Language and Gender Economics

Work in progress

Victor Gay [†]
Estefania Santacreu-Vasut [†]
Amir Shoham [‡]

Abstract

This paper studies the relation of gender marking in the language spoken at home and in countries dominant language with gender biases in labor market, political and managerial outcomes both at individual and country level, for general population and for populations of migrants. We test the hypothesis that the intensity with which languages grammar force speakers to encode gender may increase gender discrimination and/or the enforcement of gender roles and gender roles themselves. We find that gender marking has a significant, meaningful and very robust impact on gender economic outcomes both at the country and individual level. Cognitive and cultural mechanisms may explain our findings: gender marking in language may influence the formation of gender identities, the salience of deviant behavior from gender roles and gender roles themselves. We explore the relative role of countries dominant language and home language and extend the model of identity economics as developed by Akerlof and Kranton (2000) to further investigate these mechanisms.

Keywords: Culture, Credit, Gender, Identity, Labor, Language, Politics

[†] The University of Chicago
[†] Corresponding Author. Department of Economics at ESSEC Business School, THEMA, France. Avenue Bernard Hirsch, 95021, Cergy-Pontoise, France, +33 6 82 56 36 27, santacreuvasut@essec.edu
[‡] The FOX School of Business Temple University and The College of Management Academic Studies Israel. Temple University, 1801 Liacouras Walk Philadelphia, PA 19122-6083, USA and COMAS, 7 Yitzhak Rabin Blvd., RishonLeZion, 75190, Israel. amir1s@colman.ac.il

We are grateful to the participants of seminars at UC Berkeley, Ben Gurion University of the Negev, PSE, College of Management Academic studies Business School and THEMA seminars, Temple University Brown Bag Seminar, and to Pranab Bardhan, Barry Eichengreen, Ramon Ferrer-i-Cancho, Johanna Nichols and Gerard Roland. We thank Anatoli Colicev for excellent research assistance.

This paper investigates the idea that the structure of languages may influence the cognitive framework of its speakers, as grammar forces them to encode certain aspects of reality and not others and/or may contain information related to the culture of our ancestors. As Whorf (1935) argues, "We are inclined to think of language simply as a technique of expression, and not to realize that language first of all is a classification and arrangement of the stream of sensory experience which results in a certain world-order, a certain segment of the world that is easily expressible by the type of symbolic means that language employs". Among linguists, particularly among the so-called 'generativists', languages are considered part of human biology and fundamentally equal in structure, with only minor local differences (Chomsky, 1980). Sampson *et al.*, (2009) point out a new wave of linguists who think of languages as 'institutions developed as part of a society's cultural heritage and hence as differing and evolving in their levels of complexity, just as other cultural institutions do'.

Overall, current theories of coevolution point to the interplay between genetic and cultural forces to account for language diversity and change. Findings from cognitive psychology about the impact of language on cognition (Boroditsky and Gaby, 2010) indicate that there may be direct channel through which language structure influence socio-economic choices and outcomes. The study by Evans (2003) of how kinship-specific pronouns evolved in Australia is an example of a study focused on how cultural forces may shape linguistic change. As Evans and Levinson (2009) point out, 'language diversification and hybridization works just like the evolution of biological species it is a historical process, following the laws of population biology' and 'linguistic diversity is structured very largely in phylo-genetic (cultural-historical) and geographical patterns'. Christiansen and Kirby (2003) review the research on the origins and evolution of human language and argue that while one theoretical line argues that grammatical structure is the product of biological adaptation, others argue that it emerges through cultural transmission of language to hundreds (or perhaps thousands) of generations of learners.

In this paper we investigate the relation between gender marking in languages grammar and gender economics, an area of research where economists argue that culture plays an important

role (Fernandez, 2010; Alesina *et al.*,2011).

To date, the use of linguistic variables has been limited. Licht (2007) uses the grammar of pronouns as an instrumental variable to study how countries tilted more in favor of autonomy, egalitarianism, and mastery exhibit a higher rule of law, less corruption, and more democratic accountability. Mavisakalyan(2011) uses the gender of pronouns to investigate the impact of culture on females share in the labour force. Chen (2012) uses future time references marking in languages to investigate the impact future oriented decisions and outcomes like saving, debt and health related behavior. Tabellini (2008) seeks to capture the distinction between values consistent with generalized morality versus those consistent with limited morality. Altogether, norms of generalized morality promote well-functioning institutions. To control for the possibility of reverse causality, and identify the causal impact of these values on institutional outcomes, Tabellini uses the grammar of pronouns as an instrumental variable.[1] Finally, in a related paper, Santacreu-Vasut et al.(2013), we show that female/male distinctions in language grammar are positively correlated to the use of legislated quotas and sanctions today.

We survey the most comprehensive data source of language grammatical structures, an ongoing effort of linguists worldwide, called the World Atlas of Linguistic Structures (Dryer and Haspelmath, 2011) and use all four grammatical variables that are related to gender. In linguistics, a gender system is defined as a set of rules for agreements that depend on nouns of different types. These can be based on biological sex (female and male), or on other social constructs (like, age, social status, etc.) As we show in the main text there is variation in grammatical gender properties of languages both across and within linguistic families. Furthermore, gender is the most stable grammatical features of language, inherited from thousands of years distant past (Wichmann and Holman, 2009)

We consistently find that language structure is significantly, meaningfully and robustly correlated to gender biases both at the country and individual level, across general and migrant populations in the labor market, political participation and managerial roles suggesting that in-

[1]More generally, Guiso et al (2009) document that trust is affected by geographical distance between two countries, their proximity, and the commonality between their languages.

deed gender marking either shapes the construction of gender identities and may also reflect distant past culturally inherited gender roles.

1 Gender in Languages

Greville G. Corbett argues that gender " is the most puzzling of the grammatical categories". While in common parlance gender is linked to biological sex, and refers to female and male categories, in linguistics a gender system is a set of rules for agreements that depend on nouns of different types.[2] As Greville G. Corbett mentions in Chapter 30 of WALS 'In some languages gender is evident in almost every phrase, while in other languages it is absent.' (Dryer and Haspelmath, 2011). This is the variation we exploit in this paper.

There is general consensus among linguists that grammatical gender is one of the most stable feature of a language grammar, inherited from the distant past thousand of years ago, as research on the study of linguistic stability indeed confirms. In particular, Wichmann and Holman (2009) constructed a measure of stability for analyzing the linguistic features described in the World Atlas of Linguistic Structures, (Dryer and Haspelmath, 2011).[3] They define stability as, the probability that a given language remains unchanged with respect to the feature during 1000 years, that is, the feature undergoes neither internal change nor diffusion during the interval(Wichmann and Holman, 2009).

In this paper, we investigate the impact of gender marking in both the dominant language of countries and the language individuals speak at home. To do so, we exploit information in all grammatical features directly related to expressions of gender (4 out of a total 192 grammatical features classified in WALS (ibid.)). These correspond to chapter 30 (Corbett, 2011a), chapter 31 (Corbett, 2011b), chapter 32 (Corbett, 2011c) and chapter 44 (Siewierska, 2011) of WALS .

Out of these, we create 12 dummy variables, 4 individual indices and 3 aggregate indeces as detailed next for both the language spoken at home by an individual and the most spoken

[2]To avoid confusion, we will refer to 'gender' as in common parlance and to 'grammatical gender' when we refer to the broader concept of gender in linguistics.

[3]They compare their findings with categorical statements in the literature regarding the stability of the features they analyze, and there is a high degree of concordance.

language of a country.

1.1 Gender Grammatical Features and Variables

The defining characteristic of gender is agreement. A gender system is a set of rules for agreements that depend on nouns of different types. These types can be sex-based (female and male), or based on other social constructs (such as the distinction between human and animal, age, social status...)

Number Genders (NG): This variable captures how many genders are present in the language. That is, how many different types of nouns have different agreements.[4] In particular, we built three dummy variables, NG0, NG2, and NG3, corresponding to languages having 0, 2, or 3 (and more) genders, respectively.[5] A language with 2 genders, like French and Spanish, typically implies feminine versus masculine while a language with three or more genders may include neuter as the third gender, like German, or no sex related distinctions.

Sex Based (SB): This variable captures whether the gender system is linked to biological sex. We built three dummy variables, SB0, SBY, and SBN, corresponding to languages having no gender system, a sex-based gender system, or a non-sex-based gender system, respectively.[6] For example, a non-sex-based gender system might be based on the distinction between human and non-human, as in Fulfulde, a member of the Niger-Congo linguistic family, or the distinction between animate and inanimate, among others. Corbett, author of the feature about this variable in WALS, argues, for those interested in language and gender in the sociological sense, these direct reflections of biological sex in many languages, with the control group of languages with similar gender structures but without the sex component, should provide a valuable source of data. (Corbett, 2011).

Gender Assignment (GA): This variable captures how a speaker assigns nouns to the genders defined by the gender system of the language. A gender assignment system provides a set of rules to help the speaker make appropriate agreements. Assignment can depend on the meaning

[4] Gender derives from Latin genus and originally it meant kind or sort.
[5] There are languages, such as Nigerian Fula that feature 20 genders.
[6] Y and N stand for Yes and No.

(semantic) or the form of the noun. We built three dummy variables, GA0, GAL, and GAH, corresponding to languages having no gender assignment system, a semantic gender assignment or both gender assignment system that is both semantic and formal, respectively.[7] For example, a semantic assignment system is found in Kannada (a south India language), where nouns denoting male (female) humans are masculine (feminine) and all remaining nouns are neuter. In a semantic and formal assignment system, as used in Russian, in addition to semantic gender assignments, nouns that are neither masculine nor feminine are not necessarily neuter; rather they can be assigned to the masculine or feminine gender, depending on their inflectional class (whether it takes the nominative form or accusative form, etc.).

Gender Pronouns (GP): This variable captures Gender Distinctions in Independent Personal Pronouns. We build three dummy variables GP0, GPL, and GPH, corresponding to languages with no gender distinction in pronouns, gender distinction in third-person pronouns only, and gender distinction in the third-person, and also in the first and/or the second person, respectively. For example, in English, the pronominal sex-based gender system is determined by the use of she/he/it. Some languages do not have sex-based pronouns, but still have a sex-based gender system because they use 'it' but vary forms, poet/ess (i.e., they indicate the sex by changing the morphology uniquely.)

Out of the above set of dummy variables we build four individual gender intensity indices as follows:

Our first intensity index is a dummy variable, NGI, equals one for languages having two genders and equals zero otherwise (no gender or three or more genders). Our rationale is that a language with two genders, typically feminine and masculine will more pervasively use female/male distinctions than one with either no gender or three genders (which include neutral).

SBI equals one for languages having a sex-based gender system and zero otherwise. For example, Zulu, Swedish and Danish are languages with a gender system that is not sex based.

GAI equals one for languages having both semantic and formal gender assignment system and zero otherwise. For example, English assigns gender based on semantic grounds only while

[7] 'L' and 'H' stand for 'Low' and 'High'.

Spanish uses semantic and formal assignment rules. Our rationale is that a language that assigns gender on more than semantic grounds will more pervasively use female/male distinctions.

GPI equals one for languages with gender distinction in third-person pronouns and in the first and/or the second person, and equals zero otherwise. For example, English distinguishes gender in third-person pronouns only.

Finally, we construct an aggregate index for a given language as the sum of the individual indices. If the language is the dominant language spoke in a country our indices are:

$GII = NGI + SBI + GAI + GPI$ where $GII\epsilon\{0, 1, 2, 3, 4\}$. We use **?** to gather information on the distribution of speakers within countries.

For example, the GII for German is equal to 2. It has a sex-based gender system, SBI=1, and assigns gender on the basis of both semantic and formal rules, GAI=1. However, GPI=0 since German assigns gender to third person pronoun only and NGI=0 since German does have a neuter gender.

We have a sample of 84 countries for which information on the four linguistic variables is available and a total of 128 countries for which at least one of our variable is available.

We construct two additional aggregate subindices, GII_{v1} and GII_{v2}, as follows.

$GII_{v1} = NGI + SBI + GAI$

and

GII_{v2} =NGI+SBI+GPI

GII_{v1} is built as a robustness check, since the information regarding its individual components was gathered by the same researcher, Corbett.

GII_{v2} is built to overcome sample size limitations of our GII index since there is a relatively high fraction of countries for which we lack information on GAI grammatical variable.

We build similar indeces for languages spoken at home, which we call HGII, HGIIv1 and HGIIv2, where H stands for home. For 94 of the languages spoken at home by individuals in the World Value Survey we have information on at least one of our variables.

Linguistic structures vary widely across language families as well as within families. For

instance, tables 24-29 in the appendix show variations both across and within families. This is consistent with the fact that language evolve following three mechanisms: 'One is that it is the result of contact between languages. The second possibility is that it reflects a genealogical relationship among at least some of the languages, involving a feature inherited from a common ancestor.The third possible explanation for shared features within a particular geographical area is that it is at least partly coincidence.' as described in WALS, Dryer and Haspelmath (2011).

To control for the influence of geography and historical contact across societies and their languages, we perform two systematic robustness checks across our empirical analysis: (1) we control for continents, (2) we control for region as defined by UNESCO, since it captures cultural-geographical grouping of countries (Arab States, Central Eastern Europe, Central Asia East Asia and the Pacific, Latin America and the Caribbeans, South and West Asia, Sub Saharian Africa, North America and Western Europe).

2 Results

2.1 The Labor Market

In this section we study the relation between language gender marking and female labor force participation and occupation choices. We do so in three different settings, using three distinct datasets and levels of aggregation. (1) Countries dominant language gender marking correlation with the country's female labor force participation rate, share and occupational choice. (2) Individuals home language gender marking impact correlation with their labor force participation and occupational choice, using the three latest waves of the World-Value Survey (WVS) data. (3) US immigrants country of origin and mother tongue language gender marking correlation with their labor market outcomes in the US, using the New Immigration Survey database.

2.1.1 Gender in Countries Dominant Language and Labor Force

We start discussing OLS regressions to investigate the correlation between a country's dominant language gender marking on participation rates of females in the labour force and for females;

occupational choices (relative to males' choices) in services, industry and agriculture. We present results for the year 2000 but similar results are obtained using other years. We use our set of 12 gender individual dummies but similar results are obtained with the agregate indices. In all cases, our excluded dummy variable is the one corresponding to no gender marking: NG0, SB0, GA0, GP0. Therefore, we must interpret our coefficients relative to these variables.

Our set of controls includes the share of inhabitants for whom the dominant language is their mother tongue. The aim is to capture the fact that the dominant language may be less significative of a country's culture where there is less linguistic homogeneity. As research about ethno-linguistic fractionalization has shown, linguistic diversity plays an important role in economic outcomes (Mauro, 1995; Easterly and Levine, 1997). We follow Goldin (1995), and control for economic development (measured by the logarithm of GDP per capita and its square term). We control for oil production following Ross (2008), who suggests that oil, not religion, explains why women lag behind in many Middle Eastern countries. We also control for openness. Single industry studies show that the share of female employment is higher in firms that export (Baslevent and Onaran, 2004). Finally, we control for political factors such as government size, polity score on democracy, a communist past, and population size.

