, h, and so forth. In other words, in the omelette example, F consists of the six consequences tabulated in Table 1: six-egg omelette and a saucer to wash; five-egg omelette and a saucer to wash, and so on.

[5] Savage, L. J., 1954, The foundations of Statistics, Wiley, New York [p.14]

If two different acts have the same consequences in all the countries of the world, from today's point of view it would not make sense to consider them at all. Therefore, an action can be identified with its possible consequences or, more formally, an act is a function that gives consistency to every situation in the world. Now, let's use the defined symbol F in order to express the dilemma of the spouse; (a) f(good) = six-egg omelette, (b) f(rotten) = no omelette, and five good eggs destroyed.

It may be argued that the formal description of the decision thus made seems inadequate because a person may not be aware of the consequences of the acts open to him in every corner in the world. He might be ignorant, for instance, not being sure if one rotten egg will spoil a six-egg omelette. But in this case, nothing could be simpler then a culinary question and only two possible answers present themselves of whether one bad egg will spoil a six-egg omelette. Obviously, this solution works in a generalised situation, although a thorough analysis might not be without its merits.

Subjective expected utility was further detailed by Savage (1954) following some previous work by Ramsey (1926) and von Neumann has been distinguished on two parts: the personal probability distribution and a personal utility function.

1.5.1. Personal Probability Distribution

Several individuals are convinced that statements which infer about personal probability precisely mean nothing, or at any case that they mean nothing precisely. On the contrary, others hold the belief that it has no meaning for someone to analyse something that is so self-evident. An intermediate position is taken in this chapter, where a particular explanation of probability to a person is given in terms of the decision theory in the face of uncertainty. The consistency of the idea of probability, which are defined here, with should be judged by the contribution it makes to the decision theory, not by the accuracy which it analyses ordinary usage.

The first approach, possibly to find out which of the two events a person considers more probable is simply to ask him. It might even be argued, though I think that since the question concerns what is inside the person's head that, there can be no other method. Several statistical theorists believed that if we attempt to define the relative probability of a pair of events or the concept "more probable to me than" is an intuitive one as a result leads you to no ambiguity and yet admits no further analysis.

Furthermore, what if the concept was so completely intuitive, which might be characterized as a direct interrogation as a subject worthy of some behaviour of a person in the face of

uncertainty. If at the one hand the state of mind in question is not capable of manifesting itself in some sort of extraverbal behaviour, it is not forming an essential for our main interest. If, on the other hand, it does manifest itself through more material behaviour that should, at least in principle, imply the possibility of testing whether a person holds one event to be more probable than another, by some behaviour expressing and given meaning to his judgement. Several schemes of behavioural, as opposed to direct, interrogation have been suggested. The one below was suggested from Savage (1954) who takes the idea from de Finetti's paper which via the paper does not give emphasis to behavioural interrogation.

Consider this following, our ideal person has just taken two eggs from his refrigerator and holds them unbroken in his hand. Whether he thinks it more probable that the brown one is good than that the white one is. Moreover, Savage wants to address him as: "We see that you are about to open those eggs. If you will be so cooperative as to guess which of the two is good, we will pay you a dollar, should your guess prove correct. If it is incorrect, you and we are quitting, except that we will in any event exchange your two eggs for two of guaranteed goodness"[6]. This it is not fundamental to the subsequent argument but if under these circumstances the person chooses the brown one, it seems that he is corresponding well with the ordinary usage.

Nevertheless, there is a mode of interrogation found in the middle between what the Savage (1954)[7] called as behavioural and direct; one can just ask the person not how he feels about his choice but what he would do in such a situation. The theory of decision is regarded as an empirical one and the intermediation is a compromise between economy and rigor. Moreover, in theory's more important normative interpretation as a set of criteria of consistency for us to apply to our own decisions.

1.5.2 Personal Utility Function

The arithmetization of comparison among acts can -with the introduction of one mild new postulate- be extended to virtually all pairs of acts. This far-reaching comparison among acts is achieved by attaching a number U(f) to each consequence f in such a way that $f \leq g$ if and

[6] Savage, L. J., 1954, *The Foundations of Statistics*, Wiley, New York. (p.28)

[7] Savage, L. J., 1954, *The Foundations of Statistics*, Wiley, New York. (p.28)

only if the expected value of U(f) is numerically less than or equal to that of U(g), provided only that the real value functions U(f) and U(g) are essentially bounded. The act of providing can fail to be met only if there exist acts that are distinctly preferable to any fixed reward or distinctly worse than any fixed punishment.

This function U that arithntizes the relation of preference among acts will be called utility. The multiplicity of utilities is not complicated; every utility being simply related to every other. The word utility it adopted from von Neumann and Morgenstern (1944) for the economic theory and they revived the concept to which it refers in a most stimulating way: An extension of the theory of consumer preferences that incorporates a theory of behaviour toward risk variance.

The expected utility hypothesis has shown that when a consumer is faced with a choice of items or outcomes subject to various levels of chance, the optimal decision will be the one that maximizes the expected value of the utility derived from the choice made. Expected value is the summarization of the products of the various utilities and their associated probabilities.

The von Neumann Morgenstern utility function can be used in order to explain risk-averse, risk-neutral and risk-loving behaviour. For instance, a company in one year undertakes a project that has probabilities for three possible payoffs of 30$ 40$ and 50$; those probabilities are 40% 70% and 20% respectively. As a result, expected payoff from the project would be $30(0.4) + $40(0.7) + $50(0.2) = $50. The next year the firm might again undertake the same project but now the respective probabilities for payoffs will be different, but the notion is that the expected payoff is still the same. In other words, as mathematics is concerned nothing has changed.

Furthermore, it is true that the probabilities of the lowest and highest payoffs rose at the expense of the middle one, which means there is more variance (or risk) associated with the possible payoffs. The question to pose to the firm is whether it will adjust its utility derived from the project despite the project's having the same expected value from one year to the next. If the firm values both iterations of the project equally, it is said to be risk neutral. The implication is that it equally values a guaranteed payoff of $50 with any set of probabilistic payoffs whose expected value is also $50.

If the firm prefers the first year's project environment to the second, it places higher value on less variability in payoffs. In that regard, by preferring more certainty, the firm is said to be risk averse. Finally, if the firm actually prefers the increase in variability, it is said to be risk loving.

The von Neumann Morgenstern utility function adds the dimension of risk assessment to the valuation of goods services and outcomes. Such utility maximization is necessarily more subjective than when choices are subject to certainty.

1.5.3 The Challenges of SEU

The subjective expected utility theory first was developed by Savage (1954) who has been inspired by (Ramsey 1931) and de Finetti (1937), then derived by Anscombe and Aumann (1963) in an approach that combined expected utility and subjective expected utility theory.

In subjective expected utility theory, a decision maker must choose between "acts": which are denoted as uppercase letters. For instance, the consequences of an act X depend by which state's occurs, from the set S of possible states. For simplicity's sake we are going to assume that the sets of acts and states are finite. Including subjective probabilities of the states which will be denoted as p(s) then an act X will be described by a vector $(x(s_1), p(s_1);\ldots;x(s_n), p(s_n))$ (where $s_1,s_2,\ldots,s_n$).

The mathematical goal of subjective expected utility theory is to represent preferences over acts by numerical utility index u and a probability measure on the states p, such as the act X is preferred to act Y if and only if the subjective expected utility of X is larger than subjective expected utility of Y. The subjective expected utility of X is defined as

$$SEU(X) = \sum_{s \in S} p(s)\, u\,(x(s)) \qquad (4)$$

As harmless as the subjective expected utility form (4) looks, there is a long rich tradition of questioning whether it describes behavior adequately. On one hand Keynes (1921) drew the distinction between the **implications** of evidence and the **weight** of evidence, or the confidence in assessed likelihood. In Keynes's paper (1921), he express his worries whether a single probability number could express both dimensions of evidence. On the other hand, Knight (1921) distinguished "risk" or known probability and "uncertainty" and suggested that in economic returns where earned for bearing uncertainty, but not for bearing risk.

Nevertheless, the most recent attack on subjective expected utility as descriptive theory was made most directly from Ellsberg (1961) own is known as "Ellsberg paradox".

In Ellsberg's hypothesis, a decision maker must choose from an urn which contains 30 red balls and 60 balls in some combination of black and yellow. This problem is called the three-color problem. There are two pairs of acts X and Y and X' and Y'. Acts have consequences W -for win- or 0.

A plethora of people choose X > Y and Y' > X'. The number of black balls which yield a win if act Y is chosen is unknown; people prefer the less ambiguous act X. The same principle, applied to the second choice, favors Y' because exactly 60 balls yield W and vice versa for losses W < 0.

In the three-color problem, most prefer acts with a known probability of winning. As a result, they become confident when it comes to taking subjective probability into account for choices. Such a pattern is inconsistent with the sure-thing principle of subjective expected utility. Both pairs only differ in consequences when the yellow state occurs. This consequence is the same for X and Y -you win 0- and for X' and Y' -you win W-. The second axiom -the sure thing principle- assumes a state with a consequence like both acts (according to subjective expected utility X > Y if and only if X' > Y'). The common pattern X > Y and Y' > X' violates the sure-thing principle because ambiguity affects choices and the ambiguity inherent in one state, red for instance might disappear when the state is combined with an equally ambiguous state such as the yellow one.

In Ellsberg's paper (1961) another problem arises known as the two-color problem. In this problem the decision makers can use two urns. The first urn contains 50 red and black balls and the second urn 100 balls in an unknown combination of red and black. Several people prefer to bet on red from urn 1 rather than to bet on red of urn 2 and vice versa with the black but are indifferent between the two colors when betting on only one of the two urns. This pattern violates subjective expected utility theory.

