Language and Socio-Economic Structures: The Case of Culture and Gender

Estefania Santacreu-Vasut †
Amir Shoham ‡

Abstract

We provide the first systematic analysis of the relation between languages and socio-economic structures. We focus on explaining cross country variations in gender outcomes and find that gender marking in the dominant language, measured by four different variables, decreases female labor force participation rate and the share of female in the labor force by about one standard deviation. Further, it increases (decreases) the relative share of female to male in services (agriculture). These results are robust to controlling for income levels, geography, colonization and political institutions. We also find suggestive correlations between gender marking and other socio-economic outcomes.

JEL Code: A13, J16, J24, O57, Z1
Keywords: Culture, Female Labor Participation, Gender, Language, Gender, Occupation

---

† Department of Economics at ESSEC Business School and THEMA, France, santacreuvasut@essec.edu
‡ The FOX School of Business Temple University and The College of Management Academic Studies Israel, amir1s@colman.ac.il
We thank Victor Gay for excellent research assistance and Ceressec ESSEC for funding support. We are grateful to the participants of the ITFA conference 2011, PSE Economic History seminar, THEMA lunch seminar and to Guillaume Chevillon, Greville Corbett, Barry Eichengreen, Ramon Ferrer i Cancho, Johanna Nichols, Rodrigo Paillacar, Gerard Roland, and Anne Zribi-Hertz for helpful conversations.

Since North's (1990, 1993) seminal work, a growing body of research has corroborated the view that culture is a determinant of institutions and economic performance.[1] Until very recently, and despite the growing sense that culture matters, economists conducting empirical research rarely used culture as an explanatory variable of economic events, as Guiso et al. (2006) point out. The main reason is that culture is too difficult to define, measure, and potentially endogenous to economic outcomes.

To address these challenges, this paper proposes to use the grammatical structure of languages as markers of culture. We argue that culture has an exogenous or predetermined aspect that manifests itself in the grammatical structure of language. Our aim is to capture cultural influences originating in the distant past, since these have been shown to have a persistent influence on current socio-economic outcome (Guiso et al., 2008). Since the grammatical structures of languages are generally slow to change, they are predetermined, and can be used as markers of culture. Recent research argues that language is determined by coevolutionary forces of biology and culture (Christiansen and Kirby, 2003).

In this paper we focus on explaining cross country variations in gender outcomes (in particular, labour force participation and females occupational choices), an area of research where economists argue that culture plays an important role (Fernandez, 2010; Alesina *et al.*,2011).

We analyze all grammatical features of languages that are potentially related to gender, and a broad range of gender outcomes. In particular, our variables are Number of Genders, Sex-based and Non-sex-based Gender Systems, Systems of Gender Assignment and Gender Distinctions in Independent Personal Pronouns.

We consistently find that language structure matters. Even after controlling for the level of economic development, and controlling for geography, religion, and colonial history, womens participation in the labour force is lower in countries where the grammar of the dominant language has more gender markers. We extend our analysis by correlating gender intensity of the grammar to other gender outcomes such as political participation, ownership rights, household decisions, and even broader outcomes like income inequality. These correlations confirm the findings of our

[1] Among others, Licht et al., 2007; Tabellini, 2008a; Gorodnichenko and Roland, 2010; Guiso et al., 2010.

regression.

This is the first paper to provide a comprehensive analysis of the relationship between linguistic and socio-economic structures. While the idea that cultural knowledge is transmitted via language was raised by North (1993) in his Nobel prize lecture, its use has been sporadic and a comprehensive approach has been lacking.[2] We go deeper into the analysis of languages and the information contained in their structure.

Further, our analysis can be applied to deepen the study of the relationship between culture and economics. Survey-based measures of culture, like Hofstedes (1980,1983) cultural dimensions, House et al., (2004) GLOBE study, Schwartz (1994, 1999), and the World Value Survey have the advantage of measuring the attitudes and beliefs of society's culture directly but are particularly susceptible to severe endogeneity problems.

In the literature, three methods have been used to improve existing measures of culture, and isolate the impact of culture on economic outcomes. The epidemiological approach studies the behaviour of migrants living in the same country. While they carry with them the culture of their country of origin, they share a common institutional environment. Differences in their behaviour, therefore, can be attributed to their cultural origins. However, this approach is problematic because migrants are a selected pool, likely to be a non-representative sample of their home countrys culture. Alesina *et al.*(2010) studied the origins of differences in gender roles across countries using historical agricultural practices. Finally, experimental work offers a promising approach for studying whether nature or nurture is the origin of gender outcomes, such as womens taste for competition, and risk-aversion consistent behaviours. (Booth and Nolen, 2012).