The structure of dominant languages varies across countries and across linguistic families. To control for the role of geography and historical influences on the distribution of language structures across countries, we performed three types of robustness checks: geography, colonization, and religion.

For geography, we controlled distance from the equator, which may reflect Western influence as Hall and Jones (1999) argue. Following Bloom and Sachs (1998), and Gallup *et al.*(1998) in an alternative specification, we control for climate because it may influence development. Because a colonial past influences institutions and human capital (Acemoglu *et al.*, 2001), which influence development, we controlled for colonization and for its origin (English, French or Spanish), since the origin of the colonizer has been shown to impact current institutions (La Porta *et al.*, 1998, 1999). Our results regarding the origin of the colonizer are robust, and are available in the online

appendix.

Table 1 presents the results for the rate of females' participation in the labour force. Specification (1) is the baseline regression with our set of control variables only. Specifications (2)-(5) include each of our language variable dummies. Because of the correlation between our gender intensive variables, we did not include all of the dummies in the each specification. Specifications (6)-(17) include three robustness checks, two controlling for geography, and one for colonization.

Countries having a dominant language with Number of Genders equal to 2 have 17 % points lower female participation rate than countries with Number of Genders equal to 0. Note that having a Number of Genders equal to 3 or more has a smaller negative effect (approximately 5% points), but it is not significant. This is consistent with the fact that in languages with three or more genders the female/male distinction is less pervasive. Having a Sex-based Gender System decreases females' labour force participation rate by 12% points compared to baseline labour force participation in countries with no gender system. Again, this result is very significant both statistically and economically. Conversely, in a country with a Non-sex-based Gender System, females' labour force participation rate is approximately 3% points higher than their labour force participation rate than one with no gender system, but this result is not significant. Regarding Gender Assignment and Gender Pronouns, our results are intriguing because these variables include an intensity dimension. Having gender assignment that is both semantic and formal, or gender marking in a larger set of pronouns decreases the participation rate relative to not having one, more than having semantic gender assignment only or a smaller set of gender marking in pronouns. Our results are similar for another measure of females' participation in the labour market that controls for overall labor force size: the share of the labour force that is female. Further note that the Adjusted R-Squared almost doubles when we include gender language variables.

Are labour markets across countries segregated by gender? Table 2 shows the results for the distribution of females, relative to males, in service, industrial and agricultural occupations. Our results suggest that indeed they are. There is a positive female employment bias in services, and

a negative female employment bias in agriculture in countries where the dominant language is more gender intensive. Results on services are very significant. The gender intensive dummies are all significant both statistically and economically. Having a language with Number of Genders equals two, as opposed to none increases, ceteris paribus, the relative share of females employed in services by a magnitude similar to having a sex-based gender system relative to having no gender system at all. Regarding gender pronouns, our results are even stronger in magnitude.

Further it is intriguing that among gender intensive dummies those that mark gender more have a stronger impact on female employment bias in services. In the opposite direction, having a non-sex-based gender system is negatively, but not significantly, associated with females share in services. Regarding agriculture, the share of female employment in agriculture relative to male employment is reduced in countries with gender intensive languages. For instance, countries with high gender pronoun marking have a lower female negative bias by almost one standard deviation (-0.37) compared to countries with no gender pronoun marking, a very significant result.

To sum up, we find both statistically and economically significant correlations that suggest that countries whose dominant language has gender intensive features have, ceteris paribus, lower participation rate of females in the labour force, a lower share of females in the labour force, exhibit a negative bias of female employment in agriculture and a positive bias of female employment in services.

2.1.2 Gender in Language Spoken at Home and Individual Labor Market Choices

We next discuss regressions run on individuals in the three last waves of the WVS, from 1994 to 2007 for whom we have information on the language spoken at home.[8] This allow us to study individuals across 76 countries for which language data is available.

We present odd ratios of logit regressions of (1) female labor force participation and (2) whether they work in agriculture regressions on (1) the gender intensity of the language spoken at home using our three agregate indeces, HGII, HGIIv1 and HGIIv2 and (2) cultural beliefs or values regarding the role of women in society. In particular, whether in a situation of job scarcity

[8]Please refer to z-statistics provided in the table, significant levels are not marked with * system in these tables

men should have more right to a job than women, which we call MenSuperior and whether a woman has to have children to be fulfilled which we call ChildrenValue. We perform the same regressions for the sample of men as robustness check.

In all of our regressions we control for country, income decile within country, marital status, education, age, number of children, survey wave and religion and perform robustness checks controlling for continents or regions (see appendix for detailed list of variables as described in the World Value Survey).

Several results stand out. As the odd ratios presented in Table 4 are lower than one for women and higher than one for men, this suggest that gender intensity of home language decreases the relative labor force participation of women while increase men's one. The correlation of home language gender marking is significant despite the fact that we control for cultural beliefs and values. Further, women speaking a gender intensive language at home are less likely to be employed in agricultural occupation than those that do not. Results for men suggest either that they are more likely to do so or are not significant suggesting occupation related gender roles are mostly relevant for women. Note how women that strongly disagree with the statement that men should be given priority in times of job scarcity are more likely to work than those that do not, while as Table 5 shows this variable is not significantly related to male labor force decision.

In tables 6 and 7 we run logit regressions where the dependent variable is equal to 1 if the individual works in an agriculture related occupation. Results mirror findings at the country level as reported in Table 3. That is, women speaking highly gendered language at home are less likely, ceteris paribus, to work in occupations related to agriculture, as odd ratios significantly lower than one suggest. This results also relates to Alesina *et al.* (2010) who show that indeed the historical use of the plough correlates negatively with gender cultural norms towards traditional female occupations and female labour force participation today.

2.1.3 Languages travel: Study of Migrants in the USA

In this section we use the publicly available dataset version of The New Immigrant Survey, a project that spans four institutions - RAND, Princeton University, New York University, and Yale University. This allows us to study the impact of migrants language gender structure (both of their country of origin and maternal language) on their labor supply and earnings in the United States. [9]

We investigate three different labor market outcomes: the labor supply (us_work_hours) in the US, as reported in Table 8 and 9, the labor market income (us_pay) and participation in the labor force (results to be included soon). We find that migrants whose country of origin is more gender intensive are more likely to supply labor, to work more hours and earn lower labor market incomes than those from countries whose language is less gender intensive. This result is robust regardless of the gender grammatical variable we use.

Our results are not driven by migrants coming from spanish-speaking countries to work in the US as they do not change when we run our regressions excluding them. We also performed robustness checks by running our regressions on a sample of men. Further, we obtained similar results using individual's maternal language instead of migrants country of origin dominant language (see table 12).

Why does gender marking lead to lower labor force participation among women at the country and individual level for general population but the reverse across migrants living in the US?

To understand the potential mechanisms behind this finding, we use the World Value Survey since it allow us to compare the labor market behavior of individuals, and women in particular, living in low gendered countries only but speaking languages with different gender intensity at home. We focus on this subset of individuals because the United States dominant language, english, is a low gendered language. We analyze the impact of their home language on their labor supply behavior and find that results that mirror the findings from the NIS dataset.

[9] It is a "nationally representative multi-cohort longitudinal study of new legal immigrants and their children to the United States based on nationally representative samples of the administrative records, compiled by the U.S. Immigration and Naturalization Service (INS), pertaining to immigrants newly admitted to permanent residence"

Table 16 shows the impact of home language gender intensity, HGII, on labor force participation for women whose country of residence dominant language is low gendered ($GII < 2$) such as the US.[10] We perform similar analysis for individuals in low gendered countries according to two alternative measures: $GIIv1 < 1$ and $GIIv2 < 1$. These allow us to analyze the impact of gender marking in the language spoken at home for a sample of up to 17478 individuals in 29 countries whose dominant language is 'low gendered'. When we run the same set of regressions for men as control group, odd ratios for our gender marking variables are either not significant or have the opposite direction than for women. That is, women speaking highly gendered language at home living in a low gender environment are more likely to work. This suggests that what matters is the interaction between the cognitive framework of speakers and the cultural environment they live in. Gender marking in languages may influences the social constraints related to gender roles that women face as much or more than their cognitive framework.

2.1.4 The Relative Influence of Dominant versus Home Language

With a similar spirit than the analysis just presented, Table 17 shows the impact of home language gender intensity, HGII, on labor force participation for women whose country of residence dominant language is high gendered. In that case, odd ratios are very significantly smaller than one, suggesting that conditional on living in highly gender environments women that speak highly gender languages at home are less likely to work.

Table 18 and 19 present the reverse analysis. That is, fixing the language spoken at home, we compare the behavior of women in countries with different degree of gender marking in their dominant language. Among women whose home language is highly gendered ($HGII > 2$) countries dominant language gender intensity, GII, is very strongly negatively related to their labor force participation (odd ratios lower than one). Interestingly, Table 19 shows that for women whose home language is not gendered, the countries dominant language does not have a significant influence ($HGII < 2$).

[10] The exact threshold does not alter the results.

2.2 Management

In this section we study the role of women in managerial positions.[11] To do so we analyze if, conditional on being a manager, a women is less likely to manage big teams as opposed to small ones when speaking a language with high gender marking. The world value survey ask respondents if they manage a team with more or less than 10 employees.

As odd ratio lower than one in Table 13 show, gender intensity of home language decreases the relative frequency at which women managers manage big teams. The correlation of home language gender marking is significant despite the fact that we control for whether respondents believe that men make better business executives than women do, a variable we call MenSup Exec. Women that strongly disagree with this statement are, on the other hand, more likely to manage big teams.

2.3 Politics

Does gender marking in languages influences the participation of women in politics? Because quotas are one of the main determinants of increased female political participation as Jutting et al. (2006) points out, we study (1) female political participation, measured as percentage of female in parliament in the period 1972-1989, to explore the impact of dominant language gender marking before quotas were ever implemented.[12]. (2) the adoption of legislated gender political quotas, from 1971 to 2011. We determine what countries would be the first to adopt political quotas for woman using a Cox Hazard Model.

While international organizations have being pushing for quota adoption and the World Bank(2012) report recommends the use of quotas to increase women's access to political institutions, such as parliament, their application is far from widespread.

Table 14 presents coefficients from logit regressions where the dependent variable is a dummy variable equal one if a country has higher than average female political presence in parliament during the period 1972-1989. We control for electoral system, women access to run, communism

[11] Please refer to z-statistics provided in the table, significant levels are not marked with * system.
[12] The first country to implement a quota was Nepal in 1990

past, HDI, education measures such as primary or secondary enrollment, following Kenworthy and Malami (1999), Dahlerup (2003) , Paxton et al.(2010), and Tripp and Kang (2008). Gender intensity, measured using either of our three agregate measures is significantly negatively correlated with female political presence.

Table 15 presents Cox's proportional Hazard Model for the period 1971 to 2011. Female/male intensity, as measured by any of our indexes increases the probability of a country to be an early adopter of quotas.

To address the possibility that quotas are a not "window dressing" policy and to assess the effort of countries in implementing quotas we build a new dependent variable, a dummy that equals 1 if the country has legislated quotas with sanctions and 0 otherwise. The direction and significance of results is preserved when using this alternative variable.

The results of this section together are consistent with the fact that countries where female political participation is lowest are the most likely and earlier adopters of quota policies to meet the international demand for female political empowerment, such as those coming from international organizations have being pushing for quota adoption. In Santacreu et al. (2013) we show that countries that experiment the highest percentage increase in women presence in parliament pre and post quota adoption are those with highest gender intensity in their dominant language. As our results show this is not surprising since these are the countries with lower initial levels of female political participation, as evidence prior quota adoption suggests, and on the other hand, these are the countries most likely to use sanctions to enforce their quota policies as our duration model indicates.

3 The 'Linguistic-Gender Economics' Hypothesis

In this section we present the model of economics of identity from Akerlof and Kranton (2000) which provides the basic framework through which we formalize three potential mechanisms behind the findings of this paper, the 'Linguistic-Gender Economics' hypothesis.

In short, the hypothesis states that the intensity of gender marking in languages grammar

increases the salience of gender in speakers identity formation, in the salience of deviations from gender roles behavior and in the formation of the gender roles themselves. From a cognitive perspective, gender marking requires speakers to code information about gender more frequently and therefore increases its salience in speakers mental representation of themselves and of the world (reference here). From a cultural perspective, gender marking may reflect distant past culture gender roles inherited till present. We explore the consequences these may have for the choices and constrains faced by women in their economic and political life.

The model analyzes the choice between two possible activities, One and Two and expresses ideas from psychodynamic theory of personality.[13] If a person has a taste for Activity One (Two) and undertakes Activity One (Two) she earns utility V. She earns zero utility if her choice does not match her taste.

Preferences are identity-based. Suppose there are two social categories. The prescription or ideal for each of these categories is for each to choose a given activity, regardless of their tastes.

Figure 1 from their paper shows the game tree of interaction between two persons from the same social category (Person One and Person Two) with the prescription of choosing Activity One, but with different tastes (one with a taste for One and one with a taste for Two).

When a person deviates from the prescription corresponding to its social category, she suffers a loss in identity that leads to a reduction of utility equal to I_s where s stands for $self$. Further, there are identity externalities and the deviant behavior decreases the utility of the person that complies by I_o where o stands for other. In that case, the compliant person may respond to restore its identity (for example to preserve its self-image) at cost c and entail a cost equal to L to the deviant person.[14]

[13] We follow the original model notation and terminology

[14] For completeness we quote the 4 subgame perfect equilibrium as in Akerlof and Kranton(2000)

(i) Person One deters Person Two from engaging in Activity Two, when $c < I_o$ and $I_s < V < I_s + L$ (ii) Person One responds but does not deter Person Two from engaging in Activity Two, when $c < Io$ and $I_s + L < V$. (iii) Person One does not respond, and Person Two engages in Activity Two, when $c > I_o$ and $I_s > V$. (iv) Person Two does not engage in Activity Two regardless of Person Ones response, when $Is > V$.

3.0.1 Language in the Model

We next consider each of the three mechanisms to explore how gender marking in languages grammar may modify the model, influencing the choice of activity and leading to testable implications.

Language may influence identity formation, increasing identity loss in case of deviant behavior and therefore decreasing the likelihood of doing so. Further, it may influence the salience (information) of departures from gender related prescriptions. In the baseline model there is perfect information but we could extend the model to allow for imperfect signals on behavior whose precision increases with the intensity of gender marking in the language, deterring deviant behavior from risk averse individuals. Finally, even if our language variables correlation did not vanish when we included cultural beliefs or values related to gender roles in our regressions, language may be a marker of distant past culture and influence inherited gender role prescriptions. That is, it may influence gender roles themselves.

