BERKELEY
ECONOMIC
HISTORY
LABORATORY

BEHL WORKING PAPER SERIES

WP2013-03

The Grammatical Origins of Gender Roles

Victor Gay
Estefania Santacreu-Vasut
Amir Shoham

22 April 2013

Berkeley Economic History Laboratory (BEHL)
University of California, Berkeley
530 Evans Hall / MC 3880
Berkeley, CA 94602
http://behl.berkeley.edu

Berkeley Economic History Laboratory, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720

BEHL Working Paper WP2013-03 | 22 April 2013
http://behl.berkeley.edu/wp/2013-03

The Grammatical Origins of Gender Roles

Victor Gay
Estefania Santacreu-Vasut
Amir Shoham

Abstract
We investigate the relation between gender marking in grammar and female participation in the labor market, the credit market, land ownership, and politics. Cross-country and individual-level analyses reveal that women speaking languages that more pervasively mark gender distinctions are less likely to participate in economic and political life and more likely to encounter barriers in their access to land and credit. These findings are robust to a large set of controls and robustness checks. We also found that the impact of a language's gender structure remains after controlling for culture, for historical agricultural use of the plough.

Keywords: culture, gender, grammar, labor, language, occupation, politics

JEL classification: J16, J82, 017, Z1.

Victor Gay Department of Economics, University of Chicago. victorgay@uchicago.edu

Estefania Santacreu-Vasut (corresponding author) Department of Economics at ESSEC Business School, Avenue Bernard Hirsch, 95021, Cergy-Pontoise, France; THEMA; and Visiting Professor at BEHL, University of California, Berkeley. +33 (0) 1 34 43 37 19 santacreuvasut@essec.edu

Amir Shoham FOX School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122-6083, USA; and College of Management Academic Studies, Israel, COMAS, 7 Yitzhak Rabin Blvd., RishonLeZion, 75190, Israel. amirls@colman.ac.il

The Grammatical Origins of Gender Roles

Victor Gay *
Estefania Santacreu-Vasut †
Amir Shoham ‡

Abstract

We investigate the relation between gender marking in grammar and female participation in the labor market, the credit market, land ownership, and politics. Cross-country and individual-level analyses reveal that women speaking languages that more pervasively mark gender distinctions are less likely to participate in economic and political life and more likely to encounter barriers in their access to land and credit. These findings are robust to a large set of controls and robustness checks. We also find that the impact of a language's gender structure remains after controlling for culture, for the historical agricultural use of the plough and for linguistic families.

Keywords: culture, gender, grammar, labor, language, occupation, politics

* University of Chicago, victorgay@uchicago.edu
† Corresponding author. Department of Economics at ESSEC Business School, France. Avenue Bernard Hirsch, 95021, Cergy-Pontoise, France, and THEMA. +33 (0) 1 34 43 37 19, santacreuvasut@essec.edu
‡ FOX School of Business, Temple University, and College of Management Academic Studies, Israel. Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122-6083, USA; COMAS, 7 Yitzhak Rabin Blvd., RishonLeZion, 75190, Israel. amir.shoham@temple.edu

We are grateful to participants of RES 2013, PAC-DEV 2013, NEUDC 2012, RES 2012 conferences, UC Berkeley, Ben Gurion University of the Negev, Paris School of Economics, College of Management Academic Studies Business School and THEMA seminars, and Temple University Brown Bag Seminar as well as to Pranab Bardhan, Greville Corbett, Barry Eichengreen, Ramon Ferrer-i-Cancho, Johanna Nichols, and Gerard Roland. We especially thank Alberto Alesina, Paola Giuliano, and Nathan Nunn for sharing their data on plough-based historical agriculture practices. We also thank Luc Aubry and Anatoli Colicev for excellent research assistance.

1 Introduction

Women's participation in markets and political institutions varies greatly across countries. As Sen (1999) argues, agency is a key determinant of their well-being. Furthermore, low levels of women's socioeconomic participation may hinder a country's economic development (World Bank Development Report 2012), decreasing the size of markets and leading to an inefficient allocation of resources. A growing body of literature (Fernández 2010; Alesina, Giuliano, and Nunn forthcoming) analyzes the role that attitudes and beliefs play in promoting or hampering women's agency: "attitudes of the family and of the society at large toward women's economic activities and the economic and social circumstances that encourage or resist change in these attitudes" (Sen 1999).

As North (1990, 1993) points out, language is an important component of those "social circumstances" and serves as a vehicle for cultural transmission. As early as 1935, Whorf claimed that language shapes thought: "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."

For linguists—in particular, those of the *generativist* persuasion—language is part of human biology; from this perspective (see Chomsky 1980), all languages are fundamentally equal in structure and have only minor local differences. In contrast, Sampson, Gill, and Trudgill (2009) identify a new group of linguists who view 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." The current *coevolutionary* theories look to the interplay between genetic and cultural forces when accounting 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 a direct channel through which language structure influences socioeconomic choices and outcomes. Christiansen and Kirby (2003) review the research on the origins and evolution of human language; they report some authors theorizing that gram-

matical structure reflects biological adaptation while others posit that language is transmitted culturally through millennia from one generation to the next.

This paper analyzes one feature of a language's grammar: its gender system. As observed by linguist Greville G. Corbett: "In some languages gender is evident in almost every phrase, while in other languages it is absent" (quoted in Dryer and Haspelmath 2011). Is there a relation between the intensity of gender marking in languages and the economic outcomes of different genders?

In the paper's first part, we identify empirical regularities between a language's gender marking and female participation in politics and in the labor, financial, and land markets. In so doing we employ both cross-country and individual-level analysis.

We find that women's participation in the labor and credit markets and in politics is significantly lower in countries whose dominant (i.e., most spoken) language marks gender more intensively. In these countries, the occupational profile of women is biased toward services and against jobs traditionally occupied by men. In addition, women in such countries are more likely to face restrictions in their access to land ownership and to loans.

Similarly, individual-level data reveal that women who speak a language that marks gender more intensively are less likely to supply labor or to work in agricultural occupations. These results are robust to systematically controlling for geography, climate, history of colonization, religion, and continent.

In the second part we discuss our results and offer empirical analysis to help understand them. Toward this end, we first review the interdisciplinary literature on language origin, evolution, and cognitive psychology. Why does grammatical structure vary across languages? In particular, why is there such wide variation in the extent to which languages mark gender distinctions? Where and how did languages—and their grammars—originate? Does language influence cognition and/or reflect culture? These questions are much debated in the fields of linguistics and cognitive science, among others. From those debates have emerged two important findings. First, gender is one of language's most stable grammatical features and persists for thousands of years (Wichmann and Holman 2009). Second, gender marking in languages may reflect, inter alia, its speakers'

worldviews (Corbett 1999).

It is not our aim to provide definitive answers to these questions, but we do exploit three sets of evidence to investigate them. First, we incorporate studies of US immigrants' labor supply behavior as a function of their native language; this allows us to compare individuals who share a common institutional environment but speak different languages. Second, we exploit information from the World Values Survey (WVS) which documents each country's dominant language as well as the language(s) that its citizens speak at home. Further, we use data from Alesina, Giuliano, and Nunn (forthcoming) on plough-based historical agricultural practices—together with WVS data on attitudes toward gender roles—to analyze whether the impact of language's gender marking can be transmitted via cultural capital or economic specialization dating to a distant past. Finally, we study whether the impact of languages grammar remains after we control for linguistic families.

Our first main result is that a grammar's gender marking differentially influences the set of all migrants and its subset of women living in countries whose dominant language is low gendered. In particular, women speaking a high-gendered language at home but living in a low-gendered language environment (e.g., Spanish-speaking US immigrants) are more likely to be employed than are women generally. This finding suggests that the influence of language is determined by an interaction between the cognitive framework of speakers and their cultural environment. Gender marking in languages may affect the gender-related social constraints faced by women as much as (or more than) it affects their cognitive framework. Our second main result is that gender treatment in a dominant language affects the economic outcomes of women who speak that language—even after we account for historical (agricultural) practices, cultural values and linguistic families. This result suggests the existence of other channels (including, perhaps, cognition) through which the effects of language are manifested.

So far, there has been only limited use of linguistic variables in economic research. Licht, Goldschmidt, and Schwartz (2007) use the grammar of pronouns as an instrumental variable to study how countries that favor autonomy, egalitarianism, and mastery exhibit less corruption,

more democratic accountability, and a more effective rule of law. Mavisakalyan (2011) uses the gender of pronouns to investigate the impact of culture on women's share in the labor force. Chen (2013) investigates the impact of languages's markings for future time on such future-oriented decisions and outcomes as savings, debt, and health-related behavior. Tabellini (2008a and 2008b) uses the grammar of pronouns as an instrumental variable to assess the causal effect of cultural values on institutional outcomes.[1]

2 Data on Gender in Language

Corbett argues that gender "is the most puzzling of the grammatical categories" (Corbett 1991). In common parlance, gender is linked to biological sex and refers to the categories of male and female; in linguistics, however, gender is linked to a set of usage rules that depend on nouns of different types.[2] These types are typically based on male–female distinctions but can also be based on other social constructs (including distinctions related to age, social status, human versus animal, etc.). In this section we describe all four gender-related grammatical features of language as classified by linguists in the World Atlas of Language Structure (WALS; Dryer and Haspelmath 2011).[3]

2.1 Individual Intensity Indices

A language's gender system may or may not be linked to biological sex. One example of a non–sex-based gender system is one based on the distinction between human and nonhuman (as in Danish) or between animate and inanimate. The Swedish language makes gender distinctions that are not sex based. These considerations lead us to building the Sex-Based Intensity Index (SBII), which is a dummy variable set equal to 1 for languages with a sex-based gender system

[1] More generally, Guiso, Sapienza, and Zingales (2009) document that trust between the respective inhabitants of two countries is affected by their geographical proximity and the degree of commonality in their languages.

[2] To avoid confusion, we shall use *gender* as in common parlance but use *grammatical* (or *language*) *gender* when referring to the broader concept of gender in linguistics.

[3] See chapter 30 of the WALS (Corbett 2011a), chapter 31 (Corbett 2011b), chapter 32 (Corbett 2011c), and chapter 44 (Siewierska 2011). From these grammatical features we create four individual indices and three aggregate indices, as detailed in the text.

(and to 0 otherwise).

The number of genders in a language tells us how many noun types may require different forms of agreement.[4] Some languages, such as Nigerian Fula, feature twenty genders. When a language (e.g., French, Spanish) has two genders, they are usually "feminine" and "masculine." Both English and German include "neuter" as a third gender, but other languages with multiple genders involve no sex-related distinctions. We therefore build our Number Gender Intensity Index (NGII), a dummy variable set equal to 1 for languages with exactly two genders and set equal to 0 otherwise (i.e., for languages with three or more genders or with no gender distinctions).