"We did not set out to be educators or even scientists, and we do not purport that what we do is real science, but we are demonstrating a methodology by which one can engage and satisfy your curiosity"

-Adam Savage-

Chapter 2

Methodology and Data Selection

2.1 Introduction

For the matched-pairs sign and signed-rank tests, we will consider the sample as being two dependent samples or alternatively as a single sample of pairs from a bivariate population. When the conclusions to be drawn relate only to the population of differences in paired observations, usually the first step is to obtain the differences of paired observations; this leaves only a single set of observations. Therefore, this type of data can legitimately be classified as a one-sample problem. In this chapter we are going to debate further the consisting of two mutually independent random samples, so that the elements shall not only be in each sample independent, but also every element along in the first sample will be independent of every element in the second sample.

Our sample space, specifically, consists of two populations which we will call X and Y respectively, with cumulative distribution functions denoted as F_X and F_Y. The random sample of size m will be extracted from X population and another random size of n extracted independently from the Y population,

$$X_1, X_2, \ldots, X_m \text{ and } Y_1, Y_2, \ldots, Y_n$$

The hypothesis of interest in two-sample problem usually drawn from identical populations so,

$$H_0: F_Y(x) = F_X(x) \qquad \text{for all } x$$

We are willing to make assumptions according to the forms of the underlying populations and assume that the differences between the two populations occur only with respect to some parameters, such as the means or the variances. For instance, if we assume that both populations are normally distributed, it is well known that the two-sample Student t test for equality of means and the F test for equality of variances are respectively the best tests. The performances of these two tests are widely known. This does not mean that other classical tests are not good; just that they may be sensitive to violations of the fundamental model assumptions inherent in the derivation and construction of these tests.

2.2 Data and Construction of Variables

The sample data include returns from the EIKON Reuters DataStream of all stocks listed in NYSE, S&P 500 and NASDAQ. EIKON is used to obtain prices, daily return, market returns, shares outstanding, trading volume, and so on. We also obtain balance sheet information including assets, liabilities, and total equity from EIKON. We use stock prices and shares outstanding to calculate market capitalization and use daily returns to calculate distribution uncertainty for each firm in each month as well as beta, idiosyncratic volatility, skewness, and kurtosis. The market portfolio return is the value weighted index return in the EIKON. The sample period spans from January 2001 to March 2021. To be included in the final sample for a given month, at least 100 daily returns must exist in the previous 12 months.

We measure how different the empirical return distribution of a stock is from that of the benchmark portfolio. Using daily returns of each company and the market portfolio in the previous year, for each month we estimate three statistics that non-parametrically measure the distribution uncertainty: the Kolmogorov-Smirnov (KS), the Cramer-von Mises (CM), and the Kuiper (K) statistics. Before we estimate each statistic, we demean returns of each stock and the market portfolio by subtracting average returns estimated from the data of the previous year in order to control the effect of expected returns on our results. Since we measure distribution uncertainty by the difference of the return distribution of a stock from that of the market

portfolio, if we do not demean returns of each stock and the market portfolio, our proxy for distribution uncertainty merely catches the difference of expected returns of a stock and the market portfolio, not reflecting the degree of difficulty in understanding underlying distributions. Therefore, by demeaning returns, our KS, CM, and K statistics can compute the degree of difference in shapes of a stock return and the market portfolio return distributions other than the location of distributions. Since we control the size of mean for each stock return to construct KS, CM, and K, if we observe a larger return for a portfolio sorted by KS, CM, or K, it is from the difference of distribution, not from the difference of expected returns of the portfolio or risk of the portfolio.

Three statistics of KS, CM, and K measure the difference among several empirical distributions or between a given distribution and empirical distributions. In this section, we briefly introduce the definition of these three statistics adjusted for our case; a comparison between two empirical distributions.

The Kolmogorov-Smirnov (KS) statistic is used in the KS test to investigate the difference of distributions of two samples. Suppose that a first sample x1,..., xn has distribution with its cumulative distribution function F1(x) and the second sample y1,..., yn has distribution with cumulative distribution function F2(x). Then, the KS test investigate whether F1 = F2. If F1n(x) and F2n(x) are corresponding empirical cumulative distribution functions, then the KS statistic is defined as follows.

$$KS = \max_{j} F1(xj) - F2(xj) \text{ where } j = 1, 2, ..., n\,j$$

In short, the KS statistic for two samples is the maximum distance between two empirical cumulative distribution functions. The Cramer-von Mises statistic also measures how different two empirical distributions are. It is defined as follows:

$$CM = 1/n^2 \sum_{i} (n_i \sum_{j=1} t_j (F_i(x_j) - F(x_j))^2)$$

where $F(x) = 1 \sum (ni\ Fi\ (x))$, n = n1 + n2, ni is the number of observation of class i, t_j is the number of ties at the jth distinct value, and p is the number of distinct values.

The Kuiper statistic is closely related to the Kolmogorov-Smirnov statistic. The Kuiper statistic uses not only the information of maximum distance between two empirical distributions as in the Kolmogorov-Smirnov statistic, but also the information of minimum

distance between two empirical distributions. The exact formula of the Kuiper statistic is as follows.

$$K = \max(F1(xj) - F2(xj)) - \min(F1(xj) - F2(xj)) \text{where} j = 1, 2,...,n \; jj$$

"There are three types of lies; lies, damn lies, and statistics"

-Benjamin Disraeli-

Chapter 3

Results Presentation and Analysis

3.1 Portfolio Returns Sorted on Distribution Uncertainty

The first empirical investigation is whether distribution uncertainty can explain the cross-sectional variation of expected stock returns. Table 3.1 reports time series average (AR) and holding period returns (HPR) of decile portfolios formed on each of the three distribution uncertainty measures. This table has been constructed from the calculation of these measures for each sample firm over the previous month. Each month we sort stocks into 10 equal-weighted portfolios using these measures for distribution uncertainty (KS, CM, K). The initial AR represents average daily returns in percentage multiplied by 21, and HPR is the holding period return of decile portfolio rebalanced each month from 2001 to 2021. The portfolios sorted on three distribution uncertainty measures demonstrate strong variation in mean return, as shown in Table 3.1 below.

Table 3.1 Portfolio Returns Sorted on Distribution Uncertainty

This table presents equal-weighted average returns (AR) and holding period returns (HPR) for portfolios formed on each distribution uncertainty proxy within a month. We multiply daily returns by 21 to obtain monthly returns. All figures are expressed in percentage terms. The decile portfolios updated each month are formed by the sizes of Kolmogorov-Smirnov (KS), Cramer-von Mises (CM), and Kuiper (K) statistics estimated using daily demeaned individual stock return and value weighted index return over the previous 12 months. These statistics of KS, CM, and K non-parametrically measure the difference of distributions between demeaned individual stock return and demeaned market. Portfolio "S" is the portfolio of stocks with the lowest distribution uncertainty measures, Portfolio "B" is the portfolio of stocks with the highest distribution uncertainty measures, "S-B" is their difference in monthly returns, and t-statistics are reported in parentheses. ***, **, * correspond to statistical

significance at 1, 5, and 10%, respectively. The sample includes all firms listed in NYSE, S&P 500, and NASDAQ from 2001 to 2021

	KS				CM				K		
	KS	AR	HPR		CM	AR	HPR		K	AR	HPR
S	0.05	1.06	108.83		1.87	0.44	145.45		0.08	1.39	106.22
2	0.07	0.71	146.05		4.78	1.16	227.36		0.12	0.56	97.08
3	0.09	0.28	145.59		7.9	2.11	48.17		0.16	0.4	177.15
4	0.11	1.12	214.73		10.94	3.19	13.93		0.2	1.15	238.97
5	0.13	2.09	53.8		14.12	3.58	682.48		0.24	2.09	28.46
6	0.16	2.86	185.9		17.27	3.74	2163.23		0.28	2.91	94.85
7	0.18	3.65	806.31		20.37	4	317.5		0.32	3.63	641.66
8	0.2	3.9	1470.93		23.41	3.94	126.16		0.36	4.05	1824.94
9	0.22	4.24	151.38		26.75	5.04	199.71		0.4	3.99	134.25
B	0.24	5.65	140.26		29.76	5.13	187.47		0.44	4.88	191.57
9-S		5.3	0.43			5.48	0.54			5.39	0.28
t(9-S)		(345.28)***	(44.38)***			(300.32)***	(24.91)***			(427.24)***	(28.5)***
B-S		6.71	0.31			5.57	0.42			6.27	0.85
t(B-S)		(267.14)***	(21.85)***			(144.65)***	(13.72)***			(304.38)***	(46.31)***

The results show that the average returns (AR) on the decile portfolios sorted by distribution uncertainty increase monotonically in portfolio rank. The bottom decile portfolio (S) by K has 1.39% of expected return per month on average and the top decile portfolio (B) does 4.88%. The B-S spread shows 6.27% of expected return per month and t-statistic of 304.38. When a decile portfolio formed by KS, stocks (S) with at least distribution uncertainty provide 1.06% of expected return per month on average and the stocks (B) with the most distribution uncertainty do 5.65%. Furthermore, the top decile portfolio by CM seems to demonstrate considerably higher returns than the bottom decile portfolio. Because of the cross-sectional dispersion of returns being most striking between the 9th decile portfolio and the top decile portfolio (B), the calculation for the return spread of 9-S for robustness check is necessary. The results are still sustained with large, expected returns of more than 5%. Therefore, the stocks with the most distribution uncertainty have higher expected return than do stocks with the least distribution uncertainty. This implies that since investors need to spend more resources to understand unfamiliar distributions of a stock compared to that of the benchmark portfolio,

investors may require a premium for bearing distribution uncertainty. So, the results show this evidence for a positive premium for bearing distribution uncertainty.

