Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 33

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

"ANCESTRY, GENDER AND LANGUAGE - THE ROLE OF CULTURE IN
ECONOMIC DEVELOPMENT"

Presentata da: Angelina Nazarova

<table>
<tr><td>Coordinatore Dottorato</td><td>Supervisore</td></tr>
<tr><td>Maria Bigoni</td><td>Paolo Masella</td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td>Co-supervisore</td></tr>
<tr><td></td><td>Maria Bigoni</td></tr>
</table>

Esame finale anno 2022

To my family

Acknowledgements

Abstract

This thesis is a combination of research questions in development economics and economics of culture, with an emphasis on the role of ancestry, gender and language policies in shaping inequality of opportunities and socio-economic outcomes across different segments of a society.

The first chapter shows both theoretically and empirically that heterogeneity in risk attitudes can be traced to the ethnic origins and ancestral way of living. In particular, I construct a measure of historical nomadism at the ethnicity level and link it to contemporary individual-level data on various proxies of risk attitudes. I exploit exogenous variation in biodiversity to build a novel instrument for nomadism: distance to domestication points. I find that descendants of ethnic groups that historically practiced nomadism (i) are more willing to take risks, (ii) value security less, and (iii) have riskier health behavior.

The second chapter evaluates the nature of a trade-off between the advantages of female labor participation and the positive effects of female education. This work exploits a triple difference identification strategy relying on exogenous spike in cotton price and spatial variation in suitability for cotton, and split sample analyses based on the exogenous allocation of land contracts. Results show that gender differences in parental investments in patriarchal societies can be reinforced by the type of agricultural activity, while positive economic shocks may further exacerbate this bias, additionally crowding out higher possibilities to invest in female education.

The third chapter brings novel evidence of the role of the language policy in building national sentiments, affecting educational and occupational choices. In this study I focus on the case of Uzbekistan and estimate the effects of exposure to the Latin alphabet on informational literacy, education and career choices. Results show that alphabet change has an impact beyond its intended political target. It also affects people's informational literacy and the formation of certain educational and labour market trends.

JEL Classification: N30, O10, Z13, I25

Keywords: nomadism, risk aversion, risk behaviour, education, cotton, alphabet, career choices

Contents

List of Figures

List of Tables

1   "Ethnic Roots of Risk Attitudes: The Impact of Ancestral Lifestyles on Risk Taking Behaviour"

Abstract

Willingness to take risks is a cornerstone of crucial economic decisions ranging from savings and employment to education and health. This paper finds that various dimensions of risk attitudes and risk behaviour are deeply rooted in our ancestral lifestyles: nomadic and sedentary. Those communities were spatially and culturally distant from each other. One of its most distinctive features, mobility, was associated with changing landscapes, fast adaptation and constant acquire of new information. Hence, nomadic lifestyle encouraged risk-seeking behaviour and made it an important transmissible survival trait across generations. Using aggregated data from several ethnographic sources allowed to classify 1309 ethnic groups as nomadic or sedentary (including other group characteristics: inheritance rules, beliefs, hierarchy among others), at time of the first European records. Manually matching this historical data with risk proxies from up-to-date surveys covered people living in nearly 450 regions within 50 to 80 countries. To document the effect of nomadic lifestyle on risk, this study proposes a novel instrument specification exploiting initial biogeographic conditions. Namely, the shortest distance between original location of ethnic group and the first wild animals' domestication location. Looking at different specifications, this study provides strong evidence that ancestral lifestyle can explain within country differences in risk attitudes in terms of willingness to take risks, preference for security and risky health practices. It shows that risk heterogeneity can be traced back to the ethnic origins emphasizing the important role of ancestral culture in shaping risk perceptions[1].

JEL classification: Z13, N30, O12, D81

Keywords: nomadism; risk attitudes; risk behaviour; persistence

1.1 Introduction

Risk and uncertainty are cornerstones of many crucial economic decisions. The growing amount of literature emphasized the importance of individual risk attitudes in understanding economic behaviour. The uncovered components of individual risk heterogeneity predict various dimensions of economic choices, including savings and investment, schooling, employment and gender norms (Guiso & Paiella, 2005, Dohmen et al, 2010, Sutter et al, 2015, Becker et al, 2018), however, some questions remain being understudied.

One of the most puzzling questions concerns the roots of modern heterogeneity in risk attitudes. Recent evidence has shown that the large variation exists not only within individual characteristics, but also across countries. The main source of this cross-country divergence is focused on environment and exposure to conflicts (Gächter & Schulz, 2016, Voors et al, 2012, Galor & Savitsky, 2018). Yet, culture is one of the important candidates explaining country's heterogeneity across various dimensions. Therefore, this paper questions whether risk aversion could also be rooted in cultural traits. By bringing together the comparative development studies uncovering the importance of historical events and persistence of culture (Alesina et al, 2013, Spolaore & Wacziarg, 2016, Michalopoulus et al, 2018, Giuliano & Nunn, 2019), this study shows that heterogeneity in risk attitudes can be traced to the ethnic origins and be partially explained by ancestral way of living.

At the course of human history, people were always in motion. We may consider ourselves as migratory species, since our ancestors have been constantly moving at the dawn of human civilization. Some of them kept doing so even after settled life became more appealing for development and growth. Nomadic people were carriers of civilizations in the old world for thousand years. They have always been a distinctive form of living. However, nomads were not always perceived as such. It is been only the last few decades that nomads are viewed not only as a separate societal phenomenon, but as a fascinating form of human lifestyle (Cashdan, 1990, Leder & Streck, 2005). The importance of nomadism is now seen in a new light. This paper gets inspiration from the ethnographic literature and transmits those knowledge to the realm of economics in order to explain the ethnic component in risk attitudes.

Along wide stretch of history, nomads, who were not necessarily small in groups, created their own peculiar form of a lifestyle characterized by a great variety. One may instantly think of the

pastoral groups, however, other forms of nomadic lifestyle do exist and include military nomads, hunter-gatherers, tinkers and traders. All of them has the most crucial feature in common - mobility. This is the most efficient strategy applied by nomads to generate highest possible returns while exploiting resources at the highest variance of risk. Thus, any nomad can be described as a member of community without fixed habitation, which moves from place to place (Leder & Streck, 2005). This mobility has shaped nomadic living norms and created spatial and cultural distance to the settled "sedentary" communities.

The notion of sedentarism and its role in the process of urbanization have been explored in the comparative development economic literature. Yet, little attention was given to nomadism, despite the large spread of nomadic populations. The last official figures estimate that there are 120 millions of nomadic groups worldwide (Rass, 2006). The spatial allocation of nomads can be seen from the Figure 1.1 below, where yellow colour reflects the spread of nomadism.
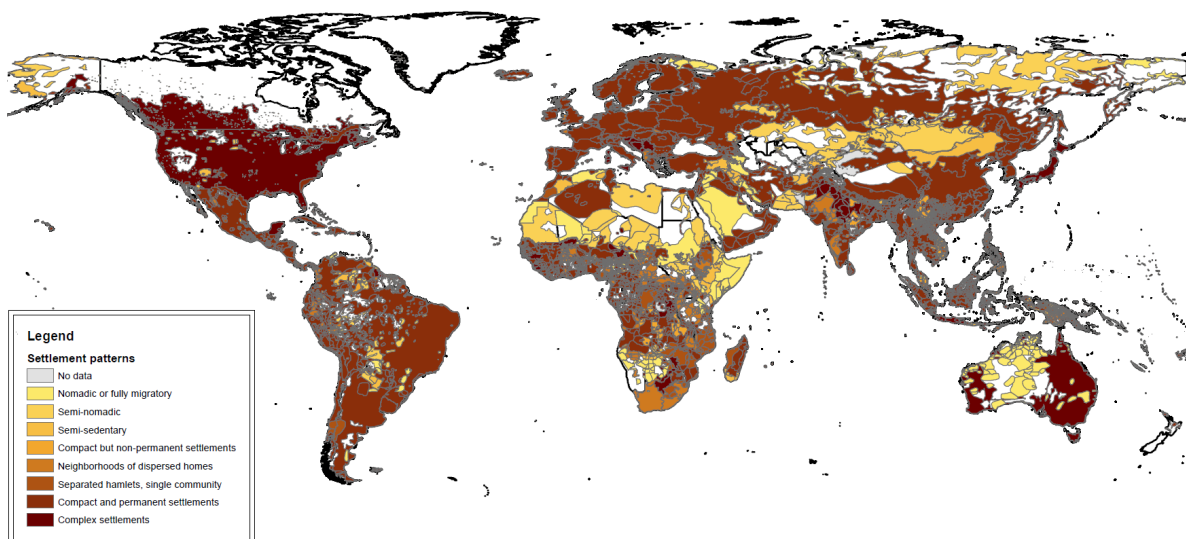
Figure 1.1 Distribution of nomads

Notes: data taken from Ethnographic Atlas (Murdock, 1967).

Evidence from anthropological literature emphasizes the advantage of nomadic lifestyle, which allows to take risks in order to generate highest returns. In particular, nomadic pastoralists seek reliability in highly risky environment: they accept the variability of productive inputs and modify their herding/social systems appropriately. Moreover, nomads do not strive for stability, they capitalize on periodic good fortune and invest in social capital. In general, nomadic groups turn instability (high variance of risk) into an asset and perceive uncertainty as an opportunity to get higher returns (Cashdan, 1990, Galaty & Bonte, 1991, Chang & Koster, 1994, Bollig & Goebel, 1997, Kraetli et al, 2013, Shaughnessy, 2018).

The key observation lying in this paper is that nomadic lifestyle has an impact on risk preferences. Exploiting resources at the high variance of risk in order to generate greatest returns associates with higher willingness to take risk and creates long-standing differences with settled communities. Mobility is associated with changing landscapes, fast adaptation and constant acquire of new information, which makes risky behaviour appealing and self-sustained. Mobile lifestyle shapes particular traits and behaviour among nomads. Attitude towards risk is part of those traits (Cashdan, 1990; Becker, 2019). This paper, hence, tests the hypothesis that descendants of nomads are more willing to take risks and practice risky health behaviour. It evaluates whether our ancestry has an impact on risk attitudes nowadays. For this purpose, I link contemporary individual-level data on various proxies for risk attitudes to a historical ethnicity-level measure of nomadism. In a broad set of within-country analysis, I find that people who descend from the ethnic groups that historically practiced nomadism (i) are more willing to take risks, (ii) value security less, and (iii) have riskier health behaviour.

The data on historical lifestyle are based on information from Murdock's Ethnographic Atlas (Murdock, 1967) and complimented by World Ethnographic Sample, Standard Cross-Cultural Sample (Murdock & White, 1969). These data are based either on early written history or first accounts from the earliest (European) observers of these cultures prior to the industrialization of those societies. It makes the database the earliest available information covering wide range of ethnic groups worldwide. There are 1309 ethnicities, each classified as nomadic or sedentary based on their lifestyle. Furthermore, the contemporary data on risk attitudes stem from two sources - the World Value Surveys and the Standard Demographic and Health Surveys (DHS). Both contain the information on respondents' ethnicity. Based on this, my empirical analysis manually links individuals' responses to their ethnic group's lifestyle from ethnographic data. This procedure generates substantial within-country variation. I cover 80 countries for risk preferences and 42 countries for risk health behaviour. Concerning the empirical strategy, it rests on two pillars. First, throughout the analysis, I compare individuals from different ethnic groups who live in the same country today, thereby holding constant the institutional environment and other factors that vary at the country level. In the same spirit, the analysis accounts for a large set of individual-level observables, ethnic-group level characteristics and geographical controls. Second, nomadism is correlated with biodiversity. The domestication of animals that were able to travel at long distances and carry loads determine the adoption of nomadic lifestyle. The closer the ethnic group was to the domestication point, the higher chance of it to become nomadic. This distance is plausibly exogenous to the risk attitudes. As described in greater detail below, the distance towards origins of domestication facilitates an instrumental variable approach.

The analysis shows that, in a sample of about 45000 people from 80 countries worldwide, ancestral nomadic lifestyle affects willingness to take risks today. Coming from nomadic background increases the willingness to take risks by 7 percentage points. Next, the analysis shows that in a sample of about 500000 people from 42 countries, nomadic background is associated with riskier health behaviour. In particular, people with nomadic ancestry are by nearly 4 percentage points more likely to smoke regularly, by 8.7 percentage points less likely to have health insurance and by 1.4 percentage points less likely to use contraception. These effects are quantitatively meaningful and account for region time fixed effects. Moreover, these results are robust across a wide range of specifications that account for (i) individual-level observables such as age, religious denomination, urban residence, educational attainment, or marital status and (ii) historical ethnic group level characteristics such as plough use, subsistence, kinship structure, or the jurisdictional hierarchy and (iii) geographical characteristics, such as temperature, humidity, elevation, distance to the coast and distance to the equator. Next, I restrict the sample for movers only, people who currently live outside their ethnic homeland. The results stay fairly unchanged in terms of direction and magnitude of the effect. To further support a causal interpretation of these results, I turn to an instrumental variable approach, which exploits the fact that nomadism was partially defined by the availability of animals that could be used for the long distance travel. Based on the data from genetic biology literature, I identify the points of first domestication and use the distance to the ethnic group's homeland as an instrument for the ethnic group's ancestral lifestyle. The resulting IV estimates are consistent with their OLS counterparts in terms of coefficient's sign and statistical significance but tend to be larger in terms of the effect size.

Relating this work to the literature, this paper adds to the research on the origins of heterogeneity in risk attitudes, such as environment or conflicts (Voors et al, 2012, Gächter & Schulz, 2016). It also relates to the literature on cultural persistence, historical events and current economic outcomes (Spolaore & Wacziarg, 2009, 2013, 2016, Xue, 2019, Giuliano & Nunn, 2019). In particular, the contribution of this paper is (i) providing the first analysis on the ethnic origins of risk attitudes; (ii) focusing on within-country variation; (iii) analyzing the effect of nomadic lifestyle from the perspective of economics; (iv) introducing the novel instrumental approach. The paper thereby provides an empirical evidence that ancestral experiences play a great role in defining our behaviour nowadays.

The rest of this paper is structured as follows. Section 1.2 shows the related literature on the origins of risk attitudes and persistence of culture and history. In Section 1.3, I motivate the expected relationship between ancestry and modern risk attitudes. In Section 1.4 I discuss the data I used and the empirical strategy. Section 1.5 presents the results. Section 1.6 brings

additional evidence from one country. Section 1.7 provides the concluding remarks. Section 1.8 is bibliography. Appendix can be found in Section 1.9.

1.2 Literature Review

Evidence from ethnographic literature pointed out several crucial characteristics that nomadic life brings to people (Cashdan, 1990, Galaty & Bonte, 1991, Bollig & Goebel, 1997, Fratkin, 1997). Generally speaking, nomads are able to manage risks more successfully than sedentary counterparts under same circumstances, since they turn instability into an asset and uncertainty of constant moving is seen as opportunity. For instance, according to Steinfeld et al (2010) and Kraetli et al (2013), mobility is production-boosting strategy, pulled by relative abundance rather than pushed by scarcity. Nomads exploit risky conditions to generate the highest returns, their perception of risk is lower comparing to their sedentary counterparts. These studies are mainly focused on observing single tribes and communities in different regions of the world. This paper is bringing these findings in the realm of economics.

Risk preferences have long been known to be too complex to be described by a single parameter (Vickrey, 1945, Friedman & Savage, 1948). Broad range of economic literature shed light on the determinants of individual risk attitudes including wealth, income, age, gender, religion and nationality (Weber et al, 1998; Guiso & Paiella, 2004; Brown, 2007; Benjamin et al, 2010; Dohmen et al, 2010, Dohmen et al, 2017). This research went beyond individual determinants of risk preferences in its attempt to explain cross-country heterogeneity (Dohmen et al, 2012; Becker et al, 2014; Galor & Savitskiy, 2018). The main source of this divergence is focused on environmental characteristics and exposure to conflicts. Adverse weather conditions, war and conflicts are among main candidates to explain country level divergence in risk (Gächter & Schulz, 2016; Voors et al, 2012; Galor & Savitsky, 2018). Yet, the roots of such heterogeneity are to be fully uncovered.

