UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"LINGUISTIC DIFFERENCES AND ECONOMIC BEHAVIOUR: THE ASSOCIATION BETWEEN CONDITIONAL CLAUSES AND WORK PREFERENCES"

RELATORE:

CH.MO PROF. LORENZO ROCCO

LAUREANDO: ALESSANDRO ASMUNDO

MATRICOLA N. 1082196

ANNO ACCADEMICO 2016 – 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

_____

Summary

1. Introduction

The aim of this work is to propose an innovative approach to study the association between conditional clauses and economic choices, paying specific attention to individual work preferences. The analysis draws its inspiration from the hypothesis of linguistic relativity proposed by Sapir and Whorf in the nineteenth century.

The linguistic relativity hypothesis originates from the intuition that differences in morphological and syntactic language structures could affect human thinking, individual preferences and therefore economic behaviour.

Scholars, in the analysis of this hypothesis, focused mainly on conceptual contents of language, in a crossroad among linguistics, psychology, economics and sociology. There exist several examples, concerning many semantic domains, about the effect of structural linguistic differences on various socio-economic factors, but not only those.

For instance, the range of words that can be used to describe colours, that change amongst languages, may affect colour perception (Davies and Corbett (1997); Roberson et al. (1999); Winawer (2007)). According to the linguistic relativity hypothesis, Italian speakers are more able in colour discriminations with respect to English ones, distinguish “blu”, “celeste” and “azzurro” against the only term “blue” in English. Linguistic spatial frames define the categorization of space in cognitive domain, shaping the individual perception (Majid et al. (2004)).

Several researchers focused on the distinction between languages that describe spatial relations with respect to their own body and those languages that refer instead to points fixed in the environment, that is to say cardinal points. In the latter case, as a matter of fact, an individual have to know where she or he is with respect to the north. According to the studies carried out by the Max Planck Institute for Psycholinguistics, this was proved to be true.

Moreover, the perception of gender role in society seems to be consistently affected by gender distinction in grammar, with noteworthy implications for gender equality in labour market, in politics or in the credit market. Santacreu-Vasut analyzed the determinants of gender political quota and sanctions related to enforcement, fundamental instruments to enhance female participation. Language results the most significant explanatory variable concerning the adoption of quotas, more than religion, economic development or political system of the country (Santacreu-Vasut et al (2013)).

There exist two possible interpretations of the Sapir-Whorf hypothesis, according to the studies on the relation between language and human thinking. The first one looks at the

relativity of languages, varying in the expression of concepts, and somehow affecting cognitivity of individuals. The second one claims that people who speak different languages with different grammatical structures could encode different preferences and therefore take different economic choices on the basis of their perception of the surrounding environment.

The linguistic determinism approach has been largely overcome by this weaker form of the linguistic relativity hypothesis. Such version claims indeed an indirect effect of grammatical and morphological language differences on cognitive habits, as memory and categorization, providing some constraints in the cognitive process. According to this interpretation, if people speaking different languages diverge in their vision of the world on the basis of the language used, this could influence individual preferences and thus decision processes.

The literature on the relationship between language and economic behaviour is still very poor. As far as we know, there exist only three studies that explicitly address the issue of the relationship between language structures and economic behaviour. Two of these were particularly inspiring for our analysis.

Chen (2013) was the first who tried to test a linguistic-saving hypothesis, analysing the effect of future time references on future oriented behaviour. In his theory, people who explicitly mark present and future events show a short-term orientation in several economic choices, in contrast to those who talk about future events substantially in the same way as about the present events. The main underlying mechanism to explain these insights is related to the perception of time and therefore to a more precise time-partitioning. Strongly marking future events affects intertemporal choices and risk-aversion.

Kovacic (2016), in addition, carried on the idea proposed by Chen and tried to investigate the correlation between the intensity of use on non-indicative moods, in contexts that involve grammatical categories regarding the expression of uncertainty, and individual attitude toward risk. The use of these *irrealis* moods, hence, leads to perceive displacement between actual and hypothetical state of the world as larger, when we describe possible situations.

The approach we opted for in our thesis is basically in line with Kovacic (2016), following the weaker linguistic relativity hypothesis. It diverts, anyway, in at least two fundamental aspects. First of all, we decided to study a sample of immigrants selected in the same hosting country, considering migration as a microevolutionary step, in order to isolate language effect, *ceteris paribus*, fromc confusing factors such as culture or customs and traditions. In addition, we abandoned the classic binomial relationship between risk-aversion and economic behaviour, placing emphasis on the concepts of risk perception and cognitive biases.

Investigating the association between conditional clauses and economic behaviour, we supposed that individual speaking languages characterised by an intensive displacement into

uncertainty perceived the world as being more unstable and mutable but, at the same time, they have a fitting and more precise perception of the actual risks they face. Moodless-language speakers, on the contrary, are affected by cognitive biases such as overconfidence that could amplify their risk perception. This may distort their attitude toward risk and affect decision processes related, for example, to work preferences or entrepreneurship.

We found that an intensive use of irrealis moods is associated with a higher propensity to be employed in a high-risk occupation and a lower probability of starting a new venture, with a positive effect on wage and salary income. This could be explained by the awareness of unmanageable risks related to hazardous jobs and, on the contrary, of a controllable risk entailed by entrepreneurship.

Our intuitions are supported by the positive relationship between the intensity of use of irrealis moods on wage and salary income. People who undertake a high-risk occupation earn higher wages because of the positive correlation between risk and salary compensation. At the same time, avoiding the creation of unprofitable businesses prevent from potential losses.

Finally, we built up a robustness check, based on an epidemiological approach, analysing our hypothesis on a sample of second-generation immigrants in the United States.

In the next section we will proceed with a broad analysis of the association between linguistic relativity and economic behaviour, together with a digression on philosophy of language, linguistics and cognitive psychology. Chapter 3 focuses in particular on our baseline assumption of association between conditional clauses and economic behaviour, analysing data and methods, the linguistic employed in the model and our empirical strategy. The robustness check is presented in Chapter 4 and in Chapter 5 we draw the conclusions of the work, examining our results.

2. The association between linguistic relativity and economic behaviour

The hypothesis that language differences can affect thoughts, and therefore preferences and economic behaviour of individuals was suggested for the first time by Sapir (1921) and Whorf and Carrol (1964) and is known as Sapir-Whorf hypothesis.

Language and thought could interact in many interesting ways. There is much debate about the existence and the intensity of this association: on the one hand is difficult to identify it, for example, for an individual who speaks several languages and should have been affected from each of them; on the other hand it is quite natural to think that individuals experience the world in the same manner regardless of their birthplace or their mother tongue. Precisely for this reason many linguists, philosophers and social scientists used to investigate the grammatical and semantic structures of languages, identifying their common and unique traits and trying to find a definitive answer about the existence of any interrelationships between language and human thinking.

Going back to the Sapir-Whorf hypothesis, there are two possible channels through which linguistic relativity could affect thought, following Geeraerts and Cuyckens (2010). The first is that languages are relative, varying in the expression of concepts, and affect in some way the cognitive level of individuals. The second suggests that people speaking different languages could encode different preferences and take different economic choices on the basis of their perception of the surrounding environment, due to different language structures.

The strongest form of the hypothesis is the linguistic determinism, stating that linguistic categories directly control and determines the cognitive process of individuals, has now largely been recognized as wrong. It was overcome by the weaker hypothesis that claims an indirect effect of language differences on cognitive habits, as memory and categorization, providing some constraints in the cognitive process.

The idea of the existence of an interconnection between language and thoughts has ancient roots in the philosophy of language. Gorgias, for example, focused on the interconnection among *logos*, *ethos* and *pathos* (language, ethics and emotions), asserting the total supremacy of the language in shaping not only human behaviour but also human mindset, affecting both emotional and intellective structure. As other sophists Gorgias argued that the world can be experienced only through the language.

Plato, in the dialogue *Cratylus*, questioned the role of language and focused on its origins. The philosopher analyzed whether words are “natural or “conventional” i.e. if language is a

system of arbitrary signs or if there is an intrinsic relations between words and the things they are representing. The underlying theory was that words are just reflecting ideas that are eternal.

Such as Plato also Saint Augustine, for example, supported the idea of words as label applied to existing things, a view which remained the main one in the Middle Age. Philosophers as Bacon, on the other side, considered the language as a veil that is hiding the truth to individual, hostages of linguistic conventions.

In the 18th century German Romantic philosophers proposed the idea of language as a fabric on thought, reflecting the existence of ethnic national characters that impede a common grammatical structure amongst languages.

On the basis of *Innere Sprachform* of Wilhelm von Humboldt (1836), Edward Sapir started from the idea that languages encoded the key to understanding how people view the world: as languages represent the reality in different ways, the speakers of different languages could perceive reality differently. Sapir first explicitly refused linguistic determinism, noticing on the other side a weak connection between language and culture. “Human beings [...] are very much at the mercy of the particular language which has become the medium of expression for their society. […] The fact of the matter is that the "real world" is to a large extent unconsciously built up on the language habits of the group” (Sapir, 1929; in Mandelbaum, 1958, p. 162)

In the Tractatus Logico Philosophicus, Wittgenstein proposed a representational theory of language: the world is a mere collection of facts, not of things, that can be represented by the language, if it has a logical form. Few years later Wittgenstein changed his mind, refusing the “picture theory” of language. This change represented the general shift of 20th century philosophy from logical positivism to behaviourism and pragmatism, considering language as a social practice, a fluid structure built up by using conventionally-determined words in a sort of “language-game”.

Benjamin Lee Whorf took up the suggestions of Sapir about linguistic relativity going through a deeper analysis of the correlation between language and thought. Human thinking has to be organized by our mind and namely by our linguistic system, capable of organizing concepts, ascribing significances, marking the time. Whorf suggested an agreement within a speech-community which codifies the surrounding world through the same patterns of language, studying Native American languages. In some way languages, as like as meaning or references, are relativized to the context in which they are spoken. According to Gumperz and Levinson “meanings are relative to contexts, and context and principles of interpretation (or at least applications of them) are relative to shared networks. [...] Differences in the use of

language imply differences in meaning or interpretation." (Gumperz and Levinson, 1996, p. 359)

Whorf had many detractors, such as Chomsky, Lenneberg and Pinker, who criticized the weakness and the poor scientific relevance of his hypothesis, pointing out a lack of causality.

Just Eric Lenneberg and Roger Brown, in 1954, argued that Whorf never reveal the interconnection between linguistic and mental phenomenon and proposed their formal hypothesis, later developed by Brown into the "weak" and "strong" formulation mentioned above, retrospectively attributed to Sapir and Whorf:

"1. Structural differences between language systems will, in general, be paralleled by nonlinguistic cognitive differences, of an unspecified sort, in the native speakers of the language.

2. The structure of anyone's native language strongly influences or fully determines the worldview he will acquire as he learns the language." (Brown, 1976, p.128)

The experiments of Brown and Lenneberg went dealt with colors perception and their codification in languages.

The next step was the theorization of universalism, usually attributed to Noam Chomsky. The hypothesis stated that certain set of grammar structural rules are innate to humans and do affect in no way cognitive process. All the languages, according to universalism, share a common underlying structure that overcomes grammatical differences.

According to Lucy (1997) there are three main currents in the analysis of linguistic relativity.

The "structure-centered approach", starting from Whorf theories, looks at language structural peculiarity and studies the relationship with human thought.

The "domain-centered approach", centered for example on colors perception, chooses a semantic domain and compares it across languages and cultures.

The "behaviour-centered approach", which is what we are interested in, looks for causality in human behaviour due to linguistic differences.

As previously mentioned, in recent years the Sapir-Whorf linguistic relativity hypothesis represented the basis for many authors in investigating the effect of language differences on human thought.

The main thread focused on conceptual contents of language, in a twist among linguistics, sociology, psychology and economics. We can find several examples, in many semantic domains, about the effect of linguistic relativity hypothesis on many socio-economic factors but not only those.

Colour perception, for example, may change on the basis of the range of words for color, differing amongst languages (Davies and Corbett (1997); Roberson et al. (1999); Winawer

(2007)). Italian, for example, distinguishes amongst three different kind of blue such as “blu”, “celeste” and “azzurro”, whereas English uses only the term “blue”. Following the linguistic relativity hypothesis Italian speakers are more able in colour discriminations with respect to English ones.

Furthermore, according to Majid et al. (2004), the linguistic spatial frames define the categorization of space in cognitive domain, shaping the individual perception.

Several researches, analyzing linguistic determinism, were conducted at the Max Planck Institute for Psycholinguistics. Stephen C. Levinson, the supervisor of the project, focused on languages that describe spatial relations with respect to the body and those that refer to points that are fixed in the environment. English, for the first case, uses as points of reference words such as front, back, right or left, whereas in some aboriginal languages of Australia, for the second case, individuals refer to north, south, west and east. In order to speak the latters, an individual have to know where he is with respect to the north. The studies of the Max Planck Institute for Psycholinguistics shown, after several analyses, that this is true.

Moreover, gender distinction in grammar, for example, seems to affect the perception of gender role in the society, with significant consequences on gender equality in the labor market, in politics or in the credit market.

Santacreu-Vasut analyzed the determinants of gender political quota and sanctions related to enforcement, two fundamental tools to increase female participation. She found that language is the most significant explanatory variable related to quota adoption, more than religion, economic development or political system of the country (Santacreu-Vasut et al (2013)).

A particularly interesting thread emerged in the last decades, focusing on differences in grammatical structures and primarily in verb structures. “Tense and aspect system” (Dahl, 1985) represents the benchmark for most of these analysis about grammatical differences in language structures, shaping a common background that defines the use of tense, moods and aspects categories in several languages, identifying a limited set of cross-linguistic category types.

Like Bell (1978), Dahl tried to overcome the Indo-European bias of every language sample, reducing the distortion due to grammatical contaminations deriving, for example, from colonisation and defining a set of 124 languages.

The main claim of Dahl’s work was that “it is possible to reduce the bewildering multitude of tense-mood-aspect categories found in the languages of the world [...] to a fairly small set of cross-linguistic categories, characterized by bundles of morphosyntactic and semantic properties, more specifically by the (proto)typical contexts in which they are used and by the

typical ways in which they are manifested morphologically or syntactically." (Dahl, 1985, p.182)

2.1 Future Time References

In "The grammar of future time reference in European languages" (2000) Dahl focused on the major semantic distinctions relevant to the marking of future time references (FTR) and on grammaticalization of the future tense in about 30 european countries, building the baseline for the analysis of Chen (2013).

If languages can affect the perception of the reality we may assume that they are capable to affect also individual preferences and thus their choices about many aspects of their habits.

Chen (2013) and Sutter et al. (2015) were the first who attempted to study the effect of languages on individual economic behaviour starting from the linguistic relativity hypothesis. Languages strongly differ in the way they encode time and they could be categorised on the basis of future time references. The two authors argue that it is the language itself that induces both long-term or short-term orientation and thus economic choices of individuals.

The discriminating factor are the future time references, that study how and when languages require the usage of future tense to mark the timing of events. We can separate languages in two categories: strong and weak FTR languages, the first requiring grammatical marking of future events, the second that do not.

The basic hypothesis, inspired by the classification of languages by Dahl (2000) and Thieroff (2000) and based on the specific verb morphology for FTR, is that languages that require a specific tense for future events, therefore grammatically separating present and future, imply less future-oriented actions by the speaker. On the other side, languages equating present and future make the latter to appear closer, in a single time horizon, reducing the discount-rate of future events and therefore inducing more future-oriented actions.

In his studies Dahl defines as "futureless" languages the ones that do not require "the obligatory use [of grammaticalized future-time reference] in prediction-based contexts" (Dahl 2000, p.12.). A prediction is a statement about the future with no intentional component but as a fact.

Thieroff instead started the analysis on future tense distinguishing between prediction-based and intention-based future time references, adding the scheduling aspect later, and focusing on different language families and areas.

Thieroff claims, like Dahl, the absence of inflectional futures for all Germanic and Finno-Ugrian languages and for the majority of the Slavic ones. A suggestive step forward is the

identification of the tendency not to distinguish present and future, neither inflectionally nor periphrastically, splitting clearly weak and strong future time reference languages.

Thieroff has a more neutral approach, talking about "weak-FTR" instead of "futureless" languages, as recalled by Chen.

Chen (2013) ran both cross-country and within-country regressions, comparing individuals with identical income, education and family structure but who speak different languages. The idea is to study the correlation between weak-FTR languages and future-oriented behaviour.

