UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"

Bachelor's Degree in Digital Economics and Business

BOUNDED RATIONALITY IN FIXED-INCOME MARKETS: A BEHAVIORAL ANALYSIS OF CROWDLENDING

Supervisor:
Prof. Francesco James Mazzocchini

Thesis candidate:
Davide Ciarcelluti

Academic Year 2025/2026

INDEX

INTRODUCTION

Economic thought and literature have always been very tied to the assumption of rationality. A relatively recent current of thought, behavioral finance, was born after doubting this assumption. Many times over the centuries, people thought to be rational and many others they discouraged this convinction.

The philosophical current of Logical Positivism stated that everything could be reduced to logical or empirical testing. Later philosophy built on the negation of this current. Physics itself moved from the deterministic paradigm of classical physics, to the much more limited perspective of quantum mechanics, where Eisenberg's Uncertainty principle definitely stated that it is intrinsically not possible to know everything about something to its deepest essence, because human intervention alters the nature of things itself. Many other fields of human knowledge accepted limitedness of human rationality and human intervention.

I think the bounded rationality paradigm is actually about this. Not much about nihilistically surrender to irrationality, on the contrary about accepting the coexistence of irrationality with rationality in human dynamics, which is everything economics is about.

The first chapter concerns the shift (still occurring) to the bounded rationality paradigm, the evidence of it found in financial market and the critics to the full rationality paradigm made from different perspectives.

The second chapter narrows the scope to fixed-income markets, which are explained in their structure and then put to the test.

The third and last chapter provides an econometric analysis of crowlending, an emerging market which is investigated under the lens of behavioral finance.

CHAPTER 1: WHAT IS BEHAVIORAL FINANCE

1. RATIONAL AGENTS

From the first outright economic scientist Adam Smith, and his Wealth of Nations, economic decision-making has been treated with a fundamental assumption: individuals and, more generally, economic agents, are rational. The degree of rationality assumed changed over the years and, predominantly, decreased, especially moving from a prescriptive to a descriptive perspective.

The model of *homo economicus* emerged as a pillar of the economics temple. As outlined in Bee & Desmarais-Trembley (2023), it found its foundations in the work of Vilfredo Pareto, who modelled the individual as an abstract one-dimensional subject that acts in a self-referential perspective, for whom the only principle considered while choosing is the utility associated with the choice (*ophelimity*). Pareto thought of *homo economicus* as of a *choice machine*, that will always choose the same thing in a certain circumstance, with the only purpose of maximizing pleasure and minimizing pain. This anthropological set of assumptions evolved and some of the assumption were relaxed, but it stayed the basis for the many economic theories that came after.

Capital markets are not exempt from *homo economicus* influence. An important clarification is that, when talking about investments, not exactly every single investor is supposed to be rational. The main assumptions regard marginal

investors, i.e. those that actively bid up and down prices in the direction of equilibrium, as opposed to the investors that instead have a secondary or passive role in the economy. Marginal investors are responsible for the reaction of the markets to new information and, more generally, to all information, while passive investors can be neglected, for they do not contribute to prices determination or at least they do not contribute to a quantitively significant extent.

From Muth (1961) we know that "[…] the economy generally does not waste information, and that expectations depend specifically on the structure of the entire system". This means that rational behavior of investors is based on the theory (i.e. economic models) that describe market dynamics and on the answer, through the predictive tools made available by the theory, of the formers that seek to profit from new information. Muth (1961) further argues that imperfections and biases in prediction are negligible as long as deviations from the rational forecasts are not correlated with each other, which appears to be the case. From this, we can understand how the assumption of rational expectations, and therefore of rational choices, keeps into consideration the possibility of deviations from rationality, but considers it neglectable, for expectations show to align with economic models.

A consequence of rational expectations is the Efficient Market Hypothesis (EMH). In 1970 Eugene Fama published his famous work on the matter, in whose introduction he states: "[…] investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at

any time "fully reflect" all available information. A market in which prices always "fully reflect" available information is called "efficient"". In Fama (1970), three versions of EMH are tested empirically, depending on the "relevant information" subset considered: in *weak form*, information set consists of historical prices; in *semi-strong form*, information set comprehends also other publicly available information (e.g. stock splits, dividend announcements); in *strong form*, the set includes even insider information. In the study, *strong form* is purely treated as a benchmark to test the two other forms (because previously proven to be null), while both *weak form* and *semi-strong form* appear to be robustly proven by evidence from data analysis. It has to be noted, however, that the categories of public information used in the mentioned testing of *semi-strong form* are rather few. Some correlation between expected returns is shown, but only in intervals that are smaller than one day. Given the results, it can be inferred that financial markets can be approximated by a random walk model, i.e. the prices of different days or longer periods are statistically independent. Since changes are always due to new information, which stems from events that are not predictable, it can be stated that price changes are random, and that this is a direct consequence of agent's rationality.

2. BOUNDED RATIONALITY

2.1. First steps towards bounded rationality

In Garbuglia (2024) it is narrated how, between the 1940s and the 1950s, a small niche of researchers started to doubt the paradigm of *perfect rationality*. Among the first pioneers of the behavioral approach we find George Katona (1901 – 1981), who reconsidered the importance of psychological variables for a better comprehension of individuals' choices. He stressed the difference between objective economic reality and how people actually perceive the latter. Being economics typically focused on the *outcomes* of decisions and psychology in the beneath *processes*, Katona was convinced that these two disciplines could have benefitted of each other's methodologies. In the construction of the *Index of Consumer Sentiment*, with the purpose of anticipating economic cycles, he personally put together psychological measurements with macroeconomic and statistical data.

Somewhat later Herbert Alexander Simon (1916 – 2001), considered the father of decision science, proposed the first models of *bounded rationality*. In these models it was shown that individuals are not able to control all the variables of the decision-making process, due to unpredictability and randomness of events, limited information, time constraints and varying cognitive capabilities. Rational maximization of utility was discarded as a decision criterion in favor of

"satisficing", namely the tendency of individuals to choose something that is good enough and not necessarily the best. This process was thought of as the result of *heuristics*, i.e. decisional patterns that individuals develop based on experience and intuition, used to arrive to a fast result, whose conclusions are very often less precise than those of optimization analysis. For instance, a general consumer who goes grocery shopping will not sit at the table performing calculations to decide which goods to buy and where to buy them, but will instead choose the place and the goods based mostly on the impression he has from previous purchases. This means that decisions are much more the result of *perception*, or a of simplified version of things, than they are of *reality*.

More recently, studies have also taken a closer look at the *motivations* behind decisions, showing that the *self-interest* theorized by Smith is valid in some contexts, while in others it is completely overlooked or even contradicted by behavior. Instances like blood donation show how remuneration, that we can interpret as a proxy for *self-interest*, can be a disincentive, seeing how people who are offered money are less prone to donate.

Even if *processes* underlying decision-making in investment have long been ignored, as they fall out of the scope of economic research, it has to be said that the possibility of behaviors that are not fully rational does have been considered. Nonetheless, its importance has been deemed neglectable and the degree of

irrationality has been underestimated. However, the contamination of psychology, and studies born from this contamination, told a whole different story.

2.2. The search for proof in financial markets

In the previous section we concluded that changes in prices can only be due to new information, to which agents in the market react decreasing or increasing the price with the aim of profiting from predicted movements. However, in Shiller (1981), contradictions to this theory emerge. The inequality that sums up market efficiency can be stated as $\sigma(p) \leq \sigma(p*)$, where p is the actual price found in the market and p* is the *ex post* price calculated from data, based on the actual value of fundamentals; both are detrended to be plotted and confronted independently of long-run growth. If expectations are rational as Muth (1961) suggests and markets are efficient, then $p = E(p*)$. But Shiller (1981)'s results show that measures of stock price volatility are too high to be attributed to new information, for it is found that $\sigma(p) > \sigma(p*)$, i.e. price is much more volatile than what it should predict, therefore price movements don't follow publicly available information. Subsequently, due to the excess volatility shown, the hypothesis that prices only react to new information appears to be wrong, to an extent that allows to comfortably exclude errors of any kind. This paves the way to the hypothesis of other factors that may be the cause of such fluctuations. Figure 1 displays, in order,

the analysis on S&P500 and Dow Jones historical data, which lead us to the same conclusion:

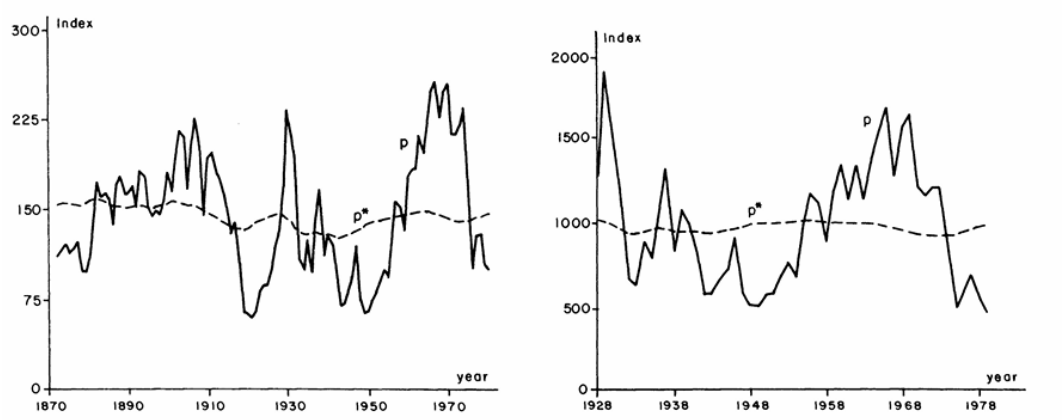

Figure 1. S&P500 and Dow Jones fundamentals vs prices

Another heavy critique to Efficient Market Hypothesis is represented by Thaler and De Bondt (1985). The underlying hypothesis of this work is that "In revising their beliefs, individuals tend to overweight recent information and underweight prior (or base rate) data", in contradiction to Bayes' rule. The phenomenon under analysis in the study is the so called price overshooting, namely an excessive price reaction to unexpected or dramatic events. If markets overreact, leading to overpriced and underpriced securities, the prices will likely tend to go back to the actual values due to mean reversion. Two fundamental hypotheses are stated in the paper: "(1) Extreme movements in stock prices will be followed by subsequent price movements in the opposite direction. (2) The more extreme the initial price movement, the greater will be the subsequent adjustment". Both hypotheses

contradict *weak-form* market efficiency, for they imply that market behavior can, to some extent, be predicted. Formally:

$$E\left(\tilde{u}_{jt} \mid F_{t-1}\right) \neq 0$$

that is, the expected value of unexpected return is greater or lesser than zero.

What is very interesting is how these hypotheses were tested. The first step was the creation of two different portfolios, on the base of cumulative past returns: *winner* portfolios, with highest cumulative past returns, and *loser* portfolios, with lowest ones, were created, considering calculations based on periods of different length. After portfolio formation, the performance of *winner* and *loser* portfolios, in terms of cumulative returns, was evaluated over holding periods of different lengths, using market performance as a benchmark. Results were found to be generally consistent with hypotheses across combinations of length (i.e. of portfolio construction periods and performance periods). In particular, *loser* portfolio generally outperformed both the *winner* portfolio and the market, while *winner* portfolio underperformed. Figure 2 displays the aforementioned results.

Looking at the graph, it is almost trivial to notice that reversal is asymmetric. From this it can be hypothesized that correction of *loser* portfolio is stronger than *winner* portfolio, i.e. *winner* tends to remain overvalued for longer.

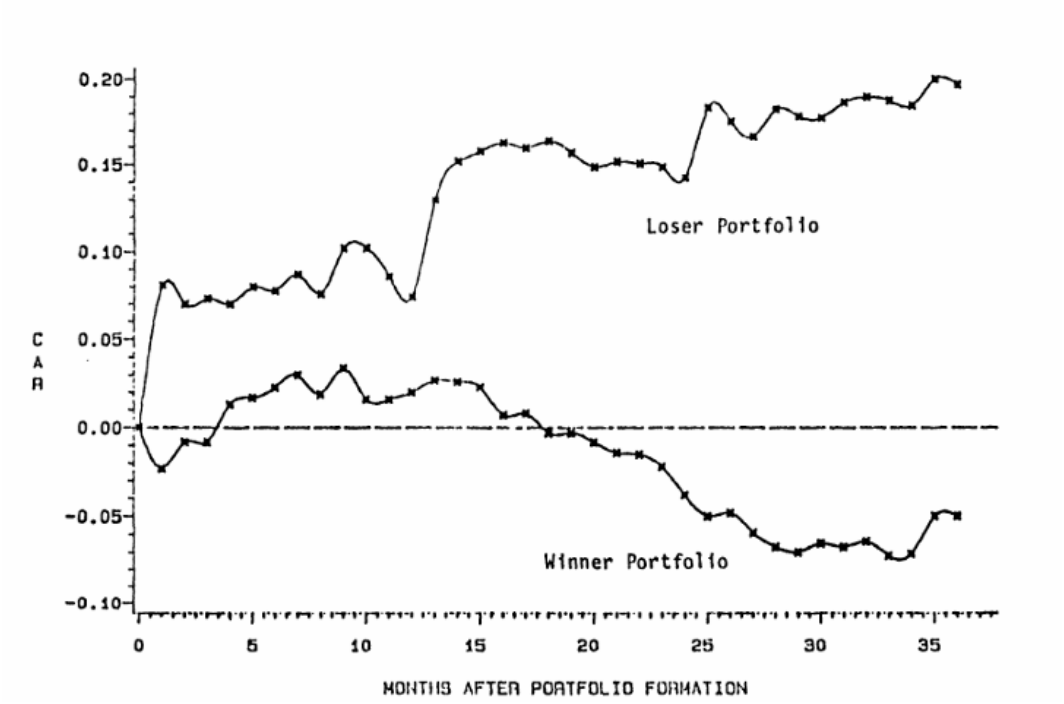

Figure 2. Portfolios' reversals over months

Therefore, from the findings of Thaler and De Bondt (1985), it can be stated that even *weak-form* Efficient Market Hypothesis fails, for evidence suggests that past data can contain information about future market behavior, and hence not all information is already contained in price.