One possible explanation can be found in different historical experiences that have been accumulated over time and caused divergence in risk attitudes (Becker et al, 2014). Similar to the way historical factors have been shaping distinctive culture norms, historical lifestyle may shape risk behaviour nowadays (Alesina et al, 2013; Guiliano & Nunn, 2017; Michalopoulus et al, 2018). Taken together, the debate regarding the origins of risk attitudes is ongoing and not fully explored. It suggests that there is a relationship between temporally distant events, persistent cultural characteristics and economic determinants. This paper attempts to provide an insight on how our ancestry shapes risk behaviour by building on the literature of cultural persistence and transmission of risk attitudes.

6

Numerous studies emphasized important role of culture and its persistence over generations. Guiso et al (2006) was one of the pioneers in defining culture and its influence on economic outcomes. Spolaore & Wacziarg (2009, 2011) uncovered a strong correlation between genetic distance and income differences across countries. They showed that lower genetic distance between two countries brings the diffusion of knowledge and technology leading to development. Starting from early 2010s literature on historical persistence of traits and preferences evolved widely. Among others, two prominent studies were conducted by Nunn & Wantchekon (2011) and by Voigtlander & Voth (2012). The first focused on the origins of mistrust in Africa while analyzing the slave trade patterns. They found that less trust today is the consequences for those groups whose ancestors were heavily raided in the time when slave trade emerged and developed mistrust to everyone who was not from their respective group. The second one showed that the raise of anti-semitic violence in Nazi Germany had medieval origins. Precisely, plague-era pogroms were strongly associated with violence occurred in 1920s, votes for the Nazi Party, deportations and attacks on synagogues. Both brought an evidence on the important role of individual ethnicity in shaping people's behaviour. Another study by Michalopoulos et al (2018) focused directly on the effect of historical ancestral lifeways on the attained level of education and wealth among the descendants in Africa. Controlling for pre-colonial characteristics, attitudes towards violence and institutional changes, they showed that individuals, whose ancestors derived a larger share of subsistence from agriculture, are more educated and wealthy today. Emphasizing the important role of agriculture in the life of our ancestors, it is crucial to mention the work by Alesina et al (2013). Their evidence of cultural persistence and evolution of individual beliefs was based on the origins of gender roles in the modern societies. Shifting cultivation is labour intensive and is using handheld tools, in particular, plough cultivation requires considerable strength involved in the process. Thus, practicing plough agriculture caused males to work in the fields while females were specializing in home-related activities. With time this division of labour generated the belief that the natural place of female is at home. These cultural views persisted and nowadays they influence entrepreneurship, labour market participation and political activity among females. In a similar way, analyzing cultural persistence, but exploiting ancestral living conditions, Giuliano & Nunn (2019) showed that adopting traditions of the previous generation is advantageous only in stable environments. When the environment is changing, the cultural heritage of the previous generation does not bring valued information which would be relevant in the current living conditions. They controlled for the environmental stability through measuring the average temperature fluctuations across 20 years generations from 500-1900. Applying epidemiological approach across countries and ethnic groups, they found that groups with ancestors who lived in fairly stable environments put a higher importance on keeping traditions nowadays. Also, these groups preserve their traditions with higher degree of persistence over time.

This paper supports the presented evidence on how ancestral beliefs are persistent and shaping our culture nowadays, but proposes to apply this way of thinking for risk attitudes.

Turning to the risk preferences, the literature grew in various dimensions. I will solely focus on the couple of papers uncovering risk heterogeneity and its transmission over generations. Prominent study on risk perception across states was done by Becker et al (2014). They studied the world-wide distribution of risk preferences and found that ancestral migration patterns (proxied by human genetic variation) bear a strong association with the between-country heterogeneity in risk. Both actual and predicted population genetic distance explains the observed heterogeneity suggesting that the specific ways in which mankind migrated out of Africa thousands of years ago has left a footprint in the contemporary cross-country distribution of economic traits. More importantly, works by Cesarini et al (2009) and Dohmen et al (2012) showed that attitudes of people, including risk aversion, are transmitted across generations and that part of transmission is genetic in nature. Moreover, recent literature points out that economic preferences, including risk preferences, are malleable by idiosyncratic experiences and by the composition of people's environment (Voors et al, 2012, Gallen et al, 2014, Kosse et al, 2016, Alan & Ertac, 2018, Rao, 2018). Therefore, the differential historical experiences which have been acquired over generations and years might have given a rise to various risk and time preferences as of today.

A key question is how risk attitudes perpetuate over time. In particular, Dohmen et al (2011) shed a light on where risk attitudes come from and found similarities in behaviour across generations, but differences across ethnic groups and countries. This paper showed how intergenerational transmission of risk attitudes from parents to children is an important channel of influence. For instance, they observed that children end up with same outcomes as their parents partially because they inherit same attitudes and follow same behaviour to make similar decisions in life. Similarly, based on cross-country survey, Fehr et al (2006) demonstrated that the sizeable difference between average willingness to take risks in the USA and in Germany (5.6 versus 4.4 respectively, on a scale from 0 to 10) could be explained by the difference in immigrants' behaviour. In particular, US immigrants are more willing to take risks than people in their country of origin. Overall, intergenerational transmission helps explaining the persistence of heterogeneity in risk attitudes over time which in turn explains a variety of well documented differences in risk behaviour across countries.

Bringing together the importance of ancestry and transmittable risk attitudes, this paper contributes to the different strands of economic literature by explaining the fraction of within-country group-level variation in risk through its ethnic origins. It emphasizes the role of nomadism in formation of lower degree of risk aversion and it brings an evidence that heterogeneity in risk attitudes nowadays can be partially explained by ancestral lifestyle.

1.3 Motivation

This section proposes an evolutionary explanation for the origin of risk aversion for nomadic and sedentary types of population. It shows how risk aversion evolves due to the difference in type of risk, which individuals belonging to a certain group are facing. Nomads are constantly moving, known for their fast adaptation and continuous update of the knowledge (Kraetli et al, 2013; Shaughnessy, 2018). For them, idiosyncratic risks hide the danger for their survival and transmission of traits to their descendants. In contrary, sedentary people learn how to cope with old age and sickness and other forms of idiosyncratic shocks better given the advantages that their lifestyle presents them. However, they are more vulnerable to aggregate shocks such as flood, war and others which threaten their survival and their non-movable assets. I exploit this difference in the types of risk associated with both lifestyles and evaluate how evolutionary perspective of transmission of traits may lead to the heterogeneity in risk behaviour. Following Robson & Samuelson (2010) and Zhang et al (2014), I present here a simple model on how risk traits were propagating through generations and how this divergence may be explained by the difference in ancestral lifestyles.

Assume that individuals live for one period, have asexual random reproduction and die. During his/her life individual i has only one decision to make: choose one of two possible actions $A1$ and $A2$. This decision results in one of two respective random quantities of offspring $q_{A1,i}$ and $q_{A2,i}$. So reproductive success of an individual is a weighted sum, where

$$\begin{cases} q_{A1,i} = \delta z_{A1} + (1-\delta)y_{A1,i} \\ q_{A2,i} = \delta z_{A2} + (1-\delta)y_{A2,i} \end{cases}$$

where (z_{A1}, z_{A2}) is assumed to be i.i.d over time and for all individuals in a given generation, hence, it represents systematic (or aggregate) risk. In other words, one person experiences the same outcome as others who chose the same action. Similarly, $(y_{A1,i}, y_{A2,i})$ is assumed to be i.i.d both over time and across person i in a given generation, hence, it represents idiosyncratic risk. Both components are described by some well-behaved probability distributions and δ is a real number between zero and one representing the share of each type of risk.

Every person chooses $A1$ with some probability $p \in [0,1]$ and $A2$ with probability $1-p$. In this specification p determines the way person chooses between $A1$ and $A2$, so p is an individual behavior. Additionally, individual i of type p produces random quantity q_i^p of offspring. Hence, $q_i^p = I_i^p q_{A1,i} + (1 - I_i^p)q_{A2,i}$ where I_i^p is the Bernoulli random variable that equals 1 with probability p and 0 otherwise. Offspring behaves in the same manner as parents. Therefore, the population could be segmented to types p. This assumption demonstrates perfect genetic trans-

mission of traits including risk attitudes from one generation to the next. Another assumption is that initial population contains an equal number of all types, so it is normalized to 1 for each type without loss of generality. Lets rewrite q_i^p as a combination of systematic and idiosyncratic risks:

$$q_i^p = \delta z_i^p + (1-\delta)y_i^p,$$
$$\text{where}$$

$$\begin{cases} z_i^p = I_i^p z_{A1} + (1-I_i^p)z_{A2} \\ y_i^p = I_i^p y_{A1,i} + (1-I_i^p)y_{A2,i} \end{cases}$$

The coefficient $\delta \in [0,1]$ shows the proportion of systematic risk in the environment. When $\delta = 1$, all risk is systematic, when $\delta = 0$, all risk is idiosyncratic, and when $0 < \delta < 1$, both risks are there. In the environment described by δ, the total number of offspring of type p in generation G is denoted by g_G^p. The average of the log population $G^{-1}log(g_G^p)$ converges in probability to the log geometric average growth rate[2]:

$$\gamma_\delta(p) = \mathbb{E}_z[log(\delta z^p + (1-\delta)\mathbb{E}_y[y^p])], \tag{1.1}$$

where $\mathbb{E}_z$ denotes the expectation taken with respect to z^p and $\mathbb{E}_y$ is expectation with respect to y^p. Let's rewrite the previous equation as following:

$$\gamma_\delta(p) = \mathbb{E}_z[log(p\phi_{A1}^\delta + (1-p)\phi_{A2}^\delta)] \tag{1.2}$$

where
$$\phi_{A1}^\delta = \delta z_{A1} + (1-\delta)\mathbb{E}_y[y_{A1}], \phi_{A2}^\delta = \delta z_{A2} + (1-\delta)\mathbb{E}_y[y_{A2}] \tag{1.3}$$

Maximising equation (2):

$$p_\delta^* = \begin{cases} 1 \text{ if } \mathbb{E}_z[\phi_{A1}^\delta/\phi_{A2}^\delta] > 1 \text{ and } \mathbb{E}_z[\phi_{A2}^\delta/\phi_{A1}^\delta] < 1 \\ 0 = \mathbb{E}_z[\frac{\phi_{A1}^\delta - \phi_{A2}^\delta}{p\phi_{A1}^\delta + (1-p)\phi_{A2}^\delta}] \text{ if } \mathbb{E}_z[\phi_{A1}^\delta/\phi_{A2}^\delta] \geq 1 \text{ and } \mathbb{E}_z[\phi_{A2}^\delta/\phi_{A1}^\delta] \leq 1 \\ 0 \text{ if } \mathbb{E}_z[\phi_{A1}^\delta/\phi_{A2}^\delta] < 1 \text{ and } \mathbb{E}_z[\phi_{A2}^\delta/\phi_{A1}^\delta] > 1 \end{cases} \tag{1.4}$$

If population consists of people facing different risks, some are exposed to idiosyncratic risk, some - to systematic, some - mix of both.

The lifestyle of the group of individuals determines the proportion of the certain type of risk to occur. I model the extreme cases for nomadic and sedentary lifestyles. I refer to the

---

[2]The proof of this equation is given in the Appendix Section 1.9.2

categorization of risks for population adopted from World Development Report (2000) in the table below.

Type of risk	Idiosyncratic Risk	Systematic Risk
Natural		Rainfall, flood, drought
Health	Illness, injury, disability, old age, death	Pandemic
Social	Crime, domestic violence	War, social uprising
Economic		Unemployment, epidemic
Political		Riots, coup d'état
Environmental		Pollution, deforestation, desertification

Table 1.1 Categorization of risks: from World Development Report (2000)

From this table I differentiate for two types of risk which are best related to particular lifestyle supported by Cashdan (1990), Kraetli et al (2013), Shaughnessy (2018). In particular, nomadic tribes are mainly exposed to the idiosyncratic type of risk, since if the person gets sick or disabled, it has much stronger effect on the well-being of the group given continuous movement and its small size. While if the aggregate shock occurs, nomads can change the location and leave the troublesome zone. The contrary could be observed for sedentary counterparts, where systematic shock hits harder taking away the opportunity to mitigate it by leaving affected area. Idiosyncratic shock, otherwise, has less effect, since injuries and illnesses are treated better in stable environments.

Now let's assume that individual preferences are described by objective function $V_\delta(z, y)$ with z, y representing different types of risk (systematic and idiosyncratic, respectively). Let (z_1, y_1) be preferred over (z_2, y_2) if and only if $V_\delta(z_1, y_1) > V_\delta(z_2, y_2)$. In order to make choice over random outcomes, person maximizes $V_\delta(z, y)$. Let's model perception of risk into equations above. When $\delta = 1$, the person faces systematic risk. The evolutionary dominant behaviour f^* is given by $f^* = argmax_f \mathbb{E}_z[log(z^f)]$, so maximization criterion is

$$V_{\delta=1}(z, y) = \mathbb{E}_z[log(z)] \tag{1.5}$$

When $\delta = 0$, the person faces idiosyncratic risk. The evolutionary dominant behaviour f^* is given by $f^* = argmax_f log(\mathbb{E}_y[y^f])$, so maximization criterion is

$$V_{\delta=0}(z, y) = \mathbb{E}_y[y] \tag{1.6}$$

This function is expected linear utility. The optimal criteria for a person is the expected value, meaning that individuals are risk neutral. The level of risk aversion is a function of systematic risk in the environment. Keeping the general case, $0 < \delta < 1$, risk contains both systematic and

idiosyncratic components. The maximization will look like:

$$V_\delta(z, y) = \mathbb{E}_z[log(\delta z + (1 - \delta)\mathbb{E}_y[y])] \quad (1.7)$$

The idiosyncratic component is giving a rise to the linear expectation in y, and the systematic component is giving a rise to a logarithmic function of z and the expectation in y. Risk aversion emerges as a consequence of systematic risk and risk neutrality or risk-seeking - as a result of idiosyncratic risk. As a result, people who are most threaten by systematic risk tend to be more risk averse than those who are vulnerable to idiosyncratic one. Implying that sedentary people tend to be more risk averse in comparison to nomads.

1.4 Data and Empirical Strategy

1.4.1 Data

Ethnographic data on ancestral ethnicity is coming from Murdock's Ethnographic Atlas (Murdock, 1967) and complimented by World Ethnographic Sample, Standard Cross-Cultural Sample (Murdock & White, 1969)[3]. These data are based either on early written history or first accounts from the earliest (European) observers of these cultures prior to industrialization of those societies. It makes the database the earliest available information covering wide range of ethnic groups worldwide. There are 1309 ethnicities, each classified as nomadic or sedentary based on their lifestyle. In addition, it contains ethnic group level information on subsistence, kinship organization, size of community, domestic organization, marital structure, religious beliefs, settlement patterns, political organizations, institutional complexity, class stratification, slavery, inheritance rules, rigidness of the society (openness and acceptance of foreigners) and is intended to reflect ancestral ways of living before colonization and industrialization, even when the exact timing of observation differs between ethnic groups. The constructed map below shows the location of centroids coming from combined ethnographic dataset.