Chen's linguist-savings hypothesis is based on two different causal mechanisms that shape individual preferences.

The first one is the bias in beliefs due to the obligatory linguistic distinction between present and future events: "a constant pressure to mark the present tense as different from the future in one's language can make the temporal future seem further away by contrast" (Roberts and Chen 2015, p.2), affecting individual's investment and saving choices.

The second mechanism is the one related to the perception of time and to a more precise time-partitioning due to strongly-marking future tense. This mechanism is related to risk-aversion and intertemporal choices, overcoming the idea that language is related to intertemporal choices – as the linguistic-savings hypothesis suggests – but not to attitude toward risk.

The effect of language differences on economic behaviour in Chen (2013) is strong, involving variables as savings, investments, years of education, healthy behaviours in smoking, drinking and lifestyle.

Individuals who speak a strong-FTR language, for example, are roughly 30% less likely to have saved in the last year, have about 39% less years by the time they retire, a 24% higher probability of having ever smoked, a 29% lower probability of being physically active, a 13% higher probability of being obese (Chen, 2013).

The effect of FTR on economic behaviour measured by Chen flows through a channel that seems independent from either cultural or cognitive differences between linguistic groups. Nevertheless, as far as Chen's paper is concerned, we cannot rule out completely the possibility that language is "only as a powerful marker of some deeper driver of intertemporal preferences". The language structure of future time references is quite old and it may be changed in years.

Many critics point out that linguistic systems for referring to future events "are more complex than the binary strong/weak future tense distinction, and there is variation amongst speakers of the same language" (Roberts and Chen 2015).

Through a mixed effect modelling, along with other analysis, Roberts and Chen tested the robustness of correlation between future time references and the propensity to save, while

controlling for the relatedness of languages. They found that weak-FTR speakers are 1.5 times more likely to save than strong-FTR speakers, that is to say a diminished effect when they account for language relatedness. The most important hint is that any correlation should control for historical and geographical relatedness.

An interesting aspect of Chen's linguistic-saving hypothesis was faced by Oded Galor. He supposed and proved, focusing on the coevolution of cultural and linguistic characteristics and their impact on development, that "pre-industrial geographical characteristics that were conducive for the emergence and progression of complementary cultural traits triggered an evolutionary process in language structures that has fostered the transmission of these cultural traits and has magnified their impact on the process of development. [...] Regional variations in pre-industrial geographical characteristics that were conducive to higher return to agricultural investment, and thus to the emergence of long-term orientation, are at the root of existing cross-language variations in the presence of the future tense." (Galor et al., 2016, p.1)

By the way, despite language coding of past human behaviour and especially ancestral traits, variations in pre-industrial geographical characteristics are contributory causes of variations across languages for what concerns the existence of future tense, grammatical gender, and politeness differences in grammar and therefore the authors identify an independent effect of linguistic differences on economic choices and outcome.

Stepping forward, according to Rubinstein (2000), "the constraint of preferences might arise from the language that the decision-maker uses to verbalize decision taken" (Rubinstein (2000, in "Anderlini L., Felli L. (2004) Book review: economics and language: five essays by Ariel Rubinstein", p. 171).

2.2 Moods, displacement and modality, context of use

A suggestive step forward in studying the relationship between linguistic variations and economic behaviour was made by Kovacic et al. in his working paper "Risk attitudes, investment behaviour and linguistic variation". The author proposes an innovative approach in his analysis, setting a new linguistic marker which classifies languages "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" (Kovacic et al., 2016, p.26) and study the association between the intensity of use of this moods and risk aversion of individuals.

According to Sutter (2015) language is related to intertemporal choices – as the linguistic-savings hypothesis suggests – but not to risk attitudes: Kovacic tries to show the opposite. His approach is in line with Chen (2013), being based on a weak version of the linguistic-saving

hypothesis but on moods and uncertainty instead of future time references and intertemporal choices.

The author suggest that individuals that speak languages where non-indicative moods are used more frequently in *irrealis* contexts "perceive the world as more mutable and uncertain with respect to speakers of languages where these forms are less frequently used or do not exist at all".

The main hypothesis is that differences in the intensity of use of *irrealis* moods (i.e. non indicative moods in *irrealis* contexts) shape the perception of facts and affect individual levels of risk aversion, since they encode different grades of uncertainty for different situations. When we use an *irrealis* mood the displacement between actual and hypothetical state of the world is perceived as larger, when we describe possible situations.

Borrowing an example from Kovacic et al. we can compare two sentences, in English and Italian, where the event should be perceived as more uncertain by the Italian speaker, despite the same event is occurring.

I think he has left. *(English)*
(Indicative mood)

Penso sia partito. *(Italian)*
(Subjunctive mood)

In the first statement of our example the speaker uses an indicative mood, whereas in the second one the speaker uses a subjunctive (i.e. *irrealis*) mood, widening the share of uncertainty perceived by the individual when he pronounces the sentence.

The concepts of displacement and modality are fundamental when approaching to linguistic relativity and grammatical structures, especially in *irrealis* contexts.

Displacement is the characteristic of human language whereby individuals can refer not only to events *hic et nunc* but can encompass also future, past, possible or impossible situations, ranging over time (from the past to the future) or contingency (unreality).

Modality as well represents, as like as futurity, a crucial dimension in displacement. Sentences can express as a matter of fact reality, predictions, intentions, scheduling, opinions, desire and so on.

Borrowing an example by Kovacic et al. (2016)

(1) Wenn es sonnig wäre, ginge ich spazieren. (German)

if it sunny be - **KONJ**, go - **KONJ** I walk

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

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

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

"I think the meeting has nished."

(3) Chodźmy do mnie na kawe. (Polish)

go - IMP to me for coffee

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

[Kovacic et al. (2016), p. 7]

Where "KONJ" represents the German Konjunktiv, "1SG" is First Singular and "IMP" is the Imperative.

Analyzing the previous sentences we notice that they are not about actual facts, the truthfulness of these expressions can be understood indeed only looking at the states of facts.

The “possible worlds” shown in the example certainly involve the idea of uncertainty in themselves.

The subsequent step in analysing languages structure is to focus on moods or, more precisely, morphological moods, i.e. “the grammatical category concerned with the expression of situations involving the "world" parameter.” (Kovacic et al. (2016), p. 9).

The *indicative* is the mood used to express an event that is actual and is not marked by uncertainty. This does not exclude, obviously, that it can assume modal functions.

The *subjunctive* is the mood typically used to express various states of unreality such as emotion, possibility, wish, opinion, obligation or action not yet occurred.

The *imperative* is the mood that forms commands or requests, like giving permission or forbidding something. It is also used for exhortation, both as present and future tense .

The *conditional* is the mood that expresses a proposition whose validity is dependent on some condition, possibly counterfactual. Not all the languages use the conditional mood for conditional sentences, such as English.

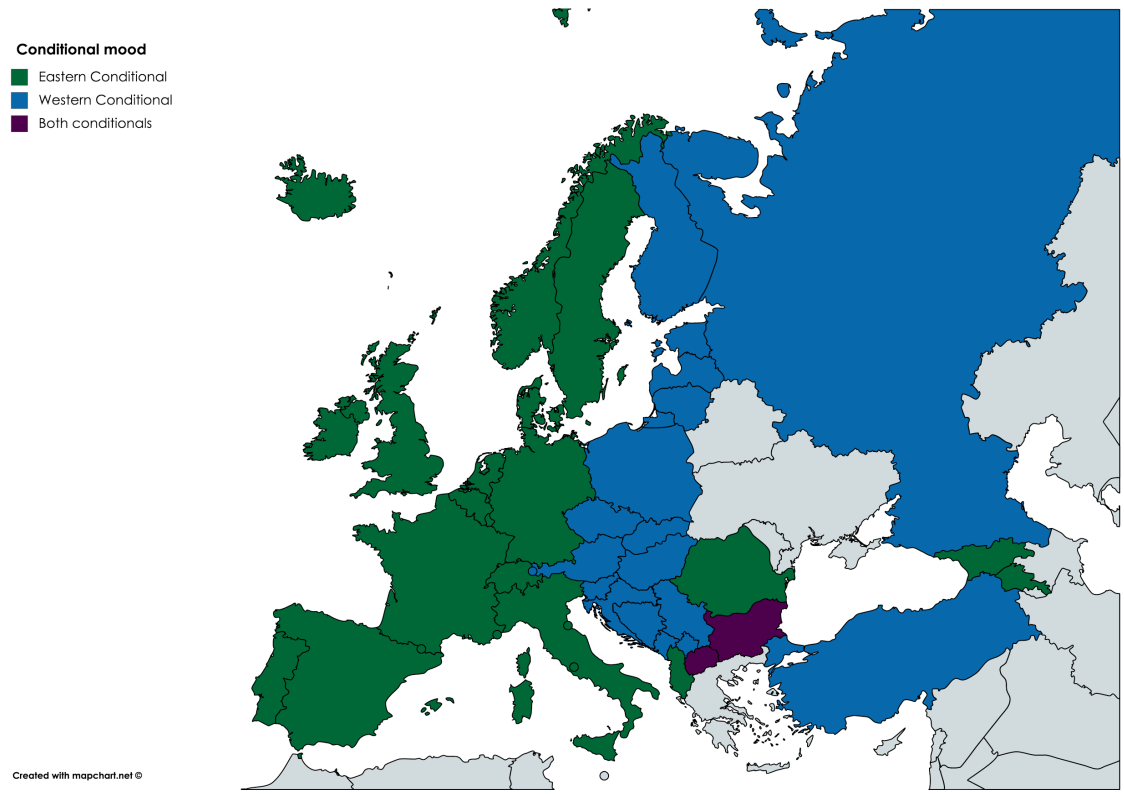
Thieroff distinguishes between two different *conditionals* in the languages of Europe: one is a mood and the other is a tense. The *Western conditional*, appearing mainly in the languages of the west of the continent, is the tense; the *Eastern Conditional*, instead, is the mood. The languages with the western conditional are almost the same which have a subjunctive

"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." (Kovacic et al. (2016), p. 9)

In addition to the number of non-indicative moods is fundamental to consider the syntactic context of use of tenses that may change their actual meaning. For what concern our work we focus on *irrealis* context, namely the one concerning possible or unreal situations.

Kovacic identifies six cross-linguistic *irrealis* contexts in which non-indicative moods are frequently used:

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.*

[Kovacic et al., 2016, p.10]

There are some other contexts involving non-indicative moods that use some adverbial clauses but they are not as frequent as the previous ones and that are not so interesting for our work.

As already mentioned, the main and most detailed study about these grammatical characteristics is “Mood in the Languages of Europe” (2010) by Rothstein and Thieroff, who tried to identify a common structure about displacement and modality in 39 european languages, despite the different linguistic family, focusing on their shared features.

The interconnection among displacement and modality, strictly related to the context of use, build a complex structure which is the base of our work in the analysis of the association between language differences and economic behaviour and namely between conditional clauses and work preferences

3. Non-indicative moods, irrealis contexts and economic choices

According to the main literature, based on the linguistic relativity hypothesis by Sapir and Whorf, there exists an association between language differences and economic behaviour, particularly with regard to future time references and long-term orientation in economic choices of individuals, as studied by Chen.

Nevertheless the debate about the consistency and the intensity of this association is still in progress, favouring the weak interpretation of the Sapir-Whorf hypothesis and underlying the indirect effect of language differences on cognitive habits, for example, as memory and categorization.

The issue becomes more fuzzy, with respect to Chen linguistic-saving hypothesis, when we look at the relationship between language and attitude toward risk. According to Sutter, for instance, "language is related to intertemporal choices – as the linguistic-savings hypothesis suggests – but not to risk attitudes" (Sutter (2015), p.14). Kovacic et al., as previously mentioned, proved the opposite. He showed a significative correlation between language and individual risk attitudes, as like as on stock ownership: speakers of languages where non-indicative moods are used more frequently to express predictions, intentions or schedules in *irrealis* contexts seem to perceive the reality as more uncertain and mutable. These individuals should hence be more risk-averse and avoid any risky behaviour with respect to low intensity users.

This way of thinking represents the basis of our thesis in analysing the association between conditional clauses and economic behaviour: work preferences should reflect risk-aversion and therefore language differences, leading individuals who speak a high-intensity *irrealis* language to choose safer and less hazardous jobs.

3.1 Data

The empirical analysis is run on a family basis on a panel of individuals in the American Community Survey by the U.S. Census Bureau. The annual survey is a representative, randomized sample of about three million American households from areas with a population greater than 65000 units and exists since 1850 as supplement or a deepened version of the basic decennial census. It collects data about several aspects of the american population such as migration, language, income, jobs and occupations, educational attainment, veterans or housing characteristics. The American Community Survey (ACS) represents a fundamental

tool for planning several private analysis or public interventions, such as the Government geographical subdivision of more than $400 billion in federal and state funds.

The respondents come from 78 countries around the world and they are all residing in the 49 federal states of the U.S.. We selected, from 2001 to 2015 (ACS, 5% sample), all the foreign born non american immigrants, aged between 25 and 55 years (2,658,128 of 4,625,036 observations), speaking 27 different eurasian languages as defined by the Dahl's categorization, including just those with a complete set of information about language and nationality. Individuals were thus chosen in the main range of the working age, after the end of the university period.

Individuals, when fill in the questionnaire of the American Community Survey, have to declare which is, if other than English, the language they speak at home as like as their proficiency in speaking English. This allowed us to select only foreign-language speakers and especially those analysed by Rothstein and Thieroff in our subsample of immigrants.

Our research analyzes the behaviour of first-generation immigrants addressing confounding effects in the same (common) country of birth, accounting for all the individual characteristics potentially affecting human behaviour and language spoken. We used also year and country fixed effect in order to account for cultural background and country of origin effect, potentially significant for first-generation immigrants.

There is still, however, a potential bias due to omitted variables, although our identification strategies should allow us to find and identify the effect of culture (opposed to culture and geography).

Throughout our analysis we controlled for individual education levels, age, number of years spent in the United States, since the migration, as a proxy of cultural integration, marital status and country of birth, as these factors may affect economic choices.

The idea underlying the model, suggested by Gay et al., is to analyse "migrations to the same country as a microevolutionary step that may uncover constraints on behavior. […] Studying the behavior of migrants allows the researcher to compare individuals that evolve in a common institutional environment." (Gay et al., 2016, p. 1-3)

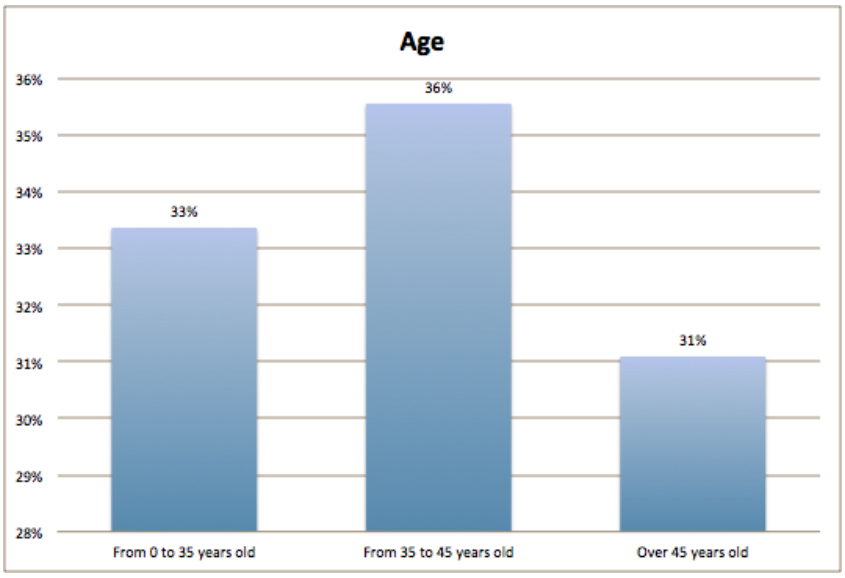
Had we analysed different countries, we could have succumbed to the risk of confusing the potential effect of grammatical differences in primary languages with innate differences between natives and immigrants. Imagine, for example, the numerous English-speaking-countries in our sample (such as the members of the United Kingdom or Canada, Australia, New Zealand and so on), each with its peculiarity for what concerns culture, society, habits and customs or economic structures.

