Assessing the subjective dimensions of risk tolerance.

Insight 2

InvestSuite

Assessing the subjective dimensions of risk tolerance.

The limited space and time available during the 'in-app' onboarding of clients creates a real challenge for robo-advisors. Collecting objective facts about investors is one thing. Understanding how they feel about risk, however, is something completely different. It requires both a scientifically sound methodology that yields consistent results and an engaging user experience that sets the tone for a long and fruitful relationship. Our research shows that, for risk profiling too, the magic is in the mix.

Onboarding retail clients to a roboadvisor essentially requires the same suitability assessment as any other service that provides investment advice and portfolio management[1]. In the ESMA guidelines we read that: **'Firms should establish, implement and maintain adequate policies and procedures -including appropriate tools- to enable them to understand the essential facts and characteristics about their clients. Firms should ensure that the assessment of information about their clients is done in a consistent way, irrespective of the means used to collect such information.'**

Understanding objective facts about clients is relatively easy, also with regards to risk. Information on a client's financial capacity, investment goal, time horizon, knowledge and experience are either available at the financial institution or can be collected in a straightforward manner. Understanding the more subjective aspect of investor risk tolerance (IRT), however, proves to be more difficult. Especially given that the same guidelines stipulate that 'self-assessment for the risk tolerance should be avoided'. Simply asking how much risk the client is willing to take, would be deemed irregular. Moreover, the procedure with which to gain an understanding of this -subjective- IRT must be 'adequate' and the assessment must be done in a 'consistent' way. This means that any method that is put in place must effectively measure what should be measured (not something else) and that outcomes shouldn't vary after re-assessments or after changing the format, design or medium used.

To translate this piece of regulation into a concrete challenge for financial institutions looking to launch a roboadvisor would be to ask: how do you make a psychometrically valid assessment of IRT, using a questionnaire on a 5,5-inch mobile screen... with the lowest possible drop-out rate?

How do you make a psychometrically valid assessment of IRT, using a questionnaire on a 5,5-inch mobile screen... with the lowest possible drop-out rate?

A PSYCHOLOGICAL CONSTRUCT WITH FOUR DIMENSIONS

A first step would be to define more specifically what IRT actually is and which factors constitute its make-up. A useful way to define IRT could be to describe it as "the maximum amount of uncertainty that someone is willing to accept when making a financial decision" (Grable, 2000, p. 625). From literature we know that, aside from socioeconomic and demographic factors such as income, gender or education level, IRT is also affected by a number of psychological factors. Four of these seem particularly relevant and have the useful quality of being behaviors that can be measured. They can help objectify the subjective:

1 Risk taking propensity (RITA)

The basic tenet here is that the more a person wants to take risk and the more experienced a person is with risky activities, the more risk this person will be able to tolerate (Lo & Repin, 2002; Lo, Repin, & Steenbarger, 2005).

2 Emotional control/loss aversion (EMCO)

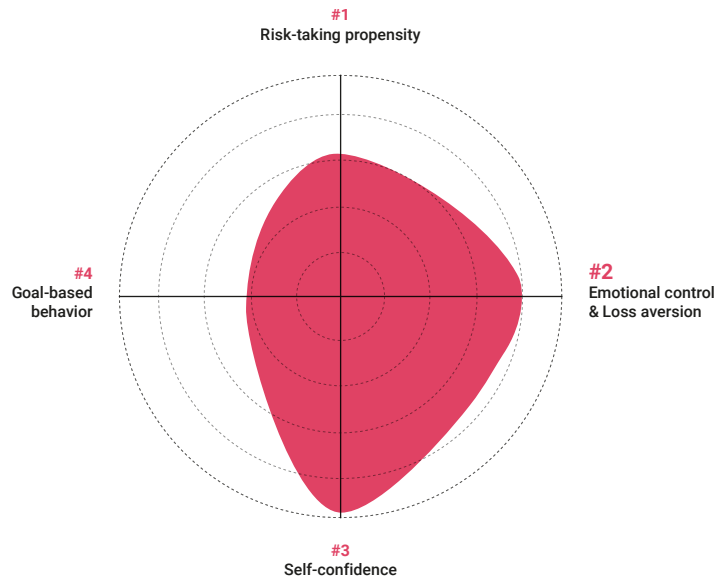
Loss aversion is generally accepted as an important factor in risk taking behavior and was famously associated with prospect theory by Kahneman and Tversky. Loss aversion bias is the tendency of people to prefer avoiding losses to acquiring equivalent gains. The higher this loss aversion bias, the less an investor will be able to tolerate risk.