To date, the use of linguistic variables has been limited to isolated instruments for specific questions. For example, Licht *et al.*(2007) use the grammar of pronouns as an instrumental variable to study how countries tilted more in favour of autonomy, egalitarianism, and mastery exhibit a higher rule of law, less corruption, and more democratic accountability. They argue

[2]North argues that 'this is the kind of learning that the individuals in a society acquire over time. Time in this context encompasses not only current experiences and learning, but also the cumulative experience of past generations that is embodied in culture. Collective learning - a term used by Hayek - consists of those experiences that have passed the slow test of time and are embodied in our language, institutions, technology, and ways of doing things. It is the transmission in time of our accumulated stock of knowledge (Hayek, 1960: 27)'.

that languages which require the explicit use of I or you signal that the person is highlighted, and autonomy is valued. They point out that culture (linguistic) has significant influence of on governance. Mavisakalyan (2011), uses the gender of pronouns to investigate the impact of culture on females share in the labour force. Chen (2012), uses future time references marking in languages to investigate the impact future oriented decisions and outcomes like saving, debt and health related behavior. Tabellini (2008b) seeks to capture the distinction between values consistent with generalized morality versus those consistent with limited morality. Altogether, norms of generalized morality promote well-functioning institutions. To control for the possibility of reverse causality, and identify the causal impact of these values on institutional outcomes, Tabellini uses the grammar of pronouns as an instrumental variable.

More generally, Guiso *et al.*(2009) document that trust is affected by geographical distance between two countries, their proximity, and the commonality between their languages. Finally, Falck *et al.* (2010) and Cavalli-Sforza (2000) argue that language acts as a type of memory that stores information, similar to a genome.[3]

An important advantage of our approach is that it can be exploited for both cross country analysis and individual analysis, because we can use the distribution of languages and speakers of different languages across countries and information about the languages spoken by migrants within a single country. In addition, linguistic diversity within countries makes it possible for us to measure dominant versus minority and colonial, inherited cultures. Finally, compared to existing culture databases that include a small sample of countries, biased towards western and more developed countries, ours includes the largest sample of countries available, 217 countries.

We proceed as follows. Section 1 presents our motivation. Section 2 our data. Section 3 our empirical analysis. Section 4 discuss and extends our analysis. Section 5 concludes proposing future avenues of research.

[3]To back up their claim they even quote Darwins Origin of Species If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the languages now spoken around the world; and if all extinct languages, and all intermediate and slowly changing dialect, were to be included, such an arrangement would be the only possible one.

1 Motivation

This paper builds on the idea that the structure of languages contains information related to the culture of our ancestors. Among linguists, particularly among the so-called generativists, languages are considered part of human biology and fundamentally equal in structure, with only minor local differences (Chomsky, 1980). Sampson *et al.*, (2009) point out a new wave of linguists who think of languages as institutions developed as part of a societys cultural heritage and hence as differing and evolving in their levels of complexity, just as other cultural institutions do.

The study by Evans (2003) of how kinship-specific pronouns evolved in Australia is an example of a study focused on how cultural forces may shape linguistic change. As Evans and Levinson (2009) point out, language diversification and hybridization works just like the evolution of biological species it is a historical process, following the laws of population biology and linguistic diversity is structured very largely in phylo-genetic (cultural-historical) and geographical patterns. Christiansen and Kirby (2003) review the research on the origins and evolution of human language and argue that while one theoretical line argues that grammatical structure is the product of biological adaptation, others argue that it emerges through cultural transmission of language to hundreds (or perhaps thousands) of generations of learners.

Overall, current theories of coevolution point to the interplay between genetic and cultural forces to account for language diversity and change. Finally, findings from cognitive psychology about the impact of language on cognition (Boroditsky and Gaby, 2010) indicate that there may be direct channel through which language structure influence socio-economic choices and outcomes.

From a methodological perspective, the main advantage of using languages grammatical structures is that they are inherited from the distant past. A recent trend in linguistic research studies the stability of grammatical structures. In particular, Wichmann and Holman (2009) constructed a measure of stability for analyzing the linguistic features described in the World Atlas of Linguistic Structures, (Dryer and Haspelmath, 2011).[4] They define stability as, the probability that a given language remains unchanged with respect to the feature during 1000 years, that is, the fea-

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

ture undergoes neither internal change nor diffusion during the interval(Wichmann and Holman, 2009).

From an economic point of view, the stability of grammatical features is not surprising, and could be related to how network externalities affect technology adoption (Katz and Shapiro, 1986). Language is a technology characterized by networks externalities, since the value of mastering a language increases with the number of speakers. Linguistic evolution can be seen as a sort of technological adaption. As Katz and Shapiro show, in the absence of sponsors, the technology superior today has a strategic advantage and is likely to dominate the market (Katz and Shapiro,1986).[5] This is likely to be the case of languages, which are characterized by an absence of an owner or sponsor (defined as an entity that has property rights to the technology and hence is willing to make investments to promote it.(Katz and Shapiro, 1986) The linguistic variables related to gender used in this paper are all categorized as very stable by Wichmann and Holman (2009).