---

[3]please see Appendix for additional details
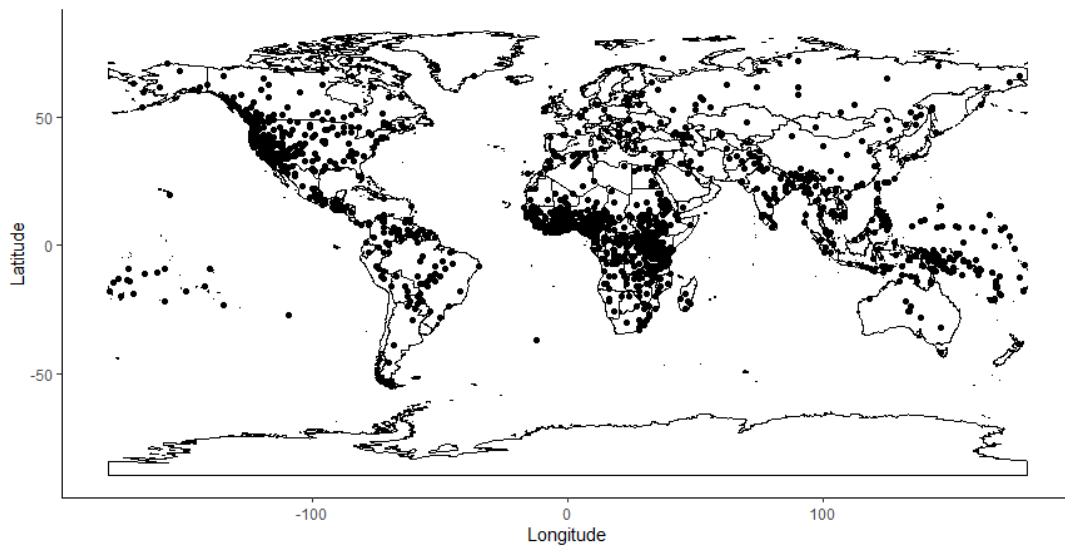
Figure 1.2 Location of centroids for each ethnicity from Ethnographic Atlas and Standard Cross-Cultural Sample

The validity of the Atlas was questioned by some anthropologists and historians (Leach, 1964, Jerven, 2011). The recent work by Bahrami-Rad, Becker and Henrich (2021) tests the concerns across a wide range of dimensions. They conclude that Atlas is a valid source of information about various human societies. Since the accounts captured in the Atlas are informative and considerably analogous to the self reported records of the descendants of the portrayed societies.

I construct my main explanatory variable - ancestral lifestyle - by aggregating together the information on settlement patterns and lifestyle. I create a dummy variable (0;1) where 1 stands for Nomadic or fully migratory and 0 - for all other different layers of sedentary life. Sedentary complexity includes several types of settlements: compact, dispersed, partially dispersed, impermanent and permanent settelmens, each of different size. I aggregate them all together to account for sedentary lifestyle. There are two additional levels: seminomadic and semisedentary, which I do not include into my main analysis, since I am solely focusing on the difference between nomadic and sedentary forms of living. Figure 1.3 shows the distribution of areas (yellow coloured) where people were practicing nomadism.
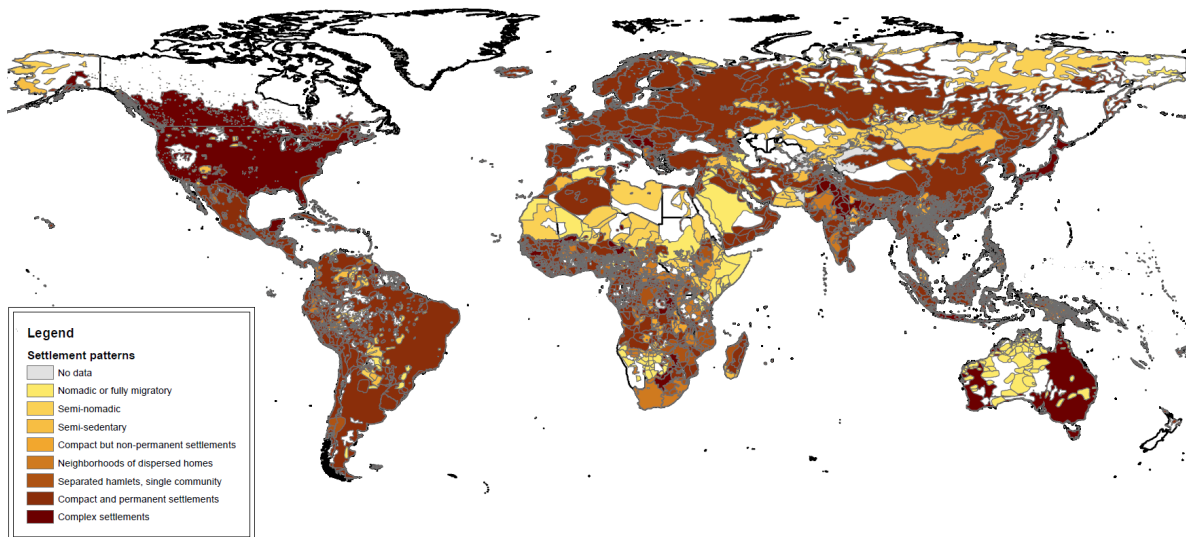
Figure 1.3 Distribution of nomadism: calculations based on Ethnographic Atlas

The contemporary individual-level data stem from the World Value Surveys (hereafter, WVS) and Demographic Health Surveys (hereafter, DHS). In order to measure risk preferences and risk behaviour I used different proxies. The proxies for willingness to take risks are taken from WVS and a set of proxies for risky health behaviour are taken from DHS. The World Values Survey (WVS) is a common questionnaire survey used to study the changing attitudes and values of individuals in almost 100 countries. It is a cross-national time series study that is conducted in waves, with each wave covering a span of five years. The WVS has conducted interviews with almost 400,000 respondents and covers a broad range of countries with very different wealth levels and cultural backgrounds.

The DHS surveys are nationally representative household surveys covering more than 90 countries worldwide. From 1984 until today, seven waves have been conducted. The country samples are quite large, with typically between 5,000 and 30,000 households being surveyed. The DHS elicits detailed household and respondent characteristics. It records not only standard socio-demographic variables but also, for example, information on housing quality, availability of electricity, literacy, access to clean water, sanitation, or types of insurance, location of households.

Importantly, for some countries and waves, both WVS and DHS contains information about respondents' ethnicity. This information allows me to match respondents in the DHS to their ancestors' ethnic group in the ethnographic dataset.

In order to link this information about ancestoral lifestyle to risk proxies for each individual I use a manual matching procedure. It involves manual correction of each ethnicity indicated in WVS and DHS datasets to directly match with the ethnographic data (please refer to Figure 1.4).

14

WVS or DHS

BY: Belorussian
Belorus
Gheg Albanian
Oyo Yoruba
Yoruba
Kalmik
Qalmyq

Byelorussians
Byelorussians
Gheg
Yoruba
Yoruba
Kalmyk
Kalmyk

**Manual correction
of each ethnicity**

Ethnographic Atlas

Byelorussians
Gheg
Yoruba
Kalmyk
...
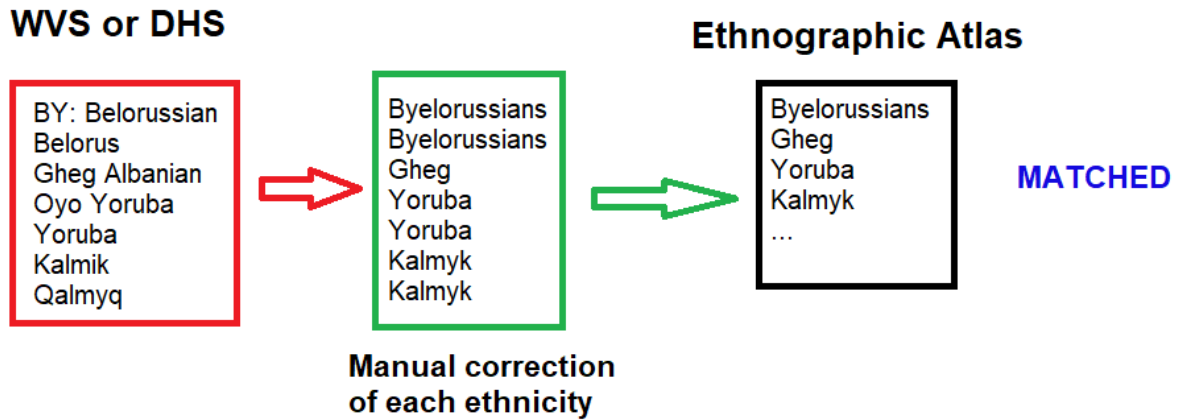
MATCHED

Figure 1.4 Matching procedure

The list of ethnicities that each individual indicated in the WVS and DHS is manually processed and corrected in order to be matched with ethnographic data with the list of ancestral ethnicities. Some corrections are made based on information in those ethnographic sources, such as changing ethnic names over time, geographical location along with language spoken and some corrections based on the country where the respondent is living in.

In total, I match almost 60000 individuals representing 104 ethnicities and nearly 500000 individuals from 332 ethnicities, correspondingly. In particular, I cover 450 regions, 80 countries in WVS and 403 regions, 42 countries in DHS over time 1981-2016. Risk preferences are measured by the following three variables: 1) TakingRisks: Whether it is important to this person taking risks (following Dohmen et al (2011)); 2) JobSecurity: Important in a job: job security (measure of loss aversion by Galor & Savitskiy (2018)); 3) FeelingSecure: It is important to this person living in a secure surroundings. Subsequently, risk behaviour is proxied as follows: 1) Smoking habits: Are you smoking regularly?; 2) Contraception use: Are you using any contraceptive method?; 3) Health insurance: Do you have a health insurance?.

The main measure of willingness to take risks is at the Schwartz scale from 1 to 6, asking whether the person is willing to take risks and engage in adventures. The distribution across sedentary and nomadic groups can be observed in Figure 1.5.
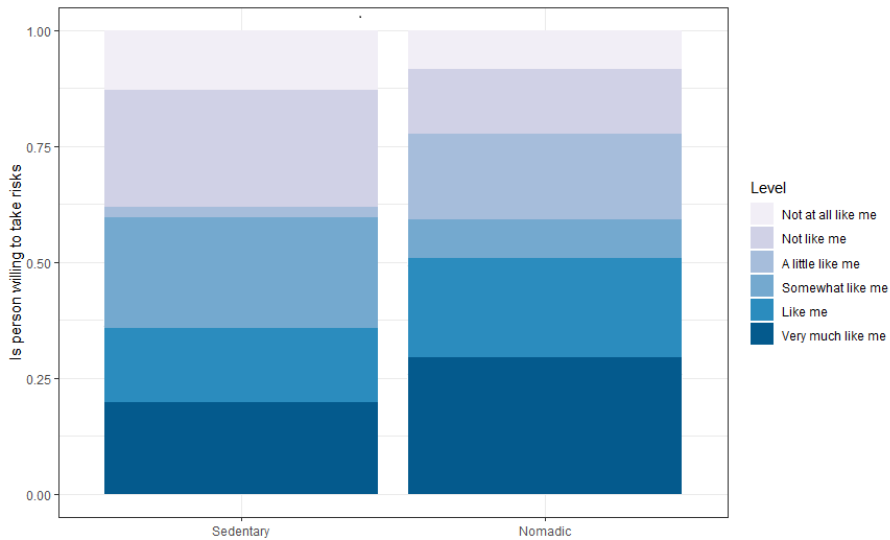
Figure 1.5 Willingness to take risks

Standardizing this measure into dummy variable from 0 to 1 and applying for people from sedentary and nomadic ethnic groups, Figure 1.5 shows that people with nomadic background associate themselves more as risk takers in contrary to their sedentary counterparts.

Another measure of risk or loss aversion taken from Galor & Savitsky (2018) which is valuing job security more than any other job characteristics including salary, career opportunities, working hours. It reflects people not being willing to take any risks and valuing security and stability above all. From Table 1.2 it can be seen that people from sedentary background are among those preferring stability of job above other factors, which signals that these individuals do not want to take risks on a daily basis.
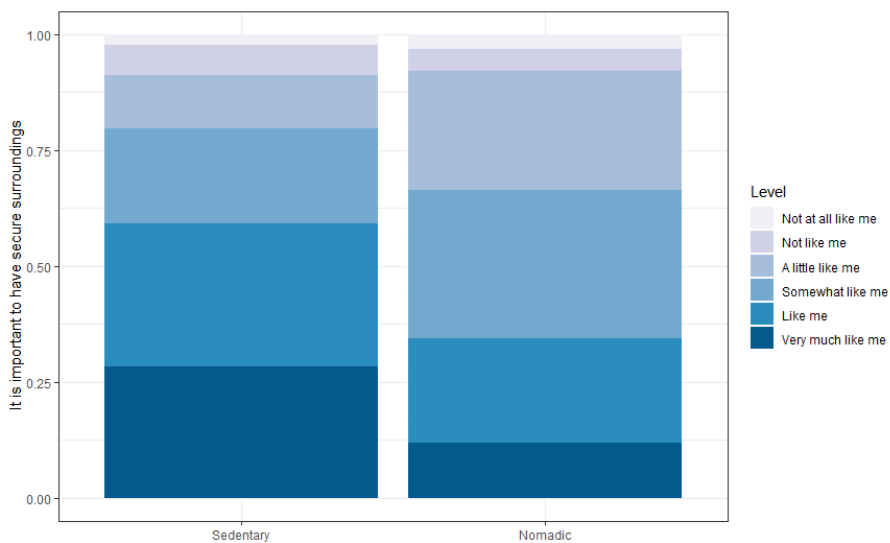
Figure 1.6 Important to have secure surroundings

The third proxy for risk is whether the person values the security of surroundings, which is calculated according to the Schwartz scale from 1 to 6 (similarly to the willingness to take risks). The allocation of preferences among two groups can be seen in the Figure 1.6. Afterwards, the measure is harmonized to (0;1) dummy for valuing security for each person. Individuals with nomadic background tend to put less importance on the security of the surroundings supporting the hypothesis that nomadic lifestyle in the past is associated with risk-taking behaviour nowadays (please refer to the Table 1.2).