2.3. Feedback theories and sunspots

Another source of market inefficiencies is shown in feedback models. As prices affect investor behavior and investor behavior affects prices, the loop thus created

may set the stage for market inefficiencies, whenever investors bid up the price based only on the expectations of further increases.

A review of the theories around this matter is proposed by Shiller (2003). The idea behind price-to-price feedback theory is exposed: “When speculative prices go up, creating successes for some investors, this may attract public attention, promote word-of-mouth enthusiasm, and heighten expectations for further price increases. The talk attracts attention to "new era" theories and "popular models" that justify the price increases. This process in turn increases investor demand and thus generates another round of price increases. If the feedback is not interrupted, it may produce after many rounds a speculative "bubble," in which high expectations for further price increases support very high current prices. The high prices are ultimately not sustainable, since they are high only because of expectations of further price increases, and so the bubble eventually bursts, and prices come falling down”. In the end, how is it possible that, notwithstanding investors rationality, financial crises happen? How is it possible that financial bubbles keep swelling and then burst?

What appears to be a satisficing explanation is that word-of-mouth and extrinsic factors do matter, and that is because investors can happen to base their expectations on factors that do not directly affect the economy directly but can affect it through self-fulfilling expectations. These factors are called so called *sunspots*.

In Marimon et al. (1993) a brilliant agentic simulation is built to investigate the role of *sunspots* in the formation of expectations. The agents used in the experiment are characterized by bounded rationality and learn from past experiences over the periods, choosing the rule to follow based only on past history and success of previous choices, which we can interpret as an heuristic-like learning. They are put in simulated economies, and have to converge to an equilibrium price for the fiat money to be used as an efficient mean of exchange for transactions. What comes out when subjecting agents to an exogenous shock, in this case a variation of generation size, is that they diverge to cyclical equilibria, in the attempt of adjusting to the change in fundamentals. But what interesting is that, even when the shock is over, the experience-based learning leads the agents to keep diverging to cyclical equilibria, instead of converging to an equilibrium price, without the presence of an actual fundamentals-related motivation to do so. The point to stress here is that the agents did not individually act following this path, but they coordinated and exchanged information to arrive at such result.

We can interpret the generation size shock as a change of the real economy (population growth, trading volumes, etc.), which makes the simulation extremely realistic. *Sunspot*-related expectations are what leads, for instance, to bank runs, when creditors withdraw money without banks failing only because expected to fail or, more importantly for the purpose of our discussion, what leads to financial bubbles where price raises because people buy and expect to resell at a higher price

(without any rational reason to do so). That is how continuous *positive feedback* can lead to bubbles that grow being already doomed to burst.

Shiller (2003) brings to the discussion also the case of Ponzi schemes, that involve "[…] a superficially plausible but unverifiable story about how money is made for investors and the fraudulent creation of high returns for initial investors by giving them the money invested by subsequent investors", which goes on and "[…] as the rounds of high returns generate excitement, the story becomes increasingly believable and enticing to investors". These schemes, of which we have real examples like those that grew large in Albania in 1996-1997, resemble the dynamics that leads to the creation of bubbles.

Finally, Shiller (2003) discusses how even models that typify investors in *smart money* – rational optimizing investors – and *ordinary investors* – irrational and heuristics-driven investors – fail to conclude that *smart money* can tie price to the value of fundamentals and, contrariwise, conclude that often *smart money*'s power is limited, due to intrinsic limitations of instruments, like short-selling, that should counter excessively bullish sentiment.

From this we can assert that irrational investors do have a significant weight in the market and their behavior cannot at all be neglected in economic models, as proven by previously discussed studies that show evidence against efficiency.

3. PROSPECT THEORY: A BEHAVIORAL APPROACH TO UTILITY

3.1. Premises about expected utility theory

An important matter in economics is to model how people decide. Under full rationality the explanatory model was the expected utility theory, mainly formalized by John von Neumann and Oskar Morgenstern in their "Theory of games and economic behavior". The theory states that people do not actually want to maximize their wealth in absolute terms, but rather their utility, according to the following formula:

$$EU = \sum p_i u(x_i)$$

This formula is built over a small set of axiomatic foundations, which are exposed below:

1. Completeness: for any two lotteries A and B:

$$A \succcurlyeq B \; or \; B \succcurlyeq A$$

   This means that for any two options, individuals will either prefer one option to the other or be indifferent between the two. There are no "cannot choose" options.

2. Transitivity: for any lotteries A, B, C:

$$if \; A \succcurlyeq B \; and \; B \succcurlyeq C \Rightarrow A \succcurlyeq C$$

Preferences are logically consistent: if one prefers tea to coffee and coffee to milk, he must also prefer tea to milk. Otherwise, choices become contradictory.

3. Independence: for any lotteries A, B, C and any $\alpha \in (0,1)$:

$$A \succcurlyeq B \Rightarrow \alpha A + (1-\alpha)C \succcurlyeq \alpha B + (1-\alpha)C$$

Preferences should not be affected by the mix with a third common option, for any common option.

4. Continuity: If $A \succeq B \succeq C$, then there exists $\alpha \in (0,1)$ such that:

$$B \sim \alpha A + (1-\alpha)C$$

No "jumps" in preference; there exist intermediate options, created from a mix of the best and worst outcomes. Two options can be found so that they are indifferent.

In short, people are assumed to evaluate their options by equally evaluating different probabilities, with no difference of any kind with respect to how lotteries are exposed or formulated.

3.2. Inconsistencies and paradoxes

However, just three years after the formalization of von Neumann-Morgenstern framework, Friedman and Savage (1948) showed the first contradictions while attempting to refine the utility framework. In the analysis of individuals' behavior, it was found that choice behavior was inconsistent: people, all other things fixed,

were found to spend money both on insurance and of gambling, which follow two opposite principles, namely the first is a payment to avoid a risk, and the second is a payment to seek a risk. Thus, individuals are apparently both *risk-averse* and *risk-seekers*, which appeared to be a paradox. Moreover, this attitude is reversely proportional to income, which is contrary to the greater utility per unit of money that people of low income have, at least theoretically.

Not much later, Allais (1953) shed light on further paradoxes arising from the formulation of the expected utility theory. In particular, it was found that the independence axiom was systematically violated in experimental context, since it was shown that individuals' choices differ even when options are designed to be equivalent in expected utility terms. Rather, their weightings were conditioned on a subjective perception, depending on the extent to which the actual probability was close to certainty, i.e. the closer to certainty, the more disproportionally overweighted the relative option.

3.3. Prospect theory

On the groundwork of the former study and some others, there was space for a new theory to explain more accurately how individuals weight their options. In Kahneman and Tversky (1979) an experiment was designed to further test the independence axiom and understand more deeply its pitfalls. This experiment consists, essentially, of a questionnaire subjected to studies, with measures taken to

account for possible biases due to the order of questions or biases of other kind. The questionnaires consist of couples of prospects between which to choose. "A prospect $(x_1, p_1; \ldots; x_n, p_n)$ is a contract that yields outcome x_i with probability P_i, where $P_1 + P_2 + \cdots P_n = 1$. To simplify notation, we omit null outcomes and use (x, p) to denote the prospect $(x, p; 0, 1 - p)$ that yields x with probability p and 0 with probability $1 - p$".

The first anomaly found is the so called *certainty effect*: "in expected utility theory, the utilities of outcomes are weighted by their probabilities" but the answers show that "[…] people overweight outcomes that are considered certain, relative to outcomes which are merely probable". It is found that reducing the probability of both prospects while keeping them equal in terms of expected utility, the choices are reversed: in the first place, the certain outcome is less convenient in terms of expected utility but chosen anyway, but the opposite happens when a 66% of receiving a sum is removed on both sides.

The second anomaly is here named *possibility effect*: the next questionnaires show how, when faced with two prospects that have very small probability, i.e. they are *possible* rather than *probable*, and are equivalent in terms of expected utility, most people choose the one that offers the larger gain.

A third mechanism that is found is the *reversion effect*, shown in the table 1:

Positive prospects				Negative prospects			
Problem 3:	(4,000, .80)	<	(3,000).	Problem 3′:	(−4,000, .80)	>	(−3,000).
$N = 95$	[20]		[80]*	$N = 95$	[92]*		[8]
Problem 4:	(4,000, .20)	>	(3,000, .25).	Problem 4′:	(−4,000, .20)	<	(−3,000, .25).
$N = 95$	[65]*		[35]	$N = 95$	[42]		[58]
Problem 7:	(3,000, .90)	>	(6,000, .45).	Problem 7′:	(−3,000, .90)	<	(−6,000, .45).
$N = 66$	[86]*		[14]	$N = 66$	[8]		[92]*
Problem 8:	(3,000, .002)	<	(6,000, .001).	Problem 8′:	(−3,000, .002)	>	(−6,000, .001).
$N = 66$	[27]		[73]*	$N = 66$	[70]*		[30]

Table 1. Preferences between positive and negative prospects

What is found here is that when the signs of the outcomes are reversed, preference orders are strongly reversed as well. Moreover, it appears that the attitude of people is *risk-averse* in the positive domain and *risk-seeking* in the negative domain.

A further observed pattern is the *isolation effect*. “In order to simplify the choice between alternatives, people often disregard components that the alternatives share, and focus on the components that distinguish them”, subsequently “[…] a pair of prospects can be decomposed into common and distinctive components in more than one way, and different decompositions sometimes lead to different preferences”.

These four effects violate systematically the independence axiom of expected utility theory. Together with deeper observations from the questionnaires, they are the starting point of a new theory of decision, *prospect theory*. “Prospect theory distinguishes two phases in the choice process: an early phase of editing and a subsequent phase of evaluation.

The editing phase consists of a preliminary analysis of the offered prospects, which often yields a simpler representation of these prospects. In the second phase, the edited prospects are evaluated and the prospect of highest value is chosen". "Editing consists of the application of several operations that transform the outcomes and probabilities associated with the offered prospect", and is composed of four main operations:

- *Coding*. "[…] people normally perceive outcomes as gains and losses, rather than as final states of wealth or welfare. Gains and losses, of course, are defined relative to some neutral reference point". "However, the location of the reference point, and the consequent coding of outcomes as gains or losses, can be affected by the formulation of the offered prospects, and by the expectations of the decision maker."
- *Combination.* "Prospects can sometimes be simplified by combining the probabilities associated with identical outcomes."
- *Segregation.* "Some prospects contain a riskless component that is segregated from the risky component."
- *Cancellation.* "The essence of the isolation effects described earlier is the discarding of components that are shared by the offered prospect."

Editing operations are assumed to be performed whenever possible, since they significantly simplify the decision process. There is not a precise order in which they are performed, and this is why the formulation of a prospect can influence

preferences. Of course, the editing operations can and do give rise to anomalies of preference.

After editing, the decision-maker will evaluate the edited prospects, and assess which one has the higher value. "The overall value of an edited prospect, denoted V, is expressed in terms of two scales, π and v. The first scale, v, associates with each probability p a decision weight $\pi(p)$, which reflects the impact of p on the over-all value of the prospect. However, π is not a probability measure, and it will be shown later that $\pi(p) + \pi(1 - p)$ is typically less than unity. The second scale, v, assigns to each outcome x a number $v(x)$, which reflects the subjective value of that outcome". It has to be stressed that "[…] outcomes are defined relative to a reference point, which serves as the zero point of the value scale. Hence, v measures the value of deviations from that reference point, i.e., gains and losses".

