# Risk attitudes, investment behavior and linguistic variation

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

April 26, 2016

**Abstract**

This paper explores the relationship between linguistic variation and individual attitudes toward risk and uncertainty. Linguistic variation refers to differences in linguistic forms across languages. According to the linguistic relativity hypothesis, differences in grammatical structures and the vocabulary may affect how speakers of distinct languages perceive and think about the world. We develop a specific linguistic marker that classifies languages according to the number of non-indicative moods in *irrealis* contexts in their respective grammars. These grammatical categories express situations involving uncertainty, and the frequency of their use may be closely related to the overall degree of uncertainty perceived by individuals. Using data from the Survey of Health, Aging and Retirement in Europe (SHARE) and World Value Survey (WVS), we show that speakers of languages where non-indicative moods are used more intensively are on average more risk averse. This evidence holds both across countries and within linguistically heterogeneous countries. The results are robust to the inclusion of additional set of regressors and several fixed-effect controls for individual characteristics. Finally, we use our linguistic marker to instrument individual attitudes toward risk in the structural model for financial assets accumulation.

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

**JEL Classification:** D14, D81, D91, C36.

***Corresponding author**. E-Mail: matija.kovacic@unive.it, Phone: +39.041.234.9138.
We thank Agar Brugiavini, Cristina Orso, Danilo Cavapozzi, 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

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. According to Chen, seeing the future as contiguous with the present encourages actions that are future-oriented because they are directly connected to the present. 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 distinction between present and future event (strong FTR languages), while others allow their speakers to talk about the future essentially in the same way as about the 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).

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. Finally, we estimate a structural 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 causal effect of risk aversion on the probability of holding risky financial assets.

The paper is structured as follows. In the next section we introduce the issue linguistic relativity and mood as well as a discussion of typological distinction used in Chen (2013). Section 3 analyzes the relationship between our linguistic markers and individual attitudes toward risk, while in Section 4 we present the results from the IV estimation of the effects of risk aversion on the probability of investing in risky financial assets. Section 5 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. 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' economic behavior. The literature on the relationship between language and economic behavior is still very poor. 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) involves a typological distinction discussed in Dahl (2000) and Thieroff (2000) whereby there are languages that employ a specific verb morphology to refer to future events, whereas other languages do not. By adopting a 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 display future time reference ("weak FTR" or "futurless" 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.

## 2.1 Future Marking

Dahl (2000) and Thieroff (2000) typology followed by Chen (2013) sorts languages into one of the two categories with respect to contexts involving prediction, such as weather forecasts - and *only* with respect to these contexts. They do this because if a language has an obligatory FTR marking, this marking is more naturally enforced in prediction-based contexts. In other words, if a language features FTR marking, it will be used at least in prediction-based contexts. Prediction-based contexts, hence, are the core environment where FTR marking occurs. However, there are other contexts denoting future time reference apart from prediction-based ones. Schedules, plans, ongoing processes having a natural *terminus* in the future are also contexts involving a future time reference. It follows that there may be different criteria from the one chosen by Dahl (2000) and Thieroff (2000) to distinguish languages relative to their use of future tenses. If we consider schedules, the English language turns to be a weak FTR language since it allows to use the present tense for situations having a natural *terminus* in the future, like in the following proposition "*The train leaves at five o'clock.*" Moreover, The World Atlas of Linguistic Structures (WALS henceforth) adopts as a typological criterion the presence of inflectional marking of future/non-future distinction (Dahl and Velupillai, 2013).[1] According to this criterion, the European languages surveyed in Dahl and Velupillai (2013) can be classified as in Table 4 (in the Appendix). Moreover, one may classify languages with respect to the expression of their future time framework assuming as a criterion the presence of *any* grammatical marking of future/non-future distinction, be it inflectional or periphrastic.[2] Following this criterion, the European languages surveyed in Dahl and Velupillai (2013) can be classified as in the third column in Table 4 (in the Appendix). Additionally, one may even classify languages with respect to the expression of future time distinguishing between languages that use a specific morphology to refer to future events - no matter what the event expresses, be it a forecast, a plan, an intention, a schedule, etc., and languages that do not use a specific morphology when referring to future events. Since almost all languages can use present morphology to express plans and intentions, a classification built on this criterion would not be particularly meaningful. And yet, such a taxonomy may be relevant as far as the conceptualization of future time reference is involved, exactly because future events are often actions deliberately decided by an agent.

[1] By inflectional marking we mean the modification of a word through bound morphemes to express grammatical categories. For instance, Italian leggerò "I will read" is made up by adding the bound morpheme - erò to the verb theme *legge* - and is therefore "inflectional". See Dryer and Haspelmath (2013).

[2] Periphrastic marking is the expression of a grammatical meaning through one or more function words (free morphemes). For instance, in English *I will read* is a periphrastic form - the auxiliary *will* is added to the base form of the verb.

From the above discussion it follows that different criteria may result in different classifications. The distinction between "strong" and "weak" FTR languages is just another one (see column 4, Table 4). From a theoretical point of view, there is no self - evident reason to prefer one classification to the others. Moreover, in view of the linguistic relativity hypothesis, it is not obvious that the classification adopted in Dahl (2000) and Thieroff (2000) should be preferred to other classifications. This notwithstanding, we consider Chen'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 *mood* and *uncertainty*. 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 concerned with the expression of uncertainty, and we relate it to the individual's perception of risk and risky behavior.

## 2.2 Displacement and Modality

Future tense is a way *displacement* is achieved in language. 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:[3]

(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.*"

[3]"KONJ" stems for German *Konjunktiv*; "1SG" for *First Singular*, and "IMP" for *Imperative*.

(2) *Penso* *che* *la* *riunione* *sia* *finita.* (**Italian**)

think - **1SG** that the meeting is - **SUBJ** finished

"*I think the meeting has finished.*"

(3) *Chodźmy* *do* *mnie* *na* *kawę.* (**Polish**)[4]

go - **IMP** to me for coffee

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

If we consider 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 in fact 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. To sum up, sentences (1) to (3) do *not* refer to actual ones. They refer to situations whose truth as of the *here* and *now* is not asserted by simply uttering them, which is what we do when uttering sentences like "*It is sunny and I am having a walk*", "*the meeting has finished*", "*I having a coffee at home with a friend*". They refer to *possible* and hence *uncertain* situations.

After Carnap (1947), philosophers of language and linguists refer to the technical notion of "possible worlds" to deal with possible situations. Possible worlds may be defined as alternative states of facts, which cannot be asserted as of the world we actually live in (the "actual world"). Since in semantics the notion of truth is crucial, and yet it does not seem to be trivially involved in examples like (1) to (3), it is generally assumed that in these examples the truth is assessed relatively to a possible world and not to the actual world. Thus, the event of going out for a walk in sentence (1) is not true in the possible worlds where it is sunny. Sentence (2) asserts that the meeting is finished in the possible worlds that are compatible with what the speaker thinks here and now. Similarly, in the world where the

[4] Swan (2002), pp 242.

speaker's invitations are accomplished, the speaker is having a coffee at home with a friend of hers.

## 2.3 Mood

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.[5] 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. In other words, the distance between the actual and alternative state of facts ("*possible worlds*") is perceived as larger when a non-indicative mood is used. In the previous example (sentence (2)), "*has finished*" is perceived as less uncertain than "*sia finita*" even though they describe the same "possible world".

