COLONIAL INFLUENCE AND THE ORIGINS OF GENDER ROLES

Ana Paula Franco Carruitero

Abstract

This study argues that current heterogeneity in female labour force participation across elevation levels in Peru may be explained by differences in historical Spanish influence. Before conquest, men and women were equally-important in the economic sphere. After conquest, women became "legal minors" without enough reasoning power to make economic decisions. Because of Spanish reluctance to settle at higher elevations, I argue that inclusive pre-conquest gender roles persist in those areas. Using elevation as an instrument for Spanish influence, I find that living in a district less exposed to Spanish influence significantly increases female labour force participation. Results are highly robust to the inclusion of a variety of demographic, geographical, contemporary and historical covariates.

1. INTRODUCTION

A clear gender gap in labour force participation (LFP) has been identified in Peru. However, this gap is not equally pronounced everywhere. In particular, female LFP exhibits heterogeneity across elevation levels, ranging from 60 to 80%, with higher participation at higher elevation levels. Male LFP does not exhibit such variation, thus the gender participation gap diminishes as elevation increases (see Figure 1).

This study argues that heterogeneity in female LFP may be explained and traced back to differences in historical cultures, and specifically to colonial Spanish influence that varied geographically across elevation[1]. According to historical sources, elevation determined the Spanish settlement during the conquest in 1532 because their international-trade-oriented economic system favored settlement in coastal cities with a port. This trend was not relevant previously, as the Incas ruled and settled across all elevation levels. In Inca culture, men and women were considered equally-important partners in the business of life, as a way of preserving the balance and harmony in the world. Conversely, Spaniards considered women to be legal minors without enough reasoning power to make economic or political decisions. Because of Spanish reluctance to settle at higher elevations, I argue that the more inclusive gender roles of the Inca were able to persist at higher elevation levels, providing a potential explanation of the gender gap heterogeneity by elevation.

This argument is built on evidence that colonial characteristics affect current economic outcomes; and that gender roles are sticky and have historical roots. According to this literature, historical agricultural practices, geography and social structures are the main drivers of current gender roles in Europe, Africa, Asia and the United States. However, the recent literature and analysis have largely ignored the Latin American experience.

Peru is an interesting setting to study historical roots in gender roles, as it is the most gender-unequal country in the region with an unexplained gender wage gap of 25% (Nopo, 2014). At the same time, the country was home to one of the most developed pre-colonial cultures on the continent and was later the center of the 16th-century Spanish conquest. Given Peru's large gender wage gap and rich history, it seems natural to consider the relationship between both factors and start exploring the role of history on female LFP.

To test the proposed argument, I rely on a two-step estimation strategy. First, after gathering evidence that Spaniards did not settle in the highest-elevation ecosystem[2], I exploit ecosystem variation as a proxy for Spanish presence. I find that living in a non-Spanish ecosystem is associated with an increase in female LFP of 10 percentage points. These results are

[1] I only focus on labour outcomes when it comes to gender roles. Additional outcomes include fertility, education, competitive attitudes, domestic violence, etc. (see Giuliano (2018) for a detailed review)

[2] Climate-specific geographic area defined by elevation, longitude and latitude with a specific crop production and livestock

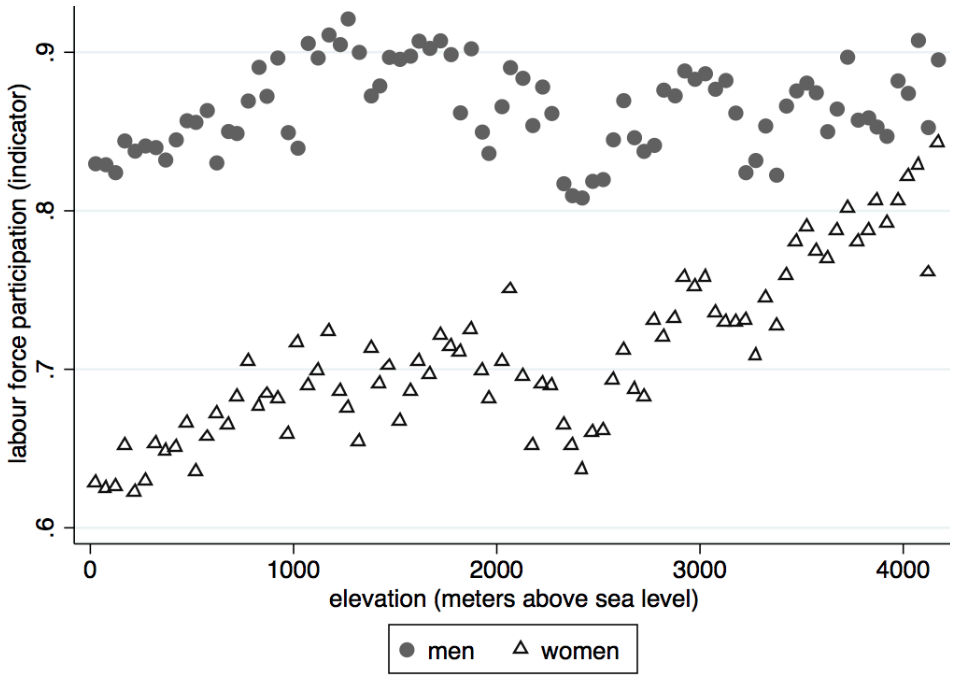
significant and robust to the inclusion of geographical and demographic covariates. However, I cannot claim a causal effect of non-Spanish presence. Doing so requires the assumption that the ecosystem only affects female LFP through Spanish presence. Ecosystems, by definition, have specific climate and agricultural practices, which have also been linked to the historical formation of gender roles. Thus, the exogeneity assumption is not met.

To deal with this problem, in a second step I narrow the analysis to ecosystems with similar agricultural practices, but different levels of exposure to Spanish influence. In this case, my proxy for Spanish exposure is the share of Spanish-speaking population at the district level in 1981. To deal with potential selection in Spanish exposure, I use elevation as an instrument. The identification strategy relies on the assumption that elevation only affects the female LFP through Spanish exposure, conditional on covariates. This is a more reasonable assumption, given that controls for agricultural practices (the main confounding factor identified in the literature) are included in the analysis. OLS estimates show that an increase in the share of Spanish-speaking population in the district by 1 percentage points is associated with a significant decrease in female LFP of 9 percentage points. 2SLS estimates remain negative and highly significant as in the OLS estimates. Results are robust to the inclusion of a variety of geographical, contemporary and historical covariates.

To my knowledge this is the first study that reports a negative effect of Spanish gender values on current female LFP. Although this study focuses on Peru, it opens the question of how other Latin American countries may have also been influenced by Spanish gender values.

The rest of the study is organized as follows. The next section explains the conceptual framework of the historical origins of gender roles. Section III explains the relationship between Inca and Spanish gender values and elevation. Section IV describes the first empirical strategy using variation in ecosystems. Subsections include data, methodology and OLS results. Section V describes the second empirical strategy using elevation as an instrument for Spanish exposure. Subsections include data, methodology, OLS and 2SLS results, and robustness checks. Section VI concludes.

Figure 1: Labour-force participation across elevation levels



2. CONCEPTUAL FRAMEWORK

Understanding why women work less than men outside the home has been one of the economists' great obsessions for the last 50 years. A key part of the puzzle are the large differences across countries. In Iceland, for instance, 80% of women work outside the home; whilst in Turkey only 20% does.

A strand of the literature supports the idea of economic factors being the main drivers of these differences – according to them, female labour supply increased with the rise of female-intensive economic sectors, the diffusion of technology that allowed housework to become less time-consuming and the availability of contraception methods (Fernández, 2007). Goldin and Katz (2002) document how the introduction of the pill in the US increased female educational attainment and labour supply by reducing the opportunity cost of engaging in long-term career investments. Attanasio et al. (2008) explore how differences in child care prices explain differences in female labour participation. Albanesi and Olivetti (2016) do the same for the period following the invention of the baby formula.

A second strand of literature argues that differences are driven by cultural beliefs about the appropriate role of women in society – the so-called cultural gender roles. It is not the market pushing women in or out of the labour markets, it is the society telling them that it is adequate (or not) to work outside the home (see Bertrand (2010) for a detailed review). Fernández and

Fogli (2009), for instance, explore work and fertility behaviour of US-born women with parents born in another country; and found a positive correlation between female labour supply and historic labour supply of the parent's country of birth (a proxy for culture). Using data from the World Values Survey, Fortin (2005) shows that the beliefs about women being housewives and men being breadwinners strongly correlates with female labour outcomes.

The cultural hypothesis is appealing because data also show large differences in beliefs about the appropriate role of women in society across countries. Taking the previous example of Iceland and Turkey, the Values Survey report that only 3.6% of Iceland's population believe that men have more right than women to work outside the home. In Turkey, 49.8% of the population believes so. This difference in beliefs could be behind the mentioned difference in female LFP.

Unlike economic factors that change rapidly, cultural beliefs tend to be persistent. As such, these must have historical roots; tracing the historical origins of gender roles has become the interest of economists during the past decade.

The first and most-studied historical roots are agricultural practices. Alessina et al. (2013) hypothesize differences in gender roles arise from differences in historical agricultural technology, in particular, the different tasks required for shifting and plowing agriculture. They claim farming with plough requires more upper-body strength than hoe-farming. The plough led to a specialization in agriculture where men oversaw the agricultural work outside the home whilst women stayed. This division fostered a belief about the appropriate role of women – inside the home. The authors find evidence to support their hypothesis: today women living in plough-societies work less outside the home (when economy has moved out of agriculture).