So, mathematically, "if $(x, p;\ y, q)$ is a regular prospect (i.e., either $p + q < 1$, or $x \geq 0 \geq y$, or $x \leq 0 \leq y$)", then the formula that describes how preferences are built will be:

$$V(x, p;\ y, q) = \pi(p)v(x) + \pi(q)v(y)$$

About the new value function that stems from this, some considerations have to be made. "Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes". This means that the same level of wealth can imply poverty or riches, depending on one's current

assets. We know that “many sensory and perceptual dimensions share the property that the psychological response is a concave function of the magnitude of physical change”. Kahneman & Tversky (1979) suggested that this principle applies in the same way to monetary changes. It is hypothesized that “the value function for changes of wealth is normally concave above the reference point ($v"(x) < 0$, for $x > 0$) and often convex below it ($v"(x) > 0$, for $x < 0$). That is, the marginal value of both gains and losses generally decreases with their magnitude.” It is important to emphasize that “A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains. The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount.” Summing up, the value function is generally defined on deviations rather than absolute magnitudes, concave for gains and convex for losses, steeper for losses than for gains, and looks like the function in Figure 3:

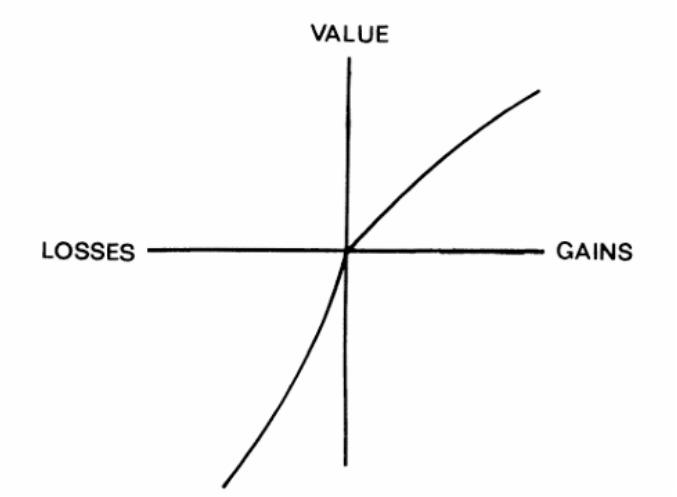

Figure 3. The new value function

In Kahneman & Tversky (1979), the weighting function π is the function that relates decision weights to stated probabilities. "Naturally, π is an increasing function of p, with $\pi(0) = 0$ and $\pi(1) = 1$. That is, outcomes contingent on an impossible event are ignored, and the scale is normalized so that $\pi(p)$ is the ratio of the weight associated with the probability p to the weight associated with the certain event." A further proposal is that very low probabilities are generally overweighted, as a consequence of *possibility effect* (note that overweighting differs from overestimation). Moreover, evidence suggests that that, for all $0 < p < 1$, subcertainty is implied, such that $\pi(p) + \pi(1 - p) < 1$ i.e. "the sum of the weights associated with complementary events is typically less than the weight associated with the certain event".

Finally, it is necessary to point out that since Prospect theory is based on evaluations based on a reference point, changes in this reference point mean alterations of preferences among two prospects. This happens whenever "gains and losses are coded relative to an expectation or aspiration level that differs from the status quo", For instance, an unexpected tax deduction from salary is perceived as a loss, not as a reduction of gain.

3.4. Cumulative prospect theory

Kahneman & Tversky (1992) further extends and formalizes Prospect theory, resulting in a new model called Cumulative prospect theory. Prospect theory

presents two main problems: it does not always satisfy stochastic dominance, and it fails to be consistent when facing multi-outcome lotteries. These problems can be solved by the rank-dependent or cumulative functional.

By functional we mean, in mathematics, a mapping that takes a *function* as an input, e.g. an entire probability distribution, and outputs its *value*. "Instead of transforming each probability separately, this model transforms the entire cumulative distribution function." Cumulative functional is separately applied to gains and losses.

"Let S be a finite set of states of nature; subsets of S are called events. It is assumed that exactly one state obtains, which is unknown to the decision maker. Let X be a set of consequences, also called outcomes. […] We assume that X includes a neutral outcome, denoted 0, and we interpret all other elements of X as gains or losses. An uncertain prospect f is a function from S into X that assigns to each state $s \in S$ a consequence $f(s) = x$ in X. […] A prospect f is then represented as a sequence of pairs (x_i, A_i), which yields x_i if A_i occurs, where $x_i > x_j \; iff \; i > j$." This latter means that events A_i are ranked from worst to best based on the ordering of the outcomes, i.e. events are not assigned a rank *a priori*. Moreover, "The positive part of f, denoted f^+, is obtained by letting $f^+(\mathrm{s}) = f(s)$ if $f(s) > 0$, and $f^+(s) = 0$ if $f(s) \leq 0$. The negative part of f, denoted f^-, is defined similarly." It is defined "[…] a number $V(f)$ such that f is preferred to or indifferent to $g \; iff \; V(f) > V(g)$" in terms of the concept of *capacity*: "A capacity W is a function that assigns to each $A \subseteq S$ a number $W(A)$ satisfying $W(\emptyset) = 0$, $W(S) = 1$, and $W(A) \geq$

$W(B)$whenever $A \supseteq B$". Cumulative prospect theory asserts that the prospect evaluation function has a form like:

$$V(f) = \sum_{i=-m}^{n} (\pi_i v(x_i))$$

considering also that:

$$V(f) = V(f^+) + V(f^-)$$

$$V(f^+) = \sum_{i=0}^{n} \pi_i^+ v(x_i), \quad V(f^-) = \sum_{i=-m}^{0} \pi_i^- v(x_i)$$

where decision weights are defined by:

$\pi_n^+ = W^+(A_n), \ \pi_{-m}^- = W^-(A_{-m}),$

$\pi_i^+ = W^+(A_i \cup \ldots \cup A_n) - W^+(A_i \cup \ldots \cup A_n), \qquad 0 \leq i \leq n - 1,$

$\pi^- = W^-(A_{-m} \cup \ldots \cup A_i) - W^-(A_{-m} \cup \ldots \cup A_{i-1}), \qquad 1 - m \leq i \leq 0$

The decision weight can be interpreted as the marginal contribution of the single event. The first term of the sum is the weight of an outcome at least as good as A_i, while the second term is the weight of an outcome strictly better than A_i. So by subtracting the second term to the first term, we "isolate" the capacity attached to i, where i is the rank of the event. With this formulation the problems of stochastic

dominance violations and failure to cope with multi-outcome prospects are both solved.

As in Kahneman & Tversky (1979) we assume that v is steeper for losses than for gains, concave for positive changes and convex for negative changes. All three reflect the principle of *diminishing sensitivity*, i.e. impact decreases as distance from the reference point decreases. The same principle appears to apply to the weighting function π, which makes it convex near 0 and concave near 1.

Besides the theoretical apparatus, an experiment was carried out to get detailed information about the value function and the weighting function. The experiment consisted of displaying a prospect and its expected value, together with a descending series of seven sure outcomes between the extremes of the former prospect. The subject had to choose between these all options. After this, a new set of seven certain outcomes was shown, spaced between a value 25% higher than lowest accepted amount and a value 25% lower than highest rejected amount. The certainty cash equivalent was estimated as the midpoint of the two choices. Results are shown in Table 2.

For nonmixed prospects (i.e. strictly positive or strictly negative) a fourfold pattern appears: *risk-averse* and *risk-seeking* preferences, respectively, for gains and for losses with high probability ($p > 0.5$); tendency is reversed, even if not perfectly, when they are low ($p < 0.1$). Table 3 shows the relative percentages.

	Probability								
Outcomes	.01	.05	.10	.25	.50	.75	.90	.95	.99
(0, 50)			9		21		37		
(0, −50)			−8		−21		−39		
(0, 100)		14		25	36	52		78	
(0, −100)		−8		−23.5	−42	−63		−84	
(0, 200)	10		20		76		131		188
(0, −200)	−3		−23		−89		−155		−190
(0, 400)	12								377
(0, −400)	−14								−380
(50, 100)			59		71		83		
(−50, −100)			−59		−71		−85		
(50, 150)		64		72.5	86	102		128	
(−50, −150)		−60		−71	−92	−113		−132	
(100, 200)		118		130	141	162		178	
(−100, −200)		−112		−121	−142	−158		−179	

Table 2. Median cash equivalents (in dollars) for all nonmixed prospects

	Gain		Loss	
Subject	$p \leq .1$	$p \geq .5$	$p \leq .1$	$p \geq .5$
Risk seeking	78[a]	10	20	87[a]
Risk neutral	12	2	0	7
Risk averse	10	88[a]	80[a]	6

Table 3. Risk attitudes towards possible gains or losses

Under this theory, assuming $X = \mathbb{R}$ and preference homogeneity, v is represented as a function of the form:

$$v(x) = \begin{cases} x^{\alpha} \text{ if } x \geq 0 \\ -\lambda(-x)^{\beta} \text{ if } x < 0. \end{cases}$$

Let c/x be the ratio of certain cash equivalent to the nonzero outcome. Figure 4 and Figure 5 plot the ratio as a function of p, respectively, for positive and negative prospects:

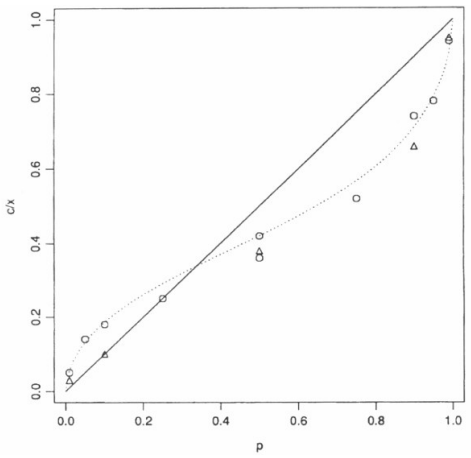

Figure 4. c/x for positive prospects

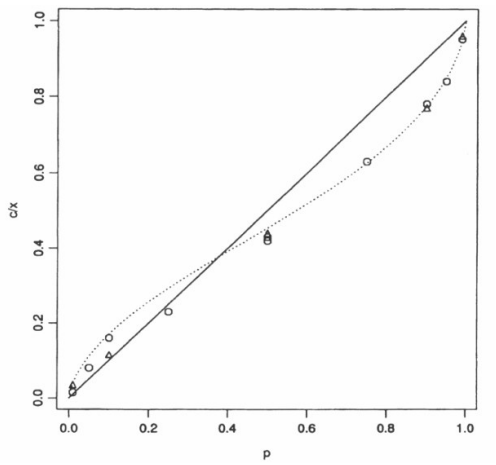

Figure 5. c/x for negative prospects

Note that the straight line represent *risk-neutral* attitude, point above represent *risk-seeking* and points below represent *risk-aversion*. These two curves can be interpreted as weighting functions, fitted using the following functional forms:

$$w^+(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \quad w^-(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}}.$$

"The curvature of the weighting function explains the characteristic reflection pattern of attitudes to risky prospects", which explains much more simply than Friedman and Savage (1948) that overweighting of high probabilities is why lotteries and insurance, even if apparently opposed to each other, are both popular

among people. Respectively, lotteries represent the *risk-seeking* tendency of people in front of gains combined with low probability, while insurance represents the opposite tendency of *risk-aversion* in front of losses that have a low probability.

The Cumulative prospect theory is, in a way, a further argument against the assumption of rational economic agents. It is shown that probabilities are consistently distorted by perception and that *risk-neutral* attitudes are to be considered exceptions rather than descriptive rules.

However, as the authors state, it has to be said that "theories of choice are at best approximate and incomplete. One reason for this pessimistic assessment is that choice is a constructive and contingent process", and that framing effects and editing operations do have a role in how the decision-making process happens. This means that even if Cumulative prospect theory outlines important decisional dynamics and effectively formalizes them in rather simple situations, it leaves much of the explanation of more complex decisions to the heuristics of choice.

4. HEURISTICS AND COGNITIVE BIASES

Time has passed since Sigmund Freud firstly hypothesized that the conscious is not the only thing that moves people, and that there is a powerful, irrational unconscious that drives choices in a completely different way.

As Epstein (1994)'s review of past literature shows, dual-process models of human cognition and behavior have recurred over the years, differing only in the mapping of the types, mechanisms and origins of the two systems from which human thinking stems; some scholars even proposed three systems instead of two. Many of these different models share common characteristics, like the division into a rational system and an emotional/intuitive system and the basic mechanisms by which they interact. Kahneman (2011) calls the two systems System 1 and System 2, and defines them as follows:

- "System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control
- "System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration"

Epstein (1994), on the other hand, calls them respectively Experiential system and Rational system, whose attributes are summarized in Table 4.

The second definition is clearly more thorough and, in a way, comprehensive of the former. We understand that the two systems are responsible for action in different contexts, act in a different way at the expense of different amounts of energy consumption. In theory they are separate, but in practice it happens that System 1 enters into action when System 2 should work alone.