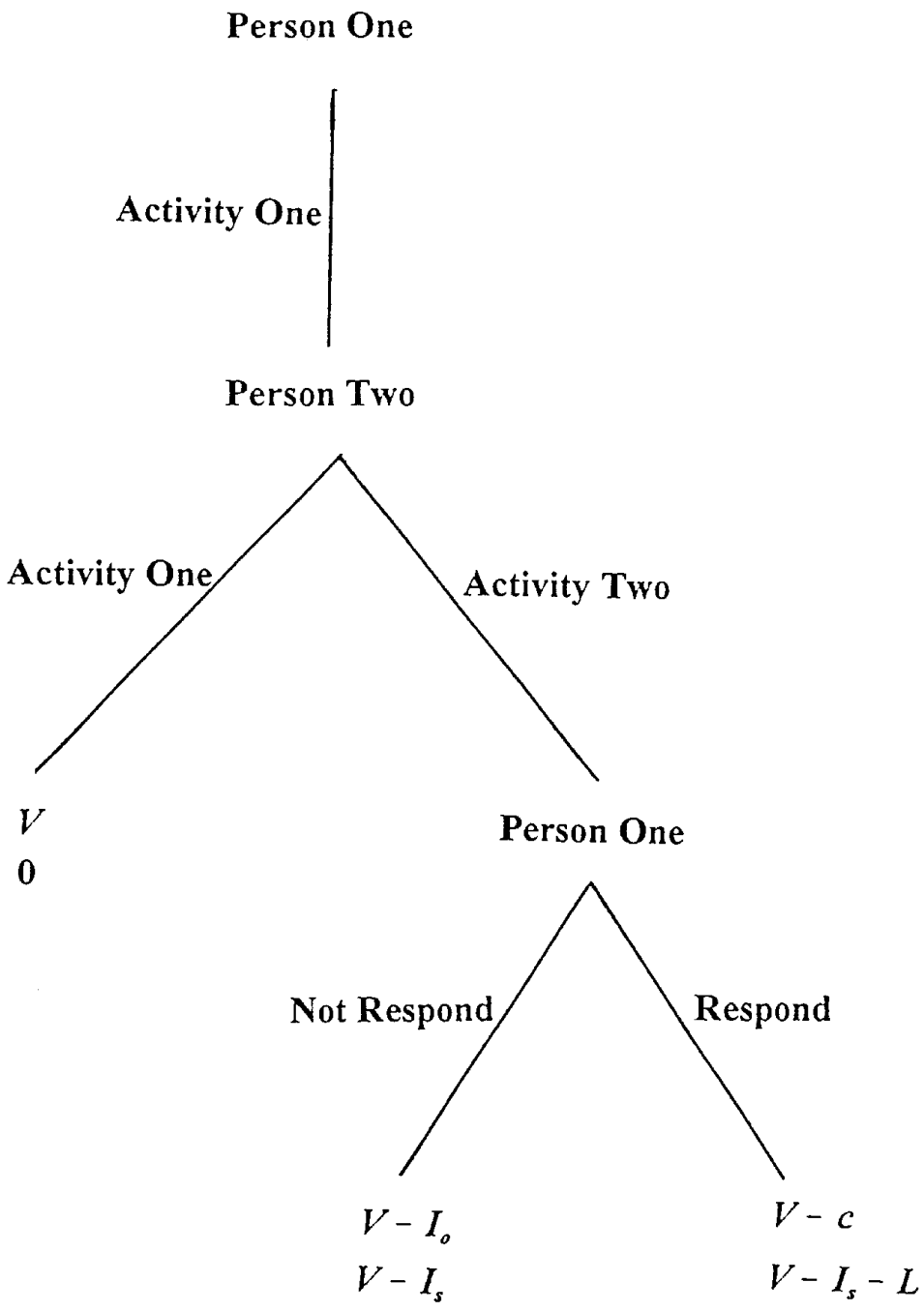

FIGURE I
Game Tree of Interaction between Person One and Person Two

4 Conclusion

Regardless of the socio-economic aspect considered (labor, occupation, political), of the language variable used from a rich set of language variables capturing female/male distinctions in languages grammar, and of the scope of analysis (country or individual level, migrants population), this paper presents evidence suggesting that the way we speak matters, specially for women and their degree and type of participation in economic life. Countries in which the dominant language marks gender more intensively exhibit significantly lower female labor force and a labor market where occupations segregated by gender. Political participation is also lowest in those countries, which are also more likely to regulate women's presence in politics since market forces fail to integrate women economically and politically. Yet, US immigrants coming from countries with highly gendered language are more likely to supply labor and earn lower wages. These results suggest both the language we think in and we speak with matter.

These results are robust to a large set of control variables including historical influences such as colonization, contemporary forces that shape the role of women today such as religion and geographical variables that may deal with part of the omitted variable bias of cross country level regressions.

To go beyond those we analyzed the impact of the language spoken at home on individual choices related to participating in the labor force, working in agricultural related occupations and managing big teams as opposed to small ones. To do so we relied on World Value survey data that allow us to exploit both within country and across country variations in language structure. Even when we included cultural values or beliefs regarding gender roles we find that women speaking gender intensive languages are home are less likely to be in the labor force, to work in agriculture and to manage big teams of employees. We also explored the influence of the language spoken at home across women living in countries with high gendered or low gendered dominant language. Further, we analyzed the influence of countries dominant language across individuals with high or low gendered home language. The results suggests a rich interaction between the language we think in and the language we speak.

Results for men were also presented and are consistent with the hypothesis we formalize at the end of the paper. That is, gender marking may either reflect gender roles inherited from the distant past, or influence the cognitive framework of speakers and the formation of their gender identity, as well as the salience of other's gender related behavior. Future work will deepen the analysis of this and other datasets to further investigate these complementary mechanisms and the fascinating relation between languages and economic outcomes.

5 Tables

Table 1: *Female Labor Force Participation Rate*

	Baseline					Baseline with Geography and Colonization											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
NG2		-17.10***				-17.33***	-17.95***	-17.68***									
		(-5.935)				(-6.181)	(-6.406)	(-6.168)									
NG3		-5.020				-3.678	-3.883	-4.579									
		(-1.652)				(-1.227)	(-1.345)	(-1.519)									
SBY			-12.98***						-12.70***	-12.65***	-13.03***						
			(-4.780)						(-4.691)	(-4.738)	(-4.790)						
SBN			3.221						3.227	3.591	3.040						
			(0.621)						(0.625)	(0.707)	(0.584)						
GAL				-6.336								-5.891	-7.700	-6.892			
				(-1.041)								(-0.991)	(-1.262)	(-1.128)			
GAH				-14.66***								-15.39***	-16.08***	-15.30***			
				(-3.626)								(-3.885)	(-4.038)	(-3.736)			
GPL					-1.526										-0.683	-0.0849	-0.496
					(-0.526)										(-0.240)	(-0.0294)	(-0.170)
GPH					-10.82***										-12.73***	-12.92***	-11.15***
					(-3.573)										(-4.185)	(-4.292)	(-3.720)
Dist_equ						-19.43***			-10.94			-21.28**			-18.91**		
						(-2.707)			(-1.479)			(-2.201)			(-2.599)		
Trop_sh							9.463***			9.849***			8.826*			8.665**	
							(2.776)			(2.727)			(1.907)			(2.551)	
Frost_days							-0.115			0.114			-0.0483			-0.0524	
							(-0.606)			(0.586)			(-0.175)			(-0.270)	
Coast_sh							0.825			5.705			1.470			1.113	
							(0.202)			(1.318)			(0.235)			(0.282)	
Landlocked							0.985			3.302			-2.924			3.027	
							(0.308)			(0.963)			(-0.554)			(0.961)	
Colo								4.498*			2.038			3.241			4.523*
								(1.878)			(0.831)			(0.984)			(1.916)
dom_sh	-10.10**	-5.190	-5.696	-8.472	0.452	-3.564	-3.948	-4.661	-5.171	-6.423	-5.657	-6.765	-8.576	-8.440	2.364	2.060	1.425
	(-2.015)	(-0.891)	(-0.946)	(-1.124)	(0.0788)	(-0.625)	(-0.705)	(-0.808)	(-0.862)	(-1.090)	(-0.938)	(-0.915)	(-1.141)	(-1.120)	(0.420)	(0.361)	(0.251)
log_inc	-118.2***	-96.01***	-97.07***	-133.5***	-116.8***	-102.0***	-89.99***	-96.67***	-103.1***	-89.67***	-98.70***	-141.4***	-142.3***	-135.3***	-116.7***	-94.09***	-119.5***
	(-5.432)	(-3.875)	(-3.742)	(-4.299)	(-4.459)	(-4.216)	(-3.688)	(-3.945)	(-3.945)	(-3.368)	(-3.789)	(-4.632)	(-4.120)	(-4.349)	(-4.573)	(-3.462)	(-4.610)
log_inc_sq	15.99***	12.61***	13.12***	17.75***	15.34***	13.84***	12.45***	12.90***	14.19***	12.58***	13.45***	19.24***	19.42***	18.14***	15.67***	12.75***	15.85***
	(5.302)	(3.703)	(3.710)	(4.156)	(4.263)	(4.139)	(3.750)	(3.827)	(3.949)	(3.484)	(3.773)	(4.555)	(4.181)	(4.229)	(4.465)	(3.456)	(4.446)
openness	-0.0542**	-0.0560**	-0.0412	-0.0441	-0.0227	-0.0691***	-0.0736***	-0.0582**	-0.0482*	-0.0608**	-0.0420	-0.0631*	-0.0670**	-0.0474	-0.0357	-0.0418	-0.0211
	(-2.205)	(-2.232)	(-1.587)	(-1.324)	(-0.894)	(-2.777)	(-3.009)	(-2.342)	(-1.838)	(-2.342)	(-1.616)	(-1.876)	(-1.996)	(-1.414)	(-1.409)	(-1.627)	(-0.839)
gov_size	-0.0142	0.108	0.114	0.0923	-0.102	0.0194	-0.132	0.0720	0.0623	-0.0803	0.0968	0.0129	-0.0922	0.0665	-0.132	-0.155	-0.131
	(-0.0982)	(0.454)	(0.467)	(0.295)	(-0.608)	(0.0831)	(-0.577)	(0.306)	(0.253)	(-0.330)	(0.394)	(0.0420)	(-0.299)	(0.212)	(-0.804)	(-0.940)	(-0.788)
oil_rents	-0.00160**	-0.000557	-0.000661	-0.000787	-0.000277	-0.000699	-0.000603	-0.000285	-0.000775	-0.000746	-0.000555	-0.000880	-0.000823	-0.000622	-0.000259	-2.45e-05	-1.31e-05
	(-2.317)	(-0.863)	(-0.994)	(-0.981)	(-0.416)	(-1.109)	(-0.990)	(-0.436)	(-1.164)	(-1.156)	(-0.819)	(-1.123)	(-1.064)	(-0.758)	(-0.399)	(-0.0377)	(-0.0195)
demo	0.271	0.769**	1.072***	1.149**	0.666*	0.864**	0.674*	0.831**	1.148***	0.801**	1.112***	1.273***	1.099**	1.167**	0.711**	0.526	0.685*
	(0.809)	(2.203)	(3.043)	(2.560)	(1.868)	(2.530)	(1.877)	(2.395)	(3.243)	(2.113)	(3.124)	(2.880)	(2.212)	(2.597)	(2.045)	(1.445)	(1.944)
log_pop	-3.645**	-3.262*	-2.630	-3.112	-1.505	-3.430*	-3.222*	-2.710	-2.744	-2.395	-2.390	-3.118	-3.345	-2.779	-1.752	-1.318	-1.029
	(-2.299)	(-1.795)	(-1.395)	(-1.319)	(-0.839)	(-1.939)	(-1.860)	(-1.488)	(-1.462)	(-1.291)	(-1.251)	(-1.353)	(-1.438)	(-1.165)	(-1.001)	(-0.749)	(-0.575)
comm	8.945***	4.538*	8.827***	6.523	9.555***	8.265***	9.467***	5.370*	11.20***	10.84***	9.350***	9.661**	10.42**	6.987	11.97***	10.99***	10.15***
	(3.577)	(1.678)	(3.352)	(1.537)	(3.454)	(2.783)	(2.983)	(1.981)	(3.646)	(3.234)	(3.449)	(2.205)	(2.151)	(1.636)	(4.197)	(3.411)	(3.689)
Constant	287.1***	252.3***	241.8***	317.5***	272.1***	263.6***	231.2***	245.4***	252.7***	217.1***	241.0***	332.8***	328.5***	316.1***	274.1***	220.9***	269.2***
	(7.392)	(5.613)	(5.021)	(5.693)	(5.679)	(6.001)	(5.075)	(5.503)	(5.214)	(4.283)	(4.997)	(6.064)	(5.080)	(5.665)	(5.870)	(4.309)	(5.684)
Observations	179	124	124	90	121	124	124	124	124	124	124	90	90	90	121	121	121
R-squared	0.293	0.478	0.446	0.443	0.473	0.511	0.556	0.494	0.457	0.502	0.449	0.476	0.510	0.450	0.504	0.531	0.491
Adj. R-squared	0.255	0.427	0.392	0.364	0.420	0.458	0.495	0.440	0.398	0.433	0.390	0.394	0.411	0.364	0.449	0.464	0.434