Following the same line, Hansen et al. (2015) hypothesize that societies with *long* histories of agriculture have more male-centric values today. Their idea is that patriarchy values originated in the Neolithic transition from gathering-hunting to agriculture created a premium on male strength. As with the plough, the situation led to a division of labor which confined women to the house. Using data for European countries and the U.S., authors find a negative correlation between length of agriculture period and current female labour supply. Using data from China, Xue (2018) proves a similar point for cotton-weaving societies, a traditionally female activity. She finds that individuals living in historically better-suited-for-cotton-weaving areas are less likely to believe men are superior to women.

Geography is another important long-term determinant of gender roles. Carranza (2014), for example, reports that soil texture determines gender roles through the strength needed for land preparation in India. She finds that loamy soil, as opposed to clayey, needs less female activities (transplanting, fertilizing, and weeding) and hypothesizes that this creates the well-known labour division. She finds that the women work less in regions with loamy soil and documents that this effect has been stable at least since 1961.

Literature also documents long-term effects of social structures on gender roles. The most studied structures are the alternative types of settlement (sedentary or nomadic), the living arrangements after marriage (matrilocality or virilocality) and modes of marriage (dowry or bride price). Alesina et al. (2016) do a comprehensive study on these factors in Africa[3]. They link current violence against women – also a strong predictor of gender roles– and pre-historical characteristics in Africa. Regarding types of settlement, they find that women whose ancestors were nomads experience and justify more violence today. Their hypothesis is that developing social structures of protection was harder in a changing environment. In the same line, women whose ancestors paid the bride-price have a lower probability of being exposed to violence today. Their idea is men valued their wives more highly when they had to pay for them.

There is little doubt about the role of historical traits on current gender roles, but there is still the question as to why 500-year-old beliefs still matter today. The institutional view argues cultural beliefs are reinforced by policies, laws and institutions. A society with traditional beliefs about gender inequality may perpetuate them by institutionalizing unequal property or voting rights (Giuliano, 2018). On the other side, the sticky-culture view believes cultural beliefs are inherently sticky. These beliefs become easy-to-follow rules-of-thumb which are useful to avoid processing information every time, and thus are hard to change (Alessina et al., 2013). As Peru is a unitary and highly-centralized State where policies are decided at the national level, I argue that the stickiness-of-culture view makes more sense in this scenario.

3. ON THE ORIGINS OF PERUVIAN GENDER ROLES

Figure 1 revealed women living at high-elevation level work more outside the home than their low-elevation counterparts. The relevant question is why is that. As traditional economic factors fall short while explaining female labour outcomes in Peru, I dig into the history of the most important gender-related sea-change event of Peruvian history, the Spanish Conquest. However, to understand the *change*, it is important to understand the *before*-conquest context.

3.1. *The Inca Gender Roles*

Before the Spaniards, the Incas ruled in the Peruvian territory. They were the last and greatest high-developed pre-Colonial culture in South America (Hyslop, 2014). Their influence encompassed current territory of Argentina, Bolivia, Chile, Colombia, Ecuador and Peru. The capital, Cusco, was located in Peru. Agriculture and livestock were the Incas' main economic activities. The economy was organized through the reciprocity principle as they did not have markets nor currency (Pease, 2007). Within a community, families produced different

[3] See also Ashraf et al. (Forthcoming), Levine and Kevane (2003), Jayachandran (2015), Lowes (2018)

crops to exchange with the others. Labour force was also distributed following the reciprocity principle. For example, if a family needed extra labour force for the harvest, their relatives were obliged to help. But they expected the same help in return (Vieira Powers, 2000). For this system to work, everyone had to participate which made labour one of the most essential parts of the community life. This included women who were considered a critical productive labour force.

In this environment, women were "partners of men in the business of life" (Vieira Powers, 2006). The Incas believed in the existence of labour complementarity between men and women which meant that activities performed by men or women were equally important. Furthermore, for certain activities, gender complementarity meant gender interdependence. It was impossible to perform certain activities without the other (Silverblatt, 1978). Labour complementarity came from a broader Inca belief of gender duality which conceived men and women as entities living in separate but equally powerful dimensions of the world (Kellogg, 2005). One could not exist without the other, as it was the only way to preserve the balance and harmony in the world. The basic economic unit of the Inca society was the couple, the "male-female labour team" (Vieira Powers, 2000). A (male or female) individual could not be taxed if he/she was not married because he/she was incomplete to be a productive member of the society (Silverblatt, 1978).

Women's labour was not confined to the private sphere. It did involve typical household-related activities such as cooking but those activities were not performed to serve their husbands and children. Women distributed their food to other households of the community as a way to reinforce the ties. Furthermore, women performed important activities outside the home. They were the weavers, brewers, traders, healers and agronomists of the society. Weaving was extremely important for the Inca society because, without written language, women's textiles recorded episodes that range from the household life to the critical events of one year (Silverblatt, 1978). Brewing the corn beer (*chicha*) was also an important female activity because *chicha* was a key ingredient for religious practices. Women also controlled the household economic surplus and led the inter-community trade.

Gender complementarity was most obvious in agriculture where women played the role of geneticist and agronomists by choosing and planting seeds (Silverblatt, 1978). Agricultural knowledge passed from generation to generation only through the line of women. They also worked the soil, oversaw the quality of crops and led the harvest. Their role was based on the religious belief that women were descendants of the earth mother (*pachamama*), the goddess related to fertility. This relationship was most obvious during the planting: women would reverence to the earth mother, offer the corn beer and ask for a productive harvest while placing the seeds in the soil. Gender complementarity implied men also played a fundamental role: breaking the soil was the male activity. Much like women with fertility, men were associated with the force god (Silverblatt, 1978).

Gender duality had implications in other dimensions of life such as access to economic resources and religious/political power. Women's inheritable resources (e.g. land, livestock) were transmitted from mothers to daughters through the female line (Vieira Powers, 2000). Having their own resources guaranteed economic independence from men. Women also had their own religious and political organizations with their female hierarchies of priestesses and officials. This was the gender panorama that Spanish conquerors found during the conquest of Peru in 1532.

3.2. *The Spanish Gender Roles*

Spanish gender roles were the opposite of Inca gender roles. Spaniards considered women as legal minors without enough reasoning power to make any economic, political or religious decision. They needed a male tutor to respond for them in any legal or economic transaction – tutors included fathers, husbands or siblings (Cook, 1998). Women could not be involved in any legal or economic transactions. Priests and political officials of the time were also positions reserved for men. These male-centric beliefs were related to Christianity, a monotheistic religion with *one* male omnipotent god, which praised a high value on female virginity (Vieira Powers, 2000).

It is not surprising Spaniards invalidated the gender-dualism and established a patriarchal male-centric society in the new world. New laws for the colonized Americas considered women as legal minors, as in Spain. With this, the old dual political and religious hierarchies disappeared. Women were removed from their positions of religious and political power (Vieira Powers, 2000).

Economic and labour dynamics also changed during this period. As the primary objective of the Spanish Crown was to obtain profits, conquerors established markets for extracting and exporting the resources to Europe – gold, silver, wine and wheat were the more common. For this colonial system to work, Spaniards made indigenous people work for free as a tribute payment. To organize this new labour force, gender roles played a crucial role. Spaniards believed women should stay at home, focusing on the reproduction, childbirth, and care of family and household; while men should work outside the home because they had the abilities to perform the skilled occupations. Thus, women were banned from prestigious occupations such as professional weaving or craft producing, which had been female activities for more than 500 years. Furthermore, women lost access and ownership of their economic resources; the structure of mother-daughter inheritance disappeared from one day to the other as female resources were transferred to the male tutors (Vieira Powers, 2000).

It is important to remark that the Spanish institutions not only deteriorated the indigenous women status, but also empowered indigenous men (Vieira Powers, 2000). Patriarchal attitudes became common amongst indigenous men to the point that historians have identified a "double jeopardy" of colonization on women (Deeds, 1977) – women were conquered twice, first by the Spaniards and then by the indigenous men. It is straightforward to assume that this new male-centric society established by the Spaniards deteriorated the average Peruvian's beliefs about the appropriate role of women in society. But Spaniards did not conquered all areas within Peru with the same "intensity". They were not interested in every area of the country equally.

3.3. *The role of elevation*

As opposed to the Incas which were an Andean society[4], Spaniards settled down in low-elevation areas. The creation of markets for the sea export of silver and other goods to Spain changed the geographic distribution of the country by enhancing the importance of coastal cities (Glave, 1989). Some goods were still produced or extracted in high-elevation areas such as the mining products; but the religious, political and economic power was based on the coast. As Spaniards were the only ones who could hold these positions of power, they settled on the coast (Glave, 1989). Just 3 years after the conquest, the capital switched from Cusco to Lima, home to the biggest port of the country. Besides economic and political reasons, it is well-documented that conquerors settled down in areas with similar geographic conditions to those at home to prevent the acquisition of unknown diseases (Acemoglu et al., 2000). Spain's average altitude is 660 MASL and the city with the highest altitude is located at 1692 MASL. Peru's average altitude is 1555 MASL and the highest altitude city is located at 4380 MASL.

Even though there is no historical evidence of a discrete elevation cut-off above which there is no Spanish influence, there is evidence about indigenous escaping from the Spaniards and hiding in the *puna*. The *puna* is the highest fertile Peruvian ecological region located between 3400 and 4300 MASL[5]. At such elevation, the climate is cold and cultivable crops only include Inca-native tubers such as the potato, oca and ulluco. For the indigenous, going there meant cultural defiance. It meant going back to their origins where only they knew how to live. In the puna, they rejected the colonial ideology and religion, and returned to their native religion (Silverblatt, 1987). The latter implies a return to the gender values.