Languages vary a great deal as for the expression of modality. In some languages many morphological categories are involved in the expression of modality; others, the most in fact, have a limited number of grammatical categories concerning mood; some others do not have any specific morphological markers for mood.[6] Among languages having mood distinction, moods can be conveyed in different ways, through inflection, that is, by means of bound morphemes, or through periphrases, by means of free morphemes (auxiliaries or *modal particles*). Most importantly, languages may vary as for the *contexts* of use of the different

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

[6]To illustrate, excluding non - finite moods, like the infinitive or the gerund in many Indo - European languages, 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 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.

moods. While in all the languages indicative is the mood used to assert and imperative to command, the semantic import of subjunctive and the conditional moods may differ a great deal from language to language (see Appendix for examples). For the sake of clarity, we will refer to subjunctive and the conditional moods as *non-indicative* moods.

## 2.4 Contexts of Use

Beside the number of non-indicative moods, languages vary with respect to syntactic contexts of use. Each syntactic context is characterized by a set of rules that govern how words are assembled into meaningful sentences. Since we are interested in possible and, hence, not real situations, we will refer to these environments as *irrealis contexts*. From a cross-linguistic viewpoint the following irrealis contexts trigger non-indicative moods more consistently:[7]

1. **complements of modal predicates** (*i.e.,* to be possible, to be likely, to be necessary): *It's probable that action should be taken to improve the well-being of the captive animals.*;

2. **complements of desiderative and volitional predicates** (*i.e.,* to want, to wish, to desire): *I wish I hadn't been late for school.*;

3. **complements of epistemic (non-factive) predicates** (*i.e.,* to think, to believe, to doubt): *I think we should keep a diverse energy portfolio.*;

4. **complements of emotive-factive predicates** (*i.e.,* to regret, to be happy, to be sad): *I regret that this joke has garnered so much attention.*;

5. **complements of declarative predicates** (*i.e.,* to say, to tell, to announce): *I said that one day in my career bad results will come.*;

6. **the protasis (the if - clause) and the apodosis (the main clause) in a conditional sentences**: *If he had studied harder, he would have passed the exam.*.

[7] A few other environments that require non-indicative moods include some adverbial clauses, which however are not as frequent as the clause types listed here.

We take the extent of use of different non-indicative moods in these syntactic contexts as a ratio of use of the different moods in a language in general. We assign value 1 to the occurrence of a non-indicative mood in each syntactic environment and 0 to indicative moods. Adding the values, we obtain an indicator of how frequently non-indicative forms are used in a language, so that languages can be ranked according to the parameter of use of non-indicative moods. Languages that do not require non-indicative moods in any *irrealis* context are called "moodless" languages.

Our linguistic mapping consists of 39 mostly European languages. Data on grammatical mood were mainly collected from Rothstein and Thieroff (2010) (RT henceforth). RT is the most complete typological survey on grammatical mood in the languages of Europe (see Appendix for details). Since not all the data we needed were included in RT, we worked out a questionnaire compiled by a number of linguists throughout Europe. They 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).[8]

Table 3 (in the Appendix) presents the total number of non-indicative moods for 39 languages considered as well as their distribution across six irrealis contexts. In languages having non-indicative moods, these moods occur in complement of modal predicates. The same holds as for clausal arguments of volitional predicates. As for non-factive attitude predicates, variability can be observed across languages. Some languages employ a non-indicative mood in this context, some languages can use both, the indicative and a non-indicative mood, some others use the indicative (despite they have a non-indicative mood). Variability also holds as for clausal argument of factive attitude predicates, where most languages use the indicative. The same is true for clausal arguments of declarative verbs. Finally, in potential and counter-factual conditional clauses, languages having a non-indicative mood usually have it in the *protasis*, whereas variability results in the *apodosis*, with many languages having an indicative mood in it. There are 6 moodless languages, eight languages use non-indicative moods in only two contexts, and seven languages use the non-indicative moods in

[8]The questionnaire was programmed using the software *Qualtrics*. It is available upon request.

three contexts. Finally, there are 14 languages with four non-indicative moods contexts and only three languages that use non-indicative moods in all of the six contexts. Significant variability of non-indicative moods across languages may hence represent a good platform for testing the linguistic relativity hypothesis in the context of several economic behaviors involving risk and uncertainty.

## 3 Linguistic Differences and Risk Attitudes

As far as mood is concerned, linguistic relativity suggests that speakers of languages where mood distinctions are not there should perceive the divide between actual and possible situation differently from speakers of languages where there are different markings for mood. Our main hypothesis stems from the mechanism according to which speakers of languages where specific non-indicative moods are used more intensively (*i.e.*, in more irrealis contexts) to express potential situations, perceive the world as being more mutable and hence more uncertain than speakers of languages where the indicative is used even to express potential situations. Because of this, more intensive users of non-indicative moods may be more risk averse and tend to avoid risky behaviors, while low intensity users may be more tolerant to risk.

To illustrate this mechanism, consider a simple economic system populated by two types of individuals, one speaking a low intensity IRR language, $W$, and another speaking a more intensive IRR language, $S$. Both types of individuals are engaged in a lottery represented by a cumulative distribution function $F(\cdot)$ with $F(x)$ being the probability of getting less than or equal to $x$. The preferences of $S$-type and $W$-type individuals are represented by the following expected utility function:

$$U_i(F) = \int u_i(x)dF(x), \quad i = \{W, S\};$$

where $u(\cdot)$ is a twice differentiable strictly concave Bernoulli utility function and $U(\cdot)$ is the

Von Neumann-Morgenstern utility function.

Let $z_x$ be a degenerate lottery that gives $x$ for sure. Assume that for any non degenerate lottery $F(\cdot)$ with expected value $E(F) = \int x dF(x)$, both types of individuals strictly prefer $z_x$ to $F$. This means that both $W$ and $S$ are generally strictly risk-averse:

$$\int u_i(x) dF(x) < u_i\left(\int x dF(x)\right) \quad for\ all\ F, \quad i = \{W, S\}.$$

Since both types of individuals are strictly risk averse, they prefer getting $E(F)$ for sure to the lottery $F$. Denoting the certainty equivalent for $W$ and $S$ as the amount such that:

$$u_i(c(F, u_i)) = \int u_i(x) dF(x), \quad i = \{W, S\},$$

the strict concavity of $u$ implies that $c(F, u_W) < E(F)$ and $c(F, u_S) < E(F)$. For all lotteries $F$, individuals of type $S$ are more risk averse than individuals of type $W$ if $c(F, u_S) < c(F, u_W)$. Individuals using non-indicative moods more intensively (type $S$) perceive the world as being more uncertain than individuals that use non-indicative moods less frequently (type $W$). As a consequence, for any given lottery $F$, type $S$ individuals will accept a lower guaranteed return with respect to $W$-type individuals. In other words, the risk preferences of high intensity IRR speakers are characterized by a higher Arrow-Pratt coefficient of absolute risk aversion:

$$ARA_S(x) = -\frac{u_S''(x)}{u_S'(x)} > -\frac{u_W''(x)}{u_W'(x)} = ARA_W(x), \quad \forall x.$$

For a given $\tilde{x}$, the same is true for the coefficient of relative risk aversion, *i.e.*,

$$RRA_S(\tilde{x}) = -\tilde{x}\frac{u_S''(\tilde{x})}{u_S'(\tilde{x})} > -\tilde{x}\frac{u_W''(\tilde{x})}{u_W'(\tilde{x})} = RRA_W(\tilde{x}).$$

## 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.[9], and in the two last waves (Waves 5 and 6) of the World Value Survey (WVS henceforth). The respondents in SHARE come from 16 European countries and Israel, speaking 17 different languages. We include all the individuals for which we have a complete information on self-declared risk aversion, as well as on socio-economic, family, cognitive and health conditions (65 985 out of 75 273 observations). 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 in SHARE speaking 10 different languages with a significant variation of IRR.[10]. As a robustness check, we consider the World Value Survey for individuals speaking 22 different languages in 39 worldwide countries (33 391 observations).