3.2 Portfolio Returns Sorted on Distribution Uncertainty and Firm Characteristics

The second empirical investigation is between distribution uncertainty and future stock returns after the control of firm characteristics. For instance, stocks which have high distribution uncertainty tend to be small and illiquid. In order to ensure that the distribution uncertainty is not being affected from these characteristics, the second empirical test will be the investigation of the profitability of portfolios sorted by distribution uncertainty after controlling for firm characteristics such as beta, size, book-to-market ratio, momentum, short-term reversal, and illiquidity. The beta of a stock for a month (BETA) is estimated by regressing the daily stock return on the value weighted index return using a previous year sample. SIZE is the natural logarithm of the market value of equity of the company (in thousands of dollars) measured by times series average of a firm's market capitalization for the most recent 12 months. Book-to-market ratio (BM) is the book value of equity divided by its market value at the end of the last fiscal year. Momentum (MOM) is the cumulative stock return over the previous eleven (11) months starting two (2) months ago in order to isolate it from short-term reversal effect. Additionally, short-term reversal (REV) being measured for each stock in month "t" as the return on the stock over the previous month and by Amihud (2002), stock illiquidity (ILLIQ) is defined as the ratio of the absolute monthly stock return to its dollar trading volume.

Table 3.2 Portfolios Returns Sorted on Distribution Uncertainty and Firm Characteristics

This table reports average returns (AR) for portfolios based on distribution uncertainty proxies and firm characteristics. We multiply daily returns by 21 to obtain monthly returns and report the monthly returns in percent. In each case, we first sort the stocks into deciles using the firm characteristics. Within each characteristic's decile, we sort stocks into ten additional portfolios based on distribution uncertainty proxy (KS, CM, K) and compute the returns on the corresponding portfolios over the subsequent month. These statistics of KS, CM, and K non-parametrically measure the difference of distributions between demeaned individual stock return and demeaned market. This table presents average returns across the firm characteristic deciles. Portfolio "S" is the portfolio of stocks with the lowest distribution uncertainty measures, Portfolio "B" is the portfolio of stocks with

the highest distribution uncertainty measure, "S-B" is their difference at 1, 5, and 10% respectively. The sample includes all firms listed in NYSE, S&P 500, and NASDAQ from 2001 to 2021.

Panel A. KS						
	BETA	**SIZE**	**BM**	**MOM**	**REV**	**ILLIQ**
S	0.23	0.06	0.24	0.07	0.32	0.13
2	0.04	0.13	0.03	0	0.09	0.16
3	0.03	0.19	0.07	0.12	0	0.17
4	0.13	1.15	0.16	0.72	0.09	0.12
5	0.18	1.91	0.2	2.83	0.17	0.43
6	0.26	3.54	0.36	4.67	0.29	0.69
7	0.49	9.01	0.47	9.12	0.46	3.02
8	0.38	16.63	0.25	13.63	0.36	8.97
9	0.31	23.13	0.7	14.23	0.7	5.45
B	0.85	21.77	0.27	35.1	0.95	24.64
B-S	0.62	21.71	0.03	35.03	0.62	24.52
t(B-S)	(4.71)***	(8.7)***	(0.9)***	(21.23)***	(2.11)***	(10.01)***
Panel B. CM						
	BETA	**SIZE**	**BM**	**MOM**	**REV**	**ILLIQ**
S	0.04	0.19	0.04	0.29	0.06	0.1
2	0.12	0.18	0.12	0.11	0.1	0.29
3	0.19	0.22	0.19	0.01	0.16	0.46
4	0.31	0.11	0.3	0.11	0.31	0.32
5	0.37	0.13	0.37	0.12	0.43	0.29
6	0.4	0.14	0.4	0.13	0.37	0.28
7	0.3	0.01	0.26	0.54	0.27	0.55
8	0.41	0.01	0.35	0.64	0.31	0.4
9	0.87	1.74	0.89	0.81	0.82	0.52
B	1.79	1.52	1.79	0.65	1.7	1.62
B-S	1.75	1.33	1.75	0.36	1.64	1.52
t(B-S)	(4.42)***	(2.69)***	(4.41)***	(0.75)***	(2.51)***	(1.32)***
Panel C. K						
	BETA	**SIZE**	**BM**	**MOM**	**REV**	**ILLIQ**
S	0.32	0.01	0.32	0.09	0.36	0.01
2	0.01	0.14	0.01	0.01	0.08	0.15
3	0.05	0.13	0.05	0.08	0.01	0.24
4	0.15	0.24	0.15	0.07	0.1	0.19
5	0.19	0.48	0.19	0.34	0.16	0.16
6	0.3	0.89	0.29	0.37	0.34	0.01

7	0.33	3.45	0.3	1.86	0.3	0.42
8	0.42	3.22	0.4	2.53	0.4	0.09
9	0.34	4.37	0.35	0.93	0.23	0.32
B	0.72	14.09	0.74	18.79	0.87	9.5
B-S	0.4	14.1	0.42	18.69	0.51	9.5
t(B-S)	(4.82)***	(4.3)***	(4.83)***	(18.91)***	(3.74)***	(2.72)***

By Balu et al. (2011) and Baltussen et al. (2013), Table 3.2 shows monthly returns averaged across the portfolios formed by two-way sorts on a stocks return's distribution uncertainty and firm characteristics. Firstly, I am going to categorise the stocks into 10 groups by firm characteristics. Secondly, within each decile portfolio sorting further stocks into decile portfolios ranked based upon our KS, CM, and K statistics, which the results will be in total of 100 portfolios. The third and last move will be the average of each distribution uncertainty portfolios across the firm characteristic's deciles. As Baltussen et al. (2013) argue, it is possible to control each form characteristic without assuming a parametric form about the relationship between distribution uncertainty and future stock returns. For each of these portfolios, we calculate average equal-weighted returns over the following month.

The first column of Panel A in Table II reports returns averaged across the ten beta deciles to produce decile portfolios with dispersion in KS. Since we average across beta deciles, the produced decile portfolios sorted by KS will include all betas. The portfolio returns for each month are calculated as an equal-weighted average of returns from strategies initiated at the end of the past month. After controlling for beta, the average return difference between the low and high KS portfolios is about 0.62% per month with a t-statistic of 4.71. It suggests that the positive relation between distribution uncertainty and future stock returns is not affected by beta. The results in Panel A show that the highest distribution uncertainty firms earn an average of 21.77%, compared to 0.06% for the smallest distribution uncertainty firms, when we control for size. The return differential between these two deciles (B-S) is 21.77% and significant (t=8.7). When controlling for book-to market ratio (BM), the return differentials between B and S are also positive and significant. When stocks are sorted based on momentum, the average return of the big-small portfolio is 35.03%, with a t-statistic of 21.23. Subsequently, the average excess return of the B-S portfolio equals 0.62% per month, with t-statistic 2.11 when controlling for short-term reversal. Finally, we see whether the illiquidity explains the higher returns for the highest distribution uncertainty stocks relative to the smallest distribution

uncertainty stocks. The average return of the B-S portfolio is 24.52% per month with a t-statistic of 10.01. These results suggest that a positive distribution uncertainty premium remains and firm characteristics do not explain the positive relation between distribution uncertainty and futures stock returns. Panel B of Table II presents average monthly returns to portfolios formed by two-way sorts on CM and firm characteristics. We find similar, confirmatory evidence in Panel B with CM as a proxy for distribution uncertainty. In Panel C, we examine the performance of K-sorted portfolios after controlling firm characteristics. The results with K are also similar to those in Panel A and Panel B. Overall, the results from these robustness tests using alternative measures of distribution uncertainty still support our hypothesis.

3.3 Alphas of Portfolios Sorted on Distribution Uncertainty

The third and last empirical investigation is to examine whether a rational risk-based approach can explain our result that the degree of distribution uncertainty provides premium. Table 3.3 shows the equal-weighted portfolios' postranking alphas estimated under three different factor specifications the capital asset pricing model (CAPM), the three factors proposed in Fama and French (1993), and the four-factor proposed in Carhart (1997).

Table 3.3 Alphas of portfolios Sorted on Distribution Uncertainty

This table reports the alphas of the CAPM, the Fama-French 3-factor model and the Carhart (1997) 4-factor models for 10 portfolios based on three proxies for distribution uncertainty. The decile portfolios updated each month and are formed by the sizes of Kolmogorov-Smirnov (KS), Cramer-von Mises (CM), and Kuiper (K) statistics estimated using daily demeaned individual stock return and value weighted index return over the previous 12 months. These statistics of KS, CM, and K are non-parametrically measures as much as differences of distributions between demeaned individual stock return and demeaned market. Alphas are from a time series regression of the daily returns on daily Rm-Rf, SMB, HML, and UMD as in Fama and French (1993) and Carhart (1997). We multiply daily alphas by 21 to obtain monthly alphas and report the monthly alphas in percentages. Portfolio "S" is the portfolio of stocks with the lowest distribution uncertainty measure, Portfolio "B" is the portfolio of stocks with the highest distribution uncertainty measure, "S-B" is their difference in monthly returns. The sample includes all firms listed in NYSE, S&P 500, and NASDAQ from 2001 to 2021.