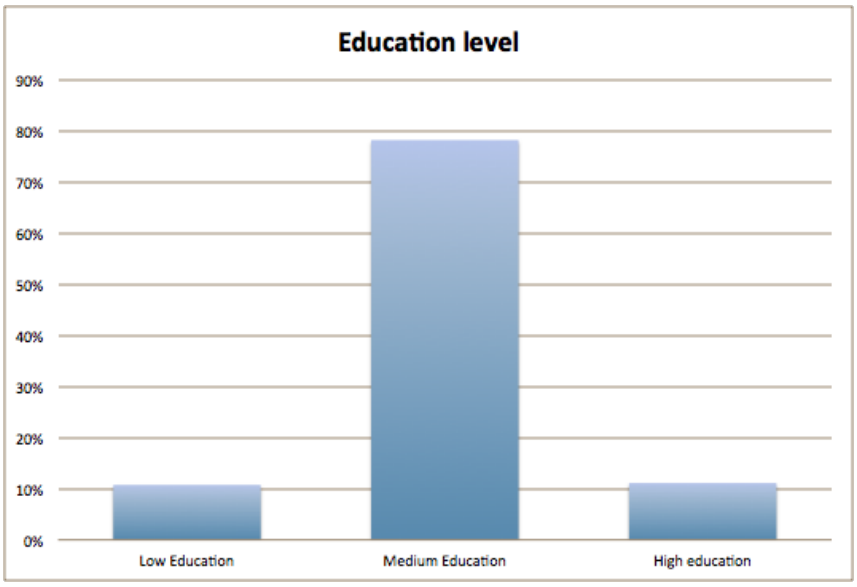
One potential concern in our model is therefore that omitted parental characteristics might bias the estimated effect of linguistic relativity on individual's behaviour. In our sample, unfortunately, ancestral characteristics are defined for too few observations with a remarkable bias due to the questionnaire compilation method that we are going to explain.

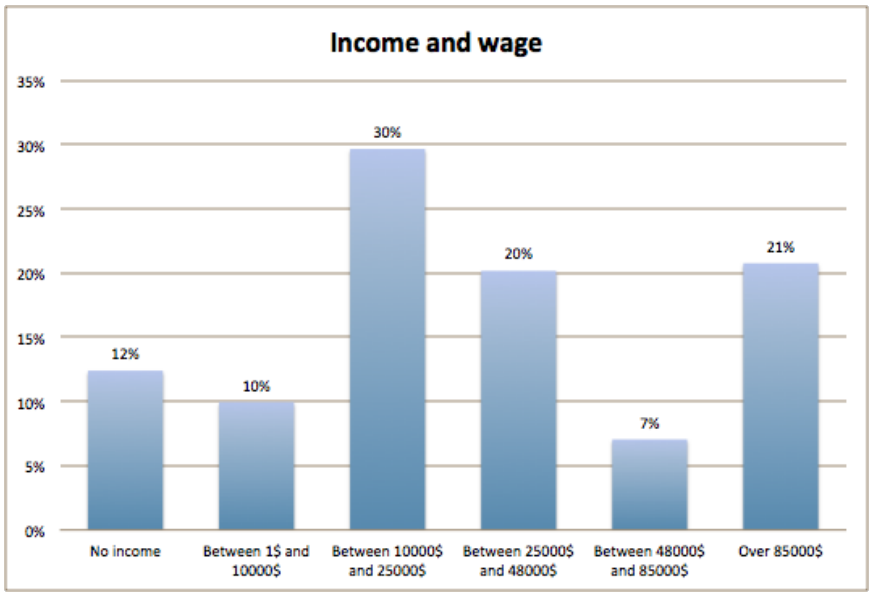
The structure of the American Community Survey questionnaire provides it to be filled in by the owner or the tenant of the house or the apartment where the interviewed family lives. Almost always this individual is the householder, namely the mother or the father, and she or he has to complete the questionnaire with all the information about the people residing in the house, sons included. This structure implies that the ancestral characteristics of individuals are defined and available only for those who still live with their parents, less than a tenth of the total sample, because they are deducted indirectly and not with an explicit question to the compiler. This forced us to drop this variables out of our regression models. We would have strongly liked to use an epidemiological approach which would have focused on second-generation immigrants in the United States but it was impossible because of this sample bias. The epidemiological approach indeed studies causality by the comparison between natives and immigrants.

Anyway, as a robustness check, we decided to study the association between conditional clauses and work preferences also on a subsample of second-generation immigrants, as we will see later.

Our main sample is quite stratified for what concerns age, sex, educational attainment, income and wages and other control variables.







We grouped individuals in family clusters, with a range of household members between two and seventeen.

All the explanatory variable in our model had been reclassified as dummy or categorical variables to set a baseline composed by male speaking a moodless language, not married, with a low level of education and that has been living in the US for less than ten years.

We set up two additional categorical variables: the first one related to the continent of birth, namely North America, Central America, South America, Europe, Asia, Africa and Australia, New Zealand, Pacific Islands, Antarctica; the second one is focused on immigrants born in Europe, splitting it in Northern Europe, Western Europe, Southern Europe, Central Europe, and Eastern Europe.

In addition we generated three interaction dummies that link the linguistic marker, which we are going to discuss later on, and respectively age, number of years spent in the United States and the level of proficiency in speaking English.

In order to analyse the association between the intensity of use or non-indicative moods in *irrealis* contexts and work preferences, we set up our main dependent variable as a dummy equal to one if the individual is employed in a hazardous occupation and zero otherwise, based on the List of Occupations Designated As High-Risk Or Safety-Sensitive by the Labor Commissioner of the State of Connecticut. The categorization document was drawn up starting from a precise request received by the Connecticut Department of Labor on behalf of individual employers in order to guide the firms to a particular attention to employees' health. The riskiness of the job, as shown by the List, is defined on the basis of physical, chemical, biological or psychological risk, ranging for example from engineer to firefighter occupation.

The intuition underlying this dependent variable is to analyse the association between intensity of use of *irrealis* moods and work preferences through the potential effect of language structures on risk-aversion.

Moreover, in order to examine in depth the relationship between language structures and economic choices we looked also at two additional dependent variables.

The first one is the logarithm of total salary and wage income of individuals of our sample, strongly and positively related to the riskiness of the job in which you are involved in. The idea is to analyse if *irrealis* use is associated to an effect on the wealth of individuals.

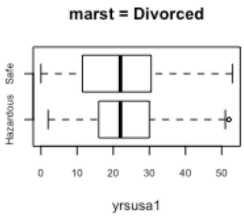
The second one is a binary variable, again, equal to one if an individual is self-employed/entrepreneur and zero otherwise, namely if he or she works for wage or is unemployed or is not in the labour force. We want to exploit it to study the effect of grammatical structures on the probability of becoming a businessman i.e. on entrepreneurship. The underlying purpose is to exploit the degree of entrepreneurship as a robustness check to measure the risk-aversion in our model. More frequent *irrealis* moods should imply a higher propensity in taking risks involving choices such as self-employment and entrepreneurship.

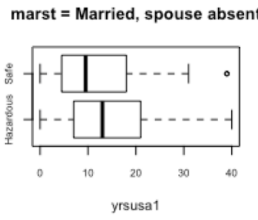
On the other side, we needed to quantify unambiguously the intensity of usage of non-indicative moods in *irrealis* contexts in all the languages of our sample. In order to identify the association of *irrealis* moods with our dependent variable we take advantage of the linguistic marker theorized by Kovacic et al. (2016), the one that we are going to analyze in depth in the following section of our thesis.

Table of summary statistics

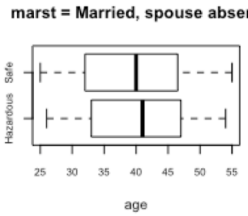
Variable	Label	Obs	Mean	Std. Dev.	Min	Max
highrisk	High risk occupation	2644984	.4149961	.4927214	0	1
irrealis	Irrealis (dummy)	1980837	.745188	.4357556	0	1
catIRR	Irrealis (cat)					
1	Intermediate use	1980837	.0589746	.2355771	0	1
2	Intensive use	1980837	.6862135	.4640309	0	1
age	Age	2658128	4.016.521	8.539.312	25	55
gender	Gender	2658128	.5127556	.4998374	0	1
married	Married	2658128	.6884458	.4631288	0	1
bpl	Birthplace	2658128	3.487.925	1.525.726	105	950
yearsUSA	Years in the USA					
1	Between 10 and 15 years	2658128	.1615122	.3680028		
2	Between 15 and 22 years	2658128	.1976632	.3982367	0	1
3	Between 22 and 30 years	2658128	.1825875	.3863281	0	1
4	More than 30 years	2658128	.1668983	.3728851	0	1
cat_incwage	Income and wage (cat)					
1	Between 1$ and 10000$	2658128	.0884092	.2838892	0	1
2	Between 10000$ and 25000$	2658128	.2209705	.4149007	0	1

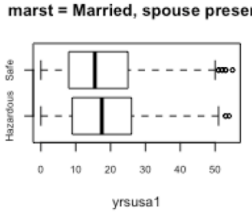
3	Between 25000$ and 48000$	2658128	.2158448	.4114072	0	1
4	Between 48000$ and 85000$	2658128	.1347478	.3414541	0	1
5	Over 85000$	2658128	.2181859	.4130144	0	1
education	Education (cat)					
1	Medium Education	1607006	.6877579	.4634082	0	1
2	High education	1607006	.2251361	.417672	0	1
N	Family components	2658128	1.803.725	.8597909	1	14
entrepreneur	Entrepreneurship	2309993	.1132947	.3169528	0	1
FTR	FTR	2292042	.9694168	.1721856	0	1

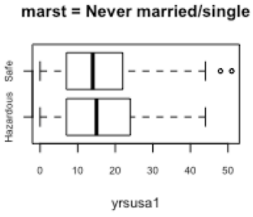

marst = Divorced
Safe
Hazardous
0 10 20 30 40 50
yrsusa1

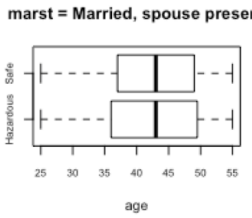

marst = Married, spouse absent
Safe
Hazardous
0 10 20 30 40
yrsusa1


marst = Divorced
Safe
Hazardous
30 35 40 45 50 55
age

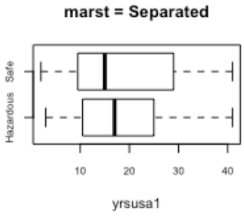

marst = Married, spouse absent
Safe
Hazardous
25 30 35 40 45 50 55
age

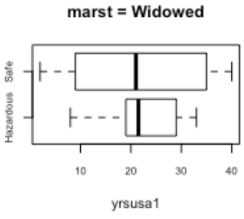

marst = Married, spouse present
Safe
Hazardous
0 10 20 30 40 50
yrsusa1


marst = Never married/single
Safe
Hazardous
0 10 20 30 40 50
yrsusa1


marst = Married, spouse present
Safe
Hazardous
25 30 35 40 45 50 55
age

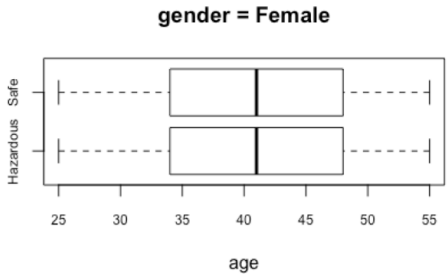

marst = Never married/single
Safe
Hazardous
25 30 35 40 45 50 55
age

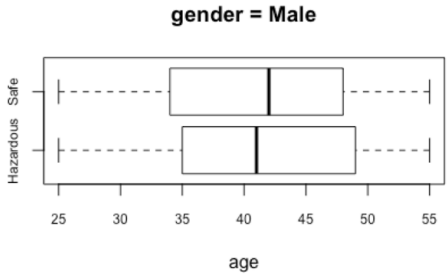

marst = Separated
Safe
Hazardous
10 20 30 40
yrsusa1

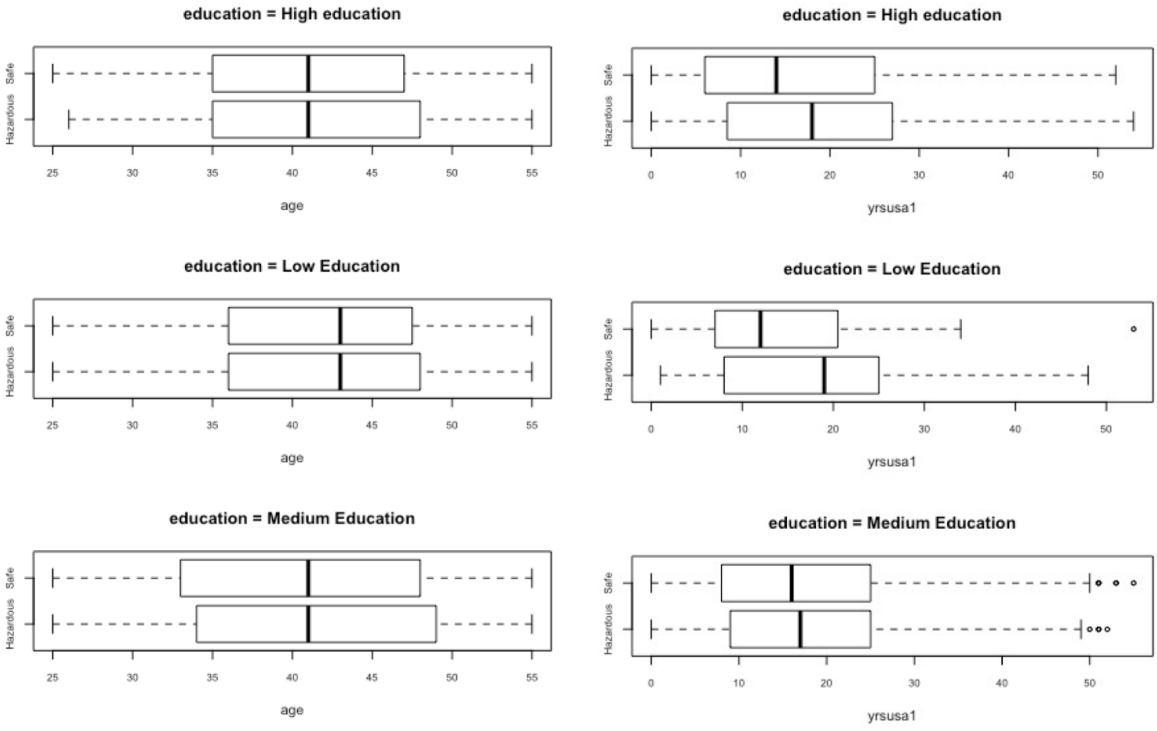

marst = Widowed
Safe
Hazardous
10 20 30 40
yrsusa1

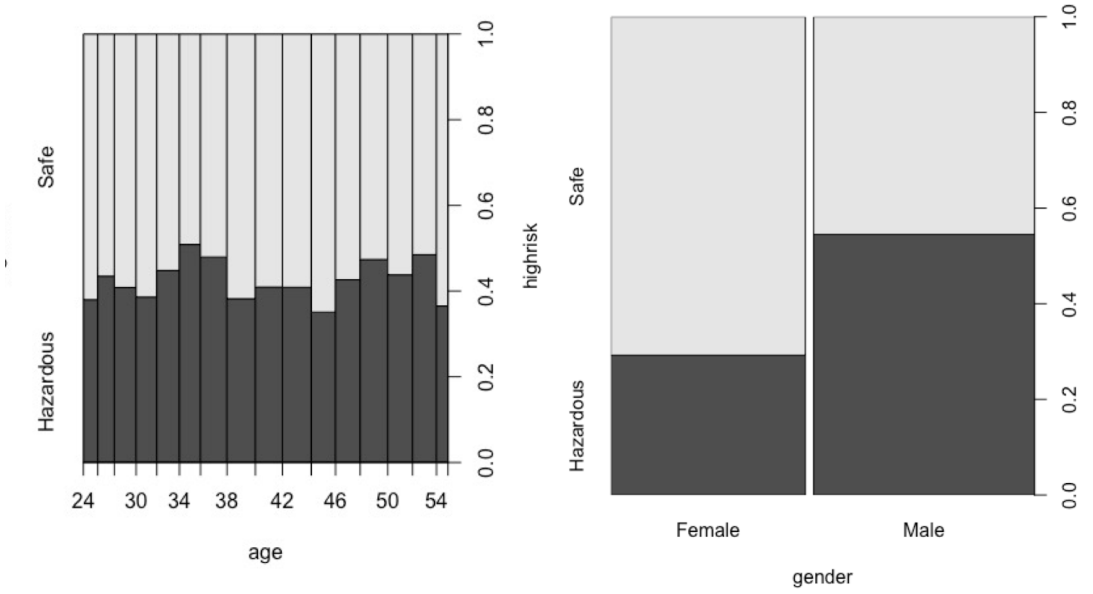

marst = Separated
Safe
Hazardous
25 30 35 40 45 50 55
age


marst = Widowed
Safe
Hazardous
42 44 46 48 50 52 54
age


gender = Female
Safe
Hazardous
25 30 35 40 45 50 55
age


gender = Male
Safe
Hazardous
25 30 35 40 45 50 55
age





3.2 The linguistic marker

Structuring our model, we started from the innovative linguistic marker proposed by Kovacic et al. (2016). The index draws its inspiration on the accurate work of grammatical analysis on moods in the European languages, made by Rothstein and Thieroff, and it is aimed at showing univocally the intensity of use of non-indicative moods in *irrealis* contexts.