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 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. We do not consider individuals who were not born in the same country or whose parents were not born in the same country, *i.e.*, the first and the second generation of immigrants are excluded in order to avoid confusing differences in an individual's primary language with differences between natives and immigrants.

We consider only individuals who are responsible for financial matters in the household (head of household) who were asked to answer a simple risk tolerance question:

*When people invest their savings they can choose between assets that give low return with*

[9]Three countries (Luxembourg, Slovenia and Estonia) are included only in Wave 5, and other three (Greece, Ireland and Poland) only in Wave 2. We exclude Ireland from our sample since we are missing the data on household income (not imputed in Wave 2).

[10]The list of linguistically heterogeneous countries includes: Spain, Belgium, Israel, Estonia, Luxembourg, and Switzerland. 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).

*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; and **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 highly risk averse. In our sample, 75.47% of individual declare to be highly risk averse, 20.43% of individuals are ready to take average financial risks, and only 4.11% of individuals are willing to take above average and substantial financial risk. We aggregate intermediate and low risk averse individuals in one, low risk averse, category.

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

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

| | **IRR Linguistic Marker** | | | | | |
|---|---|---|---|---|---|---|
| | IRR=0 | IRR=2 | IRR=3 | IRR=4 | IRR=6 | Total |
| Low and Very Low RA | 44.24 | 23.69 | 15.28 | 11.84 | 4.95 | 100.00 |
| Intermediate RA | 26.91 | 33.36 | 20.53 | 13.73 | 5.47 | 100.00 |
| High and Very High RA | 11.65 | 32.41 | 29.07 | 19.33 | 7.55 | 100.00 |
| Total (%) | 16.10 | 32.24 | 26.76 | 17.88 | 7.02 | 100.00 |
| Total (Observations) | 12 121 | 24 271 | 20 140 | 13 459 | 5282 | 75 273 |

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

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

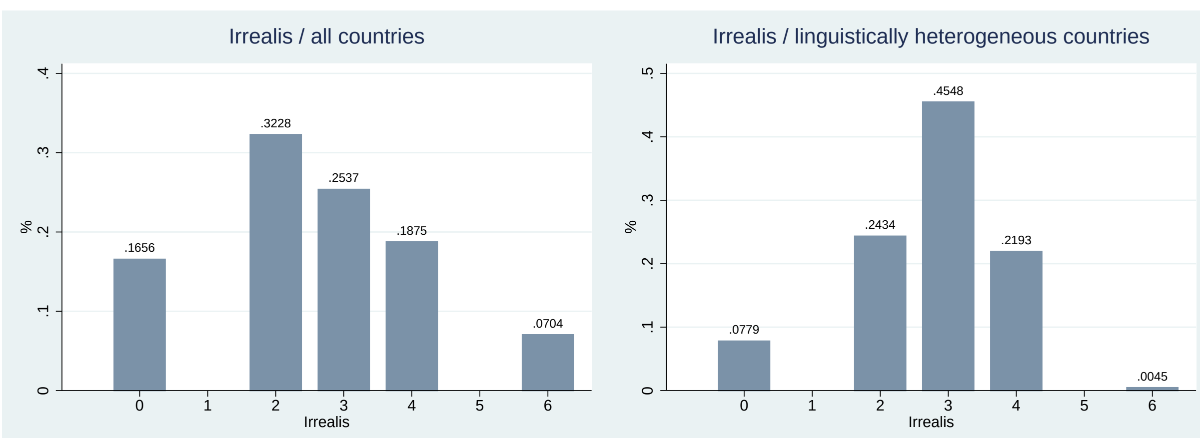
| | **Categorized IRR** | | | |
|---|---|---|---|---|
| | CatIRR0 | CatIRR1 | CatIRR2 | Total |
| Low and Very Low RA | **44.24** | 38.96 | **16.80** | 100.00 |
| Intermediate RA | **26.91** | 53.89 | **19.20** | 100.00 |
| High and Very High RA | **11.65** | 61.47 | **26.88** | 100.00 |
| Total (%) | 16.10 | 59.00 | 24.90 | 100.00 |
| Total (Obs.) | 12 121 | 44 411 | 18 741 | 75 273 |

*Source:* SHARE, Wave2 and Wave 5. All countries considered (17).

More than 44% of IRR moodless speakers classify themselves as risk lovers, 27% are moderately averse to risk, while only 11.65% declare to be highly risk averse. Strong IRR and highly risk averse speakers are almost twice as numerous with respect to strong IRR risk lovers.

Figure 1 shows the distribution of individuals 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).

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




*Source:* SHARE, Wave 2 and Wave 5. N. Observations: 75 273 (all countries); 25 356 (linguistically heterogeneous countries).

Almost half of individuals in linguistically heterogeneous countries are intermediate irrealis users, 6% of individuals do not use IRR at all, while 21% are intensive and very intensive IRR users.[11]

## 3.2 Empirical Strategy

Even though the literature on linguistic differences and individual attitudes toward risk is almost absent, there is a large 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. In addition, Dohmen et al. (2011) consider the family background and find that the fathers' education is negatively correlated with the individual's risk aversion and risky behavior. The relationship between risk aversion and income is generally negative.[12] 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.[13] 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. 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

[11] 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 six IRR are Portuguese and Icelandic.

[12] Pratt (1964) argues in favor of decreasing absolute risk aversion, that is, as wealth and income increases, individuals should become more risk tolerant in absolute terms.

[13] Dohmen et al. (2011) and Cohen and Einav (2007) draw attention to a potential problem of endogeneity as a greater willingness to take risks could *ex ante* lead to higher levels of wealth.

observe any significant effects in that sense.[14] 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)). The relationship between age and risk aversion is less clear. Barsky et al. (1997) show that individuals younger than 55 years and older than 70 years are the most risk averse. Likewise, both Halek and Eisenhauer (2001) and Riley Jr. and Chow (1992), while finding that risk aversion generally decreases in age, they show that after a cut-off of 65 years, risk aversion increases. 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. (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) does not find any significant correlation between poor health and risk tolerance.

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. The dependent variable $RA_i$ is an individual declaring being highly averse to risk taking. The empirical problem consists of estimating the following equation:

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

where:

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i. \tag{1}$$

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

[14] 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.

socio-economic characteristics of individual $i$, such as gender, marital status, family size, occupation, education and 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, we 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 speak a different language (in terms of IRR). These regressions are estimated using fixed-effect (conditional) logistic model:

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

where:

$$r_i = \alpha + \beta IRR_i + \gamma X_i + \theta Z_i + \rho CW_i + \lambda FE_i + \eta_i. \qquad (2)$$

$FE_i$ is the set of individual specific fixed-effects, such as gender, education, age, income, marital status, and number of children.