Panel A. Kolmogorov-Smirnov (KS) Statistic						
	CAPM		**Fama-French 3 Factor**		**Carhart 4 Factor**	
	Alpha	**Adj. Rsq**	**Alpha**	**Adj. Rsq**	**Alpha**	**Adj. Rsq**
S	0.6065	0.3362	0.6046	0.3395	0.6035	0.3412

2	0.6858	0.3535	0.6948	0.3582	0.6936	0.3599
3	0.8331	0.3977	0.8385	0.3976	0.8369	0.3991
4	0.9545	0.3713	0.9631	0.3744	0.9614	0.3761
5	1.0825	0.3653	1.0866	0.3638	1.0844	0.3656
6	1.1546	0.3459	1.1552	0.3434	1.1529	0.3454
7	1.2264	0.3157	1.2329	0.3155	1.2304	0.3177
8	1.3244	0.2941	1.3129	0.2895	1.3105	0.2919
9	1.3752	0.2823	1.4032	0.2844	1.4008	0.2874
B	1.4818	0.2850	1.5258	0.2856	1.5265	0.2900
B-S	0.8753		0.9212		0.9230	
Panel B. Cramer-Mises (CM) Statistic						
	CAPM		**Fama-French 3 Factor**		**Carhart 4 Factor**	
	Alpha	**Adj. Rsq**	**Alpha**	**Adj. Rsq**	**Alpha**	**Adj. Rsq**
S	0.7257	0.3679	0.7282	0.3699	0.7268	0.3717
2	0.9426	0.3809	0.9450	0.3814	0.9433	0.3831
3	1.0969	0.3658	1.0947	0.3657	1.0926	0.3674
4	1.1784	0.3531	1.1807	0.3528	1.1783	0.3547
5	1.2077	0.3125	1.2128	0.3125	1.2104	0.3147
6	1.3094	0.3145	1.3021	0.3121	1.2992	0.3142
7	1.3255	0.2951	1.3335	0.3024	1.3313	0.3052
8	1.4402	0.3027	1.4591	0.2897	1.4565	0.2922
9	1.5653	0.2723	1.5530	0.2772	1.5506	0.2795
B	1.4664	0.2626	1.4938	0.2419	1.4938	0.2437
B-S	0.7406		0.7656		0.7670	
Panel C. Kuiper (K) Statistic						
	CAPM		**Fama-French 3 Factor**		**Carhart 4 Factor**	
	Alpha	**Adj. Rsq**	**Alpha**	**Adj. Rsq**	**Alpha**	**Adj. Rsq**
S	0.6094	0.3358	0.6110	0.3380	0.6100	0.3401
2	0.7018	0.3652	0.7024	0.3653	0.7012	0.3673
3	0.8290	0.3827	0.8316	0.3860	0.8301	0.3878
4	0.9600	0.3786	0.9658	0.3802	0.9640	0.3819
5	1.0821	0.3645	1.0797	0.3639	1.0776	0.3657
6	1.1685	0.3531	1.1713	0.3525	1.1689	0.3543
7	1.2196	0.3188	1.2222	0.3176	1.2196	0.3198
8	1.3255	0.3068	1.3176	0.3052	1.3150	0.3075
9	1.3979	0.2906	1.4258	0.2882	1.4233	0.2904
B	1.5740	0.2965	1.5520	0.2845	1.5503	0.2870
B-S	0.9646		0.9410		0.9403	

The results in Panel A show that our measures for distribution uncertainty are highly correlated with alphas estimated from three different factor specifications. The magnitude of the alpha is positively related to the level of distribution uncertainty, which implies that the high distribution uncertainty portfolios earn more positive abnormal returns. All three alphas of the B-S spread are significantly positive. The CAPM alpha is 0.8753% per month, the three-factor alpha is 0.9212% per month, and the four-factor alpha is 0.9230% per month. A trading strategy with a short position in the low distribution uncertainty firms and a long position in high distribution uncertainty firms generates a monthly abnormal return of 0.9230% after controlling for the market, size, value, and momentum effects. This pattern of alphas from the three different factor specifications implies that the abnormal returns of B-S portfolios are not specific to an asset pricing models and confirms our hypothesis of distribution uncertainty premium. The results of positive alphas are also robust across various distribution uncertainty proxies.

Conclusion

This thesis investigates the significance of uncertainty of the return distribution in the cross-sectional pricing stocks. I am using proxies for distribution uncertainty of a stock return, the Kolmogorov-Smirnov (KS), Cramer-von Mises (CM), and Kuiper (K) statistics, which non-parametrically measure difference between empirical return distributions of an individual stock and a benchmark portfolio.

The results show that stocks with severe distribution uncertainty exhibit high returns on average, and the difference between returns on the portfolios with highest and lowest distribution uncertainty is 6% per month. The corresponding four-factor alphas from B-S KS, CM, K sorted portfolios are 0.76% to 0.94% a month. Nevertheless, after extensive consideration for the robustness of the empirical results we found that the impact of distribution uncertainty persists after accounting for firm characteristics such as beta, size, book-to-market ratio, momentum, short-term reversal, and illiquidity.

References (Books)

Feller, William, 1971, *An Introduction to Probability Theory and Its Application,* Volume II, John Wiley & Sons, Inc. New York.

Gibbons, Jean Dickinson and Subhabrata Chakraborti, 2010, *Nonparametric Statistical Inference*, 4th end, Marcel Dekker Inc., New York.

Knight Frank H., 1921, *Risk, Uncertainty and Profit*, 1st end, Houston Mifflin, Boston, MA.

Savage, L. J., 1954, *The Foundations of Statistics*, Wiley, New York.

References (Papers)

Anscombe F.J., Aumann R.J., 1963, A definition of subjective probability, *The Annals of mathematical Statistics*, vol. 34, No.1

Amihud Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Baltussen Guido, Sjoerd Van Bekkum, and Bart Van Der Grient, 2013, Unknown unknowns: Vol-of-vol and the cross-section of stock returns, *Working Paper Erasmus University*.

Chae Joon and Lee Eun Jung, 2018, Distribution uncertainty and expected stock returns, Finance Research Letters volume 25, 55-61.

Camerer Colin, and Martin Weber, 1992, Recent developments in modelling preferences: Uncertainty and ambiguity, *Journal of Risk and Uncertainty* 5, 325-370

Campbell John Y., Adrew W. Lo, A. Graig MacKinlay, 1997, *"The Econometrics of Financial Markets",* Princeton University Press, Princeton New Jersey

Carhart Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

Ellsberg Daniel, 1961, Risk ambiguity and the Savage axioms, *Quarterly Journal of Economics* 75, 643-669.

Fama Eugene and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fox, Craig R. And Amos Tversky, 1995, Ambiguity aversion and comparative ignorance, *Quarterly Journal of Economics* 110, 585-603

Fisher Irving, 1919, The nature of capital and income, *New York: The Macmillan Company; London: Macmillan & Co., Ltd.*

Kacperczyk Marcin and Paul Damien, 2011, Asset allocation under multivariate regime switching, *Journal of Economic Dynamics and Control* 31.

Keynes John Maynard, 1921, A treatise on Probability *London: Macmillan*.

Markowitz Harry, 1952a, Portfolio selection, *Journal of Finance* 7, 77-91.

von Neumann, J.; Morgenstern, O. (1947). Theory of Games and Economic Behaviour, 2nd ed. *University Press, Princeton.*

Neyman, J.; Pearson, E. S. (1933-02-16). "IX. On the problem of the most efficient tests of statistical hypotheses". *Phil. Trans. R. Soc. Lond. A.* 231 (694–706): 289–337

Ramsay, Frank Plumpton, 1926, "Truth and Probability", Chapter VII and "Further Considerations "(1928) and "Probability and Partial Belief" (1929) in *The Foundations of Mathematics and other Logical Essays*

Steele Katie & Stefansson H. Orri, 2015, *"Decision Theory"* Stanford Encyclopaedia of Philosophy.

Suppes, Patrick, 2002, *Representation and Invariance of Scientific Structures*, Stanford: CSLI Publications.

Sherman Roger, 1974, The psychological difference between ambiguity and risk, *Quarterly Journal of Economics* 88, 166-169.

Jack L. Treynor, 1961, Market Value, Time, and Risk, Independent, p.46

Jack L. Treynor, 1962, Toward a Theory of Market Value of Risky Assets, Independent, p.20

Turan G. Bali, Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* volume 99, p. 427-446.

Zellner Arnold, 1971, An introduction to Bayesian inference in econometrics, John Wiley and Sons *New York*.