2 The Data

We built a data set of 192 language structure variables from the World Atlas of Language Structure (WALS), (Dryer and Haspelmath, 2011), for 217 countries. Our language dataset contains information on the structure of all languages spoken as the mother tongue of a country's population. We used Encyclopedia Britannica (2010) to obtain information about mother tongues and the number of speakers. We define the dominant language as the one spoken as mother tongue by a majority of the population. Further, we also coded information about whether or not the language has a colonial origin.[6]

In this paper, we focus on exploiting the information contained in the structure of the dominant language and language variables directly related to expressions of gender (4 out of 192). These correspond to chapter 30 (Corbett, 2011a), chapter 31 (Corbett, 2011b), chapter 32 (Corbett, 2011c) and chapter 44 (Siewierska, 2011) of WALS . We created 12 dummy variables out of these,

[5] Tang and Kevoes (2008) argue that changes in economic conditions are the source of cultural dynamics, while institutions, among them language, religion and legal provide the foundation for cultural stability.

[6] In a few cases, we used other sources to complement the data as described in the appendix.

as described in the next section. We match our language dataset with socio-economic data from a variety of sources (see Table 7 in the appendix) and end up with a sample of 137 countries and 80 different dominant languages (see Tables 9 and 10 in the appendix). Table 1, and Table 8 in the appendix, present the descriptive statistics of our language and socio-economic variables respectively.

Table 1: *Descriptive Statistics Dominant Language (per country)*

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

2.1 *Gender Linguistic Variables Description*

The defining characteristic of gender is agreement. A gender system is a set of rules for agreements that depend on nouns of different types. These types can be sex-based (female and male), or based on other social constructs (such as the distinction between human and animal, age, social status...) As Greville G. Corbett mentions in Chapter 30 of WALS 'In some languages gender is evident in almost every phrase, while in other languages it is absent.' (Corbett, 2011). This is the variation we exploit in this paper.

Number Genders (NG): This variable captures how many genders are present in the language.

That is, how many different types of nouns have different agreements.[7] In particular, we built three dummy variables, NG0, NG2, and NG3, corresponding to languages having 0, 2, or 3 (and more) genders, respectively.[8] A language with 2 genders, like French and Spanish, typically implies feminine versus masculine while a language with three or more genders may include neuter as the third gender, like German, or no sex related distinctions. For our purposes, we consider NG2 as more gender intensive than NG3 which is, in turn, more gender intensive than NG0.

Sex Based (SB): This variable captures whether the gender system is linked to biological sex. We built three dummy variables, SB0, SBY, and SBN, corresponding to languages having no gender system, a sex-based gender system, or a non-sex-based gender system, respectively.[9] For example, a non-sex-based gender system might be based on the distinction between human and non-human, as in Fulfulde, a member of the Niger-Congo linguistic family, or the distinction between animate and inanimate, among others. Corbett, author of the feature about this variable in WALS, argues, for those interested in language and gender in the sociological sense, these direct reflections of biological sex in many languages, with the control group of languages with similar gender structures but without the sex component, should provide a valuable source of data. (Corbett, 2011). For our purposes, we consider SBY as more gender intensive than SBN, which is, in turn, more gender intensive than SB0.

Gender Assignment (GA): This variable captures how a speaker assigns nouns to the genders defined by the gender system of the language. A gender assignment system provides a set of rules to help the speaker make appropriate agreements. Assignment can depend on the meaning (semantic) or the form of the noun. We built three dummy variables, GA0, GAL, and GAH, corresponding to languages having no gender assignment system, a semantic gender assignment or both gender assignment system that is both semantic and formal, respectively. For our purposes, we consider GAH as more gender intensive than GAL, which is more gender intensive than GA0.[10] For example, a semantic assignment system is found in Kannada (a south India language), where

[7] Gender derives from Latin genus and originally it meant kind or sort.
[8] There are languages, such as Nigerian Fula that feature 20 genders.
[9] Y and N stand for Yes and No.
[10] 'L' and 'H' stand for 'Low' and 'High'.

nouns denoting male (female) humans are masculine (feminine) and all remaining nouns are neuter. In a semantic and formal assignment system, as used in Russian, in addition to semantic gender assignments, nouns that are neither masculine nor feminine are not necessarily neuter; rather they can be assigned to the masculine or feminine gender, depending on their inflectional class (whether it takes the nominative form or accusative form, etc.).