### 3.3 Results

Empirical estimation of equation (1) is presented in Tables 6, 7, 8 and 11. In all regression models we calculate the robust standard errors clustered by country. Models 1-5 in Table 6 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 on CatIRR2 in Model 6 indicates that strong and very strong IRR speakers are on average 16% more risk averse with respect to speakers with no indicative moods in irrealis contexts. The results confirm our initial intuition: 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 attitude with employed and self-employed individuals being significantly more risk tolerant than others. There is some evidence for married individuals being more risk averse than non married ones but this effect is not very strong. Restricting our sample to linguistically heterogeneous countries (Table 7) does not significantly alter the results: being a strong irrealis user is associated with higher aversion to risk. In order to control for possible inter-dependences between languages belonging to the same linguistic family and/or subfamily, in Table 8 we control for main linguistic families and run separate regressions for three main linguistic sub-families, namely 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 (Indo-European, Semitic, and Uralic). The effect of IRR remains strong and highly significant even within each linguistic sub-family (Models 3-5).

In Tables 9 and 10 we estimate conditional logit model with the set of additional fixed-effect controls (equation (2)). All the coefficients are reported as odds-ratios. The results show that even when comparing individuals that are identical on every dimension, speaking a high intensive IRR language is associated with significantly higher probability of being highly risk averse. These effects do not change significantly when we restrict our sample to linguistically heterogeneous countries.

Finally, Table 11 considers individuals from the World Value Survey. Since the distribution of IRR in WVS differs significantly from that in SHARE, we consider a binary coded IRR variable that equals 1 whenever IRR is greater or equal to 4 (72% of observations) and 0 otherwise. As before, risk aversion is a binary coded individual self-declared risk aversion variable equal to 1 if the individual is highly risk averse and 0 otherwise. In line 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 for unemployment which is negative and not significantly different from zero. The association between IRR and individual risk aversion seems to be very robust.

# 4 Linguistic Variation and Stock Ownership

Many economic decisions involve outcomes that are uncertain or delayed. As a consequence, individual risk and time preferences (impatience) are of fundamental importance for economic modeling. The economic literature suggests that there is an inverse relationship between risk aversion and risky asset accumulation. Moreover, risk aversion is decreasing in income and/or wealth making less risk averse and wealthier individuals more prone to accumulate financial and risky assets as an investment in future income. As for the individual time preferences, the higher the subjective discount rate, the lower the propensity to invest in financial assets with uncertain rate of return.

Consider a simple portfolio choice problem: risky assets versus risk free assets (savings). The economy is populated by $n$ individuals that differ according to their income endowments and risk aversion. Each individual is deciding the amount $AS$ of their income $x_0$ to invest in risky portfolio with uncertain rate of return $r$, and in risk free assets (bank account savings) which pay a certain rate of return. Without loss of generality, assume that the risk free rate of return is equal to zero. Suppose, for simplicity, that the investment horizon is one period. The individual $i's$ wealth at the end of the period is $\bar{x}_{t=1} = x_0 + AS \cdot r$ and s/he solves the following maximization problem:

$$\max_{AS} Eu(\bar{x}_{t=1}) = maxEu[x_0 + AS \cdot r]. \tag{3}$$

Assume that individuals are non satiated and generally risk averse, *i.e.,* $u^{'}(\cdot) > 0$ and $u^{''}(\cdot) < 0$. The necessary and sufficient first order condition for (3) is:

$$E[u^{'}(x_0 + AS \cdot r) \cdot r] = 0 \tag{4}$$

If we denote with $\overline{AS}$ the optimal level of investment in risky assets, (4) implies that risk averse individuals will have a positive amount of risky assets whenever the risky asset has a positive rate of return, *i.e.*, $Er > 0$. Analogously, the probability of investing in risky assets is 0 when $Er = 0$. If we assume:

$$u(x) = \frac{x^{1-\alpha}}{1-\alpha}, \alpha > 0, \tag{5}$$

individuals have a constant relative risk aversion (equal to $\alpha$). For all positive income levels $x$, more risk averse individuals have a lower probability of investing in risky assets with respect to less risk averse individuals (given that their income level is the same). However, absolute risk aversion decreases with $x$ which means that wealthier individuals have lower risk aversion. Wealthier individuals, hence, have lower risk aversion and invest more in risky assets, *ceteris paribus*.[15]

If the system is populated by two types of individuals, identical in all aspects except for the language they speak (less intensive IRR users on one side ($W$) and strong IRR users on the other ($S$)), and if $\alpha_S > \alpha_W$, more risk averse individuals ($S$) invest less in risky assets with respect to less risk averse individuals ($W$) for any given positive income level.

Formally, given (5) and $\alpha_S > \alpha_W$, since $RRA_S(\tilde{x}) = \tilde{x}\alpha_S > \tilde{x}\alpha_W = RRA_W(\tilde{x})$, then $AS^S(\tilde{x}) < AS^W(\tilde{x})$, for all $\tilde{x} > 0$. This result does not depend on measuring risk aversion using the absolute or the relative risk aversion measure, *i.e.*, if $ARA_S(\tilde{x}) = \alpha_S > \alpha_W = ARA_W(\tilde{x})$, then $RRA_S(\tilde{x}) = \tilde{x}\alpha_S > RRA_W(\tilde{x}) = \tilde{x}\alpha W$ and $AS^S(\tilde{x}) < AS^W(\tilde{x})$.

[15]It is not fully clear whether and to what extent economic agent exhibit increasing (IARA), decreasing (DARA) or constant absolute risk aversion (CARA). Pratt (1964) himself argues that utility function with decreasing ARA is " ... logical candidate to use when trying to describe the behavior of people who, one feels, might generally pay less for insurance against a given risk the greater their assets." (pp122-123). Likewise, it seems to be mostly contingent to situational factors whether the coefficient of relative risk aversion is increasing, decreasing or constant. However, as also mentioned in Xie (2000), the most common assumption in existing literature is that of Constant Relative Risk Aversion (CRRA) and as a consequence decreasing absolute risk aversion (DARA).

## 4.1 Empirical Strategy

Risk preferences are typically survey-elicited which makes them potentially endogenous. The empirical estimation of 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 excluded from the model that jointly determine 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. In light of the empirical evidence in Section 3, our main hypothesis stems from the mechanism according to which linguistic differences directly influence the individual perception of risk and uncertainty, and indirectly their investment decisions. The empirical problem consists of estimating the following causal relationship:

$$AS_i = \alpha + \beta RA_i + \gamma X_i + \theta Z_i + \rho CW_i + \eta_i \qquad \text{(EQ4)}$$

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, family size, occupation, education and household's income level, and $Z_i$ contains controls for cognitive ability and literacy, level of trust and health conditions. Finally, $CW_i$ denotes country-wave fixed effects.