3 Self-confidence (SECO)

There is ample academic research that proves that self-confidence has a consistent positive relationship with financial risk tolerance. A person with high self-esteem, tends to be more confident and able to take more risk than someone with low self-esteem (Grable & Joo, 2004).

4 Goal-based behavior (GOBE)

As proven in many experiments on the subject of mental accounting, people can have different risk appetites with different pools of assets or goals to achieve. They might be willing to take on considerable risk with assets they earmark for one particular goal, but not with others.



Combined, these factors shape IRT as a four-dimensional psychological construct. A simple model that facilitates a pragmatic approach to assessing people's feelings about risk.

THE RIGHT QUESTIONS FOR EACH DIMENSION

A second step would be to determine which questions to ask in order to 'adequately' assess each of the IRT dimensions. An important sidenote here is that more than one set of questions is needed to deal with, for example, re-assessments. These alternative questions should be markedly different but their outcome (the client's risk profile) must remain consistent. Together with a Danish financial institution and researchers from the Maria Curie Sklodowska University (UMCS) in Poland, we created such a main and alternative questionnaire. The process involved three steps:

Longlist of questions

Starting with desk research, we looked at both the literature on the subject and the many examples available in the financial securities industry today to compile a longlist of potentially useful questions to assess each of the four dimensions. This longlist was further trimmed, based on criteria inspired by the ESMA guidelines. For each question we asked ourselves:

- Is it clear and not misleading? (questions must be impossible to misunderstand)
- Is it definitive and not open-ended? ('no reply' should never be an option)
- Is it impartial and not orienting? (no nudges, suggestive language, priming, ...)
- Is it specific and not abstract? (the subject is investing, not gambling or outdoor sports)
- Is it personal and not generic? (using client data to personalize where possible)
- It is easy and not complex (straight answers require low cognitive load.)

A total of 74 questions passed this initial screening and were then categorized as either RITA, EMCO, SECO or GOBE.

Expert review

Three researchers in the domain of decision making and psychology from Poznan University of Economics and Business, University of Wroclaw and the Catholic University of Lublin reviewed the questions and their categorization. In this process, questions were either accepted, rejected or edited to better suit their purpose. Based on expert consensus, some of the 46 remaining questions were also re-allocated to another IRT dimension than originally planned. The peer review proved to be trustworthy, showing an interclass correlation coefficient[2] of 0.72, indicating a moderate to good consistency of peer-ratings.

Psychometric validation

The 46 questions were then tested with an online panel of n=343 Danish and n=403 Polish citizens aged 21 years or older. Soft quota were used to guarantee sufficient male/female, younger/older and experienced/inexperienced investors in the sample. The data analysis followed a 3-step procedure to create the final selection of essential IRT questions:

1. Exploratory factor analysis

The exploratory factor analysis revealed that there are, indeed, a select number of variables at play and that the experts managed to correctly relate (with only a few exceptions) the questions to the four dimensions of IRT.
We also discovered a fifth variable which we called Asset Allocation (ALLO). However, it proved impossible to formulate a specific, coherent theme with these questions. At least not one that would make sense for constructing a logical questionnaire. Based on the factor analysis we created a shortlist of 26 questions that sufficiently load the different dimensions of IRT. With these questions we are confident to measure what we are supposed to measure, and not something else.

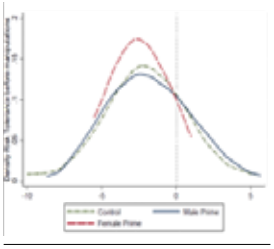
Based on the factor analysis we created a shortlist of 26 questions that sufficiently load the different dimensions of IRT. With these questions we are confident to measure what we are supposed to measure, and not something else.