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

2.2 *Preliminary Analysis: Gender Intensivity and Female in the Labor Market*

Because we aim at capturing the intensity of female/male gender distinctions in languages, and its relationship to gender outcomes, we have categorized the following set of features as gender intensive: NG2, SBY, GAH, and GPH. As Table 2 shows languages in our dataset do not necessarily exhibit all these features together, even though there is a strong positive correlation among them. Out of a sample of 86 countries for which we have information about our four language variables, less than half (41) have dominant languages that exhibit all four of these features.

Figure 1 shows the distribution of the variable Number Genders.[11]

Table 3 presents the correlation matrix of the gender intensive dummy variables (NG2, SBY, GAH, GPH) and five variables related to female participation in the labour market. In particular,

[11] Maps of the other three variables may be found in the appendix.

Table 2: *Autocorrelation Gender Intensive Variables*

	NG2	SBY	GAH	GPH
NG2	1			
SBY	0.548***	1		
GAH	0.698***	0.595***	1	
GPH	0.793***	0.480***	0.646***	1

Variables defined as in the text.

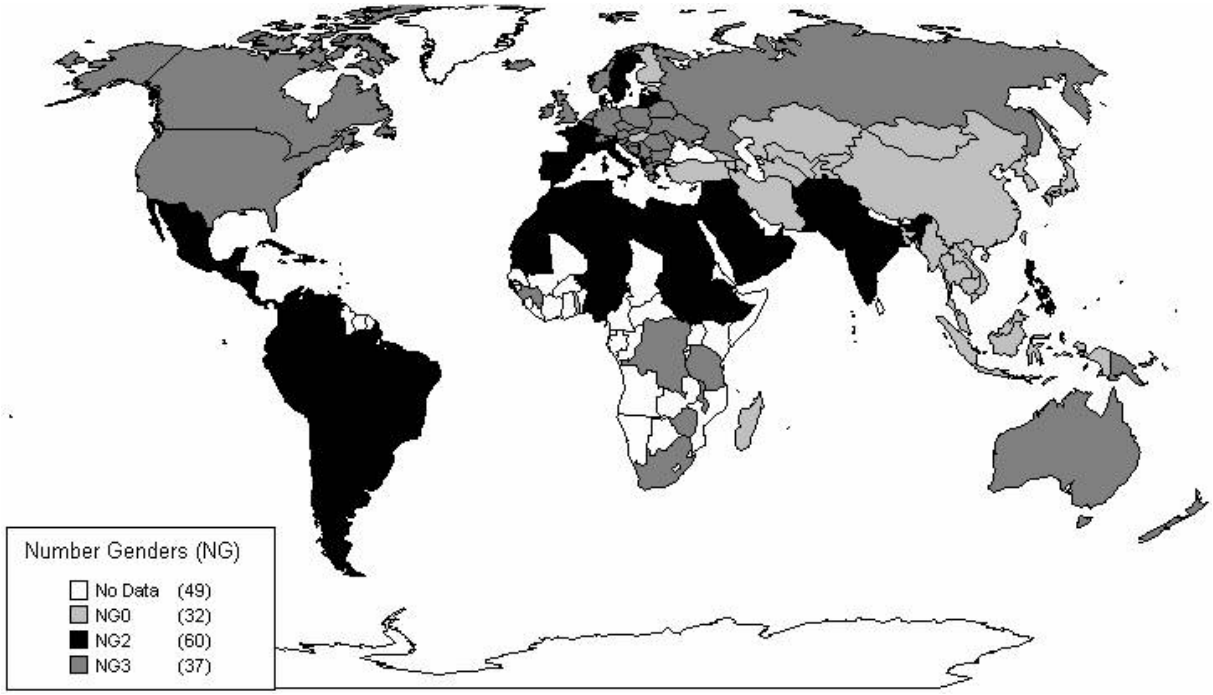

Figure 1: Map for Number Genders

the rate of females' participation in the labour force (Lab_part), the share of females in the labour force (Lab_for), as well as their occupational distribution, that is the share of females relative to share of males in services (Serv_fem), industry (Ind_fem) and agriculture (Agri_fem). Countries whose dominant language has gender intensive features exhibit lower participation rates for females in the labour force, and lower shares of females in the labour force. Regarding occupation, gender intensive languages are positively correlated with a positive female bias (relative to males)

in services, and a negative bias in agriculture; the correlations suggest no bias for females in industry. We now turn to our empirical analysis.