If areas above 3400 MASL were used to escape from the Spaniards, it is safe to assume that those were areas with non-Spanish presence. That would also imply that Inca values of gender dualism had more chances to survive at higher elevations – especially areas above the *puna* cut-off. Therefore, a first approach to exploring the effects of the Spanish influence on the current gender roles would be to compare women living in the *puna* versus women living outside the *puna*. Next section presents these results.

[4]Incas' most important cities were located above the 1500 MASL where land was more fertile. The Peruvian coast is a dessert

[5]Ecological regions are ecosystem regions with specific climate

4. THE *PUNA* AND GENDER ROLES

Drawing from the historical sources, the *puna* seems to have been a resistance area which defended the Inca values of gender duality. In this section, I test whether women currently living in the *puna* work more outside the home than their counterparts in non-*puna* areas.

4.1. *Data*

This section's data set crosses the geographic distribution of the ecoregions with current female labour outcomes. First, to obtain the geographic distribution of the ecological regions across Peru, I use the Ministry of Environment of Peru (MINAM)'s map of 8 ecological regions. The map is a shapefile that divides Peru into polygons, each showing the location of an ecoregion (see 2). Although the map is a recent database (released in 2018), the ecoregions have maintained through time, as ensure by MINAM.

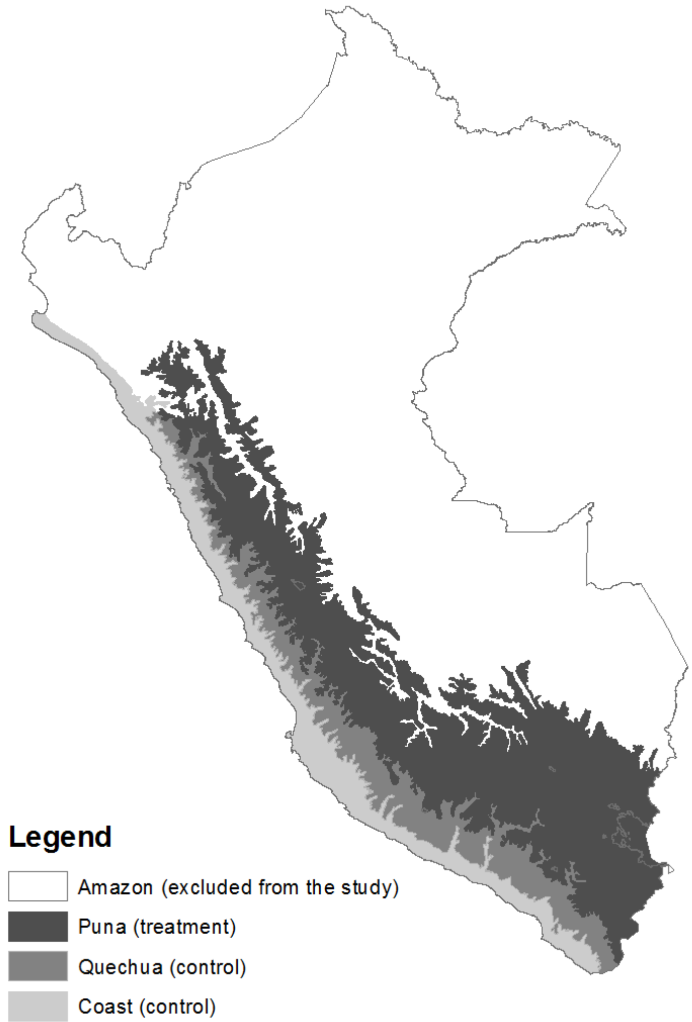
Second, to identify households living on each ecoregion, I use the National Household Survey (ENAHO) of Peru from 2007 to 2017. This is a national representative household survey on living conditions carried by the National Institute of Statistics and Informatics of Peru (INEI). Sampling is done for households at the regional level – first level of administrative division. I use a pooled cross-section data set because parts of the analysis are performed at a more disaggregated level, the district (1850 districts in 2017). Pooling allows me to gain observations and representativeness. I use data since 2007 because it is the first year in which the survey includes the latitude and longitude of the household's block; that means that I can geo-reference it. ENAHO's main modules include labour, income, expenditure, health status and health care, housing, and demographics. I focus on the labour module for my main outcome variables which collects data for individuals older than 15 years old (legal age for working in Peru). This narrows my sample.

Table 1 shows sample size according to the control variables used. First column shows the number of individuals for which I have data on labour outcomes for each year. Total sample is 833,417 individuals for the 11 years. From these individuals, I have information on labour force participation and real monthly wage. The first variable, labour supply, captures whether the individual works outside the home or not. ENAHO explicitly excludes housework. As I am dealing with a pooled data set, the monthly wage is adjusted by inflation and expressed in 2007 Peruvian soles. I also extract other socio-demographic controls from ENAHO such as main language spoken by the individual, age, educational attainment and relation to the household head. Peru has three official languages; Spanish, Quechua and Aymara. Quechua and Aymara are the native languages; originally spoken by the Incas, the former was the official language in the northern Andes and the latter in the southern Andes.

Third, dealing with spatial data demands for geographic controls to account for potential confounders (i.e. geographic variables that correlate with the ecoregions and may also correlate with female labour outcomes). Specifically, I include variable of elevation, slope and distance to rives. To obtain elevation, I use the Shuttle Radar Topography Mission (SRTM) of the NASA, which reports data on the world's elevation for 30 arc-second by 30 arc-second (approximately 1 km by 1 1km) grid cells. Elevation data is enough to further construct slope data at the same resolution quality. I then cross the SRTM data with ENAHO's households geographic location and obtain the elevation and slope of their blocks. Elevation is measured in meters above the sea level (MASL) and slope is measured in degrees at the block level. To obtain the smallest distance to a river, I use the Hydrographic Map of the National Water Authority of Peru. This map collects the location of every river in the country. I construct a variable which accounts for the smallest distance from the household's block to a major river. This variable is measured in meters and accounts for the geography of the terrain. Second column from Table 1 shows the narrowing of the sample after including these geographic controls. It is a sample of 823,413 individuals.

Finally, I further narrow my sample by excluding the individuals living in the Peruvian jungle, the Amazon. This area was not part of the Inca Empire nor was conquered by the Spaniards. Thus the Amazon does not contribute with information to answer this study's relevant questions. Column 3 of Table 1 shows the final sample of this study of 594,197 individuals.

Figure 2: Map of Ecoregions



Note: Figure shows MINAM's map excluding the Amazon/jungle ecoregions.
Source: Ministry of Environment of Peru.

4.2. *Empirical Strategy*

In this subsection, I first explain my main regression which tests whether it is true that women currently living in the *puna* work more than their counterparts living in the non-*puna* ecoregions; then I describe the estimation and identification strategy. The main equation regresses the labour outcome of interest on the interaction between the *puna* and female indicators:

Table 1: Sample size, first section

Year	Individuals (1)	(2)	(3)
2007	63,269	62,506	44,621
2008	61,657	60,955	43,255
2009	62,582	62,030	43,915
2010	61,695	61,016	43,061
2011	70,818	69,891	50,095
2012	71,171	70,371	50,603
2013	85,393	84,291	61,193
2014	85,425	84,340	61,260
2015	85,755	84,700	61,347
2016	94,958	93,726	69,264
2017	90,694	89,587	65,583
Total	833,417	823,413	594,197
Labour outcomes	yes	yes	yes
GIS controls	no	yes	yes
Exclusion of Amazon	no	no	yes

Data Source: Peruvian National Household Surveys 2007-2017.

$$y_i = \beta_0 + \beta_1 \text{ puna}_b + \beta_2 \text{ puna}_b f_i + \beta_3 f_i + \mathbf{X}_b^G \Gamma + f_i \otimes \mathbf{X}_b^G \Pi + \varepsilon_i \tag{1}$$

where i denotes individual and b denotes the block in which the individual lives; y_i is the labour outcome of interest (labour force participation or real monthly wage); puna_b is an indicator variable that equals 1 if i lives in a block b located within the puna ecoregion and 0 otherwise; f_i is a gender indicator variable that equals 1 if i is female and 0 otherwise; $\text{puna}_b f_i$ accounts for the interaction of the puna and gender indicators, it equals 1 if i lives in a b located within the puna ecoregion and is female. $\mathbf{X}_b^G$ and $f_i \otimes \mathbf{X}_b^G$ are vectors of the block's geographic controls and its interaction with the gender indicator respectively.

The parameter of interest is β_2. It captures the total effect of being a female living in the *puna* and her labour outcomes, which surely includes the effect of living in an area where the Spanish gender values did not expand. I must note that although the parameter of interest capture lower exposure to Spanish values, it is also capturing the effect of potential confounders correlated to the *puna* and the female labour outcomes. Some of them can be accounted through geographic controls but other will remain a problem. I further explore this issue in the next subsection

The geographic control variables $\mathbf{X}_b^G$ capture differences in geographical features across blocks that are located within the *puna* and those which are not. These are important because a potential concern is that the estimated differential effect of being a woman living in the *puna* on the labour outcomes (β_2) may be driven by other geographical features. It may not be living in the *puna* itself but living in a more sloppy area or being closer to a river. However, similarly to Nunn and Puga (2012), I only care about potential omitted variables that correlate both with *puna* and the labour outcomes if they affect women and men differently. To deal with this issue, I add both the geographic feature and its interaction with the female indicator as controls; $f_i \otimes \mathbf{X}_b^G$ is the matrix of geographic features with the female indicator and Π is the vector of their respective coefficients.