Experiential system	Rational system
1. Holistic	1. Analytic
2. Affective: Pleasure–pain oriented (what feels good)	2. Logical: Reason oriented (what is sensible)
3. Associationistic connections	3. Logical connections
4. Behavior mediated by "vibes" from past experiences	4. Behavior mediated by conscious appraisal of events
5. Encodes reality in concrete images, metaphors, and narratives	5. Encodes reality in abstract symbols, words, and numbers
6. More rapid processing: Oriented toward immediate action	6. Slower processing: Oriented toward delayed action
7. Slower to change: Changes with repetitive or intense experience	7. Changes more rapidly: Changes with speed of thought
8. More crudely differentiated: Broad generalization gradient; stereotypical thinking	8. More highly differentiated
9. More crudely integrated: Dissociative, emotional *complexes; context-specific processing*	9. More highly integrated: Cross-context processing
10. Experienced passively and preconsciously: We are seized by our emotions	10. Experienced actively and consciously: We are in control of our thoughts
11. Self-evidently valid: "Experiencing is believing"	11. Requires justification via logic and evidence

Note. From "Cognitive–Experiential Self-Theory: An Integrative Theory of Personality" by S. Epstein, in R. C. Curtis, *The Relational Self: Theoretical Convergences in Psychoanalysis and Social Psychology*, New York: Guilford Press. Copyright 1991 by Guilford Press. Adapted by permission.

Table 4. Characteristics of experiential systems and rational

Heuristics are defined in Kahneman (2011) as "a simple procedure that helps find adequate, though often imperfect, answers to difficult question". What this means is that the normal state of our brain is made of sensations and intuitive judgements about everything that happens to us. This is why, usually, we hate or love a person before having enough information to judge. Very often, heuristics consists of the entrance into action of substituting questions, i.e. whenever we face a complex question, our System 1 associates the question to a simpler one, to which we actually answer. Table 5 shows a few examples.

More precisely, "The target question is the assessment you intend to produce", while "the heuristic question is the simpler question that you answer instead". The

Target Question	*Heuristic Question*
How much would you contribute to save an endangered species?	How much emotion do I feel when I think of dying dolphins?
How happy are you with your life these days?	What is my mood right now?
How popular is the president right now?	How popular will the president be six months from now?
How should financial advisers who prey on the elderly be punished?	How much anger do I feel when I think of financial predators?
This woman is running for the primary. How far will she go in politics?	Does this woman look like a political winner?

Table 5. How system 1 transforms questions

creation of these questions stems from a mechanism undertaken by System 1, the so called *Mental shotgun*: System 1 performs many computations at any time, related to our primordial instincts, like controlling for possible dangers, building a three-dimensional representation of what we see, and many others; it does this independent of our control or will. After heuristic questions are such computed, there is still the need to adapt heuristic answers to the original questions. System 1 does this through *intensity matching*: if I want to express my feelings about dying dolphins, I have to express it in dollars.

On the other hand, computations of System 2, like counting the syllables of a word, are voluntary. Due to the Mental shotgun, if we want to give a complex and thorough answer, System 1 produces an heuristic answer anyway, and we have to

consciously discard it in favor of a more rigorous computation. Otherwise, like it often happens, System 2 will follow the way of least effort and endorse the heuristic answer.

An example of heuristic is the Affect heuristic: "the dominance of conclusions over arguments is most pronounced where emotions are involved", e.g. political preferences determine the arguments that you find compelling not because they are actually better but because you were already convinced of that certain thing at the start. Opinions can change, at least a little, and when, for instance, risks about something we disliked are found to be smaller, even the benefits appear greater, because of the affection increase due to the first discover.

Another important discovery disclosed in Kahneman (2011) is that of biases, that are "systematic errors" that "recur predictably in particular circumstances". Biases are a consequence of heuristics, whenever these give us a wrong conclusion, and they can be found in specific patterns of our everyday life. The heuristic is the tool, the bias is the error it produces.

An example of this is the *halo effect*. The *halo effect* consists of appreciating (or loathing) everything about a person, including the often many things that have not been observed. It is why the world representation that System 1 build appears way more simple and coherent than it is in reality. For instance, if asked "Is Mindik a good leader? She is smart and strong…" the immediate answer that our System 1 will produce will be "yes", but that would change if we put "corrupted and cruel"

immediately after the question. It may sound trivial, but any time we make such a judgement, we are biased to the extent to which we do not have enough information to make a judgement, though we judge anyways. Notwithstanding this, a spontaneous and immediate judgement is produced, and our final conviction depends on whether or not we decide to discard it.

Another example is the *confirmation bias*. The *confirmation bias* is a consequence of the fact that System 2 answers to a doubt through the research of evidence in favor rather than evidence against, whatever the question to be answered. The two questions "Is Sam outgoing?" and "Is Sam rude?" will trigger different memories related to event that can confirm the one or the other hypothesis (in case we know Sam). Moreover, experiments show that while System 2 is engaged in any activity, tendency to believe what is false increases dramatically.

More examples of biases are exposed in Kahneman (2011), but the description of those goes beyond the scope of this work. What is more interesting, in my opinion, is what hides behind heuristics and biases. An important premise regards the interaction between our brain, especially System 1, and the world: "Whenever you are conscious, and perhaps even when you are not, multiple computations are going on in your brain, which maintain and update current answers to some key questions: Is anything new going on? Is there a threat? Are things going well? Should my attention be redirected? Is more effort needed for this task? […] The assessments are carried out automatically by System 1, and one of their functions is to determine

whether extra effort is required from System 2." To carry out this continuous stream of computations, mental energy is required, and using System 1 is much more energy-efficient than using System 2.

But when is System 2 necessary? Whenever voluntary effort is involved, e.g. running at a hard pace, performing a cognitive demanding activity, repressing emotions, System 2 is activated and all these activities draw from a shared pool of mental energy. Of course, writing a book is more energy-demanding than having a conversation. It has been discovered that a voluntary effort of any kind is tiring, and "if you have had to force yourself to do something, you are less willing or less able to exert self-control when the next challenge comes around." This phenomenon is known as *ego depletion*. Experiments showed that "participants who are instructed to stifle their emotional reaction to an emotionally charged film will later perform poorly on a test of physical stamina". What is even more surprising is that the energy issue has been proven to be more than a metaphor: when System 2 activity is intense, blood glucose levels drop.

Putting everything together, it can be asserted that what allows people to make better, reasoned decisions is how much they require System 2 to work. As I said before, due to the *Mental shotgun*, System 1 continuously provides easy, intuitive answers, that are independent of will. Moreover, our instinct lead us very often to accept those heuristic answers in order to save energy. Finally, subjective confidence about a judgement is not a rational evaluation, but a sensation that

reflects coherence of the information and the related cognitive ease. Therefore, it is up to the individual to make an effort and investigate for an accurate answer whenever an intuitive one is produced, to avoid biases whenever possible and accept the mistakes we are inherently prone to make.

4.1. A closer glance: overconfidence bias

Overconfidence bias is a bias where people have a certain degree of confidence about their judgement that is higher than actual accuracy.

Moore & Healy (2008) analyzed literature about overconfidence and found out that there is not a univocal definition, and rather three main aspects are studied:

- *Overestimation*: "overestimation of one's actual ability, performance, level of control, or chance of success";
- *Overplacement*: "people believe themselves to be better than others";
- *Overprecision*: "excessive certainty regarding the accuracy of one's beliefs".

The underlying dynamics suggested by Moore & Healy (2008) is that "People often have imperfect information about their own performances, abilities, or chance of success. However, they have even worse information about others. As a result, people's estimates of themselves are regressive, and their estimates of others are even more regressive. Consequently, when performance is high, people will underestimate their own performances, underestimate others even more so, and thus

believe that they are better than others. When performance is low, people will overestimate themselves, overestimate others even more so, and thus believe that they are worse than others."

From this model we understand that *overconfidence* is only a face of the medal. On the other face we find *underconfidence*, that mirrors the three aforementioned versions .

This appears to have effects on financial markets. What is found by Grezo (2020) is that overconfidence increases trading volumes and leads to excessive trading. This finding is confirmed by Singh et al. (2024), which also finds hypothesis that it may lead to price bubbles and contribute to market booms, crashes and inefficiencies. Still in Grezo (2020), it is found that overconfident investors overreact to their private signals and underreact to public information.

Abdin et al. (2022) asserts a correlation between *overconfidence* and *self-attribution* bias, i.e. tendency of investors to give credit to themselves for success and to blame others for failure. Grezo (2020) also points out that many studies use proxies of overconfidence, and that this influences the results whenever the correlation between variables and proxies is far from the correlation of the same variables with overconfidence. It also observes that the aspect of overconfidence that is investigated, i.e. *overestimation*, *overplacement* or *overprecision*, is very often not specified and when it is it leads to different conclusions. In particular, *overplacement* is the one that has the strongest effect among direct measures, while

the other two aspects appear to have trivial or insignificant effects. Considering overconfidence in financial markets and more specifically in fixed-income markets, my research question follows: does overconfidence play a role in the size of credit spread and, if it does, which is its magnitude?

CHAPTER 2: THE FIXED-INCOME MARKET AND THE SOURCES OF RISK

1. FIXED-INCOME MARKET

In Bodie (2023), the *fixed-income* market is described as the market where, generally, *debt securities* are traded. Debt securities "are claims on a specified periodic stream of cash flows". Since they "[…] promise either a fixed stream of income or a stream of income determined by a specified formula", they are also called *fixed-income securities*,

These securities come in a considerable variety of maturities and payment provisions. They are typically classified as: *money market* securities, which are short-term, low risk and highly liquid, e.g. U.S. Treasury bills; and *capital market* securities, which are long-term and medium to high risk, e.g. Treasury notes and bonds, corporate bonds, etc.

This work will focus on *bonds*. According to Bodie (2023) "A *bond* is a security that is issued in connection with a borrowing arrangement. The borrower issues (i.e. sells) a bond to the lender for some amount of cash"; the bond is therefore an obligation the borrower has to repay the lender (i.e. the buyer of the bond). "The issuer agrees to make specified payments to the bondholder on specified dates, […] called *coupon payments* […]. When the bond matures, the issuer repays the debt by paying the bond's par value (i.e. its face value). The coupon rate of the bond

determines the interest payment:" the periodical payment is "the coupon rate times the bond's par value. The coupon rate, maturity date, and par value of the bond are part of the bond indenture, which is the contract between the *issuer* and the *bondholder*." In this treatment, the term *bond* will include both money market and capital market bonds. A further classification of the many types of bonds is beyond the scope of this discussion.

2. THE PRICE OF A BOND AND YIELD TO MATURITY

According to Bodie (2023), a bond's cash flows consist of several payments that occur all in the future. Clearly, considering the time value of money, it is reasonable that the price an investor is willing to pay depends on the value of money received in the future compared to the value of money today. The present value depends on market's interest rates, which consider a risk-free real rate plus a premium to compensate for expectations about inflation. Assuming a unique interest rate appropriate for discounting, present value of a bond will be given by the formula:

$$Bond\ value = \sum_{t=1}^{T} \frac{C_t}{(1+r)^t} = \sum_{t=1}^{T} \frac{Coupon}{(1+r)^t} + \frac{Par\ value}{(1+r)^t}$$

Where, referring to the second term, C_t is the cash flow at time t and $1/(1+r)^t$ is the discount factor at time t. Considering the series of level coupon payments as an annuity-due, the formula becomes:

$$Bond\ value = C \cdot a_{\overline{T}|r} + \frac{Par\ value}{(1+r)^t}, where\ a_{\overline{T}|r} = \frac{1-(1+r)^{-T}}{r}$$

After a bond is issued at par, it can be held to maturity or sold on the secondary market. In the latter case, depending on many factors, e.g. market interest rate variations, the price of the bond will increase or decrease, and therefore trade below or above par value. Therefore, a measure of the rate of return of the bond, that accounts for both eventual capital gains or losses. This measure it the *yield to maturity* (YTM), defined as "the interest rate that makes the present value of a bond's payments equal to its price", often interpreted as the average rate of return of a bond bought now and held to maturity. It is obtained equaling the present value to the price:

$$P = \sum_{t=1}^{T} \frac{C}{(1+YTM)^t} + \frac{F}{(1+YTM)^t}$$

The equation shows that price and YTM inversely related. As a consequence, if price increases YTM will decrease; vice versa, if price decreases, YTM will increase.