A gender assignment system provides a set of rules to help the speaker make appropriate agreements between nouns and the genders defined by the system. Assignment can depend on the meaning (semantic) or the form of the noun. For example, a semantic assignment system is used in Kannada (a language in southern India), where nouns denoting male (female) humans are masculine (feminine) and all remaining nouns are neuter. English also assigns gender on semantic grounds only. In contrast, some languages (including Russian and Spanish) have both a semantic and a formal assignment system. Thus, nouns that are neither masculine nor feminine need not be neuter; in Russian, for example, they can be assigned to the masculine or feminine gender as a function of their inflectional class (e.g., whether the noun takes the nominative or the accusative form). To capture these differences we build the Gender Assignment Intensity Index (GAII), a dummy variable set equal to 1 only for languages whose gender assignment system is both semantic and formal. This variable is otherwise set to 0 because gender is presumed to be more pervasive in the former case.

Finally, languages differ in the extent to which they distinguish gender in pronouns. For example, in English the pronominal sex-based gender system is determined by the use of she/he/it. Yet there are some languages that have a sex-based gender system but do *not* have sex-based pronouns. Hence we build the Gender Pronouns Intensity Index (GPII), a dummy variable set equal to 1 for languages with a gender distinction in third-person pronouns and also in the first and/or the second person; this variable is set equal to 0 otherwise—that is, if the language does

[4]Gender derives from the Latin *genus*, and it originally meant "kind" or 'type".

not distinguish gender in pronouns (or does so only in the third-person pronoun).

The four maps in Figure I show the distribution of intensity indices for each country's dominant language. For each index, the countries shaded in black are those for which the dominant (most spoken) language features the more sex-based gender treatment; in other words, the black-shaded countries are those for which that panel's dummy variable was set to 1.

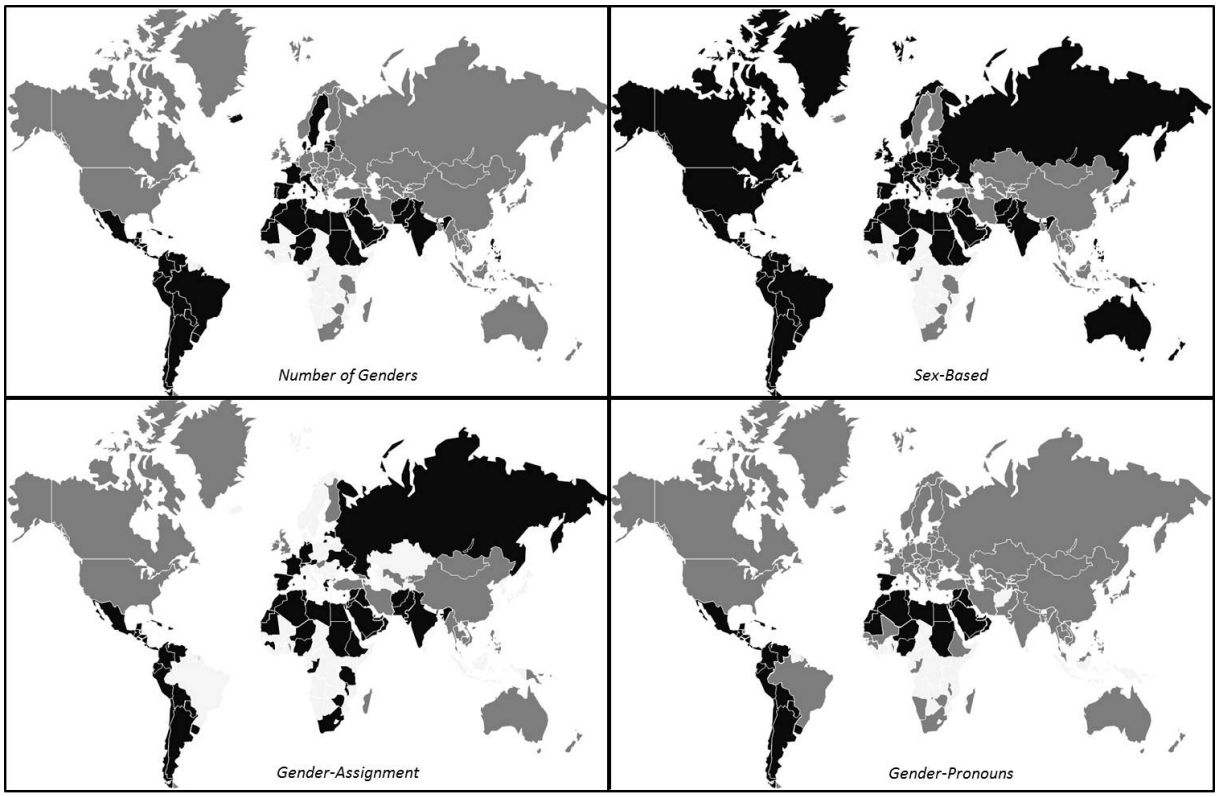

Figure I: Distribution of Individual Intensity Indices
Clockwise from upper left: NGII, SBII, GPII, and GAII; the dominant language of countries shaded in black (resp. gray) is characterized by gender treatment that is more (resp. less) sex based for that index.

As the figure indicates, languages vary in the degree to which they are gender intensive along the different indices; Table I shows the autocorrelation matrix of these four indices.[5] Table II lists the percentage of countries for which the dominant language simultaneously exhibits gender-intensive features in all possible pairings of the individual indices.

In the Appendix we present the data alphabetically (by country) in Tables XXIX–XXXII

[5] All tables in this paper observe the convention of denoting probability values as follows: $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

Table I: Autocorrelation of Intensity Indices

	NGII	SBII	GAII	GPII
NGII	1			
SBII	0.548***	1		
GAII	0.698***	0.595***	1	
GPII	0.793***	0.480***	0.646***	1

Table II: Joint Occurrences (%) of Value-1 Intensity Indices for Countries' Dominant Languages

	NGII = 1	SBII = 1	GAII = 1	GPII = 1
NGII = 1	1			
SBII = 1	42	1		
GAII = 1	50	59	1	
GPII = 1	34	34	45	1

(where "n/a" denotes "not available.")

It must be emphasized that linguistic structures vary widely across language families as well as within families. This fact is consistent with the evolution of language via 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" (Dryer and Haspelmath 2011). Throughout this study we use the four individual indices just described—in addition to the three aggregate ones described in the next section—because their respective samples do not perfectly overlap.

2.2 Aggregate Intensity Indices

The four gender-related variables described in Section 2.1 reflect different features of grammatical gender and so capture different aspects of the use intensity of male–female distinctions

in a given language. Yet there are languages whose gender system is sex-based but that either do not mark gender in pronouns or assign gender on semantic grounds only. To capture such variations, we construct an aggregate index for each language as the sum of its individual indices. Thus, our Gender Intensity Index (GII) is calculated as follows:

$$\text{GII} = \text{NGII} + \text{SBII} + \text{GAII} + \text{GPII}, \quad \text{GII} \in \{0, 1, 2, 3, 4\}.$$

For example, the GII for German is $1 + 1 + 0 + 0 = 2$. It has a sex-based gender system (SBII = 1) and assigns gender on the basis of both semantic and formal rules (GAII = 1); however, German assigns gender to third-person pronouns only (GPII = 0) and does have a neuter gender (NGII = 0).

We construct two additional aggregate indices as

$$\text{GIIV1} = \text{NGII} + \text{SBII} + \text{GAII};$$

$$\text{GIIV2} = \text{NGII} + \text{SBII} + \text{GPII}.$$

Thus GIIV1 is an aggregate index for which all information regarding its component indices was gathered by the same researcher (Corbett). The GIIV2 index excludes GAII from GII. Our motive for this exclusion is to overcome sample-size limitations of our GII index owing to the relatively high fraction of countries for which we lack information on the GAII grammatical variable.

Tables XXXIII–XXXV in the Appendix list our data sources for the socioeconomic, historical, and geographic variables used in the paper.

3 Empirical Strategy

There is general consensus among linguists that the gender system is one of a language's most stable features and is capable of surviving thousands of years (Wichmann and Holman 2009). In Section 5 we explore the linguistics literature that addresses the stability of grammatical features and that describes the emergence (and disappearance) of gender in language. Overall,

this literature strongly suggests that gender is a predetermined feature of grammar; therefore, we should not be concerned with economic outcomes determining gender systems (rather than vice versa).

Yet establishing the direction of causality does not rule out the possibility of omitted variable bias. Hence we systematically perform robustness checks across our empirical analysis that control for the influence of geography and of historical contact between societies and their languages. With regard to geography, we control for continent dummies and for distance from the equator; according to Hall and Jones (1999), the latter factor may reflect "Western influence." Following Bloom and Sachs (1998) and Gallup, Sachs, and Mellingeret (1998), who argue that climate may influence development, we also control for the country's portion that is in the tropics, the number of frost days, coastline as percentage of the border, and a dummy for being landlocked. As for historical factors, we follow Acemoglu, Johnson, and Robinson (2001) in controlling for colonization and colonial origin (English and Spanish) and follow Weber (1904/05) in controlling for religion.

Later in the paper, when investigating the potential mechanisms behind our results, we make three additional robustness checks. First, we use WVS data to control for direct measures of cultural beliefs concerning gender roles. Second, we control for the historical use of the plough, a measure constructed by Alesina, Giuliano, and Nunn (forthcoming). Third, we control for linguistic family origin.

Finally, because we may not be able to completely rule out an omitted variable bias in the cross-country results, we also analyze individual-level data (when information is available concerning what languages are spoken by individuals at home).

Specifications

For the country-level analysis we estimate the following equation(s) using OLS:

$$y_c = \beta(GenderIndex) + X_c\Gamma + \epsilon_c,$$

where y_c is an outcome of interest, c denotes country, *GenderIndex* is a placeholder for each of our intensity indices, X_c is a vector of covariates at the country level (including a constant),β is the coefficient of interest, and Γ is a vector of regression coefficients.

For individual-level analysis using the World Values Survey, we estimate a logit model of the underlying probability of a binary dependent variable, π_i as a linear function of the predictors:

$$logit(\pi_i) = \beta(GenderIndex) + X_i\Gamma;$$

here X_i is a vector of covariates at the individual level.

4 Results

4.1 Cross-Country Variations

This section investigates whether there is a statistically significant relation between variations in the intensity of gender marking in language and variations in the extent and nature of women's agency across countries where by *agency* we mean the ability of women to

> "... earn an independent income, to find employment outside the home, to have ownership rights and to have literacy and be educated participants in decision within and outside the family" (Sen 1999).

We shall present results of female participation in the labor market, the credit market, land ownership, and politics. Our goal here is to document the empirical findings in a systematic way. Their interpretation is deferred to Section 5, where we provide additional evidence to deepen the analysis.

4.1.1 Labor

This section presents the results related to female labor force participation and occupational choices relative to their male counterparts. All results reported are for the year 2000 to facilitate

comparison with Alesina, Giuliano, and Nunn (forthcoming). Results of similar magnitude and significance were obtained for the years in the period 1980–2009 and are available upon request.

We control for the share of inhabitants for whom the dominant language is their mother tongue to reflect the possibility that a dominant language is less representative when there is greater linguistic diversity. Furthermore, as shown by research on ethno-linguistic fractionalization, linguistic diversity plays an important role in economic outcomes (Mauro 1995; Easterly and Levine 1997). We follow Goldin (1995) in controlling for economic development (measured by the logarithm of GDP per capita and its square term). In addition, we control for trade openness because single-industry studies show that the share of female employment is higher in firms that export (Baslevent and Onaran 2004); we also control for government size. As Cavalcanti and Tavares (2011) demonstrate, increases in government size and female labor force participation are related. Finally, we control for a communist past (Paxton, Hughes and Painter, 2010).