	Sedentary	Nomadic	Tstats	p-value
TakingRisks	0.11	0.25	-1.48	0.93
JobSecurity	0.65	0.44	-18.52	1.00
FeelingSecure	0.83	0.69	-4.34	1.00

Table 1.2 T-statistics for main variables of interest

In order to look into another dimension of risk, namely, risky lifestyle behaviour, I constructed three health proxies described above, that are reflecting current attitude to the lifestyle and associated risk among individuals with nomadic and sedentary background. According to Cawley & Ruhm (2011), Teh et al (2019), Arni et al (2020) among others, risky health practices reflect risky behaviour and could be considered as a determinant of risk attitudes. Smoking, unsafe sex practice and not having health insurance are among most important factors of risky health behaviour described in the literature (Weber et al, 2002, Nosic & Weber, 2007, Galizzi & Miraldo, 2012, Akasaki et al, 2019). I construct these three factors from DHS data. In particular, the variable used in the analysis to capture the effect of smoking is a dummy variable which takes a value of 1 if the person smokes regularly. In order to have a better overview of smoking patterns among the descendants of nomads and sedentarists, I made a comparison based on the number of cigarettes smoked daily by the representatives of each group (see Figure 1.7). It can be seen that descendants of nomads tend to be heavy smokers comparing to their counterparts.
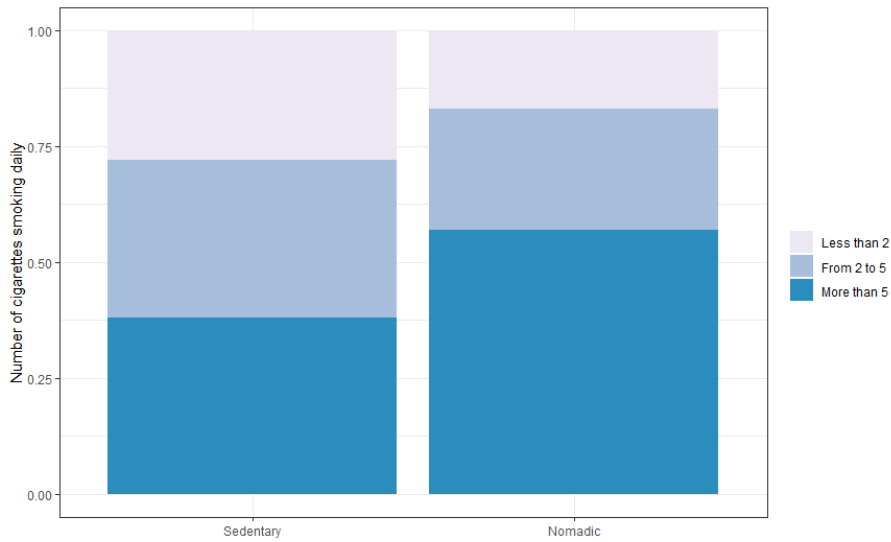
Figure 1.7 The number of cigarettes smoked daily

Comparing means for standardized variables of smoking, contraception use and having insurance (please refer to Table 1.3), it can be observed that people with nomadic background smoke regularly more on average than those with sedentary, they tend to neglect contraception and to lack insurance coverage. Looking at the types of contraception used by people with different

	Sedentary	Nomadic	Tstats	p-value
Smoking	0.22	0.39	-14.55	1.00
Contraception	0.67	0.47	27.97	1.00
Insurance	0.04	0.01	24.62	1.00

Table 1.3 T-statistics for main variables of interest

ethnic background, it is interesting to notice that descendants of nomads mainly used withdrawal, which is one of the riskiest way of contraception (please refer to Figure 1.8). They are also more likely to practice periodic abstinence, which could be referred to the nature of nomadic life. This observation supports Becker (2019), who found the strong effect of historical pastoralism on female genital cutting practices nowadays. Descendants of sedentarists tend to prefer pills and IUC, which may signal about heritage of greater dependencies on pharmaceuticals and better quality of the health system among sedentary communities.
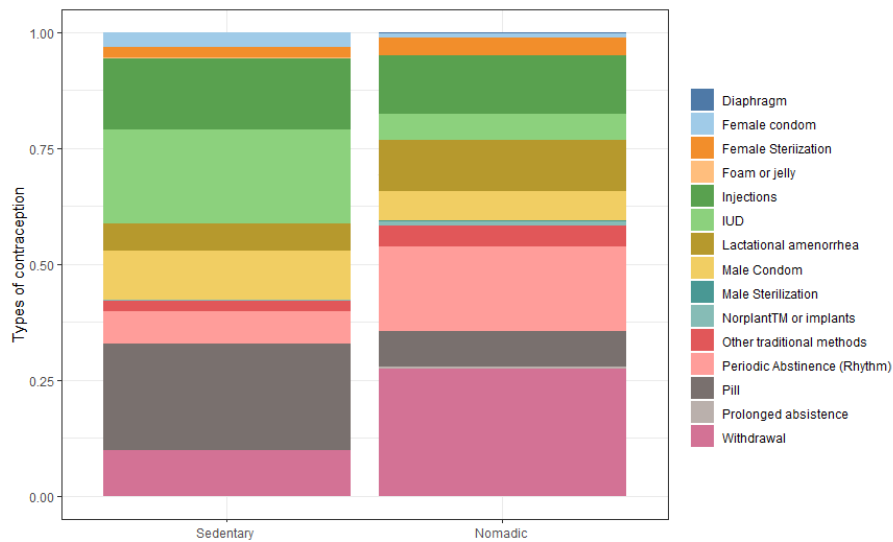
Figure 1.8 Types of contraception used

Overall, this section presented main features of the variables of interest, while the next one will describe the identification techniques used to address the question of nomadic influence on current risk attitudes and behaviour.

1.4.2 Identification Strategy

The empirical analysis overcomes considerable challenges in the identification of the causal effect of ancestral lifestyle on risk attitudes. The analysis adopts an empirical strategy that is designed to mitigate concerns about the potential role of omitted variables in the observed association between ancestral lifestyle and willingness to take risks (reverse causality is not an issue given the historical nature of main explanatory variable). Potential concerns about the role of omitted geographical, institutional, cultural, and human characteristics in the observed association between nomadic background and risk attitudes are mitigated by accounting for a large set of confounding characteristics that might have determined risk aversion. In particular the analysis accounts for: potentially confounding effects of: (i) geographical and climatic characteristics (e.g., elevation, distance to equator, historical temperature and humidity, distance to coast, level of precipitation), (ii) regional fixed effects, capturing unobserved time-invariant heterogeneity at the regional level; (iii) country fixed effects, and thus time-invariant country specific factors, (e.g., geography, institutions, history, and culture); (iv) individual characteristics (e.g., age, gender, number of siblings, religion, education level and income); (v) ethnographic characteristics (e.g., intensity of agriculture, subsistence, kinship organization, size of community, domestic organization, marital structure, religious beliefs, settlement patterns, political organizations, institutional complexity, class stratification, slavery, inheritance rules).

19

Additionally, the adoption of the epidemiological approach and the exploration of the determinants of risk attitudes among movers, e.g. people who live outside their ethnic homeland, permits the analysis to overcome two major concerns: (i) it distinguishes between the effect of ancestral lifestyle in the place of origin (rather than where they currently reside) on risk aversion, capturing the culturally embodied, intergenerationally-transmitted component of the effect of nomadic way of living; (ii) it accounts for time invariant unobserved heterogeneity in the place of living (e.g., geographical, cultural and institutional characteristics), and thus, mitigating possible concerns about the confounding effect of place-specific characteristics.

1.4.3 Baseline Specification

In this section I present the following baseline specification:

$$Y_{i,e,r,c,t} = \beta_1 Nomadic_e + \beta_2 X_{i,e,r,c,t} + \beta_3 Z_{e,t} + \beta_4 G_{e,r,c,t} + \sum_r \sigma_r Region_i + \sum_c \sigma_c Country_i + \varepsilon_{e,t}$$

where $Y_{i,e,r,c,t}$ is a binary measure of risk attitudes (proxy for willingness to take risks and risky health behaviour) of an individual i belonging to the ethnic group e in region r of country c at time t; $Nomadic_e$ depicts ethnic group being nomadic (0,1); $X_{i,e,r,c,t}$ is a set of individual socio-economic characteristics; $Z_{i,e,r,c,t}$ is a set of ethnic group characteristics; $G_{e,c,t}$ is a set of geographical and climate controls; $Region_i$ is a dummy variable for region and $Country_i$ is a dummy variable for country, where individual i is living. Standard errors are clustered at the ethnic group level[4]. The similar specification is applied for subsample of people, who are currently living outside their ethnic homelands.

1.5 Main Results

I start with baseline results for the willingness to take risks and I proceed with findings for risky health behaviour. Also, I reduce the sample to people living outside their ethnic homeland respectively and discuss the implications.

1.5.1 WVS

This section is focusing on the results from baseline specification regarding risk attitudes using proxies from WVS data.

---

[4]Clustering at region and country level didn't change the results much

WVS Baseline results

Variables	TakingRisks			JobSecurity			FeelingSecure		
	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Nomad	0.087**	0.0700**	0.067**	-0.189**	-0.187**	-0.172**	-0.104*	-0.180*	-0.171*
	(0.0290)	(0.0265)	(0.0211)	(0.0700)	(0.0797)	(0.0701)	(0.1160)	(0.0861)	(0.0791)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	-	No	Yes	-	No	Yes	-
Region FE	No	-	Yes	No	-	Yes	No	-	Yes
Time FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Mean (Y)	0.261	0.261	0.261	0.735	0.735	0.735	0.799	0.799	0.799
SD (Y)	0.367	0.367	0.367	0.441	0.441	0.441	0.400	0.400	0.400
Obs	43421	43421	43421	43421	43421	43421	43421	43421	43421

Table 1.4 WVS Baseline results

Note: Standard errors are clustered at ethnic group level and the analysis covers 104 ethnic groups in 450 regions of 80 countries. Individual controls include gender, having partner, number of children, highest education, religion, unemployed, income group, age, financial situation satisfaction, working class (e.g. middle class), urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
p <0.10,** p <0.05,*** p <0.01

From Table 1.4 it can be seen that people with nomadic background are more prone to take risks. Column 1 of the Table 1.4 shows the positive correlation between willingness to take risks and having nomadic background documents without any additional controls. In Column 2 I add individual-level controls with ethnic group and geographical controls. The inclusion of these controls makes the coefficient on ancestral lifestyle going up. In particular, having nomadic background in preindustrial times leads to a 8.7 percentage point increase in the willingness to take risks. In light of the fact that the unconditional willingness to take risks is only 16%, this effect is large. In a final step, I include country and then region time fixed effects with a broad set of variables which might be outcomes of nomadism themselves. At the individual level, I have educational attainment, marital status, a dummy for living in an urban area, age, religion, income group, number of children and if the person is unemployed. At the historical ethnic group level, I include a measure for subsistence to alleviate the concern that the effect of nomadism is not merely the result of pastoralism as opposed to the agriculture. Similarly, I include a measure for how hierarchical a society was in terms of their political organization and inheritance rules with the society to ensure that nomadic groups are not simply better organized to cope and insure themselves against risks. Most importantly, I include geographical controls to ensure that risk taking behaviour is affected by ancestral lifestyle and not associated variation in climatic

conditions. Adding the controls and fixed effects decreases the coefficient on nomadic lifestyle, but it remains large and statistically significant. It can be concluded that having a nomadic background in preindustrial times increases the willingness to take risks by approximately 7 percentage points.

Column 4 documents the negative relationship between nomadic background and loss/risk aversion. People with nomadic background tend to value less job security among other characteristics, which signals of their lower degree of loss aversion. Adding the controls and region-time fixed effects in Column 6, it shows that ancestral nomadism reduces the choice of job security as the most valuable job characteristics by 17.2 percentage points.

Similarly, people with nomadic background tend to value less the security of their surroundings (Column 7). Controlling for individual-, group-level variations and including geographical covariates, keeping region and time fixed, it demonstrates that nomadic background makes people to care about feeling secure by 17 percentage points less than their sedentary counterparts. Overall, using these proxies for the willingness to take risks shows that nomadic background in preindustrial times is associated with higher degree of risk taking today.

WVS Results for movers

The analysis of movers (people living outside their ancestral homeland) accounts for time invariant unobserved heterogeneity in the country of current residence (e.g., geographical and institutional characteristics). Moreover, since wide range of characteristics of the place of origin are distinct from those of the country of residence, the estimated effect captures the culturally-embodied, intergenerationally-transmitted effect, rather than the direct effect of geography or environment.

I define movers as people living outside their ethnic homeland. The way I allocate people is as follows: each ethnic group has location defined by centroid on the map. I follow Michalopoulos et al (2012) and draw a boundary at 200km around, anyone who is living 10 km further is considered a mover (Figure 1.14 in Appendix).

Table 1.5 demonstrates the results using baseline specification. Adding the controls and region-time fixed effects, results for movers mirror the main fidnings. Column 1 shows that nomadic background in preindustrial time increases willingness to take risks by 6 percentage points.

Similarly, in column 2 people with nomadic background tend to value job security less against other job characteristics. They also tend to put less value on the security of surroundings (column 3). The magnitude and direction of the coefficient is in line with the baseline results from complete sample.

	TakingRisks	JobSecurity	FeelingSecure
Nomad	0.06**	-0.15**	0.05**
	(0.078)	(0.067)	(0.023)
Individual Controls	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Mean (Y)	0.31	0.65	0.79
SD (Y)	0.45	0.44	0.39
# of obs	27551	27551	27551
Clustered s.e	Yes	Yes	Yes

Table 1.5 WVS Regression Results for people living outside their ethnic homeland

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, highest education, religion, unemployed, income group, age, financial situation satisfaction, working class (e.g. middle class), urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Table 1.6 shows the results for more restricted sample where I focus on the people who have nomadic or sedentary ancestors from one region, but now they are residing in another region. I am also able to identify people who are coming from different sedentary and nomadic backgrounds, but live outside both of their ethnic homelands. For instance, I observe individual 1 from ethnicity 1 (sedentary) and individual 2 from ethnicity 2 (nomadic), both living outside their ethnic homeland, in the area of ethnicity 3. The results are almost completely similar to the movers' sample. These findings capture the culturally embodied component of the effect of nomadic lifestyle and accounts for time invariant unobserved heterogeneity in the place of residence mitigating concerns about place-specific characteristics.

Overall, this is an evidence that descendants of nomads have greater willingness to take risks, lower loss aversion and they put less value on security in general. It signals that our ancestry has an impact on risk attitudes. In order to bring another dimension of risk to this analysis, I use DHS data to evaluate whether ancestry is also affecting risk behaviour.

	TakingRisks	JobSecurity	FeelingSecure
Nomad	0.09**	-0.17**	0.07**
	(0.081)	(0.063)	(0.031)
Individual Controls	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Table 1.6 People living outside their ethnic homeland in the third region

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, highest education, religion, unemployed, income group, age, financial situation satisfaction, working class (e.g. middle class), urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

1.5.2 DHS

In this section I employ the same baseline specification to analyze the effect of nomadic lifestyle on risky health behaviour. Nosic & Weber (2007), Galizzi & Miraldo (2014), Dohmen et al (2014), Teh et al (2019), Arni et al (2020) among others showed that risk preferences, in particular, willingness to take risks, are positively correlated with several determinants of risky health behaviour. For instance, smoking, drinking and having unsafe sex. Taking health indicators as a proxy of risky behaviour, I evaluate whether having a nomadic background affects risky behaviour in addition to the evidence on risk preferences discussed in the previous section.

DHS Baseline Results

Table 1.7 presents the results analogous to the baseline results for risk preferences proxied with WVS data. It can be seen that direction of the effect is supporting the hypothesis that having a nomadic ancestry makes people to adopt riskier health behaviour. Column 1 of Table 1.7 shows the positive correlation between nomadic ancestry and smoking habits. Inclusion of individual-, group-level controls alongside geographical ones makes the coefficient to double (column 2). Adding region and time fixed effects to level out regional differences, e.g. institutional or social, increases the coefficient on nomadic background even further (column 3). In particular, nomadic ancestry increases the probability of the person to be a regular smoker by 3.8 percentage points[5].

---

[5]It can be also noticed that for some cases the coefficient increased instead of going down when I add controls. The possible explanation is that these controls have an opposite effect on the dependent variable, so their inclusion increases the effect of our target one

	Smoking_habits			Health_insurance			Contraception_use		
Nomadic	0.013***	0.0201*	0.038***	-0.0302***	-0.1215*	-0.0862**	-0.069***	-0.089*	-0.014*
	0.0013	0.0104	0.013	0.0064	0.0619	0.0352	0.0025	0.049	0.006
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Group Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Geographical Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Region FE	No	No	Yes	No	No	Yes	No	No	Yes
Time FE	No	No	Yes	No	No	Yes	No	No	Yes
Mean (Y)	0.57	0.57	0.57	0.037	0.037	0.037	0.67	0.67	0.67
SD (Y)	0.23	0.23	0.23	0.189	0.189	0.189	0.46	0.46	0.46
# of obs	211990	211990	211990	202768	202768	202768	331344	331344	331344
# of regions	403	403	403	403	403	403	403	403	403
# of ethnicities	332	332	332	332	332	332	332	332	332
Clustered s.e	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.7 DHS Baseline Results

Note: Standard errors are clustered at ethnic group level and the analysis covers around 300,000 individuals in 403 regions and 332 ethnicities. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Concerning health insurance the results are similar. From column 4 it can be seen that the correlation with nomadic ancestry is negative and the coefficient goes up after adding the controls and region time fixed effects (column 5-6). The final estimate is 8.6 percentage point increase in the likelihood of not having health insurance, which is huge given the low mean of the dependant variable. Potential explanation could be that there are few regions in the world where health insurance exists as a state policy, therefore, there is a small number of people actually having health insurance. Hence, people do not have it because of lack of service and not by choice.