It is important to notice that if the price of the bond is equal to its par value, its coupon rate will equal its yield to maturity. When the coupon rate is lower than the market interest rate, investors will bid down the price and subsequently the bond will sell below par value, i.e. at a *discount*, and a built-in capital gain will be

available for those that hold it to maturity. On the other hand, when the coupon rate is higher than the market interest rate, investors will bid up the price and the bond will sell above par value, i.e. at a *premium*, and those who hold it to maturity will eventually face a capital loss. From this mechanics it appears evident that interest rates and bond prices are inversely related. However, as the bonds approach maturity, and therefore the interest difference becomes less and less relevant, prices approach par value. Figure 6 shows the price path of two 30-years maturity bonds:

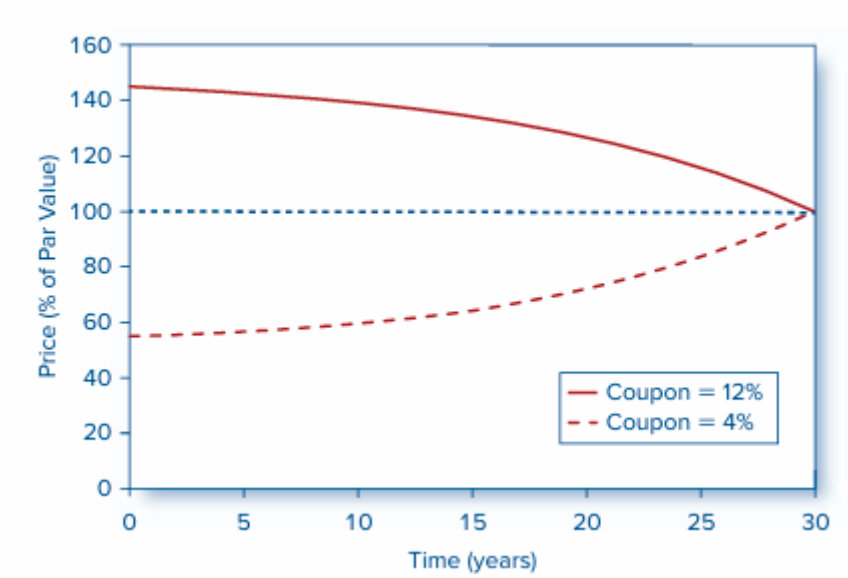

Figure 6. How bonds approach Par value

3. BOND PRICING AND THE TERM STRUCTURE

Bodie (2023) poses a further problem to face when dealing with bond pricing: if yields on different-maturity bonds differ, how should we value coupon bonds that

make payments at different times? The solution is to consider each cash flow as a stand-alone zero-coupon bond, as it is done with Treasury strips. The latter ones are independent zero-coupon securities created by the Treasury upon a request of a bond dealer, where each one is either a coupon payment or the principal payment. The total value of the strips and the value of the bond ought to be the same, otherwise there would be opportunities for arbitrage through *bond stripping* or *bond reconstitution*, and this is assumed be excluded. Therefore, the *spot rate* is defined as today's observed interest rate for a zero-coupon bond that pays interest at a future time T. So the coupons of an n-years bond will be discounted each at n year's *spot rate*. Spot rates are pivotal to price bonds accurately. To find the spot rate, one has to look at the *term structure* of interest rates, i.e. the relationship between maturity and YTM. This relationship can be visualized through the *yield curve*, which can be downward-sloping, upward-sloping or flat, depending on the expectations about future inflation and about future market interest rates. Some examples are shown hereafter in Figure 7:

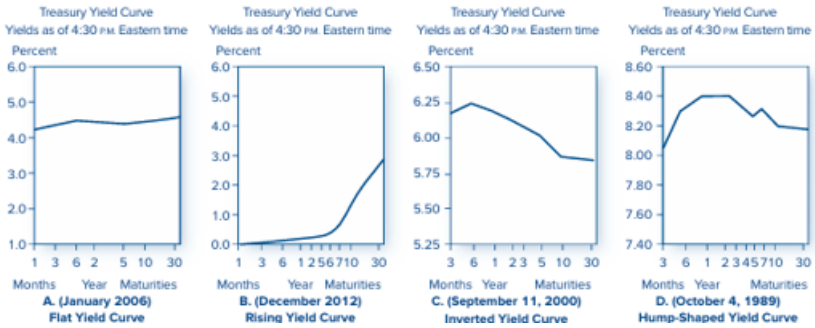

Figure 7. Different types of yield curve

4. SOURCES OF RISK AND CREDIT SPREAD

Bodie (2023) explains that, given a bond contract that is established between two parties, there is the possibility that the lender will default on the obligation. The risk of default, commonly called *credit risk*, is measured by rating agencies such as Moody's Investor Services, Standard & Poor's Corporation, and Fitch Ratings. These agencies provide financial information about firms and quality ratings of bond issues. Rating is assigned through letter grades, reflecting the assessment of safety, i.e. capacity to repay principal and interest. The top rating is AAA or Aaa. Bonds whose rating is BBB (or Baa) or above are considered *investment-grade bonds*, while lower-rated bonds are called *junk bonds* (or speculative bonds).

In general, it is assumed that government bonds are risk-free, for the likelihood of a default is very low, especially for economically strong countries like US. The same is not true for corporate bonds, since they do not have a central bank to back them, and they are therefore exposed to the possibility of insolvency. For this reason, when dealing with a corporate bond a distinction between the *stated yield*, i.e. the maximum possible yield realized with no default, and the *expected yield*, that takes into account default chances, has to be made.

To compensate for *credit risk*, corporate bonds must offer a default premium, also known as *credit spread*, i.e. offer a higher interest than that offered by government

bonds. The *credit spread*, in general, is the difference between the stated yield on a corporate bond and the yield of an otherwise-identical government bond that is (assumed to be) riskless in terms of default.

The pattern of *credit spread* is known as the *risk structure* of interest rates. This measure varies over the years, and widens in periods of crisis, when the general risk of default becomes higher. Figure 8 displays historical credit spread trends of 10-year bonds:

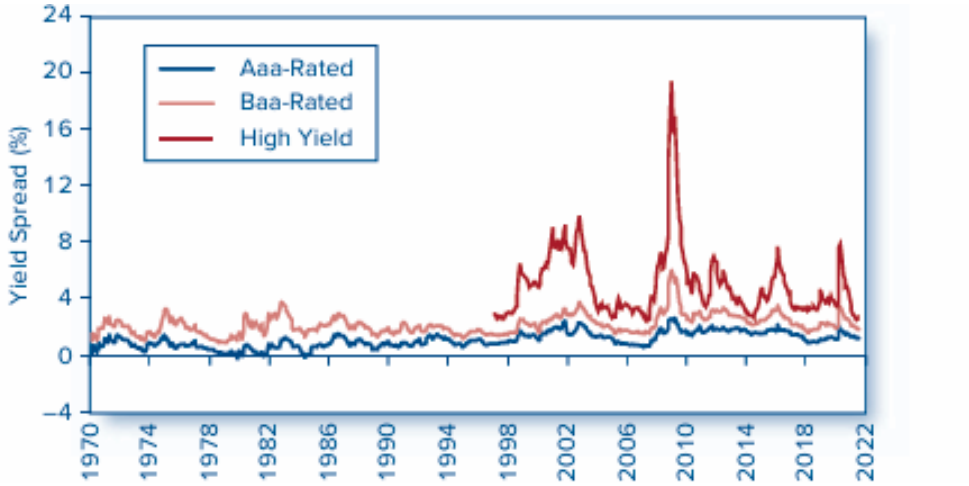

Figure 8. Credit spread of bonds over the years

A second important type of risk is *liquidity risk*, i.e. the risk related to the marketability of a bond. Liquidity is the ease with which a security can be traded in exchange for money; it is a consequence of trading volume and market depth, measured by the bid-ask spread. Holding an illiquid bond means incurring in high transaction costs at the moment of sale, due to high bid-ask spreads. Liquidity risk

increases with the maturity of a bond, and is higher for corporate bonds than for government bond, as the former ones are substantially less traded. Furthermore, in the case of long-term bond, liquidity risk is not only related to transaction costs, but also and more importantly to the time the investor has to renounce to his money for.

A third type of risk is *interest rate risk*. Whenever a bond is issued, a certain interest is promised as a compensation for the time value of the money. However, next issuances may promise a higher rate if the market rate has risen, and the previously issued bond may become subsequently less attractive for those entering the bond market, hence selling at a *discount*. If the bondholder holds the first bond to maturity, things remain unchanged, but if he has to sell it for whatever reason, e.g. at the moment of financial need, he will face a capital loss. In general, the longer the maturity, the more sensitive the bond will be to interest rate changes, and the more significant will the *interest rate risk* be.

The fourth and last type of risk here discussed is *inflation risk*, i.e. the possible reduction of the purchasing power of the money received at maturity. This risk is stronger for longer-term bonds, because they are subject to inflation, which compounds over time, for longer periods.

Generally, investors want to be compensated for the risks. The *risk premium* takes into account the many risks, but it may be complex to assess how much each risk accounts for the total premium. In particular, it is not clear which are the main determinants of *credit spread*, where maturity is fixed and therefore risks related to

maturity can be neglected. In this regard, how much does *credit risk* contribute to credit spread? Does *liquidity risk* contribute significantly? Are there more factors to consider?

Elton et al. (2001) traces credit spread back to three main causes. Firstly and obviously, default risk is empirically confirmed to be a cause but, surprisingly, it is found to account at most for 25% of the spread. A second and larger cause that is found is a *tax premium*, as a consequence of the fact that payments on corporate bonds are taxed while payments on government bonds are not. This second cause appears to account for up to 37% of credit spread. As a third cause, Elton et al. (2001) finds a premium for systematic risk. In the same way in which stock returns are related to their systematic risk, bond yields are higher when systematic risk is higher. This premium accounts for 85% of the remaining spread.

Chen et al. (2007) finds that liquidity is priced in corporate yield spread (i.e. credit spread). It asserts that that nor the credit spread nor its changes can be fully explained by structural form models. Instead "[…] more illiquid bonds earn higher yield spreads, and an improvement in liquidity causes a significant reduction in yield spreads. These results hold after controlling for common bond-specific, firm-specific, and macroeconomic variables, and are robust to issuers' fixed effect and potential endogeneity bias. […] Because investors cannot continuously hedge their risk, they demand an ex ante risk premium by lowering security prices. Therefore, for the same promised cash flows, less liquid bonds will trade less frequently,

realize lower prices, and exhibit higher yield spreads." The association between illiquidity and a large yield spread is a positive one, and it is found to explain up to 7% of yield variation for investment grade bonds, and up to 22% of yield variation for speculative grade bond (i.e. junk bonds).

It is evident that the determinants of credit spread are many and that they are different. There is a widespread consensus that structural models fail to catch details of these dynamics. Furthermore, given the behavioral approach adopted in the previous part of this work, it is relevant to investigate for causes related to sentiment, rather than on pure rationality. In this respect, Nayak (2010) shows that sentiment does in fact have a role. Market sentiment has two definitions: "(a) it denotes the level of (irrational) optimism or pessimism in projections of future cash flows and risks underlying any security, and (b) it reflects the propensity to speculate in certain securities that are more likely to be mispriced and are difficult and costly to arbitrage." Two assumptions back this definition: firstly, there exist biased investors who build expectations are not build on fundamentals; second, there are inherent limits to arbitrage, such as the borrowing cost faced when short-selling. Thus, the irrational expectations become relevant at the aggregate level, as already shown in Shiller (1981), where the deviations of price are not justified by movements of fundamentals at all. This effect in the bond market, in Nayak (2010), is investigated using credit spread as a measure. It is assumed that bonds are relatively underpriced when market sentiment is low (i.e. pessimistic) and therefore

yields increase and credit spread widens; in the same way, when market sentiment is high (i.e. optimistic) bonds are relatively overpriced, hence yields decrease and credit spread narrows. As a consequence of this, it is hypothesized that *after* a pessimistic period in which credit spread widens, prices will adjust and therefore yields will decrease, so credit spread will narrow; in the same way, *after* an optimistic period in which credit spread narrows, prices will adjust and therefore yields will increase, so credit spread will widen. Summing up, "if beginning-of-period sentiment is low, subsequent spreads are low" and if beginning-of-period sentiment is high, they are high. This view is supported by data, and the trend is stronger for speculative grade bonds. The Figure 9 and Figure 10 display the aforementioned hypotheses:

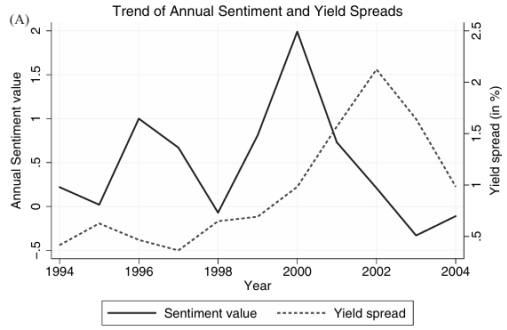

Figure 9. Annual trend

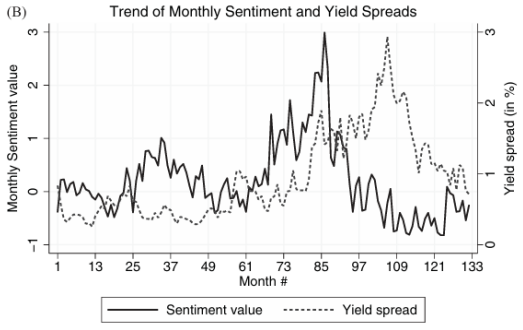

Figure 10. Monthly trend

Given what these works empirically tested, it can be concluded that credit spread is not only default-driven, but presents a variety of different determinants, including some of a behavioral kind.