Table III presents cross-country OLS regression results regarding the female labor force participation rate. This table shows that countries whose dominant language has two genders have a female participation rate that is 16.6 percentage points (pp) lower than countries whose dominant language has (no or) more than two genders. Similarly, column (2) shows that having a sex-based gender system decreases the female labor force participation rate by 13 pp % relative to the baseline value in countries with no gender system. This result, too, is significant both statistically and economically. Columns (4) (resp., and (5)) shows that, when the dominant language's having gender assignment that is both semantic and formal, (resp.,or gender marking applies to a large set of pronouns), female labor participation also decreases significantly.

Our results are virtually identical when using labor force's female share, a measure that controls for overall size of the labor force.

Table III: Female Labor Force Participation Rate: OLS Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		−0.166***	−0.134***	−0.141***	−0.150***	−0.0513***	−0.0771***	−0.0648***
		(0.0282)	(0.0309)	(0.035)	(0.0264)	(0.0111)	(0.016)	(0.0118)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	119	119	87	117	82	87	112
R^2	0.17	0.385	0.281	0.265	0.413	0.439	0.396	0.445
Adjusted R^2	0.141	0.346	0.236	0.199	0.375	0.386	0.342	0.408
F-statistic	6.323	10.36	6.59	5.686	10.09	9.302	8.448	10.7

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. Female Labor Force Participation Rate is the percentage of women in the labor force in 2000, which ranges from 13% to 91% with a mean of 0.509 and a standard deviation of 0.156; these values correspond to the sample from the baseline regression column (1). The controls include log income and log income squared, trade openness, communist history, government size, and share of speakers of the dominant language measured (year 2000).

Table IV reports the set of robustness checks as detailed in the previous section for female labor force participation rate. Owing to space constraints, throughout the paper we report robustness checks only for GIIV2. We chose this gender marking index because it is the aggregate index with the largest sample size. Results for all of the other indices are similar; see the Appendix.[6]

Table IV: Female Labor Force Participation Rate: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	−0.0697***	−0.0690***	−0.0816***	−0.0799***	−0.0600***
	(0.0123)	(0.0117)	(0.0124)	(0.0107)	(0.0121)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	112	112	112	112	112
R^2	0.46	0.565	0.532	0.597	0.574
Adjusted R^2	0.418	0.517	0.491	0.548	0.522
F-statistic	10.1	13.02	11.57	12.88	.

Notes: Geography represents distance from the equator, measured as $\sqrt{\text{latitude}}/90$. Climate includes four variables: share of the population in tropical areas, average number of frost days per unit of population, a dummy for whether the country is landlocked, and the share of population within 100 km of the coast or an ocean-navigable river. Colonization includes two dummies (set equal to 1 only if the country ever was, respectively, a British or Spanish colony). Continent and Religion are a set of dummy variables indicating the country's continent and majoritarian religion (Protestantism, Catholicism, Islamic, Jewish, Hinduism, Buddhist).

The significance and magnitude of the coefficients for these gender indices coefficients are virtually unchanged when we control for our set of robustness checks. In other words, the results are very robust to accounting for a variety of geographical and historical forces that may influence both language and economic outcomes. Each of our seven intensity indices, (as reported in the Appendix), remains significant at the 1% level for our five robustness checks.

Table V reports cross-country OLS regression results for the ratio of female to male employment in the services sector. Table VI gives results of our robustness checks.

Countries with higher levels of pronoun gender marking exhibit, on average, a ratio of female to

[6] Available upon request

Table V: Gender Ratio in Services Employment: OLS Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		0.185***	0.192***	0.0988	0.234***	0.0659**	0.0718*	0.0996***
		(0.0648)	(0.0718)	(0.0908)	(0.0806)	(0.0281)	(0.0372)	(0.0317)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	131	96	96	69	96	66	69	91
R^2	0.172	0.292	0.292	0.195	0.303	0.235	0.226	0.315
Adjusted R^2	0.132	0.235	0.236	0.102	0.247	0.143	0.137	0.258
F-statistic	3.836	7.354	8.732	3.826	5.75	4.614	5.059	7.28

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. Gender Ratio in Services Employment is the ratio of gender employment female to male employment in the services sector which ranges from 0.21 to 2.53 with a mean of 1.29 and a standard deviation of 0.363. For controls, see Notes to Table III.

Table VI: Gender Ratio in Services Employment: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	0.0860**	0.0917***	0.0289	0.0456	0.0739**
	(0.033)	(0.0337)	(0.0326)	(0.0283)	(0.0357)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	91	91	91	91	91
R^2	0.333	0.514	0.505	0.556	0.537
Adjusted R^2	0.268	0.446	0.45	0.488	0.466
F-statistic	6.013	9.654	9.781	9.745	.

Notes: For definitions of the variables, see Notes to Table IV.

male employment in the services sector that is nearly one standard deviation greater. In general, all but one of our indices have a significant and positive impact on female presence in services.

The GIIV2 variable remains highly significant and of a similar order of magnitude for our robustness checks—with the exceptions of Continent and Colonization. As for Colonization, it is noteworthy that the index SBII remains significant at the 5% level; as for Continent, the NGII remains significant at the 10% level. The other indices mirror the results for GIIV2.

Table VII reports cross-country OLS regression results for the ratio of female to male employment in the agricultural sector and Table VIII gives results of our reports robustness checks.

Table VII: Gender Ratio in Agricultural Employment: OLS Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		−0.341***	−0.252**	−0.104	−0.352***	−0.109**	−0.121**	−0.171***
		(0.0946)	(0.107)	(0.135)	(0.104)	(0.0439)	(0.0579)	(0.0445)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138	94	94	67	96	65	67	90
R^2	0.261	0.279	0.234	0.185	0.301	0.252	0.235	0.314
Adjusted R^2	0.227	0.22	0.171	0.0879	0.245	0.16	0.144	0.256
F-statistic	15.79	14.44	9.614	4.764	15.1	7.62	6.804	14.63

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. Gender Ratio in Agricultural Employment is the ratio of female to male employment in the agricultural sector, which ranges from 0 to 2.354 with a mean of 0.717 and a standard deviation of 0.459. For controls, see Notes to Table III.

All but one of our indices have a significant and negative impact on female presence in agricultural occupations. For example, countries with a sex-based gender system have a significantly lower ratio (exceeding half a standard deviation in magnitude) of female to male agricultural employment. Except for Continent and Colonization, GIIV2 remains highly significant and of a similar order of magnitude. Note that, for Continent, the NGII remains significant at the 5% level. The other indices mirror the results for GIIV2.

Overall these results suggest that gender marking in language is robustly correlated to female labor market outcomes. Indeed, women living in countries whose language marks female–male distinctions more pervasively are significantly less likely to participate in the labor market— even

Table VIII: Gender Ratio in Agricultural Employment: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	−0.146***	−0.165***	−0.0576	−0.078	−0.148**
	(0.052)	(0.047)	(0.0536)	(0.0512)	(0.057)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	90	90	90	90	90
R^2	0.347	0.48	0.531	0.558	0.468
Adjusted R^2	0.283	0.406	0.478	0.489	0.385
F-statistic	11.17	11.05	16.42	10.46	.

Notes: For definitions of the variables, see Notes to Table IV.

when we take into account historical, geographic, and religious factors. Furthermore, women are less (resp. more) likely, than men to be employed in agricultural (resp. service) occupations.[7]

4.1.2 Credit

In this section we study credit market supply- and demand-side behavior of women (relative to men) as well as the institutionally related features of female participation in financial markets. In particular, we study savings behavior, access to loans and the holding of accounts in formal financial institutions. We report results for the year 2011 because this is the only year for which the World Bank Gender Statistics provide data on credit market outcomes by gender. All the dependent variables in this section are taken from this data set.

In each of our regressions we control for "the strength of legal rights" index that ranges from 0 to 10— with higher scores indicative of laws that are better designed to expand credit access. We also control for legal origin and government size. As Hallward-Driemeier (2011) emphasizes, institutional factors play an important role in the gender gap evident in credit access. Moreover, La Porta et al. (1998) show that investor protection and ownership concentration is related to the

[7]We do not report the results for our Gender Ratio Employment Industry variable because they are not significant; however, these results are available upon request.

origins of a country's legal system. We also control for income, given that Rajan and Zingales (1998) have demonstrated a strong link between economic and financial development, and also for Protestantism (following Weber [1904/05] and, more recently, Chen [2013]). Finally, we control for the share of inhabitants for whom the dominant language is their mother tongue in order to capture the representativeness of that language.

Duflo (2012) discusses the existing literature that addresses how intra intrahousehold bargaining power affects women's savings behavior and use of financial accounts. At the macroeconomic level we can capture that power with two related variables: the share of women who were first married by age 18 (as a percentage of women aged 20–24) and the share of female-headed households (the percentage of households with a female head)—both of which are available from the World Bank Gender database. To increase sample size, we build two variables by inputting data for a window of ± 6 years around the year 2000 for married by age of 18 and similarly one of ± 2 years for female headed households.[8] Using this procedure yields 95 and 71 observations for the respective variables, which reduces to 53 and 46 observations when combined with our gender intensity indices. The small sample size prevents us from including these variables as controls in the baseline specification. Nonetheless, we present and discuss their correlation with our outcomes of interest and gender intensity indices following the main regression analysis. Table IX presents results for the gender ratio of "percentage who saved any money in the past year."

[8]The year 2000 offers the largest sample size.

Table IX: Gender Ratio for Savings: OLS Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		−0.114***	−0.106***	−0.133***	−0.0818**	−0.0415***	−0.0577***	−0.0415***
		(0.0327)	(0.0356)	(0.0368)	(0.0371)	(0.0123)	(0.0157)	(0.0139)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138	105	105	74	104	72	74	101
R^2	0.131	0.263	0.24	0.273	0.218	0.289	0.306	0.253
Adjusted R^2	0.0767	0.194	0.168	0.17	0.143	0.186	0.209	0.179
F-statistic	14.39	5.699	12.55	.	12.32	.	.	11.44

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. Gender Ratio Savings is the female/male ratio of "percentage who saved any money in the past year" (age 15+), which ranges from 0.34 to 1.27 with a mean of 0.87 and a standard deviation of 0.15. The controls include log income, legal index, government size, legal origin, a dummy for Protestantism, and the share of speakers of the country's dominant language.

Column (4) of Table IX shows that countries with a dominant language that assigns gender on both semantic and formal grounds have a lower Gender Ratio Savings (by nearly a whole standard deviation). Similarly, column (5) shows that gender marking in a large set of pronouns decreases this ratio by more than half a standard deviation. Indeed, all seven individual indices are negative and significant both statistically and economically. In short, women living in countries whose dominant language marks gender more intensively are less likely than men to save. This finding is consistent with the observation that these women are less likely to have an independent source of revenue, given their low participation in the labor market; it also could result from unequal bargaining power within the household. Table X reports the set of robustness check results.

