Risk attitudes, investment behavior and linguistic variation

Juliana Bernhofer[1], Francesco Costantini[2], and Matija Kovacic*[3]

[1]*Department of Economics, Ca' Foscari University of Venice*
[2]*Department of Humanities, University of Udine*
[3]*Department of Economics, Ca' Foscari University of Venice*

June 10, 2017†

Abstract

This paper explores the relationship between linguistic variation and individual attitudes toward risk and uncertainty. We propose an innovative linguistic marker that classifies languages according to the number of non-indicative moods in the grammatical contexts involving uncertainty. We find that more intensive users of these moods are on average more risk averse. Our marker is then used to instrument individual attitudes toward risk in the model for financial assets accumulation. In addition, by using the Chen (2013) marker as a proxy for subjective discount rate, we disentangle the effects of risk aversion and time preferences on assets accumulation.

Keywords: Language, Uncertainty, Risk Aversion, Time Preferences, Assets Accumulation, Instrumental Variables.

JEL Classification: D14, D81, D91, C36.

*E-mail addresses: juliana.bernhofer@unive.it, fnc.costantini@gmail.com, matija.kovacic@unive.it

*Corresponding author

†We thank Agar Brugiavini, Danilo Cavapozzi, Cristina Orso, Giacomo Pasini, Ca' Foscari University of Venice, Cinzia Di Novi, University of Pavia, and Chiara Strozzi, University of Modena and Reggio Emilia, for their support, useful comments and help.

1 Introduction

Consider a situation in which you have to take a decision about something that has an uncertain prospect, may it be related to sports, health or financial choices. It is clear that your personal characteristics and preferences have an influence on how you evaluate the potential outcomes, but would you argue that the language you speak reveals part of how you perceive risk?

This paper proposes an innovative approach to analyze the individual attitudes toward uncertainty and risky behavior based on the hypothesis of linguistic relativity. The basic principle of linguistic relativity stems from the idea that differences in grammatical structures and the vocabulary may induce speakers of different languages to conceptualize and experience the world differently (Hill and Mannheim, 1992). The research in support of this hypothesis has mainly concentrated on conceptual contents of languages. For instance, Majid et al. (2004) show that the individual perception of space is influenced to some extent by linguistic spatial frames which directly affect the categorization of space in the cognitive domain. On the other hand, Davies and Corbett (1997), Roberson et al. (1999), and Winawer et al. (2007) suggest that words for colors may influence color perception and ability of individuals to distinguish between different types of the same basic color. In a recent paper on cross-country differences in gender political quota, Santacreu-Vasut et al. (2013) show that pervasiveness of gender distinctions in grammar is an important correlate for individual perception of the general role of men and women in the society, which directly influences the extent of regulation of gender political quota.

If speakers of different languages vary in their *worldview* depending on the language they use, some dimensions of linguistic structures may also shape individual preferences and their economic decision-making. The literature on the relationship between linguistic differences and economic behavior, however, is still very poor. To the best of our knowledge, there is only one research article that deals explicitly with the features of linguistic differences and certain aspects of individual economic behavior. In a recent paper on the effect of language on economic behavior, Chen (2013) tests a linguistic-savings hypothesis: when people are required to speak in a distinct way about future events, they take fewer future-oriented actions. The author adopts a future time criterion from typological linguistics discussed in Dahl (2000) and Thieroff (2000), which separates languages into two broad categories: weak and strong Future Time Reference (FTR henceforth) according to how they require speakers to mark the timing of events. Some languages require an explicit verb conjugation in order to distinguish between present and future event (strong FTR languages), while others allow their speakers to talk about the future by using the same verb forms as for present events (weak FTR languages). The author then examines how these differences correlate with future oriented behavior such as saving, smoking, physical activity, and wealth accumulation by retirement. The association between weak future time reference and future oriented behavior is strong: speakers of weak FTR languages save more, accumulate more wealth by retirement, smoke less frequently and are more physically active (and, hence, less obese). According to Chen, seeing the future as overlapping with the present encourages actions that are future-oriented because they are directly connected to the present. On the other hand, by using specific verb forms that identify futurity, events are considered farther away and thus appear to be less relevant.

The approach adopted in this paper is conceptually in line with Chen (2013) since it relies on a weak version of the linguistic relativity hypothesis. However, it departs from Chen (2013) for at least two reasons. First, we propose to consider the linguistic relativity hypothesis on the background of a different grammatical property and in a different economic context, namely *mood* and *uncertainty*. We develop a new linguistic mapping based on the number of grammatical categories (moods) concerned with the expression of uncertainty. We hypothesize that speakers of languages where these specific grammatical forms are used more frequently perceive the world as being more mutable and uncertain with respect to speakers of languages where these forms are less frequently used, or do not exist at all. Our mapping offers a rigorous, and to the best of our knowledge, the first linguistic mapping related to grammatical treatment of uncertainty. Second, we analyze the correlation between our linguistic markers and individual self-declared risk aversion for a large set of individuals from the Survey of Health, Aging and Retirement in Europe (SHARE) and World Value Survey (WVS). We show that a more intensive use of grammatical forms concerned with the expression of uncertainty is strongly correlated with the individual risk preferences, even after including a rich set of explanatory and control variables, and fixed-effects for individual characteristics. The association between our linguistic marker and risk aversion remains significant also for a restricted sample of linguistically heterogeneous countries, as well as for a subsample of the first-generation immigrants. Finally, we estimate a reduced form equation of the probability of investing in risky assets. We use the linguistic marker to instrument the individual attitudes towards risk and quantify a direct effect of risk aversion on the probability of holding risky financial assets. As a final step we estimate the separate effects of time preferences and risk attitudes by using the FTR parameter as a proxy for intertemporal choice behavior, as suggested by Chen (2013). We find that the effect of risk aversion on the probability of holding risky assets is nearly two times larger than the effect of the individual discount rate.

The paper is structured as follows. In the next section we introduce the issue of linguistic relativity and mood, as well as a discussion of the typological distinction used in Chen (2013). In section 3 we exploit the relationship between our linguistic marker and individual attitudes toward risk, and estimate a direct of risk aversion and time preferences on the probability of investing in risky financial assets. Section 4 concludes.

2 Linguistic Relativity and Economic Behavior

The idea that language categories can influence thought has come to be known as *Sapir-Whorf hypothesis* after Sapir (1921) and Whorf and Carroll (1964) and boasts a long history in the philosophy of language and linguistics which can be traced back at least to Humboldt's (1836) idea of *Innere Sprachform*. Following Geeraerts and Cuyckens (2010), the hypothesis of linguistic relativity encompasses two basic notions: the first being that languages are relative as they vary in their expression of concepts, and the second being that the semantic expression of concepts influences, at least to some extent, conceptualization at the cognitive level. Therefore, speakers of distinct languages may perceive reality differently. For instance, the division of the color spectrum varies between languages. Unlike English, Italian speakers distinguish between three kinds of blues ("blu", "azzurro" and "celeste") and Russian makes an obligatory distinction between lighter blues ("goluboy") and darker blues ("siniy") (Winawer et al. (2007)). According to the linguistic relativity hypothesis, different linguistic structures will make Russians and Italians more sensitive to color discrimination than English speakers.

The linguistic relativity hypothesis has generally been interpreted according to two versions. The "strong" one, also known as *linguistic determinism*, states that linguistic categories control general cognitive variables. This version of the hypothesis, however, has generally been refuted (Pinker, 1994). The "weak" version claims that linguistic categories have some effect on cognitive habits, particularly with respect to memory and categorization. The latter version of the Sapir-Whorf hypothesis was taken to be more feasible and has inspired research on topics such as color perception, shape classification, conditional reasoning, number, space, and time categorization.

If speakers of different languages tend to think and behave differently depending on the language they use, some dimensions of linguistic structures may also shape individuals preferences. Chen (2013) represents the first attempt to analyze the impact of language differences on the cognitive domain and consequently on several aspects of individual economic behavior. The empirical analysis in Chen (2013) uses a typological distinction discussed in Dahl (2000) and Thieroff (2000) whereby there are languages that employ a specific verb morphology for FTR, whereas other languages do not. By adopting the weak version of the Sapir-Whorf hypothesis, Chen (2013) hypothesized that this typological divide has an effect on how speakers conceive time. Specifically, speakers of languages that separate the future from the present tense ("strong FTR" languages) are more prone to dissociate the future from the present compared to speakers of languages that do not employ that specific verb morphology when referring to future events ("weak FTR" or "futureless" languages). As a consequence, this may induce people to perceive the future as being more distant and, as a consequence, to undertake fewer future-oriented actions such as saving, smoking, using condoms, accumulating wealth before retirement, and taking initiatives to enhance long-run health. The association between weak FTR and future oriented behavior in Chen (2013) is strong: speakers of weak FTR languages save more, accumulate more wealth by retirement, smoke less frequently and are more physically active.

This notwithstanding, Dahl (2000) and Thieroff (2000)'s classification appears to have some limitations. First, it sorts languages into one of the two categories - strong and weak

FTR languages - within contexts involving prediction, such as weather forecasts - and *only* within these contexts[1]. However, as Chen himself points out, prediction-based contexts are not the only contexts involving FTR. Schedules, plans, ongoing processes having a natural terminus in the future also call for the use of FTR. It follows that there may be different criteria from the one chosen by Dahl and Thieroff to differentiate between languages based on their use of future tenses. If we consider schedules, for instance, English becomes a weak FTR language since it allows for the use of the present tense in reference to future situations planned according to a schedule, like in the sentence "The train leaves at five o'clock.". Chen eventually decided not to use scheduled and law-like events because he relies on the results of Dahl (1985) and Dahl (2000) which show that in many, if not most, languages they are not treated as having future time reference.

We consider Chen (2013)'s idea of linking language features to economic behavior through the linguistic relativity hypothesis appealing enough to propose a reconsideration based on a different grammatical property and in a different economic context, namely the property of mood and the perception of uncertainty. Following a weak version of the linguistic relativity hypothesis, we conjecture that individual levels of risk aversion are influenced by differences in the intensity of use of indicative versus non-indicative (*i.e., irrealis*) moods as they assign a different degree of uncertainty to possible situations. In other words, when describing possible or hypothetical situations, the displacement of the actual from the alternative state of facts is perceived as larger when an *irrealis* mood is used. According to this conjecture, in sentences (1) and (2), for example, the leaving event should be perceived as less uncertain by an English speaker than by an Italian speaker, even though they describe the same possible situation.