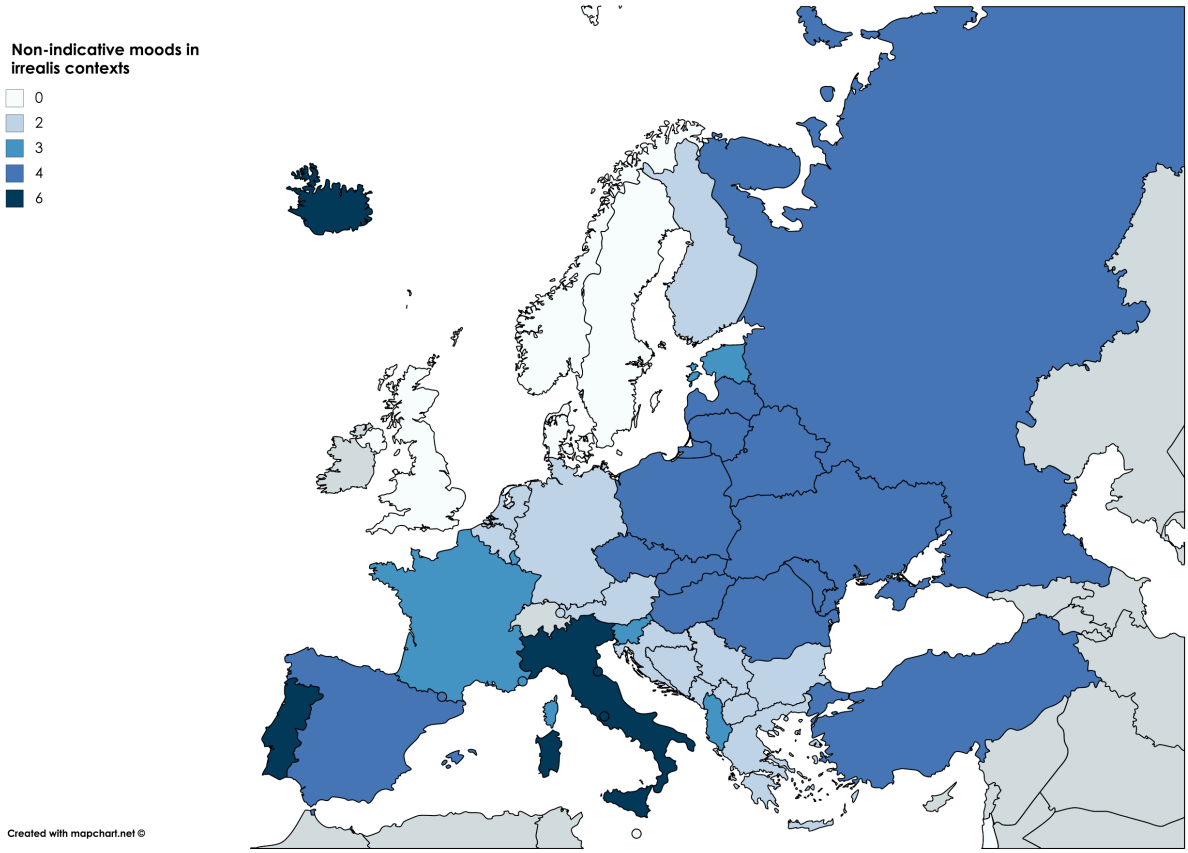
The author decided therefore to take a step forward with respect to the linguistic-saving hypothesis of Chen and his future time references. He dealt with *irrealis* as a categorical

variable, considering the use of non-indicative moods in some specific contexts as a paradigm of use of different moods in each language.

Although the survey by Rothstein and Thieroff probably represents the most detailed study about grammatical structures in Europe, their data has been supplemented by Kovacic et al. who "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)" (Kovacic et al. (2016), p. 10)

As set out above there are six contexts characterized by a more consistent use of *irrealis* moods, with a cross-linguistic point of view, including: complements of modal predicates (representing possibility, likelihood, necessity); complements of desiderative and volitional predicates (representing wish or desire); complements of epistemic predicates (representing thinking, belief or doubt); complements of emotive-factive predicates (representing regrets, happiness or sadness); complements of declarative predicates); the protasis (i.e. the dependent clause) and the apodosis (the main clause) in conditional sentences.

These contexts, taken as a ratio of use of *irrealis* moods, allow Kovacic to assign "value 1 to the occurrence of a non-indicative mood in each syntactic environment and 0 to indicative moods" (Kovacic et al. (2016), p.11) if it is used in the same situation. The summation of

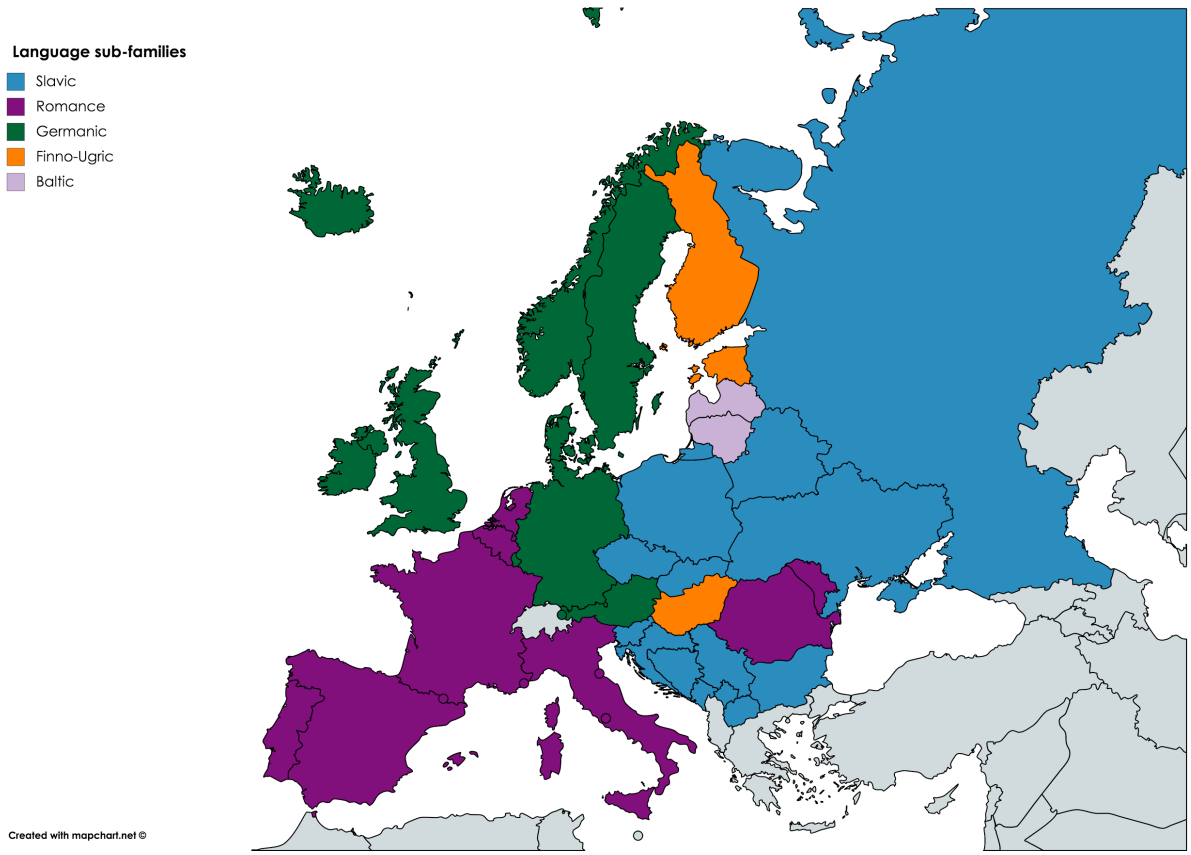


these values provides an univocal indicator of the frequency of use of non-indicative moods in each language of the sample, ranging from a minimum of zero to a maximum of six. The 39 languages analysed by Rothstein and Thieroff and later by Kovacic can therefore be ranked on the basis of this index. The ones in which the linguistic marker takes a zero value are called "moodless".

Assuming no qualitative differences among *irrealis* contexts, the summation was calculated using a uniform weighting function. In our thesis we focused then on 31 European and Eurasian languages, 27 of which are present in our subsample.

But let us now have a look at our work in relation to the analysis set out by Rothstein and Thieroff in Moods in the Languages of Europe. The two linguists analyzed in detail these languages with a particular attention to their common grammatical structure in displacement and modality, despite the different linguistic family, focusing on their shared features.

Examining in depth the usage of moods, linked to displacement and modality, we notice a strong variability amongst languages for what concerns the context of use, regardless of whether non-indicative moods are present or not. Some languages employ a non-indicative mood in these contexts, other languages can use both, either indicative or non-indicative moods, whereas others use the indicative despite the existence of non-indicative moods in their grammatical structure.



Looking briefly at their mood structures, all the European languages (except Maltese) have a non-indicative imperative mood, most of them have at least one non-indicative non imperative mood, the conditional or the subjunctive.

Kovacic identifies six moodless languages (Danish, English, Hebrew, Maltese, Norwegian, Swedish), eight languages using non-indicative moods in only two contexts (Bulgarian, Croatian, Dutch, Finnish, German, Greek, Macedonian, Serbian), seven languages using non-indicative moods in three contexts (Albanian, Basque, Catalan, Estonian, French, Slovenian, Welsh), fourteen languages with four non-indicative moods contexts (Arabic, Belarusian, Czech, Hungarian, Irish, Latvian, Lithuanian, Polish, Romanian, Russian, Slovak, Spanish, Turkish, Ukrainian) and in the end just three languages that use non-indicative moods in all of the six contexts (Italian, Icelandic, Portuguese).

This strong variability in use of *irrealis* moods is the base of our work in investigating and testing linguistic-relativity hypothesis applied to economic behaviour and economic choices related to risk and uncertainty.

The following graph shows the composition of our sample of languages with respect to the intensity of use of non-indicative moods in *irrealis* contexts, according to the linguistic marker created by Kovacic.

Figure 1 shows the distribution of immigrants on the basis of the intensity of use of *irrealis* in their native languages, regardless of their birthplace. More than a quarter of the sample are moodless speakers, i.e. no *irrealis* moods in their language; only 6% of the sample are intermediate *irrealis* user, i.e. the linguistic marker assumes a value between 1 and 3; in the end almost 70% are intensive *irrealis* users, with a linguistic marker value included between 4 and 6.

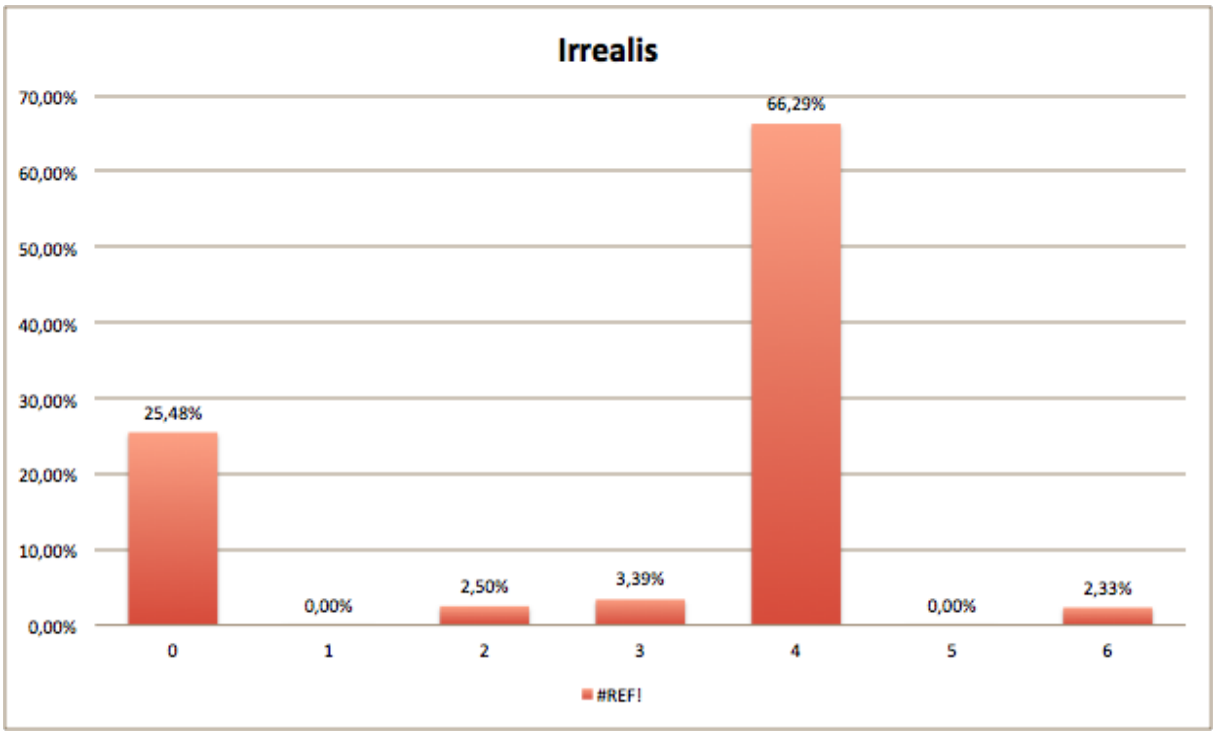


Figure 1: distribution of *irrealis* users in the sample on the basis of the linguistic marker

Source: American Community Survey, 2001-2015, language spoken at home by first-generation immigrants aged between 25 and 55 years.

The distribution of *irrealis* linguistic marker in our sample is extremely polarized around values 0 and 4 of the index because of the predictable massive presence of English-speaking and Spanish-speaking individuals. Those who speak these languages, in fact, represent almost the eighty-seven percent of the total sample. We therefore generated an *irrealis* dummy (*irrealis* henceforth) as main explanatory variable. Our binary variable takes value equal to 0 if the language spoken by the individual is a moodless language and 1 if the language is characterized by the use of non-indicative moods in at least one of the six cross-linguistic contexts mentioned in the previous section, identified as relevant for what concerns displacement and modality in the expression of possible situations.

A step forward was made generating an additional categorical variable (*catIRR* henceforth) which relates to variety in displacement and modality. This variable takes value 0 for moodless languages, value 1 for intermediate *irrealis* use languages (i.e. where the linguistic marker is included between 1 and 2) and value 2 for intensive *irrealis* use languages (i.e. where the linguistic marker ranges between 4 and 6).

Irrealis moods

	Modeless	Intermediate use	Intensive use	Total
Safe occupation	311,292	68,101	716,078	1,095,471
	61.93%	58.56%	53.02%	55.62%
Hazardous occupation	191,392	48,183	634,575	874,15
	38.07%	41.44%	46.98%	44.38%
Total	502,684	116,284	1,350,653	1,969,621
	100.00%	100.00%	100.00%	100.00%

3.3 Methods

The aim of this work is to investigate the association between conditional clauses, and particularly of the intensity of use of non-indicative moods in *irrealis* contexts, and work preferences of immigrants in the United States, with regard to the riskiness of their employment.

According to Kovacic et al. (2016) the intensity of use of *irrealis* moods has a positive correlation to risk aversion: individuals speaking a language where *irrealis* moods are used with greater intensity in expressing possible or unreal events should have a perception of the surrounding world as more unstable or mutable. This should affect their perception of the risks and therefore on their prudence and their aversion to risk, conditioning the economic behaviour of individuals with respect to those who are characterized as low intensity *irrealis* language speakers.