Estimation and identification Equation (1) is estimated through OLS. As mentioned above, β_2 captures the relationship between being a female living in the *puna* and the labour outcomes. For it to capture the causal effect of less-intense Spanish influence, there is one *key* identifying assumption. There must not be any omitted variables which correlate both with $\text{puna}_b f_i$ and y_i besides those included as covariates (i.e. no relevant confounders are driving the results). The interaction of being female and living in the *puna* must be as good as randomly assigned conditioned on observables – $\text{E}(\varepsilon_i | \text{puna}_b f_i) = 0$. This assumption implies that, besides conquest, no other omitted variable affected gender roles (and in consequence the female labour outcomes) differentiated by ecoregion.

The ideal experimental setting to ensure causality would be to randomly assign comparable households (in the observable and unobservable characteristics related to gender equality) to the *puna* and the non-*puna* ecoregions, just before the conquest. Such approach is not feasible. A second-best approach would be to explore if pre-determined characteristics related to gender roles of those living in the *puna* and the non-*puna* before the conquest are balanced.

Unfortunately, I do not have that data access. However, I can draw on historical sources to understand if characteristics related to gender roles were similar in *puna* and non-*puna* ecoregions. As mentioned before, historical formation of gender roles are associated with agricultural practices and technology (Giuliano, 2018). This imposes a challenge to my identifying assumption because a change of ecoregion implies a change in ecosystem, which has a specific climate and crop production. It is plausible that other agricultural-related characteristics of the *puna* may have also affected gender roles.

As has been already explained, men and women were equally important parts of the agricultural cycle because each one had a particular job to fulfill. Due to religious beliefs, women were the chosen ones when it came to touching the soil to promote fertility, thus they prepared the soil, place the seeds and did the harvest. Men broke (but did not touch) the soil using two different tools depending on the soil type – the hoe or the foot-plow (*taclla*). The hoe was used in the valley/low-elevation soils where plots were farmed continuously. It was easy to use and required the work of one man (Morlon et al., 1992). Conversely, the *taclla* was a more-advanced native instrument used to dig up the tough and compacted land of periodically-rested land. The *taclla* was used to sow native tubers such as the potato, the crops of the *puna* (Gade and Rios, 1972).

The *taclla* differs from most agricultural tools as it requires a three-people team because of the heaviness of the task; usually, a man and two women are involved (Gade and Rios, 1972). The plough-man locates in the center and breaks the soil while the two women (or a woman and a child) turn the clods of soil aside (see Figure 3). Setting gender-mixed teams was the efficient way of dealing with a heavy task and respecting the religious beliefs at the same time. This implies women had more agricultural work outside the home in areas that require the use of *taclla*. As *taclla* is associated with the *puna* ecoregion crops, β_2 could be also capturing the effect of women having a more important role in agriculture in those areas before the arrival of the Spaniards. Thus I cannot claim it is capturing causal effects of non-Spanish presence on female labour outcomes. It identifies a combined effect of both.

Figure 3: Use of *taclla*

Note: Figure shows one man and two women sowing the potato using a *taclla*.
Source: (Gade and Rios, 1972)

4.3. *Results*

4.3.1. *OLS estimates*

Column 1 of Table 2 reports the baseline estimates of the *puna* effect on the probability of working, and its differential effect for women. The coefficient estimate for living in the *puna* interacted with the female indicator is positive and statistically significant at the 1% level. Being a woman living in the *puna* increases the probability of working outside the home by 10 percentage points (pp). This regression includes year and region fixed effects. Year fixed effects are indicator variables for each year of the pooled data set taking 2007 as the base. Region fixed effects are also indicator variables for each of the 25 regions of the country.

The coefficient estimate is statistically significant and similar in magnitude after controlling for contemporary demographic controls and the geographic controls as shown in columns 2-5. These demographic controls include typical Mincer equation variables of the individual – age and schooling. And the geographic controls include slope and the smallest distance to a river. To ensure that I am actually capturing the *puna* effect and not the effect of other geographic characteristic, I include the interaction of the geographical variables and the female indicator[6]. These estimates are not only statistically significant but also economically relevant. Being a women living in the *puna* implies an increase of 6-10 p.p. in the probability of working outside the home which accounts for 0.13 and 0.22 standards deviations of the labour supply. More importantly, as being female reduces the labour supply in about 20 p.p., being female and living in the *puna* counteracts almost half of this negative effect.

Table 8 of Appendix 7.1 also shows a positive and significant association between *puna* and the real monthly wage. With the caveat of sample selection in mind, this regression is important because it supports the hypothesis about female outcomes improving. Not only female work outside the improves but these women also earn more than their counterparts in non-*puna* regions.

[6] I do not include elevation because an ecoregion is defined precisely by longitude, latitude and elevation. Controlling directly for elevation would take the variation that comes from the correlation between labour supply and elevation to identify β_2

Table 2: OLS estimates: *Puna* and female labour force participation

	Dependent variable:				
	Labour force participation, 2007-2017				
	(1)	(2)	(3)	(4)	(5)
Mean of dep. var.	0.709	0.709	0.709	0.709	0.709
(Std. Dev.)	0.45	0.45	0.45	0.45	0.45
Puna x Female	0.101***	0.106***	0.107***	0.106***	0.062***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.010)
Puna	0.040***	0.007	-0.005	0.006	-0.048***
	(0.008)	(0.014)	(0.013)	(0.014)	(0.012)
Female	-0.197***	-0.193***	-0.192***	-0.193***	-0.201***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Baseline controls:					
Year fixed effects		yes	yes	yes	yes
Region fixed effects		yes	yes	yes	yes
Demographic controls		yes	yes	yes	yes
GIS controls:					
Slope			yes		yes
Slope x Female			yes		yes
Distance to river				yes	yes
Dist. river x Female				yes	yes
No. observations	594,355	594,197	594,197	594,197	594,197
R^2	0.044	0.051	0.052	0.051	0.054

Note: OLS estimates are reported with robust standard errors in parenthesis. Regressions are at the individual level. "Labour force participation" is an indicator variable that equals one if the individual works outside the home and zero otherwise. The mean of this variable is 0.679. "Puna" is an indicator variables that equals 1 if the household's block is located within the *puna* ecoregion. "Year and region fixed effects" include fixed effects for the 25 regions and for the 11 years (2007-2017). "Demographic controls" include the typical Mincer Equation controls, schooling and age, of the individual. Regarding GIS controls, "elevation (x100)" reports the elevation of the household's block measured in hundreds of meters above the sea level. "Slope" reports the slope of the household's block measured in degrees. "Distance to river" reports the smallest distance from the household's block to the nearest river. "Robust standard errors clustered at the district level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

5. THE SPANISH INFLUENCE AND GENDER ROLES

The previous section establishes the existence of a female-*puna* effect on the female labour outcomes – that is, women living in the puna ecoregion work more outside the home today. However, I could not claim that it is due to a non-Spanish presence in the *pua* because there was another potential mechanism, the agricultural Inca practices (use of the *taclla*). In this section I further explore the long-term effects of the Spanish influence on the current gender roles (measured as female labour supply) by "eliminating" the influence of Inca agricultural practices. To do so, I only concentrate on the ecoregions where the Inca agricultural practices were homogeneous and there were different levels of exposure to Spanish influence, i.e. within non-*puna* ecoregions where *taclla* was not needed.

5.1. *Data*

This section draws from the same data as the previous section; I still need information on the household labour outcomes (pooled data set ENAHO from 2007 to 2017), the ecoregions geographic distribution and the geographic controls. First, this section excludes households living in the *puna* because it is impossible to disentangle the Incas agricultural practices and the non-Spanish presence there. Column 4 of Table 3 shows the final sample for this section. Additionally, I need variation of the Spanish influence across towns/cities located within the non-*puna* ecoregion. The ideal data set would include a map of the distribution of the Spaniards during the conquest before the independence of Peru in 1821. With it, I could obtain the proportion of Spaniards as a direct proxy of their influence. The more Spaniards, the more its influence on the cultural values. Unfortunately, I do not have access to this historical data set.

To measure the Spanish cultural influence, I use the distribution of Spanish-speaking population at the district level in 1981 as proxy. Language is a core value of a group's culture which implies a common sensory experience, a common classification and understanding the world (values, customs, faith) (Edwards, 1985). Based on such definition, if the Spanish language was more widespread in certain district, then the Spanish cultural influence was also more widespread in the same district. And that includes the Spanish beliefs about gender roles.

The share of Spanish-speaking population at the district level comes from the Peruvian 1981 Population Census carried by the INEI. It includes the the total number of Spanish, *Quechua* and *Aymara* speakers at the district level. The latter two are native languages. I create a variable of the share of Spanish-speaking population at the district level that ranges from 0 to 1. The districts are the smaller administrative divisions in Peru. These were defined after the conquest by the Spaniards; before them, there were no clear administrative boundaries, communities/extended-families (Incas called them *ayllus*) divided the population. Actually, as communities had households at different elevation floors, using these as treatment to identify the *puna* effect did not make sense.. As Spaniards established said boundaries and distributed following them, I explore variation at this level.