(1) *I think he has left.* (English)

(2) *Penso sia partito.* (Italian)
Think-1SG is-SUBJ left
"I think he has left"

The former expresses the leaving situation by resorting to the indicative mood, while the latter has to use a subjunctive (*irrealis*) mood. In general, by using *irrealis* more

[1]The reason for this choice is due to the fact that if a language has an obligatory FTR marking, which is naturally enforced in prediction-based contexts.

intensely, speakers move from the region of certainty to that of uncertainty, in other words, their latent area of the unknown is greater with respect to their peers who speak a less *irrealis*-intensive language. They are expected to be more risk averse as the semantic salience of their region of uncertainty increases.

For this purpose we develop a specific linguistic marker defined on the number of non-indicative moods used in *irrealis* contexts, *i.e.*, contexts that involve grammatical categories concerned with the expression of uncertainty and we relate it to the individual's perception of risk and risky behavior. In what follows we describe the definitions of displacement, modality and mood more in depth, providing also some applied examples and contexts that define our *Irrealis* indicator.

2.1 Displacement and Modality

By displacement semanticists mean the specific characteristic of human language whereby language expressions do not only refer to the here and now but are able to range over future, past, potential, possible and even impossible situations (Hockett (1960), Hockett and Altmann (1968)). In that sense, futurity is an instance of displacement within the temporal dimension. Another crucial dimension of displacement is modality, the grammatical category that indicates whether a sentence expresses a fact, a command, a condition, an opinion, a desire, etc. Consider for instance the following sentences:[2]

(1)	*Wenn*	*es*	*sonnig*	*wäre,*	*ginge*	*ich*	*spazieren.*	(German)
	If	it	sunny	be-KONJ	go-KONJ	I	walk	

"If it were sunny, I would go for a walk."

(2)	*Penso*	*che*	*la*	*riunione*	*sia*	*finita.*	(Italian)
	Think-**1SG**	that	the	meeting	is	finished	

"I think the meeting has finished."

(3)	*Chodźmy*	*do*	*mnie*	*na*	*kawę.*	(Polish)[3]
	Go-**IMP**	to	me	for	coffee	

"Let's go to my place for a coffee."

[2] "KONJ" stems for German *Konjunktiv*; "1SG" for *First Singular*, and "IMP" for *Imperative*.
[3] Swan (2002), pp 242.

By observing sentences (1), (3), and the embedded clause in (2), we notice that they do not concern actual facts, the truth or falsity of the expressions can be decided simply by considering whether the state of facts described in the sentences is true (or false). Sentence (1) does not assert that it is sunny and that the speaker is having a walk; sentence (2) does not assert that the meeting is finished. It may be finished, and the speaker probably believes that it has, but one's belief may turn out to be wrong when actual states of facts are taken into consideration. Sentence (3) does not assert that the speaker is at home having a coffee with the hearer. Sentences (1) to (3) do not refer to actual facts, differently from sentences like "It is sunny today", "I am having a walk", "the meeting has finished", "I'm having a coffee at home with a friend". They refer to possible situations or "possible worlds" (Carnap, 1947), not to real ones. Possible worlds represent alternative state of facts, which cannot be asserted as of the world we actually live in (the "actual world"), and as such they involve the notion of uncertainty.

2.2 Mood and *Irrealis* context of use

Mood is the grammatical category concerned with the expression of situations involving the "world" parameter. What grammarians call *indicative*, for instance, is the mood generally used to assert that a proposition is true as of the actual world.[4] To express possible situations languages can use moods other than the indicative. In sentence (1), for instance, the verbs are in the so-called *Konjunktiv II*. The embedded clause in (2) is in the *Subjunctive*. Sentence (3) is in the *Imperative*. In sentence (2), the English language uses an indicative while Italian uses a non-indicative mood (subjunctive). The difference between indicative and non-indicative moods lies in the fact that they assign a different degree of uncertainty to possible situations.

Some languages have a wide range of morphological moods, some, the most in fact, have a limited number of grammatical categories concerning mood, which are basically the indicative, the imperative and the subjunctive/conditional and others do not have any specific morphological markers for mood[5]. Most importantly, languages may vary as

[4] This does not exclude that the indicative may have modal functions, too.

[5] With the exclusion of nonfinite moods, like the infinitive or the gerund, most Romance languages have four moods according to traditional grammars: the indicative, the subjunctive, the conditional and the imperative. Most Slavic languages have three moods: the indicative, the conditional and the imperative. German has three moods, too: the indicative, the "Konjunktiv" and the imperative. Northern

for the contexts of use of different moods. While in all languages the indicative is the mood used to assert a state of fact and imperative the one to command, the other moods (e.g. subjunctive, conditional, etc.) have different functions and may be used in contexts that vary from language to language. The contexts where *irrealis* moods are used more consistently from a cross-linguistic viewpoint include the following:

- complements of modal predicates (e.g. to be possible, to be likely, to be necessary, to be probable);

  It is probable that these events were coincidences.

- complements of volitional predicates (e.g. to want, to wish, to desire);

  I wish I hadn't been late for school.

- complements of epistemic (non-factive) predicates (e.g. to think, to believe, to doubt);

  I think we should keep a diverse energy portfolio.

- complements of emotive factive predicates (e.g. to regret, to be happy, to be sad);

  I regret that this joke has garnered so much attention.

- complements of declarative predicates (e.g. to say, to tell, to announce);

  I said that one day in my career bad results will come.

- the protasis (the if-clause) in conditional sentences;

  If he had studied harder, he would have passed the exam.

- the apodosis (the main clause) in a conditional sentences.

  If he had studied harder, he would have passed the exam.

Germanic languages have only two moods: the indicative and the imperative-subjunctive mood is also mentioned in some traditional grammars, but it has only residual uses and is no longer productive.

For the purpose of our index, we take the extent of use of different *irrealis* moods in these syntactic contexts as a ratio of use of the different moods in a language in general. We assign a value of 1 to the occurrence of a non-indicative mood in a particular syntactic environment and 0 otherwise. By addition we obtain an indicator (IRR henceforth) of how frequently *irrealis* forms are used in a language, so that languages can be ranked according to the intensity of use of *irrealis* moods.[6]. Finally, languages that do not require *irrealis* moods in any of the context above are called "moodless" languages.

Our linguistic mapping covers 39 languages as listed in Table 3 in Appendix A. Data on grammatical mood were mainly collected from Rothstein and Thieroff (2010) (RT henceforth) as it is the most comprehensive typological survey on grammatical mood in the languages of Europe (see Appendix A for further details). Since not all the data we needed were included in RT, we also needed to rely collect some additional primary data. For this purpose we worked out a questionnaire compiled by a number of linguists throughout Europe[7]. We contacted linguistic experts who were asked to provide a translation of various sentences into their native language and to produce, for each sentence, explanations on which mood they were using in their versions (indicative versus other non-indicative moods to be described).

Table 3 in Appendix A shows that six languages are moodless, whereas three languages use *irrealis* moods in all of the six context. The remaining 27 languages range from two to four contexts in which they employ *irrealis* moods. Thus, significant variation of IRR across languages may represent a good platform for testing the linguistic relativity hypothesis in the context of several economic behaviors involving risk and uncertainty.

3 Linguistic Variation, Risk Attitudes and Investment Behavior

Linguistic relativity suggests that speakers of languages without mood distinctions should perceive the divide between actual and possible situations differently from speakers of languages that have specific markings for mood. Our main hypothesis stems from the

[6]Since there is no qualitative difference between contexts in defining IRR, the index sum was calculated using a uniform weighting function.

[7]Full version available upon request

idea that speakers of languages in which non-indicative moods are used more intensively to express potential situations, perceive the world as being more mutable and hence more uncertain. As a consequence, they are expected to be more risk averse than others, and to engage less in any kind of activity with an uncertain outcome.

To illustrate this mechanism, consider a simple economic system populated by two types of individuals, one speaking a low intensity *IRR* language, L, and another speaking a more intensive *IRR* language, H. Both are endowed with the same level of initial wealth, W_0, and decide on the amount A of their wealth to be invested in a risky asset with an uncertain rate of return $\tilde{r}$. The part of wealth that has not been invested in risky asset pays a risk-free interest r_F. The investment horizon is one period and the individuals solve the following maximization problem:

$$\max_{A_i} EU_i(\widetilde{W}_{t=1,i}) = \max EU_i[(1 + r_F)(W_0 - A_i) + (1 + \tilde{r})A_i], \quad i \in \{L, H\}; \qquad (1)$$

where $\widetilde{W}_{t=1,i}$ is the individual $i's$ wealth at the end of the period, $U(\cdot)$ is a twice-differentiable strictly concave Bernoully utility function, and EU follows the axioms of the Von Neumann - Morgenstern expected utility specification. Since $U' > 0$ and $U'' < 0$, both the L-type and the H-type are strictly risk averse. If we assume that their preferences are represented by

$$U_i(\widetilde{W}) = \frac{\widetilde{W}^{1-\gamma_i}}{1 - \gamma_i}, \; \gamma_i > 0, \quad i \in \{L, H\};$$

with γ_i being the parameter of the individuals' constant relative risk aversion, and for simplicity set r_F equal to zero, the maximization problem in (1) yields the following first order condition:

$$EU_i^{'} = \frac{E\tilde{r}}{(W_0 + A_i E\tilde{r})^{\gamma_i}} = 0. \qquad (2)$$

As long as $E\tilde{r} > 0$, both types of individuals choose to invest a positive amount of wealth in risky assets, *i.e.,* $A_i^* > 0$. Moreover, for any W_0, more risk averse individuals invest less with respect to less risk averse ones.

Following our initial intuition, individuals speaking a more intensive *IRR* language (H-type) should be more cautious and, hence, more risk averse with respect to L-type individuals, *ceteris paribus*. As a consequence, for any given level of wealth, H-type individuals invest less in risky assets with respect to L-type individuals, *i.e.,* $A^*_H < A^*_L$. At this point we can state the following testable hypotheses:

Hypothesis 1 ***Direct Effect of Irrealis on Risk Aversion***
Speakers of languages characterized by a more pronounced displacement into uncertainty (H-type) are more risk averse with respect to speakers of languages with a weaker displacement into uncertainty (L-type), ceteris paribus, i.e., $\gamma_H > \gamma_L$.

Hypothesis 2 ***Indirect Effect of Irrealis on Investment Behavior***
If $\gamma_H > \gamma_L$*, speakers of languages with a more pronounced displacement into uncertainty invest less in risky assets, ceteris paribus, i.e.,* $A^*_H < A^*_L$.

In other words, the intensity of displacement into uncertainty directly influences the individuals' attitudes toward risk, and indirectly their propensity to invest in risky assets. The purpose of the proposed mechanism is not to assert that the intensity of displacement into uncertainty represents the main driver of individuals' attitudes toward risk. What our intuition suggests is that this particular linguistic feature may bias the individuals' perception of uncertainty, and consequently strongly correlate with their degree of risk aversion.