Table 3: *Female Labor Market and Gender Intensity*

	NG2	SBY	GAH	GPH
Lab_part	-0.570***	-0.467***	-0.409***	-0.561***
Lab_for	-0.612***	-0.367***	-0.429***	-0.642***
Serv_fem	0.182	0.356***	0.106	0.17*
Ind_fem	0.090	-0.044	-0.017	0.075
Agri_fem	-0.221**	-0.229**	0.016	-0.315***

***p-value< 0.01, **p-value< 0.05, *p-value< 0.10
Our variables are defined as in the text.

3 Empirical analysis

This section presents OLS regressions for participation rates of females in the labour force, and for females' occupational choices (relative to males' choices) in services, industry and agriculture on our gender dummy variables, as previously defined, a set of control and robustness checks of the variables for the year 2000. In all cases, our excluded dummy variable is the one corresponding to no gender marking: NG0, SB0, GA0, GP0. Therefore, we must interpret our coefficients relative to these variables. The definitions, sources and descriptive statistics for our variables may be found in the appendix.

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

show that the share of female employment is higher in firms that export (Baslevent and Onaran, 2004). Finally, we control for political factors such as government size, polity score on democracy, a communist past, and population size.

The structure of dominant languages varies across countries and across linguistic families. To control for the role of geography and historical influences on the distribution of language structures across countries, we performed three types of robustness checks: geography, colonization, and religion. However, because controlling for religion significantly reduced our sample size, we did not include these results. They are available in the online appendix.[12] For geography, we controlled distance from the equator, which may reflect Western influence as Hall and Jones (1999) argue. Following Bloom and Sachs (1998), and Gallup *et al.*(1998) in an alternative specification, we control for climate because it may influence development. Because a colonial past influences institutions and human capital (Acemoglu *et al.*, 2001), which influence development, we controlled for colonization and for its origin (English, French or Spanish), since the origin of the colonizer has been shown to impact current institutions (La Porta *et al.*, 1998, 1999). Our results regarding the origin of the colonizer are robust, and are available in the online appendix.

Table 4 presents the results for the rate of females' participation in the labour force. Specification (1) is the baseline regression with our set of control variables only. Specifications (2)-(5) include each of our language variable dummies. Because of the correlation between our gender intensive variables, we did not include all of the dummies in the each specification. Specifications (6)-(17) include three robustness checks, two controlling for geography, and one for colonization.

[12] See $http://works.bepress.com/santacreu_vasut/$

Table 4: *Female Labor Force Participation Rate*

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

t-statistics in parentheses

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

Countries having a dominant language with Number of Genders equal to 2 have 17 % points lower female participation rate than countries with Number of Genders equal to 0. Note that having a Number of Genders equal to 3 or more has a smaller negative effect (approximately 5% points), but it is not significant. This is consistent with the fact that in languages with three or more genders the female/male distinction is less pervasive. Having a Sex-based Gender System decreases females' labour force participation rate by 12% points compared to baseline labour force participation in countries with no gender system. Again, this result is very significant both statistically and economically. Conversely, in a country with a Non-sex-based Gender System, females' labour force participation rate is approximately 3% points higher than their labour force participation rate than one with no gender system, but this result is not significant. Regarding Gender Assignment and Gender Pronouns, our results are intriguing because these variables include an intensity dimension. Having gender assignment that is both semantic and formal, or gender marking in a larger set of pronouns decreases the participation rate relative to not having one, more than having semantic gender assignment only or a smaller set of gender marking in pronouns. Our results are similar for another measure of females' participation in the labour market that controls for overall labor force size: the share of the labour force that is female (see Table 11 in the appendix). Further note that the Adjusted R-Squared almost doubles when we include gender language variables.

Are labour markets segregated by gender? Table 5 shows the results for the distribution of females, relative to males, in service, industrial and agricultural occupations. Our results suggest that indeed they are. There is a positive female employment bias in services, and a negative female employment bias in agriculture in countries where the dominant language is more gender intensive. Results on services are very significant. The gender intensive dummies are all significant both statistically and economically. Having a language with Number of Genders equals two, as opposed to none increases, ceteris paribus, the relative share of females employed in services by a magnitude similar to having a sex-based gender system relative to having no gender system at all. Regarding gender pronouns, our results are even stronger in magnitude.

Table 5: *Occupation Profile, Language and Gender*

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

t-statistics in parentheses

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

Further it is intriguing that among gender intensive dummies those that mark gender more have a stronger impact on female employment bias in services. In the opposite direction, having a non-sex-based gender system is negatively, but not significantly, associated with females share in services. Regarding agriculture, the share of female employment in agriculture relative to male employment is reduced in countries with gender intensive languages. For instance, countries with high gender pronoun marking have a lower female negative bias by almost one standard deviation (-0.37) compared to countries with no gender pronoun marking, a very significant result. In a similar vein, Alesina *et al.* (2010) show that indeed the historical use of the plough correlates negatively with gender cultural norms towards traditional female occupations and female labour force participation today.