The authors focused first on the association between *irrealis* moods and risk aversion and then on the effect of the formers on stock ownership. The results presented by Kovacic are robust and highly significant, with a 12% higher probability of being risk averse for individuals who speak a language characterized by an intensive use of *irrealis* moods and an inverse causal relationship between risk aversion and asset accumulation with a coefficient of 11%.

The intuition that underlies the baseline model of our thesis is to analyse the association between the intensity of displacement into uncertainty and work preferences of individuals,

looking at employment in hazardous occupation. Two additional models look further at the association between *irrealis* moods and the propensity to become an entrepreneur as well as the effect on wage and salary income for a specific subsample of immigrant in the United States.

The idea, suggested ba Gay, Hicks and Santacreu-Vasut (2016) is to study migration as a microevolutionary step, taking advantage from the analysis of individuals with different historical, linguistic, social, economic and cultural background in a common and shared institutional environment as an host country.

We make no claim of being exhaustive in our analysis of the association above. This work does not want to identify the intensity of use of non-indicative moods or displacement into modality as the unique main driver of individual occupational choices.

We are fully aware of the intricacy of human thought and of the shaping of preferences. We simply suppose that certain linguistic differences and peculiarities could affect the perception of uncertainty, basically twisting the attitude toward risk of economic agents as like as their preferences.

Our baseline model, as previously touched upon, examine the association between the frequency of displacement into uncertainty in languages and the particular aspect of the propensity to be employed in hazardous occupations, as defined by the Labor Commissioner of the State of Connecticut.

The purpose is that our main dependent variable should clearly represent a paradigm of the individual's' attitude towards risk, if they are prudent and aware of the degree of riskiness of the job that they are involved in.

In all our models we used year dummies in order to capture the potential influence of aggregate trends of the time series, with a view to macroeconomic tendency between 2001 and 2015. On the other side we employed also a set of country-of-birth dummies for all the 79 country of our sample, in order to eliminate the effect of exogenous factors in the country of origin and therefore to account for cultural background and history of individuals before the year of migration in the United States.

The country-of-birth dummies, above all, represent a fundamental tool when you are dealing with immigrants and particularly with first generation immigrants. They are essential, indeed, to bear in mind the potential confounding effects in investigating a causal effect of differences in *irrealis* use amongst languages on economic behaviour which could be caused by innate characteristics such as social, institutional and cultural background as well as customs and traditions of the home country. Languages, indeed, could relate in two ways: a common ancestor (vertical dependence) and language contact (horizontal dependence) (Roberts et al.

(2015)). An epidemiological approach would have solved the problem of addressing ancestral characteristics, distinguishing noticeably native and host country effect, but as we mentioned above we had to abandon this way because of the structure of data and the lack of informations provided by the American Community survey.

The baseline model, as well as all the robustness checks, studies a sample of individuals aged between 25 and 55 years, namely the most relevant subset of the working age as ruled by the United States Department of Labor (more specifically 16 years to be allowed to work, up to 67 years for retirement, by 2025).

The organisation of our analysis required to structure a panel in order to control for any time-invariant differences between individuals and to account for unobservable characteristics that could bias the estimated coefficients, such as cultural factors, religion, gender, race and so on, controlling for individual heterogeneity

Given the structure of the yearly data obtained by the American Community Survey, we needed to identify univocally the observations in our dataset. In each year of the sample the personal identifying serial numbers are the same and repeated, generating a misrepresentation due to overlapping of different individuals. For this reason we generated a household serial number, grouping individuals on the basis of family clusters to exploit an univocal identification and to analyse variation at the actual household level. This model allow us to study several information correlated amongst families, identifying unobservable individual characteristics such as, for example, husband and wife sorting.

The fundamental reason of this choice, above all, is to generate and to work with fixed-effect models, employed in all our regressions. We selected this design in order to analyse as accurately as possible causal differences and changes within these family cluster, overcoming the shortage of parental characteristics, setting aside time-invariant traits of individuals and assuming a homogeneous environment to examine the association between language structures and economic behaviour.

As we were saying, the lack of detailed informations about ancestral and parental characteristics al well as other unobservables traits, regarding for example the language spoken by parents, their occupational status, their job preferences, their educational attainment or merely their wage and salary income, is the main reason that lead us to employ a fixed-effect linear model.

3.3.1 The association between irrealis and work preferences, entrepreneurship, wage and salary income

In our baseline model, in order to analyse the association between irrealis and work preferences, we regressed our high risk occupation dummy on the main explanatory variable, namely the irrealis dummy counterposing moodless languages to those with a non-indicative mood used in at least one of the six irrealis context mentioned in the previous section.

$$\text{highrisk} = \begin{cases} 1 \text{ high risk occupation} \\ 0 \text{ low risk occupation} \end{cases}$$

Then we tried to exploit more in detail the variability in the *irrealis* linguistic marker, substituting our explanatory dummy with the categorical variable (catIRR) which relates to linguistic differences in displacement and modality. CatIRR takes value 0 for moodless languages, value 1 for intermediate irrealis use languages (i.e. where the linguistic marker is between 1 and 2) and value 2 for intensive irrealis use languages (i.e. where the linguistic marker is between 4 and 6).

Both explanatory variables that represent the linguistic marker, *irrealis* and *catIRR*, were used in the same way to try to explain the variability in our sample of immigrants for what concerns either the probability of becoming an entrepreneur or the logarithm of wage and salary income.

All the regressions control for age, gender, educational attainment, marital status and years spent in the United States since the migration. In our robustness checks we ran several regressions on specific subsamples: age groups, years of migration, on the basis of the continent of birth and particularly, if Europe, on the european region of provenance.

3.4 Empirical strategy

In our first set of regressions, ran on our ACS subsample of immigrants, we tried to define the probability that an individual *i*, living in the household *j* would choose a hazardous occupation or not, HR_{ij},

$$HR_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 irrealis_{ij} + hhid_j + \varepsilon_{ij}$$

where X_{ij} is a vector of individual socio-economic and demographic characteristics, including year dummies, such as age, gender, marital status, educational attainment and years of migration, $irrealis_{ij}$ is our linguistic marker dummy, identifying the existence and the use of non-indicative moods in *irrealis* contexts, $hhid_{ij}$ is a household fixed effect and ε_{it} is an error term.

The dependent variable HR_{it} is equal to 1 for individuals employed in high risk occupations and 0 otherwise.

In the drafting of our work we faced a literature on the potential correlation between grammatical differences and economic behaviour that is still weak and not developed enough. Despite this lack of robustness all the studies about this interconnection reveal a quite reliable evidence of the association between attitudes toward risk and many aspects of human life. Several works are focused on the effect of risk-aversion on a wide range of socio-economic characteristics, namely, for example, trust, income and wages, educational attainment, health conditions, occupational status, numerousness of the family or number of sons.

Indirectly, given the supposed relationship between displacement into uncertainty and attitudes toward risk, we receive interesting signals and hints related to the interpretation of our model.

Scrolling through the literature in relation to the topic of this thesis, for example, we notice a generally negative relationship between income and risk-aversion. Anyway, according to Barsky et al. (1997) and Hartog et al. (2002) this relationship may not be linear. This hypothesis is supported by several authors that suppose and find that the aversion to risk decreases in wealth: the more you have, the more you can afford (Riley Jr. and Chow (1992), Cohen and Einav (2007), Guiso and Paiella (2008), Dohmen et al. (2011)). Nonetheless there could be a problem of reverse causality between wealth and risk-aversion since a higher level of wealth may be explained by a stronger willingness to risk.

With regard to the employment status, the level of individual risk-aversion results significantly correlated with entrepreneurship and self-employment. Businessmen prove to be consistently more tolerant to risk with respect to employees. (see Hartog et al. (2002), Siegel and Hoban (1991)). Surprisingly on the other hand, according to Hartog et al. (2002) and Dohmen et al. (2011), there seems not to exist a correlation between unemployment status and individual attitude toward risk.

Going over, the effect of educational attainment on risk-aversion seems to be clear: as the years of education increase the individual results more risk tolerant with respect to uneducated ones. (Dohmen et al. (2011), Lin (2009), Bellante and Green (2004), Riley Jr. and Chow (1992)).

Furthermore, even though the number of children positively correlates with the level of risk-aversion, the numerousness of the family seems to have the opposite effect, increasing risk tolerance (Siegel and Hoban (1991), Dohmen et al. (2011), Lin (2009)).
In the end, according to Barsky et al. (1997), Jianakoplos and Bernasek (1998)), Cohen and Einav (2007), Dohmen et al. (2011) women result to be more risk averse with respect to men.
Other factors possibly influencing individual attitudes toward risk, such as health status, religion or cognitive decline, were not addressed in our thesis.

3.5 Results

The empirical estimations exposed in the previous paragraph are presented in detail from Table 1 to Table 14 in the Appendix B of this thesis.
Models in Tables 3 to Table 14 analyse some specific subsamples of regressions on the three dependent variables. These are based on year spent in the United States, age groups, continents of birth and in particular, if Europe, the european region of provenance.
A higher intensity of use of non-indicative moods in irrealis contexts, and generally of displacement into uncertainty, is associated indeed to a higher propensity to be employed in high risk occupations, with respect to speakers of moodless languages. Females, on average more risk-averse with respect to males, are less involved in high-risk occupations, being married is generally associated with a higher propensity to hazardous jobs and there is a positive association between the years spent in the United States and the riskiness of the job. Educational attainment has, in line with the literature, a positive effect on wages and salaries but confirms the negative effect on entrepreneurship.
The results obtained using the categorical variable representing the linguistic marker (catIRR) are coherent to the baseline ones, showing a consistent coefficients both for the intermediate use of non-indicative moods and for the intensive use of irrealis moods.
In order to examine in depth the association between conditional clauses and work preferences we chose to analyse also effect of irrealis use on individual wage and salary income. Consistently with the results which shows a positive effect of irrealis moods on the riskiness of the job done, individuals speaking a language with a higher displacement into uncertainty earn better wages because of the positive correlation between risk and salary compensation.
Looking at the third and last outcome variable of our model, we examined the potential association between the frequency of use of conditional clauses and the probability of becoming an entrepreneur and forming a new venture. Entrepreneurship is usually considered as one of the exemplary choice involving taking risks, since it entails proposing, developing

and running a new business. It intrinsically implicates the investment of venture capital, choices under uncertainty, that is to say business risk. This is the only case, in our work, in which the degree of displacement and modality has a negative effect on economic choices of individuals.

In Tables 3 to Table 14 we decided to restrict our sample. The sub-samples are built on the basis of a) years lived in the United States since the migration; b) age groups; c) continent of birth and particularly, if Europe, on the european region of provenance. The results, if statistically significant, are consistent with the baseline ones.

According to the work by Kovacic et al. (2016), the first to analyse the economic implications of displacement into uncertainty, individuals who use to speak a language characterised by the existence of non-indicative moods in at least one of the six contexts of uncertainty previously mentioned are on average 12% more risk averse with respect to moodles-language speakers. In their robustness check the authors identify also a consistent effect of irrealis moods on investment choices and stock ownership, with a negative correlation between the linguistic marker and the amount invested in shares and financial products.

4. Robustness check

In our main robustness check we chose an epidemiological approach, as a robustness check, on a sample of second-generation immigrants, as postulated by Fernandez (2010), to study variations in the association between conditional clauses and work preferences across different immigrant groups in the same country, the United States.

Individuals of the same group therefore share a common institutional and economic environment but are characterised by different cultures of origin and thus with potentially differing social beliefs. This allows one to separate the effect of culture and language from the original economic and institutional environment. The approach recalls the one used by epidemiologists who compare health outcomes from immigrants and natives in order to identify the genetic contribution to diseases descending from the physical environment.

Along with the example proposed by Fernandez (2010), focusing on weaknesses and strengths of the epidemiological approach, we start from a medical issue, namely heart disease incidence in two different countries, the origin and the host ones. If we notice a convergence in heart disease in natives and immigrants it is probable that the difference between the two countries is not due to genetic causes but to environment. No convergence, on the other side, does not necessarily imply the opposite. As a matter of fact, different levels of heart disease could derive from several reasons: selection in immigration could relate with a specific health outcome, for example, early immigration could cause a kind of immunity or a slow cultural assimilation could occur (e.g. unhealthy behaviours such as no varying diet or sedentary life hailing from the source country).

According to Fernández first-generation immigrants are more likely to suffer several confounding factors with respect to second-generation ones such as the significance of ties with non immigrants' family members, the degree of proficiency of the host country level and the effects of the ancestral culture of their country of birth.

Furthermore there exist a significant evidence of transmission and persistency of human attitudes and cultural traits from parents to children, inserted in a shared environment. Dohmen et al. (2006) focused on the intergenerational transmission of risk and trust attitudes, identifying a robust effect with heterogeneity due to the family structure. (suggestivo, difficile da individuare nel nostro lavoro)

This leaded us to analyze just second-generation immigrants in our model. We chose this strategy not only because of the intergenerational transmission of ancestral characteristics, which vary across immigrant groups reflecting country of origin culture, but also because, as mentioned above, individuals living in the same country face the same institutional and

economic environment. “The idea is thus that individuals from different cultures will take different actions despite facing identical environments.” (Fernández (2010), p.12)

We exploit variations in language spoken by different second-generation immigrants, with the same parents’ country of birth, to single out the effect of a language structure on economic choices from the effect of ancestral cultural traits.

Taking the cue from Galor et al., in contrast to the previous studies that used an epidemiological approach, we mapped univocally second-generation immigrants cultural traits with parental country of origin, overcoming several potential biases due to important omitted ancestral variables.

When using an epidemiological approach, an important strategy is to use an ancestors’ country of birth fixed effect to reveal unobservable cultural characteristics such as social norms or customs and traditions that could have been transmitted between the first and the second generation of immigrants.

The epidemiological approach is exposed to several critiques. For sure parents are not the only conveyors of culture, since the local environment, for example, clearly affect the formation of children. Individuals are strongly influenced by the surrounding, with a strong influence of social norms built by institutions, schools or neighbourhoods et cetera on culture. On the other hand the effect of mother tongue could have been mitigated in second-generation immigrants because of integration in the hosting country but also because probably our individuals use to speak English fluently when they are not at home, balancing out the effect of their original language.

4.1 Data and methods

The empirical analysis, as in the baseline model, is run on a family basis on a panel of individuals in the American Community Survey by the U.S. Census Bureau.

The respondents come from 78 countries around the world and they are all residing in the 49 federal states of the U.S.. We selected, from 2001 to 2015 (ACS, 5% sample), all the foreign born non american immigrants, aged between 15 and 65 years (516,567 observations), speaking 27 different eurasian languages as defined by the Dahl’s categorization, including just those with a complete set of information about language and nationality.

In our analysis we controlled for individual characteristics such as educational attainment, age, years spent in the United States since the migration, as a proxy of cultural integration, marital status.

Above all we controlled for parental characteristics such as wage and salary income, educational attainment and the involvement in hazardous occupation of mother and father of

individuals. As in the baseline model we employed year dummies, in order to capture the potential influence of aggregate trends of the time series, and country-of-birth dummies, to eliminate the effect of exogenous factors in the country of origin and therefore to account for cultural background and history of parents.

Our regressions are fixed effect models, grouping individuals on the basis of family clusters to analyse as accurately as possible causal differences and changes within these family cluster. Starting by our sample of second-generation immigrants we try to define the probability that an individual i with an ancestors' country c speaking a specific language l would choose a hazardous occupation or not, HR_{icl},

$$HR_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 irrealis_{ij} + hhid_{ij} + \varepsilon_{ij}$$

where X_{it} is a vector of individual socio-economic and demographic characteristics, including year dummies (e.g. age, gender, marital status, educational attainment and years of migration), $irrealis_{ij}$ is the linguistic marker dummy, identifying the use of non-indicative moods in *irrealis* contexts, $hhid_{ij}$ is a household fixed effect and ε_{it} is an error term.

The main dependent variable HR_{ij}, as in the baseline model, is equal to 1 for individuals employed in high risk occupations and 0 otherwise. The other two outcome variables are entrepreneurship and the logarithm of salary and wage income.

4.2 Results

The empirical estimations of the preceding paragraph are presented in detail in Table 15, Table 16 and Table 17 of the Appendix C of the thesis.