As regards to contraception, the behaviour of people with nomadic background tend to be riskier compared with sedentary counterparts. Column 7 shows that the correlation is negative, while the coefficient increases after adding controls (column 8). The interesting point is that the effect of nomadic ancestry decreases drastically when I introduce region time fixed effects. It could be due to the regional differences in norms and contraception behaviour, respectively. Some regions may not support contraception informally due to cultural or social views regardless of the state of legislation on this matter.

Baseline results for risk behaviour are in line with the evidence on risk preferences. Particularly, nomadic ancestry has an impact on risk attitudes nowadays. Risk taking behaviour imposed by nomadic lifestyle holds over generations and is well pronounced among descendants of nomads. Yet one may think that DHS results are not compatible with WVS, since the number of ethnic groups increases substantially and may affect the findings. For this reason I restrict DHS sample to exactly same ethnic groups covered in WVS and redo the evaluation for risk health behaviour proxies. Table 1.8 shows the results for 104 ethnic groups used in WVS. I report only results

	Smoking_habits		Health_insurance		Contraception_use	
Nomadic	0.018**	0.029**	-0.068**	-0.065**	-0.071*	-0.017*
	0.009	0.011	0.0517	0.033	0.041	0.005
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Time FE	No	Yes	No	Yes	No	Yes
# of obs	65323	65323	34791	34791	64901	64901
# of ethnicities	104	104	104	104	104	104
Clustered s.e	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.8 DHS Baseline Results: same sample as WVS

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

with included controls and fixed effects skipping simple correlation for better comparison. It can be seen that the results are very much in line with the whole sample, magnitude and direction of the effect hold and support the hypothesis established above.

DHS Results for movers

This section uses moving as a setting to identify the portable component of cultural influence on current outcomes. First, I check if these results hold among those people, who have moved to the current place of residence as it is indicated in DHS data. Second, I focus on movers, who are living outside their original ethnic homeland as I defined above.

Table 1.9 presents the results for people who have moved to the current place of residence as indicated in DHS data. They are in line with the baseline results from the whole sample. I will focus on columns 3, 6, 9 since they include all controls and region time fixed effects. It can be seen that the coefficient of interest has the same direction across all three variables. Magnitude stay the same for smoking and contraception usage. For instance, nomadic ancestry increases the likelihood to be a regular smoker by 4.2 percentage points and decreases usage of contraception by 1.5. However, the coefficient for health insurance decreases a lot and lost its significance. One possible explanation for decrease in the coefficient is that, potentially, movers are people moved for a better quality of life, so the choice of having health insurance goes in line with risk attitudes and most probably we are comparing people who have an opportunity to get insurance,

	Smoking_habits			Health_insurance			Contraception_use		
Nomadic	0.009***	0.026*	0.042**	-0.028***	-0.028*	-0.025	-0.013***	-0.022*	-0.015*
	0.002	0.0105	0.016	0.007	0.015	0.027	0.001	0.0127	0.0113
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Group Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Geographical Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Region FE	No	No	Yes	No	No	Yes	No	No	Yes
Time FE	No	No	Yes	No	No	Yes	No	No	Yes
Mean (Y)	0.54	0.54	0.54	0.04	0.04	0.04	0.66	0.66	0.66
SD (Y)	0.22	0.22	0.22	0.19	0.19	0.19	0.47	0.47	0.47
# of obs	102103	102103	102103	87870	87870	87870	161954	161954	161954
# of regions	347	347	347	347	347	347	347	347	347
# of ethnicities	270	270	270	270	270	270	270	270	270
Robust s.e	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.9 DHS Results for Movers

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
p <0.10,** p <0.05,*** p <0.01

but make their choices according to risk preferences. Yet, statistical significance is probably lost due to the shrinkage of the sample and introduction of fixed effects.

	Smoking_habits			Health_insurance			Contraception_use		
Nomadic	0.007**	0.028**	0.033**	-0.011***	-0.044*	-0.092	-0.09**	-0.005*	-0.015*
	0.0014	0.0102	0.0117	0.001	0.02	0.064	0.002	0.0021	0.005
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Group Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Geographical Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Region FE	No	No	Yes	No	No	Yes	No	No	Yes
Time FE	No	No	Yes	No	No	Yes	No	No	Yes
Mean (Y)	0.19	0.19	0.19	0.03	0.03	0.03	0.8	0.8	0.8
SD (Y)	0.39	0.39	0.39	0.16	0.16	0.16	0.39	0.39	0.39
# of obs	26487	26487	26487	26487	26487	26487	26487	26487	26487
# of regions	193	193	193	193	193	193	193	193	193
# of ethnicities	139	139	139	139	139	139	139	139	139
Robust s.e	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.10 DHS Results for GeoMovers (more than 10 km from ancestral homeland)

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
p <0.10,** p <0.05,*** p <0.01

Despite the fact that movers's sample was giving the opportunity to find the portable component of culture, due to the data limitation it is not possible to track whether people have moved far from their origins or just changed the house. For this reason, using the geolocation of households

helps to better identify those respondents, who live outside their respective ethnic homeland. These individuals, so-called geomovers, are defined in the same way as described before. Each ethnic homeland is defined by the centroid and 200 km buffer around it. Everyone living 10 km further from the homeland are considered geomovers. 10 km is chosen due to the official perturbation of borders by 5-10 km done by DHS data constructors.

Table 1.10 shows the results for geomovers, which are very close both in magnitude and direction with the baseline ones. From column 3 it can be seen that the probability to be a regular smoker is increased by 3.3 percentage points for descendant of nomadic lifestyle. Similarly, nomadic ancestry is negatively associated with having health insurance and decrease the likelihood of safe sex practices by 1.5 percentage points.

Table 1.11 demonstrates the results for more restricted sample where I focus on people who have nomadic or sedentary ancestors from one region, but now they are residing in another region. The results are almost completely similar to the both movers' samples. These findings capture the culturally embodied component of the effect of nomadic lifestyle and accounts for time invariant unobserved heterogeneity in the place of residence mitigating concerns about place-specific characteristics.

	Smoking_habits	Health_insurance	Contraception_use
Nomadic	0.041**	-0.089	-0.055*
	(0.012)	(0.071)	(0.021)
Individual Controls	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Table 1.11 GeoMovers: nomadic and sedentary descendants in the third region

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
p <0.10,** p <0.05,*** p <0.01

Overall, these findings bring the evidence that nomadic ancestry not only affects risk preferences, but also risk behaviour. In both cases, the findings suggest strong negative correlation between nomadic lifestyle and risk aversion. In particularly, descendants of nomads are more willing to take risks and exhibit riskier health behaviour. However, given the nature of the data collection and multiple ethnographic sources, one may be concerned about measurement error problems,

which may cause downward bias in OLS.

Another natural question to ask is how nomadism originated. This is important, since one may be concerned that similar factors determining lifestyles also determine individual outcomes. For instance, certain cultures may be more militant or belligerent, which induce practicing nomadism to conquer other lands. Thus, these people would more likely to choose nomadic way of living and be risk takers, but nomadism itself would not be relevant. In this case, OLS results may overstate the effect of nomadism. In order to mitigate these concerns, I apply instrumental variable approach that will be described in more details in the next section.

1.5.3 IV: The importance of origins of biodiversity

An important question when interpreting the results presented in the preceding sections is what generates the variation in ancestral lifestyle. It has been hypothesized that an important determinant of whether the tribe became nomadic was the characteristics of animals that were present in the same habitual areas (Vigne et al, 2011). To lead the nomadic life one needs a support that makes mobility sustainable.
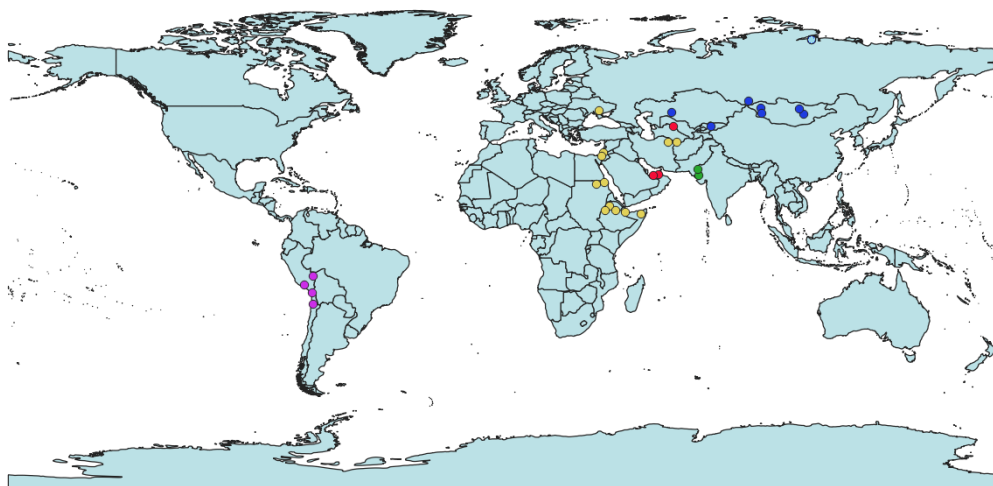
Figure 1.9 Original domestication points

Notes: each dot corresponds to the centroid of the location where the first domestication took place. Purple – llamas/pacos, blue – horses, red – camels, green – elephants, yellow – donkeys, onagers. Data taken from Rossel et al, 2008, Larson & Fuller, 2014, Almathen et al, 2016, Kaczensky et al, 2018, Khan, 2019.

Indeed, in different parts of the world known nomadic tribes extensively used animals who can travel at long distances and are of great endurance: horses, camels, donkeys/onagers, llamas and elephants. Thus, in the areas where those animals were first domesticated during early Neolithic time (mid-Helocene period), it was higher probability that people would choose to pursue nomadic lifestyle.

Since 2006, the study of the domestication process through complete genome sequence has become possible, and it has been associated with the detection of selection in a large number of genomic loci that have likely evolved by selective pressures (Carneiro et al, 2014; Larson et al, 2014). Therefore, in order to test my argument, I collected the data from genetic biology literature (Rossel et al, 2008, Larson & Fuller, 2014, Almathen et al, 2016, Kaczensky et al, 2018, Khan, 2019), which dissect genetic codes in the modern animals to track their existence back to the common ancestor of one specie. Knowing the common ancestor made it possible to locate the origins of where the domestication took place for the first time (please refer to Figure 1.9).
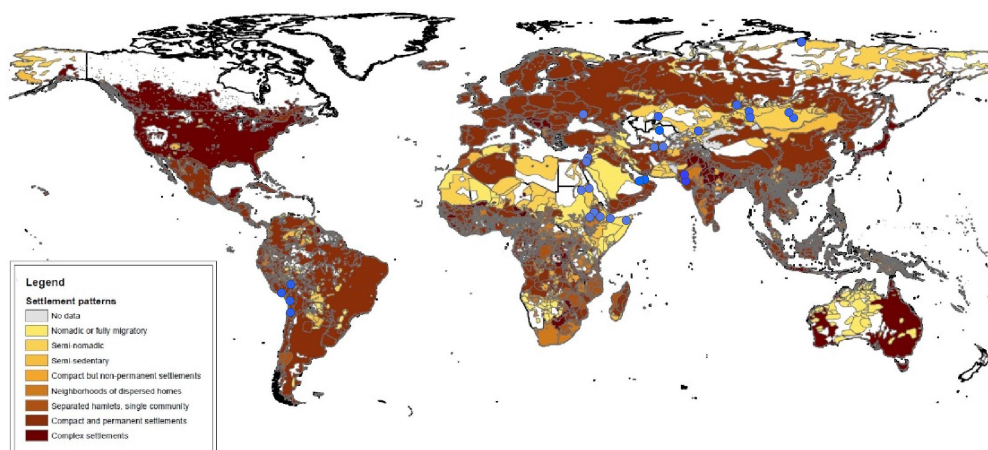
Figure 1.10 Distribution of nomads and original locations of domestication

Notes: data taken from Rossel et al, 2008, Larson & Fuller, 2014, Almathen et al, 2016, Kaczensky et al, 2018, Khan, 2019 - for domestication points; from Ethnographic Atlas - for distribution of nomads.

Importantly for this identification strategy, the origins of domestication strongly influences the likelihood that and the extent to which an ethnic group practiced nomadism in pre-industrial times. More specifically, the closer an ethnic group was to the point of domestication, the more likely it will adopt nomadism. Figure 1.10 depicts this relationship by overlaying the map of distribution of nomadic tribes with first domestication sites.
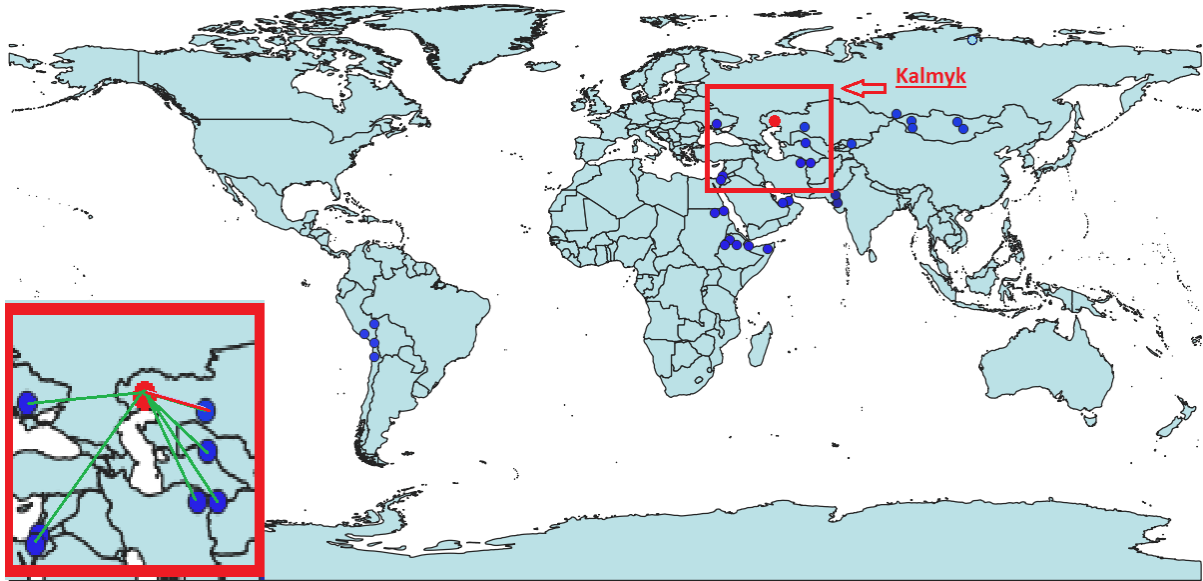
Figure 1.11 Shortest distance as an instrument

Notes: data taken from Rossel et al, 2008, Larson & Fuller, 2014, Almathen et al, 2016, Kaczensky et al, 2018, Khan, 2019 - for domestication points.

Since I know the location of each ethnic group, I am able to calculate the set of Euclidian distances between the centroid of certain ethnic group location and of the first domestication point of each animal that was used for travelling and carrying loads at long distances. Afterwards, I choose the shortest distance in the set which will be linked to particular ethnic group in the dataset that I have constructed (refer to Figure 1.11).