5. PEER TO PEER LENDING

Peer to peer (P2P) lending is a form of alternative finance performed through digital means. According to Morse et al. (2015), "P2P lending platforms provide a facility creating a marketplace where investors who wish to lend funds can find potential borrowers and provide credit through P2P Agreements. Platforms use online technologies to facilitate efficient interactions between investors and borrowers that would be difficult or costly to achieve in other ways, hence opening up the feasibility of direct lending to a wider range of investors and borrowers than before." P2P platforms undertake a set of additional operational functions: verifying borrower characteristics; assessing credit quality to ensure proportional interest rates; payment processing on both sides; making available data to inform investment decisions; collecting debt in case of arrears or default; etc. Additional services may be provided, like determining the interest rate for investors, auto-allocation of investments to loans or even providing investors with secondary markets to exit investments more easily.

Morse et al. (2015) explains that due to the fact that P2P lending business models evolved over time, there is very much heterogeneity in how platforms seek for investors, in the types of loan and, of course, in the added services available. Their innovative approaches are not constrained by legacy systems, given their relatively recent entry to the market, so they can introduce changes more rapidly and cheaply than traditional intermediaries.

"From the perspective of borrowers, P2P lending platforms provide sources of finance that primarily compete with those offered by banks. In these competitive markets, P2P lenders can typically be expected to be 'price-takers', in that they lend at the going rate. […] On the investor side, P2P lending platforms give retail and institutional investors the opportunity to fund loans directly. Investors therefore essentially own a part of the cash flows of a lending business, tied to specific loans through the P2P Agreement. For the investors, this represents a novel asset class".

Table 6 compares P2P lending with three other forms of exposure to lending.

The risks and liquidity constraints offer the investor considerable advantages in terms of return, meaning that they can earn a higher return than that obtained from products such as bank savings account.

But why does P2P lending ultimately survive in the market? Morse et al. (2015) reports that P2P loans process has been found to be faster and characterized by a more simple application process. On the other hand, there is no evidence that they

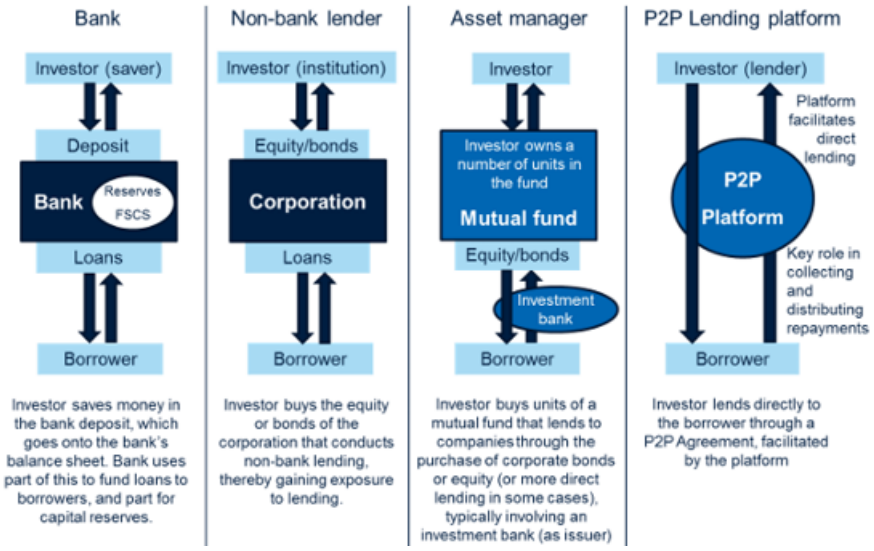

Table 6. Comparation of P2P lending dynamics with other types of investment

offer better rates. Moreover, from the investor side, P2P lending is a completely new offer in terms of its combination of risk-return trade-off and liquidity. Figure 11 displays this relative to other types of debt investment:

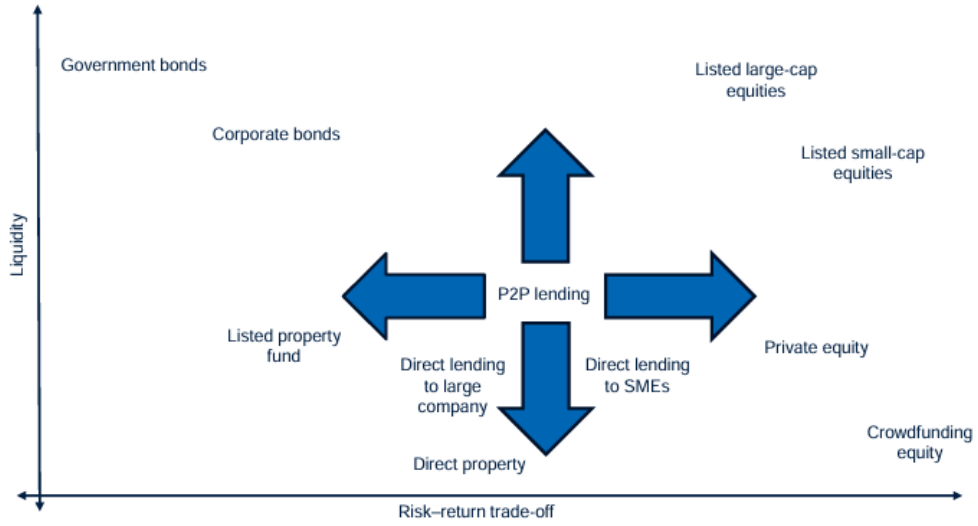

Figure 11. Investments liquidity and risk-return trade-off

It appears that P2P lending has a risk-return trade-off higher than bonds, as well as a lower liquidity. Therefore, it can be expected that the typical investor involved is oriented to long-term investments. According to Morse et al. (2015), risk management is operated through functions including: credit risk assessment and interest rate management; diversification; smoothing returns; providing liquidity; managing platform risk, i.e. the risk that a P2P platform collapses. Platforms have incentives to manage risk due to revenue impacts and reputational impacts which, in turn, affect long term survival.

From a valuation perspective, crowdlending can be treated analogously to bonds, summing the present value of future cash flows. Specifically, the installments received during the period of the loan are comparable to the coupons of a bond, and the possibility of selling the loan in the secondary market (whenever platforms make it possible to do so) makes the loan similar to a fixed-income security that can be sold when facing a more convenient market rate. However, the very scarce trading volume, consequential to the very illiquid nature of this type of investment, makes the impact of the trades much less powerful with respect to the fixed-income market. The riskiness of a P2P loan with a certain grade can be measured, as it is done with corporate bonds, through credit spread with respect to a risk-free benchmark, possibly a government bond with the same maturity. Foo et al. (2017) is an example in this regard. This spread must take into account the lower liquidity

and many risks related to this type of investment, including platform risk that is not found when dealing with bonds. However, it is reasonable to think that much of the spread is due to default risk, since, as it is found in Wang & Li (2022), a correlation is found between high interest rates and probability of default. Moreover, according to Foo et a. (2017), systemic risk and general uncertainty have a meaningful role as well as general uncertainty. Due to the scarce regulation and the high heterogeneity of the platforms, risks related to moral hazard and adverse selection may be significant as stated by Wang & Li (2023), and therefore are to be included.

Little literature has so far developed on the topic, and it is desirable that research will move in the direction of analyzing more deeply the risk determinants of P2P credit spread, along with the investigation of behavioral factors that do or may influence this recently developed market.

CHAPTER 3: EMPIRICAL ANALYSIS

1. DATA

1.1. Data collection

The first goal of the empirical analysis was to find out which are the main variables crowlending dynamics stem from. In order to do this, data was collected from the crowlending.it website.

Figure 12. Crowdlending.it

As Figure 12 shows, projects are displayed on the website. There are about 197 projects available to possibly invest in, with different conditions in terms of risk, rating, maturity and others. Figure 13 below shows how these variables appear on the website. The data so displayed was collected through *web scraping*, a coding technique whose aim is extracting information from websites, with different levels of complexity. The web scraping was carried out through Python programming language, in particular using the library Selenium.

In Figure 14 follows an extract of the code. The first block of code is aimed at "presenting" the algorithm as a search engine, about which every detail is specified, e.g. name, version, etc. After that, the second block is the one that actually enters the website and a couple of lines to reject cookies are written as well.

Figure 13. Typical display of information about a project

```
options = Options()
options.add_experimental_option("detach", True)
# We keep this standard to avoid bot detection flags
options.add_argument("user-agent=Mozilla/5.0 (Windows NT 10.0; Win64; x64) " \
"AppleWebKit/537.36 (KHTML, like Gecko) Chrome/122.0.0.0 Safari/537.36")

driver=webdriver.Chrome(options=options)
driver.maximize_window() # Important: ensures elements are visible
driver.get("https://www.crowdlender.it/it/finanziamenti")
wait=wdw(driver, 8)
refuse_button = driver.find_element(By.XPATH, "//div[contains(text(), 'Rifiuta tutto')]")
refuse_button.click()#Click the button to reject cookies

links = driver.find_elements(By.XPATH, "//div[contains(@class, 'text-center mb-2')]//a")
url_list = []
for link in links:
    url = link.get_attribute("href")
    if url: # Ensure it's not a null link
        url_list.append(url)

driver.get(url_list[0])
```

Figure 14. Entering the website and getting the links

The “for loop” shown in the third block is the real pivot of scraping, where all the links to the single projects are got in order to scrape them one by one in a second time. As shown in Figure 15, a list containing the links is created:

```
variables = ['Project_name', 'Raised_capital', 'Target', 'Max_cap', 'Interest_rate', 'Maturity', 'Rating', 'Minimum_investment' ]
print(variables)
all_data = []

for final_url in url_list:
    driver.get(final_url)
    try:
        project_name = driver.find_element(By.XPATH, "//div[contains(@class, 'text-center')]/h1").text.strip().replace(',', ' ')
    except:
        project_namec = 'NA'
```

Figure 15. Scraping the values of every variable

The for loop iteration shown in Figure 15 has the goal of scraping the values of every variable we are interested in. The loop consists of entering every single link in the aforementioned list, and of finding, through Xpath, the values of the variables contained in the HTML code of the web page. A try/except structure is also used, considering that the older projects are found not to have values of variables such as rating, to avoid bugs and instead keep the observations as not available (NA).

Xpath, which stands for XML path language, is a powerful query language that can target specifically the nodes of an HTML document to find a specific item, at any point in the tree. It is here used both for finding the links of Figure 14 and for scraping the values in Figure 15.

```
    row_data = [project_name, raised_cap, target, max_cap, interest, maturity, rating, min_investment]
    all_data.append(row_data)
    print(row_data)


df=pd.DataFrame(all_data, columns=variables)
df.to_csv(r'C:\Users\39347\OneDrive\Documenti\2. Uni\Tesi e tirocinio\crowdlending data.csv', index=False)
```

Figure 16. How the data framework was created

As shown in Figure 16, after collecting data in lists, a list of lists is created, still within the for loop, to create something that looks like a table. Once this process is completed, through a function from Pandas library, the list of lists is converted to a data frame where the names of the columns are those of the list “variables” shown in Figure 15, and then exported as a CSV file.

1.2. Data cleaning

The main problem with the observations collected is that their data type after scraping was “string”. When dealing with data in general, and especially when performing quantitative analysis such as regression analysis, it is fundamental to convert the observations into the correct data type. Moreover, excess text was contained in every information, since values were scraped without filtering. How the data appeared is displayed in Figure 17:

	A	B	C	D	E	F	G	H
1	Project_name	Raised_capital	Target	Max_cap	Interest_rate	Maturity	Rating	Minimum_investment
2	Supernova - Rispescia tranche 2	30.222,07â,¬	500,00â,¬	186.559,65â,¬	12,00%	24 mesi	5 - RISCHIO MEDIO ALTO	500,00â,¬
3	Edil Power Corporation Tranche 2	25.000,00â,¬	10.000,00â,¬	25.000,00â,¬	10,00%	18 mesi	5 - RISCHIO MEDIO ALTO	1.000,00â,¬
4	Edil Power Corporation	346.000,00â,¬	300.000,00â,¬	370.000,00â,¬	10,00%	18 mesi	5 - RISCHIO MEDIO ALTO	1.000,00â,¬
5	Setteduezero Costruzioni	142.000,00â,¬	100.000,00â,¬	140.000,00â,¬	10,00%	21 mesi	5 - RISCHIO MEDIO ALTO	1.000,00â,¬
6	Supernova - Rispescia	113.440,35â,¬	100.000,00â,¬	300.000,00â,¬	12,00%	24 mesi	5 - RISCHIO MEDIO ALTO	500,00â,¬