Our empirical exercise proceeds in two steps. We first test a direct association between our linguistic marker and the individuals' self-declared risk aversion. We consider one particular aspect of preferences related to financial risk that the individuals are willing to take when making investments. As a robustness check, we also consider another aspect of individual risk preferences related to adventure risk taking. As a next step, we specify a two stage empirical model in which we use our linguistic marker to instrument risk aversion in the reduced form equation for the probability of holding risky assets. In addition, we disentangle the effects of risk and time preferences by using the *FTR* linguistic marker as a proxy for the individuals' subjective discount rate.

3.1 Data and Methods

Our empirical analysis is run on individuals in the Survey on Health, Aging and Retirement in Europe (SHARE henceforth), Wave 2 - release 2.6.0 and Wave 5 - release 1.0.0.[8] The respondents in SHARE come from 16 European countries and Israel, speaking 16 different languages.[9] We include all the individuals for which we have a complete information on self-declared risk aversion, asset holdings, as well as on socio-economic, family, cognitive and health conditions. In addition to the entire set of countries, we run separate regressions for individuals living in linguistically heterogeneous countries. There are in total 6 linguistically heterogeneous countries speaking 10 different languages with a significant variation of IRR.[10] In order to test the robustness of our baseline models, we also consider the two last waves (Waves 5 and 6) of the World Value Survey (WVS henceforth) for individuals speaking 22 different languages in 47 countries.[11]

As for risk attitude in SHARE, only individuals who are responsible for financial matters in the household (heads of household) are considered. They were asked to answer a simple risk tolerance question:

[8]The variable of financial risk preferences in SHARE, release 2.6.0., is not present in Waves 1 and 4. Wave 3 is a retrospective survey with a different methodology. For those respondents that had been interviewed in both waves, values of the more recent Wave 5 were imputed.

[9]List of countries and languages in Appendix B.

[10]The list of linguistically heterogeneous countries in Appendix B. There are 10 different languages spoken in these countries (IRR in parentheses): Arabic (4), Catalan (3), Dutch (2), Estonian (3), French (3), German (2), Hebrew (0), Italian (6), Russian (4) and Spanish (4).

[11]List of countries and languages in Appendix B.

When people invest their savings they can choose between assets that give low return with little risk to lose money, for instance a bank account or a safe bond, or assets with a high return but also a higher risk of losing, for instance stocks and shares. Which of the statements on the card comes closest to the amount of financial risk that you are willing to take when you save or make investments?

(1) Take substantial financial risk expecting to earn substantial returns;

(2) Take above average financial risks expecting to earn above average returns;

(3) Take average financial risk expecting to earn average returns;

(4) Not willing to take any financial risk.

Individuals who answered (1) and (2) are considered as risk lovers. The intermediate risk takers are those who answered (3) while all the individuals who answered (4) are considered as risk averse. In our sample, 75.13% of individuals declare to be risk averse, 20.70% of individuals are ready to take average financial risks, and only 4.17% of individuals are willing to take above average or substantial financial risk.

Differently from SHARE, the question on risk aversion of the WVS refers to adventure risk taking.[12] The risk preference questions in SHARE and WVS, hence, refer to two different types of risk attitudes.[13] As a consequence, the distribution of respondents in WVS differ from the one in SHARE with roughly 19% of individuals being highly averse to any adventure and risk taking and 22% of respondents that classify as risk lovers. In addition, the information on stock holdings is available only in SHARE. Since the features of risk preferences elicited in SHARE fit particularly well with the nature of our research question, and WVS does not contain any information on individuals' asset holdings, we use SHARE as a primary source of data.

There is a substantial difference in the language variable treatment between SHARE and WVS. While in WVS the individuals are asked to declare the language they normally speak at home, in SHARE the individuals living in countries with two or more official

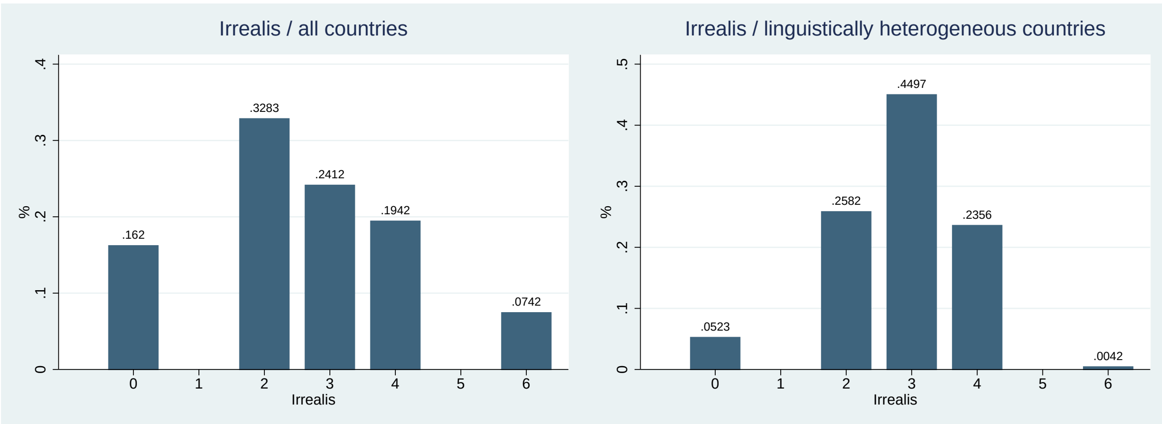
[12]Moreover, the information on risk attitude in WVS includes all the respondents and not only the heads of household.

[13]The risk preference question from WVS is reported in Appendix B.

languages are given the possibility to choose whether to compile the questionnaire in one language or another. We assume that the language in which the questionnaire is compiled is also an individual's primary language. In order to avoid confusing differences in an individual's primary language with differences between natives and immigrants, we exclude the first-generation immigrants from our main analysis. However, as an additional robustness check we also test the association between IRR and individuals' attitude towards risk on the subsample of first-generation immigrants in SHARE who were assigned with the language of their country of origin.[14]

Figure 1 shows the distribution of native speakers by IRR for the entire set of countries (left-hand side figure) and for a restricted set of linguistically heterogeneous countries (right-hand side figure).[15]

Figure 1: Distribution of IRR users in all countries and in linguistically heterogeneous countries.



Source: SHARE, Wave 2 and Wave 5. N. Observations: 79 204 (all countries); 24 785 (linguistically heterogeneous countries).

Roughly 16% of all respondents are moodless speakers, slightly more than 50% are intermediate IRR users, while 27% are intensive and very intensive IRR users. Similarly, the majority of individuals in linguistically heterogeneous countries are intermediate IRR users, 5% do not use IRR at all, while 24% classify as intensive and very intensive IRR

[14]This additional pool of immigrants consists of 5 525 respondents living in 16 countries. As their original language, we considered the official language spoken in the country of birth, if available, or the language spoken by more than 80% of the population. Immigrants from linguistically heterogeneous countries, such as Switzerland and Belgium, were excluded from the analysis.

[15]Distribution of IRR by country in Appendix A, Table 4.

users.[16]

Finally, Tables 1 and 2 show the distribution of individual attitudes toward risk for each value of IRR in the entire set of countries. Table 1 considers the entire range of IRR (from 0 to 6), while in Table 2 we classify IRR in 3 different categories: CatIRR0 contains no IRR, CatIRR1 refers to intermediate IRR usage (2 or 3 IRR) and CatIRR2 represents an intensive and very intensive IRR usage (4 or 6 IRR).[17]

Table 1: **Risk Aversion by IRR (%)**

	IRR Linguistic Marker					
	IRR=0	IRR=2	IRR=3	IRR=4	IRR=6	Total
Risk Lovers	44.6	23.5	14.6	12.1	5.2	100.0
Risk Neutral	26.6	33.6	19.9	14.2	5.7	100.0
Risk Averse	11.5	33.0	27.6	20.0	8.0	100.0
Total (%)	16.0	32.7	25.4	18.5	7.4	100.0
Total (Obs.)	12 831	26 005	19 107	15 381	5 880	79 204

Source: SHARE, Wave 2 and Wave 5. All countries considered (17).

Table 2: **Risk Aversion by IRR (categorized) (%)**

	Categorized IRR			
	CatIRR0	CatIRR1	CatIRR2	Total
Risk Lovers	44.6	38.1	17.3	100.0
Risk Neutral	26.6	53.5	19.9	100.0
Risk Averse	11.5	60.6	28.0	100.0
Total (%)	16.0	58.2	25.8	100.0
Total (Obs.)	12 831	45 112	21 261	79 204

Source: SHARE, Wave 2 and Wave 5. All countries considered (17).

More than 44% of risk prone individuals also speak a moodless language. Half of the respondents with an intermediate risk attitude also classify as intermediate IRR users. Finally, those who declare to be risk averse are mostly either intermediate or high IRR users.

[16] Italian is the only European language in our sample with the maximum number of IRR (6). Other European languages not included in our sample with IRR=6 are Portuguese and Icelandic.

[17] None of the languages in our dataset has IRR=1 and IRR=5. However, generally these values are admissible.

3.2 Irrealis and Risk Aversion: Empirical Strategy

Our first set of regressions examines the relationship between individual attitudes toward risk and the IRR linguistic marker associated to the individual's primary language (*Hypothesis 1*). The dependent variable RA_i is equal to 1 for an individual declaring to be averse to taking risks and 0 otherwise. The empirical problem consists of estimating the following probit model:

$$P(RA_i) = \Phi(r_i) \equiv \int_{-\infty}^{r_i} \phi(r_i) dr_i$$

where $\phi(r_i)$ is the standard normal density

$$\phi(r_i) = (2\pi)^{-1/2} exp(-r_i^2/2)$$

and

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i. \tag{3}$$

Our main variable of interest IRR_i denotes the number of non-indicative moods in *irrealis* contexts in the individual $i's$ primary language. X_i is the vector of demographic and socio-economic characteristics of individual i, such as gender, marital status, family size, occupation, education and the household's income level. Z_i contains controls for cognitive ability and literacy, level of trust and health conditions. Finally, CW_i is a country-wave fixed effect.

In addition to the probit analysis, we also estimate our model with a set of fixed-effect controls for individual demographic and socio-economic characteristics. In such a way we compare individuals identical on these dimensions, but who differ in their IRR usage. These regressions are estimated using a fixed-effect (conditional) logistic model:

$$P(RA_i) = \frac{exp(r_i)}{1 + exp(r_i)}$$

where:

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \lambda FE_i + \eta_i. \quad (4)$$

FE_i is the set of individual-specific fixed-effects which are divided into a set of exogenous determinants (Age and Gender, Country and Wave) and two groups of factors potentially endogenous to individual preferences toward risk and risky behavior (Income and Education, Marital Status and Number of Children).