Appendix

Load Pre-process

```
library(dplyr)
library(twosamples)

#data <- readRDS("data.RDS")

#################################
## close prices
#################################
sp <- readxl::read_excel("SP500.xlsx", sheet = 2)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
```

```
        Name = as.Date(Name, origin = "1899-12-30"),
        Name = as.character(Name)) %>%
  rename("Date" = "Name")

nyse <- readxl::read_excel("NYSE.xlsx", sheet = 1)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
        Name = as.Date(Name, origin = "1899-12-30"),
        Name = as.character(Name)) %>%
  rename("Date" = "Name")

nasdaq <- readxl::read_excel("NASDAQ.xlsx", sheet = 1)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
        Name = as.Date(Name, origin = "1899-12-30"),
        Name = as.character(Name)) %>%
  rename("Date" = "Name")

################################
## get firm codes
################################
sp_codes <- readxl::read_excel("SP500.xlsx", sheet = 2, col_names = FALSE)[1:2,] %>%
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", ""))

nyse_codes <- readxl::read_excel("NYSE.xlsx", sheet = 1, col_names = FALSE)[1:2,]
%>%
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", ""))

nasdaq_codes <- readxl::read_excel("NASDAQ.xlsx", sheet = 1, col_names =
FALSE)[1:2,] %>%
```

```
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", ""))

#########################################################
## market (benchmark) portfolio - composite indexes
#########################################################
bench_sp <- readxl::read_excel("SP500.xlsx", sheet = 1)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
         Name = as.Date(Name, origin = "1899-12-30"),
         Name = as.character(Name)) %>%
  rename("Date" = "Name")

bench_nyse <- readxl::read_excel("NYSE.xlsx", sheet = 2)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
         Name = as.Date(Name, origin = "1899-12-30"),
         Name = as.character(Name)) %>%
  rename("Date" = "Name")

bench_nasdaq <- readxl::read_excel("NASDAQ.xlsx", sheet = 2)[-c(1,2),] %>%
  mutate(Name = as.numeric(Name),
         Name = as.Date(Name, origin = "1899-12-30"),
         Name = as.character(Name)) %>%
  rename("Date" = "Name")



## merging returns with respective composite indexes (based on date)
sp <- left_join(sp, bench_sp, by = "Date")
nyse <- left_join(nyse, bench_nyse, by = "Date")
nasdaq <- left_join(nasdaq, bench_nasdaq, by = "Date")

sp <- sp %>%
```

```
    tidyr::gather("stock", "close", -c("Date", "S&P 500 COMPOSITE - PRICE INDEX")) %>%
    tidyr::separate(Date, c("year", "month", "day"))
  names(sp)[4] <- "cpi"

  nyse <- nyse %>%
    tidyr::gather("stock", "close", -c("Date", "NYSE COMPOSITE - PRICE INDEX"))%>%
    tidyr::separate(Date, c("year", "month", "day"))
  names(nyse)[4] <- "cpi"

  nasdaq <- nasdaq %>%
    tidyr::gather("stock", "close", -c("Date", "NASDAQ COMPOSITE - PRICE INDEX"))%>%
    tidyr::separate(Date, c("year", "month", "day"))
  names(nasdaq)[4] <- "cpi"

  rm(bench_nasdaq, bench_nyse, bench_sp)



  ## merge respective codes to each dataset
  sp <- left_join(sp, sp_codes, by = c('stock' = 'Name'))
  nyse <- left_join(nyse, nyse_codes, by = c('stock' = 'Name'))
  nasdaq <- left_join(nasdaq, nasdaq_codes, by = c('stock' = 'Name'))

  rm(sp_codes, nyse_codes, nasdaq_codes)

  ## calculate returns - firms and market
  sp <- sp %>%
    group_by(stock) %>%
    mutate(cpi = as.numeric(cpi),
           cpi = as.vector(quantmod::Delt(cpi)),
           close = as.numeric(close),
           return = as.vector(quantmod::Delt(close)))
```

```
nyse <- nyse %>%
  group_by(stock) %>%
  mutate(cpi = as.numeric(cpi),
         cpi = as.vector(quantmod::Delt(cpi)),
         close = as.numeric(close),
         return = as.vector(quantmod::Delt(close)))

nasdaq <- nasdaq %>%
  group_by(stock) %>%
  mutate(cpi = as.numeric(cpi),
         cpi = as.vector(quantmod::Delt(cpi)),
         close = as.numeric(close),
         return = as.vector(quantmod::Delt(close)))



## mean center stock returns and marker portfolio based on previous year
## (2002[first year available] cannot be therefore mean centered)

sp_rm <- sp %>%
  group_by(stock, year) %>%
  summarise(avg_return = mean(return, na.rm = TRUE),
            avg_cpi = mean(cpi, na.rm = TRUE)) %>%
  mutate(year = as.numeric(year),
         year = year-1)

sp <- sp %>%
  mutate(year = as.numeric(year)) %>%
  left_join(sp_rm, by = c("year", "stock")) %>%
  mutate(return = return - avg_return,
         cpi = cpi - avg_cpi) %>%
  select(-c(avg_return, avg_cpi)) %>%
  mutate(exchange = "sp")
```

```
nasdaq_rm <- nasdaq %>%
  group_by(stock, year) %>%
  summarise(avg_return = mean(return, na.rm = TRUE),
            avg_cpi = mean(cpi, na.rm = TRUE)) %>%
  mutate(year = as.numeric(year),
         year = year-1)

nasdaq <- nasdaq %>%
  mutate(year = as.numeric(year)) %>%
  left_join(sp_rm, by = c("year", "stock")) %>%
  mutate(return = return - avg_return,
         cpi = cpi - avg_cpi) %>%
  select(-c(avg_return, avg_cpi)) %>%
  mutate(exchange = "nasdaq")

nyse_rm <- nyse %>%
  group_by(stock, year) %>%
  summarise(avg_return = mean(return, na.rm = TRUE),
            avg_cpi = mean(cpi, na.rm = TRUE)) %>%
  mutate(year = as.numeric(year),
         year = year-1)

nyse <- nyse %>%
  mutate(year = as.numeric(year)) %>%
  left_join(sp_rm, by = c("year", "stock")) %>%
  mutate(return = return - avg_return,
         cpi = cpi - avg_cpi) %>%
  select(-c(avg_return, avg_cpi)) %>%
  mutate(exchange = "nyse")

## combine all 3
data <- rbind(sp, nyse, nasdaq)
rm(nasdaq, nyse, sp, nasdaq_rm, sp_rm, nyse_rm)
```

Table 1

```
################################
#####  K-S  ###################

data %>%
  group_by(stock, month) %>%
  filter(!is.na(cpi)) %>%
  mutate(ks = ks_stat(return, cpi)) %>%
  ungroup %>%
  mutate(decile = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
  group_by(stock) %>%
  mutate(ar = sum(return),
         hpr = sum(FinCal::hpr(first(return), last(return)))) %>%
  ungroup %>%
  group_by(decile) -> ks


ks %>%
  summarise(ks = round(mean(ks), 2),
            ar = round(abs(mean(ar)), 2),
            hpr = round(abs(mean(hpr, na.rm = TRUE))*100, 2)) -> ks_tb1

ks %>% group_split() -> ks_rest


dif1_ks <- round(abs(mean(ks_rest[[9]]$ar) - mean(ks_rest[[1]]$ar)), 2) ## S-9 decile
difference - AR
tt1_ks <- paste0("(", round(abs(t.test(ks_rest[[1]]$ar, ks_rest[[9]]$ar)$statistic), 2), ")***")
## S-9 decile t.test - AR
```

```
    dif2_ks <- round(abs(mean(ks_rest[[10]]$ar) - mean(ks_rest[[1]]$ar)), 2) ## B-S decile
difference - AR
    tt2_ks <- paste0("(", round(abs(t.test(ks_rest[[1]]$ar, ks_rest[[10]]$ar)$statistic), 2),
")***")   ## B-S decile t.test - AR

    dif3_ks <- round(abs(mean(ks_rest[[9]]$hpr) - mean(ks_rest[[1]]$hpr)), 2) ## S-9 decile
difference - HPR
    tt3_ks <- paste0("(", round(abs(t.test(ks_rest[[9]]$hpr, ks_rest[[1]]$hpr)$statistic), 2),
")***")   ## S-9 decile t.test - HPR

    dif4_ks <- round(abs(mean(ks_rest[[10]]$hpr) - mean(ks_rest[[1]]$hpr)), 2) ## B-S decile
difference - HPR
    tt4_ks <- paste0("(", round(abs(t.test(ks_rest[[10]]$hpr, ks_rest[[1]]$hpr)$statistic), 2),
")***")   ## B-S decile t.test - HPR


    data.frame(decile = c("9-S", "t(9-S)", "B-S", "t(B-S)"),
               ks = c("","","",""),
               ar = c(dif1_ks, tt1_ks, dif2_ks, tt2_ks),
               hpr = c(dif3_ks, tt3_ks, dif4_ks, tt4_ks)) -> ks_rest

    tb_ks <- rbind(ks_tb1, ks_rest)

    rm(dif1_ks, dif2_ks, dif3_ks, dif4_ks, tt1_ks, tt2_ks, tt3_ks, tt4_ks, ks, ks_tb1, ks_rest)


    ################################
    #####  CM  ###################

    data %>%
      group_by(stock, month) %>%
      filter(!is.na(cpi)) %>%
      mutate(cm = cvm_stat(return, cpi)) %>%
      ungroup %>%
```

```
    mutate(decile = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
    group_by(stock) %>%
    mutate(ar = sum(return),
           hpr = sum(FinCal::hpr(first(return), last(return)))) %>%
    ungroup %>%
    group_by(decile) -> cm


  cm %>%
    summarise(cm = round(mean(cm), 2),
              ar = round(abs(mean(ar)), 2),
              hpr = round(abs(mean(hpr, na.rm = TRUE))*100, 2)) -> cm_tb1

  cm %>% group_split() -> cm_rest


  dif1_cm <- round(abs(mean(cm_rest[[9]]$ar) - mean(cm_rest[[1]]$ar)), 2) ## S-9 decile
difference - AR
  tt1_cm <- paste0("(", round(abs(t.test(cm_rest[[1]]$ar, cm_rest[[9]]$ar)$statistic), 2),
")***")   ## S-9 decile t.test - AR

  dif2_cm <- round(abs(mean(cm_rest[[10]]$ar) - mean(cm_rest[[1]]$ar)), 2) ## B-S decile
difference - AR
  tt2_cm <- paste0("(", round(abs(t.test(cm_rest[[1]]$ar, cm_rest[[10]]$ar)$statistic), 2),
")***")   ## B-S decile t.test - AR

  dif3_cm <- round(abs(mean(cm_rest[[9]]$hpr) - mean(cm_rest[[1]]$hpr)), 2) ## S-9 decile
difference - HPR
  tt3_cm <- paste0("(", round(abs(t.test(cm_rest[[9]]$hpr, cm_rest[[1]]$hpr)$statistic), 2),
")***")   ## S-9 decile t.test - HPR

  dif4_cm <- round(abs(mean(cm_rest[[10]]$hpr) - mean(cm_rest[[1]]$hpr)), 2) ## B-S
decile difference - HPR
```

```
    tt4_cm <- paste0("(", round(abs(t.test(cm_rest[[10]]$hpr, cm_rest[[1]]$hpr)$statistic), 2),
")***")   ## B-S decile t.test - HPR


    data.frame(decile = c("9-S", "t(9-S)", "B-S", "t(B-S)"),
               cm = c("","","",""),
               ar = c(dif1_cm, tt1_cm, dif2_cm, tt2_cm),
               hpr = c(dif3_cm, tt3_cm, dif4_cm, tt4_cm)) -> cm_rest

    tb_cm <- rbind(cm_tb1, cm_rest)

    rm(dif1_cm, dif2_cm, dif3_cm, dif4_cm, tt1_cm, tt2_cm, tt3_cm, tt4_cm, cm, cm_tb1,
cm_rest)


    ###################################
    #####  Kuiper  ###################

    data %>%
      group_by(stock, month) %>%
      filter(!is.na(cpi)) %>%
      mutate(k = kuiper_stat(return, cpi)) %>%
      ungroup %>%
      mutate(decile = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
      group_by(stock) %>%
      mutate(ar = sum(return),
             hpr = sum(FinCal::hpr(first(return), last(return)))) %>%
      ungroup %>%
      group_by(decile) -> k


    k %>%
      summarise(k = round(mean(k), 2),
```

```
            ar = round(abs(mean(ar)), 2),
            hpr = round(abs(mean(hpr, na.rm = TRUE))*100, 2)) -> k_tb1

    k %>% group_split() -> k_rest


    dif1_k <- round(abs(mean(k_rest[[9]]$ar) - mean(k_rest[[1]]$ar)), 2) ## S-9 decile
difference - AR
    tt1_k <- paste0("(", round(abs(t.test(k_rest[[1]]$ar, k_rest[[9]]$ar)$statistic), 2), ")***")
## S-9 decile t.test - AR

    dif2_k <- round(abs(mean(k_rest[[10]]$ar) - mean(k_rest[[1]]$ar)), 2) ## B-S decile
difference - AR
    tt2_k <- paste0("(", round(abs(t.test(k_rest[[1]]$ar, k_rest[[10]]$ar)$statistic), 2), ")***")
## B-S decile t.test - AR

    dif3_k <- round(abs(mean(k_rest[[9]]$hpr) - mean(k_rest[[1]]$hpr)), 2) ## S-9 decile
difference - HPR
    tt3_k <- paste0("(", round(abs(t.test(k_rest[[9]]$hpr, k_rest[[1]]$hpr)$statistic), 2), ")***")
## S-9 decile t.test - HPR

    dif4_k <- round(abs(mean(k_rest[[10]]$hpr) - mean(k_rest[[1]]$hpr)), 2) ## B-S decile
difference - HPR
    tt4_k <- paste0("(", round(abs(t.test(k_rest[[10]]$hpr, k_rest[[1]]$hpr)$statistic), 2),
")***")   ## B-S decile t.test - HPR


    data.frame(decile = c("9-S", "t(9-S)", "B-S", "t(B-S)"),
            k = c("","","",""),
            ar = c(dif1_k, tt1_k, dif2_k, tt2_k),
            hpr = c(dif3_k, tt3_k, dif4_k, tt4_k)) -> k_rest

    tb_k <- rbind(k_tb1, k_rest)
```

```
rm(dif1_k, dif2_k, dif3_k, dif4_k, tt1_k, tt2_k, tt3_k, tt4_k, k, k_tb1, k_rest)
```