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

4 Further Analysis and Discussion

In Figure 2, we plot female's participation in the labour force (%) and a variable that summarizes gender intensity. In particular we build an index equal to the sum of values of the gender intensive dummy variables, GN2, SBY, GAH, and GAL for each country. Its possible values range from 0, for countries whose dominant language have none of these features to 4, for countries whose dominant language have all of them.

Apart from labour market participation and occupation structure, is gender marking in grammar related to other socio-economic variables?

Table 6 shows that indeed there are strong and significant correlations between our set of gender intensive variables and non-gender focused measures such as Inequality (Gini and income share held by highest 10%). This is not surprising since the concept of gender in linguistics, and

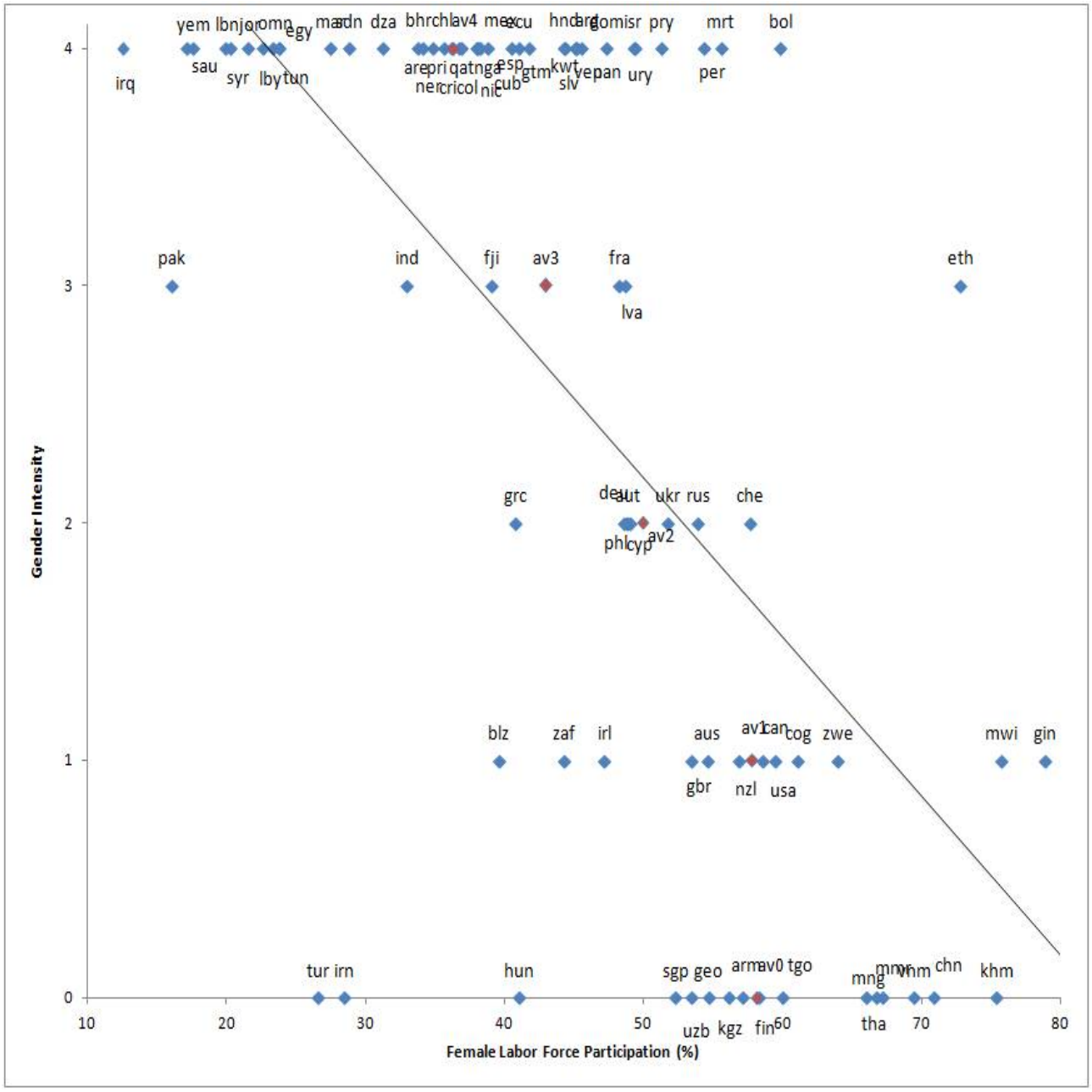

Figure 2: Gender Intensity Index

The red dots correspond to the average Labor Force Participation for a given level of Gender Intensity. They are denoted by AV4 for Gender Intensity of 4, and so on. (see the text for a description of the variables)

our variables encompass not only female/male distinctions but also a set of rules of agreements that depend on nouns of different types, based social constructs such as social status and age. Therefore, in addition of our set of gender intensive variables we include two additional language

variables: SB0 and SBN. That is, whether the language lacks a gender system, and whether the gender system is not sex-based. As we can see, there is a negative correlation between having no gender system and inequality, although it is not significant and the reverse is true for languages with gender system (and higher for non sex-based).