Even though the literature on linguistic differences and individual attitudes toward risk is very scarce, there is a large body of evidence on the relationship between attitudes toward risk and several socio-demographic and behavioral characteristics, such as educational attainment, income and wealth endowments, occupational status, age, family size, cognitive and health conditions, and trust. Bellante and Green (2004), Dohmen et al. (2011), Lin (2009) and Riley Jr. and Chow (1992) for instance show that a higher level of education increases risk tolerance. The relationship between risk aversion and income is generally negative.[18] Regarding the occupational status, self-employment correlates significantly with the level of individual risk attitude with entrepreneurs being significantly more risk tolerant than others (Hartog et al. (2002), Siegel and Hoban (1991)). On the other hand, Dohmen et al. (2011) and Hartog et al. (2002) show that an individual's unemployment status does not seem to be relevant to their attitude toward risk. Moreover, a higher level of risk aversion for married individuals is found in Cohen and Einav (2007) and Halek and Eisenhauer (2001), whereas Bellante and Green (2004) and Hartog et al. (2002) do not observe any significant effects in that sense.[19] The number of children is found to increase the level of risk aversion (Dohmen et al. (2011) and Lin (2009)), but family size has a negative effect making individuals more risk tolerant (Siegel and Hoban (1991) and Lin (2009)). Finally, women are found to be more risk averse than men (Barsky et al. (1997), Dohmen et al. (2011), Halek and Eisenhauer (2001), Hartog et al.

[18] However, Barsky et al. (1997) and Hartog et al. (2002) show that this relationship is not linear. Similarly, Guiso and Paiella (2008), Riley Jr. and Chow (1992), and Dohmen et al. (2011) find that the level of risk aversion decreases in wealth. However, Dohmen et al. (2011) and Cohen and Einav (2007) point out that that there may be a reverse causality between attitudes toward risk and wealth since a greater willingness to take risks may *ex ante* lead to higher levels of wealth.

[19] One should consider, as pointed out by Halek and Eisenhauer (2001), that more risk averse individuals could also have an ex ante preference for marriage instead of the opposite flow of causality.

(2002), Cohen and Einav (2007) and Jianakoplos and Bernasek (1998)). Other factors influencing individual risk attitudes include a poor health status and cognitive decline which are shown to play a role in Bellante and Green (2004) and Bonsang and Dohmen (2015) respectively, whereas Hartog et al. (2002) do not find any significant correlation between poor health and risk tolerance.

3.2.1 Results

The empirical estimations of Equation (3) are presented in Tables 7, 8, 9, 12 and 14, Appendix C. In all regression models we calculate the robust standard errors clustered by country. Models 1-5 in Table 7 consider IRR as a limited discrete variable. Model 6 includes the categorized version of IRR. The coefficients associated to IRR are highly significant in all model specifications. The coefficient of CatIRR2 in Model 6 indicates that strong and very strong IRR speakers are on average 12% more risk averse with respect to speakers with no indicative moods in IRR contexts. The results confirm our initial intuition (*Hypothesis 1*): there is a strong association between IRR and the individual attitudes toward risk. Females are on average more risk averse than men, while higher levels of education are generally associated with lower risk aversion. In line with the existing literature we find that wealthier individuals are on average less risk averse than poorer ones, while being unemployed is positively correlated to individual risk aversion. Regarding the occupational status, employment correlates significantly with the level of individual risk attitudes with employed and self-employed individuals being significantly more risk tolerant than others. Also, married individuals and those in registered partnerships are slightly more risk averse than others. Restricting our sample to linguistically heterogeneous countries (Table 8) does not significantly alter the results: a more intensive use of IRR is associated with higher aversion to risk, *ceteris paribus*. In order to control for possible inter-dependences between languages belonging to the same linguistic family and/or subfamily, in Table 9 we run separate regressions for main linguistic families (Indo-European, Semitic, and Uralic) and for three linguistic sub-families (Slavic, Romance and Germanic). Coefficients in Models 1 and 2 indicate a strong and significant association between IRR and individual risk aversion even after controlling for main linguistic families. The effect of IRR remains strong and highly significant even within each linguistic sub-family (Models 3-5).

In Tables 10, 11 and 13 we estimate a conditional logit model with the set of additional fixed-effect controls (Equation (4)). All coefficients are reported as odds-ratios. The results show that even when comparing individuals that are identical on every dimension, speaking a higher-intensive IRR language is associated with significantly higher probability of being risk averse. For instance, the coefficient on IRR in Model 5 shows that among individuals of the same age, gender, income, educational attainment, marital status and number of children, a marginal increase in IRR translates into 21% higher probability of being risk averse. These effects do not change significantly when we restrict our sample to linguistically heterogeneous countries.

In order to rule out the possibility that the effect of IRR is driven by some unobserved country- or region-specific cultural and/or social factors, in Table 12 we run our baseline regressions on the subsample of first-generation immigrants. In such a way, we are able to compare individuals with different cultural and linguistic backgrounds living in the same environment independent of their country of origin. Moreover, by considering only the first-generation immigrants, we significantly increase the variability of IRR within each host country (Table 5, Appendix A). The association between IRR and risk aversion is positive and robust. Immigrants speaking intensive IRR languages (CatIRR2) are on average 11% more risk averse with respect to moodless speakers, while intermediate IRR users have on average 9% more chance to be risk averse. These effects are very similar to those obtained in Tables 7 and 8. Moreover, even when comparing only immigrants of the same age, gender, income and educational attainments in Table 13, those with a higher IRR are significantly more risk averse.[20]

Finally, Table 14 considers individuals from the World Value Survey. As before, risk aversion is a binary coded individual self-declared risk aversion variable equal to 1 if the individual is highly averse to any adventure and risk in general, and 0 otherwise. In all model specifications we control for country and wave fixed effects, as well as for the individuals' subjective health status. Differently from SHARE, WVS does not contain information on individuals' cognitive abilities and number of chronic diseases. In line

[20]Differently from the models in Tables 10 and 11, here we not consider fixed effects for marital status and number of children since there are very few observations within each group defined with respect to these dimensions as well.

with the results from SHARE, the association between IRR and risk aversion is positive and statistically significant. All the other coefficients have the expected sign, except the unemployment variable which is not significantly different from zero.

3.3 Irrealis, Risk Aversion and Stock Ownership: Empirical strategy

The empirical estimation of the causal relationship between risk aversion and asset accumulation may suffer both from reverse causality and an omitted variables problem. Since the returns to financial assets represent a certain form of income, and income and risk aversion are negatively correlated, assets accumulation and risk aversion may be simultaneously determined. Moreover, there are several unobservables that may jointly influence the individual attitudes toward risk, making the self-declared risk aversion variable correlated with the error term. In order to make accurate predictions, we need reliable instruments for measuring individual risk preferences.

According to our hypotheses, linguistic differences directly influence the individual perception of risk and uncertainty (*Hypothesis 1*) and indirectly their investment decisions (*Hypothesis 2*). In other words, language (IRR) affects investment decisions through its direct impact on risk aversion. In light of the empirical evidence in Section 3.2.1 which strongly supports *Hypothesis 1*, the IRR linguistic marker may represent a suitable instrument for individual risk preferences in the reduced form equation for the propensity of holding risky assets.

The empirical problem of linking risk preferences to investment decision making, hence, consists of estimating the following causal relationship:

$$AS_i = \alpha + \beta RA_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i \tag{5}$$

where RA_i denotes the individual i's risk aversion, X_i is the vector of demographic and socio-economic individual characteristics, such as gender, marital status, number of children, occupation, education and household's income level, and Z_i contains controls for cognitive ability, level of trust and health conditions. Finally, CW_i denotes the vector of country-wave controls.

In the first stage we estimate the effects of IRR, socio-economic characteristics and location on individual self-declared risk aversion:

$$RA_i = \alpha + \pi_{i1} IRR_i + \pi_{i2} X_i + \pi_{i3} Z_i + \pi_{i4} CW_i + \zeta_i \tag{6}$$

where IRR_i denotes the number of non-indicative moods in IRR contexts in the individual $i's$ language. By plugging the first stage fitted values in the second stage equation we obtain the reduced form model for asset accumulation:

$$AS_i = \alpha + \beta \widehat{RA_i} + \gamma X_i + \theta Z_i + \rho CW_i + error_i \tag{7}$$

The economic model presented at the beginning of this section suggests an inverse causal relationship between risk aversion and asset accumulation. If the prediction of the model is correct, the empirical validation of (7) should yield a negative coefficient on $\widehat{RA_i}$.

In addition to risk aversion, the individual time preferences represent another fundamental driver of intertemporal decision-making. In order to disentangle the effects of time and risk preferences on asset accumulation, we extend (6) and (7) and consider the FTR linguistic marker from Chen (2013) as a proxy for the individual subjective discount rate. Regardless of the nature of investment choices (investment in risky assets versus savings), separating the effects of risk aversion and intertemporal preferences is not an easy task. Epstein and Zin (1989), for instance, develop a theoretical model flexible enough to allow for the separation between the intertemporal preferences and the attitudes toward risk. The authors propose a class of utility functions that allows each dimension to be parameterized separately and show that this utility representation is equivalent to the CRRA utility whenever the agent's coefficient of risk aversion is inversely related to the time preference parameter. However, Andreoni and Sprenger (2012a) and Andreoni and Sprenger (2012b)) agree that these two aspects of individual preferences cannot be considered as perfect substitutes but also claim that they cannot be completely separated.[21]

[21] In a similar manner, Andersen et al. (2008) stress that the assumption of risk neutrality for individuals that are instead risk neutral, result in upward-biased discount rate estimates. They also point out that the parameter values for risk and time preferences must come from the same population.

Even though the separability of risk and time preferences remains an open question from a theoretical point of view, an empirical implementation of FTR and IRR, the first as a proxy for individual discount rate and the second as an instrument for risk aversion, may represent an interesting attempt to disentangle the effects of these fundamental aspects of individuals preferences on the propensity to invest in risky assets.