Figure 17. Observations before cleaning

Besides the “Project_name” variable, it is clear that the remaining variables are intended to be quantitative but they are not in practice. In order to remove every excess symbol or word, revert commas and points when referring to numbers and to convert them into a “float” data type, Python was used again, along with Pandas library.

```python
import pandas as pd

df = pd.read_csv(r'C:\Users\39347\OneDrive\Documenti\2. Uni\Tesi e tirocinio\crowdlending data.csv')
print(df.head(1).T)

df['Raised_capital'] = df['Raised_capital'].str.replace('.', '').str.replace(',', '.', regex=True).str.replace('€', '', regex=True)
df['Target'] = df['Target'].str.replace('.', '').str.replace(',', '.', regex=True).str.replace('€', '', regex=True)
df['Max_cap'] = df['Max_cap'].str.replace('.', '').str.replace(',', '.', regex=True).str.replace('€', '', regex=True)
df['Minimum_investment'] = df['Minimum_investment'].str.replace('.', '').str.replace(',', '.', regex=True).str.replace('€', '', regex=True)
df['Interest_rate'] = df['Interest_rate'].str.replace(',', '.', regex=True).str.replace('%', '', regex=True)
df['Maturity'] = df['Maturity'].str.replace(' mesi', '', regex=True)
df['Rating'] = df['Rating'].str[0]

# 1. Define the columns that should be floats/integers
numeric_cols = [
    'Raised_capital',
    'Target',
    'Max_cap',
    'Minimum_investment',
    'Interest_rate',
    'Maturity',
    'Rating'
]

# 2. Convert them all to numeric data types
df[numeric_cols] = df[numeric_cols].apply(pd.to_numeric, errors='coerce')

df['Interest_rate'] = df['Interest_rate'] / 100

df = df.rename(columns={
    'Maturity': 'Maturity_Months',
    'Interest_rate': 'Promised_yield'
})

print(df.head(1).T)
df.to_csv(r'C:\Users\39347\OneDrive\Documenti\2. Uni\Tesi e tirocinio\crowdlending data cleaned.csv', index=False)
```

Figure 18. Data cleaning algorithm

Figure 18 shows the code used to perform what previously explained. Every single variable was cleaned from excess symbols and words, and correctly converted to a numerical data type. Finally, the cleaned data was exported as a new CSV file ready to be analyzed. The final result follows in Figure 19:

	A	B	C	D	E	F	G	H
1	Project_name	Raised_capital	Target	Max_cap	Promised_yield	Maturity_Months	Rating	Minimum_investment
2	Supernova - Rispescia tranche 2	30222.07	500.0	186559.65	0.12	24	5.0	500.0
3	Edil Power Corporation Tranche 2	25000.0	10000.0	25000.0	0.1	18	5.0	1000.0
4	Edil Power Corporation	346000.0	300000.0	370000.0	0.1	18	5.0	1000.0
5	Setteduezero Costruzioni	142000.0	100000.0	140000.0	0.1	21	5.0	1000.0
6	Supernova - Rispescia	113440.35	100000.0	300000.0	0.12	24	5.0	500.0

Figure 19. Observations after cleaning

2. REGRESSION MODEL

2.1. Data visualization and descriptive analysis

The last step before performing a regression analysis for our model is to deeply understand the data. To do this and for the following analysis, , we firstly visualize

the general summary statistics of all the present variables. From Figure 20 and Figure 21 we can already see that the number of Null values is very high: the Rating variable presents 90 missing values.

```
Project_name       Raised_capital        Target          Max_cap        Promised_yield
Length:197         Min.   :  10931   Min.   :    50   Min.   :  18314   Min.   :0.05000
Class :character   1st Qu.:  52842   1st Qu.: 40000   1st Qu.:  78863   1st Qu.:0.09500
Mode  :character   Median : 100014   Median : 50000   Median : 220000   Median :0.10250
                   Mean   : 161025   Mean   : 75395   Mean   : 424350   Mean   :0.09866
                   3rd Qu.: 181079   3rd Qu.:100000   3rd Qu.: 556622   3rd Qu.:0.10500
                   Max.   :2001000   Max.   :800000   Max.   :2000000   Max.   :0.13500
                   NA's   :5         NA's   :5        NA's   :5
```

Figure 20. Summary statistics 1

```
Maturity_Months     Rating      Minimum_investment
Min.   : 4.0    Min.   :3.000   Min.   :    50
1st Qu.:12.0    1st Qu.:5.000   1st Qu.:    50
Median :17.0    Median :5.000   Median :    50
Mean   :16.1    Mean   :4.972   Mean   :  1911
3rd Qu.:17.0    3rd Qu.:5.000   3rd Qu.:    50
Max.   :60.0    Max.   :6.000   Max.   :100000
                NA's   :90
```

Figure 21. Summary statistics 2

The two ways this problem was solved were the removal of all the rows with null values on one hand and the removal of the Rating variable on the other and null rows only based on the other variables' nulls. This two processes, shown in Figure 22, culminated in two different datasets.

```
projects_full <- projects %>% drop_na(Raised_capital, Target, Max_cap)
projects_full <- projects_full %>% select(-Rating)
projects <- projects %>% drop_na(Rating, Raised_capital, Target, Max_cap)
```

Figure 22. Creation of two different datasets

Secondly, distributions of the variables are plotted to check whether they are skewed or whether they present particular traits that need to be offset.

The first three variables that were analyzed are Raised capital, Target capital and Maximum capital, that contain respectively how much capital was raised, the target of the financing campaign, and the maximum capital receivable for the campaign (which can exceed the target). These three variables are expressed in euros and, similarly to how monetary quantities usually behave, they show in Figure 23 a

distribution similar to a lognormal one, i.e. heavily right-skewed. As expected when plotting the distribution of the logarithm of the variables, it appears closer to a normal one. Therefore, these variables were treated with a logarithm to ensure the results of the analysis to be reliable.

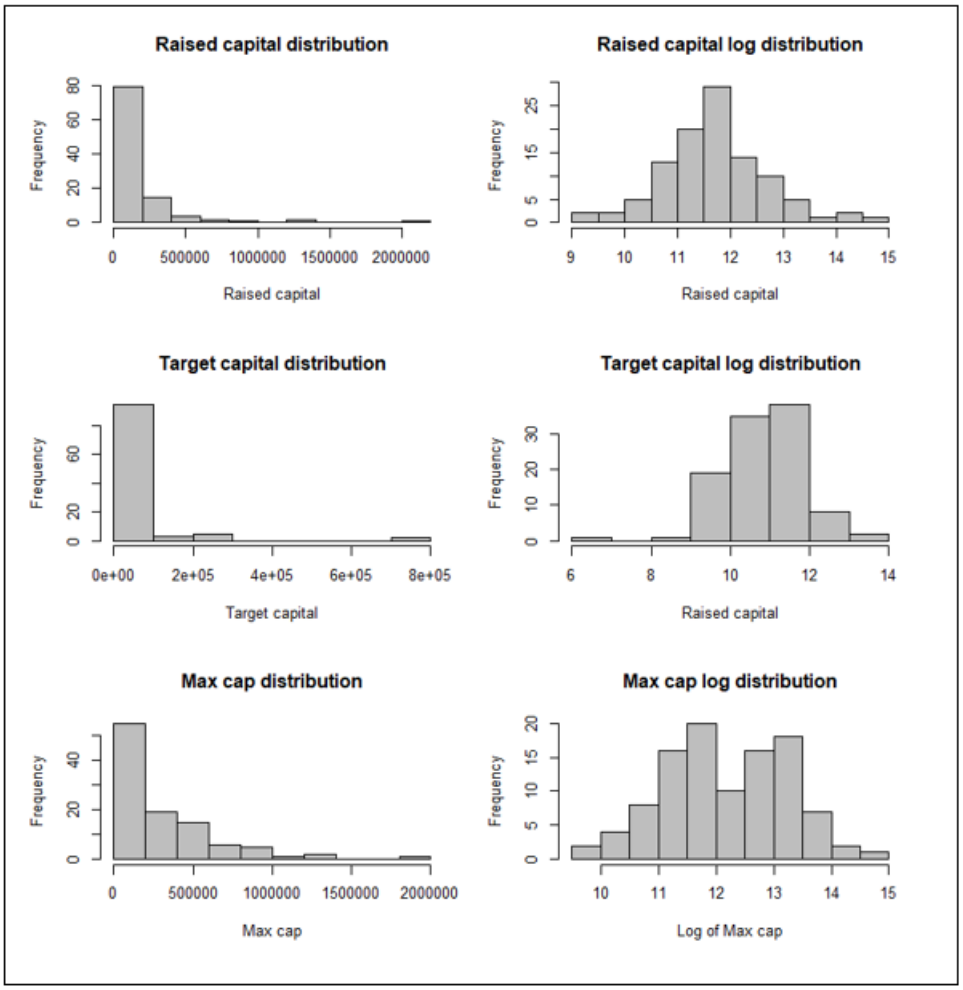

Figure 23. Distributions and log distributions

What is more difficult to manage is what comes out in Figure 24, when displaying the distribution of minimum investment, that contains the minimum amount to be invested in each project. Here, both in the case

of the variables as it is and of its

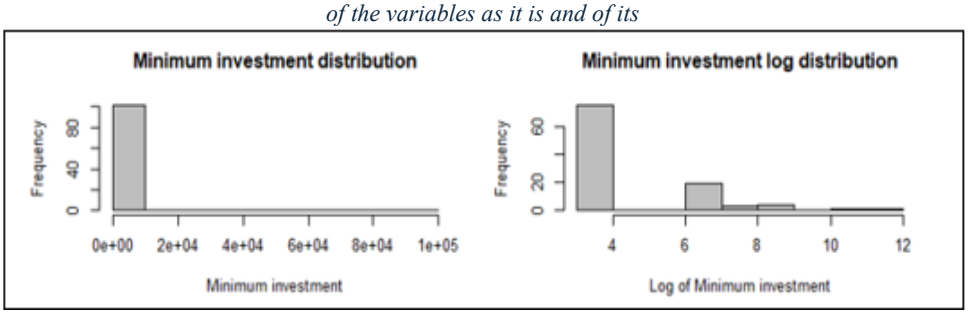

Figure 24. Minimum investment distributions

logarithm, the biggest part of the values appears to be in the first interval, with a huge distance from the others.

Also here, two possible ways were found to solve this issue: the first is to use the log version of the variable. The only alternative is to create a dummy variable which separates the projects that fall in the first group from the others, so as to later assess whether belonging to the first group results in a variation of the dependent variable. The code implemented to create the dummy variable in both the datasets is shown in Figure 25.

```
projects <- projects %>%
  mutate(High_Min_Inv = as.integer(Minimum_investment > 50))
projects_full <- projects_full %>%
  mutate(High_Min_Inv = as.integer(Minimum_investment > 50))
```

Figure 25. Creation of a dummy for Minimum_investment

About the Promised Yield variable, it contains a decimal number that represents, in a way, the yield to maturity. Being it decimal, it wouldn't make any sense to apply a log transformation, since it would simply return negative values. Figure 26 shows its distribution:

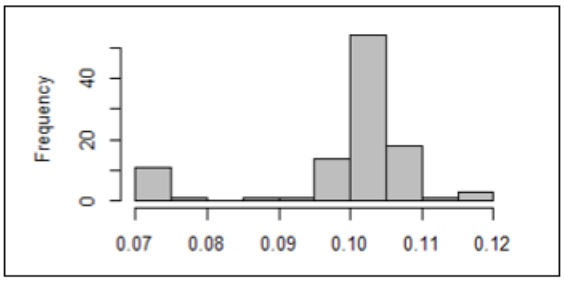

Figure 26. Promised yield distribution

Another variable to consider is the Maturity, expressed in months, which contains the maturity of each crowdlending project, i.e. how much each project lasts. As figure 27 shows on the left, this variable appears to be right-skewed, with a large distance between the major part of the distribution and some very heterogeneous values. Nonetheless, the log transformation appears to partly fix this problem, reducing the distance between the values. Moreover, being the relationships

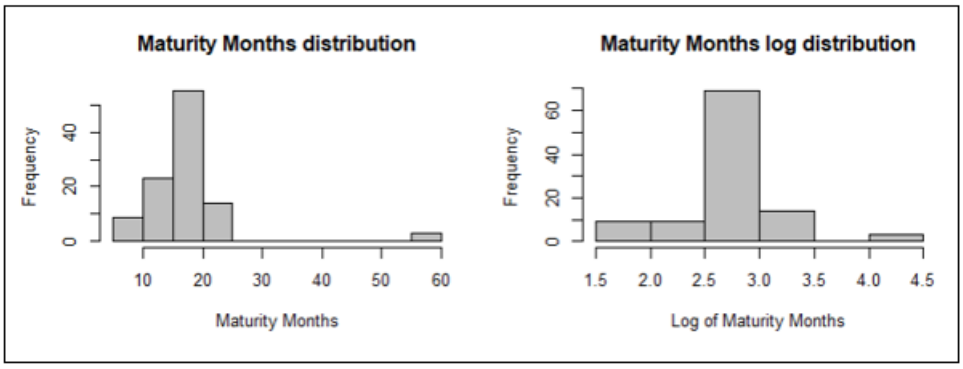

Figure 27. Maturity distributions

between maturity and returns and between maturity and capital raised characterized by diminishing marginal utility, the log transformation makes perfect sense.