t-statistics in parentheses

***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 2: *Female Share of the Labor Force*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(10)	(11)	(12)	(14)	(15)	(16)	(18)	(19)	(20)
NG2		-9.151***				-9.192***	-9.071***	-9.558***									
		(-5.852)				(-5.875)	(-5.704)	(-6.238)									
NG3		-2.082				-1.823	-1.771	-1.800									
		(-1.257)				(-1.087)	(-1.074)	(-1.112)									
SBY			-6.723***						-6.729***	-6.423***	-6.780***						
			(-4.521)						(-4.495)	(-4.309)	(-4.586)						
SBN			2.125						2.124	2.712	1.948						
			(0.748)						(0.744)	(0.956)	(0.689)						
GAL				-3.459								-3.358	-4.329	-3.945			
				(-1.056)								(-1.023)	(-1.271)	(-1.211)			
GAH				-7.624***								-7.793***	-7.899***	-8.183***			
				(-3.501)								(-3.562)	(-3.551)	(-3.748)			
GPL					0.102										0.149	0.472	0.837
					(0.0649)										(0.0943)	(0.294)	(0.541)
GPH					-5.050***										-5.141***	-4.990***	-5.223***
					(-2.997)										(-2.974)	(-2.914)	(-3.188)
Dist_equ						-3.962			0.265			-4.902			-1.045		
						(-0.988)			(0.0643)			(-0.918)			(-0.259)		
Trop_sh							4.504**			4.590**			4.123			4.870**	
							(2.324)			(2.275)			(1.596)			(2.522)	
Frost_days							0.0653			0.174			0.0938			0.165	
							(0.602)			(1.589)			(0.608)			(1.521)	
Coast_sh							2.054			4.521*			2.735			1.389	
							(0.882)			(1.861)			(0.782)			(0.637)	
Landlocked							1.009			2.017			-0.539			1.681	
							(0.536)			(1.019)			(-0.183)			(0.936)	
Colo								3.359**			2.064			2.839			3.382***
								(2.612)			(1.535)			(1.616)			(2.715)
Dom_sh	-4.564*	-2.058	-2.344	-3.743	-0.983	-1.727	-2.244	-1.643	-2.357	-3.229	-2.267	-3.348	-4.613	-3.703	-0.886	-1.469	-0.269
	(-1.760)	(-0.651)	(-0.709)	(-0.922)	(-0.320)	(-0.544)	(-0.705)	(-0.533)	(-0.709)	(-0.980)	(-0.690)	(-0.820)	(-1.100)	(-0.922)	(-0.285)	(-0.466)	(-0.0898)
Log_inc	-59.37***	-41.34***	-44.06***	-62.72***	-58.57***	-42.70***	-38.14***	-42.29***	-43.91***	-37.99**	-45.91***	-64.54***	-64.58***	-64.34***	-58.66***	-49.81***	-61.29***
	(-5.265)	(-3.054)	(-3.096)	(-3.752)	(-4.092)	(-3.138)	(-2.698)	(-3.204)	(-3.030)	(-2.524)	(-3.234)	(-3.830)	(-3.351)	(-3.881)	(-4.079)	(-3.244)	(-4.396)
Log_inc_sq	8.323***	5.767***	6.345***	8.689***	8.010***	6.036***	5.534***	6.048***	6.318***	5.603***	6.698***	9.032***	9.059***	9.036***	8.041***	6.979***	8.489***
	(5.324)	(3.100)	(3.269)	(3.779)	(4.071)	(3.210)	(2.879)	(3.330)	(3.170)	(2.746)	(3.448)	(3.873)	(3.495)	(3.953)	(4.062)	(3.343)	(4.423)
Openness	-0.0199	-0.0237*	-0.0149	-0.0211	-0.0108	-0.0263*	-0.0318**	-0.0252*	-0.0147	-0.0244*	-0.0156	-0.0255	-0.0305	-0.0239	-0.0114	-0.0191	-0.00924
	(-1.545)	(-1.743)	(-1.049)	(-1.176)	(-0.787)	(-1.900)	(-2.283)	(-1.897)	(-1.016)	(-1.683)	(-1.106)	(-1.372)	(-1.629)	(-1.342)	(-0.818)	(-1.353)	(-0.695)
Gov_size	0.00805	0.1000	0.105	0.0633	-0.0870	0.0822	0.0151	0.0740	0.107	0.0446	0.0885	0.0450	-0.000314	0.0406	-0.0897	-0.109	-0.112
	(0.104)	(0.778)	(0.787)	(0.376)	(-0.885)	(0.633)	(0.116)	(0.589)	(0.785)	(0.328)	(0.663)	(0.265)	(-0.00182)	(0.243)	(-0.904)	(-1.095)	(-1.169)
Oil_rents	-0.00188***	-0.00134***	-0.00140***	-0.00159***	-0.00136***	-0.00137***	-0.00136***	-0.00113***	-0.00139***	-0.00142***	-0.00129***	-0.00161***	-0.00160***	-0.00144***	-0.00136***	-0.00125***	-0.00116***
	(-5.272)	(-3.819)	(-3.837)	(-3.675)	(-3.818)	(-3.887)	(-3.942)	(-3.236)	(-3.788)	(-3.950)	(-3.489)	(-3.715)	(-3.709)	(-3.300)	(-3.799)	(-3.506)	(-3.289)
Demo	0.482***	0.732***	0.911***	0.963***	0.722***	0.752***	0.608***	0.780***	0.909***	0.683***	0.952***	0.992***	0.856***	0.979***	0.725***	0.591***	0.739***
	(2.771)	(3.867)	(4.727)	(3.983)	(3.787)	(3.949)	(2.969)	(4.207)	(4.645)	(3.220)	(4.922)	(4.063)	(3.084)	(4.085)	(3.779)	(2.930)	(3.987)
Log_pop	-1.606*	-1.372	-0.981	-1.006	-0.627	-1.403	-1.314	-0.947	-0.979	-0.837	-0.726	-1.007	-0.996	-0.713	-0.634	-0.491	-0.235
	(-1.908)	(-1.392)	(-0.950)	(-0.791)	(-0.639)	(-1.423)	(-1.337)	(-0.972)	(-0.943)	(-0.811)	(-0.699)	(-0.792)	(-0.767)	(-0.561)	(-0.644)	(-0.499)	(-0.243)
Comm	8.206***	6.468***	9.027***	6.855***	9.637***	7.240***	7.226***	7.130***	8.969***	8.157***	9.572***	7.578***	7.261***	7.262***	9.785***	8.532***	10.17***
	(6.335)	(4.388)	(6.255)	(2.999)	(6.215)	(4.339)	(3.909)	(4.887)	(5.252)	(4.265)	(6.478)	(3.131)	(2.685)	(3.189)	(5.898)	(4.613)	(6.696)
Constant	150.3***	118.9***	116.5***	155.5***	143.9***	121.4***	108.4***	114.4***	116.2***	100.7***	116.0***	159.1***	156.1***	154.3***	144.2***	122.9***	142.8***
	(7.462)	(4.852)	(4.414)	(5.179)	(5.541)	(4.927)	(4.109)	(4.780)	(4.328)	(3.515)	(4.421)	(5.248)	(4.324)	(5.189)	(5.523)	(4.271)	(5.659)
Observations	176	123	123	90	119	123	123	123	123	123	123	90	90	90	119	119	119
R-squared	0.417	0.636	0.605	0.586	0.670	0.639	0.660	0.657	0.605	0.632	0.613	0.591	0.609	0.600	0.671	0.691	0.692
Adj. R-square	0.385	0.599	0.566	0.528	0.637	0.599	0.613	0.619	0.562	0.581	0.571	0.527	0.530	0.537	0.633	0.646	0.657

t-statistics in parenthesis
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 3: *Occupation Profile, Language and Gender*

	Agriculture (Agri_fem)					Industry (Ind_fem)					Services (Serv_fem)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
NG2		-0.220*					0.0220					0.256***			
		(-1.986)					(0.267)					(3.289)			
NG3		0.0545					-0.143*					0.0627			
		(0.480)					(-1.672)					(0.749)			
SBY			-0.0992					-0.0280					0.193***		
			(-0.964)					(-0.381)					(2.660)		
SBN			0.125					-0.362**					-0.0114		
			(0.526)					(-2.330)					(-0.0827)		
GAL				-0.152					-0.191					0.191	
				(-0.660)					(-1.135)					(1.192)	
GAH				-0.0291					-0.0749					0.183*	
				(-0.174)					(-0.626)					(1.698)	
GPL					0.0556					-0.0692					0.134
					(0.521)					(-0.682)					(1.611)
GPH					-0.370***					0.113					0.422***
					(-3.308)					(1.088)					(4.999)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.717*	2.734	3.733	3.931	1.769	-2.989*	1.170	1.369	0.978	1.287	-0.354	1.144	0.706	0.824	1.840
	(1.911)	(1.102)	(1.473)	(1.238)	(0.789)	(-1.880)	(0.795)	(0.930)	(0.565)	(0.785)	(-0.274)	(0.878)	(0.537)	(0.539)	(1.414)
Observations	126	89	89	61	88	146	105	105	75	103	131	101	101	73	98
R-squared	0.357	0.398	0.353	0.345	0.454	0.102	0.192	0.201	0.183	0.180	0.202	0.385	0.362	0.329	0.433
Adj. R-square	0.307	0.312	0.261	0.198	0.375	0.0422	0.0965	0.107	0.0401	0.0807	0.143	0.309	0.283	0.209	0.361

t-statistics in parentheses

***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 4: Women in the Labor Force, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.16884	1.328758	1.235439	1.16995	1.32238	1.237578	1.166882	1.310543	1.22176	1.162817
	(5.67)	(5.47)	(5.52)	(5.61)	(5.39)	(5.5)	(5.65)	(6.29)	(5.7)	(6.42)
ChildrenValue	0.8260911	0.753388	0.825106	0.896751	0.762978	0.823114	0.890944	0.775471	0.839307	0.891429
	(-2.93)	(-3.02)	(-2.61)	(-2.13)	(-2.66)	(-2.96)	(-2.18)	(-2.96)	(-2.9)	(-2.59)
HGII		0.888808	0.866686	0.898413						
		(-1.77)	(-2.68)	(-1.9)						
HGIIv1					0.871515	0.855965	0.928561			
					(-1.8)	(-2.43)	(-1.41)			
HGIIv2								0.821587	0.780708	0.872327
								(-2.57)	(-3.34)	(-2.12)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	88640	26459	25418	25418	28428	27356	27356	33748	32196	32196
Pseudo R-sq2	0.1444	0.1	0.116	0.1362	0.0975	0.1134	0.1361	0.1071	0.1191	0.1381
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-statistics in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Controls: country, income decile within country, marital status, education, age, number of children, survey wave and religion.

Table 5: Men in Labor Force, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MenSuperior	1.029842	0.990275	1.032042	1.025843	0.990015	1.03025	1.017188	0.999257	1.029358
	(0.68)	(-0.26)	(0.88)	(0.66)	(-0.31)	(0.92)	(0.48)	(-0.02)	(0.99)
ChildrenValue	0.941691	0.95979	0.878437	0.995412	1.011539	0.913881	0.95765	0.974257	0.929026
	(-0.71)	(-0.59)	(-1.6)	(-0.06)	(0.16)	(-1.08)	(-0.62)	(-0.41)	(-1.05)
HGII	1.103422	1.101379	1.093287						
	(2.94)	(2.62)	(1.35)						
HGIIv1				1.105629	1.098883	1.067764			
				(2.84)	(2.46)	(1.03)			
HGIIv2							1.069534	1.045753	0.955737
							(1.76)	(1.23)	(-0.82)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	22858	22307	22307	24760	24165	24165	29513	28475	28475
Pseudo R-sq	0.135	0.1343	0.1367	0.1387	0.1372	0.1404	0.1547	0.1523	0.1555
Continent	no	yes	no	no	yes	no	no	yes	no
Region	no	no	yes	no	no	yes	no	no	yes

z-stats in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level. Controls: country, income, marital status, education, age, number of children, survey wave and religion.

Table 6: Women in Agricultural Occupation, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.102064	1.215749	1.298338	1.202199	1.258549	1.315802	1.184573	1.235818	1.247591	1.166782
	(1.51)	(2.47)	(3.85)	(2.64)	(3.33)	(4.13)	(2.88)	(3.54)	(3.44)	(2.91)
ChildrenValue	1.768183	1.344292	1.209051	1.345542	1.831214	1.585326	1.396493	1.733168	1.461982	1.370591
	(3.28)	(1.97)	(1.57)	(2.07)	(2.49)	(2.34)	(2.30)	(3.17)	(2.71)	(2.53)
HDI	0.0467213	0.1522971	0.8615059	0.000546	0.24986	0.887922	0.0004783	0.196654	0.139063	.0004863
	(-1.78)	(-1.8)	(-0.1)	(-2.68)	(-1.02)	(-0.06)	(-3.21)	(-1.25)	(-0.9)	(-3.64)
HGII		0.7919264	0.8092585	0.92667						
		(-2.11)	(-1.97)	(-0.67)						
HGIIv1				0.663081	0.743469	0.8017462				
				(-2.94)	(-2.39)	(-1.93)				
HGIIv2								0.637421	0.725725	0.7983778
								(-3.17)	(-2.09)	(-1.58)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	45954	26558	26558	26558	28553	28553	28553	33463	33463	33463
Pseudo R-sq	0.1377	0.1547	0.1675	0.1751	0.1728	0.1802	0.2089	0.1715	0.1812	0.2114
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-stats in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Controls: country, income, marital status, education, age, number of children, survey wave and religion.

Table 7: Men in Agriculture Occupations, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MenSuperior	1.028064	1.059686	1.020172	1.03176	1.057267	1.008821	1.033534	1.04449	1.003782
	(0.42)	(1.29)	(0.56)	(0.57)	(1.26)	(0.25)	(0.66)	(0.91)	(0.10)
ChildrenValue	1.437079	1.375024	1.369794	1.534811	1.42337	1.348919	1.37176	1.274213	1.229634
	(2.82)	(2.96)	(3.13)	(2.8)	(2.97)	(2.89)	(2.34)	(2.34)	(2.17)
HDI	0.033806	0.013723	0.001421	0.035775	0.014335	0.0005798	0.025252	0.00566	0.0002363
	(-3.37)	(-4.35)	(-5.36)	(-3.28)	(-4.01)	(-5.95)	(-3.85)	(-4.86)	(-7.34)
HGII	1.016715	1.006217	1.080909						
	(0.27)	(0.1)	(1.34)						
HGIIv1				0.957881	0.98473	0.9567394			
				(-0.49)	(-0.2)	(-0.62)			
HGIIv2							0.939341	0.989582	0.9299764
							(-0.71)	(-0.12)	(-0.85)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	23237	23237	23237	25171	25171	25171	29554	29554	29554
Pseudo R-sq	0.1405	0.1540	0.1561	0.1481	0.154	0.1657	0.1426	0.1506	0.1636
Continent	no	yes	no	no	yes	no	no	yes	no
Region	no	no	yes	no	no	yes	no	no	yes

z-stats in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level. Controls: country, income, marital status, education, age, number of children, survey wave and religion.