Table 6: *Correlation Language and Other Socio-Economic Variables*

		Language Variables					
		NG2	SB0	SBY	SBN	GAH	GPH
Economic Outcomes							
Inequality	Inc_ineq	.392**	-.146	.075	.093	.259*	.550**
	Gini	.371**	-.133	.097	.047	.192	.530**
Other Gender Outcomes							
Political Part.	Fem_seats	-.151	-.109	-.035	.249**	-.026	-.220*
	Fem_minis	-.078	-.125	-.049	.316**	-.151	-.250**
Ownership Rights	Fem_loan	.185	-.223*	.028	.339**	.255*	.016
	Fem_land	.151	-.156	-.022	.316**	.228	.046
	Fem_prop	.111	-.178	.043	.235*	.125	-.014
Household	Fertil	.320**	-.132	-.007	.250**	.296**	.304**
	Marr	.185	-.153	.012	.222	.275	.081
Human Capital	Lit_rate	-.205*	.141	-.026	-.233*	-.175	-.139

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

Does language structure capture only past cultural influences? Do the languages we speak shape or influence the way we think today? Cognitive scientists currently study cross-linguistic differences in thought related to time, colours, objects, and events. Boroditsky *et al.* (2003), for example, study how grammatical gender influences the way speakers of different languages think about inanimate objects. As Boroditsky argues, 'These 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'. Our current findings show that language structures have economic consequences as well. In particular, this paper focuses on the impact of language gender marking for gender-related outcomes, labour force participation and labour market segregation. Yet, language structure may be related to other important economic outcomes, such as inequality and savings. Our ongoing research seeks to answer these questions. Can our language shape our economy through its impact on cognition?

Using microeconomic data on first, second, and third generation migrants who speak different languages seems a promising avenue for continuing research that addresses the question of whether language is related to culture or cognition, or both, and its economic impact.

5 Conclusion

This paper presented the first systematic analysis of the relationship between language and socio-economic structures. We argue that culture has an exogenous or predetermined aspect that manifests itself in the grammar of a language. Our aim is to capture cultural influences originating in the distant past, since these have been shown to have persistent influence (Guiso *et al.*, 2008).

We focused on all language variables that are related to gender marking, that is, agreement rules governing how nouns of different types function. These can be sex-based or not. We find that, even after controlling for economic development, geography, religion, political institutions, and colonial past, countries where the dominant language marks gender more intensively have significantly lower rate of female participation in labour force, females' share of the labour force, and a more segregated labour market. Female presence is higher than the male presence in services, and lower in agriculture. Further analysis suggests that studying the relationship of other aspects of languages' structure and their consequences for socio-economic outcomes such as inequality, and women's political and economic rights is a promising avenue for research.

While current research points to an interplay of biological and cultural forces in the origin and evolution of languages, language may have a direct influence on our economy via its impact on cognition. Cutting-edge cognitive science is currently studying whether and how language shapes thought. Because languages travel with migrants, studying their behaviour may be a productive research direction for deepening our understanding of these fascinating questions.

6 References

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

401.

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

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

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

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

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

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

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

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

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

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

Corbett, G. G. (2011a). 'Number of genders', in (M. S. Dryer, and M. Haspelmath ed.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library, chapter 30.

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

Library, chapter 31.

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

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

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

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

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

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

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

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

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

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

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

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

Guiso, L., Sapienza, P. and Zingales, L. (2009). 'Moral and social constraints to strategic default

on mortgages', NBER Working Paper No. 15145.

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

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

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

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

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

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

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

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

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

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

Marshall, M. G., Jaggers, K. and Gurr, T. D. (2011). *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010*, Center for International Development and Conflict Management, University of Maryland.

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

pp. 681-712.