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

$$RA_i = \alpha + \pi_{i1} IRR_i + \pi_{i2} X_i + \pi_{i3} Z_i + \pi_{i4} CW_i + \zeta_i \qquad \text{(EQ5)}$$

where $IRR_i$ denotes the number of non-indicative moods in irrealis 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:

$$AS_i = \alpha + \beta\widehat{RA_i} + \gamma X_i + \theta Z_i + \rho CW_i + error_i \tag{EQ6}$$

The economic theory suggests there is an inverse causal relationship between risk aversion and asset accumulation: everything else equal, higher risk aversion reduces the individual propensity to invest in risky assets. If the prediction of the theory is correct, the empirical validation of $EQ6$ should yield a negative coefficient on $\widehat{RA_i}$.

## 4.2 Results

Empirical estimation of EQ5 is presented in Table 12. Only linguistically heterogeneous countries are considered. In such a way we can compare the individuals living in the same (or very similar) environments but speaking different languages. As a robustness check we run our regressions on the full set of countries (Table 13). For the two-stage empirical model in EQ5 and EQ6 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 12 and 13 confirm the strength of our instrument. In all model specifications, the $F$-statistic is significantly 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. 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 the country fixed effects (which capture the institutional and other country specific heterogeneity), 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 effect of linguistic variation on the propensity to invest in risky assets through omitted variables.

Table 14 shows the second stage estimates (EQ6) from a recursive bivariate probit model (Model 1). The dependent variable (asset accumulation) equals 1 whenever individuals hold some money in stocks or shares (listed or unlisted on stock market), and 0 otherwise. The information on individual asset holding comes from the Survey of Health, Aging and Retirement in Europe (SHARE). The WVS dataset does not contain any information on individuals' asset holdings. 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 results a stronger instrument than $CatIRR$.[16] 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 individual subjective discount rate.

The instrumented risk aversion is highly significant in all model specification. 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 3%. Income and wealth are positively associated with assets accumulation. Married couples invest more while having more children is negatively related to asset holding. Finally, since females are more risk averse than men, they also invest less in risky assets. Finally, the estimated coefficients in the last model show that individuals with high subjective discount rate invest 4% less in risky assets with respect to low discounting individuals. The effect of risk aversion on asset accumulation is almost three times larger than the effect of individual discount rate.

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

## 5 Conclusions

This paper proposes an innovative approach for analyzing the 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 concerned with the expression of uncertainty. Our empirical exercise consists in testing the hypothesis that speakers of languages where 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 frequently used, or do not exist at all. The association between our linguistic markers and risk aversion seems to be very robust. The individuals speaking languages where non-indicative moods are used more intensively in grammatical contexts involving uncertainty have on average 16% 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. These effects do not change when we restrict our sample to linguistically heterogeneous countries.

We also investigate the direction of causality from risk aversion to stock asset accumulation using the linguistic marker as an instrument for the individuals' self-declared risk aversion. The results 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 individual subjective discount rate. The effect of risk aversion on asset accumulation is almost three times larger than the effect of individual discount rate.

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. Since linguistic variation is seen as a trait of individual identity, and can be exploited as a source of identity and 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 the individual risk and time preferences, and consequently their economic behavior.

# References


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. (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.

Dahl, Ö. and Velupillai, V. (2013). *Tense and Aspect*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

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.

Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

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 Jounal 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.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32(1):122–136.

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.

Xie, D. (2000). Power risk aversion utility functions. *Annals of Economics and Finance*, 1:265–282.

# Appendix A: Linguistic Mapping

Table 3: Linguistic Mapping

| Language | Family | Sub-Family | #Moods | a | b | c | d | e | f | g | $\sum$ Non Ind. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | $\sum$ Non Ind. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 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: Future Marking

| Language | Inflectional | Future Marking | FTR |
|---|---|---|---|
| Basque | Yes | Yes | Strong |
| Bulgarian | No | Yes | Strong |
| English | No | Yes | Strong |
| Finnish | No | No | Weak |
| French | Yes | Yes | Strong |
| German | No | Yes | Weak |
| Greek | No | Yes | Strong |
| Hungarian | No | Yes | Strong |
| Latvian | Yes | Yes | Strong |
| Maltese | No | Yes | Strong |
| Portuguese | No | Yes | Strong |
| Romanian | No | Yes | Strong |
| Russian | No | Yes | Strong |
| Spanish | Yes | Yes | Strong |
| Swedish | No | Yes | Weak |

**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 5: 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 Irrealis Moods | 0.166 | 0.372 | 0 | 1 | 84455 |
| 2 Irrealis Moods | 0.323 | 0.468 | 0 | 1 | 84455 |
| 3 Irrealis Moods | 0.254 | 0.435 | 0 | 1 | 84455 |
| 4 Irrealis Moods | 0.187 | 0.39 | 0 | 1 | 84455 |
| 6 Irrealis Moods | 0.07 | 0.256 | 0 | 1 | 84455 |
| No Irrealis Moods | 0.166 | 0.372 | 0 | 1 | 84455 |
| 2 or 3 Irrealis Moods | 0.577 | 0.494 | 0 | 1 | 84455 |
| 4 or 6 Irrealis 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 |
| Writing | 3.592 | 1.156 | 1 | 5 | 74574 |
| # Chronic | 1.743 | 1.581 | 0 | 14 | 83987 |

**List of countries**: Austria, Germany, Sweden, Netherlands, Spain (LH), Italy, France, Denmark, Greece, Switzerland (LH), Belgium (LH), Israel (LH), Czech Republic, Poland, Luxembourg (LH), Slovenia, Estonia (LH). "LH" stems fot *linguistically heterogeneous countries*.

# Appendix C: Regression Tables

Table 6: *Probit Model: Risk Aversion, Marginal Effects*

| Risk Aversion (d) | RA 1 | RA 2 | RA 3 | RA 4 | RA 5 | RA 6 |
|---|---|---|---|---|---|---|
| IRR | 0.038*** | 0.033*** | 0.030*** | 0.029*** | 0.027*** | |
| | (0.009) | (0.008) | (0.009) | (0.009) | (0.009) | |
| CatIRR1 (d) | | | | | | 0.129*** |
| | | | | | | (0.014) |
| CatIRR2 (d) | | | | | | 0.161*** |
| | | | | | | (0.005) |
| Age | 0.006*** | 0.004*** | 0.004*** | 0.003*** | 0.003*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Female | 0.107*** | 0.086*** | 0.089*** | 0.089*** | 0.092*** | 0.093*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.008) | (0.008) |
| Low Edu. | | 0.076*** | 0.071*** | 0.070*** | 0.064*** | 0.064*** |
| | | (0.006) | (0.006) | (0.006) | (0.007) | (0.007) |
| High Edu. | | -0.091*** | -0.084*** | -0.082*** | -0.075*** | -0.075*** |
| | | (0.009) | (0.009) | (0.009) | (0.008) | (0.008) |
| Income | | -0.016*** | -0.016*** | -0.015*** | -0.014*** | -0.015*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Owner | | -0.030*** | -0.028*** | -0.026*** | -0.022*** | -0.022*** |
| | | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Married | | | 0.012 | 0.011 | 0.014** | 0.014** |
| | | | (0.006) | (0.007) | (0.007) | (0.007) |
| HH Size | | | 0.003 | 0.002 | 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 | | | | 0.034*** | 0.033*** | 0.033*** |
| | | | | (0.007) | (0.008) | (0.008) |
| Unemployed | | | | 0.050*** | 0.050*** | 0.051*** |
| | | | | (0.009) | (0.009) | (0.009) |
| Disabled | | | | 0.060*** | 0.044*** | 0.045*** |
| | | | | (0.010) | (0.010) | (0.010) |
| Homemaker | | | | 0.030*** | 0.028*** | 0.027*** |
| | | | | (0.008) | (0.010) | (0.010) |
| *Country, Wave dummy* | Yes | Yes | Yes | Yes | Yes | Yes |
| *Cognitive, Health* | No | No | No | No | Yes | Yes |
| *N. Observations* | 74506 | 74148 | 73338 | 71907 | 65985 | 65985 |
| *N. Countries* | 17 | 17 | 17 | 17 | 17 | 17 |

**Notes**: *The dependent variable is "High Risk Aversion". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired. StrongFTR refers to Future Time Oriented classification used in Chen (2013).*
** $p < 0.05$, *** $p < 0.01$.