Table X: Gender Ratio for Savings: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	−0.0417***	−0.0447***	−0.0460***	−0.0407**	−0.0289**
	(0.0143)	(0.0156)	(0.0163)	(0.0163)	(0.0144)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	101	101	101	101	101
R^2	0.253	0.268	0.26	0.272	0.348
Adjusted R^2	0.17	0.158	0.168	0.153	0.25
F-statistic	10.51	8.952	9.243	10.63	.

Notes: For definitions of the variables, see Notes to Table IV.

The significance and magnitude of the coefficients for our gender indices coefficients are virtually unchanged when we control for this set of robustness checks. Except for SBII, which becomes marginally insignificant when we control for Continent, all other indices and robustness checks mirror the results for GIIV2—remaining strongly significant (at the 1% and 5% levels) and similar in order of magnitude.

Table XI reports the odds ratios from cross cross-country *logit* regressions of a variable that captures restrictions on women's access to loans from the World Bank Gender Statistics. In

particular, the dependent variable in this regression is a dummy set equal to 1 if women are either fully or partially restricted (and to 0 zero otherwise).

Table XI: Women's Restriction in Loan Access: Logit Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		11.31**	5.689**	6.722*	8.077**	1.573	1.925*	2.074**
		(11.05)	(4.981)	(6.566)	(8.51)	(0.456)	(0.758)	(0.75)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	95	66	66	54	66	52	54	63
Pseudo-R^2	0.389	0.36	0.323	0.343	0.349	0.301	0.335	0.323
χ^2	33.07	18.79	20.43	18.68	20.55	17.98	19.99	18.76

Notes: The odds ratios from logit regressions are reported with robust standard errors given in parentheses; the unit of observation is a country. Restriction in Access to Loans is an indicator variable set equal to 1 only if women are either partially or fully restricted in their access to loans. For controls see Notes to Table IX.

The odds ratios reported in Table XI all exceed 1 which means that women in countries whose dominant language marks gender more intensively are much more likely to face some restriction in their access to credit. These results are significant not only economically but also (except for the aggregate GII) statistically. Table XII presents results for the set of robustness checks.

Table XII: Women's Restriction in Loan Access: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	2.025*	2.644**	2.530**	2.542**	1.691
	(0.758)	(1.106)	(0.953)	(1.128)	(0.604)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	63	63	63	63	58
Pseudo-R^2	0.328	0.486	0.391	0.397	0.311
χ^2	17.88	22.98	24.85	314	19.3

Notes: For definitions of the variables, see Notes to Table IV.

As Table XII reports, the coefficient for GIIV2 remains greater than 1 one and significant

throughout (except when we control for religion). The rest of our indices are also robust to the inclusion of controls for climate, geography and continent but they lose significance (except for NGII) when we control for religion.

Next we analyze an institutional feature of female participation in the credit: the holding of an account at a financial institution, as reported in the World Bank Gender Statistics. Table XIII presents cross-country OLS regressions for the female/male ratio of "percentage holding an account at a formal financial institution".

Table XIII: Gender Ratio for Financial Account: OLS Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		−0.146***	−0.112*	−0.218**	−0.164***	−0.0515**	−0.0718**	−0.0587***
		(0.0501)	(0.0598)	(0.0839)	(0.052)	(0.021)	(0.0274)	(0.0198)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138	105	105	74	104	72	72	101
R^2	0.241	0.332	0.3	0.351	0.354	0.326	0.326	0.345
Adjusted R^2	0.194	0.269	0.234	0.26	0.292	0.228	0.228	0.28
F-statistic	13.06	7.208	11.15	8.925	.	.	.	8.797

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. The dependent variable is the female/male to male ratio of "percentage holding an account at a formal financial institution" which ranges from 0.169 to 0.177 with a mean of 0.83 and a standard deviation of 0.22. For controls see Notes to Table IX.

As column (4) of the table shows, the gender ratio for a formal financial account is one standard deviation smaller in magnitude for countries whose dominant language assigns gender on both semantic and formal grounds. Similarly, columns (6) and (7) show that our aggregate intensity indices have a large and very significant impact. Indeed, all seven individual indices are negative and are both statistically and in economically significant. Women living in countries whose language marks gender more intensively are less likely to hold accounts in the formal financial industry, a result that is consistent with these women being less likely to save and obtain loans.

Table XIV reports robustness check results with GIIV2 for the gender ratio of financial account holding. As the table shows, GIIV2 remains strongly significant at the 1% level for all robustness

Table XIV: Gender Ratio for Financial Account: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	−0.0568***	−0.0480**	−0.0790***	−0.0678***	−0.0601***
	(0.0202)	(0.0211)	(0.0225)	(0.0244)	(0.0211)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	101	101	101	101	101
R^2	0.346	0.425	0.377	0.429	0.432
Adjusted R^2	0.273	0.339	0.3	0.336	0.347
F-statistic	7.968	7.363	8.994	9.187	.

Notes: For definitions of the variables, see Notes to Table IV.

checks except for Climate (it is significant at the 5% level).

Overall, then, gender marking bears a strong relation to female participation in the credit market, as indicated by their savings rates, access to loans, and holding of financial accounts. Moreover, our results are robust to a set of geography, climate, religion, continent, and colonization controls.

Table XV reports the correlation matrix for the credit market outcomes studied in this section and two measures of household bargaining power: marriage by age 18 and female-headed households.

Table XV: Credit Market Outcomes and Household Bargaining Power: Pairwise Correlations

	Savings	Financial Account	Loan Restrictions	Married by 18	Female Head
Savings	1				
Financial Account	0.5367***	1			
Loan restrictions	−0.1743*	−0.2681***	1		
Married by 18	−0.0955	−0.1817	0.4660***	1	
Female Head	0.2592**	0.3891***	−0.2300*	−0.3203***	1

As the literature has established, in countries with a higher proportion of women married by age of 18 women are less likely to participate in the credit market. In contrast, there is more women's participation in countries with a higher proportion of female-headed households.

Table XVI: Household Bargaining Power and Gender Intensity Indices: Pairwise Correlations

	Married by 18	Female Head
Married by 18	1	
Female Head	−0.3203***	1
NGII	0.1847	−0.2460*
SBII	0.0118	−0.0319
GAII	0.2748*	0.0315
GPII	0.0577	−0.1143
GII	0.1279	−0.1489
GIIV1	0.1562	−0.1522
GIIV2	0.1236	−0.1762

Table XVI reports the correlation between the household bargaining variables and our gender intensity indices. Gender intensity in language is positively correlated with marriage by age 18 and negatively correlated with the share of female-headed households. This result suggests that one channel through which gender marking may influence female participation in the credit market is gender roles within the household.

4.1.3 Land

One cause of women's relative lack of access to credit is their lack of collateral assets. Although this deficiency can in turn be explained by having no independent source of revenue, (e.g.,as labor income), women also face restrictions in terms of property ownership. In this section, we investigate the relation between male–female distinctions in language and a measure of the restrictions faced by women seeking to own land. Such restrictions—in the form of inheritance laws, for instance—are indeed a barrier to female agency. For example, Dyson and Moore (1983) argue that inheritance laws have a negative effect on female autonomy in northern India.

Table XVII reports the odds ratios from cross-country logit regressions in which the dependent variable is a dummy set equal to 1 if women are either fully or partially restricted with respect to land ownership (and to 0 otherwise), as measured in the World Bank Gender Statistics. That the odds ratios in this table are much greater than 1 one implies that women in gender-intensive linguistic environments are more likely to face restrictions in land ownership. All of our gender

Table XVII: Womens Restriction in Access to Land:-Country Logit Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		10.23**	4.103	37.12***	8.090*	2.289**	2.874**	2.057*
		(11.18)	(3.586)	(51.93)	(8.896)	(0.821)	(1.363)	(0.782)
Observations	101	73	73	54	73	52	54	70
Pseudo-R^2	0.389	0.37	0.326	0.373	0.361	0.326	0.338	0.345
χ^2	37.09	25.08	23.76	18.92	21.56	21.63	22.11	23.89

Notes: The odds ratios from logit regressions are reported with robust standard errors given in parentheses. Restriction in Access to Land is an indicator variable set equal to 1 only if women are either partially or fully restricted in their access to land. For controls, see Notes to Table IX.

indices, (excepting only SBII), are significant.

Table XVIII: Womens Restriction in Access to Land: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
GIIV2	2.089*	2.266**	2.835**	2.481**	2.381**
	(0.786)	(0.839)	(1.254)	(1.072)	(0.983)
Controls	Yes	Yes	Yes	Yes	Yes
Geography	Yes				
Climate		Yes			
Colonization			Yes		
Continent				Yes	
Religion					Yes
Observations	70	70	70	70	69
Pseudo-R^2	0.345	0.421	0.378	0.419	0.362
χ^2	24.08	27.79	29.96	29.58	30.44

Notes: For the definitions of variables, see Notes to Table IV.

Table XVIII reports on our set of robustness check regressions for the GIIV2. This index is robust—statistically and in order of magnitude—to the inclusion of geography, climate, colonization, religion, and continent controls. Coefficients for the other indices (see the Appendix), remain fundamentally unchanged in magnitude and significance from the baseline regressions when we control for geography. They become even more significant when we control for climate and for colonization, with SBII also becoming significant in both cases. Furthermore, all seven indices are significant when controlling for continent and religion.

4.1.4 Politics

In this section we study countries adoption of legislated gender political quotas, from 1971 to 2011. Gender quotas are one of the main determinants of increased female political participation today, as Jutting et al. (2006) points out. Indeed, when we look at the direct impact of our gender intensity variables on the share of political positions held by women in 2000 we do not find statistically significant results, similar to Alesina Giuliano, and Nunn (forthcoming).

We use survival analysis to investigate the time it takes for a country to adopt gender political quotas during the period as function of the intensity of gender marking of their dominant language. The dependent variable is an indicator variable equal to one if a country adopts legislated quotas at time t. Following Kenworthy and Malami (1999), Dahlerup (2003) , Paxton et al.(2010), and Tripp and Kang (2008) we control for electoral family (3 dummies: majoritarian, combined, proportional), number of years since suffrage rights were granted to all women, communist past and income per capita. Table XIX presents odd ratios of a Cox's proportional Hazard Model for the period 1971 to 2011.

Table XIX: Gender Legislated Political Quotas: Cox Hazard Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		2.040*	1.615	3.073**	3.339***	1.767***	1.609**	1.631**
		(0.796)	(0.666)	(1.705)	(1.308)	(0.344)	(0.360)	(0.312)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,490	2,399	2,399	1,676	2,365	1,576	1,676	2,244
Pseudo-R^2	0.0201	0.0437	0.0379	0.0672	0.0569	0.0960	0.0737	0.0539
χ^2	7.196	12.60	10.40	12.31	16.03	20.61	15.67	13.48

Notes: The odds ratios are reported with robust standard errors given in parentheses. For controls, see text.

Since they are significant and bigger than one, it means that countries who mark gender more intensively in their language are more likely 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 repeated the analysis using 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, as table XX shows.