3.3.1 Results

The empirical estimation of (6) is presented in Table 15, Appendix C. Only linguistically heterogeneous countries are considered which allows for the comparison of individuals living in the same (or a very similar) environments, but speaking different languages.[22] For the two-stage empirical model in (6) and (7) to work, the IRR linguistic marker must satisfy three basic requirements: a) it must be correlated with the endogenous variable (instrument relevance), b) uncorrelated with the error term (independence), and c) it should not have any direct impact on the probability of holding assets other than through its first stage impact on risk aversion (exclusion restriction). The first stage test statistics in Tables 15 confirm the strength of our instrument. In all model specifications, the F-statistic is higher than a commonly used threshold (10 or 16). Our instrument, hence, is highly correlated with the endogenous variable even after controlling for the effect of other regressors, and for the FTR linguistic marker in Models 5 and 6. Moreover, the Hansen J statistics in Models 3 and 4 confirm that our models are correctly specified and that both $CatIRR$ are exogenous. Even though the exogeneity of the instrument cannot be directly tested, there is no reason to suspect that there is any reverse effect of the propensity to invest in risky assets on the instrument. Since we control for country fixed effects (which capture institutional and other country-specific heterogeneities), trust, education, income, occupational status and health conditions (which may be influenced to some extent by linguistic variation), the exclusion restriction should not be violated. In other words, we can rule out any direct effects of linguistic variation on the propensity to invest in risky assets through omitted variables.

Table 16 shows the second stage estimates from a recursive bivariate probit model. The dependent variable (asset accumulation) equals 1 whenever individuals hold some

[22] As a robustness check we also run our regressions on the full set of countries. The estimation results available upon request.

money in stocks or shares (listed or unlisted on the stock market), and 0 otherwise. Only marginal effects are reported. In all regression models we control for country and wave fixed effects, cognitive abilities and individual health conditions. We report the estimated coefficients for a non-categorized version of the instrument only, since it proves to be a stronger instrument than $CatIRR$.[23] To obtain a direct effect of individual time preferences on asset accumulation, we run separate regressions using the FTR linguistic marker (Chen (2013)) as a proxy for the individual subjective discount rate (Model 2). In order to estimate the separate effects of risk aversion and time preferences on the propensity to invest in risky assets, we reestimate a recursive bivariate probit model using the FTR parameter as a marker for intertemporal choice preferences and the IRR marker as an instrument for risk aversion (Model 3).

The instrumented risk aversion is highly significant. Without controlling for time preferences, for an individual with average characteristics of the population, being highly risk averse reduces the probability of holding risky assets by approximately 11%. All the other coefficients have the expected sign. Increasing the education level from medium to high increases the probability of investing in risky assets by 2%. Income and wealth are positively associated with asset accumulation. Being married or in registered partnership does not correlate significantly with the probability of holding risky assets, whereas having more children is negatively related to asset holding. Also, since females are more risk averse than men, they also invest less in risky assets. As for the individual time preferences, the estimated coefficient on FTR marker in Model 2 shows that individuals with a high subjective discount rate invest 4% less in risky assets with respect to low discounting individuals.

The separate effects of risk aversion and time preferences are shown in Model 3. The effect of the individual discount rate is negative and highly significant while the coefficient on risk aversion reduces by 3 percentage points. The effect of risk aversion in our reduced form model is almost two times larger than the effect of the individual discount rate.

[23]Second stage estimation with $CatIRR$ available upon request.

4 Conclusions

This paper proposes an innovative approach to analyzing individual attitudes toward uncertainty and asset accumulation based on the *Sapir-Whorf* hypothesis of linguistic relativity. We develop a specific linguistic marker defined on the basis of the number of non-indicative moods used in *Irrealis* contexts, *i.e.,* contexts that involve grammatical categories related to the expression of uncertainty. Our empirical exercise consists in testing the hypothesis that speakers of languages in which non-indicative moods are used more frequently perceive the world as being more mutable and uncertain with respect to speakers of languages where these forms are less common, or do not exist at all. The association between our linguistic markers and risk aversion is robust to different model specifications. Individuals speaking languages where non-indicative moods are used more intensively have on average a 12% higher probability of being strongly averse to risk. Even when we compare individuals that are identical on every other dimension, such as gender, education, age, income, marital status, and number of children, a more intensive use of non-indicative moods is associated with significantly higher levels of risk aversion. Moreover, these effects are robust to the restriction of our sample to linguistically heterogeneous countries, and for a subsample of first-generation immigrants.

The approach adopted in this paper is, to the best of our knowledge, the first non-experimental attempt to measure a direct and unbiased effect of risk aversion and individual time preferences on investment in risky financial assets. The results indicate that there is a significant variation in risk attitudes both, across individuals living in the same country and speaking a different language, and across countries. Using our linguistic marker as an instrument for the individuals' self-declared risk aversion we show that being highly risk averse reduces the probability of holding risky financial assets by 11%. In addition to risk preferences, we run separate regressions using the FTR linguistic marker (Chen (2013)) as a proxy for the individual subjective discount rate. We find that both measures are relevant determinants in the decision of investing in stocks. In line with our hypotheses. the level of risk aversion and the preference for current consumption have a negative impact on risky asset holdings. Moreover, by using appropriate techniques, we were able to benefit from the orthogonality of the FTR marker and the *Irrealis* marker which ultimately allowed us to show that the impact of risk aversion is higher in absolute terms, than the impact of the individual discount rate. Since linguistic variation is seen as a trait of individual identity, and can hence be exploited as a source of identity and as a cultural marker, not only at the individual but also at the group level, the results obtained in this paper also shed light on the importance of non-economic factors in shaping individual risk and time preferences, and consequently their economic behavior.

References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

Andreoni, J. and Sprenger, C. (2012a). Estimating Time Prefernces from Convex Budgets. *American Economic Review*, 102(7):3333–3356.

Andreoni, J. and Sprenger, C. (2012b). Risk Preferences Are Not Time Preferences. *American Economic Review*, 102(7):3357–3376.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2):537–579.

Bellante, D. and Green, C. A. (2004). Relative risk aversion among the elderly. *Review of Financial Economics*, 13(3):269–281.

Bonsang, E. and Dohmen, T. (2015). Risk attitude and cognitive aging. *Journal of Economic Behavior & Organization*, 112:112–126.

Breu, W. (2010). Mood in Albanian. *Mood in the Languages of Europe*, pages 447–472.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Number v. 3 in Meaning and Necessity: A Study in Semantics and Modal Logic. University of Chicago Press.

Chen, M. K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. *American Economic Review*, 103(2):690–731.

Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.

Dahl, Ö. (1985). Tense and aspect systems.

Dahl, Ö. (2000). The grammar of future time reference in European languages. In Dahl, Ö., editor, *Tense and Aspect in the Languages of Europe*, pages 309–328. Walter de Gruyter, Berlin.

Davies, I. R. L. and Corbett, G. G. (1997). A cross-cultural study of colour grouping: evidence for weak linguistic relativity. *British journal of psychology*, 88:493–517.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Epstein, L. G. and Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica*, 57(4):937–969.

Geeraerts, D. and Cuyckens, H. (2010). *The Oxford handbook of cognitive linguistics*. Oxford University Press.

Giorgi, A. (2009). Toward a syntax of the subjunctive mood. *Lingua*, 119(12):1837–1858.

Guiso, L. and Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic Association*, 6(6):1109–1150.

Halek, M. and Eisenhauer, J. G. (2001). Demography of Risk Aversion. *The Journal of Risk and Insurance*, 68(1):1–24.

Hartog, J., Ferrer-i Carbonell, A., and Jonker, N. (2002). Linking Measured Risk Aversion to Individual Characteristics. *Kyklos*, 55(1):3–26.

Hill, J. H. and Mannheim, B. (1992). Language and World View. *Annual Review of Anthropology*, 21(1):381–406.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203:88–96.

Hockett, C. F. and Altmann, S. A. (1968). A note on design features. *Animal communication: Techniques of study and results of research. Indiana University Press.*, 203:61–72.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2):231–270.

Jianakoplos, N. and Bernasek, A. (1998). Are Women More Risk Averse. *Economic Inquiry*, 10(4):307–324.

Laca, B. (2010). Mood in Spanish. In Thieroff, R. and Rothstein, B., editors, *Mood in the languages of Europe*, Studies in Language Companion Series, pages 198–220. John Benjamins.

Lin, F.-T. (2009). Does the risk aversion vary with different background risk of households? *International Research Journal of Finance and Economics*, 34(34):69 – 82.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3):108–114.

Pinker, S. (1994). The Language Instinct: The New Science of Language and Mind. *Penguin Books*, page 483.

Riley Jr., W. B. and Chow, K. V. (1992). Asset Allocation and Individual Risk Aversion. *Financial Analysts Journal*, 48(6):32–37.

Roberson, D., Davidoff, J., and Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, 71(1):1–42.

Rothstein, B. and Thieroff, R. (2010). *Mood in the Languages of Europe.* Studies in language companion series. John Benjamins.

Santacreu-Vasut, E., Shoham, A., and Gay, V. (2013). Do female/male distinctions in language matter? Evidence from gender political quotas. *Applied Economics Letters*, (September 2012):37–41.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech.* Harvest books. Harcourt, Brace.

Siegel, F. and Hoban, J. (1991). Measuring risk aversion: Allocation, leverage, and accumulation. *Journal of Financial Research*, 14(1):27–35.

Swan, O. E. (2002). *A Grammar of Contemporary Polish.* Slavica.

Thieroff, R. (2000). On the Areal Distribution of Tense-Aspect Categories in Europe. In Dahl, Ö., editor, *Tense and Aspect in the Languages of Europe*, pages 265–308. Walter de Gruyter, Berlin.

Whorf, B. L. and Carroll, J. B. (1964). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* MIT language antropology. M.I.T. Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):7780–7785.