As in our baseline model with first-generation immigrants, a higher intensity of displacement into uncertainty, i.e. more non-indicative moods employed in irrealis contexts, positively correlates to a higher propensity to be employed in hazardous jobs, with respect to speakers of moodless languages. This effect is mitigated, but still strongly significant, compared to the baseline model.

The overall effect of displacement into uncertainty on wage and salary income and entrepreneurship is consistent and statistically significant too. The effect of irrealis moods on entrepreneurship, however, turns to be positive conversely to the coefficient identified in the baseline model.

When we account for parental characteristics of second-generation immigrants we find several interesting hints.

A higher mother's or father's educational attainment, for example, consistently reduces the probability of being employed in a high risk occupation.

Parental wage and salary income is positively correlated to the probability of conducting a high risk occupation. The association with individual's income is equally positive and statistically significant for parental incomes up to 48000$. On the contrary they have a consistently negative effect on the propension to start a new venture.

Going further, on the other hand, individuals with parents working in hazardous jobs are more likely to be hired in a risk-similar employment.

The results obtained in the models which exploit the irrealis categorical variable representing the linguistic marker (catIRR) are coherent to the baseline ones, showing a consistent coefficient both for the intermediate use of non-indicative moods and for the intensive use of irrealis moods.

5. Conclusions

The aim of this work is to take one step further in the investigation of the intriguing association between differences in language structures and individual attitude toward risk, starting from the Sapir-Whorf hypothesis of linguistic relativity. The existence and the intensity of the interaction between language and human thought is still on debate and our purpose is to add another piece in this analysis.

In order to identify the intensity of displacement into uncertainty we adopt the specific linguistic marker proposed by Kovacic et al. (2016). It is based and structured on the number of non-indicative moods, such as Konjunctive, Conditional or Imperative, used in irrealis contexts. These are the contexts that require the use of grammatical structures straightly related to the expressions of uncertainty. This index is inspired by the study of Rothstein and Thieroff, which analyses tenses and moods in 39 different European languages, and is aimed at showing univocally the intensity of displacement into uncertainty.

The empirical strategy of this thesis aspires to test the hypothesis that there exist an association between the intensity of displacement into uncertainty in language and work preferences of individuals with regards to the riskiness of the job. On the one hand, in our view, irrealis-intensive speakers could perceive the surrounding world with a stronger degree of mutability, a trait that could be observed looking at individual risk-aversion. On the other hand, a more precise language where displacement between actual and hypothetical states of the world is perceived as larger could enhance the capability of individuals in interpreting reality.

In our opinion a wider point of view is therefore required in the interpretation of the association between conditional clauses and economic behaviour of individuals. Testing only for risk-aversion is probably too restrictive in the analysis of human behaviour because of the limiting assumption that individuals face and evaluate the same risks in the same identical way.

The perception of risk is indeed based on individual judgement and evaluation of hazardous events, activities or technologies and could strongly differ amongst economic agents, not necessarily corresponding to the degree of risk-aversion.

One of the first attempts to analyse the drivers of risk perception is the psychological approach championed by Paul Slovic in 1987. His psychological analysis focuses on technological evolution, in particular on nuclear and chemical technologies, and the displacement between actual and perceived risk related to potential damages to earth and humans. The psychometric approach adopted by the scholar is useful to create a sort of

taxonomy of hazards in order to identify the characteristics that lead to specific valuations of risk and analyse human responses to potentially damaging events.

The main insight is that individuals tend to be intolerant to all those risks that are perceived as not manageable or uncontrollable, implying a distribution of risks and benefits that is not equitable. Slovic expose the paradox of people persisting in extreme fear of nuclear accident but almost fearless of the risk of car accidents, despite the tremendously huge difference in death rates. The same reasoning could explain, for example, an irrational fear of genetically modified organisms.

The next step in our work was therefore to analyse the relation between linguistic differences and economic choices through the gateway of risk perception, focusing on the drivers of the attitude toward risk. Since risk perception influences risk-taking or risk-averse actions we need to identify what leads to variation in people's risk perception and how this is related to the intensity of displacement into uncertainty.

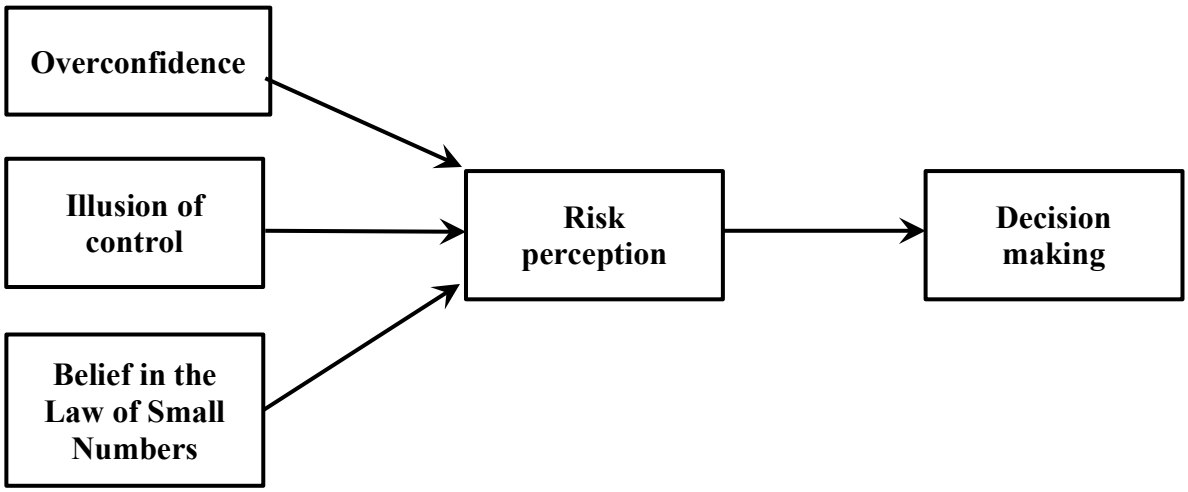
According to scholars, cognitive biases could be the answer that we were looking for. Cognitive biases are "common types of mental shortcuts used to make judgments" (Simon et al. (2000), p.1) and they could consistently affect the risk perception and therefore the decision making process. We start from the assumption that an intensive use of non-indicative moods in irrealis contexts is related to a clearer sense of the world, perceived as mutable but in a more precise way. Hence intensive irrealis-language speakers are more accurate in individual risk assessment, relying on a more correct risk perception. On the other hand, moodless speakers neither look for nor precisely understand informations because of their limited cognitive capacity that leads them to employ heuristic or simplifying strategies that cause cognitive biases when assessing risks.

There exist several cognitive biases that may influence this processes but we hold just three of these to be true in affecting work preferences, wage and salary income and the decision to create a new venture: overconfidence, above all, the illusion of control and the belief in the law of small numbers.

Moore and Healy defined overconfidence in three different ways: "1) overestimation of one's actual performance, 2) overplacement of one's performance relative to others, and 2) excessive precision in one's beliefs" , i.e. an unwarranted certainty (Moore and Healy (2008), p. 2). Overconfident individuals may consider their assumptions as facts, emphasizing the degree of uncertainty associated a specific choice and erroneously considering a certain action as risky despite the actual probability of being so. This has a direct effect, for example, on tasks performed or on job seeking.

Illusion of control is a self-explanatory expression. It concerns the inclination of individuals to behave as they have the total control when actually they had none, overestimating the role of skills as crucial factors when chance is fundamental instead. Overconfidence is someway related to the overestimation of the informations regarding actual facts whereas the illusion of control concerns an overestimation of skills and subsequently of her ability to predict and deal with future possible events.

The third and last cognitive bias we account for is the belief in the law of small numbers. This instance regards individuals drawing based on a small and restricted sample of informations, rising what he or she knows to absolute and ignoring, for example, base-rate probabilities. The main issue related to this bias is that it leads people to consider spot or infrequent events, especially as smaller is the sample, as sufficient to the assessment of risk, twisting individual



perception of the latter. Let's think, for example, about people who are very frightened by airplane crashes and consider them more dangerous than smoking, despite tobacco causes many more deaths than the former.

For the reasons exposed above, cognitive biases may be the gateway to explain the results of our empirical exercise and understand the different behaviour of irrealis speakers.

Cognitive biases, and particularly the belief in the law of small numbers, lead individuals speaking a moodless language to consider potential risks related to hazardous occupations as an ungovernable risk that can not be submitted to her or his direct control. On the other hand, because of bias due to overconfidence and illusion of control, moodless-language speakers feel completely able to manage the risk of starting a new venture and becoming an entrepreneur.

We can shape a defined distinction between risks related to hazardous occupation, perceived as external and potentially ineluctable, and risks related to entrepreneurship, perceived as completely manageable and containable by the individual.

Our empirical results are completely coherent with this interpretation. As we mentioned in the previous chapter, individuals speaking a language characterised by non-indicative moods, employed in at least one of the six irrealis contexts, show a clearer cognition of the surrounding world and of the displacement between actual and possible events. This more precise perception of the risk, linked to language spoken, leads people to be more prudent and cautious and to abstain from starting a new unsuccessful venture. At the same time this individuals are aware of the actual risk deriving by a hazardous occupation, regardless of their degree of risk-aversion, and could choose it consciously.

These intuitions are confirmed by the positive effect of the intensity of use of irrealis moods on wage and salary income. Individuals undertaking a high risk occupation are remunerated with higher wages because of the positive correlation between risk and salary compensation. In the same way, avoiding the creation of idle ventures prevent from potential losses.

We also investigated the association between conditional clauses, work preferences, wage and salary income and entrepreneurship exploiting an epidemiological approach in order to take advantage of variations in language spoken by a sample of second-generation immigrants, with the same parents' country of birth, in order to distinguish the effect of a grammatical structure on economic choices from the effect of ancestral cultural traits.

The results of our robustness check are in line with the baseline ones.

The epidemiological approach is still exposed to numerous critiques. Certainly parents are not the only conveyors of culture. Local environment, as an instance, may affect the formation of children. There certainly exists a strong influence on culture deriving from social norms built by institutions, schools or neighbourhoods et cetera. The effect of mother-tongue, at the same time, could result as mitigated in second-generation immigrants because of gradual integration in the hosting country but also because individuals probably use to speak English fluently when they are not at home, balancing out the effect of their original language.

It would be important, moreover, to further develop our model, focusing on linguistic characteristics of individuals in our sample to address language relatedness. We may indeed identify two ways in which languages could relate: a vertical dependence, deriving from a common ancestor, or a horizontal dependence, due to language contact. (Gay et al., 2016) Structuring and employing a linguistic-area fixed effect could address the problem of language relatedness.

The approach we adopted in this thesis, as far as we know, is one of the first attempt to analyse empirically the association amongst language, risk perception and economic choices. As long as language, in some way, is a marker for individual identity it proved to be an important instrument in the investigation of non-economic factors that influence human behaviour, starting from risk assessment and risk perception.

Appendix A

Linguistic mapping

Language	Family	Sub-family	# Moods	∑ Non Ind
Albanian	Indo-euro	-	>2	3
Arabic	Semitic	-	2	4
Basque	Isolate	-	2	3
Belarusian	Indo-euro	Slavic	1	4
Bulgarian	Indo-euro	Slavic	1	2
Catalan	Indo-euro	Romance	1	3
Croatian	Indo-euro	Slavic	1	2
Czech	Indo-euro	Slavic	0	4
Danish	Indo-euro	Germanic	0	0
Dutch	Indo-euro	Germanic	0	2
English	Indo-euro	Germanic	0	0
Estonian	Indo-euro	Finno-ugric	2	3
Finnish	Uralic	Finno-ugric	2	2
French	Indo-euro	Romance	1	3
German	Indo-euro	Germanic	1	2
Greek	Indo-euro	-	1	2
Hebrew	Semitic	-	0	0

Hungarian	Uralic	Finno-ugric	2	4
Icelandic	Indo-euro	Germanic	1	6
Irish	Indo-euro	Celtic	2	4
Italian	Indo-euro	Romance	2	6
Latvian	Indo-euro	Baltic	1	4
Lithuanian	Indo-euro	Baltic	1	4
Macedonian	Indo-euro	Slavic	1	2
Maltese	Semitic	-	0	0
Norwegian	Indo-euro	Germanic	0	0
Polish	Indo-euro	Slavic	1	4
Portuguese	Indo-euro	Romance	2	6
Romanian	Indo-euro	Romance	1	4
Russian	Indo-euro	Slavic	1	4
Serbian	Indo-euro	Slavic	1	2
Slovak	Indo-euro	Slavic	1	4
Slovenian	Indo-euro	Slavic	1	3
Spanish	Indo-euro	Romance	1	4
Swedish	Indo-euro	Germanic	0	0
Turkish	Ural-altaic	Turkic	>2	4
Ukrainian	Indo-euro	Slavic	1	4

Welsh	Indo-euro	Celtic	1	3

List of the languages

English, Swedish, Danish, Norwegian, Hebrew/Israeli, German, Dutch, French, Greek, Albanian, Serbo-Croatian, Yugoslavian, Slavonian, Finnish, Italian, Spanish, Portuguese, Rumanian, Russian, Ukrainian, Ruthenian, Little, Russian, Czech, Polish, Slovak, Lithuanian, Persian, Iranian, Farsi, Other, Persian, Magyar/Hungarian, Turkish, Arabic.

List of birthplaces

Afghanistan, Albania, Atlantic Islands, Australia and New Zealand, Austria, Belgium, Bulgaria, Cambodia (Kampuchea), Canada, Central America, China, Cuba, Cyprus, Czechoslovakia, Denmark, England, Estonia, Finland, France, Germany, Greece, Guam, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel/Palestine, Italy, Japan, Jordan, Korea, Kuwait, Laos, Latvia, Lebanon, Lithuania, Malaysia, Mexico, Nepal, Netherlands, Norway, USSR/Russia, Pacific Islands, Philippines, Poland, Portugal, Puerto Rico, Romania, Saudi Arabia, Scotland, Singapore, Spain, Sweden, Switzerland, Syria, Thailand, Turkey, U.S., Virgin Islands, United Arab Emirates, United Kingdom, Vietnam, Wales, West Indies, Yemen, Arab Republic, (North), Yugoslavia

Appendix B: Baseline Regression Tables

Table 1: Fixed Effect Linear Model. High Risk Occupation, Income and Wage and Entrepreneurship on Irrealis Dummy

VARIABLES	(1) High risk occupation	(2) Income and wage	(3) Entrepreneurship
Irrealis dummy	0.037***	0.032**	-0.009**
	(0.006)	(0.016)	(0.004)
Age	0.002***	0.006***	0.002***
	(0.000)	(0.000)	(0.000)
Female	-0.441***	-0.572***	-0.004***
	(0.001)	(0.003)	(0.001)
Marital status	0.017***	0.046***	0.008***
	(0.003)	(0.006)	(0.002)
Education level = 1, Medium Education	0.004*	0.068***	-0.004**
	(0.003)	(0.007)	(0.002)
Education level = 2, High education	-0.010	0.570***	-0.049***
	(0.007)	(0.018)	(0.005)
Years in the USA, categorical = 1, Between 10 and 15 years	0.023***	0.117***	0.001
	(0.003)	(0.006)	(0.002)
Years in the USA, categorical = 2, Between 15 and 22 years	0.040***	0.199***	-0.000
	(0.003)	(0.007)	(0.002)
Years in the USA, categorical = 3, Between 22 and 30 years	0.049***	0.239***	-0.004*
	(0.003)	(0.008)	(0.002)
Years in the USA, categorical = 4, More than 30 years	0.030***	0.278***	-0.010***
	(0.004)	(0.010)	(0.003)
Constant	0.652***	9.840***	-0.078**
	(0.059)	(0.135)	(0.036)
Observations	801,109	575,087	675,047
R-squared	0.305	0.189	0.005
Number of hhid	439,017	383,043	418,196
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Family clusters	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, Years in the United States = 0 (less than 10 years)

Table 2: Fixed Effect Linear Model. High Risk Occupation, Income and Wage and Entrepreneurship on Categorical Irrealis (catIRR)