Table XX: Gender Legislated Political Quotas with Sanctions: Cox Hazard Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coeffcient		3.751***	6.942**	4.106	2.395	2.165*	2.422	2.042***
		(1.903)	(6.824)	(4.292)	(1.290)	(0.931)	(1.737)	(0.512)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,630	2,539	2,539	1,791	2,484	1,679	1,791	2,363
Pseudo-R^2	0.0530	0.0753	0.0779	0.0789	0.0584	0.120	0.105	0.0832
χ^2	7.597	17.42	9.640	9.330	10.49	13.43	10.99	16.50

Notes: See Notes to Table XIX.

The fact that countries where female political participation is lowest are the most likely and earlier adopters of quota policies is consistent with the increasing international demand from international organizations for female political empowerment. Indeed, 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. In countries where language gender intensity is high women are relatively deprived in their access to land, credit and labor markets, increasing the need of using regulatory instruments to promote female political participation. Table XXI reports robustness checks of our results.

Table XXI: Gender Legislated Political Quotas and Sanctions: Robustness Checks

		Quota				Sanctions				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GIIV2	1.630***	1.700***	1.598**	1.600*	1.250	1.979***	1.906***	1.848**	2.581***	1.516
	(0.308)	(0.339)	(0.323)	(0.393)	(0.244)	(0.492)	(0.429)	(0.508)	(0.877)	(0.407)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography	Yes					Yes				
Climate		Yes					Yes			
Colonization			Yes					Yes		
Continent				Yes					Yes	
Religion					Yes					Yes
Observations	2,244	2,236	2,244	2,244	2,244	2,363	2,355	2,363	2,363	2,363
Pseudo-R^2	0.0539	0.0713	0.0580	0.0810	0.103	0.0840	0.106	0.120	0.181	0.145
χ^2	13.82	18.80	20.49	14.72	675.7	17.67	25.09	9207	948.7	17749

Notes: For the definitions of robustness check variables, see Notes to Table IV. For description of controls see the main text.

4.2 Individual-Level Data

In this section we use individual-level data from the last three waves of the World Value Survey, which span the period 1994–2007. These data allows us to exploit information on the language that female respondents speak at home together with their labor force participation choice and their occupation in the agricultural sector. We do not use earlier waves because they do not include information regarding the language the respondent speaks at home. We control for age, age squared, primary and secondary education (via dummies), number of children, marital status (single, married, divorced, widowed) and survey wave fixed effects. We cluster standard errors at the country level and perform three robustness checks for individual religion, migrant father, and migrant mother.

Table XXII presents the odds ratios from logit regressions of female labor force participation (indicator variable set equal to 0 when the respondent is neither employed nor seeking employment). The odds ratios reported in this table are all less than 1 (albeit not all significant). In other words, women who speak a language that marks gender more intensively are less likely to supply labor.

Table XXII: Female Labor Participation and Household Language:-Level Logit Regressions
Gender Index

		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		0.542***	0.855	0.776	0.522***	0.861*	0.844	0.797**
		(0.114)	(0.225)	(0.194)	(0.115)	(0.0766)	(0.0988)	(0.0869)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	89,383	47,584	48,474	58,257	59,796	59,796	60,842	48,474
Pseudo-R^2	0.114	0.156	0.153	0.167	0.167	0.155	0.166	0.149
χ^2	542.7	417.5	433.7	525.8	556.4	558.4	526.7	444

Notes: The odds ratios from logit regression estimates are reported with robust standard errors (in parentheses) clustered at the country level. The dependent (indicator) variable is set equal to 1 when the respondent is employed full-time or part-time or is self-employed or unemployed (and to 0 otherwise). Indices for gendered grammar correspond to the language spoken at home. For controls, see the main text.

Table XXIII presents the odds ratios from logit regressions of female employment in agricultural occupations among working-age women in the labor force. As the table shows, women

speaking a gender-intensive language at home are less likely (conditional on being in the labor force) than other women to be employed in an agricultural occupation. That is, the odds ratios are significantly less than 1 for all seven gender-marking indices.

Table XXIII: Female Agricultural Employment and Household Language -Level Logit Regressions

		Gender Index						
		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient		0.279***	0.270***	0.302***	0.232***	0.641***	0.563***	0.549***
		(0.0936)	(0.102)	(0.123)	(0.0774)	(0.0857)	(0.0973)	(0.0908)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84,451	45,257	46,145	55,646	57,183	57,183	58,173	46,145
Pseudo-R^2	0.106	0.135	0.133	0.124	0.109	0.119	0.116	0.113
χ^2	450.4	242.7	239.4	336.8	343.8	348.4	359.6	239.5

Notes: The dependent (indicator) variable is set equals to 1 only for respondents employed in an agricultural occupation—conditional on being part of the labor force. See also Notes to Table XIX.

Table XXIV reports our robustness checks for the individual-level logit regressions of female labor force participation and female agricultural employment. Our first robustness check is religion. In particular, we control for a set of dummy variables set equal to 1 if the individual self-identifies as either a Buddhist, Catholic, Hindu, Jew, Muslim, Orthodox, or Protestant (and to 0zero otherwise). Our second robustness check captures the migrant status of the respondent's parents via a pair of dummy variables set equal to 1 if the father and mother are migrants (respectively) migrants. We have information on the migrant status of respondent families only for the last wave of the WVS.

The individual-level data results presented in this section closely track our previously given results at the country level. In sum: women speaking a highly gendered language at home are less likely, ceteris paribus, to be part of the labor force or (if they are in the labor force) to work in agricultural occupations.

Table XXIV: Individual-Level Data: Robustness Checks

	Labor		Agriculture	
	(1)	(2)	(3)	(4)
GIIV2	0.799***	0.739	0.555***	0.442***
	(0.0695)	(0.138)	(0.0766)	(0.12)
Controls	Yes	Yes	Yes	Yes
Religion	Yes	No	Yes	No
Migrant	No	Yes	No	Yes
Observations	48,474	16,091	55,646	16,091
Pseudo-R^2	0.185	0.156	0.149	0.188
χ^2	.—	307.7	.—	169.7

Notes: See the text for controls and definitions.

5 Discussion and Further Analysis

Two main questions emerge from these empirical findings. First, what exactly does the gender marking in a language's grammar reflect? Second, why do languages differ in their extent of gender marking? More specifically, we should like to know whether grammatical structure affects the speaker's cognitive framework and whether cultural values can be crystalized in grammar—as well as how languages originate and why male–female distinctions are pervasive in some languages yet entirely absent in others.

In this section we discuss interdisciplinary literature and provide additional empirical analysis related to these questions.

Wichmann and Holman (2009) construct a measure of stability for analyzing the linguistic features described in the World Atlas of Linguistic Structures, (Dryer and Haspelmath, 2011). They compare their findings with categorical statements in the literature regarding the stability of analyzed features and find a high degree of concordance. Wichmann and Holman 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). Gender is classified as an extremely stable feature of language. Indeed, "the term *gender* was first used in the 5th century BC by the Greek philosopher Protagoras, [who] . . . divided Greek nouns into three classes: 'feminine', 'masculine' and 'inanimate' (nowadays

called 'neuter')" (Aikhenvald 2004).

From an economic perspective, the stability of grammatical features is not surprising and could be related to how network externalities affect technology adoption (Katz and Shapiro, 1986). Language can be viewed as a technology characterized by networks externalities because the value of mastering a language increases with the number of its speakers. Linguistic evolution can thus be seen as a type of technological adaption. According to Katz and Shapiro "in the absence of sponsors, the technology [that is] superior today has a strategic advantage and is likely to dominate the market".[9] This dynamic is likely to obtain in the case of languages, which are not owned or sponsored (in this context, an owner/sponsor is"an entity that has property rights to the technology and hence is willing to make investments to promote it"; (Katz and Shapiro, [1986].

As for the origins of language and grammatical structure, Johansson (2005) reviews the interdisciplinary literature. From a purely evolutionary perspective, a feature of language will be selected (or not) on the basis of its utility to speakers. That being said, a feature's current function may differ from its original function. The advent of agriculture postdates language, which probably evolved among humans when they lived in modest-sized tribes of hunter-gatherers. (Daly and Wilson 1999; see also Nesse and Williams 1994 [cited in Johansson 2005]). This remark of Johansson (2005) is noteworthy: "the division of labor between hunting men and gathering women, with the organized cooperation and exchange of food that it entails, is invoked by Quiatt (2001) as an explanation for the emergence of language." (2005).

Other forces driving language evolution might include its conferring an advantage for reproductive success (Deacon, 1997). However, the most widely accepted explanation is the sociopolitical hypothesis, which Aiello (1998) extends by "arguing that social complexity drove the need for language, but also that social intelligence provided the cognitive structure for language." Factors proposed in the literature as exerting selective pressure for the emergence of language include hunting, tool making, sex, child care, teaching, and social relations.

[9] Tang and Koveos (2008) argue that changes in economic conditions are the source of cultural dynamics, whereas institutions (including language, religion, and the law) provide the foundation for cultural stability.

Since grammatical features often force us to encode certain aspects of reality to the exclusion of other aspects, it follows that applying an evolutionary perspective to grammar formation could well reveal selective pressures to codify the most relevant or salient aspects of our ancestors' reality, including culture and past economic specialization.

Under this hypothesis, we are led to ask: Why should economic specialization and/or culture of the distant past influence today's female participation in society? One possible answer is that culture is an extremely slow-moving institution, (cf. Roland 2004) and that a language's grammar serves as a marker of this persistent effect. An alternative (though not necessarily exclusive), hypothesis is that, irrespective of why heterogeneity exists in male–female distinctions across languages, such distinctions shape the gender consciousness of speakers as well as the perceptions of individuals belonging to a speaker's community.

Cognitive scientists currently study cross-linguistic differences that are related to time, colors, objects, and events. For example, Boroditsky, Schmidt, and Phillips*et al.* (2003) examine how grammatical gender influences the way speakers of different languages think about inanimate objects. As the authors point out such "questions touch on all the major controversies in the study of mind, with important implications for politics, law and religion. Yet very little empirical work had been done on these questions until recently."

Hence we provide three pieces of additional evidence that address aspects of the hypotheses just described. First, we investigate whether male–female distinctions in language have similar effects when we control for culture (via the World Values Survey). Second, we use the plough measure constructed by Alesina, Giuliano, and Nunn (forthcoming) to investigate whether gendered in language still matters after we control for economic specialization in the distant past; note that the current configuration of languages' gender marking was established prior to the 17th century which is the earliest period for which plough data has been collected. Third, we analyze the labor market behavior of a sample of migrants (with heterogeneous linguistic backgrounds) who are living in the United States as well as the behavior of a subsample of the WVS respondents whose household language differs from their country's dominant one.

5.1 Grammar and Culture

Table XXV presents the odds ratios from logit regression estimates in which the dependent (indicator) variable is set equal to 1 if the respondent is in the labor force (whether employed or unemployed) and set to 0 otherwise, from the World Value Survey last three waves.

We control for the baseline set of controls and for two variables that capture the respondent beliefs concerning gender roles as measured in the World Value Survey: Menpriority, (set to 1 only for respondents who believe that men should have priority when jobs are scarce) and Children-priority, (set equal to 1 only for respondents who believe that a woman needs children to fulfill her natural function).