Table 2

```
#####################################################
############ Firm Characteristics ##################
#####################################################
char_ids_sp <- readxl::read_excel("NC.xlsx", sheet = 5, col_names = FALSE)[1:2,] %>%
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  tidyr::separate(Name, c("stock", "char"), " - ") %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", "")) %>%
  na.omit

sp <- readxl::read_excel("NC.xlsx", sheet = 5)[-c(1,2),] %>%
  rename("year" = "Name") %>%
  tidyr::gather("stock", "value", -"year") %>%
  tidyr::separate(stock, c("stock", "char"), " - ") %>%
  left_join(char_ids_sp, by = c("stock", "char")) %>%
  select(-stock) %>%
  filter(!is.na(Code))

char_ids_nyse <- readxl::read_excel("NC.xlsx", sheet = 4, col_names = FALSE)[1:2,] %>%
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  tidyr::separate(Name, c("stock", "char"), " - ") %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", "")) %>%
  na.omit
```

```
nyse <- readxl::read_excel("NC.xlsx", sheet = 4)[-c(1,2),] %>%
  rename("year" = "Name") %>%
  tidyr::gather("stock", "value", -"year") %>%
  tidyr::separate(stock, c("stock", "char"), " - ") %>%
  left_join(char_ids_nyse, by = c("stock", "char")) %>%
  select(-stock) %>%
  filter(!is.na(Code))

char_ids_nasdaq <- readxl::read_excel("NC.xlsx", sheet = 3, col_names = FALSE)[1:2,]
%>%
  t() %>%
  as_tibble() %>%
  janitor::row_to_names(row_number = 1) %>%
  tidyr::separate(Name, c("stock", "char"), " - ") %>%
  mutate(Code = stringr::str_replace(Code, "\\s*\\([^\\)]+\\)", "")) %>%
  na.omit

nasdaq <- readxl::read_excel("NC.xlsx", sheet = 3)[-c(1,2),] %>%
  rename("year" = "Name") %>%
  tidyr::gather("stock", "value", -"year") %>%
  tidyr::separate(stock, c("stock", "char"), " - ") %>%
  left_join(char_ids_nasdaq, by = c("stock", "char")) %>%
  select(-stock) %>%
  filter(!is.na(Code))


rm(char_ids_nasdaq, char_ids_nyse, char_ids_sp)

chars <- rbind(sp, nyse, nasdaq) %>%
  distinct(year, char, Code, .keep_all = TRUE) %>%
  mutate(year = as.numeric(year))
```

```
rm(nasdaq, nyse, sp)

## market capitalization - FIRM SIZE
size_dt <- chars %>%
  group_by(Code) %>%
  filter(char == "MARKET CAPITALIZATION") %>%
  mutate(value = log(as.numeric(value))) %>%
  filter(!is.na(value)) %>%
  left_join(select(data, year, Code, return, month, cpi), by = c("year", "Code")) %>%
  ungroup %>%
  filter(!is.na(cpi)) %>%
  mutate(decile = cut(x = value, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
  group_by(decile, month, stock) %>%
  mutate(ks = ks_stat(return, cpi),
         cm = cvm_stat(return, cpi),
         k = kuiper_stat(return, cpi)) %>%
  ungroup %>%
  mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
         decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
         decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))

size_ks <- size_dt %>%
  group_by(decile1) %>%
  summarise(size = abs(mean(return))*1000)

size_ks_t <- size_dt %>%
  group_by(decile1) %>%
  group_split()
```

```
    dif2_ks <- round(abs(mean(size_ks_t[[10]]$return)*1000 -
mean(size_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_ks <- paste0("(", round(abs(t.test(size_ks_t[[1]]$return,
size_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR

    size_cm <- size_dt %>%
      group_by(decile2) %>%
      summarise(size = abs(mean(return))*1000)

    size_cm_t <- size_dt %>%
      group_by(decile2) %>%
      group_split()

    dif2_cm <- round(abs(mean(size_cm_t[[10]]$return)*1000 -
mean(size_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_cm <- paste0("(", round(abs(t.test(size_cm_t[[1]]$return,
size_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR



    size_k <- size_dt %>%
      group_by(decile3) %>%
      summarise(size = abs(mean(return))*1000)

    size_k_t <- size_dt %>%
      group_by(decile3) %>%
      group_split()

    dif2_k <- round(abs(mean(size_k_t[[10]]$return)*1000 -
mean(size_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_k <- paste0("(", round(abs(t.test(size_k_t[[1]]$return,
size_k_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR
```

```
    size <- data.frame(decile = as.character(size_ks$decile1),
                       ks = as.character(round(size_ks$size, 2)),
                       cm = as.character(round(size_cm$size, 2)),
                       k = as.character(round(size_k$size, 2))) %>%
      as_tibble() %>%
      tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
      tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

    rm(size_ks, size_cm, size_k, size_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
size_cm_t, size_ks_t, size_k_t)




    ## BOOK VALUE-OUT SHARES-FISCAL - BM
    bm_dt <- chars %>%
      filter(char == "BOOK VALUE-OUT SHARES-FISCAL") %>%
      mutate(value = as.numeric(value)) %>%
      filter(!is.na(value)) %>%
      left_join(select(data, year, Code, return, month, cpi), by = c("year", "Code")) %>%
      ungroup %>%
      filter(!is.na(cpi)) %>%
      mutate(decile = cut(x = value, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
      group_by(decile, month, stock) %>%
      mutate(ks = ks_stat(return, cpi),
             cm = cvm_stat(return, cpi),
```

```
        k = kuiper_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
        decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
        decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))

  bm_ks <- bm_dt %>%
    group_by(decile1) %>%
    summarise(bm = abs(mean(return))*1000)

  bm_ks_t <- bm_dt %>%
    group_by(decile1) %>%
    group_split()

  dif2_ks <- round(abs(mean(bm_ks_t[[10]]$return)*1000) -
abs(mean(bm_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
  tt2_ks <- paste0("(", round(abs(t.test(bm_ks_t[[1]]$return,
bm_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR

  bm_cm <- bm_dt %>%
    group_by(decile2) %>%
    summarise(bm = abs(mean(return))*1000)

  bm_cm_t <- bm_dt %>%
    group_by(decile2) %>%
    group_split()

  dif2_cm <- round(abs(mean(bm_cm_t[[10]]$return)*1000) -
abs(mean(bm_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
```

```
    tt2_cm <- paste0("(", round(abs(t.test(bm_cm_t[[1]]$return,
bm_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    bm_k <- bm_dt %>%
      group_by(decile3) %>%
      summarise(bm = abs(mean(return))*1000)

    bm_k_t <- bm_dt %>%
      group_by(decile3) %>%
      group_split()

    dif2_k <- round(abs(mean(bm_k_t[[10]]$return)*1000) -
abs(mean(bm_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_k <- paste0("(", round(abs(t.test(bm_k_t[[1]]$return, bm_k_t[[10]]$return)$statistic),
2), ")***")   ## B-S decile t.test - AR


    bm <- data.frame(decile = as.character(bm_ks$decile1),
                     ks = as.character(round(bm_ks$bm, 2)),
                     cm = as.character(round(bm_cm$bm, 2)),
                     k = as.character(round(bm_k$bm, 2))) %>%
      as_tibble() %>%
      tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
      tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

    rm(bm_ks, bm_cm, bm_k, bm_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
bm_cm_t, bm_ks_t, bm_k_t)


    ## Momentum - MOM
```

```
    mom_dt <- data %>%
      group_by(Code, month, exchange) %>%
      mutate(month = as.numeric(month)) %>%
      filter(!is.na(cpi)) %>%
      mutate(mom = cumsum(return)) %>%
      mutate(decile = cut(x = mom, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
      ungroup %>%
      distinct(year, month, day, stock, return, .keep_all = TRUE) %>%
      group_by(decile, stock, month) %>%
      mutate(ks = ks_stat(return, cpi),
             cm = cvm_stat(return, cpi),
             k = kuiper_stat(return, cpi)) %>%
      ungroup() %>%
      mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
             decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
             decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))



    mom_ks <- mom_dt %>%
      group_by(decile1) %>%
      summarise(mom = abs(mean(return))*1000)

    mom_ks_t <- mom_dt %>%
      group_by(decile1) %>%
      group_split()

    dif2_ks <- round(abs(mean(mom_ks_t[[10]]$return)*1000) -
abs(mean(mom_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
```

```
    tt2_ks <- paste0("(", round(abs(t.test(mom_ks_t[[1]]$return,
mom_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR

    mom_cm <- mom_dt %>%
      group_by(decile2) %>%
      summarise(mom = abs(mean(return))*1000)

    mom_cm_t <- mom_dt %>%
      group_by(decile2) %>%
      group_split()

    dif2_cm <- round(abs(mean(mom_cm_t[[10]]$return)*1000) -
abs(mean(mom_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_cm <- paste0("(", round(abs(t.test(mom_cm_t[[1]]$return,
mom_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    mom_k <- mom_dt %>%
      group_by(decile3) %>%
      summarise(mom = abs(mean(return))*1000)

    mom_k_t <- mom_dt %>%
      group_by(decile3) %>%
      group_split()

    dif2_k <- round(abs(mean(mom_k_t[[10]]$return)*1000) -
abs(mean(mom_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_k <- paste0("(", round(abs(t.test(mom_k_t[[1]]$return,
mom_k_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    mom <- data.frame(decile = as.character(mom_ks$decile1),