Table 8: US Labor Supply (working hours) and Home Country Dominant Language, NIS data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours
age	87.21***	83.06***	82.62***	84.15***	85.04***	85.40***	86.31***	76.52***	78.31***	76.82***
	(26.96)	(28.02)	(26.37)	(28.55)	(27.56)	(25.96)	(28.18)	(23.43)	(22.26)	(23.34)
age2	-104.9***	-100.7***	-98.16***	-102.4***	-99.22***	-98.03***	-101.6***	-93.29***	-93.43***	-93.63***
	(29.79)	(31.01)	(29.24)	(31.86)	(30.66)	(28.79)	(31.61)	(25.32)	(24.05)	(25.21)
pre_school	11.18	14.36*	15.82**	12.63	-8.748	-5.624	-9.878	18.61***	17.27***	18.23***
	(7.742)	(7.593)	(7.379)	(7.952)	(7.232)	(7.056)	(7.256)	(6.682)	(6.310)	(6.718)
us_school	64.32***	60.00***	60.77***	60.06***	17.83	21.40	20.11	60.78***	60.56***	60.87***
	(13.14)	(14.52)	(13.57)	(14.82)	(22.08)	(21.03)	(22.36)	(17.43)	(15.96)	(17.50)
curr_us_school	-298.3***	-346.5***	-319.6***	-346.2***	-353.2***	-329.4***	-351.9***	-370.4***	-340.9***	-370.8***
	(69.94)	(67.65)	(66.58)	(68.03)	(66.55)	(68.15)	(67.49)	(69.55)	(68.98)	(69.40)
married	-215.9***	-264.1***	-245.1***	-269.3***	-264.0***	-245.7***	-271.8***	-287.7***	-268.7***	-288.1***
	(70.74)	(79.71)	(80.85)	(81.03)	(72.45)	(73.04)	(74.88)	(69.98)	(70.22)	(69.97)
nb_child	-52.74***	-59.05***	-67.20***	-57.51***	-54.13***	-60.34***	-50.71***	-51.09***	-60.67***	-51.39***
	(15.36)	(16.75)	(18.61)	(16.31)	(14.90)	(16.44)	(14.69)	(18.56)	(20.49)	(18.64)
pre_work	90.25*	112.4**	127.7***	115.6**	115.8***	130.7***	117.0***	93.29*	104.2*	96.54*
	(48.14)	(46.51)	(46.19)	(47.92)	(42.20)	(42.81)	(43.68)	(55.79)	(53.22)	(56.57)
GII		64.25**			76.87***			44.92		
		(25.07)			(18.30)			(29.99)		
giiv1			102.8***			106.5***			102.5***	
			(33.80)			(27.40)			(36.55)	
giiv2				79.89**			85.46***			55.05
				(32.19)			(26.06)			(34.18)
Constant	-855.0*	-932.2*	-1,010**	-929.9*	-1,142**	-1,195**	-1,111**	-539.1	-633.7	-553.9
	(459.7)	(519.0)	(486.9)	(501.3)	(510.1)	(482.8)	(497.7)	(485.8)	(470.7)	(481.8)
Observations	2,990	2,536	2,820	2,536	2,466	2,750	2,466	2,328	2,610	2,328
R-squared	0.063	0.078	0.077	0.077	0.114	0.107	0.111	0.082	0.082	0.083
Adjusted R-square	0.0605	0.075	0.0742	0.0741	0.109	0.103	0.106	0.0765	0.0769	0.0770
F test	26.35	37.04	33.21	37.18	54.82	63.25	170.9	37.21	38.43	37.21
English	no	no	no	no	yes	yes	yes	no	no	no
Race	no	no	no	no	no	no	no	yes	yes	yes

Robust standard errors in parentheses, clustered at the country level.
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 9: US Labor Supply (working hours) and Home Country Dominant Language, NIS data: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours
age	83.72***	84.48***	84.54***	67.66***	66.55***	68.47***	77.16***	78.98***	77.30***
	(26.83)	(25.40)	(27.18)	(24.01)	(22.67)	(23.76)	(22.64)	(21.75)	(22.61)
age2	-97.53***	-97.16***	-99.02***	-78.93***	-75.90***	-80.67***	-89.01***	-89.63***	-89.18***
	(29.80)	(28.15)	(30.30)	(26.26)	(24.64)	(26.07)	(24.54)	(23.54)	(24.50)
pre_school	-7.057	-4.585	-7.637	-3.711	-1.234	-4.107	-5.735	-4.207	-5.830
	(7.943)	(7.470)	(8.045)	(6.005)	(5.996)	(5.975)	(7.011)	(6.764)	(7.035)
us_school	11.26	15.99	13.24	17.81	20.82	19.49	10.25	13.18	10.55
	(18.68)	(17.86)	(18.86)	(20.06)	(19.12)	(20.04)	(22.90)	(21.63)	(22.90)
curr_us_school	-344.5***	-319.3***	-342.2***	-346.1***	-325.8***	-346.0***	-375.7***	-351.1***	-375.7***
	(70.55)	(71.64)	(71.00)	(76.29)	(74.96)	(77.30)	(70.21)	(70.50)	(70.19)
married	-250.4***	-232.8***	-251.7***	-238.5***	-214.0***	-245.8***	-292.3***	-274.7***	-292.6***
	(73.58)	(74.31)	(74.24)	(56.85)	(59.12)	(58.30)	(61.96)	(61.39)	(61.92)
nb_child	-54.46***	-60.35***	-53.46***	-55.12***	-62.76***	-51.76***	-47.97***	-54.59***	-48.19***
	(15.26)	(16.90)	(15.35)	(17.09)	(19.55)	(16.94)	(15.69)	(17.08)	(15.75)
pre_work	95.39**	112.0**	96.66**	121.0***	136.7***	121.3***	103.1**	107.9**	104.8**
	(42.85)	(43.13)	(43.21)	(36.17)	(38.16)	(36.55)	(47.12)	(46.13)	(47.46)
english_speak_1	777.2***	709.8***	752.7***	639.2***	591.4***	620.8***	851.5***	781.8***	848.3***
	(100.8)	(101.5)	(98.59)	(109.1)	(113.6)	(106.0)	(104.7)	(107.6)	(102.4)
english_speak_2	423.5***	413.4***	405.2***	303.4***	313.1***	288.7***	472.8***	469.6***	470.3***
	(93.35)	(101.0)	(96.27)	(98.68)	(101.6)	(98.15)	(97.00)	(98.33)	(95.97)
english_speak_3	356.7***	312.9***	350.3***	289.1***	247.1***	282.5***	348.8***	301.3***	348.1***
	(76.94)	(86.44)	(78.93)	(74.79)	(84.93)	(76.28)	(76.59)	(86.73)	(76.64)
GII	91.62***			65.20***			16.52		
	(23.72)			(12.03)			(20.07)		
giiv1		108.1***			96.04***			36.66	
		(36.00)			(19.18)			(30.44)	
giiv2			119.0***			68.63***			22.35
			(32.37)			(22.88)			(20.44)
Constant	-1,146**	-1,097***	-1,150**	-1,019**	-435.9	-995.5**	-705.7	-729.3	-715.5
	(438.5)	(408.3)	(433.8)	(477.7)	(480.1)	(459.1)	(536.3)	(520.6)	(534.8)
Observations	2,466	2,750	2,466	2,466	2,750	2,466	2,259	2,541	2,259
R-squared	0.126	0.116	0.124	0.147	0.139	0.145	0.124	0.115	0.124
Adjusted R-square	0.118	0.11	0.117	0.141	0.134	0.139	0.117	0.109	0.117
F test	143.5	72.34	146.5	50.57	269.4	339.4	181.0	60.31	192.6
English	yes	yes	yes	yes	yes	yes	yes	yes	yes
Race	no	no	no	no	no	no	yes	yes	yes
Religion	yes	yes	yes	no	no	no	no	no	no
Visa	no	no	no	yes	yes	yes	no	no	no

Robust standard errors in parentheses, clustered at the country level.
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 10: US Pay and Home Country Dominant Language, NIS data

VARIABLES	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay
age	-0.425	-0.354	-0.349	-0.391	-0.324	-0.261	-0.282	-0.433	-0.577	-0.421
	(2.584)	(2.856)	(2.658)	(2.861)	(2.991)	(2.599)	(3.011)	(3.136)	(2.734)	(3.120)
age2	1.399	1.528	1.350	1.624	1.501	1.321	1.507	1.756	1.675	1.766
	(3.588)	(3.934)	(3.684)	(3.941)	(4.114)	(3.637)	(4.135)	(4.336)	(3.808)	(4.321)
pre_school	1.732***	1.778**	1.626***	1.955**	1.213	0.925	1.325	2.173*	1.734*	2.241*
	(0.522)	(0.770)	(0.561)	(0.825)	(0.802)	(0.603)	(0.803)	(1.149)	(0.905)	(1.180)
us_school	2.570**	2.774*	2.345**	2.883*	3.407	2.145	3.382	3.803*	2.950*	3.856*
	(1.018)	(1.555)	(1.008)	(1.574)	(2.214)	(1.431)	(2.165)	(2.205)	(1.495)	(2.221)
curr_us_school	-6.734	-0.861	-1.055	-0.817	-7.162	-4.495	-7.203	3.218	2.343	3.429
	(6.782)	(4.261)	(3.135)	(4.197)	(9.801)	(6.679)	(9.441)	(5.150)	(3.831)	(5.218)
married	3.452	2.045	1.017	2.435	6.692	3.494	7.103	3.358	1.836	3.698
	(2.792)	(3.369)	(2.987)	(3.332)	(5.449)	(3.849)	(5.288)	(4.262)	(3.723)	(4.222)
nb_child	-0.252	-0.123	-0.142	-0.0914	-0.295	-0.0925	-0.253	-0.344	-0.319	-0.336
	(0.666)	(0.935)	(0.767)	(0.958)	(0.907)	(0.800)	(0.932)	(0.984)	(0.832)	(1.005)
pre_work	5.419	4.252	3.062	3.959	3.797	2.917	3.625	6.341	4.920	6.138
	(5.278)	(5.950)	(4.178)	(5.851)	(5.605)	(3.774)	(5.609)	(7.754)	(5.633)	(7.715)
GII		-6.478**			-6.825***			-7.318*		
		(2.517)			(2.193)			(4.287)		
giiv1			-5.070			-4.727			-2.975	
			(3.873)			(3.398)			(7.592)	
giiv2				-8.972**			-9.255***			-8.484*
				(3.545)			(2.997)			(4.382)
Constant	-13.57	0.519	-2.294	0.455	-6.796	-12.66	-8.729	-34.58	-26.00	-35.68
	(44.39)	(48.51)	(39.79)	(48.62)	(52.55)	(37.47)	(52.63)	(64.25)	(47.33)	(64.37)
Observations	362	302	347	302	287	332	287	283	328	283
R-squared	0.028	0.051	0.036	0.053	0.067	0.043	0.068	0.062	0.043	0.064
Adjusted R-square	0.00630	0.0221	0.0102	0.0243	0.0260	0.00722	0.0276	0.00958	-0.00290	0.0111
F test	3.891	7.764	2.561	6.841	12.84	7.425	12.02	18.27	7.817	12.58
English	no	no	no	no	yes	yes	yes	no	no	no
Race	no	no	no	no	no	no	no	yes	yes	yes

Robust standard errors in parentheses, clustered at the country level.
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 11: US Pay and Home Country Dominant Language, NIS data: Robustness Checks

VARIABLES	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay	us_pay
age	-1.919	-1.192	-1.968	-1.820	-1.896	-1.783	-0.368	-0.324	-0.629
	(3.007)	(2.587)	(3.011)	(2.180)	(1.931)	(2.215)	(3.336)	(3.320)	(2.699)
age2	3.677	2.648	3.786	3.232	3.206	3.237	1.690	1.662	1.840
	(4.101)	(3.621)	(4.126)	(3.161)	(2.868)	(3.203)	(4.571)	(4.552)	(3.778)
pre_school	1.349	0.980	1.415	0.853	0.415	0.915	1.272	1.308	0.919
	(0.904)	(0.629)	(0.910)	(0.973)	(0.808)	(0.949)	(0.938)	(0.946)	(0.648)
us_school	4.180	2.582	4.160	4.168*	2.669	4.135*	4.430	4.458	2.880
	(2.518)	(1.716)	(2.479)	(2.382)	(1.751)	(2.340)	(2.749)	(2.743)	(1.914)
curr_us_school	-6.505	-3.962	-7.189	-5.859	-4.154	-5.497	-5.412	-5.161	-1.579
	(10.33)	(6.844)	(9.786)	(8.225)	(6.116)	(7.852)	(12.52)	(12.44)	(8.579)
married	7.084	4.227	7.338	9.208	7.910	9.449	8.298	8.595	4.452
	(4.731)	(3.476)	(4.509)	(6.232)	(5.465)	(5.965)	(6.367)	(6.335)	(4.522)
nb_child	-0.661	-0.320	-0.640	-0.288	-0.211	-0.243	-0.426	-0.414	-0.220
	(0.835)	(0.897)	(0.836)	(1.002)	(0.881)	(1.040)	(0.955)	(0.976)	(0.836)
pre_work	5.021	4.197	4.971	5.234	4.115	4.809	5.896	5.775	5.013
	(5.374)	(4.132)	(5.353)	(5.868)	(4.201)	(5.845)	(7.099)	(7.100)	(5.020)
GII	-8.038***			-5.287***			-6.722*		
	(2.211)			(1.484)			(3.829)		
giiv1		-3.927			-2.634			-2.316	
		(2.666)			(2.361)			(7.129)	
giiv2			-11.30***			-7.545***			-7.840*
			(3.005)			(1.883)			(3.870)
Constant	-5.648	-6.880	19.62	16.53	13.67	17.22	-52.29	-39.38	-54.04
	(60.17)	(40.10)	(52.89)	(33.23)	(22.29)	(33.15)	(72.50)	(46.15)	(72.89)
Observations	287	332	287	287	332	287	268	313	268
R-squared	0.082	0.055	0.084	0.095	0.071	0.097	0.077	0.050	0.078
Adjusted R-square	0.0135	-0.00591	0.0155	0.0374	0.0212	0.04	0.0104	-0.00804	0.0118
F test				19.05	8.245	42.70	32.42	7.581	55.31
English	yes	yes	yes	yes	yes	yes	yes	yes	yes
Race	no	no	no	no	no	no	yes	yes	yes
Religion	yes	yes	yes	no	no	yes	no	no	no
Visa	no	no	no	yes	yes	no	no	no	no

Robust standard errors in parentheses, clustered at the country level.
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 12: US Work Hours Robustness Check: Maternal Language