Table 7: *Probit Model: Risk Aversion, Marginal Effects. Linguistically Heterogeneous Countries*

| Risk Aversion (d) | RA 1 | RA 2 | RA 3 | RA 4 | RA 5 | RA 6 |
|---|---|---|---|---|---|---|
| IRR | 0.033*** | 0.028*** | 0.026*** | 0.026*** | 0.023*** | |
| | (0.008) | (0.007) | (0.008) | (0.007) | (0.008) | |
| CatIRR1 (d) | | | | | | 0.109*** |
| | | | | | | (0.013) |
| CatIRR2 (d) | | | | | | 0.138*** |
| | | | | | | (0.006) |
| Age | 0.005*** | 0.003*** | 0.003*** | 0.002*** | 0.001*** | 0.001*** |
| | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) |
| Female | 0.084*** | 0.068*** | 0.070*** | 0.072*** | 0.074*** | 0.074*** |
| | (0.006) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Low Edu. | | 0.074*** | 0.074*** | 0.074*** | 0.066*** | 0.066*** |
| | | (0.006) | (0.005) | (0.005) | (0.006) | (0.006) |
| High Edu. | | -0.086*** | -0.082*** | -0.079*** | -0.066*** | -0.065*** |
| | | (0.013) | (0.012) | (0.012) | (0.010) | (0.010) |
| Income | | -0.014*** | -0.014*** | -0.013*** | -0.012*** | -0.012*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Owner | | -0.027*** | -0.027*** | -0.026*** | -0.022*** | -0.022*** |
| | | (0.004) | (0.005) | (0.004) | (0.005) | (0.005) |
| Married | | | 0.011 | 0.009 | 0.013** | 0.012** |
| | | | (0.006) | (0.005) | (0.005) | (0.005) |
| HH Size | | | -0.008** | -0.008** | -0.009*** | -0.009*** |
| | | | (0.004) | (0.003) | (0.003) | (0.003) |
| Trust People | | | -0.009*** | -0.009*** | -0.009*** | -0.009*** |
| | | | (0.001) | (0.001) | (0.001) | (0.001) |
| Retired | | | | 0.033*** | 0.034*** | 0.033*** |
| | | | | (0.012) | (0.012) | (0.013) |
| Unemployed | | | | 0.045*** | 0.040*** | 0.040*** |
| | | | | (0.009) | (0.012) | (0.012) |
| Disabled | | | | 0.046*** | 0.030** | 0.031** |
| | | | | (0.016) | (0.014) | (0.014) |
| Homemaker | | | | 0.018 | 0.018 | 0.015 |
| | | | | (0.010) | (0.012) | (0.013) |
| *Country, Wave dummy* | Yes | Yes | Yes | Yes | Yes | Yes |
| *Cognitive, Diseases* | No | No | No | No | Yes | Yes |
| *N. Observations* | 25356 | 25023 | 24664 | 24136 | 22967 | 22967 |
| *N. Observations* | 6 | 6 | 6 | 6 | 6 | 6 |

**Notes**: *The dependent variable is "High Risk Aversion". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired. StrongFTR refers to Future Time Oriented classification used in Chen (2013).*
** $p < 0.05$, *** $p < 0.01$.

Table 8: *Probit Model: Risk Aversion, Marginal Effects. Controls for Linguistic Families and Separated Regression for Linguistic Sub-families.*

| Risk Aversion (d) | RA1 | RA2 | RA3 | RA4 | RA5 |
|---|---|---|---|---|---|
| IRR | 0.028*** | | 0.129*** | 0.032*** | 0.134*** |
| | (0.009) | | (0.006) | (0.005) | (0.003) |
| CatIRR1 (d) | | 0.122*** | | | |
| | | (0.017) | | | |
| CatIRR2 (d) | | 0.160*** | | | |
| | | (0.006) | | | |
| Age | 0.003*** | 0.003*** | 0.004*** | 0.002*** | 0.003*** |
| | (0.000) | (0.000) | (0.001) | (0.001) | (0.000) |
| Female | 0.092*** | 0.093*** | 0.056*** | 0.062*** | 0.133*** |
| | (0.008) | (0.008) | (0.004) | (0.005) | (0.007) |
| Low Edu. | 0.064*** | 0.064*** | 0.073*** | 0.056*** | 0.069*** |
| | (0.007) | (0.007) | (0.004) | (0.009) | (0.011) |
| High Edu. | -0.075*** | -0.075*** | -0.108*** | -0.065*** | -0.079*** |
| | (0.008) | (0.008) | (0.011) | (0.005) | (0.011) |
| Income | -0.014*** | -0.015*** | -0.008*** | -0.013*** | -0.018*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.002) |
| Owner | -0.022*** | -0.022*** | -0.017*** | -0.016*** | -0.026*** |
| | (0.003) | (0.003) | (0.002) | (0.004) | (0.005) |
| Married | 0.014** | 0.014** | -0.007 | 0.006 | 0.022 |
| | (0.007) | (0.007) | (0.018) | (0.007) | (0.012) |
| HH Size | 0.001 | 0.001 | 0.007 | 0.001 | 0.007** |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.003) |
| Trust People | -0.012*** | -0.012*** | -0.011*** | -0.008*** | -0.015*** |
| | (0.001) | (0.001) | (0.003) | (0.002) | (0.001) |
| Retired | 0.033*** | 0.033*** | 0.043*** | 0.015 | 0.036*** |
| | (0.008) | (0.008) | (0.007) | (0.014) | (0.014) |
| Unemployed | 0.050*** | 0.051*** | 0.080*** | 0.027 | 0.061*** |
| | (0.009) | (0.009) | (0.019) | (0.017) | (0.013) |
| Disabled | 0.044*** | 0.044*** | 0.054** | 0.038** | 0.048*** |
| | (0.010) | (0.010) | (0.025) | (0.015) | (0.015) |
| Homemaker | 0.028*** | 0.027*** | 0.097*** | 0.030*** | 0.013 |
| | (0.010) | (0.010) | (0.008) | (0.010) | (0.013) |
| *Country, Wave dummy* | Yes | Yes | Yes | Yes | Yes |
| *Cognitive, Diseases* | Yes | Yes | Yes | Yes | Yes |
| *Linguistic family* | Yes | Yes | Only | Only | Only |
| and sub-family | | | Slavic | Romance | Germanic |
| *N. Observations* | 65985 | 65985 | 11229 | 19438 | 28227 |
| *N. Countries* | 17 | 17 | 5 | 6 | 8 |

**Notes:** *The dependent variable is "High Risk Aversion". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
** $p < 0.05$, *** $p < 0.01$.