Appendix A: Linguistic Mapping

Table 3: Number of non-indicative moods (IRR) by country

Language	Family	Sub-Family	#Moods	a	b	c	d	e	f	g	IRR
Albanian	Indo-Euro	—	>2	1	1	0	0	0	0	1	3
Arabic	Semitic	—	2	1	1	1	1	0	0	0	4
Basque	Isolate	—	2	1	1	0	0	0	0	1	3
Belorussian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Bulgarian	Indo-Euro	Slavic	1	1	1	0	0	0	0	0	2
Catalan	Indo-Euro	Romance	1	1	1	0	0	0	1	0	3
Croatian	Indo-Euro	Slavic	1	0	0	0	0	0	1	1	2
Czech	Indo-Euro	Slavic	0	1	1	0	0	0	1	1	4
Danish	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Dutch	Indo-Euro	Germanic	0	0	0	0	0	0	1	1	2
English	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Estonian	Indo-Euro	Finno-Ugric	2	0	1	0	0	0	1	1	3
Finnish	Uralic	Finno-Ugric	2	0	0	0	0	0	1	1	2
French	Indo-Euro	Romance	1	1	1	0	1	0	0	0	3
German	Indo-Euro	Germanic	1	0	0	0	0	0	1	1	2
Greek	Indo-Euro	—	1	0	1	0	0	0	0	1	2
Hebrew	Semitic	—	0	0	0	0	0	0	0	0	0
Hungarian	Uralic	Finno-Ugric	2	1	1	0	0	0	1	1	4
Icelandic	Indo-Euro	Germanic	1	1	1	1	0	1	1	1	6
Irish	Indo-Euro	Celtic	2	1	1	0	0	0	1	1	4
Italian	Indo-Euro	Romance	2	1	1	1	1	0	1	1	6
Latvian	Indo-Euro	Baltic	1	1	1	0	0	0	1	1	4
Lithuanian	Indo-Euro	Baltic	1	1	1	0	0	0	1	1	4
Macedonian	Indo-Euro	Slavic	1	1	1	0	0	0	0	0	2
Maltese	Semitic	—	0	0	0	0	0	0	0	0	0
Norwegian	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Polish	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Portuguese	Indo-Euro	Romance	2	1	1	1	1	0	1	1	6
Romanian	Indo-Euro	Romance	1	1	1	0	0	0	1	1	4
Russian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Serbian	Indo-Euro	Slavic	1	0	0	0	0	0	1	1	2
Slovak	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Slovenian	Indo-Euro	Slavic	1	0	1	0	0	0	1	1	3

Language	Family	Sub-Family	# Mood	a	b	c	d	e	f	g	IRR
Spanish	Indo-Euro	Romance	1	1	1	0	1	0	1	0	4
Swedish	Indo-Euro	Germanic	0	0	0	0	0	0	0	0	0
Turkish	Ural-Altaic	Turkic	>2	1	1	1	0	0	1	0	4
Ukrainian	Indo-Euro	Slavic	1	1	1	0	0	0	1	1	4
Welsh	Indo-Euro	Celtic	1	1	0	0	0	0	1	1	3

Notes: *Contexts: a = Modal; b = Desire; c = Attitude (non factive); d = Attitude (factive); e = Declarative; f = Protasis (counterfactual conditional); g = Apodosis (counterfactual conditional).*

Table 4: % Irrealis by Country: Natives (LH countries in bold)

	IRR=0	IRR=2	IRR=3	IRR=4	IRR=6	Total
Austria	0.0	100.0	0.0	0.0	0.0	100.0
Germany	0.0	100.0	0.0	0.0	0.0	100.0
Sweden	100.0	0.0	0.0	0.0	0.0	100.0
Netherlands	0.0	100.0	0.0	0.0	0.0	100.0
Spain	0.0	0.0	**30.6**	**69.4**	0.0	100.0
Italy	0.0	0.0	0.0	0.0	100.0	100.0
France	0.0	0.0	100.0	0.0	0.0	100.0
Denmark	100.0	0.0	0.0	0.0	0.0	100.0
Greece	0.0	100.0	0.0	0.0	0.0	100.0
Switzerland	0.0	**74.9**	**21.8**	0.0	**3.3**	100.0
Belgium	0.0	**52.3**	**47.7**	0.0	0.0	100.0
Israel	**62.3**	0.0	0.0	**37.7**	0.0	100.0
Czech Republic	0.0	0.0	0.0	100.0	0.0	100.0
Poland	0.0	0.0	0.0	100.0	0.0	100.0
Luxembourg	0.0	**46.3**	**53.7**	0.0	0.0	100.0
Slovenia	0.0	0.0	100.0	0.0	0.0	100.0
Estonia	0.0	0.0	**98.6**	**1.4**	0.0	100.0
Total	15.1	33.3	24.4	19.7	7.5	100.0

Table 5: % Irrealis by Country: First-Generation Immigrants

	IRR=0	IRR=2	IRR=3	IRR=4	IRR=6	Total
Austria	3.6	49.9	4.1	39.2	3.3	100.0
Germany	2.2	42.0	0.3	53.0	2.5	100.0
Sweden	19.7	52.8	0.5	25.7	1.4	100.0
Netherlands	10.4	54.7	0.0	31.1	3.8	100.0
Spain	7.0	13.0	8.0	69.0	3.0	100.0
Italy	5.0	22.5	15.0	50.0	7.5	100.0
France	3.4	8.1	7.0	61.4	20.2	100.0
Denmark	37.5	28.1	4.7	23.4	6.3	100.0
Greece	3.4	8.5	10.2	76.3	1.7	100.0
Switzerland	5.1	43.0	17.2	13.9	20.8	100.0
Belgium	4.2	22.3	29.0	17.6	27.0	100.0
Israel	2.4	7.6	1.0	88.1	0.9	100.0
Czech Republic	0.8	10.0	0.4	88.3	0.4	100.0
Poland	0.0	28.1	7.8	64.1	0.0	100.0
Slovenia	0.0	87.8	4.1	1.5	6.6	100.0
Estonia	0.0	0.6	2.3	97.1	0.0	100.0
Total	3.6	21.4	6.0	62.0	7.0	100.0

List of languages surveyed in Rothstein and Thieroff's (2010): 36 Languages

1. seven Germanic languages (Icelandic, Norwegian, Swedish, Danish, English, Dutch and German);

2. six Romance languages (French, Portuguese, Spanish, Catalan, Italian, Rumanian);

3. three Celtic languages (Irish, Breton, Welsh);

4. ten Slavic languages (Russian, Polish, Czech, Slovak, Sorbian, Bosnian, Croatian, Serbian, Bulgarian and Macedonian);

5. two Baltic languages (Latvian and Lithuanian)

6. three other Indo-European languages (Greek, Albanian and Armenian);

7. three Finno-Ugric languages (Finnish, Estonian and Hungarian);

8. four other non-Indo-European languages (Turkish, Maltese, Georgian and Basque).

Regarding the number of finite moods in the languages of Europe, Thieroff (2000) outlines some typological generalizations. First, all the languages have a distinct imperative mood while only one language, Maltese, does not have any non-indicative non-imperative mood. Seven languages (Norwegian, Swedish, Danish, Dutch, English, Irish and Welsh), have one non-indicative non-imperative mood, the subjunctive, but are in the process of losing it. In these languages subjunctive has a very limited use, often restricted to formulaic, almost unproductive forms. As a consequence, the use of indicative in these languages has spread in semantic domains where in previous stages the subjunctive was used. We consider those languages as "moodless". Most languages spoken in Europe have one non-indicative non-imperative mood, the subjunctive or the conditional. This group includes languages such as Breton, Bosnian, Bulgarian, Catalan, Croatian, Czech, French, Georgian, German, Greek, Icelandic, Italian, Latvian, Lithuanian, Macedonian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Sorbian, Spanish. Belarusian, Slovenian, and Ukrainian are not surveyed in RT, but can also be added to this group of languages. Notice that Thieroff follows the view that Romance conditional mood is not an independent mood (Iatridou (2000), Laca (2010)) and classifies it as part of the indicative paradigm. However, opinions vary on this point. Giorgi (2009), for instance, claims that the conditional is a mood of its own. If we assume conditional mood is an independent mood, Romance languages may be classified as two-mood languages. Four languages in RT have two non-indicative non-imperative moods: Armenian (subjunctive and debitive), Estonian (conditional and jussive), Finnish (conditional and potential) and Hungarian (subjunctive and conditional). Finally, two languages are traditionally considered as having more than two non-indicative non-imperative moods: Albanian (Breu, 2010) and Turkish (Optative, Irrealis, Necessitative, Conditional and Possibilitative).

Semantic import of subjunctive and the conditional moods: Examples

Italian subjunctive, for instance, is inflectional, as it adds a bound morpheme to the verb:

(1) *io vengo* *(che) io venga*
I come-IND.1SG (that) I come - SUBJ.1SG

German "Konjunktiv II", on the other hand, is an example of periphrastic mood (at least in weak verbs) formed by an auxiliary and the infinitive:

(2) ich würde kaufen
I **AUX** buy
"I would buy"

Conditional mood in Greek is periphrastic as well, but it resorts to an invariable modal particle (*tha*) followed by the verb in the imperfect or the pluperfect tense (Holton et al. 2004: 123, 151).

(3) tha éhana
PRT lose-IMPF.1SG
"I would lose".

For instance, in counter-factual conditional sentences, (e.g. "If it had not rained, I would have gone for a walk") German has the Konjunktiv II in both the if - clause and in the main clause.

(4) Wenn es nicht geregnet hätte, ginge ich spazieren.
If it not rained AUX.KONJ, go.KONJ I walking
"If it had not rained, I would have gone for a walk."

Italian has the subjunctive in the if-clause only, while the conditional must be used in the main clause.

(5) Se non avesse piovuto, avrei fatto una passeggiata.
If not AUX.SUBJ rained, AUX.KONJ made a walk
"If it had not rained, I would have gone for a walk."

French has the imperfect in the if-clause and the conditional in the main clause.

(6) Si il n'avait pas plu, j'aurais fait une promenade.
If it not AUX.IMP N. rained, I AUX.COND made a walk
"If it had not rained, I would have gone for a walk."

In argument clauses of desire verbs, for instance, some languages (like Italian, see (10)) use the subjunctive, while others (like German, see (11)) use the indicative.

(7) Spero che tu stia bene.
Hope-1SG that you be-SUBJ well.
"I hope you are fine".

(8) Ich hoffe dass es dir gut geht.
Hope-1SG that you be-SUBJ well.
"I hope you are fine".

In argument clauses of declarative verbs, some languages use obligatorily the indicative, while others can optionally use the subjunctive.

(9) Mi ha detto che sta bene.
Me has told that stays well
"He told me he's fine."

(10) Er sagte mir, es gehe ihm gut.
He told me, it goes him well
"He told me he's fine."