Final category	Initial category	Question	Factor RITA	EMCO	SECO	GOBE	ALLO
Risk-taking propensity **RITA**	Risk-taking propensity	RT005	.679				
	Risk-taking propensity	RT056	.678				
	Risk-taking propensity	RT039	.674				
	Risk-taking propensity	RT026	.661				
	Risk-taking propensity	RT204	.582				
	Risk-taking propensity	RT202	.558				
	Risk-taking propensity	RT101	.525				
Emotional control **EMCO**	Emotional control	RT018		.697			
	Loss aversion	RT009		.687			
	Emotional control	RT017		.684			
	Emotional control	RT205		.535			
	Emotional control	RT030		.518			
Self-confidence **SECO**	Self-confidence	RT102-Knowledge			.838		
	Self-confidence	RT102-Experience			.838		
	Self-confidence	RT037			.724		
	Self-confidence	RT102-Intuition			.708		
	Self-confidence	RT042			.644		
Goals based behavior **GOBE**	Risk-taking	RT014				.650	
	Goals based behavior	RT027				.645	
	Goals based behavior	RT203				.608	
	Goals based behavior	GO001				.548	
	Goals based behavior	RT206				.503	
Asset allocation **ALLO**	Risk-taking propensity	RT048					.640
	Loss aversion	RT021					.568
	Financial status	RT010					.564
	Risk-taking propensity	RT045					.546

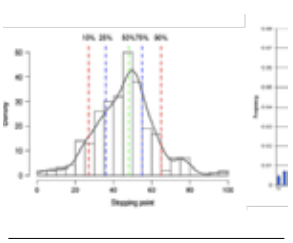
2. Reduction-with-best-fit ananlysis
From literature we know that IRT distributions ideally look like the examples below. They show:

- a Gaussian distribution with a strong bunching near the median.
- a slightly right skewed distribution with more observations on the left (low risk) than on the right side (high risk) of the median.
- a smooth curve with few local maximums or minimums, when based on large sample sizes.

When we calculated the survey results from the 26 selected questions (using unified scales where 1 is very small and 100 is very large) the distribution did indeed look as expected.

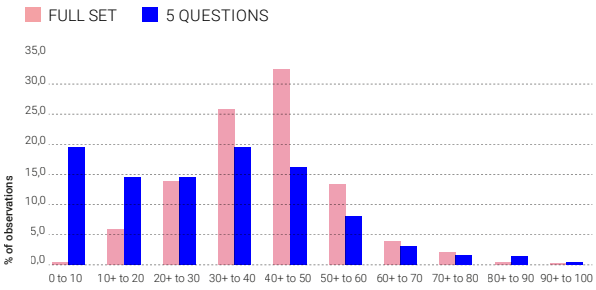


F. D'Acunto. Risk tolerance of men and women. VOX CEPR Policy Portal. 20 September 2015, voxeu.org



P. Crosetto, A. Filipin. The "bomb" risk elicitation task. Journal of Risk and Uncertainty (2013) 47:31-65

The pink bars in the graph below illustrate this beautifully. The blue bars on the other hand, show the distribution based on a set of 5 IRT questions from the Danish financial institution we partnered with for the research. The reason for this unusual distribution could be that, even though our reviewers had attributed these 5 questions to 3 different dimensions, the factor analysis grouped them all into one single factor. In other words: these 5 questions are 5 different ways of asking the exact same thing. This insight strengthens our view that IRT is not one-dimensional, but -as literature indicates- a multidimensional psychological construct.

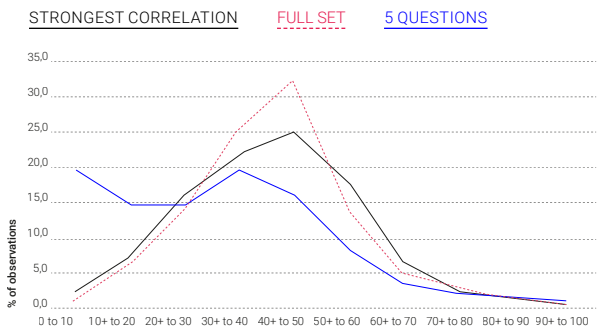


In reality no financial institution will ever ask 26 questions on IRT alone, to onboard a client to a robo-advisor. The questionnaire would become too long and too many potential clients would simply drop out of the process. So the question is: how do we ask as few questions as possible, while minimizing the error of placing an individual into a risk tolerance class, other than the one he or she would end up in, when we use all 26 IRT questions?

How do we ask as few questions as possible, while minimizing the error of placing an individual into a risk tolerance class, other than the one he or she would end up in, when we use all 26 IRT questions?