Table 9: *Fixed effects: Conditional Logit model. Odds-Ratios. All Countries.*

| Risk Aversion (d) | RA 1 | RA 2 | RA 3 | RA 4 | RA 5 |
|---|---|---|---|---|---|
| IRR | 1.270*** | 1.267*** | 1.242*** | 1.243** | 1.236** |
| | 0.075 | 0.071 | 0.079 | 0.119 | 0.126 |
| Retired | | 1.283*** | 1.274*** | 1.232*** | 1.178*** |
| | | 0.052 | 0.053 | 0.055 | 0.060 |
| Unemployed | | 1.511*** | 1.470*** | 1.384*** | 1.353*** |
| | | 0.070 | 0.071 | 0.108 | 0.105 |
| Disabled | | 1.581*** | 1.510*** | 1.507*** | 1.347*** |
| | | 0.099 | 0.097 | 0.144 | 0.118 |
| Homemaker | | 1.278*** | 1.277*** | 1.157*** | 1.096 |
| | | 0.072 | 0.074 | 0.061 | 0.063 |
| Owner | | 0.865*** | 0.870*** | 0.891*** | 0.899*** |
| | | 0.016 | 0.016 | 0.017 | 0.016 |
| Trust People | | | 0.922*** | 0.916*** | 0.920*** |
| | | | 0.005 | 0.006 | 0.006 |
| *Cognitive, Diseases* | No | No | No | No | Yes |
| **Fixed Effects**: | | | | | |
| *Sex x Age* | Yes | Yes | Yes | Yes | Yes |
| *Country x Wave* | No | Yes | Yes | Yes | Yes |
| *Income x Education* | No | Yes | Yes | Yes | Yes |
| *MarStatus x Num.Child* | No | No | No | Yes | Yes |
| *N. Observations* | 57770 | 56630 | 56169 | 30875 | 29414 |
| *N. Countries* | 17 | 17 | 17 | 17 | 17 |

**Notes**: *The dependent variable is "High Risk Aversion". The method of estimation is Conditional Logit Model. Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
** $p < 0.05$, *** $p < 0.01$.

Table 10: *Fixed effects: Conditional Logit model. Odds-Ratios. Linguistically Heterogeneous Countries.*

| Risk Aversion (d) | RA 1 | RA 2 | RA 3 | RA 4 | RA 5 |
|---|---|---|---|---|---|
| IRR | 1.270*** | 1.267*** | 1.244*** | 1.245** | 1.235* |
| | 0.079 | 0.076 | 0.083 | 0.128 | 0.135 |
| Retired | | 1.340*** | 1.324*** | 1.220** | 1.167** |
| | | 0.091 | 0.090 | 0.097 | 0.089 |
| Unemployed | | 1.527*** | 1.493*** | 1.443*** | 1.365** |
| | | 0.089 | 0.099 | 0.196 | 0.200 |
| Disabled | | 1.456*** | 1.398** | 1.630** | 1.473** |
| | | 0.199 | 0.192 | 0.345 | 0.255 |
| Homemaker | | 1.157 | 1.146 | 0.982 | 0.961 |
| | | 0.094 | 0.094 | 0.091 | 0.087 |
| Owner | | 0.825*** | 0.829*** | 0.878*** | 0.887*** |
| | | 0.025 | 0.024 | 0.038 | 0.037 |
| Trust People | | | 0.929*** | 0.923*** | 0.928*** |
| | | | 0.008 | 0.011 | 0.010 |
| *Cognitive, Diseases* | No | No | No | No | Yes |
| **Fixed Effects**: | | | | | |
| *Sex x Age* | Yes | Yes | Yes | Yes | Yes |
| *Country x Wave* | No | Yes | Yes | Yes | Yes |
| *Income x Education* | No | Yes | Yes | Yes | Yes |
| *MarStatus x Num.Child* | No | No | No | Yes | Yes |
| *N. Observations* | 17815 | 17453 | 17257 | 8066 | 7934 |
| *N. Countries* | 6 | 6 | 6 | 6 | 6 |

**Notes**: *The dependent variable is "High Risk Aversion". The method of estimation is Conditional Logit Model. Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: *Probit Model: Risk Aversion, Marginal Effects. World Value Survey*

| Risk Aversion (d) | RA 1 | RA 2 | RA 3 | RA 4 | RA 5 |
|---|---|---|---|---|---|
| Strong_IRR (d) | 0.091*** | 0.090** | 0.090*** | 0.088*** | 0.087** |
| | (0.034) | (0.035) | (0.035) | (0.033) | (0.034) |
| Age | 0.006*** | 0.006*** | 0.006*** | 0.005*** | 0.005*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Female (d) | 0.105*** | 0.107*** | 0.106*** | 0.106*** | 0.100*** |
| | (0.005) | (0.006) | (0.006) | (0.006) | (0.007) |
| Low Education (d) | | -0.000 | -0.007 | -0.009 | -0.011 |
| | | (0.008) | (0.008) | (0.008) | (0.008) |
| High Education (d) | | -0.025*** | -0.017** | -0.014 | -0.013 |
| | | (0.008) | (0.008) | (0.008) | (0.008) |
| Income | | | -0.011*** | -0.012*** | -0.012*** |
| | | | (0.002) | (0.002) | (0.002) |
| Married (d) | | | | 0.057*** | 0.054*** |
| | | | | (0.007) | (0.008) |
| HH Size | | | | 0.005** | 0.005** |
| | | | | (0.003) | (0.002) |
| Retired (d) | | | | | 0.019 |
| | | | | | (0.014) |
| Unemployed (d) | | | | | -0.003 |
| | | | | | (0.013) |
| Homemaker (d) | | | | | 0.020** |
| | | | | | (0.009) |
| *Country Wave FE* | Yes | Yes | Yes | Yes | Yes |
| *Cognitive, Diseases* | NA | NA | NA | NA | NA |
| *N. Observations* | 36684 | 34488 | 33647 | 33452 | 33391 |
| *N. Countries* | 39 | 39 | 39 | 39 | 39 |

**Notes**: *The dependent variable is "High Risk Aversion". The method of estimation is Probit. Robust standard errors in parentheses. Reference categories: Low Irrealis, Male, Not Married, Medium Education, Employed.*
** $p < 0.05$, *** $p < 0.01$.