	Being nomadic
The shortest distance to the point of domestication	-0.011*** (0.0012)
Controls	Yes
Region FE	Yes
Time FE	Yes
Obs	43274
F-statistic	87.34

Table 1.12 Instrumental variable analysis: First stage

Note: Standard errors are clustered at ethnic group level.
p <0.10,** p <0.05,*** p <0.01

Therefore, in order to provide evidence that the relationship between nomadism and current risk attitudes is causal, I use the shortest distance to the point of domestication as an instrument for nomadic lifestyle. The assumption underlying the exclusion restriction is that origins of domestication do not affect modern risk attitudes through channels other than nomadism. Table 1.12 provides the first stage regression. The first stage F-statistic is 87.34 and for the exclusion restriction to hold it needs to be true that location of domestication does not affect risk attitudes

through channels other than nomadism. Generally speaking, I do not expect the direct effect of the distance between ethnic group location and domestication point on the risk attitudes nowadays (around 6000 years after). However, one may be concerned whether environmental or climate conditions were driving the effect rather than the lifestyle. Looking at the Figure 1.12, it can be seen that the estimations made recently with ice samples from Greenland cores show the fairly stable temperature around 4000BC (corresponds to 6BP), when most of domestication happened, and nowadays.
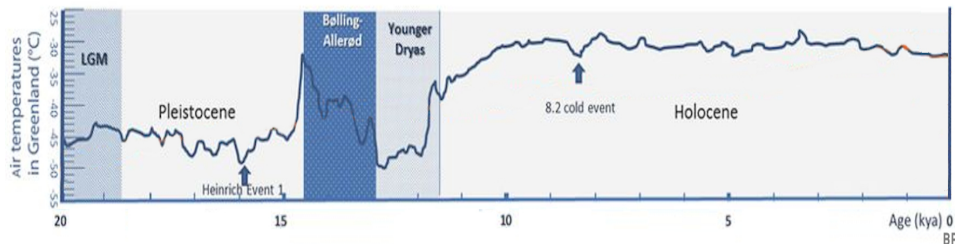
Figure 1.12 Evolution of the temperature in the Post-Glacial period
Source: Zalloua & Matisoo-Smith, 2017

Using the data from National Climatic Data Center and estimations from different parts of the world, Figure 1.13 also demonstrates that the temperature was fluctuating no more than within 0.5 degree during the last 5-6 thousands years.
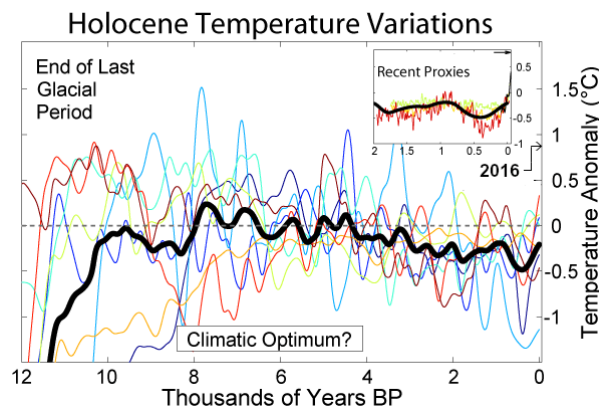
Figure 1.13 Holocene Temperature variations
Source: National Climatic Data Center

Therefore, adding climate controls from 1835 till nowadays aims to alleviate the concern that distance towards domestication point may affect current risk attitudes, because of climatic and environmental conditions of the past.

Another potential concern is that domestication was endogenous given the skills of people indigenous to the area to which the species was native. In other words, whether nomads were

more likely to tame large animals due to their skills rather than it was a random event. On this I follow Diamond (1997, 2002), who argues that although domestication arose in only a few areas of the world and in certain places of those areas earlier than in others, it was a biogeograhic luck overall. The roots of animal domestication included the ubiquitous tendency of all people to try to tame or manage wild animals (including such unlikely candidates as ospreys, hyenas and grizzly bears). Diamond (1997, 2002) questioned the selectivity of domestication in two ways: whether the difficulty lied with the species itself or with people indigenous to the area where species were native. He claims that there are six independent lines of evidence which are converging to prove that the difficulty lied with the species itself, not with local people. Particularly, the quick acceptance of introduced Eurasian domesticates by non-Eurasian people; the fast ancient domestication of the most valued species; the repeated independent domestication of many of them; the failure of even modern animal breeders to make a considerable contribution to the short list of valuable domesticates; the identification of certain reasons preventing the domestication of many species.

To go more into details, Diamond also provided an example on how European horse breeders who settled in South Africa in the 1600s tried to domesticate zebra (like African herders did for previous millennia). After several centuries they gave up, because zebras are vicious, have the bad habit of biting a handler and not letting go until the handler is dead. Moreover, they have better peripheral vision than horses, making it impossible even for professional rodeo cowboys to lasso, since they see the coming rope and flick away their head.

Given divergence across species, Diamond classified six main obstacles for wild mammal to be domesticated: the diet which is hard to be provided by humans, slow growth rate and long birth spacing (e.g. gorillas), nasty disposition (grizzly bears and rhinoceroses), reluctance to breed in captivity (pandas and cheetahs), lack of follow-the leader dominance hierarchies (bigorn sheep and antelope) and tendency to panic in enclosures or when facing predators (deer, gazelles). There are many species which satisfy five out of six criteria, but were not domesticated. Recent development in genetic biological literature (Larson & Fuller, 2014; Blaustein, 2015) showed that domestication happened due to certain genetically random generated characteristics and supported Diamond's claim that this is a biogeographic luck to be in a place where species were successfully domesticated. Last, but not least, this instrument addresses the concern of some groups being belligerent and having culture or spirit of conquests. However, domestication of animals, that are capable of long distance travel, captures the possibility of groups to travel far and conquer new lands. Overall, I believe that presented evidence is a strong motivation towards exogeneity of this instrument.

Table 1.13 and 1.14 present the results. Column 2, 6, 10 show the OLS coefficient from baseline results including all controls and fixed effects. Columns 4, 8, 12 present the respective second

	TakingRisks: OLS		TakingRisks: IV		JobSecurity: OLS		JobSecurity: IV		FeelingSecure: OLS		FeelingSecure: IV	
Variables	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
Nomad	0.087**	0.070**	0.163***	0.171***	-0.189**	-0.187**	-0.371***	-0.373***	-0.104	-0.181*	-0.327***	-0.326***
	(0.029)	(0.026)	(0.081)	(0.079)	(0.070)	(0.079)	(0.077)	(0.086)	(0.116)	(0.086)	(0.141)	(0.142)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs	43421	43421	43274	43274	43421	43421	43274	43274	28476	28476	23698	23698

Table 1.13 WVS Baseline Results

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, highest education, religion, unemployed, income group, age, financial situation satisfaction, working class (e.g. middle class), urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

stage IV coefficients. The IV estimates confirm the results from the OLS regressions in terms of coefficient sign and statistical significance for both risk preferences and risk behaviour. Historical nomadism predicts risk attitudes of people nowadays. In terms of magnitude, the IV coefficients are consistently larger than their OLS counterparts for set of risk-taking proxies, while IV coefficients on the determinants of risky health behaviour are mostly consistent with OLS results. One potential explanation for the difference in coefficient size is the non-perfect take-up of the treatment. It could be that those groups who 'received treatment', did so, because they live closer to domestication points, but they do not practice nomadism. Another plausible explanation that could result in downward bias in the OLS coefficients is measurement error in defining ancestral lifestyles. The measure of lifestyle and subsistence for each society was collected by a different ethnographers in different ways, which possibly induced biased OLS results. Moreover, the binary nature of the variable may cause the measurement error alone.

	Smoking habits: OLS		Smoking habits: IV		Health insurance: OLS		Health insurance: IV		Contraception use: OLS		Contraception use: IV	
Variables	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
Nomad	0.0201*	0.038**	0.041***	0.045***	-0.0715**	-0.0862**	-0.087	-0.096	-0.089*	-0.014*	-0.079**	-0.077**
	(0.0104)	(0.013)	(0.015)	(0.016)	(0.0619)	(0.0352)	(0.124)	(0.189)	(0.049)	(0.006)	(0.047)	(0.045)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs	211990	211990	211990	211990	95262	95262	95262	95262	313344	313344	313344	313344

Table 1.14 DHS Baseline Results

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, wealth group, age, having bank account, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, belief in God, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator and distance to the coast.
$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Overall, IV approach supports the story in which the ethnic groups who were close to the origins of animal domestication, were more likely to become nomadic, and that nomadism then conferred portable characteristics on descendants of these groups which made them less risk averse even after moving away from their homelands. As expected, IV approach also provides more precise estimation by correcting existing biases in OLS. This identification approach increases the confidence in the conclusion derived from OLS, and provides further support for a causal interpretation of the results.

1.6 Evidence from Russia

This section compliments the main analysis across countries and regions by zooming into one country with great variety of different ethnicity and several geographical zones, namely, Russia. Focusing on Russia would address the concern of imperfect accounting for the time-varying differences across countries and regions. It also brings an additional evidence to the influence of nomadic lifestyle on current determinants of risk taking behaviour. The same empirical strategy exploiting instrumental variable approach discussed above is used for this analysis. The data comes from the Russia Longitudinal Monitoring Survey (RLMS, hereafter), which is a series of nationally representative surveys designed to monitor the effects of Russian reforms on the health and economic welfare of households and individuals in the Russian Federation including relevant community- and region-level data.

Similarly to the main section, these data were manually matched with the ethnographic information at the ethnic group level, which produced successful match of 132 ethnic groups. Overall, the data covers 39 regions within Russia. The peculiarity of this analysis lies in the different set of proxies for risk taking behaviour, which provides better understanding of the nature of risk taking. Particularly, I focus on life insurance, engagement into entrepreneurship, perception of security and insurance of physical assets, such as car or flat, which proxy risk taking behaviour and are complimentary to the main analysis.

In particular, I use the following questions: 1) Have voluntary insurance: life insurance; 2) Engagement into entrepreneural activity; 3) Importance of security; 4) Have voluntary insurance: flat, car or dacha. I construct dummy variables to keep the uniformed specification. From Table 1.15 it can be seen that the direction and the magnitude of the effect is in line with the main analysis. As expected, having nomadic background has strong negative impact on the probability to get insurance for life and for physical assets, around 2 percentage points less probable. The result on security is similar to the main analysis, where nomadic background is associated with

caring less about security of surroundings. More importantly, nomadic background is strongly associated with the probability to be engaged into entrepreneurship, which is one of the main determinants of risk taking behaviour (Rauch et al, 2004, Komulainen et al, 2009, Becker, 2020). Having nomadic ancestry increases the chance that the person becomes an entrepreneur by 10 percentage points. Hence, these results bring additional support to the established cross-country evidence regarding the effect of nomadic ancestry on risk taking behaviour.

	Life insurance	Entrepreneurship	Security	Physical assets insurance
Variables	Col1	Col2	Col3	Col4
Nomadic	-0.021***	0.104**	-0.321*	-0.015**
	(0.00704)	(0.0530)	(0.191)	(0.00615)
Individual Controls	Yes	Yes	Yes	Yes
Group Controls	Yes	Yes	Yes	Yes
Geographical Controls	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
# of obs	41,353	17,853	4,334	39,192
# of ethnicities	132	132	132	132
# of regions	39	39	39	39

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 1.15 RLMS IV Results

Note: Standard errors are clustered at ethnic group level. Individual controls include gender, having partner, number of children, years of education, religion, occupation, income, age, urban; group controls include domestic organization, inheritance rules, mean size of the community, jurisdictional hierarchy, subsistence; geographical controls include historical average temperature and humidity, elevation, distance to the equator.

1.7   Conclusion

This paper provides an evidence that heterogeneity in risk attitudes can be understood through the prism of ethnic roots. In particular, difference in ancestral lifestyle has created certain traits that differentiate people in their perception of risk. This paper shows that people having nomadic ancestry tend to be more willing to take risks and practice riskier health behaviour than those with sedentary background. This relationship holds at individual level bot country and region fixed effects across more than 40 thousand people in 450 regions of 80 countries in risk preference sample and across more than 300 thousand people in 403 regions of 42 countries for risk health behaviour sample. Moreover, it is robust to a wide range of specifications with a large number of controls. Furthermore, by making use of the fact that variation in nomadic lifestyle is determined by biodiversity, this paper suggests a novel instrumental variable approach by using location of domestication of the animals used for long distance travel as an instrument

for the adoption of nomadic lifestyle. These findings show that IV coefficients are similar in sign and significance to the initial OLS results with a small divergence. They conclude that people, whose ancestors historically practiced nomadism (i) are more willing to take risks, (ii) value security less, and (iii) have riskier health behaviour. The paper thereby points out that risk attitudes could be rooted in culture and can be explained by the difference in the lifestyle that our ancestors have led.

1.8 Bibliography

Akasaki,. M, Ploubidis, G.,B., Dodgeon, B., Bonell, C.,P., (2019). The clustering of risk behaviours in adolescence and health consequences in middle age. Journal of Adolescence, 77:188-197

Alan, S., Ertac, S., (2018). Fostering Patience in the Classroom: Results from A Randomized Educational Intervention. Journal of Political Economy, 126 (5). 1865 - 1911

Alesina, A., Giuliano, P. (2010). The power of the family. Journal of Economic growth, 15(2), 93-125.

Alesina, A., Giuliano, P., Nunn, N. (2013). On the origins of gender roles: Women and the plough. The Quarterly Journal of Economics, 128(2), 469-530.

Alesina, A., Michalopoulos, S., Papaioannou, E. (2016). Ethnic inequality. Journal of Political Economy, 124(2), 428-488.

Algan, Y., Cahuc, P. (2010). Inherited trust and growth. American Economic Review, 100(5), 2060-92.

Almathen, F., Charruau, P., Mohandesan, E., Mwacharo, J. M., Orozco-terWengel, P., Pitt, D., Abdussamad, A. M., Uerpmann, M., Uerpmann, H.-P., De Cupere, B., Magee, P., Alnaqeeb, M. A., Salim, B., Raziq, A., Dessie, T., Abdelhadi, O. M., Banabazi, M. H., Al-Eknah, M., Walzer, C., Burger, P. A. (2016). Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. Proceedings of the National Academy of Sciences of the United States of America, 113(24), 6707–6712.

Arni, P., Dragone, D., Götte, L., Ziebarth, N., (2020). Biased Health Perceptions and Risky Health Behaviors: Theory and Evidence, IZA Discussion Papers, No. 13308, Institute of Labor Economics (IZA), Bonn

Bahrami-Rad, D, A. Becker, and J Henrich, (2021). Tabulated Nonsense? Testing the Validity of the Ethnographic Atlas. Economic Letters 204: 109880

Becker, A., (2019). On the Economic Origins of Restrictions on Women's Sexuality. CESifo Working Paper Series, No. 7770

Becker, A., (2020). Pastoralism and Female Entrepreneurship. Mimeo. [Online] Available from: https://sites.google.com/site/ankebeckerecon/EntrepreneurshipJanuary2020.pdf

Becker, A., Falk, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., (2018). Global Evidence on Economic Preferences. Quarterly Journal of Economics, 133(4), pp. 1645–1692

Becker, Anke, Benjamin Enke, and Armin Falk, (2020). Ancient Origins of the Global Variation in Economic Preferences. AEA Papers and Proceedings 110: 319–323

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. (2010). Social Identity and Preferences. American Economic Review, 100 (4): 1913-28.

Blau, F. D., Kahn, L. M. (2015). Substitution between individual and source country characteristics: Social capital, culture, and us labor market outcomes among immigrant women. Journal of Human Capital, 9(4), 439-482.

Blaustein, R., (2015). Unraveling the Mysteries of Animal Domestication. BioScience 65: 7–13

Bollig, M., Göbel, B., (1997). Risk, uncertainty, and pastoralism: An introduction. In Risk and uncertainty in pastoral societies, 5-21. Nomadic Peoples (NS) 1. Oxford: Berghahn.