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours	us_work_hours
age	87.21***	74.38**	75.08**	75.27**	70.43**	71.34**	72.06**	68.59***	69.16***	70.43***
	(26.96)	(29.00)	(29.16)	(28.35)	(27.98)	(28.32)	(27.55)	(25.19)	(25.00)	(24.62)
age2	-104.9***	-90.20***	-91.04***	-89.12***	-81.66**	-83.19***	-81.92***	-83.96***	-84.54***	-83.73***
	(29.79)	(31.98)	(32.19)	(31.34)	(30.61)	(31.05)	(30.11)	(27.70)	(27.42)	(27.09)
pre_school	11.18	11.85	10.62	11.54	-11.09	-11.86	-9.394	15.40**	14.76*	14.63**
	(7.742)	(7.829)	(8.206)	(7.445)	(7.265)	(7.365)	(6.824)	(7.505)	(7.491)	(6.989)
us_school	64.32***	57.06***	56.30***	59.18***	5.025	6.570	12.39	59.76***	59.66***	62.15***
	(13.14)	(12.32)	(12.17)	(12.32)	(17.17)	(17.25)	(18.14)	(15.74)	(15.79)	(15.56)
curr_us_school	-298.3***	-341.0***	-340.6***	-348.3***	-353.8***	-353.4***	-366.4***	-360.7***	-360.7***	-367.2***
	(69.94)	(66.07)	(66.00)	(64.35)	(67.75)	(68.36)	(65.98)	(68.13)	(67.94)	(65.94)
married	-215.9***	-228.4**	-226.3**	-219.6**	-264.0***	-264.5***	-253.6***	-257.1***	-255.5***	-249.5***
	(70.74)	(87.29)	(87.58)	(87.75)	(70.42)	(70.57)	(72.34)	(82.01)	(82.76)	(81.93)
nb_child	-52.74***	-58.12***	-59.16***	-66.52***	-47.77***	-47.12***	-54.44***	-50.53**	-51.67**	-59.81***
	(15.36)	(15.86)	(16.00)	(18.28)	(12.74)	(12.73)	(14.46)	(19.35)	(19.63)	(22.05)
pre_work	90.25*	109.9*	109.8*	124.5**	125.8**	123.0**	140.4**	83.88	86.25	101.3*
	(48.14)	(55.33)	(55.02)	(55.57)	(53.41)	(53.42)	(54.97)	(60.70)	(60.93)	(60.34)
gii		73.48***			92.58***			27.11		
		(27.28)			(22.06)			(33.13)		
giiv1			105.3***			119.1***			56.15	
			(34.22)			(28.40)			(35.44)	
giiv2				99.21***			107.5***			51.26
				(35.60)			(28.06)			(37.77)
Constant	-855.0*	-777.0	-803.0	-823.1	-911.7*	-906.0*	-922.5*	-301.8	-353.1	-393.1
	(459.7)	(532.6)	(531.8)	(520.1)	(535.6)	(537.5)	(525.4)	(487.3)	(484.7)	(477.1)
Observations	2,990	2,071	2,071	2,221	1,938	1,938	2,088	1,879	1,879	2,029
R-squared	0.063	0.070	0.072	0.069	0.115	0.113	0.104	0.074	0.075	0.072
Adjusted R-square	0.0605	0.0664	0.0675	0.0652	0.109	0.108	0.0991	0.0665	0.0675	0.0653
F test	26.35	23.82	25.03	22.47	183.8	189.8	48.70	26.44	24.59	18.42
English	no	no	no	no	yes	yes	yes	no	no	no
Religion	no	no	no	no	no	no	no	no	no	no
Race	no	no	no	no	no	no	no	yes	yes	yes
Visa	no							no	no	no

Robust standard errors in parentheses, clustered at the country level.
***p-value< 0.01, **p-value< 0.05, *p-value< 0.10

Table 13: Women as Managers of More Ten Employees, World Value Survey

	1	2	3	4	5	6	7
MenSup Exec	1.135621	1.488039	1.630351	1.492508	1.645223	1.337471	1.458923
	(0.97)	(2.90)	(3.56)	(3.10)	(3.85)	(2.89)	(3.34)
HGII		0.743122	0.6514932				
		(-2.19)	(-2.98)				
HGIIv1				0.6964699	0.64406		
				(-2.15)	(-2.40)		
HGIIv2						0.668463	0.5213442
						(-2.56)	(-3.41)
Controls	yes	yes	yes	yes	yes	yes	yes
N	748	449	407	470	428	575	533
Pseudo R-sq	0.0689	0.1352	0.1655	0.1254	0.1409	0.0954	0.1125
Continent	no	no	yes	no	yes	no	yes

z-stats in parentheses. Odd ratios of Logit regression are reported.

Robust standard errors clustered at country level.

Dependent variable: 1 if female managers manage teams of more 10 employees, 0 if teams less 10

Controls: country, income, marital status, education, age, number of children, survey wave and religion.

Table 14: Women Presence in Parliaments 1972-1989

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
commun	4.040***	5.039***	4.325***	4.291***	5.116***	4.282***	3.852***	6.023*	4.254***
	(6.121)	(3.222)	(2.843)	(5.312)	(3.116)	(2.950)	(2.818)	(1.841)	(2.758)
women_st	0.0204	0.0187	0.0127	0.0200	0.0186	0.0122	0.0305	0.0326	0.0333
	(1.056)	(0.868)	(0.764)	(1.109)	(0.857)	(0.679)	(1.468)	(0.899)	(1.380)
demo_fh_1	-0.195	-0.353	-0.273	-0.0957	-0.325	-0.310	-0.263	-0.387	-0.386
	(-0.665)	(-1.011)	(-0.627)	(-0.379)	(-0.967)	(-0.637)	(-0.789)	(-0.891)	(-0.830)
GII	-0.329**	-0.249*	-0.485**						
	(-2.298)	(-1.869)	(-2.437)						
GIIv1				-0.484***	-0.348*	-0.636***			
				(-3.006)	(-1.914)	(-3.349)			
GIIv2							-0.410***	-0.314***	-0.552***
							(-5.211)	(-2.848)	(-4.926)
N	964	970	905	1,026	1,019	967	1,274	1,250	1,215
Pseudo R^2	0.416	0.412	0.452	0.411	0.406	0.437	0.414	0.476	0.429
prim_en	yes	no	no	yes	no	no	yes	no	no
HDI	no	yes	no	no	yes	no	no	yes	no
second_en	no	no	yes	no	no	yes	no	no	yes

Coefficients are reported for logit regression

Dependent variable: 1 if percentage women in parliament above the mean

Robust Standard Errors in parenthesis, clustered by continent

All regressions control for the electoral system

*** $p < 0.01$, $**$ $p < 0.05$, * $p < 0.1$

Table 15: Adoption of Legislated Political Quotas for Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pol System	-0.724	0.0157	-0.316	-0.325	-0.0474	0.204	-0.0831	0.148	-0.207	-0.267
	(-1.578)	(0.0446)	(-0.937)	(-0.919)	(-0.0875)	(0.355)	(-0.186)	(0.258)	(-0.491)	(-0.475)
HDI	-0.0627	-3.119	-3.427**	-2.226	-3.309	-2.337	-3.308*	-2.842	-1.951	-1.319
	(-0.126)	(-1.610)	(-2.314)	(-1.468)	(-1.502)	(-0.832)	(-1.655)	(-1.113)	(-1.044)	(-0.559)
Women_vote	-0.00823	0.0162	0.000783	0.000709	0.0175	0.0291	-9.68e-05	0.0221	-0.000234	0.00133
	(-0.583)	(0.997)	(0.0808)	(0.0414)	(1.286)	(1.502)	(-0.00967)	(1.280)	(-0.0132)	(0.0723)
GII		1.167***			1.192***	0.707***				
		(3.129)			(2.795)	(3.138)				
GIIv1			1.134**				1.124**	0.783***		
			(2.316)				(2.044)	(2.797)		
GIIv2				0.880***					0.899***	0.437*
				(3.569)					(3.206)	(1.920)
Religion	No	No	No	No	No	Yes	No	Yes	No	Yes
Colonization	No	No	No	No	Yes	No	Yes	No	Yes	No
N	3,007	1,560	1,666	2,107	1,560	1,560	1,666	1,666	2,107	2,107
Psedo R-sq	0.0119	0.140	0.0893	0.0761	0.141	0.215	0.0966	0.207	0.0787	0.130
Chi-square	7.911	22.92	42.99	63.30	48.56	48.5	62.71	62.01	3248	322.9

Coefficients of Cox Hazard Model

Dependent variable: 1 for the year country adopts quotas, 0 otherwise

Robust Standard Errors in parenthesis, clustered by region

*** $p < 0.01$, $**$ $p < 0.05$, * $p < 0.1$

Table 16: Women in the Labor Force in Low Gendered Environments, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.28555	1.359658	1.324435	1.248835	1.375418	1.334489	1.258235	1.239441	1.22241	1.158756
	(5.32)	(5.44)	(5.72)	(6.60)	(5.77)	(6.01)	(6.62)	(6.00)	(5.81)	(5.20)
ChildrenValue	0.7216379	0.6917062	0.7938771	0.9011216	0.7486862	0.8237536	0.9157115	0.7394401	0.8184465	0.87807
	(-2.92)	(-2.66)	(-1.77)	(-1.21)	(-2.11)	(-1.50)	(-1.16)	(-2.46)	(-2.24)	(-2.15)
HGII		1.599147	1.173254	1.167549						
		(4.06)	(2.27)	(2.49)						
HGIIv1					1.748884	1.217353	1.151206			
					(3.53)	(2.20)	(2.60)			
HGIIv2								1.608049	1.415981	1.031935
								(2.25)	(1.40)	(0.24)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	23577	9262	9262	9262	10013	10013	10013	17478	17478	17478
Pseudo R-sq2	0.0993	0.1409	0.1509	0.1593	0.1284	0.1402	0.1492	0.1361	0.1407	0.1600
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-statistics in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Conditional on $GII < 2$, $GIIv1 < 1$ and $GIIv2 < 1$ respectively.
Controls: country, income decile within country, marital status, education, age, number of children, survey wave and religion.

Table 17: Women in the Labor Force in High Gendered Environments, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.260174	1.306224	1.16249	1.150718	1.236598	1.127158	1.114106	1.336899	1.185064	1.156769
	(4.80)	(4.18)	(4.12)	(4.44)	(3.89)	(3.40)	(3.58)	(4.38)	(4.40)	(4.37)
ChildrenValue	0.8670618	0.8280415	0.9052165	0.9621345	0.8478305	0.9152039	0.977711	0.8441516	0.9452552	0.9770677
	(-2.92)	(-2.65)	(-1.70)	(-0.58)	(-2.52)	(-1.54)	(-0.36)	(-2.80)	(-0.94)	(-0.35)
HGII		0.8250906	0.8446493	0.8610057						
		(-2.36)	(-2.54)	(-1.84)						
HGIIv1					0.7995639	0.8358389	0.9047729			
					(-3.23)	(-2.82)	(-1.34)			
HGIIv2								0.7570201	0.7787856	0.8392087
								(-4.98)	(-4.69)	(-2.26)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	31293	15308	14267	14267	18415	17343	17343	16270	14718	14718
Pseudo R-sq2	0.0852	0.0921	0.1221	0.1212	0.0994	0.1201	0.1289	0.0929	0.1145	0.1138
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-statistics in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Conditional on $GII > 2$, $GIIv1 > 1$ and $GIIv2 > 1$ respectively.
Controls: country, income decile within country, marital status, education, age, number of children, survey wave and religion.

Table 18: Women speaking High Gendered Home Language in Labor Force, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.270951	1.306765	1.196291	1.158692	1.269812	1.173672	1.133932	1.321883	1.188368	1.152701
	(5.11)	(5.86)	(5.70)	(4.67)	(5.13)	(4.80)	(3.92)	(6.43)	(6.59)	(5.16)
ChildrenValue	0.8272403	0.8357664	0.9447441	0.9842292	0.8657762	0.9516412	0.9865435	0.7780201	0.9360562	0.9639336
	(-2.47)	(-3.19)	(-1.08)	(-0.32)	(-2.74)	(-0.96)	(-0.28)	(-3.21)	(-1.27)	(-0.84)
GII		0.7956224	0.7406077	0.8073701						
		(-3.07)	(-4.94)	(-3.89)						
GIIv1					0.6987634	0.6616277	0.802278			
					(-2.79)	(-3.86)	(-2.69)			
GIIv2								0.7926059	0.7004257	0.7752223
								(-2.68)	(-5.39)	(-3.36)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	32908	20769	20769	20769	22117	22117	22117	23439	23439	23439
Pseudo R-sq2	0.0774	0.0900	0.1135	0.1213	0.0907	0.1092	0.1220	0.0837	0.1132	0.1194
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-statistics in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Conditional on $HGII > 2$, $HGIIv1 > 1$ and $HGIIv2 > 1$ respectively.
Controls: country, income decile within country, marital status, education, age, number of children, survey wave and religion.

Table 19: Women speaking Low Gendered Home Language in the Labor Force, World Value Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MenSuperior	1.395665	1.362494	1.314149	1.252862	1.318814	1.288122	1.224071	1.293776	1.259109	1.177099
	(5.45)	(5.84)	(5.72)	(6.69)	(6.73)	(7.07)	(10.88)	(6.70)	(9.07)	(6.63)
ChildrenValue	0.6476533	0.6124365	0.7938771	0.8343825	0.7548569	0.7304637	0.7915402	0.8174556	0.8184465	0.8126452
	(-2.46)	(-3.25)	(-1.77)	(-1.76)	(-1.61)	(-1.71)	(-1.50)	(-1.55)	(-2.24)	(-1.99)
GII		1.328429	1.02764	1.048927						
		(4.89)	(0.44)	(0.84)						
GIIv1					1.28385	0.8189997	0.9965621			
					(1.56)	(-3.67)	(-0.05)			
GIIv2								1.231013	0.9456409	0.8674714
								(1.75)	(-0.53)	(-2.48)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	10954	9755	9755	9755	6193	6191	6193	8764	8761	8764
Pseudo R-sq2	0.0903	0.1041	0.1175	0.1262	0.1182	0.1276	0.1492	0.0929	0.1108	0.1271
Continent	no	no	yes	no	no	yes	no	no	yes	no
Region	no	no	no	yes	no	no	yes	no	no	yes

z-statistics in parentheses. Odd ratios of Logit regression are reported. Robust standard errors clustered at country level.
Conditional on $HGII < 2$, $HGIIv1 < 1$ and $HGIIv2 < 1$ respectively.
Controls: country, income decile within country, marital status, education, age, number of children, survey wave and religion.

6 References

Acemoglu, D., Johnson, S., and Robinson J. A. (2001). 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, vol. 91(5), pp. 1369-401.

Alesina, A., Giuliano, P. and Nunn, N. (2011). 'Fertility and the plough', *American Economic Review*, vol. 101(3), pp. 499-503.

Baslevent, C. and Onaran, O. (2004). 'The effect of export-oriented growth on female labor market outcomes in Turkey', *World Development*, vol. 32(8), pp. 1475-93.

Bloom, D. and Sachs, J. D. (1998). 'Geography, demography, and economic growth in Africa', *Brookings Papers on Economic Activity*, vol. 29(2), pp. 207-95.

Booth, A. L. and Nolen, P. (2012). 'Gender differences in risk behaviour: does nurture matter?', Economic Journal, vol. 122(558): F56-78.

Boroditsky, L. and Gaby, A. (2010). 'Remembrances of times East: Absolute spatial representations of time in an Australian aboriginal community', *Psychological Science*, vol. 21(11), pp. 1635-9.

Boroditsky, L., Schmidt, L., and Phillips, W. (2003). 'Sex, syntax, and semantics', in (D. Gentner and S. Goldin-Meadow (ed.), *Language in Mind: Advances in the Study of Language and Cognition*, MIT Press: Cambridge, MA.