To solve this issue, we decided to:

- Leave out the 5th dimension (ALLO) because the questions lack a clear theme.
- For each of the 4 remaining factors (RITA, EMCO, SECO and GOBE) select 2 questions such that the corresponding IRT scores have the highest correlation with the full set of 26 questions.
- The graph below shows the distribution of IRT scores we measured for 3 possible scenarios (first survey, n=746). The pink dotted line shows the distribution using all 26 questions over all 5 dimensions. The black line shows the distribution using the results from the 2 strongest correlating questions in each of the 4 relevant dimension. The blue line shows the distribution using the 5 questions proposed by the financial institution.



The research showed that when we apply this reduction-with-best-fit, 72% of potential investors would have a lower than average (50 on a 100 point scale) and 28% a higher than average IRT. The distributions, of course, look different for various subsets of the population. Experienced investors were more prone to risk taking than inexperienced investors. Likewise, the average male investor was willing to accept more risk than the average female investor.

3. Internal consistency analysis

We use the internal consistency of the 4x2 combinations of questions as a measure to guarantee the overall reliability of the questionnaire. A combination with a high internal consistency coefficient (Cronbach's alpha) indicates that the questions are closely related and that people with a similar profile respond to them in a similar way.

In general, the higher the number of questions in the combination, the higher its diagnostic quality. More questions simply yield more solid data. We did, however, find enough combinations of just 2 questions within each of the 4 dimensions of the construct with 'acceptable' to 'good' internal consistency coefficients[3].

It is important to note that the questions in each combination must be distinctively different. Simply repeating a question and (for example) changing the scale for the answers, would no doubt lead to a similar response but the diagnostic quality of this combination would be lost.

Based on our research, we can now create a multitude of alternative combinations with a solid numerical quality. Different pairs of questions that yield the same results and that can be used (for example) for re-assessments.

Knowing which questions to ask is one thing. Another is to know how to ask them in an engaging way, without compromising the reliability of the outcome.

A HUMAN-CENTRIC DESIGN

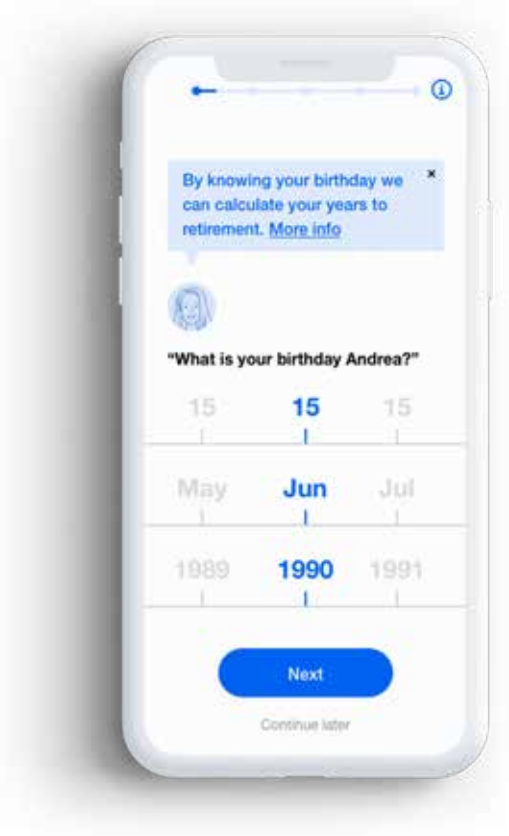
The third step is about applying form to function. It involves designing the user experience of the actual mobile screens, embedding the IRT questions in the end-to-end questionnaire and running a new test to measure the consistency of the outcomes.

UX Design

Unlike in a traditional advisory context, where risk is assessed in face-to-face meetings with clients, robo-advisors must create a digital user experience around the profiling. Ideally, this experience engages the client and creates a state of flow, resulting in high quality data and a low drop-out rate. Because risk profiling is one of those early interactions between the client and the financial institution, the experience should also feel 'on brand' and not like some tedious, administrative obligation. For our research, we created two UX concepts to suit two very different client segments:

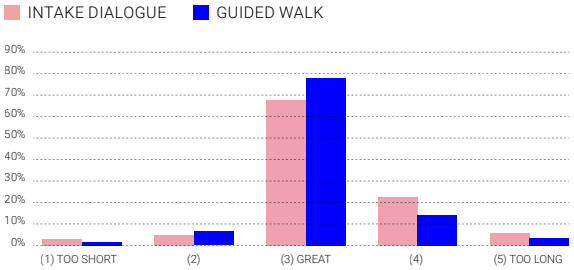
- The first we called 'the guided walk' and was designed for inexperienced investors. Here we used a video host to guide the client throughout the session. When a client clicks the host's picture, an overlay appears with, for example, a clarification of the question, the reason for asking it or an explanation of how the input will be used. A combination of Instagram-like video snippets and plain text is used to provide this type of help and context.
- The second we called 'the intake dialogue' and was designed for more experienced investors. Here we created a fluent Q&A conversation that mimics a chat with an advisor. Questions fly in and out of the screen, one at a time. The client can respond immediately via thumb-friendly input mechanisms to facilitate the data entry.

With these UX concepts we aim to take control of the flow and the friction involved. In the 'guided walk' we decrease the speed to increase the client's comfort. In the 'intake dialogue' we increase the speed to decrease the client's annoyance. Our research has shown that a 'guided walk' of 23 questions takes on average 370 seconds to complete. The 'intake dialogue' with an equal number of questions took 336 seconds. The majority of respondents found the survey length 'just great'. However, one in five



experienced investors would want the 'intake dialogue' to run a bit faster still.

Apart from creating the right user experience, data quality is another major concern when using a digital risk profiler. A number of design principles can be applied to avoid cognitive errors or biased responses. Still, specific controls need to be in place as well.



Controls

As described above, we consider IRT to be a multidimensional psychological construct where four factors are at play. Each of these has two questions which load that factor and have good internal consistency. Still, inconsistent answers occur and must be dealt with in real time. To do this, we first order the answers for each question. Next, a score from 0 to 1 is attached to the answers by distributing them evenly over the unit interval. Mathematically, we define inconsistency as when the score attached to the two answers differs by more than one third. In case of such inconsistency, a third question is instantly triggered to confirm the score. The average of the three questions will then be taken as the truth.

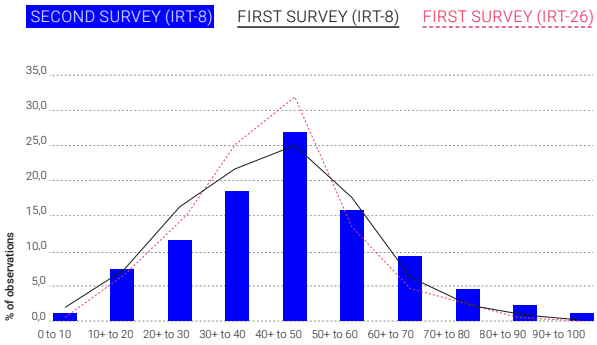
Our research has shown that inconsistencies occur in about 20% of risk profiling sessions. The example below shows the distribution of answers to two questions with different scales (a 3-point scale and a 4-point scale). 75% of respondents answer within the consistency limitations we had set. For 25% of them an additional question is required. To maintain the reliability of the IRT score, this question must obviously load the same factor and the new combination must be internally consistent.

QUESTION 1

QUESTION 2	0	1/2	1
0	22.8%	9.5%	0.8%
1/3	12.5%	9.0%	2.0%
2/3	10.8%	18.8%	6.3%
1	2.3%	0.3%	5.3%

Testing the design

The result from a second survey, where n=400 respondents were presented with the 2x4 IRT questions embedded in a fully designed online risk profiler, shows a similar bell-shaped distribution with the same characteristics we observed after our first survey. Respondents to the designed questionnaire, however, proved slightly more comfortable with risk.



Here, 66% of respondents have a lower and 34% a higher-than-average IRT score. The only differences with the first survey is that respondents were asked less questions (just 23 of which 8 on IRT) and that they were designed as either a 'guided walk' (for 64% of the sample) or an 'intake dialogue' (for 36% of the sample).

CONTINUOUS IMPROVEMENT

At Investsuite we are committed to continuously improve the quality of our risk profiler. We will conduct new research to further improve the numerical quality of IRT questions, aiming for higher load factors for all dimensions of the psychological construct and greater internal consistency coefficients for the combinations of questions we use to measure them. We will also test innovative design concepts to create truly personal user experiences at key moments in the client's financial life. Our goal is to turn what is still viewed by many as an 'obligation to be compliant' into the basis of a prosperous, long-term and dynamic client relationship.