Lastly, the Rating variable contains a rating of the riskiness of each project. Being the variable naturally linear, it is not necessary to apply a log transformation. In figure 28, the distribution is displayed:

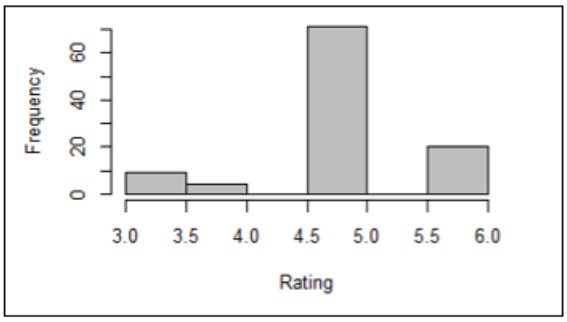

Figure 28. Rating distribution

2.2. Regression analysis

To establish relationships among the variables, a regression analysis was conducted through R programming language. The relationship that allows to best catch crowdlending dynamics is that obtained by considering Raised capital as a dependent variable and the other variables as explanatory variables:

$$\log(K_{raised}) = \beta_0 + \beta_1 \cdot j + \beta_2 \cdot \log\left(K_{target}\right) + \beta_3 \cdot \log(K_{\max}) + \beta_4 \cdot \log(t) + \beta_5 \cdot R + \beta_6 \cdot \log(I_{min}) + \varepsilon$$

where:

$j = promised\ yield$

$K_{raised} = Raised\ capital$

$K_{target} = Target\ capital$

$K_{\max} = Maximum\ raisable\ capital$

$t = Maturity\ (months)$

$R = Rating$

$I_{min} = Minimum\ investment$

$\varepsilon = Statistical\ error$

At the same time, other regression coefficients were calculated, by treating minimum investment as a dummy, as previously explained. The new regression equation therefore was:

$$\log(K_{raised}) = \beta_0 + \beta_1 \cdot j + \beta_2 \cdot \log(K_{target}) + \beta_3 \cdot \log(K_{\max}) + \beta_4 \cdot \log(t) + \beta_5 \cdot R + \beta_6 \cdot \mathrm{D}_{\text{minimum investment}} + \varepsilon$$

Follows in Figure 29 the code implemented to calculate the coefficients:

```
regression1 <- lm_robust(log(Raised_capital) ~ Promised_yield +
                           log(Target) +
                           log(Max_cap) +
                           log(Maturity_Months) +
                           Rating +
                           log(Minimum_investment),
                         data=projects)


regression2 <- lm_robust(log(Raised_capital) ~ Promised_yield +
                           log(Target) +
                           log(Max_cap) +
                           log(Maturity_Months) +
                           Rating +
                           High_Min_Inv,
                         data=projects)
```

Figure 29. R script for regression analysis

Even regressions using the second dataset (without the Rating variable and with more observations) were computed. However, being their adjusted R^2 significantly lower and having them led to the very similar results, they will not be shown nor discussed in detail. Among the two regressions shown above, the one that best fits data is the one that contains the dummy variable. We therefore show the results in figure 30:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	CI Lower	CI Upper	DF
(Intercept)	3.2085	1.05946	3.028	3.150e-03	1.1057	5.3112	97
Promised_yield	-14.7305	13.36562	-1.102	2.731e-01	-41.2576	11.7965	97
log(Target)	0.3181	0.05976	5.323	6.573e-07	0.1995	0.4367	97
log(Max_cap)	0.4995	0.05207	9.593	1.014e-15	0.3962	0.6029	97
log(Maturity_Months)	-0.2133	0.18496	-1.153	2.517e-01	-0.5804	0.1538	97
Rating	0.1875	0.15616	1.201	2.327e-01	-0.1224	0.4975	97
High_Min_Inv	0.3586	0.10821	3.313	1.296e-03	0.1438	0.5733	97

Figure 30. Results of the regression

Results are pretty straightforward:

- the value of the intercept is the second highest, which indicates that the starting base of Raised capital K_{raised} would have a value of $e^{3.2085} \approx 24.74$, which is not realistic and will never happen. Therefore, it is just a mathematical anchor for the regression line;
- Promised yield j appears to have the largest coefficient in absolute value: a 1 percentage point increase in j would mean a 14.731% decrease in raised capital; however, the result is not statistically significant, so we cannot actually reliably infer any information about this coefficient;
- K_{target} entails a 0.318% increase in K_{raised} for a 1% increase;
- $K_{\max}$ results in a $\approx$ 0.5% increase in K_{raised} for a 1% increase;
- Maturity months t would cause a 0.213% decrease for a 1% increase in K_{raised}; however, also this coefficient is not statistically significant, and no inference can be reliably made as well;
- Rating R would entail a 18.75% increase in K_{raised} for an absolute increase of 1; in reality, no inference can be made about this coefficient for it is not statistically significant;
- Finally, having a minimum investment greater than 50 means a 35.86% higher capital raised.

These relationships are true on *average*, and not for all the coefficients. As already said, and as the confidence intervals and the p-values suggest, Promised yield j,

Maturity $\log(t)$ and Rating R are not statistically significant. This means we cannot confidently assert that these relationships are true even out of the sample, i.e. we cannot make any reliable inference about magnitude or sign or even existence of these relationship due to the nonsignificance.

2.3. Robustness check

To check for robustness of the functional form, a reversal of the functional form was tested. Therefore, the four regressions previously computed (two per dataset) were recomputed setting Promised yield j as the dependent variable. From an economic point of view, it is much more correct to use the previous model to investigate and understand the behavior of investors, which *results* in the raised capital. This secondo model has the only purpose of verifying whether the nonsignificancy remains such when the relationship is reversed. Therefore, the functional form of this alternative model will be:

$$j = \beta_0 + \beta_1 \cdot \log(K_{raised}) + \beta_2 \cdot \log(K_{target}) + \beta_3 \cdot \log(K_{\max}) + \beta_4 \cdot \log(t) + \beta_5 \cdot R + \beta_6 \cdot \mathrm{D}_{\text{minimum investment}} + \varepsilon$$

Results are displayed in figure 31:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	CI Lower	CI Upper	DF
(Intercept)	0.063901	0.011436	5.5879	2.105e-07	0.041205	0.0865980	97
log(Raised_capital)	-0.003162	0.002007	-1.5754	1.184e-01	-0.007145	0.0008215	97
log(Target)	0.000245	0.001459	0.1679	8.670e-01	-0.002651	0.0031412	97
log(Max_cap)	0.001374	0.001655	0.8305	4.083e-01	-0.001910	0.0046580	97
log(Maturity_Months)	0.002880	0.005054	0.5698	5.701e-01	-0.007151	0.0129109	97
Rating	0.009534	0.002126	4.4847	2.005e-05	0.005315	0.0137531	97
High_Min_Inv	-0.001182	0.002906	-0.4069	6.849e-01	-0.006949	0.0045846	97

Figure 31. Results of the alternative regression

What comes out from the computation is that every variable but Rating R is not statistically significant, and R has a very marginal impact on j. From these results, it can be stated that, firstly, the correct relationship is that built in the previous

model where K_{raised} was the dependent variable and that, secondly, the nonsignificance generally stays unchanged, except for R.

2.4. Discussion

How can this regression model be generally interpreted? The nonsignificance of variables such as rating, maturity and yield means that they fail to capture the essential dynamics of crowdlending. This means that, ultimately, even if crowdlending presents similarities with bonds in the cash flow structure and in general common characteristics, when delving into a deeper comparison, the nature of the two investments is evidently diverse.

The first point to consider is how price determination happens. Bonds, obviously depending on type and maturity, are more or less frequently traded in the secondary market, and they are traded actively, which means that to a certain extent prices follow the laws of equilibrium, and investors are somewhere between price makers and price takers. On the other hand, crowdlending prices are determined by the platform, there is no negotiation, and the secondary market, whenever there is one, is much less active. Therefore, it is much harder to "correct" prices that are wrongly determined, and investors act much more as pure price takers. This is also why the two types of investments are so distant in terms of liquidity.

The second point to consider is that the market of crowdlending is much smaller and irregular than that of bonds, and this makes the whole equilibrium dynamics of crowdlending weaker. Even if the dynamics were comparable and the volumes of trade were equivalent relative to the size of the markets, being the crowdlending market much smaller it would be hard to get close to equilibrium with the same velocity found in the bond market. Moreover, secondary markets of crowdlending (where they do exist) are specific to the platform and disconnected from one another, which makes size and frequency of pricing mistakes much harder to compensate.

The third and most important point is about behavioral considerations. The regression analysis presented above signals something very clear: rational determinants of prices do not matter. They do not explain why things happen in this way. This might be due to the small size of the sample, but it more likely has to do with the bounded rationality of the actors. As it was reported in the previous chapter, when discussing Wang & Li (2023), information asymmetries do constitute an issue when dealing with crowdlending platforms, which makes decisions structurally less rational *a priori*. After acknowledging how little rational stock and bond markets are, it is reasonable to hypothesize that crowdlending markets amplify irrationality due to previous two points considered. However, this and the previous points need to be empirically proven and remain, therefore, more theories than laws.

CONCLUSION

Crowdlending is a very recent phenomenon and presents new market dynamics. In the analysis here presented a clear and thorough figure of how prices and yields are determined does not emerge. What instead emerges is that this new type of investment is far from being similar to the well-known and well-studied bond market. Variables that represent key determinants in the bond market are here found to be statistically nonsignificant, so it is much clear what does not determine what happens in crowdlending, rather than what does in fact determine why quantities behave as they do.

The points presented in the discussion are not supported by proof and are to be intended mostly as aspects to consider in future research, even if I personally consider them reasonable in light of economic intuition and in light of what reported about previous studies in the first two chapters.

Main points to investigate when dealing with crowdlending are about which variables determine its dynamics and which other variables hurdle quantities to move towards equilibrium. After that, it would be interesting to study how legislation could be made to standardize this type of investment and make it as efficient as possible from an economic point of view.

BIBLIOGRAPHY

Abdin, S. Z. ul, Qureshi, F., Iqbal, J., & Sultana, S. (2022). *Overconfidence bias and investment performance: A mediating effect of risk propensity*. *Borsa Istanbul Review, 22*(4), 780–793.

Allais, M. (1953). *Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine*. *Econometrica, 21*(4), 503–546.

Bee, M., & Desmarais-Tremblay, M. (2023). *The birth of homo œconomicus: The methodological debate on the economic agent from J. S. Mill to V. Pareto*. *Journal of the History of Economic Thought, 45*(4), 503–527.

Bodie, Z., Kane, A., & Marcus, A. J. (2023). *Investments* (13th ed.). McGraw-Hill.

Chen, L., Lesmond, D. A., & Wei, J. (2007). *Corporate yield spreads and bond liquidity*. *The Journal of Finance, 62*(1), 119–149.

De Bondt, W. F. M., & Thaler, R. H. (1985). *Does the stock market overreact? The Journal of Finance, 40*(3), 793–805.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). *Explaining the rate spread on corporate bonds*. *The Journal of Finance, 56*(1), 247–277.

Epstein, S. (1994). *Integration of the cognitive and the psychodynamic unconscious*. *American Psychologist, 49*(8), 709–724.

Fama, E. F. (1970). *Efficient capital markets: A review of theory and empirical work*. *The Journal of Finance, 25*(2), 383–417.

Foo, J. Y.-H., Lim, L.-H., & Wong, K. S.-W. (2023). *Discovering latent macroeconomic effects in peer-to-peer lending data*. *The Journal of FinTech, 3*(1–2).

Friedman, M., & Savage, L. J. (1948). *The utility analysis of choices involving risk*. *Journal of Political Economy, 56*(4), 279–304.

Garbuglia, F. (2024). *Che cos'è l'economia comportamentale*. Carocci.

Grežo, M. (2020). *Overconfidence and financial decision-making: A meta-analysis*. *Review of Behavioral Finance, 13*(3), 276–296.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). *Prospect theory: An analysis of decision under risk*. *Econometrica, 47*(2), 263–291.

Marimon, R., Spear, S. E., & Sunder, S. (1993). *Expectationally driven market volatility: An experimental study*. *Journal of Economic Theory, 61*(1), 74–103.

Moore, D. A., & Healy, P. J. (2008). *The trouble with overconfidence*. *Psychological Review, 115*(2), 502–517.

Muth, J. F. (1961). *Rational expectations and the theory of price movements*. *Econometrica, 29*(3), 315–335.

Nayak, S. (2010). Investor sentiment and corporate bond yield spreads. *Review of Behavioural Finance, 2*(2), 59–80.

Shiller, R. J. (1981). *Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review, 71*(3), 421–436.

Shiller, R. J. (2003). *From efficient markets theory to behavioral finance*. *Journal of Economic Perspectives, 17*(1), 83–104.

Singh, D., Malik, G., & Jha, A. (2024). *Overconfidence bias among retail investors: A systematic review and future research directions*. *Investment Management and Financial Innovations, 21*(1), 302–316.

The Economics of Peer-to-Peer Lending. (2016, September). *The economics of peer-to-peer lending*. Prepared for the Peer-to-Peer Finance Association. Oxera Consulting LLP.

Tversky, A., & Kahneman, D. (1992). *Advances in prospect theory: Cumulative representation of uncertainty*. *Journal of Risk and Uncertainty, 5*(4), 297–323.

Wang, J., & Li, R. (2023). *Asymmetric information in peer-to-peer lending: Empirical evidence from China*. *Finance Research Letters, 51*, 103452.