```

```
                ks = as.character(round(mom_ks$mom, 2)),
                cm = as.character(round(mom_cm$mom, 2)),
                k = as.character(round(mom_k$mom, 2))) %>%
      as_tibble() %>%
      tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
      tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

    rm(mom_ks, mom_cm, mom_k, mom_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
mom_cm_t, mom_ks_t, mom_k_t)



    ## BETA coefficient
    beta_dt <- data %>%
      group_by(stock, month) %>%
      filter(!is.na(cpi)) %>%
      mutate(beta = lm(return ~ cpi)$coefficients[2]) %>%
      ungroup %>%
      mutate(decile = cut(x = beta, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
      group_by(decile) %>%
      ungroup() %>%
      group_by(decile, stock, month) %>%
      mutate(ks = ks_stat(return, cpi),
             cm = cvm_stat(return, cpi),
             k = kuiper_stat(return, cpi)) %>%
      ungroup() %>%
      mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
             decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
```

```
        decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))


    beta_ks <- beta_dt %>%
      group_by(decile1) %>%
      summarise(beta = abs(mean(return))*1000)

    beta_ks_t <- beta_dt %>%
      group_by(decile1) %>%
      group_split()

    dif2_ks <- round(abs(mean(beta_ks_t[[10]]$return)*1000) -
abs(mean(beta_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_ks <- paste0("(", round(abs(t.test(beta_ks_t[[1]]$return,
beta_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR

    beta_cm <- beta_dt %>%
      group_by(decile2) %>%
      summarise(beta = abs(mean(return))*1000)

    beta_cm_t <- beta_dt %>%
      group_by(decile2) %>%
      group_split()

    dif2_cm <- round(abs(mean(beta_cm_t[[10]]$return)*1000) -
abs(mean(beta_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_cm <- paste0("(", round(abs(t.test(beta_cm_t[[1]]$return,
beta_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    beta_k <- beta_dt %>%
```

```
    group_by(decile3) %>%
    summarise(beta = abs(mean(return))*1000)

  beta_k_t <- beta_dt %>%
    group_by(decile3) %>%
    group_split()

  dif2_k <- round(abs(mean(beta_k_t[[10]]$return)*1000) -
abs(mean(beta_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
  tt2_k <- paste0("(", round(abs(t.test(beta_k_t[[1]]$return,
beta_k_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


  beta <- data.frame(decile = as.character(beta_ks$decile1),
                     ks = as.character(round(beta_ks$beta, 2)),
                     cm = as.character(round(beta_cm$beta, 2)),
                     k = as.character(round(beta_k$beta, 2))) %>%
    as_tibble() %>%
    tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
    tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

  rm(beta_ks, beta_cm, beta_k, beta_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
beta_cm_t, beta_ks_t, beta_k_t)


  ## short-term reversal - REV

  rev_dt <- data %>%
    mutate(month = as.numeric(month)-1) %>%
    filter(!is.na(cpi) & month != 0) %>%
    distinct(year, month, day, stock, return, .keep_all = TRUE) %>%
```

```
    group_by(stock, month) %>%
    mutate(MonthlyReturn = last(return) / first(return) - 1) %>%
    ungroup %>%
    mutate(decile = cut(x = MonthlyReturn, breaks = 10, labels = c("S", "2", "3", "4", "5", "6",
"7", "8", "9", "B"))) %>%
    group_by(decile) %>%
    ungroup() %>%
    group_by(decile, stock, month) %>%
    mutate(ks = ks_stat(return, cpi),
           cm = cvm_stat(return, cpi),
           k = kuiper_stat(return, cpi)) %>%
    ungroup() %>%
    mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
           decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
           decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))



  rev_ks <- rev_dt %>%
    group_by(decile1) %>%
    summarise(rev = abs(mean(return))*1000)

  rev_ks_t <- rev_dt %>%
    group_by(decile1) %>%
    group_split()

  dif2_ks <- round(abs(mean(rev_ks_t[[10]]$return)*1000) -
abs(mean(rev_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
  tt2_ks <- paste0("(", round(abs(t.test(rev_ks_t[[1]]$return,
rev_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR
```

```
    rev_cm <- rev_dt %>%
      group_by(decile2) %>%
      summarise(rev = abs(mean(return))*1000)

    rev_cm_t <- rev_dt %>%
      group_by(decile2) %>%
      group_split()

    dif2_cm <- round(abs(mean(rev_cm_t[[10]]$return)*1000) -
abs(mean(rev_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_cm <- paste0("(", round(abs(t.test(rev_cm_t[[1]]$return,
rev_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    rev_k <- rev_dt %>%
      group_by(decile3) %>%
      summarise(rev = abs(mean(return))*1000)

    rev_k_t <- rev_dt %>%
      group_by(decile3) %>%
      group_split()

    dif2_k <- round(abs(mean(rev_k_t[[10]]$return)*1000) -
abs(mean(rev_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_k <- paste0("(", round(abs(t.test(rev_k_t[[1]]$return, rev_k_t[[10]]$return)$statistic),
2), ")***")   ## B-S decile t.test - AR


    rev <- data.frame(decile = as.character(rev_ks$decile1),
                      ks = as.character(round(rev_ks$rev, 2)),
                      cm = as.character(round(rev_cm$rev, 2)),
```

```
                k = as.character(round(rev_k$rev, 2))) %>%
    as_tibble() %>%
    tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
    tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

  rm(rev_ks, rev_cm, rev_k, rev_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
rev_cm_t, rev_ks_t, rev_k_t)



  ## stock illiquidity (ILLIQ)
  illiq_dt <-  chars %>%
    filter(char == "TRADING VOLUME") %>%
    mutate(value = as.numeric(value)) %>%
    filter(!is.na(value)) %>%
    left_join(select(data, year, Code, return, month, cpi), by = c("year", "Code")) %>%
    ungroup %>%
    filter(!is.na(cpi)) %>%
    group_by(stock, month) %>%
    mutate(MonthlyReturn = last(return) / first(return) - 1) %>%
    mutate(value = abs(MonthlyReturn)/value,
        decile = cut(x = value, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile, month, stock) %>%
    mutate(ks = ks_stat(return, cpi),
        cm = cvm_stat(return, cpi),
        k = kuiper_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile1 = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B")),
        decile2 = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")),
```

```
        decile3 = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B")))


    illiq_ks <- illiq_dt %>%
      group_by(decile1) %>%
      summarise(illiq = abs(mean(return))*1000)

    illiq_ks_t <- illiq_dt %>%
      group_by(decile1) %>%
      group_split()

    dif2_ks <- round(abs(mean(illiq_ks_t[[10]]$return)*1000) -
abs(mean(illiq_ks_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_ks <- paste0("(", round(abs(t.test(illiq_ks_t[[1]]$return,
illiq_ks_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR

    illiq_cm <- illiq_dt %>%
      group_by(decile2) %>%
      summarise(illiq = abs(mean(return))*1000)

    illiq_cm_t <- illiq_dt %>%
      group_by(decile2) %>%
      group_split()

    dif2_cm <- round(abs(mean(illiq_cm_t[[10]]$return)*1000) -
abs(mean(illiq_cm_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
    tt2_cm <- paste0("(", round(abs(t.test(illiq_cm_t[[1]]$return,
illiq_cm_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


    illiq_k <- illiq_dt %>%
```

```
    group_by(decile3) %>%
    summarise(illiq = abs(mean(return))*1000)

  illiq_k_t <- illiq_dt %>%
    group_by(decile3) %>%
    group_split()

  dif2_k <- round(abs(mean(illiq_k_t[[10]]$return)*1000) -
abs(mean(illiq_k_t[[1]]$return)*1000), 2) ## B-S decile difference - AR
  tt2_k <- paste0("(", round(abs(t.test(illiq_k_t[[1]]$return,
illiq_k_t[[10]]$return)$statistic), 2), ")***")   ## B-S decile t.test - AR


  illiq <- data.frame(decile = as.character(illiq_ks$decile1),
                      ks = as.character(round(illiq_ks$illiq, 2)),
                      cm = as.character(round(illiq_cm$illiq, 2)),
                      k = as.character(round(illiq_k$illiq, 2))) %>%
    as_tibble() %>%
    tibble::add_row(decile = "B-S", ks = as.character(dif2_ks), cm = as.character(dif2_cm), k
= as.character(dif2_k)) %>%
    tibble::add_row(decile = "t(B-S)", ks = tt2_ks, cm = tt2_cm, k = tt2_k)

  rm(illiq_ks, illiq_cm, illiq_k, illiq_dt, dif2_ks, tt2_ks, dif2_cm, tt2_cm, dif2_k, tt2_k,
illiq_cm_t, illiq_ks_t, illiq_k_t)


  ks <- data.frame(Decile = beta$decile,
                   BETA = beta$ks,
                   SIZE = size$ks,
                   BM = bm$ks,
                   MOM = mom$ks,
                   REV = rev$ks,
```

```
            ILLIQ = illiq$ks)

cm <- data.frame(Decile = beta$decile,
            BETA = beta$cm,
            SIZE = size$cm,
            BM = bm$cm,
            MOM = mom$cm,
            REV = rev$cm,
            ILLIQ = illiq$cm)

k <- data.frame(Decile = beta$decile,
            BETA = beta$k,
            SIZE = size$k,
            BM = bm$k,
            MOM = mom$k,
            REV = rev$k,
            ILLIQ = illiq$k)



rm(beta, size, bm, mom, rev, illiq, chars, data)
```