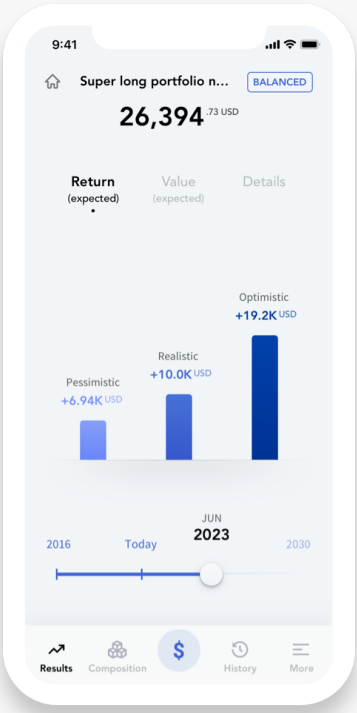
NOTES

[1] ESMA Guidelines on certain aspects of the MiFID II suitability requirements 06/11/2018 ESMA35-43-1163: Robo advice means the provision of investment advice or portfolio management services (in whole or in part) through an automated or semi-automated system used as a client-facing tool.

[2] Depending on the approach, a good strength shows a coefficient above 0.6 (Cicchetti, 1994) or 0.75 (Koo & Li, 2016).

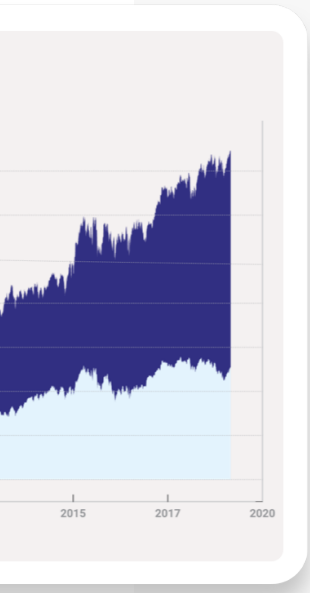
[3] We consider Cronbach alpha coefficients between 0.70 and 0.79 to be acceptable and between 0.80 and 0.89 to be good.

Discover our suite of products:



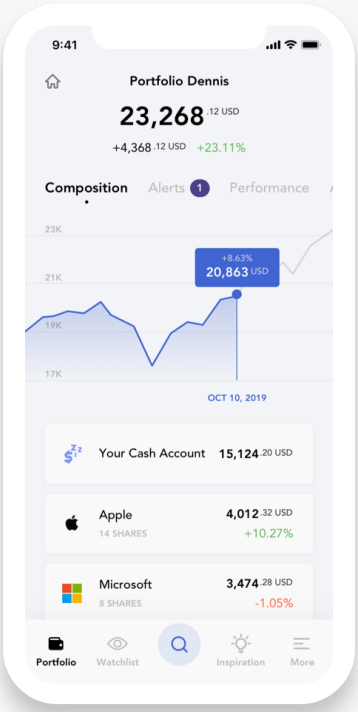
Robo Advisor

A low-cost, customisable digital wealth management tool that can deliver hyper personalised portfolios for Goal Based Investing



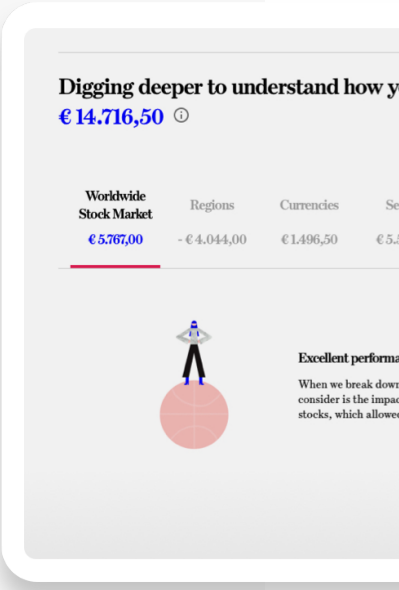
Portfolio Optimizer

The next-generation quant tools that provide cost-effective solutions for more efficient portfolio management



Self Investor

A best-in-class powerful white-label execution-only platform for easy investing



StoryTeller

A worldwide first new way to tell the story on portfolio performance

InvestSuite