As the table shows, our variables for linguistic gender marking linguistic variables remain highly significant and fundamentally unchanged. This could suggest that gender marking in grammar reflects more than cultural values.

Table XXV: Female Labor Participation and Culture: Individual-Level Logit Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GIIV2		0.797**		0.801**		0.797**		0.819**
		(0.0869)		(0.0827)		(0.0869)		(0.0823)
Menpriority			0.600***	0.591***				
			(0.064)	(0.0603)				
Childrenpriority							0.801**	0.816**
							(0.0757)	(0.0812)
Controls								
Observations	89,383	58,257	87,498	57,194	89,383	58,257	58,471	38,165
Pseudo-R^2	0.114	0.167	0.123	0.174	0.114	0.167	0.13	0.178
χ^2	542.7	525.8	548.8	556.8	542.7	525.8	460.9	620.2

Notes: See Notes to Table XIX.

5.2 Grammar and the Plough

Alesina, Giuliano, and Nunn (forthcoming) show that historical agricultural practices—in particular, the use of the plough —are negatively correlated with current cultural norms about female labor force participation and traditionally "female" occupations and female labour force participation today.

Table XXVI: Labor Market Outcomes and Historical Use of the Plough

	Female Labor Force Participation Rate				Gender Ratio Employment Agriculture				Gender Ratio Employment Services			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GIIV2		−0.0648***		−0.0667***		−0.171***		−0.141***		0.0996***		0.0796***
		(0.0118)		(0.0112)		(0.0445)		(0.0442)		(0.0317)		(0.0272)
Plough			−0.118***	−0.0924***			0.355***	0.485***			−0.312***	−0.394***
			(0.0301)	(0.0307)			(0.0981)	(0.132)			(0.0836)	(0.0904)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	112	172	111	138	90	135	89	131	91	128	90
R^2	0.17	0.445	0.26	0.498	0.261	0.314	0.355	0.451	0.172	0.315	0.295	0.505
Adjusted R^2	0.141	0.408	0.229	0.459	0.227	0.256	0.32	0.396	0.132	0.258	0.253	0.456

Notes: The OLS estimates are reported with robust standard errors given in parentheses;the unit of observation is a country. The controls include log income and log income squared, trade openness, communism, government size, and share of speakers of the dominant language measured (year 2000).

This finding supports the Boserup (1997) hypothesis that gender roles originated from an economically grounded division of labor across genders. Similarly, Bardhan (1974) suggests that the heterogeneous role of women (across regions in India) may be due to the type of cultivation (i.e., wheat-versus rice-growing regions).

Table XXVI presents OLS regressions of female labor market outcomes that control for use of the plough. We also present results of plough regressions in which our gender marking indices are excluded. As the table shows, the plough indeed has a strongly significant and negative impact on the rate of female labor force participation rate. In fact, these figures echo the results presented by Alesina, Giuliano, and Nunn (forthcoming). Even so controlling for the plough does not reduce the significance and magnitude of the coefficients of gender marking language variables.

Table XXVII: Credit Market Measures and Historical Use of the Plough

	Savings				Formal Account			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GIIV2		−0.0415***		−0.0388***		−0.0587***		−0.0629***
		(0.0139)		(0.0141)		(0.0198)		(0.0202)
Plough			0.0312	0.0394			−0.0307	−0.0618
			(0.0418)	(0.0453)			(0.0559)	(0.0596)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138	101	135	99	138	101	135	99
R^2-squared	0.131	0.253	0.137	0.26	0.241	0.345	0.24	0.353
Adjusted R^2	0.0767	0.179	0.0752	0.176	0.194	0.28	0.185	0.28

Notes: The OLS estimates are reported with robust standard errors given in parentheses; the unit of observation is a country. The controls include log income, legal index, government size, legal origin, the share of speakers of the dominant language, and a dummy for Protestantism.

Table XXVII presents cross-country OLS regressions for credit market–related variables when we control for use of the plough. As the table shows, in this case the significance and magnitude of the coefficients for our gender-marking variables are not reduced. Furthermore, it is interesting that the plough does not have a significant impact on women's credit market participation.

Table XXVIII presents cross-country OLS regressions for the variables capturing women's restrictions in loan and land access when we control for historical use of the plough. As the table shows, controlling for the use of the plough does not reduce the significance and magnitude of

Table XXVIII: Loan and Land Restrictions and Use of the Plough: Cross-Country OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GIIV2		2.074**		2.142**		2.057*		2.296**
		(0.75)		(0.777)		(0.782)		(0.886)
Plough			0.596	1.726			1.254	2.745
			(0.406)	(1.581)			(0.797)	(2.034)
Observations	95	63	93	62	101	70	99	69
Pseudo-R^2	0.389	0.323	0.382	0.321	0.389	0.345	0.376	0.354

Notes: The odds ratios from logit regressions are reported with robust standard errors in parentheses. For controls see Notes to Table XXIV .

the coefficients for our gender gender-marking language variables. Much as with its influence on credit market, the plough does not have a significantly affect restrictions on women's land ownership. These findings suggest that a fruitful direction for future research is to analyze the relation between historical agricultural practices and the gender systems of different languages' grammars.

5.3 US Immigrants

In this section we use the publicly available dataset version of the New Immigrant Survey (NIS) data set, a multi-cohort longitudinal study of new legal immigrants and their children to the United States based on nationally representative samples.[10] This data set allows us to study, for working-age (i.e., 18–65) female migrants, the effect of their languages' gender structure on their labor supply in the United States. We run OLS regressions with robust standard errors clustered at the country-of-origin level (using sample weights provided by the NIS). Our dependent variable is the number of hours worked annually in the United States in the respondent's "self-designated main (current) occupation." We control for age, age squared, years of schooling completed before migrating, schooling in the United States, a dummy for current US schooling, a dummy for married spouse present in the United States, hours worked annually before migrating, and number of children.

The results are given in Table XXIX. This table shows that, except for GAII, all gender

[10]For more information, please see $http://nis.princeton.edu/index.html$.

Table XXIX: US Working Hours of Female Migrants: OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Index		NGII	SBII	GAII	GPII	GII	GIIV1	GIIV2
Coefficient		183.8**	341.9***	109.1	169.7**	64.25**	102.8***	79.89**
		(84.34)	(72.78)	(104.1)	(82.46)	(25.07)	(33.80)	(32.19)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,990	2,823	2,823	2,536	2,829	2,536	2,820	2,536
R^2	0.063	0.072	0.081	0.073	0.072	0.078	0.077	0.077
Adjusted R^2	0.0605	0.0690	0.0784	0.0695	0.0686	0.0750	0.0742	0.0741
F-statistic	26.35	30.05	49.02	33.81	32.44	37.04	33.21	37.18

Notes: The OLS estimates are reported with robust standard errors (in parentheses) clustered at the country-of-origin of origin level. The unit of observation is an individual (working-age female migrants in the United States). For controls, see the main text.

intensity indices are positively and significantly correlated to the US labor supply of migrant women. In other words, female immigrants to the United States work more hours if the source country's language marks gender more frequently. Table XXVII presents robustness checks results for visa type, race/ethnicity, religion, and English mastery.

Table XXX: US Working Hours of Female Migrants: Robustness Checks

	(1)	(2)	(3)	(4)
GIIV2	93.37***	102.5***	101.9***	106.5***
	(23.46)	(36.55)	(36.71)	(25.68)
Controls	Yes	Yes	Yes	Yes
Visa	Yes	No	No	No
Ethnicity	No	Yes	No	No
Religion	No	No	Yes	No
English	No	No	No	Yes
Observations	2,820	2,610	2,820	2,750
R^2	0.117	0.082	0.085	0.107
Adjusted R^2	0.112	0.0769	0.0796	0.103
F-statistic	85.05	27.58	23.72	62.67

Notes: Visa includes dummies for the visa type (family, employment, etc.) Ethnicity includes dummies for ethnicity (Hispanic, Indian, White, etc.) Religion: includes dummies for the country's most prevalent religion (Catholic, Protestant, Islamic, etc.) English includes dummies indicating (four) levels of English mastery.

Why does a language's greater extent of gender marking lead to *lower* labor force participation among women (at both the country and individual level) for the general population but to *higher*

labor force participation among migrants living in the United States?

To understand the potential mechanisms behind this finding, we use the WVS and compare the labor market participation of women living in countries with low-gendered languages (GII $\leq$2) who speak languages with different gender grammatical structure at home. We focus on this subset of individuals to capture the situation of migrants arriving to the United States, whose dominant language (English) is a low gendered language. We analyze the impact of the immigrants' home language on their labor supply behavior and obtain results that are consistent with our previous findings.

In particular, Table XXX reports the odds ratios from logit regression estimates (with robust standard errors, in parentheses, clustered at the country level) for the subsample of female respondents living in countries whose dominant language is low gendered (GII $\leq$2) but who speak a language with different gender grammatical structure at home. The dependent (indicator) variables is set equal to 1 if the respondent is employed full-or part-time or is either self-employed or unemployed (and is set to 0 otherwise). The grammar language variables correspond to the language spoken at home.

Table XXXI: Women in Low-Gendered Dominant Language Countries: WVS

	(1)	(2)	(3)
Gender Index	GII	GIIV1	GIIV2
Coefficient	1.281***	1.096	1.361***
	(0.0639)	(0.0830)	(0.0792)
Observations	1,689	1,350	3,816
Pseudo-R^2	0.178	0.172	0.165
χ^2	306.5	226.1	692.3

Notes: See description and definitions in the main text.

The odds ratios here are greater than 1, which mirrors our previous results concerning migrants to the United States. That is, women who reside in a low-gendered language environment are more likely to be employed if their household language is a more gender-intensive one. It is not surprising to obtain different results for migrants given that they are a self-selected group and so will naturally differ from baseline groups. Furthermore, migrants may less attached to their

culture and even migrate to escape their country of origin cultural constraints.

5.4 Grammar and Linguistic Families

Table XXXII presents cross country OLS results for female labor force participation controlling for language families origin and grammatical gender marking indices. We also analyzed the regressions presented in the paper controlling for (1) Indo-European family origin, the most represented family in the sample, in addition to the baseline set of economic controls and gender grammatical indices, (2) clustering standard errors at the family level and (3) restricting the sample to countries with Indo-European languages. The later results are included in the appendix available upon request.

The results of these suggest that both language families origin and gender marking matter. Namely, both are very significant. Therefore, the impact of gender marking intensity on socio-economic outcomes comes both from across and within linguistic family variations in languages grammatical structure.