The oldest distribution of Spanish-speaking population across districts which I have access to is 1981, which has the caveat of not being a perfect measure of the original Spanish influence. The share of Spanish-speaking population has steadily increase in the past 200 years. I use this data under one key assumption: the cumulative growth rate of Spanish-speaking population since 1821 has been homogenous across districts. This implies that districts with the lowest share of Spanish-speaking population in 1981 were also the ones with the lowest share of Spanish-speaking population during the conquest. The assumption seems plausible after taking into account that Peruvian great migration from the Andes to the coastal cities peaked in 1985 after the rise of the terrorist organization Shining Path.

The internal conflict started in the Andes in 1980 and, by 1985, three of the most populated regions were declared emergency zones because of the outrageous number of victims. The attacks led to a great migration which destabilized the ratios of native-language speakers over Spanish-speakers. Andean districts had a higher share of native-language speakers. As the conflict intensified, Andean population migrated to the coastal cities due to the lack of security of their land, the high payments demanded by the terrorists and the anarchy created by the continued murder of local authorities (Delgado Guembes, 1999). Noticeably, these dynamics pushed down the share of Spanish-speaking population in the coastal cities. However the 1981 Census does not capture them as it was collected before the intensification of the terrorist attacks, asumming homogenous cumulative growth of Spanish-speaking population plausible.

Table 3: Sample size, second section

	No. individuals			
Year	(1)	(2)	(3)	(4)
2007	63,269	62,506	44,621	22,892
2008	61,657	60,955	43,255	21,940
2009	62,582	62,030	43,915	22,362
2010	61,695	61,016	43,061	21,925
2011	70,818	69,891	50,095	25,158
2012	71,171	70,371	50,603	25,533
2013	85,393	84,291	61,193	31,404
2014	85,425	84,340	61,260	31,245
2015	85,755	84,700	61,347	31,797
2016	94,958	93,726	69,264	38,162
2017	90,694	89,587	65,583	35,907
Total	833,417	823,413	594,197	308,325
Labour outcomes	yes	yes	yes	yes
GIS controls	no	yes	yes	yes
Exclusion of Amazon	no	no	yes	yes
Exclusion of the *puna*	no	no	no	yes

Data Source: Peruvian National Household Surveys 2007-2017.

5.2. *Empirical Strategy*

This subsection first explains the main regression which tests whether the female labour supply is higher in districts with less Spanish influence. Then I describe the estimation and discuss the identification assumptions. The basic regression follows:

$$y_i = \alpha_0 + \alpha_1 \text{ spain}_d + \alpha_2 \text{ spain}_d f_i + \mathbf{X}_d^G \Theta + f_i \otimes \mathbf{X}_d^G \Phi + \upsilon_i \tag{2}$$

where i denotes individual and d denotes the district in which the individual lives; y_i is the labour outcome of interest; spain$_d$ is the share of Spanish-speaking population in 1981 in the district d that ranges from 0 to 1; f_i is the gender indicator variable that equals 1 if i is female and 0 otherwise; spain$_d f_i$ is the interaction between the share of Spanish-speaking population in the district and the gender indicator. $\mathbf{X}_d^G$ and $f_i \otimes \mathbf{X}_d^G$ are vectors of the districts's geographic covariates and its interaction with the gender indicator respectively.

The parameter of interest is α_2 which captures the differentiated effect by gender of the Spanish influence on the labour outcome of interest under proper assumptions. As before, the geographic covariates $\mathbf{X}_d^G$ capture differences in geographical features across districts; however the source of concern is the omission of relevant geographical features which affect the labour supply differently by gender. That is why I include the interaction of the covariates with the female indicator, $f_i \otimes \mathbf{X}_d^G$, which allows the covariates to differ by gender.

Estimation and identification The simplest identification strategy would be to estimate equation 2 through OLS where α_2 would have a causal interpretation if the conditional independence assumption (CIA) holds; namely, if the interaction spain$_d f_i$ is as good as randomly assigned – $E(\upsilon_i | \text{spain}_d f_i) = 0$. In words, the unobserved female-labour-supply-relevant variables must be uncorrelated with the Spanish settlement proxy. Similarly as before, the ideal experiment in this setting would be to randomly assign a different number of Spaniards to districts which originally had the same gender roles and compare the female labour outcomes years later. Such an experiment could be approximated if the pre-conquest variables related to the female LFP were balanced across districts. However, evidence suggests it may not be the case.

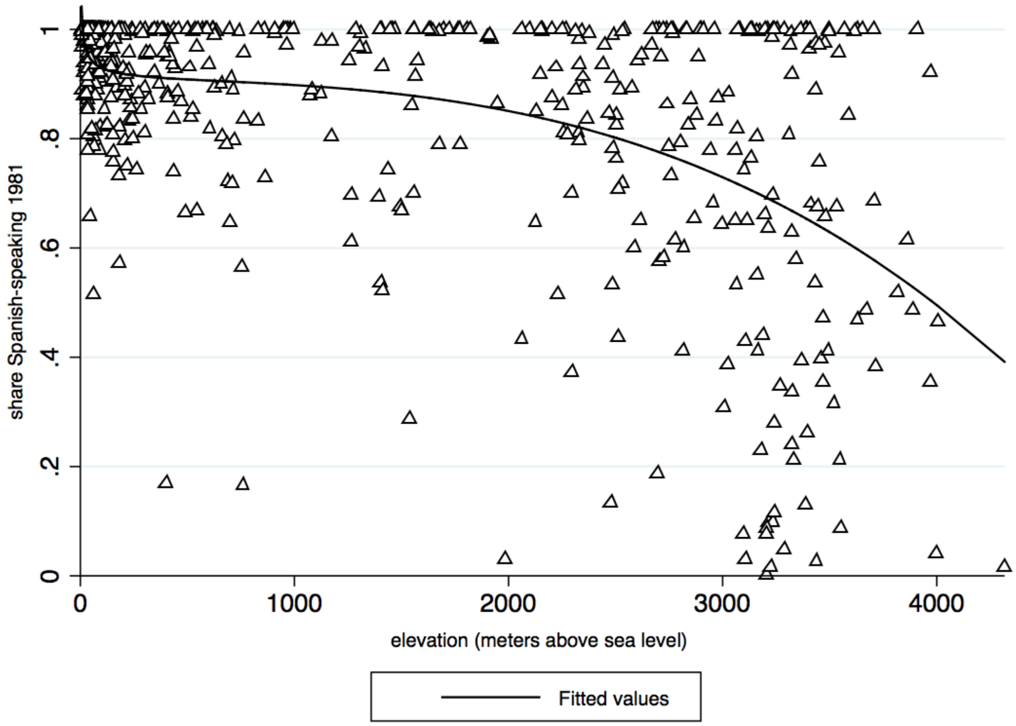
There is evidence that Spanish settlers were more likely to migrate to richer areas (Acemoglu et al., 2000). If Spaniards decided to settle down in richer areas where women were more educated, then women today could also be more educated and work more. The channel would have been female education and not the Spanish influence. This is a possibility. Another scenario where Spanish settlement would yield unbalanced gender beliefs would be if they established using the Inca infrastructure as those had already modified the gender beliefs There is evidence that living closer to the Inca road system increased the female labour supply due to their participation in the inter-community trade. As a first step for dealing with the endogeneity issue, I will include variables for which I do have data as controls (interacted with the female indicator) in my regression.

However, I do not have data on *all* relevant individual characteristics before the conquest. Thus these become omitted variables and potentially bias my coefficient estimates. Furthermore, a direct violation of CIA comes from the construction of the variable of interest – spain_d is not the actual proportion of Spaniards in the districts in 1821, the year of the Independence. Due to data limitations, it has a measurement error which will bias the results by definition. Therefore I need a valid instrument to account for the omitted variables bias.

Spaniards had a very specific objective while distributing their population – to extract and export the goods to Spain. They favored the low-elevation coastal cities due to climate, crops and access to ports; and neglected the high-elevation areas as the *puna*. It is well-documented that the *puna* was not conquered because of its high-elevation. Therefore, the closer to the *puna*, the higher the elevation, the lower the probability of conquest. I will use the elevation as an instrument.

Claiming causal effects requires two conditions from the instrument. First, the relevance condition; the elevation at the district level must have a clear effect on the levels of Spanish influence proxied by the proportion of Spanish-speaking population. Figure 4 plots the share of Spanish-speaking population in the districts in 1981 against the elevation of the district's capital measured in meters above the sea level. It supports the relevance condition. Districts located between 0 and 1000 MASL have 90% of Spanish-speakers on average. Conversely, the average for districts located above 3000 MASL is 20%. Moreover, Table 4 shows the first stage results from the regression of the share of Spanish-speaking population on elevation (in hundreds of MASL). The coefficient on elevation is highly significant (F=90.22) and explains 18.4% of the variation in the proportion of Spanish-speaking population across districts.

Figure 4: Spanish-speaking population and elevation



The second condition is the exclusion restriction; the instrument has to be as good as randomly assigned and only affect the female labour outcomes through the Spanish settlement. In other words, without Spanish Conquest, the current female labour supply would be uncorrelated with elevation in the non-*puna* region. Besides agricultural practices/technology, literature on the formation of historical gender roles points to language, pre-industrial societal characteristics (e.g. matrilineality, bride-price), family structures (e.g. strong/weak family ties) and religion as the main determinants of female labour force participation (see Giuliano (2018) for a comprehensive review).