Appendix B: Summary Statistics

Table 6: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Risk aversion	3.703	0.576	1	4	75273
Strong_RA	0.755	0.43	0	1	75273
No IRR Moods	0.166	0.372	0	1	84455
2 IRR Moods	0.323	0.468	0	1	84455
3 IRR Moods	0.254	0.435	0	1	84455
4 IRR Moods	0.187	0.39	0	1	84455
6 IRR Moods	0.07	0.256	0	1	84455
No IRR Moods	0.166	0.372	0	1	84455
2 or 3 IRR Moods	0.577	0.494	0	1	84455
4 or 6 IRR Moods	0.258	0.437	0	1	84455
Strong_IRR	0.258	0.437	0	1	84455
Income	4.506	2.871	0	9	82720
Owner	0.479	0.5	0	1	84455
EduCat	0.793	0.771	0	2	84396
Trust People	5.796	2.413	0	10	81503
Married	0.719	0.45	0	1	84452
HH Size	2.195	1.402	0	17	83403
AgeCat	62.17	10.738	40	90	84454
Sex	0.556	0.497	0	1	84455
Retired	0.543	0.498	0	1	84455
Employed	0.274	0.446	0	1	84455
Unemployed	0.028	0.164	0	1	84455
Disabled	0.037	0.19	0	1	84455
Homemaker	0.093	0.29	0	1	84455
Adl	0.278	0.941	0	6	84151
Iadl	0.444	1.238	0	7	84151
Reading	3.717	1.113	1	5	73848
# Chronic	1.743	1.581	0	14	83987

List of countries - SHARE: Austria, Germany, Sweden, Netherlands, Spain (LH), Italy, France, Denmark, Greece (Wave 2 only), Switzerland (LH), Belgium (LH), Israel (LH), Czech Republic, Poland (Wave 2 only), Luxembourg (Wave 5 only) (LH), Slovenia (Wave 5 only), Estonia (Wave 5 only) (LH). Ireland not considered as no information on income is available for release 2.6.0. "LH" stems fot *linguistically heterogeneous countries.*

List of languages - SHARE (native population): Arabic, Catalan, Czech, Danish, Dutch, Estonian, French, German, Greek, Hebrew, Italian, Polish, Russian, Slovenian, Spanish, Swedish.

List of languages - SHARE (first-generation immigrants): Albanian, Arabic, Belorussian, Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Italian, Latvian, Lithuanian, Macedonian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swedish, Turkish, Ukrainian.

List of countries - WVS: Andorra, Argentina, Australia, Bahrain, Brazil, Bulgaria, Belarus, Canada, Chile, Colombia, Cyprus, Ecuador, Estonia, Finland, Germany, Hungary, Iraq, Italy, Kazakhstan, Jordan, Kyrgyzstan, Lebanon, Libya, Mexico, Moldova, Morocco, Netherlands, New Zealand, Nigeria, Norway, Peru, Poland, Qatar, Romania, Russia, Singapore, South Africa, Spain, Sweden, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Egypt, United States, Burkina Faso, Uruguay, Yemen, (former) Serbia and Montenegro.
Counties with with less than 100 observations excluded from the analysis (Algeria, Azerbaijan, Armenia, Ethiopia, Georgia, Ghana, Hong-Kong, India, Iran, Malaysia, Mali, Rwanda, Slovenia, Zimbabwe, Uzbekistan and Zambia).

List of languages - WVS: Arabic, Bulgarian, Catalan, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Swedish, Turkish, Ukrainian.
Languages with less than 100 speakers excluded from the analysis (Albanian, Basque, Belorussian, Croatian, Macedonian, Maltese and Slovak).

Risk preference question from WVS:

> *Now I will briefly describe some people. (...) would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you?*
>
> Description: **Adventure and taking risks are important to this person; to have an exciting life.**
>
> (1) Very much like me;
>
> (2) Like me;
>
> (3) Somewhat like me;
>
> (4) A little like me;
>
> (5) Not like me;
>
> (6) Not at all like me.

Appendix C: Regression Tables

Table 7: *Probit Model: Risk Aversion, Marginal Effects. Cross-Country Analysis: All Countries.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5	RA 6
IRR	0.037***	0.028***	0.025***	0.024***	0.021***	
	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)	
CatIRR1						0.088***
						(0.011)
CatIRR2						0.118***
						(0.006)
Age	0.007***	0.004***	0.004***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.109***	0.088***	0.092***	0.091***	0.091***	0.091***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Low Education		0.076***	0.072***	0.070***	0.064***	0.064***
		(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
High Education		-0.093***	-0.085***	-0.083***	-0.077***	-0.076***
		(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Income		-0.015***	-0.016***	-0.015***	-0.014***	-0.014***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Owner		-0.040***	-0.040***	-0.039***	-0.037***	-0.037***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Married			0.020***	0.018***	0.018***	0.018***
			(0.007)	(0.007)	(0.007)	(0.007)
Num. Children			0.003	0.003	0.002	0.002
			(0.003)	(0.003)	(0.003)	(0.003)
Trust People			-0.012***	-0.012***	-0.012***	-0.012***
			(0.001)	(0.001)	(0.001)	(0.001)
Retired				0.032***	0.031***	0.031***
				(0.007)	(0.007)	(0.007)
Unemployed				0.044***	0.042***	0.042***
				(0.010)	(0.010)	(0.010)
Permanently sick or disabled				0.060***	0.045***	0.045***
				(0.010)	(0.009)	(0.009)
Homemaker				0.029***	0.026***	0.025***
				(0.009)	(0.009)	(0.009)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	No	No	No	No	Yes	Yes
N. Observations	69913	68987	68297	67478	67303	67303
N. Countries	17	17	17	17	17	17

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories for dichotomous: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 8: *Probit Model: Risk Aversion, Marginal Effects. Cross-Country Analysis: Linguistically Heterogeneous Countries.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5	RA 6
IRR	0.032***	0.024***	0.022***	0.022***	0.019***	
	(0.009)	(0.006)	(0.007)	(0.007)	(0.006)	
CatIRR1						0.076***
						(0.009)
CatIRR2						0.106***
						(0.007)
Age	0.005***	0.003***	0.003***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Female	0.086***	0.071***	0.074***	0.075***	0.075***	0.075***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Low Education		0.077***	0.077***	0.077***	0.070***	0.070***
		(0.007)	(0.005)	(0.005)	(0.005)	(0.006)
High Education		-0.091***	-0.086***	-0.083***	-0.074***	-0.072***
		(0.014)	(0.013)	(0.012)	(0.012)	(0.012)
Income		-0.012***	-0.013***	-0.012***	-0.012***	-0.012***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Owner		-0.054***	-0.056***	-0.053***	-0.051***	-0.052***
		(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Married			0.020***	0.019***	0.020***	0.019***
			(0.006)	(0.006)	(0.006)	(0.006)
Num. Children			-0.008**	-0.008**	-0.008**	-0.008**
			(0.004)	(0.003)	(0.003)	(0.003)
Trust People			-0.009***	-0.009***	-0.008***	-0.008***
			(0.001)	(0.001)	(0.001)	(0.001)
Retired				0.026**	0.027***	0.026***
				(0.011)	(0.010)	(0.010)
Unemployed				0.031***	0.030**	0.030**
				(0.012)	(0.012)	(0.012)
Permanently sick or disabled				0.048***	0.032**	0.032**
				(0.016)	(0.012)	(0.012)
Homemaker				0.012	0.010	0.007
				(0.012)	(0.012)	(0.012)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	No	No	No	No	Yes	Yes
N. Observations	22577	22337	22038	21743	21677	21677
N. Countries	6	6	6	6	6	6

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories for dichotomous variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 9: *Probit Model: Risk Aversion, Marginal Effects. Cross-Country Analysis: Controls for Linguistic Families and Separated Regression for Linguistic Sub-families.*

Risk Averse (d)	RA1	RA2	RA3	RA4	RA5
IRR	0.022***		0.133***	0.028***	0.120***
	(0.008)		(0.011)	(0.002)	(0.002)
CatIRR1		0.082***			
		(0.012)			
CatIRR2		0.117***			
		(0.006)			
Age	0.003***	0.003***	0.004***	0.002***	0.003***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Female	0.091***	0.091***	0.058***	0.064***	0.126***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.007)
Low Education	0.064***	0.064***	0.075***	0.055***	0.070***
	(0.007)	(0.007)	(0.003)	(0.011)	(0.010)
High Education	-0.077***	-0.076***	-0.107***	-0.070***	-0.078***
	(0.009)	(0.009)	(0.017)	(0.006)	(0.013)
Income	-0.014***	-0.014***	-0.009***	-0.013***	-0.017***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Owner	-0.037***	-0.037***	-0.019**	-0.033***	-0.043***
	(0.007)	(0.007)	(0.010)	(0.009)	(0.013)
Married	0.018***	0.018***	-0.002	0.009	0.027***
	(0.007)	(0.007)	(0.019)	(0.006)	(0.009)
Num. Children	0.002	0.002	0.010**	0.001	0.007**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Trust People	-0.012***	-0.012***	-0.012***	-0.008***	-0.014***
	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)
Retired	0.031***	0.031***	0.040***	0.014	0.039***
	(0.007)	(0.007)	(0.005)	(0.013)	(0.014)
Unemployed	0.042***	0.042***	0.080***	0.021	0.048***
	(0.010)	(0.010)	(0.013)	(0.015)	(0.018)
Permanently sick or disabled	0.045***	0.045***	0.047	0.042***	0.044***
	(0.009)	(0.009)	(0.024)	(0.015)	(0.015)
Homemaker	0.026***	0.025***	0.099***	0.027***	0.022
	(0.009)	(0.009)	(0.014)	(0.007)	(0.015)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	Yes	Yes	Yes	Yes	Yes
Control: Linguistic family and sub-family	Yes	Yes	Only Slavic	Only Romance	Only Germanic
N. Observations	67303	67303	10549	19937	29668
N. Countries	17	17	5	6	8

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories for dichotomous variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 10: *Fixed effects: Conditional Logit model. Odds-Ratios. Within-Country Analysis: All Countries.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	1.239***	1.231***	1.205***	1.214**	1.213**
	(0.065)	(0.063)	(0.072)	(0.100)	(0.106)
Retired		1.274***	1.264***	1.225***	1.201***
		(0.053)	(0.053)	(0.049)	(0.049)
Unemployed		1.467***	1.428***	1.341***	1.316***
		(0.079)	(0.077)	(0.116)	(0.118)
Disabled		1.562***	1.500***	1.467***	1.323***
		(0.098)	(0.097)	(0.136)	(0.110)
Homemaker		1.269***	1.269***	1.175***	1.159**
		(0.082)	(0.085)	(0.072)	(0.070)
Owner		0.818***	0.831***	0.816***	0.817***
		(0.037)	(0.037)	(0.049)	(0.050)
Trust People			0.922***	0.917***	0.919***
			(0.006)	(0.006)	(0.006)
Control: Cognitive, Health	No	No	No	No	Yes
Fixed Effects:					
Sex x Age	Yes	Yes	Yes	Yes	Yes
Country x Wave	Yes	Yes	Yes	Yes	Yes
Income x Education	Yes	Yes	Yes	Yes	Yes
MarStatus x Num.Child	No	No	No	Yes	Yes
N. Observations	53958	52629	52217	28718	28633
N. Countries	17	17	17	17	17