Cavalli-Sforza, L. L. (2001). *Genes, Peoples, and Languages*, Berkeley : University of California Press.

Chen, M. K. (2012). 'The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement Assets', Under Review.

Chomsky, N. (1980). *Rules and Representations*, New York: Columbia University Press.

Christiansen, M. H. and Kirby, S. (2003). 'Language evolution: The hardest problem in science?', in (M.H. Christiansen and S. Kirby ed.), *Language Evolution: The States of the Art*, Oxford: Oxford University Press.

Corbett, G. G. (2011a). 'Number of genders', in (M. S. Dryer, and M. Haspelmath ed.), *The*

World Atlas of Language Structures Online, Munich: Max Planck Digital Library, chapter 30.

Corbett, G. G. (2011b). 'Sex-based and non-sex-based gender systems ', in (M. S. Dryer, and M. Haspelmath ed.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, chapter 31.

Corbett, G. G. (2011c). 'Systems of gender assignment ', in (M. S. Dryer, and M. Haspelmath ed.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, chapter 32.

Dryer, M. S. and Haspelmath, M. (ed.). (2011). *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library.

Easterly, W. and Levine, R. (1997). 'Africa's growth tragedy: Policies and ethnic Divisions', *The Quarterly Journal of Economics*, vol. 112(4), pp. 1203-50.

Evans, N. (2003). 'Context, culture, and structuration in the languages of Australia', *Annual Review of Anthropology*, vol. 32, pp. 14-40.

Evans, N. and Levinson, S. C. (2009). 'The myth of languages universals: Language diversity and its importance for cognitive science', *Behavioral and Brain Sciences*, vol. 32(5), pp. 429-92.

Falck, O., Heblich, S., Lameli, A. and Suedekum, J. (2010). 'Dialects, cultural identity, and economic exchange', IZA Discussion Papers No. 4743.

Fernández, R. (2010). 'Does Culture Matter?', NBER Working Papers No. 16277.

Gallup, J. L., Sachs, J. D. and Mellinger, A. D. (1998). 'Geography and economic development', NBER Working Paper No. 6849.

Goldin, C. (1995). 'The U-shaped female labour force function in economic development and economic history', in (T. P. Shultz ed.), *Investment in Womens Human Capital and Economic Development*, pp. 61-90, Chicago: University of Chicago Press.

Gorodnichenko, Y. and Roland, G. (2010). 'Culture, institutions and the wealth of nations', NBER Working Papers No. 16368.

Guiso, L., Sapienza, P. and Zingales, L. (2006). 'Does culture affect economic outcomes?', *Journal of Economic Perspectives*, vol. 20(2), pp. 23-48.

Guiso, L. Sapienza, P. and Zingales, L. (2008). 'Long term persistence', CEPR Discussion Paper No. 6981.

Guiso, L., Sapienza, P. and Zingales, L. (2009). 'Moral and social constraints to strategic default on mortgages', NBER Working Paper No. 15145.

Guiso, L., Sapienza, P. and Zingales, L. (2010). 'Civic capital as the missing link', NBER Working Papers No. 15845.

Hall, R. E. and Jones, C. I. (1999). 'Why do some countries produce so much more output per worker than others?', *The Quarterly Journal of Economics*, vol. 114(1), pp. 83-116.

Heston, A., Summers, R., and Aten, B. (2011). Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices, University of Pennsylvania.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*, Newbury Park, CA: Sage.

Hofstede, G. (1983). 'The cultural relativity of organizational practices and theories', *Journal of International Business Studies*, vol. 14(2), pp. 75-89.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (2004). *Leadership, Culture, and Organizations: The GLOBE Study of 62 societies*, Thousand Oaks, CA: Sage Publications.

Katz, M. L. and Shapiro, C. (1986). 'Technology adoption in the presence of network externalities', *Journal of Political Economy*, vol. 94(4), pp. 822-41.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R. W. (1998). 'Law and finance', *Journal of Political Economy*, vol. 106(6), pp. 1113-55.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R. W. (1999), 'The quality of government ', *Journal of Law, Economics and Organizations*, vol. 15(1), pp. 222-79.

Licht, A. N., Goldschmidt, C. and Schwartz, S. H. (2007). 'Culture rules: The foundations of the rule of law and other norms of governance', *Journal of Comparative Economics*, vol. 35(4), 659-88.

Marshall, M. G., Jaggers, K. and Gurr, T. D. (2011). *Polity IV Project: Political Regime Charac-*

teristics and Transitions, 1800-2010, Center for International Development and Conflict Management, University of Maryland.

Mauro, P. (1995). 'Corruption and growth', *The Quarterly Journal of Economics*, vol. 110(3), pp. 681-712.

Mavisakalyan, A. (2011). 'Gender in language and gender in employment ', ANUCBE School of Economics Working Paper.

North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance*, New York: Cambridge University Press.

North, D. C. (1993). 'Douglass C. North Prize Lecture'.

Ross, M. (2008). 'Oil, Islam, and women', *American Political Science Review*, vol. 102(1) , pp. 107-23.

Sampson, G., Gill, D. and Trudgill, P. (ed.) (2009). *Language complexity as an evolving variable.* Oxford: Oxford University Press.

Schwartz, S. H. (1994). 'Are there universal aspects in the structure and contents of human values?', *Journal of Social Issues*, vol. 50(4), pp. 19-45.

Schwartz, S. H. (1999). 'A theory of cultural values and some implications for work, *Applied Psychology*, vol. 48(1), pp. 23-47.

Siewierska, A. (2011). 'Gender distinctions in independent personal pronouns', in (M. S. Dryer, and M. Haspelmath ed.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, chapter 44.

Tabellini, G. (2008a). 'Presidential address institutions and culture', *Journal of the European Economic Association*, vol. 6(2-3), pp. 255-94.

Tabellini, G. (2008b). 'The scope of cooperation: Values and incentives', *The Quarterly Journal of Economics*, vol. 123(3), pp. 905-50.

Tang, L. and Koveos, P. (2008). 'A framework to update Hofstede's cultural value indices: Economic dynamics and institutional stability,' *Journal of International Business Studies*, vol. 39(6), 1045-63.

Wichmann, S. and Holman, E. W. (2009). 'Assessing temporal stability for linguistic typological features', Mnchen: LINCOM Europa.

7 Appendix

7.1 New Immigrant Survey Appendix

Available online at $https://sites.google.com/site/santacreuvasut/working-papers/language-and-economics$

7.2 World Value Survey Appendix

Table 20: Data description World Value Survey (WVS)

Variable	**Definition**	**WVS**
Dependent		
LFP	Labor force participation, dummy variable equal to 1 if in the labor force and 0 otherwise	Variable x028
Cultural Variables		
MenSuperior	Men should have more right to a job than women, dummy variable equal to 1 if disagree 0 otherwise	Variable c001
ChildrenValue	A woman has to have children to be fulfilled, dummy variable equal to 1 if agree 0 otherwise	Variable d019
MenSupExec	Men make better business executives than women do, dummy variable equal to 1 if agree 0 otherwise	Variable d078
Controls		
S003	Country	Variable s003
x047cs	Income decile within country	Variable x047cs
x007	Marital status. "Are you currently? 1.Married 2. Living together as married 3. Divorced 4. Separated 5. Widowed 6. Single 7. Divorced, Separated or Widowed. 8. Living apart but steady relation	Variable x007
x025r	Education level. Three modalieties 1.Lower 2.Middle 3.Upper	Variable x025r
x003	Age "This means that you are __ years old"	Variable x003
x011	Number of Children. " Have you had any children? If yes how many?"	Variable x011
S002	Survey wave, 3 waves(1994-1999, 1999-2004, 2005-2007) were considered	Variable s002
F025	Religion. "Do you belong to religious denomination? If yes to which one?"	Variable f025
Continent	Asia,Africa,America, Pacific. Europe was the omitted continent.	**Unesco**
Region	Arab States, Central-Eastern Europe, Central Asia, East Asia and the Pacific, Latin America and the Caribeans, South and West Asia, Sub Saharian Africa. Western Europe and North America were the omitted variables	**Unesco**

7.3 Individual Gender Dummy Variables

7.4 Tables Variations Gender Structure Across Languages

Table 21: *Country Dominant Language Gender Dummies Dataset*

Country	Dominant Language	Speakers (%)	Number Genders			Sex Based			Gender Assignment			Gender Pronouns		
			NG0	NG2	NG3	SB0	SBY	SBN	GA0	GAL	GAH	GP0	GPL	GPH
Afghanistan	Pashto	52	0	1	0	0	1	0	0	0	1	na	na	na
Albania	Albanian	98	0	1	0	0	1	0	na	na	na	0	1	0
Algeria	Arabic	72	0	1	0	0	1	0	0	0	1	0	0	1
Argentina	Spanish	97	0	1	0	0	1	0	0	0	1	0	0	1
Armenia	Armenian	93	1	0	0	1	0	0	1	0	0	1	0	0
Australia	English	81	0	0	1	0	1	0	0	1	0	0	1	0
Austria	German	92	0	0	1	0	1	0	0	0	1	0	1	0
Azerbaijan	Azerbaijani	89	1	0	0	1	0	0	na	na	na	1	0	0
Bahrain	Arabic	68	0	1	0	0	1	0	0	0	1	0	0	1
Bangladesh	Bengali	95	1	0	0	1	0	0	na	na	na	1	0	0
Belarus	Belorussian	66	0	0	1	0	1	0	na	na	na	0	1	0
Belgium	Dutch	59	0	0	1	0	1	0	na	na	na	0	1	0
Belize	English	51	0	0	1	0	1	0	0	1	0	0	1	0
Bermuda	English	100	0	0	1	0	1	0	0	1	0	0	1	0
Bolivia	Spanish	88	0	1	0	0	1	0	0	0	1	0	0	1
Bosnia and H.	Serbian-Croatian	92	0	0	1	0	1	0	na	na	na	0	1	0
Brazil	Portuguese	98	0	1	0	0	1	0	na	na	na	0	1	0
Brunei	Malay	80	1	0	0	1	0	0	na	na	na	na	na	na
Bulgaria	Bulgarian	83	0	0	1	0	1	0	na	na	na	0	1	0
Cambodia	Khmer	89	1	0	0	1	0	0	1	0	0	1	0	0
Canada	English	60	0	0	1	0	1	0	0	1	0	0	1	0
Chile	Spanish	90	0	1	0	0	1	0	0	0	1	0	0	1
China	Mandarin	72	1	0	0	1	0	0	1	0	0	0	1	0
Columbia	Spanish	99	0	1	0	0	1	0	0	0	1	0	0	1
Congo, Rep.	Kongo	51	0	0	1	0	0	1	0	0	1	0	1	0
Costa Rica	Spanish	97	0	1	0	0	1	0	0	0	1	0	0	1
Croatia	Serbian-Croatian	96	0	0	1	0	1	0	na	na	na	0	1	0
Cuba	Spanish	100	0	1	0	0	1	0	0	0	1	0	0	1
Cyprus	Greek	74	0	0	1	0	1	0	0	0	1	0	1	0
Czech Rep.	Czech	94	0	0	1	0	1	0	na	na	na	0	1	0
Denmark	Danish	95	0	1	0	0	0	1	na	na	na	0	1	0
Dominican Rep.	Spanish	98	0	1	0	0	1	0	0	0	1	0	0	1
Ecuador	Spanish	93	0	1	0	0	1	0	0	0	1	0	0	1
Egypt	Arabic	99	0	1	0	0	1	0	0	0	1	0	0	1
El Salvador	Spanish	100	0	1	0	0	1	0	0	0	1	0	0	1
Eritrea	Tigrinya	49	na	na	na	na	na	na	na	na	na	0	0	1
Estonia	Estonian	65	1	0	0	1	0	0	na	na	na	1	0	0
Ethiopia	Oromo (Harar)	30	0	1	0	0	1	0	0	0	1	0	1	0
Fiji	Hindi	81	0	1	0	0	1	0	0	0	1	1	0	0
Finland	Finnish	92	1	0	0	1	0	0	1	0	0	1	0	0
France	French	94	0	1	0	0	1	0	0	0	1	0	1	0
Georgia	Georgian	71	1	0	0	1	0	0	1	0	0	1	0	0
Germany	German	91	0	0	1	0	1	0	0	0	1	0	1	0
Greece	Greek	98	0	0	1	0	1	0	0	0	1	0	1	0
Guatemala	Spanish	65	0	1	0	0	1	0	0	0	1	0	0	1
Guinea	Fula	39	0	0	1	0	0	1	0	0	1	0	1	0
Honduras	Spanish	97	0	1	0	0	1	0	0	0	1	0	0	1
Hong Kong	Cantonese	89	1	0	0	1	0	0	1	0	0	na	na	na
Hungary	Hungarian	99	1	0	0	1	0	0	1	0	0	1	0	0
Iceland	Icelandic	96	0	0	1	0	1	0	0	0	1	na	na	na
India	Hindi	40	0	1	0	0	1	0	0	0	1	1	0	0
Indonesia	Javanese	39	1	0	0	1	0	0	na	na	na	na	na	na
Iran	Persian	46	1	0	0	1	0	0	1	0	0	0	1	0
Iraq	Arabic	77	0	1	0	0	1	0	0	0	1	0	0	1
Ireland	English	98	0	0	1	0	1	0	0	1	0	0	1	0
Israel	Hebrew	63	0	1	0	0	1	0	0	0	1	0	0	1
Italy	Italian	94	0	1	0	0	1	0	na	na	na	0	1	0
Japan	Japanese	99	1	0	0	1	0	0	na	na	na	0	1	0
Jordan	Arabic	98	0	1	0	0	1	0	0	0	1	0	0	1
Kazakhstan	Kazakh	46	1	0	0	1	0	0	na	na	na	1	0	0
Kiribati	Kiribati	99	na	na	na	na	na	na	na	na	na	1	0	0
Korea South	Korean	100	1	0	0	1	0	0	na	na	na	0	1	0
Kuwait	Arabic	78	0	1	0	0	1	0	0	0	1	0	0	1
Kyrgystan	Kirghiz	60	1	0	0	1	0	0	1	0	0	1	0	0
Laos	Lao	53	1	0	0	1	0	0	na	na	na	1	0	0
Latvia	Latvian	56	0	1	0	0	1	0	0	0	1	0	1	0
Lebanon	Arabic	91	0	1	0	0	1	0	0	0	1	0	0	1
Libya	Arabic	96	0	1	0	0	1	0	0	0	1	0	0	1
Lithuania	Lithuanian	84	0	1	0	0	1	0	na	na	na	0	1	0
Macau	Cantonese	86	1	0	0	1	0	0	1	0	0	na	na	na
Macedonia	Macedonian	67	0	0	1	0	1	0	na	na	na	0	1	0
Madagascar	Malagasy	99	1	0	0	1	0	0	1	0	0	1	0	0
Malawi	Chichewa	59	0	0	1	0	0	1	0	0	1	1	0	0
Malaysia	Malay	58	1	0	0	1	0	0	na	na	na	1	0	0
Mali	Bambara	32	na	na	na	na	na	na	na	na	na	1	0	0
Malta	Maltese	90	0	1	0	0	1	0	0	0	1	na	na	na
Mauritania	Arabic	82	0	1	0	0	1	0	0	0	1	0	0	1
Mexico	Spanish	92	0	1	0	0	1	0	0	0	1	0	0	1