Bover, O., Coronado, E., Velilla, P. (2014). The Spanish survey of household finances (EFF): description and methods of the 2011 wave.

Breuer, W., Salzmann, A. J. (2012). National culture and corporate governance. In Corporate Governance (pp. 369-397). Springer, Berlin, Heidelberg.

Brown, A.S. (2007), Identifying risk using a new risk assessment tool: the missing piece of the jigsaw in medical device risk assessment, Clinical Risk, Vol. 13 No. 1, pp. 56-9.

Buggle, J. C., Durante, R. (2017). Climate risk, cooperation, and the co-evolution of culture and institutions.

Cashdan, E., (1990). Risk and uncertainty in tribal and peasant economies. Westview, Boulder

Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, et al. (2014) The Genomic Architecture of Population Divergence between Subspecies of the European Rabbit. PLoS Genet 10(8): e1003519.

Carroll, C. D., B.-K. Rhee, and C. Rhee (1994). Are there Cultural Effects on Saving? Some Cross-Sectional Evidence. The Quarterly Journal of Economics 109(3), 685–699.

Carroll, C. D., B.-K. Rhee, and C. Rhee (1999). Does Cultural Origin Affect Saving Behavior? Evidence from Immigrants. Economic Development and Cultural Change 48(1), 33–50.

Cawley, J., Ruhm, C., (2011). The Economics of Risky Health Behaviors. NBER Working Paper No. 17081

Cervellati, M., Sunde, U., Zimmermann, K., (2017). Demographic dynamics and long-run development: insights for the secular stagnation debate. Journal of Population Economics, 30, pp. 401 – 432

Cesarini, D., Dawes, C.,Johannesson, M.,Lichtenstein, P., Wallace, B., (2009). Genetic Variation in Preferences for Giving and Risk Taking, The Quarterly Journal of Economics, Volume 124, Issue 2, Pages 809–842

Chanda, A., Cook, C. J., Putterman, L. (2014). Persistence of fortune: Accounting for population movements, there was no post-Columbian reversal. American Economic Journal: Macroeconomics, 6(3), 1-28.

Chang, C., & Koster, H., (1994). Pastoralists at the periphery: Herders in a capitalstic world. Tucson: University of Arizona Press

Chatty, D. (1972). Structural forces of pastoral nomadism, with special reference to camel pastoral nomadism (No. 16).

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. American Economic Review, 103(2), 690-731.

Christelis, D., Georgarakos, D., Jappelli, T. (2015). Wealth shocks, unemployment shocks and consumption in the wake of the Great Recession. Journal of Monetary Economics, 72, 21-41.

Comin, D., Easterly, W., Gong, E. (2010). Was the wealth of nations determined in 1000 BC?. American Economic Journal: Macroeconomics, 2(3), 65-97.

Cronqvist, H., Siegel, S., (2015). The Origins of Saving Behavior. Journal of Political Economy, 123(1), 123–169.

Desmet, K., Ortuño-Ortín, I., Wacziarg, R. (2012). The political economy of linguistic cleavages. Journal of development Economics, 97(2), 322-338.

Diamond, J., (1997). Guns, Germs and Steel. A Short History of Everybody for the Last 13,000 Years. 480 pp. London: Jonathan Cape

Diamond, J., (2002). Evolution, consequences and future of plant and animal domestication. Nature 418, 700–707

Dohmen, T., Falk, A., Huffman, D., Sunde, U., (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? American Economic Review, 100 (3): 1238-60.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G., (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioural Consequences. Journal of the European Economic Association, 9, issue 3, p. 522-550.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., (2012). The Intergenerational Transmission of Risk and Trust Attitudes, The Review of Economic Studies, Volume 79, Issue 2, Pages 645–677

Dohmen, T., Falk, A., Huffman, D., Sunde, U., (2017). Risk Attitudes across the Life Course, The Economic Journal, Volume 127, Issue 605, Pages F95–F116

Edgerton, R. B. (1971). The individual in cultural adaptation: A study of four East African peoples.

Falk, A., Becker, A., Dohmen, T. J., Enke, B., Huffman, D. (2015). The nature and predictive power of preferences: Global evidence.

Fehr, E., Fischbacher, U., Naef, M., Schupp, J., and Wagner, G., (2006). A Comparison of Risk Attitudes in Germany and the U.S.. Mimeo, Institute for Empirical Research in Economics, University of Zurich.

Fernandez, R. (2007). Alfred Marshall lecture women, work, and culture. Journal of the European Economic Association, 5(2-3), 305-332.

Fernandez, R., Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. American Economic Journal: Macroeconomics, 1(1), 146-77.

Fratkin, E. (1997). Pastoralism: Governance and development issues, Annual Review of Anthropology 26: 235-261

Friedman, M., and Savage, L. J., (1948). The Utility Analysis of Choice Involving Risk. Journal of Political Economy, 56, no. 4: 279-304.

Fuchs-Schündeln, N., Masella, P., Paule-Paludkiewicz, H. (2017). Cultural Determinants of Household Saving Behavior. Working Paper.

Furtado, D., Marcén, M., Sevilla, A. (2013). Does culture affect divorce? Evidence from European immigrants in the United States. Demography, 50(3), 1013-1038.

Gächter, S., Schulz, J., (2016). Intrinsic honesty and the prevalence of rule violations across societies. Nature 531, 496–499. [Online] Available from: https://doi.org/10.1038/nature17160

Galaty, J.G., Bonte, P., (1991). Herders, Warriors, and Traders: Pastoralism in Africa (1st ed.). Routledge. [Online] Available from: https://doi.org/10.4324/9780429045615

Galizzi, M.M., Miraldo, M., (2012). Are You What You Eat? Experimental Evidence on Health Habits and Risk Preferences. Discussion Paper No. 2012/04, Healthcare Management Group, Imperial College Business School

Galor, O., Özak, Ö. (2016). The agricultural origins of time preference. American Economic Review, 106(10), 3064-3103.

Galor, Oded and Savitskiy, Viacheslav, (2018), Climatic Roots of Loss Aversion, No 2018-1, Working Papers, Brown University, Department of Economics. [Online] Available from: https://EconPapers.repec.org/RePEc:bro:econwp:2018-1

Giuliano, P. (2007). Living arrangements in Western Europe: Does cultural origin matter?. Journal of the European Economic Association, 5(5), 927-952.

Giuliano, P., Nunn, N. (2013). The transmission of democracy: from the village to the nation-state. American Economic Review, 103(3), 86-92.

Giuliano, P., Nunn, N. (2017). Understanding cultural persistence and change (No. w23617). National Bureau of Economic Research.

Goldschmidt, W., Kunkel, E. J. (1971). The structure of the peasant family. American Anthropologist, 73(5), 1058-1076.

Guin, B.,(2017). Culture and household saving. European Central Bank Working paper series, No 2069

Guiso, Luigi and Paiella, Monica, (2005). The role of risk aversion in predicting individual behaviour. No 546, Temi di discussione (Economic working papers), Bank of Italy, Economic Research and International Relations Area. [Online] Available from: https://EconPapers.repec.org/RePEc:bdi:wptemi:td_546_05

Guiso, L., Sapienza, P., Zingales, L. (2006). Does culture affect economic outcomes?. Journal of Economic perspectives, 20(2), 23-48.

Guiso, L., Sapienza, P., Zingales, L. (2016). Long-term persistence. Journal of the European Economic Association, 14(6), 1401-1436.

Haliassos, M., Jansson, T., Karabulut, Y. (2016). Incompatible European partners? Cultural predispositions and household financial behavior. Management Science, 63(11), 3780-3808.

Jerven, M., (2011) A clash of disciplines? Economists and historians approaching the African past. Econ. Hist. Dev. Reg., 26, pp. 111-124

Kaczensky, P., Kovtun, E., Habibrakhmanov, R. et al. (2018). Genetic characterization of free-ranging Asiatic wild ass in Central Asia as a basis for future conservation strategies. Conservation Genetics 19, 1169–1184

Khan, N., (2019. The genomic origins of modern horses revealed by ancient DNA: from early domestication to modern breeding. Mimeo. University of Copenhagen.

Komulainen, K., Korhonen, M., Räty, H., (2009). Risk-taking abilities for everyone? Finnish entrepreneurship education and the enterprising selves imagined by pupils. Gender and Education - GEND EDUC. 21. 631-649

Kosse, F., Deckers, T., Schildberg-Hörisch, H., Falk, A., (2016). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. Working papers series IZA DP No. 9861

Kraetli, S., Huelsebusch, C., Brooks, S. and B. Kaufmann, (2013). Pastoralism: a critical asset for foodsecurity under global climate change. Animal Frontiers, 3:42-50.

Larson, G., Fuller, D., (2014). The Evolution of Animal Domestication. Annual Review of Ecology, Evolution, and Systematics 2014 45:1, 115-136

Larson, Greger Piperno, D.R. Allaby, Robin Purugganan, Michael Andersson, Leif Arroyo-Kalin, Manuel Barton, Loukas Vigueira, Cindy Denham, Tim Dobney, Keith Doust, Andrew Gepts, Paul Gilbert, M Gremillion, Kristen Lucas, Leilani Lukens, Lewis Marshall, Fiona Olsen, Kenneth Pires, J. Fuller, Dorian. (2014). Current perspectives and the future of domestication studies. Proceedings of the National Academy of Sciences of the United States of America, 111

Leach, E. (1964). Reply to Raoul Naroll's on ethnic unit classification Curr. Anthropol., 5, pp. 283-312

Leder, S., Streck, B., (2005). Nomadic and Sedentary Peoples: A Misleading Dichotomy? The Bedouin and Bedouinism in the Arab Past Shifts and Drifts in Nomad-Sedentary Relations. Wiesbaden Reichert 401 419

Masella, P. (2013). National identity and ethnic diversity. Journal of Population Economics, 26(2), 437-454.

Michalopoulos, S., Papaioannou, E. (2016). The long-run effects of the scramble for Africa. American Economic Review, 106(7), 1802-48.

Michalopoulos, S., Putterman, L., Weil, D. N. (2018). The influence of ancestral lifeways on individual economic outcomes in Sub-Saharan Africa (No. w21907). National Bureau of Economic Research.

Murdock, G. P. (1967). Ethnographic atlas. University of Pittsburgh Press

Murdock G. P., White D. R., (1969). The standard cross-cultural sample. Ethnology, 8, pp. 329-369

Nosic, A., Weber, M., (2007). Determinants of Risk Taking Behavior: The role of Risk Attitudes, Risk Perceptions and Beliefs. Mimeo

Nunn, N. (2012). Culture and the historical process. Economic History of Developing Regions, 27(sup1), S108-S126.

Nunn, N., Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. American Economic Review, 101(7), 3221-52.

Ortigueira, S., Siassi, N. (2013). How important is intra-household risk sharing for savings and labor supply?. Journal of Monetary Economics, 60(6), 650-666.

Rauch, A., Wiklund, J., Freese, M., Lumpkin, G.T.(2004). Entrepreneurial orientation and businessperformance: Cumulative empirical evidence. Paper presented at the 23rd Babson College Entrepreneurship Research Conference. Glasgow, UK

Rass, N., (2006). Policies and strategies to address the vulnerability of pastoralists in sub-Saharan Africa. PPLPI Working Paper No. 37. Pro-Poor Livestock Policy Initiative, FAO, Rome.

Rao, J., (2018). Fundamental Functionings of Landowners: understanding the relationship between land ownership and wellbeing through the lens of 'capability'. Land Use Policy, 72, pp. 74-84

Robson, A., J., and Samuelson, L., (2010). The evolutionary foundations of preferences. In Jess Benhabib, Alberto Bisin, and Matthew Jackson, editors, The Social Economics Handbook. Elsevier, New York.

Rossel, S., Marshall, F., Peters, J., Pilgram, T., Adams, M., O'Connor, D., (2008). Domestication of the donkey: Timing, processes, and indicators. Proceedings of the National Academy of Sciences, 105 (10) 3715-3720

Sadalla, E., Stea, D., (1982). Nomads: Behavioural and psychological adjustments to sedentarization. Geographical Research Forum, No.5, University of California, Los Angeles, U.S.A.

Shaughnessy, S., (2018). How do East Africa's pastoralists deal with risk? Cambridge Judge Business School and McKinsey Co's Global Risk Practice. manuscript.

Spolaore, E., Wacziarg, R. (2009). The Diffusion of Development. The Quarterly Journal of Economics, vol. 124, no. 2, pp. 469-529.

Spolaore, E., Wacziarg, R. (2013). How deep are the roots of economic development?. Journal of Economic Literature, 51(2), 325-69.

Spolaore, E., Wacziarg, R. (2016). Ancestry, language and culture. In The Palgrave handbook of economics and language (pp. 174-211). Palgrave Macmillan, London.

Spolaore, E., Wacziarg, R. (2016). The Diffusion of Institutions. in Wilson, D. S. and A. Kirman, eds. Complexity and Evolution: Toward a New Synthesis for Economics. Strüngmann Forum Reports, vol. 19, chapter 9, pp. 147-166. Cambridge, MA: The MIT Press.

Spolaore, E., Wacziarg, R. (2008). Ancestry and Development: New Evidence. Journal of Applied Econometrics, vol. 33, no. 5, pp. 748-762.

Spooner, B. (1972). Population growth: Anthropological implications. Mass., MIT Press.

Steinfeld, H, Mooney, HA, Schneider, F, Neville, LE, (2010). Livestock in a changing landscape: volume 1, drivers, consequences, and responses. Island Press, Washington

Sutter, M., Angerer, S., Rützler, D., Lergetporer, P. (2015). The effect of language on economic behavior: Experimental evidence from children's intertemporal choices.

Swidler, N. (1973). The political context of Brahui sedentarization. Ethnology, 12(3), 299-314.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. Science, 344(6184), 603-608.

Teh, C.H., Teh, M.W., Lim, K.H. (2019). Clustering of lifestyle risk behaviours and its determinants among school-going adolescents in a middle-income country: a cross-sectional study. BMC Public Health 19, 1177

Vickrey, W., (1945). Measuring Marginal Utility by Reactions to Risk. Econometrica, 13,. 319–333.

Vigne, J. -D., Carrere, I., Briois, F., Guilaine, J., (2011). The early process of mam- mal domestication in the Near East. Current Anthropology 52(supplement 4): S255-271.

Voigtländer, N., Voth, H. J. (2012). Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany. The Quarterly Journal of Economics, 127(3), 1339-1392.

Voigtländer, N., Voth, H. J. (2012). The three horsemen of riches: Plague, war, and urbanization in early modern Europe. Review of Economic Studies, 80(2), 774-811.

Voors, Maarten J., Eleonora E. M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." American Economic Review, 102 (2): 941-64.

Wärneryd, K. E. (1999). The psychology of saving. Books.

Weber, E., Hsee, C., Sokolowska, J., (1998). What Folklore Tells Us about Risk and Risk Taking: Cross-Cultural Comparisons of American, German, and Chinese Proverbs. Organizational Behavior and Human Decision Processes 75(2):170-186

Weber, E. U., Hsee, C. K. (1998). Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards risk. Management Science, 44, 1205–1217

Weber, E. U., Blais, A.-R., Betz, E. (2002). A Domain specific risk-attitude scale: Measuring risk perceptions and risk behaviors. Journal of Behavioral Decision Making, 15, 263–290

World Values Survey Association, (2009). World Values Survey 1981-2014, Official Aggregate. [online] Available from: <www.worldvaluessurvey.org>

Xue, M., (2019). High-Value Work and the Rise of Women: The Cotton Revolution and Gender Equality in China. Revise and Resubmit, Review of Economic Studies

Zalloua, P., Matisoo-Smith, E., (2017). "Mapping Post-Glacial expansions: The Peopling of Southwest Asia". Scientific Reports. 7: 40338.