Table XXXII: Female Labor Force Participation and Linguistic Families

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Index		NGII	SBII	GPII	GAII	GII	GIIV1	GIIV2
Coefficient		-0.0759***	-0.0384	-0.0835***	-0.0478	-0.0310**	-0.0443**	-0.0404***
		(0.0224)	(0.0453)	(0.0313)	(0.0381)	(0.0141)	(0.0219)	(0.013)
Linguistic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162	119	119	118	87	83	87	113
R-squared	0.46	0.523	0.489	0.505	0.513	0.549	0.537	0.537
Adjusted R-squared	0.409	0.464	0.426	0.443	0.441	0.479	0.469	0.477

Notes: The OLS estimates are reported with robust standard errors (in parentheses). The unit of observation is a country. Linguistic controls include a set of dummy variables equal to 1 if language family is Afro-Asiatic, Niger-Congo, Altaic, Austronesian, Austro-Asiatic, Nilo-Saharan Sino-Tibetan, Uralic, Kartvelian, Macro-Ge, Japanese, Korean, Tai-Kadai or Indo-European. For the sample in Column (8), 63 of the countries have a Indo-European language, 22 Afro-Asiatic, 7 Altaic and 5 have a Niger-Congo language.

6 Conclusion

This paper presents novel empirical relations between language's grammatical gender markings and female economic and political participation. At both the country and the individual level, women who speak in a language—or who live in countries with a dominant language—that marks gender more intensively are less likely to participate in economic and political life and more likely to face formal and informal barriers when seeking to access credit and land. These findings are relevant to several long-debated questions about the origin of language, its evolution, and its effect on speakers.

To control for omitted variable bias, we perform a systematic and exhaustive set of robustness checks that aim to capture the influence of history, geography, religion, and colonization on women's economic outcomes. Our results are strongly robust to these controls.

So that we may further investigate and better understand the nature of our findings, we also analyze the impact of language grammar in a sample of migrants living in the United States, and therefore facing similar institutional environments. Their labor supply behavior suggests that the effect of an individual's own spoken language interacts with the language spoken by neighbors and other residents. We then show that language influences female labor force participation even after we control for current beliefs and values, for historical agricultural practices (i.e., use of the plough), as suggested by Alesina, Giuliano, and Nunn (forthcoming) and for linguistic families.

The direct and possibly cognitive influence of a language on its speakers and on economic life may have important policy implications. For instance, understanding this connection would facilitate the debate regarding the need to implement quotas or to opt for market forces to drive women economic participation up and thereby increase overall economic prosperity.

7 References

Aikhenvald, Alexandra Y . (2004) Gender and Noun Class. In Geert Booij, Christian Lehmann Joachim Mugdan (eds.), Morphology: a Handbook on Inflection and Word Formation volume 2, 1031-1045. Berlin: Mouton de Gruyter.

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.

Aiello, L. C. (1998). 'The foundations of human language'. In Jablonski, N. G. Aiello, L. C., (Eds), The origin and diversification of language. San Francisco: California Academy of Sciences

Aikhenvald, Alexandra Y . (2004) Gender and Noun Class. In Geert Booij, Christian Lehmann Joachim Mugdan (eds.), Morphology: a Handbook on Inflection and Word Formation volume 2, 1031-1045. Berlin: Mouton de Gruyter.

Alberto Alesina, A., Paola Giuliano, P., and Nathan Nunn, N. (forthcoming). "On the Origins of Gender Roles: Women and the Plough," Quarterly Journal of Economics, forthcoming.

Bardhan, Pranab. (1974) 'On Life and Death Questions' Economic and Political Weekly 9(32-34) pg 1293-1304.

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.

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.

Boserup, Ester. (1970) Womans Role in Economic Development, Lon- don: George Allen and

Unwin Ltd.

Cavalcanti T. V. De V. and Jos Tavares, 2011. 'Women Prefer Larger Governments: Growth, Structural Transformation, And Government Size,' Economic Inquiry, Western Economic Association International, vol. 49(1), pages 155-171, 01.

Chen, M. K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. Forthcoming: American Economic 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. (1991) Gender (Cambridge Textbooks in Linguistics). Cambridge: Cambridge 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.

Corbett, Greville C. (1991) Gender (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Daly, M., Wilson, M. I. (1999). Human evolutionary psychology and animal behaviour. Animal Behaviour, 57(3), 509-519.

Deacon, T.W. (1997). "What makes the human brain different?" Annual Review Anthropology 26: 337-57.

Doepke,Tertilt, & Voena 2011 The Economics and Politics of Womens Rights cited in T-XXXIII

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

Duflo, E. (2012) 'Womens Empowerment and Economic Development', Journal of Economic Literature, Vol. 50, No. 4: 1051-79

Dyson, Tim and Mick Moore (1983) On Kinship Structure, Female Autonomy, and Demographic Behavior in India Population and Development Review Vol.9, No. 1, pp 35-60.

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.

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.

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

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.

Hallward-Driemeier, Mary Rijkers, Bob Waxman, Andrew, (2011). 'Ladies first ? firm-level evidence on the labor impacts of the East Asian crisis,' Policy Research Working Paper Series 5789, The World Bank.

Johansson, Sverker (2005) 'Origins of language : constraints on hypotheses'. Converging evidence in language and communication research vol. 5, Amsterdam : Benjamins.

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.

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.

Nesse, R. M. Williams, G. C. (1994). Why we get sick. London: Vintage

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

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

Paxton, Pamela, Melanie Hughes, and Matthew Painter. (2010). 'Growth in Womens Political Representation: A Longitudinal Exploration of Democracy, Electoral System, and Gender Quotas.' European Journal of Political Research 49:25-52.

Quiatt, D. (2001). Language and evolution in the Middle Stone Age. In Gyori, G., (Ed), Language Evolution. Biological, linguistic and philosophical perspectives. Frankfurt am Main: Peter Lang

Rajan, R., Zingales, L., 1998. Financial dependence and growth. American Economic Review 88, 559586.

Roland, Gerard. 2004 Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions, Studies in Comparative International Development

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.

Sen, Amartya. 1999 Development as Freedom. Oxford: Oxford University Press.

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, Guido (2008a) The Scope of Cooperation: Values and Incentives, Quarterly Journal of Eocnomics 123(3): 905-950.

Tabellini, G. (2008b) 'Institutions and Culture. Presidential Address European Economic Association. Journal of the European Economic Association 6: 255-294.

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.

Weber, Max, Die protestantische Ethik und der Geist des Kapitalismus, Archiv fu r Sozialwissenschaften und Sozialpolitik, 1904/05, 20-21, 1784; 1110. English Translation: The Protestant Ethic and the Spirit of Capitalism, London: Allen Unwin, 1930.

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

World Bank Development Report 2012 cited in 1 of 1

8 Appendix

Table XXXIII: Intensity Indices for the Dominant Language by Country (A–F)

Country	Language	Speakers (%)	NGII	SBII	GAII	GPII
Afghanistan	Pashto	52	1	1	1	n/a
Albania	Albanian	98	1	1	n/a	0
Algeria	Arabic	72	1	1	1	1
Argentina	Spanish	97	1	1	1	1
Armenia	Armenian	93	0	0	0	0
Australia	English	81	0	1	0	0
Austria	German	92	0	1	1	0
Azerbaijan	Azerbaijani	89	0	0	n/a	0
Bahrain	Arabic	68	1	1	1	1
Bangladesh	Bengali	95	0	0	n/a	0
Belarus	Belorussian	66	0	1	n/a	0
Belgium	Dutch	59	0	1	n/a	0
Belize	English	51	0	1	0	0
Bermuda	English	100	0	1	0	0
Bolivia	Spanish	88	1	1	1	1
Bosnia and Herzegovina	Serbian-Croatian	92	0	1	n/a	0
Brazil	Portuguese	98	1	1	n/a	0
Brunei	Malay	80	0	0	n/a	n/a
Bulgaria	Bulgarian	83	0	1	n/a	0
Cambodia	Khmer	89	0	0	0	0
Canada	English	60	0	1	0	0
Chile	Spanish	90	1	1	1	1
China	Mandarin	72	0	0	0	0
Columbia	Spanish	99	1	1	1	1
Congo, Rep.	Kongo	51	0	0	1	0
Costa Rica	Spanish	97	1	1	1	1
Croatia	Serbian-Croatian	96	0	1	n/a	0
Cuba	Spanish	100	1	1	1	1
Cyprus	Greek	74	0	1	1	0
Czech Rep.	Czech	94	0	1	n/a	0
Denmark	Danish	95	1	0	n/a	0
Dominican Rep.	Spanish	98	1	1	1	1
Ecuador	Spanish	93	1	1	1	1
Egypt	Arabic	99	1	1	1	1
El Salvador	Spanish	100	1	1	1	1
Eritrea	Tigrinya	49	n/a	n/a	n/a	1
Estonia	Estonian	65	0	0	n/a	0
Ethiopia	Oromo (Harar)	30	1	1	1	0
Fiji	Hindi	81	1	1	1	0
Finland	Finnish	92	0	0	0	0
France	French	94	1	1	1	0

Table XXXIV: Intensity Indices for the Dominant Language by Country (G–M)

Country	Language	Speakers (%)	NGII	SBII	GAII	GPII
Georgia	Georgian	71	0	0	0	0
Germany	German	91	0	1	1	0
Greece	Greek	98	0	1	1	0
Guatemala	Spanish	65	1	1	1	1
Guinea	Fula	39	0	0	1	0
Honduras	Spanish	97	1	1	1	1
Hong Kong	Cantonese	89	0	0	0	n/a
Hungary	Hungarian	99	0	0	0	0
Iceland	Icelandic	96	0	1	1	n/a
India	Hindi	40	1	1	1	0
Indonesia	Javanese	39	0	0	n/a	n/a
Iran	Persian	46	0	0	0	0
Iraq	Arabic	77	1	1	1	1
Ireland	English	98	0	1	0	0
Israel	Hebrew	63	1	1	1	1
Italy	Italian	94	1	1	n/a	0
Japan	Japanese	99	0	0	n/a	0
Jordan	Arabic	98	1	1	1	1
Kazakhstan	Kazakh	46	0	0	n/a	0
Kiribati	Kiribati	99	n/a	n/a	n/a	0
Korea, South	Korean	100	0	0	n/a	0
Kuwait	Arabic	78	1	1	1	1
Kyrgystan	Kirghiz	60	0	0	0	0
Laos	Lao	53	0	0	n/a	0
Latvia	Latvian	56	1	1	1	0
Lebanon	Arabic	91	1	1	1	1
Libya	Arabic	96	1	1	1	1
Lithuania	Lithuanian	84	1	1	n/a	0
Macau	Cantonese	86	0	0	0	n/a
Macedonia	Macedonian	67	0	1	n/a	0
Madagascar	Malagasy	99	0	0	0	0
Malawi	Chichewa	59	0	0	1	0
Malaysia	Malay	58	0	0	n/a	0
Mali	Bambara	32	n/a	n/a	n/a	0
Malta	Maltese	90	1	1	1	n/a
Mauritania	Arabic	82	1	1	1	1
Mexico	Spanish	92	1	1	1	1
Moldova	Romanian	62	0	1	n/a	0
Mongolia	Khalkha	84	0	0	0	0
Morocco	Arabic	65	1	1	1	1
Myanmar (Burma)	Burmese	69	0	0	0	0

Table XXXV: Intensity Indices for the Dominant Language by Country (N–S)