Table 12: *IV Stocks: First Stage Estimation and Test Statistics, Linguistically Heterogeneous Countries*

| Risk Aversion | FS1_ML | FS2_ML | FS3_ML | FS4_ML |
|---|---|---|---|---|
| IRR | 0.033*** | 0.031*** | | |
| | (0.005) | (0.005) | | |
| CatIRR1 (d) | | | 0.160*** | 0.155*** |
| | | | (0.030) | (0.030) |
| CatIRR2 (d) | | | 0.185*** | 0.177*** |
| | | | (0.029) | (0.029) |
| Owner | -0.035*** | -0.031*** | -0.035*** | -0.031*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| Income | -0.014*** | -0.013*** | -0.014*** | -0.013*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Low Education | 0.075*** | 0.066*** | 0.075*** | 0.066*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| High Education | -0.097*** | -0.089*** | -0.095*** | -0.087*** |
| | (0.009) | (0.010) | (0.009) | (0.010) |
| Married | 0.017*** | 0.022*** | 0.016** | 0.021*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| HH Size | -0.007*** | -0.010*** | -0.007*** | -0.010*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Age | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Female | 0.083*** | 0.087*** | 0.083*** | 0.087*** |
| | (0.006) | (0.007) | (0.006) | (0.007) |
| Retired | 0.033*** | 0.035*** | 0.032*** | 0.034*** |
| | (0.009) | (0.009) | (0.009) | (0.009) |
| Unemployed | 0.039** | 0.035** | 0.039** | 0.035** |
| | (0.016) | (0.016) | (0.016) | (0.016) |
| Disabled | 0.051*** | 0.043*** | 0.052*** | 0.044*** |
| | (0.014) | (0.014) | (0.014) | (0.014) |
| Homemaker | 0.014 | 0.014 | 0.012 | 0.012 |
| | (0.011) | (0.012) | (0.012) | (0.012) |
| Trust People | -0.008*** | -0.008*** | -0.008*** | -0.009*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.742*** | 0.858*** | 0.693*** | 0.809*** |
| | (0.032) | (0.037) | (0.039) | (0.043) |
| *Country Wave FE* | Yes | Yes | Yes | Yes |
| *Cognitive, Chronic* | No | Yes | No | Yes |
| *N. Observations* | 16509 | 15405 | 16509 | 15405 |
| *N. Countries* | 6 | 6 | 6 | 6 |
| *Strong Instrument* | **50.86** | **42.06** | **24.27** | **21.54** |
| *Endogenous RA* | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| *Overidentification* | — | — | **0.7549** | **0.7597** |

**Notes**: *The dependent variable is "High Risk Aversion (d)". The method of estimation is ivreg2 (only the first stage estimates reported). Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
** $p < 0.05$, *** $p < 0.01$.

Table 13: *IV Stocks: First Stage Estimation and Test Statistics. All Countries.*

| Risk Aversion | FS1_All | FS2_All | FS3_All | FS4_All |
|---|---|---|---|---|
| IRR | 0.029*** | 0.027*** | | |
| | (0.005) | (0.005) | | |
| CatIRR1 (d) | | | 0.152*** | 0.148*** |
| | | | (0.031) | (0.031) |
| CatIRR2 (d) | | | 0.170*** | 0.162*** |
| | | | (0.029) | (0.029) |
| Owner | -0.030*** | -0.025*** | -0.030*** | -0.025*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Income | -0.016*** | -0.016*** | -0.017*** | -0.016*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Low Education | 0.061*** | 0.055*** | 0.061*** | 0.055*** |
| | (0.004) | (0.005) | (0.004) | (0.005) |
| High Education | -0.098*** | -0.094*** | -0.097*** | -0.093*** |
| | (0.005) | (0.006) | (0.005) | (0.006) |
| Married | 0.020*** | 0.023*** | 0.019*** | 0.023*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| HH Size | 0.002 | 0.002 | 0.002 | 0.002 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Age | 0.003*** | 0.002*** | 0.003*** | 0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Female | 0.096*** | 0.102*** | 0.096*** | 0.102*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Retired | 0.049*** | 0.048*** | 0.049*** | 0.047*** |
| | (0.006) | (0.006) | (0.006) | (0.006) |
| Unemployed | 0.051*** | 0.053*** | 0.051*** | 0.053*** |
| | (0.011) | (0.011) | (0.011) | (0.011) |
| Disabled | 0.070*** | 0.061*** | 0.071*** | 0.062*** |
| | (0.010) | (0.010) | (0.010) | (0.010) |
| Homemaker | 0.027*** | 0.025*** | 0.026*** | 0.024*** |
| | (0.007) | (0.008) | (0.007) | (0.008) |
| Trust People | -0.010*** | -0.010*** | -0.010*** | -0.010*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.634*** | 0.727*** | 0.542*** | 0.635*** |
| | (0.024) | (0.103) | (0.038) | (0.107) |
| *Country Wave FE* | Yes | Yes | Yes | Yes |
| *Cognitive, Chronic* | No | Yes | No | Yes |
| *N. Observations* | 51698 | 45991 | 51698 | 45991 |
| *N. Countries* | 17 | 17 | 17 | 17 |
| *Strong Instrument* | **40.19** | **33.00** | **18.64** | **16.57** |
| *Endogenous RA* | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| *Overidentification* | — | — | **0.8836** | **0.8616** |

**Notes**: *The dependent variable is "High Risk Aversion (d)". The method of estimation is ivreg2 (only the first stage estimates reported). Robust standard errors in parentheses. Reference categories: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
** $p < 0.05$, *** $p < 0.01$.

Table 14: *IV Stocks: Bivariate Probit, Marginal effects; Linguistically Heterogeneous Countries*

| Pr(Risky Assets) | **MEff_IRR** 2nd Stage RA | **MEff_FTR** Direct MEff | **MEff_FTR** Direct MEff |
|---|---|---|---|
| **RiskAversion** (d) | **-0.107*** ** | | |
| **(instrumented)** | (0.025) | | |
| **Strong_FTR** (d) | | **-0.045*** ** | **-0.039*** ** |
| **(High disc. rate)** | | (0.002) | (0.004) |
| Owner (d) | 0.019*** | 0.024*** | 0.023*** |
| | (0.005) | (0.006) | (0.007) |
| Income | 0.009*** | 0.012*** | 0.011*** |
| | (0.001 ) | (0.001) | (0.001) |
| Low Education (d) | -0.024*** | -0.041*** | -0.035*** |
| | (0.006) | (0.004) | (0.003) |
| High Education (d) | 0.018*** | 0.048*** | 0.037*** |
| | (0.007) | (0.010) | (0.009) |
| Married (d) | 0.014*** | 0.011 | 0.009 |
| | (0.005) | (0.008) | (0.008) |
| Num. Children | -0.009*** | -0.008*** | -0.008*** |
| | (0.002) | (0.002) | (0.002) |
| Age | 0.001*** | 0.000 | 0.001** |
| | (0.000) | (0.000) | (0.000) |
| Female (d) | -0.016*** | -0.031*** | -0.031*** |
| | (0.005) | (0.004) | (0.004) |
| Retired (d) | 0.022*** | 0.022*** | 0.019*** |
| | (0.006) | (0.007) | (0.007) |
| Unemployed (d) | -0.013 | -0.018 | -0.016 |
| | (0.012) | (0.014) | (0.015) |
| Disabled (d) | -0.011 | -0.031*** | -0.018** |
| | (0.012) | (0.005) | (0.007) |
| Homemaker (d) | 0.041*** | 0.028*** | 0.037*** |
| | (0.002) | (0.004) | (0.004) |
| Trust People | 0.002*** | 0.004*** | 0.004** |
| | (0.001) | (0.001) | (0.002) |
| *Country/Wave FE* | Yes | Yes | Yes |
| *Cognitive, Chronic* | Yes | No | Yes |
| *N. Observations* | 15405 | 16509 | 15405 |
| *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: No Irrealis Moods, Male, Not Married, Medium Education, Retired.*
** $p < 0.05$, *** $p < 0.01$.