I do not worry about agricultural practices/technology because individuals living in this section's sub-region only used the hoe for the soil breaking. Regarding language, religion and family structures, these were common for all communities which were part of the Inca Empire during the pre-conquest period. *Quechua* was the Inca common language and the religious gender-dualism based on the Earth Goddess and the Force God was also a common practice in the Empire. Family structures based on gender dualism where women inherited resources from the female line and men inherited the male-line resources were the rule in the Inca communities. These communities were widespread across the whole Peruvian territory but the Amazon – area excluded from this study since the beginning.

Overall, historical evidence seems reassuring but may not be enough to take the exclusion

Table 4: First-stage 2SLS estimates: Spanish settlement and elevation

	Spanish settlement (% of Spanish-speakers in district)	
	1981	2007-2017
	(1)	(2)
Mean dep. var.	0.782	0.850
Elevation (x100)	-0.010***	-0.008***
	(0.001)	(0.001)
Constant	0.907***	0.960***
	(0.019)	(0.012)
No. districts	398	538
R^2	0.184	0.229
F	90.22	159.07

Note: First-stage 2SLS estimates are reported with robust standard errors in parenthesis. Regression are at the district level. "Spanish settlement" is measured as the share of Spanish-speaking population. In the first column it comes from the 1981 Population Census. This is the instrument used in the second stage. In column 2 it comes from the pooled ENAHO 2007-2017. It ranges from 0 to 1. "Elevation" is the mean elevation of the district, measured in hundreds of MASL. *, ** and *** indicate significance at the 10%, 5% and 1% levels.
Data Source: Peruvian Population Censuses 1981-2007, Peruvian National Household Surveys 2007-2017.

restriction for granted. Using elevation as instrument to provide causal estimates of the effect of the Spanish influence on current female labour supply requires a strong assumption: the difference between high-elevation and low-elevation areas is correlated with current gender roles only through its effect on the Spanish settlement. If one is willing to believe that, then IV estimates are informative.

Table 5: OLS estimates: Spanish influence and female labour force participation

	Labour force participation, 2007-2017				
	(1)	(2)	(3)	(4)	(5)
Mean of dep. var.	0.679	0.679	0.679	0.679	0.679
Share of Spanish-speakers x Female	-0.103***	-0.113***	-0.118***	-0.113***	-0.118***
	(0.020)	(0.020)	(0.021)	(0.020)	(0.021)
Share of Spanish-speakers	0.002	0.014	0.017	0.010	0.015
	(0.019)	(0.030)	(0.029)	(0.030)	(0.029)
Female	-0.104***	-0.088***	-0.081***	-0.088***	-0.082***
	(0.018)	(0.018)	(0.020)	(0.019)	(0.021)
Baseline controls:					
Year fixed effects		yes	yes	yes	yes
Region fixed effects		yes	yes	yes	yes
Demographic controls		yes	yes	yes	yes
GIS controls:					
Slope			yes		yes
Slope x Female			yes		yes
Distance to river				yes	yes
Dist. river x Female				yes	yes
No. observations	308,435	308,325	308,325	308,325	308,325
R^2	0.043	0.055	0.056	0.055	0.057

Note: OLS estimates are reported with robust standard errors in parenthesis. Regressions are at the individual level. "Labour force participation" is an indicator variable that equals one if the individual works outside the home and zero otherwise. The mean of this variable is 0.679. "Share of Spanish-speakers" is the estimated proportion of Spanish-speaking population in 1981 in the district. It ranges from 0 to 1. "Year and region FE" include fixed effects for the 25 regions and for the 11 years (2007-2017). "Demographic controls" include the typical Mincer Equation controls: schooling and age. Robust standard errors clustered at the district level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

5.3. *Results*

5.3.1. OLS estimates

Columns 1-5 of Table 5 report OLS estimates of Spanish influence on female labour supply under different specifications controlling for year and region fixed effects, and demographic and geographical controls. Coefficients range from -0.103 to -0.118, are always negative and statistically significant at the 1% level. This implies a higher the share of Spanish-speaking population is associated with a lower the probability of women working outside the home. This is not true for the male population; there is no significant association between an increase in the share of Spanish-speakers and male LFP.

The coefficient associated to the interaction is not only significant but sizable. Using column 5 estimates, I find that increasing the share of Spanish-speaking population by 1 p.p. is associated with a reduction in the female probability of working by around 10 p.p., which equals 17.2% of the sample mean for female labour supply (0.586) and 20.3% of its standard deviation (0.492). Moreover, this coefficient is of same magnitude as the negative effect for being female which ranges between -0.081 and -0.104.

That being said, the R^2 of the regressions is around 0.05 which implies that included variables only explain 5% of the variation in LFP across individuals. This is consistent with the literature of persistence of gender roles. Alessina et al. (2013)'s historical plough differences only explain 6% of differences in female LFP. Xue (2018) historical cotton-weaving production explains between 3 and 5% of gender roles variation. Thus an important conclusion is that historical gender roles can be associated with current differences in female LFP but are clearly not the single most important factor (Burnette, 2018).

Finally, I cannot claim causal effects of the estimates. As mentioned in the previous section, a potential concern with OLS estimates are the omitted variables. It is possible that communities with less equal gender-role attitudes attracted Spanish settlements. A first approach to account for this potential bias is to control for observable characteristics potentially related to gender roles and to Spanish settlement. I do this is the next subsection.

5.3.2. *Robustness checks*

First, Spanish settlement may have been driven by the previous infrastructure, especially the Inca road system. This was a 30000-km-long system of transportation and communication built by the Incas to distribute and exchange their goods across communities (Glave, 2009). There is preliminary evidence of a positive and strong correlation between living closer to the Inca road and current female labour supply and schooling driven by the inter-community trade (Lavado and Franco, 2019). Women had this role during the Inca Empire. Those who lived closer to the road engaged almost exclusively to inter-community trade, spending less time working at home. Given its importance during the conquest when Spaniards used it to get to the capital of the Empire (Glave, 1989), the Spanish-exposure coefficient may be picking the effect of the living closer to the Inca road.

Information on the historical location of the Inca Road system is taken from the Ministry of Culture of Peru. The shape file includes the location of every main and secondary roads part of the system. I compute the air distance (in meters) from the household's block to the nearest point of the Inca Road. Column 2 of 6 shows the coefficient of interest is remains sizable and significant with this additional covariate and its interaction with the female indicator.

Another important determinant of differences in gender roles could be the presence of conflicts. The evidence regarding the direction of the relationship is mixed (Alessina et al., 2013). Conflicts tend to push societies towards vertical and patriarchal structures because women have been traditionally excluded from wars and conflicts. However, women always replace men during these periods. The relative scarcity of men increases the female labour supply.

For example, Goldin (1991) reports an increase in the female labour supply in the U.S. during World War II. The most important conflict fought in Peruvian Andes has been the 20-year-long internal conflict provoked by the rise of the terrorist organization Shining Path in 1980. Several indigenous fought back and three of the most populated regions in the Andes were declared emergency zones because of the attacks. It is a fact that intense Spanish settlement encouraged indigenous rebellions. As these may have fostered stronger Spanish anti-conflict institutions, Spanish settlement and the internal conflict could be inversely associated and thus my estimates would be capturing the effect of living in battlefield areas.

Terrorism data comes from Truth and Reconciliation Commission of Peru (TRC) – a commission created by the Ministry of Women and Vulnerable Populations of Peru to record human rights abuses during the internal conflict period. It has detailed information on the number of victims per geographical regions. I created two variables at the district level. One is a terrorism indicator which takes the value of one if the district had any victim at some point during the conflict. The other variable records the terrorism victims at the district level. Column 3 shows the coefficient estimates while controlling for these covariates and its interaction with the female indicator. Estimate on the interaction of the puna and female indicators continues to be significant and sizable.

Next I turn to the current main economic activity, mining. Mines have particular labour rules regarding women – they cannot participate, it is forbidden. As with everything involving land, Andean indigenous believe mines are related to the earth goddess. They anthropomorphize the mine as a woman's womb. In their words, during the dynamite explosion, the mine shakes like a woman giving birth to what was gestated for many years (Salazar-Soler, 2006). Miners even have a ritual to "baptize" the extracted piece of mineral. This is the reason why women cannot enter the mine – the earth goddess would get jealous. Therefore women living in mining cities have less outside-home working options.

Spanish settlement may have been associated with the location of mines. It is possible that a group of profit-seeker conquerors preferred these regions and settled there. If so, my estimates would be capturing the effect of living in a mining area. To test it, I gather data on the mining production at the district level during the 2007-2017 period from the Ministry of Energy and Mines of Peru. I created two variables, a mining indicator which takes the value of 1 if the district had any mining project during the period; and a mining production variable which sums the total production during the period. The latter is measured in millions of USD. Reassuringly, column 4 of Table 6 reports coefficients remain statistically significant and economically relevant after the inclusion of mining covariates and its interaction with the female indicator.

The last factor I consider are the current agricultural practices. The role of agriculture while shaping gender roles has been widely discussed. It may be true that women work more today because agriculture is the dominant practice in the districts with less exposure to Spanish influence. If it is, then I would be capturing the effect of current agricultural practices. Using the ENAHO dataset, I obtain the share of individuals working in agriculture-related occupations at the district level in the 2007-2017 period. Column 5 shows results are robust to the inclusion of this variable and its interaction with the female indicator. In column 6, I include all covariates in one specification and the Spanish-influence estimate on the female labour supply remains highly significant, only decreasing slightly. Overall, the estimate is robust across specifications; it is always negative, significant and ranges from -0.113 to -0.127.