Country	Language	Speakers (%)	NGII	SBII	GAII	GPII
Namibia	Ndonga	60	n/a	n/a	n/a	0
Nepal	Nepali	50	n/a	n/a	n/a	0
Netherlands	Dutch	96	0	1	n/a	0
New Zealand	English	91	0	1	0	0
Nicaragua	Spanish	98	1	1	1	1
Niger	Hausa	53	1	1	1	1
Nigeria	Hausa	21	1	1	1	1
Norway	Norwegian	97	0	1	n/a	0
Oman	Arabic	77	1	1	1	1
Pakistan	Panjabi	48	1	1	1	0
Palau	Palauan	83	0	0	0	0
Panama	Spanish	77	1	1	1	1
Papua New Guinea	Papuan Lang.	78	0	1	0	n/a
Paraguay	Spanish	55	1	1	1	1
Peru	Spanish	80	1	1	1	1
Philippines	Tagalog	29	1	1	0	0
Poland	Polish	98	0	1	n/a	0
Portugal	Portuguese	99	1	1	n/a	n/a
Puerto Rico	Spanish	85	1	1	1	1
Qatar	Arabic	40	1	1	1	1
Romania	Romanian	89	0	1	n/a	0
Russia	Russian	81	0	1	1	0
Samoa	Samoan	99	n/a	n/a	n/a	0
Saudi Arabia	Arabic	95	1	1	1	1
Senegal	Wolof	48	n/a	n/a	n/a	0
Serbia	Serbian-Croatian	75	0	1	n/a	0
Singapore	Mandarin	77	0	0	0	0
Slovakia	Slovak	86	0	1	n/a	0
Slovenia	Slovene	88	0	1	n/a	0
Somalia	Somali	98	n/a	n/a	n/a	n/a
South Africa	Zulu	24	0	0	1	0
Spain	Spanish	74	1	1	1	1
Sudan	Arabic	49	1	1	1	1
Sweden	Swedish	90	1	0	n/a	0
Switzerland	German	64	0	1	1	0
Syria	Arabic	90	1	1	1	1

Table XXXVI: Intensity Indices for the Dominant Language by Country (T–Z)

Country	Language	Speakers (%)	NGII	SBII	GAII	GPII
Taiwan	Xiamen	67	0	0	n/a	n/a
Tajikistan	Tajik	62	0	0	n/a	0
Tanzania	Nyamwezi-Sukuma	21	0	0	1	n/a
Thailand	Thai	53	0	0	0	0
Togo	Ewe	23	0	0	0	0
Tunisia	Arabic	99	1	1	1	1
Turkey	Turkish	88	0	0	0	0
Turkmenistan	Turkmen	77	0	0	n/a	0
Ukraine	Ukrainian	65	0	1	1	0
United Arab Emirates	Arabic	42	1	1	1	1
United Kingdom	English	97	0	1	0	0
United States	English	82	0	1	0	0
Uruguay	Spanish	97	1	1	1	1
Uzbekistan	Uzbek	76	0	0	0	0
Venezuela	Spanish	96	1	1	1	1
Vietnam	Vietnamese	87	0	0	0	0
Yemen	Arabic	100	1	1	1	1
Zimbabwe	Shona	72	0	0	1	0

Table XXXVII: Definitions and Data sSources and definitions: Institutions

Dominant Llanguage		Source: Britannica World Data in Britannica Year Book 2010 (World Data)
fem_loan 2009	Women's access to bank loans (from 0 = no restriction to 1 = full restriction)	World Bank Gender Statistics
Credit Doepke_Tertilt_Voena	Women's access to bank loans (0 = full restriction, 1 = no restriction)	ADD to RefsDoepke,Tertilt, and Voena (2011), The Economics and Politics of Womens Rights
fem_land 2009	Women's access to land (from 0 = no restriction to 1 = full restriction)	World Bank Gender Statistics
Land Doepke_Tertilt_Voena	Women's access to land (0 = full restriction, 1 = no restriction)	Doepke,Tertilt, and Voena (2011), The Economics and Politics of Womens Rights
fem_prop 2009	Women's access to property other than land (from 0 = no restriction to 1 = full restriction)	World Bank Gender Statistics
Prop Doepke_Tertilt_Voena	Women's access to property other than land (from 0 = full restriction to 1 = no restriction)	Doepke,Tertilt, and Voena (2011), The Economics and Politics of Womens Rights
sex_ratio 2009	Sex ratio at birth (females per 1,000 males)	World Bank Gender Statistics

Table XXXVIII: Definitions and Data sSources and definitions: Gender socio-Socioeconomics vVariables

GDP per capita (2009)	PPP converted GDP per capita, G-K method, at current prices, in obscureintertiol $	Penn World Table 7.0
GDP per capita (2005)	PPP converted GDP per capita, G-K method, at current prices, in intertiol $	Penn World Table 7.0
lab_part 2009	Female Llabor participation rate, female (% of female population ageds 15+)	World Bank Development Indicators
lab_for 2009	Female Llabor force, female (% of total labor force)	World Development Indicators
agri_fem 2009	Agricultural employment, ratio of female to male	World Development Indicators
agri_fem 2005	Agricultural employmentEmployees, agriculture, ratio of gender employment female to male	World Development Indicators
ind_fem 2009	Industrial employment, ratio of female to male	World Development Indicators
ind_fem 2005	Industrial employment,Employees, industry ratio of gender employment female to male	World Development Indicators
serv_fem 2009	Services employment, ratio of female to male	World Development Indicators
serv_fem 2005	Services Eemploymentees, services, ratio of gender employment female to male	World Development Indicators
agri_fem 2009	Employees, agriculture, ratio of gender employment female to male	World Development Indicators
ind_fem 2009	Employees, industry ratio of gender employment female to male	World Development Indicators
serv_fem 2009	Employees, services, ratio of gender employment female to male	World Development Indicators
prim_enrol	Enrollment in primary OK?education, Rratio of female to male primary enrollment (%)	World Development Indicators
sec_enrol	Enrollment in secondary education, Rratio of female to male secondary enrollment (%)	World Development Indicators
tert_enrol	Enrollment in tertiary education, Rratio of female to male tertiary enrollment (%)	World Development Indicators
Global Gender Gap Index 2009		World Economic Forum, The Global Gender Gap Report 2012
school_ratio	Expected years of schooling, ratio of female to male	World Bank Gender Statistics
dom_work	HoursTime spent performingin domestic work, ratio of female to male (hours)	World Bank Gender Statistics

Table XXXIX: Definitions and Data sSources and definitions: Language Variables

NGII	Intensity index dummy: (1 if NG2 = 1 ; 0 otherwise)	Dryer and, Matthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
SBII	Intensity index dummy: (1 if SBIOK?Y = 1 ; 0 otherwise)	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
GAII	Intensity index dummy: (1 if GAH = 1 ; 0 otherwise)	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
GPII	Intensity index dummy: (1 if GPH = 1 ; 0 otherwise)	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
GII	Aggregate index: GII = NGI + SBI + GAI + GPI	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
GIIV1	Aggregate sub-subindex: GIIV1 = NGI + SBI + GAI	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
GIIV2	Aggregate sub-subindex: GIIV2 = NGI + SBI + GPI	Dryer, andMatthew S. & Haspelmath, Martin (eds.). (2011). The World Atlas of Language Structures Online
dom_sh	Share of speakers of the dominant language in the population	Britannica World Data in Britannica Year Book 2010 (World Data)

Table XL: Control vVariables

comm	Dummy for if communist regime in the past	
dist_equ	Distance from equator ($\sqrt{\text{latitude}SQRT(Latitude)}/90$)	Portlan
trop_sh	Share of the population in tropical areas	Portlan
frost_days	Average number of frost days per unit of population	Masters and McMillan (2001)ADD t
coast_sh	Share of population within 100 km of the coast (or an ocean-navigable river)	Portlan
landlocked	Dummy forif the country beingis landlocked	CIA wWorld fFa
colo	Dummy forif colonized country	
eng_col	Dummy forif English colonization	
fren_col	Dummy forif French colonization	
spa_col	Dummy forif Spanish colonization	
prot_du	Dummy forif pProtestantism asis the main religion	Barro,
cat_du	Dummy forif chCatholicism asis the main religion	Barro
mus_du	Dummy forif iIslam asis the main religion	Barro,
jew_du	Dummy forif jJudaism asis the main religion	Barro,
hind_du	Dummy forif Hinduism asis the main religion	Barro,
budd_du	Dummy forif bBuddhism asis the main religion	Barro,
Main religion > 50% ?	Dummy for whetherif main religion is followed by more than half the50% of population	Barro,

Table XLI: More Control vVariables

lit_rate	Literacy rate, (%adult female (% of females ageds 15+) and above)	World Bank Gender Statistics
log_inc	lLog of GDP per capita (PPP current $)	Penn World Table 7.0
log_inc_sq	sSquared log of GDP per capita (PPP current $)	Penn World Table 7.0
openness	Oppenness at current prices in % (Eexports plus imports, divided by GDP)	Penn World Table 7.0
gov_size	Government consumption share of PPP, converted to GDP per capita at current prices (%)	Penn World Table 7.0,0
oil_rents	Annual oil and natural gas production, minus country-specific extraction costs	Ross (2008)
demo	Democracy index,: Ffrom 0 (least democratic) to 10 (most democratic)	Polity 4 pProject
log_pop	Log of population (in millions of people)	Penn World Table 7.0
urb_sh	uUrban population share	World Bank Gender Statistics
legor_uk	Dummy for Englishif origin of country'sthe legal system is English	La Porta et al., (1999)
legor_fr	Dummy for Frenchif origin of the country's legal system is French	La Porta et al., (1999)
legor_ge	Dummy for Germanif origin of the country's legal system is German	La Porta et al., (1999)
legor_sc	Dummy for Scandinavianif origin of the country's legal system is Scandinavian	La Porta et al., (1999)
legor_so	Dummy for socialistif origin of the country's legal system is socialist	La Porta et al., (1999)
please confirm or reviseNo entry here?	**Female Llabor force with tertiary education, female (% of female labor force)**	World Bank Gender Statistics
Ploughw	Ploughw index	Alesinao, Giuliano, and Nunn (forthcoming2011), Fertility and the plough

Table XLII: Financial Independencet vVariables: Female/Male Ratio (aged 15+)

Ratio of Account at a formal financial institution (female to male, age 15+)	World Bank Gender Statistics
Ratio of Account used for business purposes (female to male, age 15+)	World Bank Gender Statistics
Ratio of Account used to receive wages (female to male, age 15+)	World Bank Gender Statistics
Ratio of Loan from a financial institution in the past year (female to male, age 15+)	World Bank Gender Statistics
Ratio of Loan from a private lender in the past year (female to male, age 15+)	World Bank Gender Statistics
Ratio of Loan from an employer in the past year (female to male, age 15+)	World Bank Gender Statistics
ratio of Loan from family or friends in the past year (female to male, age 15+)	World Bank Gender Statistics
ratio of Outstanding loan for health or emergencies (female to male, age 15+)	World Bank Gender Statistics
ratio of Outstanding loan to purchase a home (female to male, age 15+)	World Bank Gender Statistics
ratio of Personally paid for health insurance (female to male, age 15+)	World Bank Gender Statistics
ratio of Saved any money in the past year (female to male, age 15+)	World Bank Gender Statistics
Strength of legal rights index (from 0 = weak to 10 = strong)	World Bank Gender Statistics