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit Model. Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 11: *Fixed effects: Conditional Logit model. Odds-Ratios. Within-Country Analysis: Linguistically Heterogeneous Countries.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	1.239***	1.227***	1.204***	1.213**	1.212**
	(0.069)	(0.070)	(0.078)	(0.109)	(0.114)
Retired		1.298***	1.275***	1.222***	1.212***
		(0.084)	(0.086)	(0.063)	(0.052)
Unemployed		1.369***	1.352***	1.316	1.311
		(0.128)	(0.138)	(0.214)	(0.219)
Disabled		1.446***	1.401**	1.596***	1.464***
		(0.206)	(0.199)	(0.240)	(0.155)
Homemaker		1.085	1.073	0.945	0.940
		(0.119)	(0.121)	(0.098)	(0.094)
Owner		0.688***	0.696***	0.685***	0.686***
		(0.029)	(0.029)	(0.067)	(0.068)
Trust People			0.931***	0.926***	0.928***
			(0.009)	(0.011)	(0.011)
Control: Cognitive, Health	No	No	No	No	Yes
Fixed Effects:					
Sex x Age	Yes	Yes	Yes	Yes	Yes
Country x Wave	Yes	Yes	Yes	Yes	Yes
Income x Education	Yes	Yes	Yes	Yes	Yes
MarStatus x Num.Child	No	No	No	Yes	Yes
N. Observations	15632	15309	15151	7033	7019
N. Countries	6	6	6	6	6

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit Model. Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 12: *Probit Model: Risk Aversion, Marginal Effects. Cross-Country Analysis: All Countries, First-Generation Immigrants.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5	RA 6
IRR	0.024***	0.016***	0.017***	0.016***	0.015***	
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	
CatIRR1						0.074***
						(0.022)
CatIRR2						0.102***
						(0.022)
Age	0.005***	0.003***	0.003***	0.003***	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.076***	0.054***	0.060***	0.056***	0.056***	0.056***
	(0.015)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)
Low Education		0.071***	0.067***	0.066***	0.057***	0.058***
		(0.017)	(0.016)	(0.017)	(0.016)	(0.016)
High Education		-0.062***	-0.059***	-0.054***	-0.048***	-0.048***
		(0.020)	(0.017)	(0.017)	(0.016)	(0.016)
Income		-0.014***	-0.015***	-0.014***	-0.013***	-0.013***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Owner		-0.034***	-0.035***	-0.032***	-0.027**	-0.026**
		(0.010)	(0.009)	(0.011)	(0.011)	(0.011)
Married			0.019	0.017	0.017	0.017
			(0.013)	(0.014)	(0.015)	(0.015)
Num. Children			0.002	0.002	0.001	0.002
			(0.005)	(0.005)	(0.005)	(0.005)
Trust People			-0.008***	-0.007***	-0.007***	-0.007***
			(0.002)	(0.002)	(0.002)	(0.002)
Retired				0.032	0.031	0.031
				(0.025)	(0.026)	(0.026)
Unemployed				0.103***	0.100***	0.100***
				(0.029)	(0.028)	(0.028)
Permanently sick or disabled				0.017	0.006	0.005
				(0.025)	(0.028)	(0.027)
Homemaker				0.034	0.030	0.028
				(0.026)	(0.024)	(0.024)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	No	No	No	No	Yes	Yes
Control: Origin Immigrants	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	4607	4540	4438	4386	4378	4378
N. Countries	16	16	16	16	16	16

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories for dichotomous variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 13: *Fixed effects: Conditional Logit model. Odds-Ratios. Within-Country Analysis: First-Generation Immigrants.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	1.115***	1.112***	1.124**	1.121**	1.107**
	(0.046)	(0.041)	(0.054)	(0.053)	(0.055)
Retired		1.121	1.158	1.152	1.127
		(0.251)	(0.253)	(0.248)	(0.243)
Unemployed		2.042**	2.401**	2.463**	2.467***
		(0.733)	(0.829)	(0.892)	(0.824)
Disabled		1.154	1.076	1.058	0.941
		(0.457)	(0.442)	(0.444)	(0.424)
Homemaker		1.256	1.297	1.265	1.246
		(0.399)	(0.432)	(0.424)	(0.398)
Owner		0.908	0.979	0.979	0.981
		(0.070)	(0.084)	(0.081)	(0.095)
Trust People			0.938***	0.938***	0.941***
			(0.020)	(0.021)	(0.020)
Married				1.127	1.149
				(0.236)	(0.250)
Num. Children				0.972	0.963
				(0.034)	(0.032)
Control: Cognitive, Health	No	No	No	No	Yes
Fixed Effects:					
Sex x Age	Yes	Yes	Yes	Yes	Yes
Country x Wave	Yes	Yes	Yes	Yes	Yes
Income x Education	Yes	Yes	Yes	Yes	Yes
N. Observations	1139	1125	1091	1091	1086
N. Countries	16	16	16	16	16

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Conditional Logit Model. Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 14: *Probit Model: Risk Aversion, Marginal Effects. Cross-Country Analysis: World Value Survey.*

Risk Averse (d)	RA 1	RA 2	RA 3	RA 4	RA 5
IRR	0.025***	0.023***	0.024***	0.023***	0.023***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Age	0.004***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.080***	0.077***	0.074***	0.074***	0.073***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Low Education		0.033***	0.018**	0.016**	0.016**
		(0.008)	(0.008)	(0.008)	(0.008)
High Education		-0.022***	-0.009	-0.007	-0.005
		(0.005)	(0.004)	(0.004)	(0.005)
Income			-0.010***	-0.010***	-0.010***
			(0.001)	(0.001)	(0.001)
Trust People			-0.023***	-0.023***	-0.023***
			(0.006)	(0.006)	(0.007)
Married				0.023***	0.022***
				(0.004)	(0.004)
Num. Children				0.003**	0.003***
				(0.001)	(0.001)
Retired					0.034***
					(0.010)
Unemployed					0.009
					(0.011)
Homemaker					0.008
					(0.005)
Other (students included)					-0.023***
					(0.008)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes
Control: Cognitive	NA	NA	NA	NA	NA
Control: Health	Yes	Yes	Yes	Yes	Yes
N. Observations	64904	64422	59237	59129	58222
N. Countries	47	47	47	47	46

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed.*

Table 15: *IV Stocks: First Stage Linear Estimation and Test Statistics, Linguistically Heterogeneous Countries.*

Risk Averse (d)	FS1_ML	FS2_ML	FS3_ML	FS4_ML	FS5_ML	FS6_ML
IRR	0.030***	0.028***			0.038***	0.036***
	(0.005)	(0.005)			(0.006)	(0.006)
CatIRR1			0.137***	0.131***		
			(0.031)	(0.032)		
CatIRR2			0.164***	0.156***		
			(0.030)	(0.031)		
FTR					-0.028**	-0.027**
					(0.012)	(0.012)
Age	0.002***	0.001***	0.002***	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.090***	0.090***	0.090***	0.090***	0.090***	0.090***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Low Education	0.076***	0.070***	0.076***	0.070***	0.076***	0.070***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
High Education	-0.103***	-0.098***	-0.101***	-0.095***	-0.102***	-0.097***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Income	-0.013***	-0.012***	-0.013***	-0.013***	-0.013***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Owner	-0.063***	-0.062***	-0.063***	-0.062***	-0.064***	-0.063***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Married	0.021***	0.021***	0.020***	0.021***	0.020***	0.021***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Num. Children	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Retired	0.026***	0.027***	0.025***	0.026***	0.025***	0.026***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Unemployed	0.024	0.024	0.024	0.024	0.024	0.024
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Disabled	0.046***	0.038**	0.047***	0.038**	0.046***	0.038**
	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)
Homemaker	0.005	0.004	0.002	0.001	0.002	0.001
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Trust People	-0.007***	-0.007***	-0.008***	-0.007***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Control: Country/Wave	Yes	Yes	Yes	Yes	Yes	Yes
Control: Cognitive, Health	No	Yes	No	Yes	No	Yes
N. Observations	14899	14842	14899	14842	14899	14842
N. Countries	6	6	6	6	6	6
Strong Instrument	**39.13**	**32.66**	**18.92**	**15.63**	**44.16**	**37.52**
Endogenous RA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Overidentification	—	—	**0.5809**	**0.5394**	—	—

Notes: *The dependent variable is "Risk Averse (d)". The method of estimation is ivreg2 (only the first stage estimates reported). Robust standard errors in parentheses. Reference categories for dichotomous variables: CatIRR0 (no IRR Moods), Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed, Weak FTR (FTR = 0).*

Table 16: *IV Stocks: Bivariate Probit, Marginal effects; Linguistically Heterogeneous Countries.*

Pr(Risky Assets)	**MEff_IRR** 2nd Stage RA	**MEff_FTR** Direct MEff	**MEff_FTR** 2nd Stage RA
RiskAversion (d)	**-0.108*****		**-0.077*****
(instrumented)	(0.030)		(0.027)
FTR (d)		**-0.044*****	**-0.044*****
(High disc. rate)		(0.004)	(0.006)
Age	0.001***	0.001	0.001***
	(0.000)	(0.000)	(0.000)
Female	-0.018***	-0.033***	-0.021***
	(0.006)	(0.004)	(0.006)
Low Education	-0.025***	-0.036***	-0.028***
	(0.006)	(0.004)	(0.006)
High Education	0.022***	0.043***	0.032***
	(0.008)	(0.008)	(0.008)
Income	0.010***	0.011***	0.010***
	(0.001)	(0.002)	(0.001)
Owner	0.032***	0.043***	0.033***
	(0.007)	(0.003)	(0.007)
Married	0.008	0.004	0.006
	(0.006)	(0.008)	(0.006)
Num. Children	-0.009***	-0.009***	-0.009***
	(0.002)	(0.002)	(0.002)
Retired	0.023***	0.020**	0.021***
	(0.007)	(0.009)	(0.007)
Unemployed (d)	-0.013	-0.017	-0.014
	(0.013)	(0.015)	(0.013)
Disabled	-0.015	-0.020	-0.016
	(0.012)	(0.010)	(0.012)
Homemaker (d)	0.035***	0.031***	0.030***
	(0.010)	(0.005)	(0.010)
Trust People	0.003***	0.003**	0.002**
	(0.001)	(0.001)	(0.001)
Control: Country/Wave	Yes	Yes	Yes
Control: Cognitive, Health	Yes	Yes	Yes
N. Observations	14842	14842	14842
N. Countries	6	6	6

Notes: *The dependent variable is "Has Stocks (d)". The method of estimation is Recursive Bivariate Probit (only second stage reported). Robust standard errors in parentheses. Reference categories for dichotomous variables: Male, Not Married (divorced, separated, widowed), Medium Education, Employed or Self-Employed, Weak FTR (FTR = 0).*