Table 6: Robustness of OLS estimates: Spanish settlement and female labour force participation

	Dep. variable: Labour force participation, 2007-2017					
	(1)	(2)	(3)	(4)	(5)	(6)
Mean of dep. var.	0.679	0.679	0.679	0.679	0.679	0.679
Spanish settlement x Female	-0.118***	-0.113***	-0.127***	-0.124***	-0.113***	-0.118***
	(0.021)	(0.021)	(0.025)	(0.022)	(0.020)	(0.023)
Spanish settlement	0.015	0.013	0.019	0.023	0.012	0.017
	(0.029)	(0.029)	(0.030)	(0.030)	(0.024)	(0.025)
Female	-0.082***	-0.084***	-0.072***	-0.075***	-0.074***	-0.065***
	(0.021)	(0.020)	(0.024)	(0.021)	(0.020)	(0.022)
Historical controls:						
Distance to Inca Road		yes				yes
Dist. Inca Road x Female		yes				yes
Contemporary controls:						
Terrorism indicator			yes			yes
Terrorism indicator x Female			yes			yes
Terrorism victims			yes			yes
Terrorism victims x Female			yes			yes
Mining indicator				yes		yes
Mining indicator x Female				yes		yes
Mining production				yes		yes
Mining production x Female				yes		yes
Agricultural share					yes	yes
Agricultural share x Female					yes	yes
Baseline controls	yes	yes	yes	yes	yes	yes
No. observations	308,325	308,325	308,325	308,325	308,325	308,325
R^2	0.057	0.057	0.057	0.057	0.061	0.061

Note: OLS estimates are reported with robust standard errors clustered at the district level in parenthesis. Regressions are at the individual level. "Labour force participation" is an indicator variable that equals one if the individual works outside the home and zero otherwise. The mean of this variable is 0.679. "Spanish settlement" is the estimated proportion of Spanish-speaking population in 1981 in the district. It ranges from 0 to 1. "Baseline controls" include region fixed effects for the 25 regions, year fixed effects for the 11 years (2007-2017), the typical Mincer Equation controls– schooling and age, and the GIS controls. "Distance to the Inca Road" is the smallest distance between the household's location and any part of the Inca Road system measured in meters. "Terrorism indicator" equals 1 if the region had more than the median number of terrorism's victims; and the "terrorism victims" reports the number of victims in the region during the conflict. The "mining indicator" equals 1 if the district has had any mining project between 2007 and 2017; and the "mining production" reports the sum of investment in mining projects in millions of USD during the same period of time. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

5.3.3. 2SLS estimates

Reassuringly, OLS estimates are robust but do not account for unobservable confounders – namely, I cannot test if conquerors decided to settle in areas where the gender values were less equal or where women were less educated at the time. This subsection presents results accounting for this.

Table 7 show the second-stage 2SLS estimates of the effect of Spanish influence instrumented by the elevation on the female labour supply. Column 1 and 2 show estimates with and without baseline controls – year and region fixed effects, demographic and geographic controls. The effect is negative and highly significant as in the OLS estimates. Columns 3-5 report different specifications controlling for the robustness-checks variables – the Inca road, the terrorism variables and the mining variables. As with OLS estimates, the effect remains negative and significant at the 1% level.

It is worth noting that magnitudes of these coefficients are consistently higher than the OLS estimates. A potential explanation for this is selection arising from the endogeneity of Spanish settlement. All else equal (e.g. at a same elevation), Spaniards were more likely to establish in more developed communities. These were probably the Inca centers of development which were the more gender-equal because of the Inca values. It does make sense that the values are more rooted in the capital. These are also more likely to be more gender-equal today and have a higher female participation in the labour market. This selection introduces a positive relationship between Spanish settlement and female labour force participation, biasing the negative OLS estimates toward zero.

6. CONCLUDING REMARKS

This study provides evidence that exposure to historical Spanish influence, which is strongly and inversely associated with high-elevation levels, reduces female LFP.

OLS estimates show that a 1 percentage point increase in Spanish exposure, measured by the share of Spanish-speaking population, is associated with a significant decrease in female LFP of 10 percentage points. A coefficient almost the size of the negative effect for the female indicator. These estimates are highly robust to the inclusion of a variety of geographical covariates such as slope and distance to rivers; contemporary covariates such as mining production, share of population occupied in agriculture-related jobs and the exposure to terrorism conflicts; and historical covariates such as closeness to the Inca road system.

2SLS estimates, using elevation as an instrument of the share of Spanish speaking population, confirm OLS results. Coefficients remain negative and highly significant, and are also consistently higher than OLS estimates. A potential explanation for this could be selection due to endogeneous Spanish settlement in more developed communities which were also more gender-equal. These results open the question of how other Latin American countries may have also been influenced by Spanish gender values.

Table 7: Second-stage 2SLS and reduced-form estimates: Spanish settlement and female labour force participation

	Dependent variable: Labour force participation, 2007-2017				
	(1)	(2)	(3)	(4)	(5)
Mean of dep. var.	0.679	0.679	0.679	0.679	0.679
	Panel A. Second-stage 2SLS estimates				
Spanish settlement x Female	-0.149***	-0.256***	-0.260***	-0.311***	-0.319***
	(0.035)	(0.058)	(0.065)	(0.081)	(0.092)
Spanish settlement	-0.170***	-2.118	-2.375	-1.940*	-2.199
	(0.053)	(1.321)	(1.611)	(1.155)	(1.413)
	(0.053)	(1.698)	(1.003)	(2.044)	(2.329)
Female	-0.065**	0.043	0.046	0.105	0.112
	(0.030)	(0.053)	(0.058)	(0.073)	(0.081)
Baseline controls		yes	yes	yes	yes
Historical controls (robust.)			yes		yes
Contemporary controls (robust.)				yes	yes
No. observations	308,435	308,325	308,325	308,325	308,325
	Panel B. Reduced-form estimates				
Elevation (x100) x Female	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Elevation (x100)	-0.174***	-2.081	-2.333	-1.913*	-2.171
	(0.053)	(1.269)	(1.546)	(1.112)	(1.359)
	(0.053)	(1.698)	(1.003)	(2.044)	(2.329)
Female	-0.192***	-0.196***	-0.197***	-0.185***	-0.179***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Baseline controls		yes	yes	yes	yes
Historical controls (robust.)			yes		yes
Contemporary controls (robust.)				yes	yes
No. observations	402,094	402,094	402,094	402,094	402,094
R^2	0.060	0.060	0.061	0.064	0.065

Note: Panel A reports 2SLS estimates using elevation as instrument for Spanish settlement. Panel B reports the reduced-form estimates. Robust standard errors, clustered at the district level, are shown in parenthesis. Regression are at the individual level. "Labour force participation" is an indicator variable that equals one if the individual works outside the home and zero otherwise. The mean of this variable is 0.679. "Spanish settlement" is the estimated proportion of Spanish-speaking population in 1981 in the district. It ranges from 0 to 1. "Elevation (x100)" reports the mean elevation of the district measured in hundreds of meters above the sea level. "Baseline controls" include region fixed effects for the 25 regions, year fixed effects for the 11 years (2007-2017) and the typical Mincer Equation controls: schooling and age. "Historical controls" include the distance to the Inca Road and its interaction with the female indicator. It is the smallest distance between the household's location and any part of the Inca Road system measured in meters. "Contemporary controls" include the terrorism indicator, the mining indicator, the mining production variable and the agricultural share variable; and their interactions with the female indicator. The "terrorism indicator" equals 1 if the region had more than the median number of terrorism's victims; and the "terrorism victims" reports the number of victims in the region during the conflict. The "mining indicator" equals 1 if the district has had any mining project between 2007 and 2017; and the "mining production" reports the sum of investment in mining projects in millions of USD during the same period of time. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

References

ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2000): "The Colonial Origins of Comparative Development: An Empirical Investigation," *The National Bureau of Economic Research.*

ALBANESI, S. AND C. OLIVETTI (2016): "Gender Roles and Medical Progress," *Journal of Political Economy*, 124.

ALESINA, A., B. BRIOISCHI, AND E. LA FERRARA (2016): "Violence Against Women: A Cross-cultural Analysis for Africa," Working paper, Department of Economics, Harvard University.

ALESSINA, A., P. GIULIANO, AND N. NUNN (2013): "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 128.

ASHRAF, N., N. BAU, N. NUNN, AND A. VOENA (Forthcoming): "Bride Price and Female Education," *Journal of Political Economy.*

ATTANASIO, O., H. LOW, AND V. SÁNCHEZ-MARCOS (2008): "Explaining Changes in Female Labor Supply in a Life-Cycle Model," *American Economic Review*, 98, 1517–52.

BERTRAND, M. (2010): "New Perspectives on Gender," *Handbook of Labor Economics*, 4b.

BURNETTE, J. (2018): "Gender in Economic History," .

CARRANZA, E. (2014): "Soil Endowments, Female Labor Force Participation, and the Demographic Deficit of Women in India," *American Economic Journal: Applied Economics*, 6, 197–225.

COOK, N. D. (1998): *Born to Die. Disease and New World Conquest, 1492-1650*, Cambridge: Cambridge University Press.

DEEDS, S. M. (1977): "Double Jeopardy: Indian Women in Jesuit Missions of Nueva Vizcaya," in *Indian Women of Early Mexico*, ed. by S. W. Susan Schroeder and R. Haskett, Norman: University of Oklahoma Press, 255–272.

DELGADO GUEMBES, C. (1999): *El Proceso de Pacificación Interna*, Lima: Instituto Peruano de Economía.

EDWARDS, J. R. (1985): *Language, Society and Identity*, Oxford: Basil Blackwell.