Table 22: *Country Dominant Language Gender Dummies Dataset (continued)*

Country	Dominant Language	Speakers (%)	Number Genders			Sex Based			Gender Assignment			Gender Pronouns		
			NG0	NG2	NG3	SB0	SBY	SBN	GA0	GAL	GAH	GP0	GPL	GPH
Moldova	Romanian	62	0	0	1	0	1	0	na	na	na	0	1	0
Mongolia	Khalkha	84	1	0	0	1	0	0	1	0	0	1	0	0
Morocco	Arabic	65	0	1	0	0	1	0	0	0	1	0	0	1
Myanmar	Burmese	69	1	0	0	1	0	0	1	0	0	0	1	0
Namibia	Ndonga	60	na	na	na	na	na	na	na	na	na	0	1	0
Nepal	Nepali	50	na	na	na	na	na	na	na	na	na	1	0	0
Netherlands	Dutch	96	0	0	1	0	1	0	na	na	na	0	1	0
New Zealand	English	91	0	0	1	0	1	0	0	1	0	0	1	0
Nicaragua	Spanish	98	0	1	0	0	1	0	0	0	1	0	0	1
Niger	Hausa	53	0	1	0	0	1	0	0	0	1	0	0	1
Nigeria	Hausa	21	0	1	0	0	1	0	0	0	1	0	0	1
Norway	Norwegian	97	0	0	1	0	1	0	na	na	na	0	1	0
Oman	Arabic	77	0	1	0	0	1	0	0	0	1	0	0	1
Pakistan	Panjabi	48	0	1	0	0	1	0	0	0	1	1	0	0
Palau	Palauan	83	1	0	0	1	0	0	1	0	0	1	0	0
Panama	Spanish	77	0	1	0	0	1	0	0	0	1	0	0	1
Papua New Guinea	Papuan Lang.	78	0	0	1	0	1	0	0	1	0	na	na	na
Paraguay	Spanish	55	0	1	0	0	1	0	0	0	1	0	0	1
Peru	Spanish	80	0	1	0	0	1	0	0	0	1	0	0	1
Philippines	Tagalog	29	0	1	0	0	1	0	0	1	0	1	0	0
Poland	Polish	98	0	0	1	0	1	0	na	na	na	0	1	0
Portugal	Portuguese	99	0	1	0	0	1	0	na	na	na	na	na	na
Puerto Rico	Spanish	85	0	1	0	0	1	0	0	0	1	0	0	1
Qatar	Arabic	40	0	1	0	0	1	0	0	0	1	0	0	1
Romania	Romanian	89	0	0	1	0	1	0	na	na	na	0	1	0
Russia	Russian	81	0	0	1	0	1	0	0	0	1	0	1	0
Samoa	Samoan	99	na	na	na	na	na	na	na	na	na	1	0	0
Saudi Arabia	Arabic	95	0	1	0	0	1	0	0	0	1	0	0	1
Senegal	Wolof	48	na	na	na	na	na	na	na	na	na	1	0	0
Serbia	Serbian-Croatian	75	0	0	1	0	1	0	na	na	na	0	1	0
Singapore	Mandarin	77	1	0	0	1	0	0	1	0	0	0	1	0
Slovakia	Slovak	86	0	0	1	0	1	0	na	na	na	0	1	0
Slovenia	Slovene	88	0	0	1	0	1	0	na	na	na	0	1	0
Somalia	Somali	98	na	na	na	na	na	na	na	na	na	na	na	na
South Africa	Zulu	24	0	0	1	0	0	1	0	0	1	0	1	0
Spain	Spanish	74	0	1	0	0	1	0	0	0	1	0	0	1
Sudan, The	Arabic	49	0	1	0	0	1	0	0	0	1	0	0	1
Sweden	Swedish	90	0	1	0	0	0	1	na	na	na	0	1	0
Switzerland	German	64	0	0	1	0	1	0	0	0	1	0	1	0
Syria	Arabic	90	0	1	0	0	1	0	0	0	1	0	0	1
Taiwan	Xiamen	67	1	0	0	1	0	0	na	na	na	na	na	na
Tajikistan	Tajik	62	1	0	0	1	0	0	na	na	na	1	0	0
Tanzania	Nyamwezi-Sukuma	21	0	0	1	0	0	1	0	0	1	na	na	na
Thailand	Thai	53	1	0	0	1	0	0	1	0	0	1	0	0
Togo	Ewe	23	1	0	0	1	0	0	1	0	0	1	0	0
Tunisia	Arabic	99	0	1	0	0	1	0	0	0	1	0	0	1
Turkey	Turkish	88	1	0	0	1	0	0	1	0	0	1	0	0
Turkmenistan	Turkmen	77	1	0	0	1	0	0	na	na	na	1	0	0
Ukraine	Ukrainian	65	0	0	1	0	1	0	0	0	1	0	1	0
United Arab Emirates	Arabic	42	0	1	0	0	1	0	0	0	1	0	0	1
United Kingdom	English	97	0	0	1	0	1	0	0	1	0	0	1	0
United States	English	82	0	0	1	0	1	0	0	1	0	0	1	0
Uruguay	Spanish	97	0	1	0	0	1	0	0	0	1	0	0	1
Uzbekistan	Uzbek	76	1	0	0	1	0	0	1	0	0	1	0	0
Venezuela	Spanish	96	0	1	0	0	1	0	0	0	1	0	0	1
Vietnam	Vietnamese	87	1	0	0	1	0	0	1	0	0	1	0	0
Yemen	Arabic	100	0	1	0	0	1	0	0	0	1	0	0	1
Zimbabwe	Shona	72	0	0	1	0	0	1	0	0	1	1	0	0

na: Not available

Table 23: *Descriptive Statistics Dominant Language (per country)*

	N	N (dummy=1)	Mean	Std. Dev.
Number Genders	128			
NG0		33	0.26	0.44
NG2		57	0.45	0.50
NG3		38	0.30	0.46
Sex Based	128			
SB0		33	0.26	0.44
SBY		87	0.68	0.47
SBN		8	0.06	0.24
Gender Assignment	93			
GA0		20	0.22	0.41
GAL		10	0.11	0.31
GAH		63	0.68	0.47
Gender Pronouns	124			
GP0		33	0.27	0.44
GPL		49	0.40	0.49
GPH		42	0.34	0.48

Table 24: Linguistic variations across families (Individual indexes)

Family	N^C	N^L	NG0	NG2	NG3	SB0	SBY	SBN
Ind-Eu	67	34	0.06	0.48	0.45	0.06	0.91	0.03
Afr-Asi	23	5	0	1	0	0	1	0
Nig-Cong	10	10	0.14	0	0.86	0.14	0	0.86
Alt	7	7	1	0	0	1	0	0
Austr	7	7	0.80	0.20	0	0.80	0.20	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Family	N^C	N^L	GA0	GAL	GAH	GP0	GPL	GPH
Ind-Eu	67	34	0.05	0.17	0.79	0.11	0.59	0.30
Afr-Asi	23	5	0	0	1	0	0.05	0.95
Nig-Cong	10	10	0.14	0	0.86	0.56	0.44	0
Alt	7	7	1	0	0	1	0	0
Austr	7	7	0.67	0.33	0	1	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Table 25: Variation across families individual intensity indexes

Family	N^C	N^L	NGI	SBI	GAI	GPI
Indo-European	67	34	0.48	0.91	0.79	0.30
Afro-Asiatic	23	5	1	1	1	0.95
Niger-Congo	10	10	0	0	0.86	0
Altaic	7	7	0	0	0	0
Austronesian	7	7	0.20	0.20	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Table 26: Variation across families GII

Family	N	N	GII=0	GII=1	GII=2	GII=3	GII=4
Indo-European	67	34	0.05	0.18	0.18	0.13	0.48
Afro-Asiatic	23	5	0	0	0	0.05	0.95
Niger-Congo	10	10	0.17	0.83	0	0	0
Altaic	7	7	1	0	0	0	0
Austronesian	7	7	0.67	0	0.33	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Our sample contains 12 different families. We only show 5 families, gathering 91% of the countries in our sample

Table 27: Variation within the Indo-European family individual indexes

Genus	N^C	N^L	NG0	NG2	NG3	SB0	SBY	SBN
Romance	25	5	0	0.92	0.08	0	1	0
Germanic	16	7	0	0.13	0.88	0	0.88	0.13
Slavic	12	10	0	0	1	0	1	0
Indic	5	4	0.25	0.75	0	0.25	0.75	0
Iranian	3	3	0.67	0.33	0	0.67	0.33	0
Baltic	2	2	0	1	0	0	1	0
Greek	2	1	0	0	1	0	1	0
Albanian	1	1	0	1	0	0	1	0
Armenian	1	1	1	0	0	1	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Genus	N^C	N^L	GA0	GAL	GAH	GP0	GPL	GPH
Romance	25	5	0	0	1	0	0.21	0.79
Germanic	16	7	0	0.64	0.36	0	1	0
Slavic	12	10	0	0	1	0	1	0
Indic	5	4	0	0	1	1	0	0
Iranian	3	3	0.50	0	0.50	0.50	0.50	0
Baltic	2	2	0	0	1	0	1	0
Greek	2	1	0	0	1	0	1	0
Albanian	1	1	na	na	na	0	1	0
Armenian	1	1	1	0	0	1	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Table 28: Variation within the Indo-European family individual intensity indexes

Genus	N^C	N^L	NGI	SBI	GAI	GPI
Romance	25	5	0.92	1	1	0.79
Germanic	16	7	0.13	0.88	0.36	0
Slavic	12	10	0	1	1	0
Indic	5	4	0.75	0.75	1	0
Iranian	3	3	0.33	0.33	0.5	0
Baltic	2	2	1	1	1	0
Greek	2	1	0	1	1	0
Albanian	1	1	1	1	na	na
Armenian	1	1	0	0	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Table 29: Variation within the Indo-European family iGII

Genus	N^C	N^L	GII=0	GII=1	GII=2	GII=3	GII=4
Romance	25	5	0	0	0	0.05	0.95
Germanic	16	7	0	0.70	0	0.30	0
Slavic	12	10	0	0	1	0	0
Indic	5	4	0	0	0	1	0
Iranian	3	3	1	0	0	0	0
Baltic	2	2	0	0	0	1	0
Greek	2	1	0	0	1	0	0
Albanian	1	1	na	na	na	na	na
Armenian	1	1	1	0	0	0	0

N^C: the number of countries for which the dominant language belongs to the family

N^L: the number of different languages belonging to the family

Table 30: *Sources and Definition*

Variable	Definition	Source
Dependent		
Lab_part	Labor force participation rate, female 2000 (% of female population ages 15-64)	International Labour Organization
Lab_for	Labor force, female 2000 (% of total labor force)	International Labour Organization
Serv_fem	Employees, services, ratio of female to male 2000 (% of female employment/% of male employment)	International Labour Organization
Ind_fem	Employees, industry, ratio of female to male 2000 (% of female employment/% of male employment)	International Labour Organization
Agri_fem	Employees, agriculture, ratio of female to male 2000 (% of female employment/% of male employment)	International Labour Organization
Controls		
Dom_sh	Share of the countrys population that speaks the countrys dominant language	Encyclopedia Britannica Book of the Year (2010)
Log_inc	Logarithm of PPP Converted GDP Per Capita, G-K method, at current prices (in I$) 2000.	Heston et al. (2011)
Log_inc_sq	Logarithm of PPP Converted GDP Per Capita, G-K method, at current prices squared (in I$) 2000.	Heston et al. (2011)
Openness	Openness at current prices (Import plus Exports as a share of GDP, in %) 2000	Heston et al. (2011)
Gov_size	Government Consumption Share of PPP Converted GDP Per Capita at current prices, 2000 (%)	Heston et al. (2011)
Oil_rents	Total rents from oil and gas divided by midyear population in constant 2000 $.	Ross (2008)
Demo	Index of democracy from 0 to 10 (the closer to 10, the more democratic the country) 2000	Marshall et al. (2011)
Log_pop	Logarithm of the countrys population 2000	Heston et al. (2011)
Comm	Dummy equal to 1 if the country has been under a communist regime in the 20th century, 0 if not	Barro (2007)
Robustness Checks		
Dist_equ	Distance from equator	Hall and Jones (1999)
Trop_sh	Share of the population in tropical climate zones	Physical Geography Data, Portland University (2009)
Frost_days	Average number of frost days per unit of population	Physical Geography Data, Portland University (2009)
Coast_sh	Share of population within 100km of the coast or an ocean-navigable river	Physical Geography Data, Portland University (2009)
Landlocked	Countrys landlocked status (dummy equals to 1 if the country is landlocked, 0 if not)	CIA World Factbook (2011)
Colo	Colonization (dummy equal to 1 if the country has been colonized in the past, 0 if it hasnt)	LaPorta et al. (1999)
Gender and Economic Outcomes†		
Fem_seat	Proportion of seats held by women in national parliaments (%)	UN, Women's Indicators and Statistics database
Fem_minis	Proportion of women at a ministerial position(%)	UN Women's Indicators and Statistics database
Fem_loan	Women's access to bank loans (between 0=full and 1=impossible) in 2009	Gender, Institutions and Development Database 2009
Fem_land	Women's access to land ownership (between 0=full and 1=impossible) in 2009	Gender, Institutions and Development Database 2009
Fem_prop	Women's rights to own property other than land (between 0=full and 1=no) in 2009	Gender, Institutions and Development Database 2009
Fertil	Fertility rate, total (births per woman)	UN Population Division. World Population Prospects
Marr	Women who were first married by age 18 (% of women ages 20-24)	Demographic and Health Surveys by Macro Internation
Lit_rate	Literacy rate, adult female (% of females ages 15 and above)	UNESCO Institute for Statistics.
Inc_ineq	Income share held by highest 10%	World Bank, Development Research Group
Gini	GINI index	World Bank, Development Research Group.

We completed Marshall et al. (2011) data by givin to Commonwealth countries the same score as the UK.

†For gender and economic outcomes (except Fem_loan, Fem_land and Fem_prop) we took year 2000 when possible and if not the closest year between 1997 and 2003.