VARIABLES	(1) High risk occupation	(2) Income and wage	(3) Entrepreneurship
Irrealis moods, categorical = 1, Intermediate use	0.036**	0.073*	-0.016
	(0.016)	(0.039)	(0.010)
Irrealis moods, categorical = 2, Intensive use	0.037***	0.026	-0.008*
	(0.006)	(0.016)	(0.005)
Age	0.002***	0.006***	0.002***
	(0.000)	(0.000)	(0.000)
Female	-0.441***	-0.572***	-0.004***
	(0.001)	(0.003)	(0.001)
Marital status	0.017***	0.046***	0.008***
	(0.003)	(0.006)	(0.002)
Education level = 1, Medium Education	0.004*	0.068***	-0.004**
	(0.003)	(0.007)	(0.002)
Education level = 2, High education	-0.010	0.570***	-0.049***
	(0.007)	(0.018)	(0.005)
Years in the USA, categorical = 1, Between 10 and 15 years	0.023***	0.117***	0.001
	(0.003)	(0.006)	(0.002)
Years in the USA, categorical = 2, Between 15 and 22 years	0.040***	0.199***	-0.000
	(0.003)	(0.007)	(0.002)
Years in the USA, categorical = 3, Between 22 and 30 years	0.049***	0.239***	-0.004*
	(0.003)	(0.008)	(0.002)
Years in the USA, categorical = 4, More than 30 years	0.030***	0.279***	-0.010***
	(0.004)	(0.010)	(0.003)
Constant	0.652***	9.841***	-0.078**
	(0.059)	(0.135)	(0.036)
Observations	801,109	575,087	675,047
R-squared	0.305	0.189	0.005
Number of hhid	439,017	383,043	418,196
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controlss	YES	YES	YES
Family clusters	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: catIRR = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, Years in the United States = 0 (less than 10 years)

Table 3: Fixed Effect Linear Model. High Risk Occupation on Irrealis Dummy, based on Years Lived in the United States

VARIABLES	(1)	(2)	(3)
	High Risk Occupations		
Irrealis dummy	0.036***	0.024**	0.052*
	(0.010)	(0.009)	(0.028)
Age	0.001***	0.003***	0.003**
	(0.000)	(0.000)	(0.001)
Female	-0.450***	-0.447***	-0.407***
	(0.002)	(0.002)	(0.006)
Marital status	0.010***	0.026***	0.052***
	(0.004)	(0.004)	(0.020)
Education level = 1, Medium Education	0.004	0.002	0.022
	(0.004)	(0.004)	(0.014)
Education level = 2, High education	-0.048***	0.001	0.128***
	(0.010)	(0.011)	(0.037)
Constant	0.559***	0.591***	0.661***
	(0.112)	(0.127)	(0.139)
Observations	406,599	472,769	77,151
R-squared	0.304	0.317	0.278
Number of hhid	257,811	305,506	63,794
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Female	YES	YES	YES
Family clusters	YES	YES	YES
Years in the USA	t < 10	10 < t < 30	t > 30

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education. Years in the United States categories are less than 10 years/betweeen 10 and 30 years/more than 30 years.

Table 4: Fixed Effect Linear Model. High Risk Occupation on Irrealis Dummy, based on Age Groups

VARIABLES	High Risk Occupations		
Irrealis dummy	0.012	0.038***	0.033**
	(0.012)	(0.013)	(0.015)
Female	-0.488***	-0.448***	-0.399***
	(0.002)	(0.002)	(0.003)
Marital status	0.010**	0.023***	0.029***
	(0.005)	(0.007)	(0.011)
Education level = 1, Medium Education	-0.007	-0.001	0.012*
	(0.006)	(0.006)	(0.006)
Education level = 2, High education	-0.013	-0.023	0.019
	(0.016)	(0.014)	(0.015)
Constant	0.747***	0.721***	0.906***
	(0.091)	(0.159)	(0.136)
Observations	283,248	299,424	218,437
R-squared	0.345	0.326	0.274
Number of hhid	191,021	220,938	162,987
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Female	YES	YES	YES
Family clusters	YES	YES	YES
Age	25-35 yrs	35-45 yrs	45-55 yrs

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education. Age groups are 25-35 years/35-45 years/45-55 years.

Table 5: Fixed Effect Linear Model. High Risk Occupation on Irrealis Dummy, based on Continent of Birth

VARIABLES	North America	Central America	South America	Europe	Asia	Africa	Australia, New Zealand, Pacific Island, Antartica
Irrealis dummy	0.059*	0.026***	0.019	0.031	0.001	0.021	0.112
	(0.035)	(0.007)	(0.031)	(0.024)	(0.036)	(0.057)	(0.307)
Age	0.000	0.002***	0.004***	0.003***	0.003***	0.002	0.004
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
Female	-0.284***	-0.488***	-0.369***	-0.268***	-0.168***	-0.198***	-0.166***
	(0.009)	(0.001)	(0.005)	(0.005)	(0.008)	(0.015)	(0.050)
Marital status	0.055**	0.015***	0.017	0.034*	-0.005	0.091***	-0.142
	(0.024)	(0.003)	(0.011)	(0.018)	(0.021)	(0.035)	(0.231)
Education level = 1, Medium Education	0.031	0.006**	0.033*	-0.003	0.078***	0.104**	0.022
	(0.026)	(0.003)	(0.019)	(0.023)	(0.024)	(0.049)	(0.253)
Education level = 2, High education	0.088**	0.013	0.006	-0.053**	0.172***	0.085	0.028
	(0.044)	(0.009)	(0.025)	(0.027)	(0.030)	(0.058)	(0.265)
Constant	0.997***	0.607***	0.424***	-0.262***	-0.598***	0.247***	0.858**
	(0.213)	(0.073)	(0.039)	(0.099)	(0.072)	(0.085)	(0.401)
Observations	24,055	604,795	53,967	69,566	35,128	12,210	1,293
R-squared	0.145	0.353	0.221	0.148	0.081	0.083	0.086
Number of hhid	17,792	315,024	35,977	47,516	26,344	9,281	1,097
Year fixed effect	YES	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES	YES
female	YES	YES	YES	YES	YES	YES	YES
Continent of birth	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables is the high risk occupation dummy. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Table 6: Fixed Effect Linear Model. High Risk Occupation on Irrealis Dummy, based on European Region of Provenance

VARIABLES	Europe	Northern Europe	Western Europe	Southern Europe	Central Europe	Eastern Europe
Irrealis dummy	0.031	0.112	0.016	0.006	0.020	0.033
	(0.024)	(0.307)	(0.036)	(0.069)	(0.050)	(0.050)
Age	0.003***	0.004	0.010	0.003*	0.003***	0.003***
	(0.001)	(0.006)	(0.011)	(0.002)	(0.001)	(0.001)
Female	-0.268***	-0.166***	-0.021	-0.363***	-0.365***	-0.232***
	(0.005)	(0.050)	(0.050)	(0.011)	(0.009)	(0.007)
Marital status	0.034*	-0.142	-0.306	-0.013	0.071**	0.070**
	(0.018)	(0.231)	(0.343)	(0.041)	(0.033)	(0.029)
Education level = 1, Medium Education	-0.003	0.022	1.178***	-0.073**	0.061	0.139***
	(0.023)	(0.253)	(0.143)	(0.030)	(0.067)	(0.048)
Education level = 2, High education	-0.053**	0.028	0.991***	-0.061	-0.086	0.151***
	(0.027)	(0.265)	(0.145)	(0.053)	(0.071)	(0.054)
Constant	-0.262***	0.858**	-1.772***	0.543***	0.299	0.217*
	(0.099)	(0.401)	(0.612)	(0.125)	(0.190)	(0.123)
Observations	69,566	1,293	1,072	12,054	17,668	26,104
R-squared	0.148	0.086	0.073	0.251	0.250	0.116
Number of hhid	47,516	1,097	889	8,136	12,579	17,380
Year fixed effect	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES
Female	YES	YES	YES	YES	YES	YES
European region		YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables is the high risk occupation dummy. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Table 7: Fixed Effect Linear Model. High Risk Occupation on Categorical Irrealis (catIRR) based on Year Lived in the United States

VARIABLES	(1)	(2)	(3)
	High risk occupation		
Irrealis moods, categorical = 1, Intermediate use	0.038	0.007	-0.068
	(0.024)	(0.028)	(0.074)
Irrealis moods, categorical = 2, Intensive use	0.036***	0.026***	0.067**
	(0.010)	(0.010)	(0.029)
Age	0.001***	0.003***	0.003**
	(0.000)	(0.000)	(0.001)
Female	-0.450***	-0.447***	-0.407***
	(0.002)	(0.002)	(0.006)
Marital status	0.010***	0.026***	0.051***
	(0.004)	(0.004)	(0.020)
Education level = 1, Medium Education	0.004	0.002	0.022
	(0.004)	(0.004)	(0.014)
Education level = 2, High education	-0.048***	0.001	0.128***
	(0.010)	(0.011)	(0.037)
Constant	0.559***	0.591***	0.661***
	(0.112)	(0.127)	(0.138)
Observations	406,599	472,769	77,151
R-squared	0.304	0.317	0.278
Number of hhid	257,811	305,506	63,794
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Family clusters	YES	YES	YES
female	YES	YES	YES
years in the USA	t < 10	10 < t < 30	t > 30

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: catIRR = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, Years in the United States = 0 (less than 10 years)

Table 8: Fixed Effect Linear Model. Entrepreneurship on Irrealis Dummy, based on Year Lived in the United States

VARIABLES	(1)	(2) Entrepreneurship	(3)
Irrealis dummy	-0.004	-0.008	-0.000
	(0.006)	(0.007)	(0.023)
Age	0.002***	0.003***	0.003***
	(0.000)	(0.000)	(0.001)
Female	0.009***	-0.006***	-0.037***
	(0.001)	(0.001)	(0.005)
Marital status	0.006***	0.016***	0.025*
	(0.002)	(0.003)	(0.015)
Education level = 1, Medium Education	-0.003	-0.003	-0.002
	(0.003)	(0.003)	(0.013)
Education level = 2, High education	-0.045***	-0.057***	-0.062**
	(0.008)	(0.008)	(0.025)
Constant	-0.016	-0.194*	-0.133
	(0.055)	(0.100)	(0.119)
Observations	334,653	402,589	66,924
R-squared	0.004	0.006	0.021
Number of hhid	228,308	285,692	57,788
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Family clusters	YES	YES	YES
female	YES	YES	YES
years in the USA	t < 10	10 < t < 30	t > 30

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education. Years in the United States categories are less than 10 years/betweeen 10 and 30 years/more than 30 years.

Table 9: Fixed Effect Linear Model. Entrepreneurship on Irrealis Dummy, based on Continent of Birth

VARIABLES	(1) North America	(2) Central America	(3) South America	(4) Europe	(5) Asia	(6) Africa	(7) Australia, New Zealand, Pacific Islands, Antartica
Irrealis dummy	-0.001	-0.009*	0.017	0.006	-0.005	-0.039	0.031
	(0.028)	(0.005)	(0.024)	(0.019)	(0.039)	(0.035)	(0.027)
Age	0.003***	0.002***	0.003***	0.003***	0.004***	0.003**	-0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.006)
Female	-0.024***	0.003***	0.023***	-0.052***	-0.071***	-0.040***	-0.034
	(0.006)	(0.001)	(0.004)	(0.003)	(0.007)	(0.011)	(0.034)
Marital status	0.002	0.008***	0.010	0.010	0.045**	0.052**	-0.014
	(0.014)	(0.002)	(0.008)	(0.013)	(0.020)	(0.025)	(0.254)
Education level = 1, Medium Education	-0.000	-0.006***	-0.019	0.017	0.078**	-0.062*	0.022
	(0.021)	(0.002)	(0.016)	(0.019)	(0.033)	(0.034)	(0.023)
Education level = 2, High education	0.029	-0.031***	-0.074***	-0.083***	-0.005	-0.063	0.071
	(0.034)	(0.006)	(0.022)	(0.023)	(0.040)	(0.039)	(0.087)
Constant	0.071	0.145**	0.036	1.003***	0.072	0.069	0.142
	(0.108)	(0.058)	(0.030)	(0.061)	(0.065)	(0.062)	(0.185)
Observations	19,763	504,609	47,812	62,278	28,766	10,586	1,148
R-squared	0.012	0.003	0.007	0.026	0.044	0.019	0.009
Number of hhid	15,835	302,132	33,596	45,109	23,342	8,463	1,004
Year fixed effect	YES	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES	YES
female	YES	YES	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES	YES	YES
Continent of birth	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables is the high risk occupation dummy. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Table 10: Fixed Effect Linear Model. Entrepreneurship on Irrealis Dummy, based on European Region of Provenance

VARIABLES	(1) Europe	(2) Northern Europe	(3) Western Europe	(4) Southern Europe	(5) Central Europe	(6) Eastern Europe
Irrealis dummy	0.006	0.031	0.327	-0.114	0.047	0.081**
	(0.019)	(0.027)	(0.287)	(0.075)	(0.042)	(0.035)
Age	0.003***	-0.001	-0.000	0.004***	0.004***	0.002**
	(0.000)	(0.006)	(0.005)	(0.001)	(0.001)	(0.001)
Female	-0.052***	-0.034	-0.016	-0.081***	-0.024***	-0.068***
	(0.003)	(0.034)	(0.036)	(0.009)	(0.008)	(0.005)
Marital status	0.010	-0.014	0.241	0.075**	-0.026	0.014
	(0.013)	(0.254)	(0.246)	(0.035)	(0.028)	(0.019)
Education level = 1, Medium Education	0.017	0.022	0.241*	0.019	0.098	0.015
	(0.019)	(0.023)	(0.123)	(0.027)	(0.074)	(0.033)
Education level = 2, High education	-0.083***	0.071	0.071	-0.107**	0.035	-0.101***
	(0.023)	(0.087)	(0.087)	(0.047)	(0.078)	(0.039)
Constant	1.003***	0.142	-0.339	0.091	-0.151	0.040
	(0.061)	(0.185)	(0.405)	(0.106)	(0.106)	(0.073)
Observations	62,278	1,148	985	10,485	15,720	23,651
R-squared	0.026	0.009	0.077	0.049	0.013	0.035
Number of hhid	45,109	1,004	835	7,654	11,906	16,554
Year fixed effect	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES
female	YES	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES	YES
European region		YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables is the high risk occupation dummy. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Table 12: Fixed Effect Linear Model. Income and Wage on Irrealis Dummy, based on Years Lived in the United States

VARIABLES	(1)	(2)	(3)
	Income and wage		
Irrealis dummy	0.004	0.018	0.178**
	(0.025)	(0.025)	(0.074)
Age	0.007***	0.008***	0.006**
	(0.000)	(0.000)	(0.003)
Female	-0.594***	-0.589***	-0.525***
	(0.005)	(0.004)	(0.015)
Marital status	0.032***	0.059***	0.162***
	(0.008)	(0.009)	(0.049)
Education level = 1, Medium Education	0.059***	0.061***	0.111***
	(0.011)	(0.010)	(0.038)
Education level = 2, High education	0.495***	0.566***	0.840***
	(0.029)	(0.028)	(0.090)
Constant	9.916***	10.049***	9.622***
	(0.182)	(0.317)	(0.462)
Observations	285,65	342,286	56,936
R-squared	0.172	0.202	0.190
Number of hhid	204,499	256,759	50,342
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual controls	YES	YES	YES
Family clusters	YES	YES	YES
Female	YES	YES	YES
Years in the USA	t < 10	10 < t < 30	t > 30

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable the logarithm of wage and salary income. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education. Years in the United States categories are less than 10 years/betweeen 10 and 30 years/more than 30 years.