FERNÁNDEZ, R. (2007): "Women, Work, and Culture," *Journal of the European Economic Association*, 2, 305–332.

Fernández, R. and A. Fogli (2009): "Culture: An Empirical Investigation of Beliefs, Work and Fertility," *American Economic Journal: Macroeconomics*, 1, 146–177.

Fortin, N. (2005): "Gender role attitudes and women's labour market outcomes across OECD countries," *Oxford Review of Economic Policy*, 21, 416–438.

Gade, D. W. and R. Rios (1972): "Chaquitaclla: the native footplough and its persistence in Central Andean agriculture," *Tools and Tillage*, 2, 3–15.

Giuliano, P. (2018): "Gender: A Historical Perspective," in *Oxford Handbook on the Economics of Women*, ed. by L. M. A. Susan L. Averett and S. D. Hoffman, New York: Oxford University Press.

Glave, L. M. (1989): *Trajinantes. Caminos indígenas en la sociedad colonial siglos XVI/XVII*, Lima: Instituto de Apoyo Agrario.

——— (2009): "Propiedad de la tierra, agricultura y comercio, 1570-1700: el gran despojo," *Compendio de Historia Económica del Perú. Tomo 2*.

Goldin, C. (1991): "The Role of World War II in the Rise of Women's Employment," *American Economic Review*, September 1991.

Goldin, C. and L. F. Katz (2002): "The Power of the Pilla: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, 110.

Hansen, C. W., P. Jensen, and C. Skovsgaard (2015): "Modern Gender Roles and Agricultural History: The Neolithic Inheritance," *Journal of Economic Growth*, 20, 365–404.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47.

Hyslop, J. (2014): *Qhapaq Ñan*, Lima: Petroperú.

Jayachandran, S. (2015): "The Roots of Gender Inequality in Developing Countries," *Annual Review of Economics*, 7, 63–88.

Kellogg, S. (2005): *Weaving the Past: A history of Latin America indigenous women from the pre-Hispanic period to the present*, New York: Oxford University Press.

Lavado, P. and A. P. Franco (2019): "Long-term effects of the Inca Road," Unpublished, Working paper, Universidad del Pacifico.

Levine, D. and M. Kevane (2003): "Are Investments in Daughters Lower When Daughters Move Away?" *World Development*, 31, 1065–1084.

Lowes, S. (2018): "Matrilineal Kinship and Spousal Cooperation: Evidence from the Matrilineal Belt," Working Paper, Harvard University.

Morlon, P., D. Hervé, and R. Reau (1992): "Una herramienta, un simbolo, un debate : la "chaquitaclla" y su persistencia en la agricultura andina," in *Comprender la agricultura campesina en los Andes Centrales*, ed. by P. Morlon, Lima: Instituto francés de Estudios Andinos.

Nopo, H. R. (2014): *New century, Old Disparities: Gender and Ethnic Earnings Gaps in Latin America and The Caribbean*, Washington: Inter-American Development Bank.

Nunn, N. and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *The Review of Economics and Statistics*, 94, 20–36.

Pease, F. (2007): *Los Incas*, Lima: Universidad Catolica del Peru.

Salazar-Soler, C. (2006): *Supay Muqui, dios del socavón. Vida y mentalidades mineras*, Lima: Fondo Editorial del Congreso del Perú.

Silverblatt, I. (1978): "Andean Women in the Inca Empire," *Feminist Studies*, 4, 36–61.

——— (1987): *Moon, Sun, and Witches. Gender Ideologies and Class in Inca and Colonial Peru*, Princeton, New Jersey: Princeton University Press.

Vieira Powers, K. (2000): "Andeans and Spaniards in the Contact Zone. A Gendered Collision," *American Indian Quarterly*, 24, 511–536.

——— (2006): *Women in the Crucible of Conquest: The gendered genesis of Spanish-American society, 1500-1600*, New Mexico: University of New Mexico Press.

Xue, M. M. (2018): "High-Value Work and the Rise of Women: The Cotton Revolution and Gender Equality in China," Unpublished.

7. APPENDIX

7.1. Wage regressions

Wage and puna Table 8 reports baseline estimates of the *puna* effect on the monthly wage and its differential effect for women. The coefficient estimate for the interaction between living in the *puna* and being female is positive and statistically significant at the 1% level. Columns 1-5 show that the coefficient is robust to the inclusion of the year and region fixed effects, the contemporary demographic controls and the geographic controls as in the labour supply regression. This regression is important because it supports the hypothesis about female outcomes improving. Working more outside the home could be due to necessity. Women could be working both within and outside the house, and could be accepting low-paid jobs. Results indicate the opposite: women work more and earn more. However, there is a caveat to have in mind. This regression has a selection problem as monthly wage data only exists for individuals who work (Heckman, 1979). This implies coefficients reported in columns 1-5 are biased estimates of the random treatment of living in the *puna* on wages. To solve this problem, I would need a source of exogenous variation for the labour supply. Not having one poses a limitation on my wage regression estimates.

Wage and Spanish influence Table 9 reports OLS estimates of Spanish influence on the monthly wage equation under different specifications controlling for year, region, and demographic and geographical controls. In the baseline regression, coefficient has the expected negative direction, an increase in the Spanish settlement reduces female wages by 100 PER soles. However, this estimate loses it significance after controlling for slope and its interaction – there is no relationship between Spanish settlement and female wages. This implies baseline coefficient was capturing the effect of living in a less sloppy area (in which Spaniards settled).

Table 8: OLS estimates: *Puna* and monthly wage

	Dependent variable:				
	Real monthly wage (PER soles), 2007-2017				
	(1)	(2)	(3)	(4)	(5)
Mean of dep. var.	818.46	818.46	818.46	818.46	818.46
(Std. Dev.)	1192.04	1192.04	1192.04	1192.04	1192.04
Puna x Female	117.977***	250.284***	221.239***	252.103***	222.808***
	(21.462)	(25.194)	(28.517)	(25.351)	(28.555)
Puna	-512.829***	-158.307***	-123.607***	-162.212***	-127.075***
	(47.889)	(35.048)	(35.364)	(35.214)	(35.599)
Female	-466.681***	-417.966***	-441.584***	-403.062***	-426.050***
	(15.094)	(13.580)	(13.572)	(15.636)	(15.001)
Baseline controls:					
Year fixed effects		yes	yes	yes	yes
Region fixed effects		yes	yes	yes	yes
Demographic controls		yes	yes	yes	yes
GIS controls:					
Slope			yes		yes
Slope x Female			yes		yes
Distance to river				yes	yes
Dist. river x Female				yes	yes
No. observations	421,320	421,260	421,260	421,260	421,260
R^2	0.069	0.212	0.213	0.212	0.214

Note: OLS estimates are reported with robust standard errors in parenthesis. Regressions are at the individual level. "Monthly wage" is a continuous variable which report the real monthly wage (in 2007 PER soles) of individuals working at the time of the survey. "Puna" is an indicator variable that equals 1 if the household's block is located within the *puna* ecoregion. "Year and region fixed effects" include fixed effects for the 25 regions and for the 11 years (2007-2017). "Demographic controls" include the typical Mincer Equation controls, schooling and age, of the individual. Regarding GIS controls, "elevation (x100)" reports the elevation of the household's block measured in hundreds of meters above the sea level. "Slope" reports the slope of the household's block measured in degrees. "Distance to river" reports the smallest distance from the household's block to the nearest river. "Robust standard errors clustered at the district level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

Table 9: OLS estimates: Spanish influence and monthly wage

	Dependent variable:				
	Real monthly wage (PER soles), 2007-2017				
	(1)	(2)	(3)	(4)	(5)
Mean of dep. var.	989.16	989.10	989.10	989.10	989.10
(Std. Dev.)	1327.18	1327.11	1327.11	1327.11	1327.11
Share Spanish-speakers x Female	79.864	-109.304**	-47.123	-105.337*	-39.159
	(60.343)	(54.477)	(58.313)	(54.837)	(58.891)
Share of Spanish-speakers	-107.119	-33.927	-66.282	-56.923	-93.546
	(130.332)	(144.773)	(148.082)	(149.557)	(153.784)
Female	-527.072***	-312.106***	-395.653***	-300.726***	-385.288***
	(58.394)	(50.939)	(57.224)	(51.595)	(57.198)
Baseline controls:					
Year fixed effects		yes	yes	yes	yes
Region fixed effects		yes	yes	yes	yes
Demographic controls		yes	yes	yes	yes
GIS controls:					
Slope			yes		yes
Slope x Female			yes		yes
Distance to river				yes	yes
Dist. river x Female				yes	yes
No. observations	209,312	209,268	209,268	209,268	209,268
R^2	0.030	0.168	0.170	0.169	0.171

Note: OLS estimates are reported with robust standard errors in parenthesis. Regressions are at the individual level. "Monthly wage" is a continuous variable which report the real monthly wage (in 2007 PER soles) of individuals working at the time of the survey. "Share of Spanish-speakers" is the estimated proportion of Spanish-speaking population in 1981 in the district. It ranges from 0 to 1. "Year and region fixed effects" include fixed effects for the 25 regions and for the 11 years (2007-2017). "Demographic controls" include the typical Mincer Equation controls, schooling and age, of the individual. Regarding GIS controls, "elevation (x100)" reports the elevation of the household's block measured in hundreds of meters above the sea level. "Slope" reports the slope of the household's block measured in degrees. "Distance to river" reports the smallest distance from the household's block to the nearest river. "Robust standard errors clustered at the district level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.