Table 3

```
chars <- chars %>%
  tidyr::spread("char", "value") %>%
```

```
    select(year, Code, `BOOK VALUE-OUT SHARES-FISCAL`, `MARKET CAPITALIZATION`)


  data %>%
    ungroup() %>%
    distinct(year, month, day, Code, return, cpi) %>%
    filter(!is.na(cpi)) %>%
    left_join(chars, by = c("year", "Code")) -> tb3



  tb3 %>%
    group_by(Code, month) %>%
    mutate(ks = ks_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) -> three_f_ks

  three_f_ks <- three_f_ks %>%
    tibble::add_row(decile = "B-S", alpha = three_f_ks$alpha[10]-three_f_ks$alpha[1],
adj_rsq = NA_integer_)


  tb3 %>%
    group_by(Code, month) %>%
    mutate(cm = cvm_stat(return, cpi)) %>%
    ungroup %>%
```

```
    mutate(decile = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) -> three_f_cm

  three_f_cm <- three_f_cm %>%
    tibble::add_row(decile = "B-S", alpha = three_f_cm$alpha[10]-three_f_cm$alpha[1],
adj_rsq = NA_integer_)



  tb3 %>%
    group_by(Code, month) %>%
    mutate(k = kuiper_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi +  as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) -> three_f_k

  three_f_k <- three_f_k %>%
    tibble::add_row(decile = "B-S", alpha = three_f_k$alpha[10]-three_f_k$alpha[1], adj_rsq
= NA_integer_)



  tb3 %>%
```

```
    group_by(Code, month) %>%
    mutate(ks = ks_stat(return, cpi)) %>%
    mutate(mom = cumsum(return)) %>%
    ungroup %>%
    mutate(decile = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi + mom + as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi + mom + as.numeric(`MARKET
CAPITALIZATION`) + as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) ->
four_f_ks

  four_f_ks <- four_f_ks %>%
    tibble::add_row(decile = "B-S", alpha = four_f_ks$alpha[10]-four_f_ks$alpha[1], adj_rsq
= NA_integer_)



  tb3 %>%
    group_by(Code, month) %>%
    mutate(cm = cvm_stat(return, cpi)) %>%
    mutate(mom = cumsum(return)) %>%
    ungroup %>%
    mutate(decile = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi + mom + as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi + mom + as.numeric(`MARKET
CAPITALIZATION`) + as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) ->
four_f_cm
```

```
four_f_cm <- four_f_cm %>%
  tibble::add_row(decile = "B-S", alpha = four_f_cm$alpha[10]-four_f_cm$alpha[1],
adj_rsq = NA_integer_)


tb3 %>%
  group_by(Code, month) %>%
  mutate(k = kuiper_stat(return, cpi)) %>%
  mutate(mom = cumsum(return)) %>%
  ungroup %>%
  mutate(decile = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
  group_by(decile) %>%
  summarise(alpha = lm(return ~ cpi + mom + as.numeric(`MARKET CAPITALIZATION`) +
as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`))$coefficients[2],
            adj_rsq = summary(lm(return ~ cpi + mom + as.numeric(`MARKET
CAPITALIZATION`) + as.numeric(`BOOK VALUE-OUT SHARES-FISCAL`)))$adj.r.squared) ->
four_f_k

four_f_k <- four_f_k %>%
  tibble::add_row(decile = "B-S", alpha = four_f_k$alpha[10]-four_f_k$alpha[1], adj_rsq =
NA_integer_)


data %>%
  group_by(stock, month) %>%
  filter(!is.na(cpi)) %>%
  mutate(ks = ks_stat(return, cpi)) %>%
```

```
    ungroup %>%
    mutate(decile = cut(x = ks, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi)$coefficients[2],
          adj_rsq = summary(lm(return ~ cpi))$adj.r.squared) -> capm_f_ks

  capm_f_ks <- capm_f_ks %>%
    tibble::add_row(decile = "B-S", alpha = capm_f_ks$alpha[10]-capm_f_ks$alpha[1],
adj_rsq = NA_integer_)



  data %>%
    group_by(stock, month) %>%
    filter(!is.na(cpi)) %>%
    mutate(cm = cvm_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile = cut(x = cm, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8",
"9", "B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi)$coefficients[2],
          adj_rsq = summary(lm(return ~ cpi))$adj.r.squared) -> capm_f_cm

  capm_f_cm <- capm_f_cm %>%
    tibble::add_row(decile = "B-S", alpha = capm_f_cm$alpha[10]-capm_f_cm$alpha[1],
adj_rsq = NA_integer_)



  data %>%
    group_by(stock, month) %>%
    filter(!is.na(cpi)) %>%
```

```
    mutate(k = kuiper_stat(return, cpi)) %>%
    ungroup %>%
    mutate(decile = cut(x = k, breaks = 10, labels = c("S", "2", "3", "4", "5", "6", "7", "8", "9",
"B"))) %>%
    group_by(decile) %>%
    summarise(alpha = lm(return ~ cpi)$coefficients[2],
              adj_rsq = summary(lm(return ~ cpi))$adj.r.squared) -> capm_f_k

  capm_f_k <- capm_f_k %>%
    tibble::add_row(decile = "B-S", alpha = capm_f_k$alpha[10]-capm_f_k$alpha[1], adj_rsq
= NA_integer_)




  panel1_ks <- list(CAPM = capm_f_ks,
                    `Fama-French 3 Factor` = three_f_ks,
                    `Carhart 4 Factor` = four_f_ks)

  panel2_cm <- list(CAPM = capm_f_cm,
                    `Fama-French 3 Factor` = three_f_cm,
                    `Carhart 4 Factor` = four_f_cm)

  panel3_k <- list(CAPM = capm_f_k,
                   `Fama-French 3 Factor` = three_f_k,
                   `Carhart 4 Factor` = four_f_k)



  rm(list=setdiff(ls(), c("panel1_ks", "panel2_cm", "panel3_k")))
```