Table 13: Fixed Effect Linear Model. Income and Wage on Irrealis Dummy, based on Continent of Birth

VARIABLES	(1) North America	(2) Central America	(3) South America	(4) Europe	(5) Asia	(6) Africa	(7) Australia, New Zealand, Pacific Islands, Antartica
Irrealis dummy	0.077	0.036*	-0.076	0.078	0.066	0.002	0.525
	(0.123)	(0.018)	(0.075)	(0.063)	(0.156)	(0.150)	(0.387)
Age	0.016***	0.009***	0.010***	0.013***	0.026***	0.019***	0.017
	(0.003)	(0.000)	(0.001)	(0.002)	(0.003)	(0.005)	(0.019)
Female	-0.431***	-0.589***	-0.554***	-0.600***	-0.557***	-0.544***	-0.594***
	(0.025)	(0.003)	(0.013)	(0.012)	(0.025)	(0.040)	(0.133)
Marital status	0.064	0.040***	0.078***	0.174***	0.002	0.074	-0.116
	(0.064)	(0.006)	(0.025)	(0.045)	(0.078)	(0.090)	(0.666)
Education level = 1, Medium Education	0.103	0.071***	0.099**	0.074	0.125	-0.154	-0.329
	(0.098)	(0.007)	(0.047)	(0.063)	(0.132)	(0.168)	(0.276)
Education level = 2, High education	1.094***	0.482***	0.582***	0.679***	0.786***	0.311*	1.379***
	(0.147)	(0.023)	(0.070)	(0.078)	(0.161)	(0.179)	(0.423)
Constant	9.369***	9.584***	9.747***	10.501***	9.245***	9.624***	9.987***
	(0.412)	(0.215)	(0.098)	(0.154)	(0.220)	(0.241)	-1042
Observations	17,025	430,052	39,692	53,45	24,644	9,159	988
R-squared	0.131	0.185	0.172	0.210	0.188	0.183	0.337
Number of hhid	14,154	277,932	29,535	40,769	20,647	7,569	877
Year fixed effect	YES	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES	YES
Female	YES	YES	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES	YES	YES
Continent of birth	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable the logarithm of wage and salary income. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Table 14: Fixed Effect Linear Model. Income and Wage on Irrealis Dummy, based on European Region of Provenance

VARIABLES	(1) Europe	(2) Northern Europe	(3) Western Europe	(4) Southern Europe	(5) Central Europe	(6) Eastern Europe
Irrealis dummy	0.078	0.525	1.475***	0.204	-0.026	0.183
	(0.063)	(0.387)	(0.275)	(0.161)	(0.125)	(0.143)
Age	0.013***	0.017	0.022	0.015***	0.012***	0.014***
	(0.002)	(0.019)	(0.025)	(0.005)	(0.003)	(0.002)
Female	-0.600***	-0.594***	-0.800***	-0.647***	-0.739***	-0.478***
	(0.012)	(0.133)	(0.174)	(0.027)	(0.027)	(0.016)
Marital status	0.174***	-0.116	-0.528	-0.005	0.250***	0.160**
	(0.045)	(0.666)	(0.535)	(0.104)	(0.087)	(0.067)
Education level = 1, Medium Education	0.074	-0.329	0.008	0.008	-0.154	0.362***
	(0.063)	(0.276)	(0.072)	(0.072)	(0.325)	(0.134)
Education level = 2, High education	0.679***	1.379***	1.009**	0.795***	0.337	0.724***
	(0.078)	(0.423)	(0.482)	(0.141)	(0.338)	(0.148)
Constant	10.501***	9.987***	9.376***	10.898***	9.915***	9.365***
	(0.154)	-1042	-1192	(0.656)	(0.609)	(0.300)
Observations	53,45	988	860	9,069	13,199	20,375
R-squared	0.210	0.337	0.249	0.270	0.252	0.168
Number of hhid	40,769	877	756	6,906	10,557	14,979
Year fixed effect	YES	YES	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES
Family clusters	YES	YES	YES	YES	YES	YES
Female	YES	YES	YES	YES	YES	YES
Years in the USA	YES	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES	YES
European region		YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable the logarithm of wage and salary income. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education.

Appendix C: Robustness Check Regression Tables

Table 15: Fixed Effect Linear Model. High Risk Occupation, Income and Wage and Entrepreneurship on Irrealis Dummy, Second-generation Immigrants

VARIABLES	(1) High risk occupation	(2) Income and wage	(3) Entrepreneurship
Irrealis dummy	0.013***	0.085***	0.006**
	(0.003)	(0.018)	(0.002)
Sex	-0.126***	-0.218***	-0.014***
	(0.002)	(0.011)	(0.001)
Age	0.023***	0.115***	0.002***
	(0.000)	(0.001)	(0.000)
Marital status	0.060***	0.238***	0.001
	(0.007)	(0.021)	(0.003)
Education level = 1, Medium Education	0.083***	0.056	-0.013
	(0.013)	(0.071)	(0.009)
Education level = 2, High education	0.142***	0.972***	-0.022**
	(0.015)	(0.074)	(0.010)
Mother income and wage = 1, Btw 1$ and 10000$	0.016***	-0.089***	-0.006***
	(0.003)	(0.018)	(0.002)
Mother income and wage = 2, Btw 10000$ and 25000$	0.019***	0.069***	-0.010***
	(0.002)	(0.014)	(0.002)
Mother income and wage = 3, Btw 25000$ and 48000$	0.025***	0.111***	-0.009***
	(0.003)	(0.016)	(0.002)
Mother income and wage = 4, Btw 48000$ and 85000$	0.022***	0.028	-0.007**
	(0.004)	(0.025)	(0.003)
Mother income and wage = 5, Over 85000$	0.018***	0.024	-0.003
	(0.005)	(0.041)	(0.005)
education_mom = 1	-0.006*	0.008	0.003
	(0.003)	(0.017)	(0.002)
education_mom = 2	-0.018***	-0.178***	0.007**
	(0.004)	(0.022)	(0.003)
education_pop = 1	-0.017***	-0.008	-0.005**
	(0.003)	(0.016)	(0.002)
education_pop = 2	-0.025***	-0.165***	-0.002
	(0.004)	(0.022)	(0.003)
Hazardous occupation dummy, father	0.029***	0.044***	0.002
	(0.002)	(0.012)	(0.002)
Hazardous occupation dummy, mother	0.014***	-0.010	-0.002
	(0.002)	(0.012)	(0.001)
Constant	-0.246***	6.063***	0.014
	(0.074)	(0.274)	(0.027)
Observations	160,302	55,387	62,603
R-squared	0.172	0.326	0.019
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual control	YES	YES	YES
Parental characteristics	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, No income, with father and mother with low education and employed in a highrisk occupation.

Table 16: Fixed Effect Linear Model. High Risk Occupation, Income and Wage and Entrepreneurship on on Categorical Irrealis (catIRR), Second-generation Immigrants

VARIABLES	(1) High risk occupation	(2) Income and wage	(3) Entrepreneurship
Irrealis moods, categorial = 1, Intermediate use	0.003	0.030	0.011*
	(0.007)	(0.041)	(0.006)
Irrealis moods, categorial = 2, Intensive use	0.014***	0.094***	0.005*
	(0.003)	(0.019)	(0.002)
Sex	-0.126***	-0.218***	-0.014***
	(0.002)	(0.011)	(0.001)
Age	0.023***	0.115***	0.002***
	(0.000)	(0.001)	(0.000)
Marital status	0.060***	0.238***	0.001
	(0.007)	(0.021)	(0.003)
Education level = 1, Medium Education	0.083***	0.056	-0.013
	(0.013)	(0.071)	(0.009)
Education level = 2, High education	0.143***	0.973***	-0.022**
	(0.015)	(0.074)	(0.010)
Mother income and wage = 1, Btw 1$ and 10000$	0.016***	-0.088***	-0.006***
	(0.003)	(0.018)	(0.002)
Mother income and wage = 2, Btw 10000$ and 25000$	0.019***	0.069***	-0.010***
	(0.002)	(0.014)	(0.002)
Mother income and wage = 3, Btw 25000$ and 48000$	0.025***	0.111***	-0.009***
	(0.003)	(0.017)	(0.002)
Mother income and wage = 4, Btw 48000$ and 85000$	0.022***	0.028	-0.007**
	(0.004)	(0.025)	(0.003)
Mother income and wage = 5, Over 85000$	0.018***	0.024	-0.003
	(0.005)	(0.041)	(0.005)
education_mom = 1	-0.006*	0.008	0.004
	(0.003)	(0.017)	(0.002)
education_mom = 2	-0.019***	-0.178***	0.007**
	(0.004)	(0.022)	(0.003)
education_pop = 1	-0.017***	-0.008	-0.005**
	(0.003)	(0.016)	(0.002)
education_pop = 2	-0.025***	-0.164***	-0.002
	(0.004)	(0.022)	(0.003)
Hazardous occupation dummy, father	0.029***	0.044***	0.002
	(0.002)	(0.012)	(0.002)
Hazardous occupation dummy, mother	0.014***	-0.010	-0.002
	(0.002)	(0.012)	(0.001)
Constant	-0.246***	6.062***	0.014
	(0.074)	(0.274)	(0.027)
Observations	160,302	55,387	62,603
R-squared	0.172	0.326	0.019
Year fixed effect	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES
Individual control	YES	YES	YES
Parental characteristics	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: catIRR = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, No income, with father and mother with low education and employed in a highrisk occupation.

Table 17: Fixed Effect Linear Model. High Risk Occupation on Irrealis Dummy, based on Age Groups, Second-generation Immigrants

VARIABLES	(1) <16 years	(2) 16-19 years	(3) 19-24 years	(4) > 24 years
Irrealis dummy	-0.001	0.004	0.042***	0.018*
	(0.002)	(0.005)	(0.010)	(0.010)
Sex	-0.016***	-0.100***	-0.301***	-0.266***
	(0.001)	(0.003)	(0.006)	(0.007)
Age	0.037***	0.080***	0.022***	0.001
	(0.001)	(0.002)	(0.002)	(0.001)
Marital status	-0.001	0.085***	0.043***	0.035***
	(0.008)	(0.018)	(0.013)	(0.010)
Education level = 1, Medium Education	-0.051*	-0.030	0.053**	0.117***
	(0.029)	(0.029)	(0.024)	(0.018)
Education level = 2, High education		-0.235***	0.159***	0.229***
		(0.029)	(0.032)	(0.020)
Mother income and wage = 1, Btw 1$ and 10000$	0.007***	0.019***	0.004	-0.002
	(0.002)	(0.005)	(0.010)	(0.013)
Mother income and wage = 2, Btw 10000$ and 25000$	0.004***	0.018***	0.009	0.000
	(0.001)	(0.004)	(0.008)	(0.010)
Mother income and wage = 3, Btw 25000$ and 48000$	0.004**	0.014***	0.019**	0.012
	(0.002)	(0.004)	(0.009)	(0.011)
Mother income and wage = 4, Btw 48000$ and 85000$	0.009***	0.010*	0.010	-0.019
	(0.003)	(0.006)	(0.015)	(0.018)
Mother income and wage = 5, Over 85000$	-0.001	0.009	-0.005	0.014
	(0.003)	(0.008)	(0.023)	(0.026)
education_mom = 1	-0.003	0.001	-0.002	-0.000
	(0.002)	(0.005)	(0.009)	(0.010)
education_mom = 2	-0.004	-0.004	-0.033***	-0.013
	(0.003)	(0.006)	(0.012)	(0.014)
education_pop = 1	-0.003	-0.018***	-0.018**	-0.007
	(0.002)	(0.005)	(0.009)	(0.010)
education_pop = 2	-0.003	-0.019***	-0.028**	-0.030**
	(0.003)	(0.006)	(0.012)	(0.013)
Hazardous occupation dummy, father	0.003***	0.018***	0.046***	0.058***
	(0.001)	(0.003)	(0.007)	(0.008)
Hazardous occupation dummy, mother	0.001	0.014***	0.030***	0.027***
	(0.001)	(0.003)	(0.007)	(0.009)
Constant	-0.521***	-1.131***	-0.008	0.894***
	(0.034)	(0.144)	(1,046.701)	(0.071)
Observations	60,801	54,846	24,064	20,591
R-squared	0.029	0.077	0.127	0.089
Year fixed effect	YES	YES	YES	YES
Country-of-birth fixed effect	YES	YES	YES	YES
Individual control	YES	YES	YES	YES
Parental characteristics	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variables are defined in the header. The method of estimation is a fixed effect linear model. Robust standard errors in parentheses. Reference categories for dichotomous: irrealis = 0 (no irrealis moods), Male, Not Married (divorced, separated, widowed), Low education, No income, with father and mother with low education and employed in a highrisk occupation

References

Anderlini L., Felli L. (2004). "Book review: economics and language: five essays by Ariel Rubinstein". *Cambridge University Press, Cambridge, 2000. Economica, 71 (281). pp. 169-171*

Baltagi B. H. (2005). "Econometric Analysis of Panel Data". *Third edition, John Wiley & Sons Ltd, London*

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.

Bell A. (1978). "Language samples". *Universals of human languages, Volume 1: Method – theory, 123 – 156. Stanford: Stanford University Press*

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.

Brown R., Lenneberg E. (1954). "A study in language and cognition". *The Journal of Abnormal and Social Psychology, Vol 49(3)*

Brown, R. (1976). "In Memorial Tribute to Eric Lenneberg". *Cognition*

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

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

Chomsky N. (1965). "Aspects of the Theory of Syntax". *MIT Press*

Connecticut Department of Labor - List of Occupations Designated As High-Risk Or Safety-Sensitive by the Labor Commissioner of the State of Connecticut (03/10) - www.ctdol.state.ct.us/wgwkstnd/highrisk.htm

Dahl Ö. (1985). "Tense and Aspect Systems". *Oxford: Basil Blackwell.*

Dahl Ö. (2000). "The Grammar of Future Time Reference in European Languages". *Berlin: Mouton de Gruyter.*

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

Dohmen T. et al. (2006). "The Intergenerational Transmission of Risk and Trust Attitudes". *IZA Discussion Paper No. 2380*

Dohmen, T., Falk, A., Human, 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.

Fernández R. (2008). "Culture and Economics". In *The New Palgrave Dictionary of Economics. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008. The New Palgrave Dictionary of Economics Online.* http://www.dictionaryofeconomics.com/article?id=pde2008_E000282

Fernández R. (2010). "Does Culture Matter?". *NBER Working Paper No. 16277*

Galor O. et al. (2016) "Geographical Origins and Economic Consequences of Language Structures". *CESifo Working Paper Series No. 6149*

Gay V. et al. (2016) "Migration as a Window into the Coevolution between Language and Behavior". *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11), Forthcoming.*

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

Gino F., Sharek Z., Moore D.A. (2011). "Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration". *Organizational Behavior and Human Decision Processes. 114 (2): 104–114*

Gorman S. (2013). "How do we perceive risk?: Paul Slovic's landmark analysis". *The Pump Handle, a water cooler for public health crowd. Science Blog* http://scienceblogs.com/thepumphandle/2013/01/16/how-do-we-perceive-risk-paul-slovics-landmark-analysis-2/

Guiso L., Sapienza P., Zingales L. (2006). "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives, 20(2): 23-48.*

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

Gumperz, J. J., and Levinson, S. C. (1996). "Rethinking linguistic relativity". *Cambridge, UK: Cambridge University Press.*

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.

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

Kahneman D., Tversky A. (1982). “Judgment under uncertainty: Heuristics and biases.” *(D. Kahneman, P. Slovic, & A. Tversky, Eds.). Cambridge, UK: Cambridge University Press.*

Kovacic M. et al. (2016). “Risk attitudes, investment behavior and linguistic variation”. *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 34/15.*

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

Lucy J. (1997). "Linguistic Relativity", *Annual Review of Anthropology*

Majid A., Bowerman M., Kita S., Haun D. B. M., Levinson S. C. (2004). “Can language restructure cognition? The case for space”. *Trends in Cognitive Sciences*

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

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., Braisby N. (1999). “Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks”. *Cognition*

Roberts S., Winters J., Chen K. (2015). “Future Tense and Economic Decisions: Controlling for Cultural Evolution”. PLoS ONE 10(7)

Rothstein B., 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)*

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

Sapir, E. (1929). "The status of linguistics as a science". *Language 5. 207-14. Reprinted in The selected writings of Edward Sapir in language, culture, and personality, ed. by D. G. Mandelbaum, 160-6. Berkeley: University of California Press.*

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

Simon M., Houghton S.M., Aquino K. (2000). “Cognitive biases, risk perception, and venture formation: How individuals decide to start companies”. *Journal of Business Venturing, 15(2), 113–134.*

Simon M., Shrader R.C. (2012). “Entrepreneurial actions and optimistic overconfidence: The role of motivated reasoning in new product introductions”. *Journal of Business Venturing, 27(3), 291–309.*

Slovic, P. (1987). “Perception of Risk”. *Science 236 (17 April), pp. 280-285*

Sutter M. (2015). “The Effect of Language on Economic Behavior: Experimental Evidence from Children’s Intertemporal Choices”. *IZA Discussion Paper No. 9383*

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

Winawer J., Witthoft N., Frank M. C., Wu L., Wade A. R., Boroditsky L. (2007). “Russian blues reveal e ects of language on color discrimination”. *Proceedings of the National Academy of Sciences of the United States of America*

Zhang S.X., Cueto J. (2017). “The Study of Bias in Entrepreneurship”. *Entrepreneurship Theory and Practice 41.3, 419-454*