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

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

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

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

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

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

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

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

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

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

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

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

7 Data Appendix

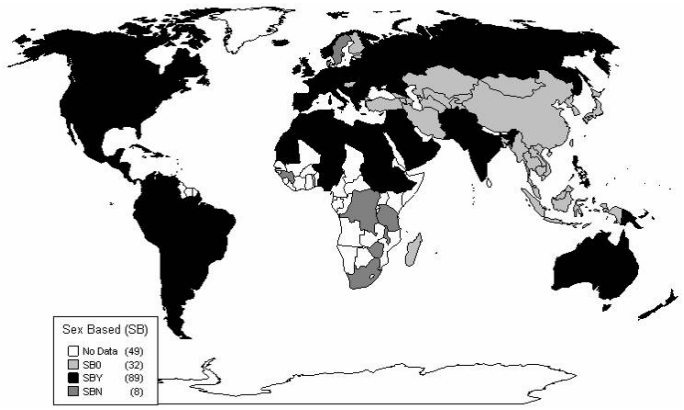

Figure 3: Map Sex Based

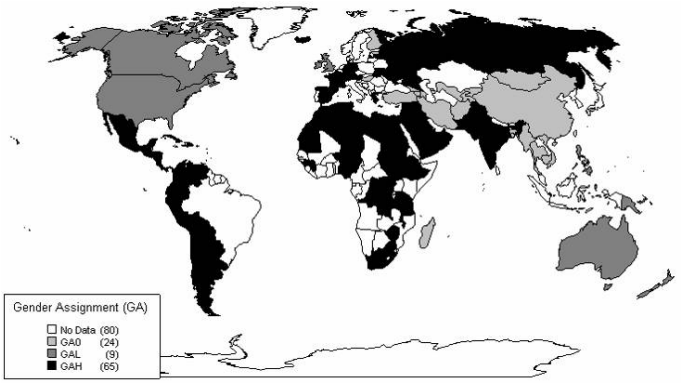

Figure 4: Map Gender Assignment

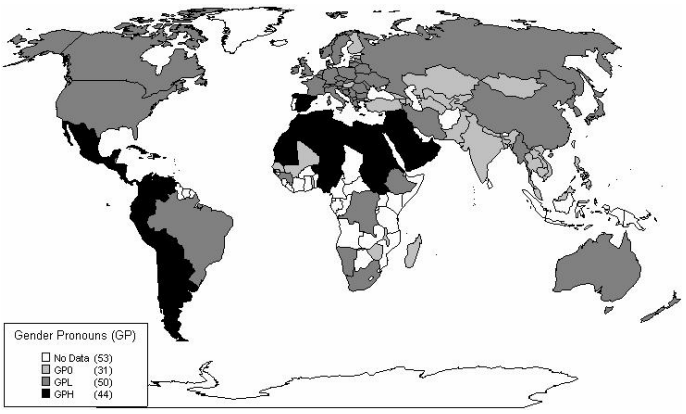

Figure 5: Map Gender Pronouns

Table 7: *Sources and Definition*

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

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

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

Table 8: *Descriptive Statistics*

	N	Minimum	Mean	Std. Dev.	Maximum
Dependent					
Lab_part	180	12.60	51.04	15.46	90.50
Lab_for	177	12.26	39.84	8.78	53.15
Serv_fem	131	0.21	1.30	0.36	2.53
Ind_fem	146	0.02	0.57	0.46	4.10
Agri_fem	126	0.00	0.71	0.42	2.33
Controls					
Dom_sh	189	0.18	0.74	0.24	1.00
Log_inc	189	2.28	3.63	0.58	4.74
Log_inc_sq	189	5.21	13.48	4.18	22.42
Openness	189	2.01	90.62	49.64	377.68
Gov_size	189	1.69	12.56	9.10	56.05
Oil_rents	189	0.00	431.15	1789.61	18057.62
Demo	189	0	5.75	4.00	10
Log_pop	188	1.29	3.72	0.88	6.10
Comm	189	0	0.22	0.42	1
Robustness Checks					
Dist_equ	189	0	0.28	0.19	0.75
Trop_sh	189	0	0.53	0.49	1
Frost_days	189	0	7.23	9.55	29.88
Coast_sh	189	0	0.62	0.38	1
Landlocked	189	0	0.21	0.41	1
Colo	189	0	0.67	0.47	1
Eng_col	189	0	0.26	0.44	1
Fren_col	189	0	0.14	0.35	1
Spa_col	189	0	0.11	0.32	1
Gender and Economic Outcomes					
Fem_seat	181	0	11.22	8.80	43
Fem_minis	170	0	8.15	8.11	43
Fem_loan	123	0	0.26	0.29	1
Fem_land	122	0	0.34	0.32	1
Fem_prop	123	0	0.26	0.30	1
Fertil	188	0.93	3.24	1.71	7.73
Marr	95	1.80	28.55	17.14	76.60
Lit_rate	144	9.40	73.72	25.28	100.00
Inc_ineq	138	21.28	32.21	7.44	55.19
Gini	137	24.70	40.82	9.27	64.30

Table 9: *Language Gender Dummies Dataset*

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

Table 10: *Language Gender Dummies Dataset (continued)*

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

na: Not available

Table 11: *Female Share of the Labor Force*

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

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