Zhang R., Brennan T.,J., Lo, A.,W., (2014). Variety Is the Spice of Life: Irrational Behavior as Adaptation to Stochastics Environments. Technical Report (MIT Laboratory for Financial Engineering, Cambridge, MA)

1.9   Appendix

1.9.1   Matching Procedure

- Manually corrected ethnicity names and disentangled from country name

    e.g. BY: Belorussian – Byelorussians

- Manual direct match with ethnographic data
- Manual indirect match based on old original names and/or language/dialect name within language family defined by Glottolog

  e.g. Gheg Albanian – Gheg, Oyo Yoruba – Yoruba

- Glottolog database classification (Max-Planck-Institut für evolutionäre Anthropologie): Languages are classified into 243 families and 189 isolates, i.e., one-member families.

1.9.2   Definition of movers and Differential selection into migration
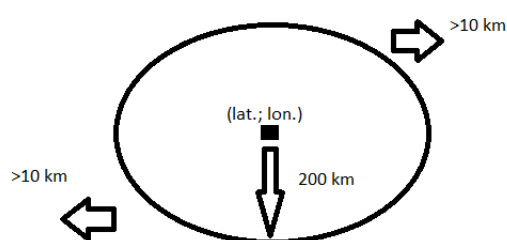
Figure 1.14 Definition of movers

	Smoking_habits	Health_insurance	Contraception_use
Nomadic	0.038**	-0.086**	-0.013*
	0.013	0.035	0.006
Geomover_nomad	0.01	-0.011	-0.05
	0.017	0.007	0.031
All Controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
# of obs	211990	202768	331344
# of regions	403	403	403
# of ethnicities	332	332	332

Table 1.16 Differential Selection into migration

1.9.3 Summary Statistics

Variables	Obs	Mean	SD	Min	Max
Nomadic	59,004	0.147651	0.354757	0	1
TakingRisks	59,004	0.160532	0.367101	0	1
JobSecurity	26,350	0.735332	0.441165	0	1
Secure	28,476	0.799867	0.400107	0	1
Male	59,004	0.477069	0.499478	0	1
HavePartner	59,004	0.594248	0.491041	0	1
NumChildren	57,791	1.837362	1.797377	0	8
HighestEdu	53,161	4.952014	2.118004	1	8
Religion	60,212	26.72436	13.78908	1	53
Age	59,004	39.92753	15.97234	15	98
Unemployed	59,004	0.091672	0.288565	0	1
IncomeDecile	55,152	4.79823	2.304536	1	10
FinStatSatis	59,004	0.311216	0.462994	0	1
LifeSatis	59,004	0.462172	0.498571	0	1
MiddleClass	59,004	0.534371	0.498822	0	1
Immig2nd	60,212	0.031555	0.174814	0	1
marital	52,408	3.690543	2.536686	1	7
dom_organization	52,408	4.529328	2.254636	1	8
crop	52,408	4.854412	0.615117	1	6
mean_size_community	52,408	7.716436	1.304112	1	9
hierarchy	52,408	2.008167	0.823159	1	4
god	52,408	3.598725	1.003393	1	5
games	52,408	5.454549	0.798415	1	6
animal_husbandry	52,408	5.841627	0.708078	1	7
subsistence	52,408	4.780778	0.72563	1	7
class_stratification	52,408	4.325313	1.227905	1	6
slavery	52,408	1.690105	0.953575	1	4
inheritance	52,408	6.171825	1.604382	1	8
rigid_society	52,408	1.969909	0.170839	1	2
tempmean	52,408	7.757191	10.56836	-14.0492	27.35118

Table 1.17 WVS Summary Statistics

Variables	Obs	Mean	SD	Min	Max
smoking_habits	211990	.577291	.2332311	0	1
contraception	500999	.6730652	.4690937	0	1
health_insurance	204104	.0372114	.18928	0	1
Nomadic	500999	0.138928	0.345872	0	1
gender	500999	0.131044	0.337449	0	1
age	500999	29.42893	10.05232	13	91
place_of_residence	500999	1.391909	0.488177	1	2
years_edu	292981	10.59451	3.556208	1	20
religion	433703	61.84139	22.74654	1	119
hhmembers	493085	6.390963	3.431769	1	20
occupation	484749	21.03046	11.11183	1	38
wealth_group	271661	3.069833	1.450098	1	5
marital_status	498675	4.965392	1.832859	1	8
have_bankaccount	409931	2.945181	1.485959	1	6
moved_in_life	433872	0.519176	0.499633	0	1
marital	494369	2.269829	2.014956	1	7
dom_organization	492007	5.393899	2.517769	1	8
crop	482097	5.713031	0.659259	1	6
mean_size_community	348437	5.884435	2.334325	1	8
hierarchy	487439	2.926268	0.63956	2	4
god	399938	2.726888	1.243947	1	4
games	181950	7.171206	1.405336	2	8
subsistence	494369	6.228016	0.943477	1	9
class_stratification	462953	3.389318	1.585731	1	5
slavery	477497	1.962982	0.67999	1	3
inheritance	453114	5.340204	1.583504	1	7
change_precipitation	353696	-5.23669	22.78458	-249.758	126.95
change_aridity	353696	-2.25335	7.756695	-96.4113	39.6302
change_meantemp	353696	0.516357	0.31204	-0.875	1.65
drought	283347	4.913791	2.484065	1	10
change_vegetation	353696	112.7564	178.2146	-749.2	1326.667
change_malaria	353696	-0.11303	0.12905	-0.56167	0.293968
luminosity	350264	3.730686	9.704129	0	118.9728
proximity_to_national_borders	350848	77985.4	85857.12	5.407586	571976.6
proximity_to_the_water	350848	109552.8	115293.1	0	798912.5
change_rainfall	353696	-47.0298	304.9399	-3191.27	1208.3
terrain	350054	1.492726	2.143685	0	22.80425

Table 1.18 DHS Summary Statistics

1.9.4 The proof of Equation 1.1

The amount of descendants of type p in generation G comprises to the summation of all descendants from the people of this type in the previous generation $G-1$.

$$g_G^p = \sum_{i=1}^{g_{G-1}^p} q_{i,G}^p = \delta \sum_{i=1}^{g_{G-1}^p} z_{i,G}^p + (1-\delta) \sum_{i=1}^{g_{G-1}^p} y_{i,G}^p = \delta(z_{A1,G} \sum_{i=1}^{g_{G-1}^p} I_{i,G}^p + z_{A2,G} \sum_{i=1}^{g_{G-1}^p} (1-I_{i,G}^p) + \tag{1.8}$$

$$+(1-\delta)(\sum_{i=1}^{g_{G-1}^p} I_{i,G}^p y_{A1,i,G} + \sum_{i=1}^{g_{G-1}^p} (1-I_{i,G}^p) y_{A2,i,G}) \tag{1.9}$$

Since the amount of descendants is infinitely increasing, applying the Law of Large Numbers produces the following:

$$g_G^p \overset{\rho}{=} g_{G-1}^p(\delta(pz_{A1,G} + (1-p)*z_{A2,G}) + (1-\delta)(p\mathbb{E}_y[y_A1] + (1-\delta)\mathbb{E}_y[y_A2] = \tag{1.10}$$

$$= g_{G-1}^p(\delta(pz_{A1,G} + (1-p)z_{A2,G}) + (1-\delta)(p\mathbb{E}_y[y^p]) \tag{1.11}$$

Without much loss of generality it can be assumed that $g_0^p = 1$. Using backward recursion, the next step is the following:

$$g_G^p \overset{\rho}{=} \prod_{G=1}^p (\delta(pz_{A1,G} + (1-p)z_{A2,G}) + ((1-\delta)(\mathbb{E}_y[y^p]) \tag{1.12}$$

Applying the Law of Large Numbers and logs:

$$\frac{1}{G}log(g_G^p) \overset{\rho}{=} \frac{1}{G} \sum_{G=1}^G log(\delta(pz_{A1,G} + (1-p)z_{A2,G}) + (1-\delta)\mathbb{E}_y[y^p]) \overset{\rho}{\longrightarrow} \tag{1.13}$$

$$\overset{\rho}{\longrightarrow} \mathbb{E}_z[log(\delta z^p + (1-\delta)\mathbb{E}_y[y^p])] \tag{1.14}$$

Equation 1.15 is the same as Equation 1.1 as was intended to be proven.

2

"Education vs Labour: Land Contracts and Gendered Parental Investments in Tajikistan"

Abstract

Female empowerment is crucial factor of economic development. The existing literature emphasized the positive effects of female labour participation and female education. However, these two results may be not equally possible in a context, where there is a trade-off between labor participation and education. This paper focuses on the context of Tajikistan and on quantifying the existing gender gap in intrahousehold spending as driven by gender-specific tasks in agriculture (e.g. cotton picking), exogenous price shocks, and types of land contracts. Since cotton harvesting is not mechanized, it is mainly done by women and children, which increases women employment but potentially affects girls' education. To test which effect prevails, I exploit household's data from the Tajikistan Living Standards Surveys, a triple difference identification strategy relying on exogenous spikes in cotton prices and spatial variation in suitability for cotton, and split sample analyses based on the exogenous allocation of state-owned versus small-size inheritable household owned land rights. Results show that girls experienced lower educational spending, which further dropped in cotton areas following the cotton price spike since they were more engaged into harvesting. Also, the findings demonstrate that the eldest girl being the mostly discriminated, and the level of bias being proportional to the number of siblings. Results are driven by households owning private land, where the costs of sending girls in the field are far less than costs of investing in their education. The household head is more interested in investing into the male descendant who will inherit the land and take care of parents, further deepening the gender gap in the long term. Overall, this paper provides the evidence that gender differences in parental investments in patriarchal societies can be reinforced by the type of agricultural activity, while positive economic shocks may further exacerbate this bias, additionally crowding out higher possibilities to invest in female education[1].

JEL classification: I0, D1, Q1
Keywords: cotton; education; gendered investment; gender bias

---

"Any society that fails to harness the energy and creativity of its women is at a huge disadvantage in the modern world"

Tian Wei

2.1 Introduction

Gender equality implies a wide range of economic benefits for the society. In particular, empowering girls from the start is proven to have long lasting and compounding benefits over the course of their lives. Policies targeting gender equality bring more stability and promote society's well-being and progress (Sen & Sengupta, 1983, Rose, 2000, Jayachandran & Kuziemko, 2011, Asfaw & Maggio, 2017). When girls have the opportunity to develop skills and knowledge that they need to make a change in their families and communities, it has far stretching benefits for the development of their country: from education and labour market to the conflict reduction. Yet, discrimination of women within the household exists in many contexts across developing world. It starts at a very young age, even before they are born, and it further limits their oppotunities in the future (Jayachandran, 2015, Evans et al, 2020).

Nowadays, there is much higher chance for a girl to be "out-of-school", to be involved into arranged or forced marriage, to face discrimination in her political rights and to be a victim of violence (Khanna et al, 2016). The roots of such gender discrimination lie in the acquisition of human capital. There has been a well-documented evidence of the gender bias in educational spending (Subramanian & Deaton, 1991, Aldermand & King, 1998, Lancaster et al, 2008, Azam & Kingdon, 2013). It was shown that various socioeconomic factors contribute to this phenomenon, such as social norms, costs of education, perceived returns to schooling and availability of schools. In particular, there is a cultural perception that sons are heirs to the property and family name, and the ones taking care of old parents. Moreover, according to Ebenstein (2013), patrilocality[2] is an important factor affecting sex ratios in developing countries. While parents put greater expectations in sons in general, in Tajikistan sons will remain living with parents and create an extended family. Despite inheritance laws being gender neutral, boys are the most probable heir to the parental assets, particularly, house and land. The main reason lies in cultural norms where girls after marriage belong to a new family, thus, leaving them property would mean giving it to another family. As a result, patrilineal approach to property inheritance exists in wide range of developing countries including Tajikistan (CEDAW, 2005).

---

[2]Residence of a couple with the husband's family.

51

Another factor that magnifies gender inequality in education is child labour. 152 million children worldwide are involved in child labour, which considerably reduces school attendance (UNESCO Institute for Statistics, 2021). In addition, parental decision on child work often relies on social norms and perception about the roles that girls should play in the household and community. Thus, daughters are much more likely to skip school in favour of taking care of the household and this trend is worsened by the widespread engagement into agricultural activities in developing countries (Akramov & Shreedhar, 2012).

Agriculture activities are usually seen as a potential way of female empowerment in the developing world. Since female labour force participation is increasing in the areas with the most population engaged into non-mechanized agriculture, which in turns lead to increase in women's income and potentially bargaining power. Women empowerment overall has a positive effect on female education, health, as well as human capital outcomes of the other family members including daughters (Qian, 2008, Duflo, 2012, Xue, 2019, Ashraf et al , 2020). However, both the advantages of female labor participation and the positive effects of female education may be not equally possible when there is a trade-off between labor participation and education. Yet, there is a limited understanding of the main drivers behind this trade-off. Hence, this paper aims to contribute to this question by evaluating to which extent gender-specific task in agriculture induces bias in intrahousehold educational spending.

This paper adds to the literature by focusing on the unexplored context of Tajikistan, where 67% of population is engaged in agriculture, and by quantifying the existing gender gap in intrahousehold spending as driven by gender-specific tasks in agriculture (e.g. cotton picking), exogenous price shocks, and types of land contracts. Tajikistan is an interesting case to study, since despite its 96% literacy rate, relatively high school enrollment rate, educational quality is rather low and actual attendance is not monitored. In the predominantly agricultural country with child labour being used in many sectors and activities, exploring the drivers of poor educational outcomes is highly relevant for the policy making (ILO, 2010). Moreover, the International Monetary Fund (2011) addressed insufficient enrollment of girls back in its 2010 report regarding the challenges of Tajikistan education system. According to the latest information from the UNESCO Institute for Statistics (UNICEF, 2020), 1.9 % of children aged 7-10 years old and, 5.6 % of children aged 11-15 years old were out of school, while in lower secondary education this indicator is equal to 9.1 % for girls, compared to 2.4 % of boys. Although secondary education in Tajikistan is free, the school enrollment rate plummeted drastically for students aged 15-16, for whom the enrollment is no longer mandatory. This is particularly true for girls. For instance, only 64 % of girls continue upper secondary education in contrast to the 86 % of boys.

To unravel the potential sources of this imbalance, this paper looks further into the case of cotton harvesting in Tajikistan. This context is particularly suited because cotton harvesting in

Tajikistan is not mechanized, harvesting is mainly done by women and children. Given that boys will take care of parents in the future and inherit parental property, latter prioritise the sons' education over the daughters'. The cotton season coincides with the beginning of school, which leads to the missed classes and falling back the program. By looking into the educational spending rather than attendance[3], this paper detects the gender bias and its relation to cotton picking. Despite the fact that cotton season increases options for women employment, it may have a negative impact on the girls' education.

In order to test which of these effects prevails, this study exploits household's data from the Tajikistan Living Standards Surveys 2007-2009-2011, a triple difference identification strategy relying on exogenous spikes in cotton prices (2010/2011) and spatial variation in suitability for cotton (from FAO), and split sample analyses based on the exogenous allocation of different land contracts (state-owned versus small-size inheritable household farms). The results show that educational spending in girls drop by 35% in cotton areas following cotton price spikes, and that the spending in private tutoring drop by 12% following the same price spikes. This effect is even strengthened by an increasing number of siblings. Results are driven by households owning private land, since girls bring more profit being in the fields rather than in school. Overall, this project brings causal evidence that gender differences in parental investments in patriarchal societies can be reinforced by the type of agricultural activity, while positive economic shocks may further exacerbate this bias, additionally crowding out higher possibilities to invest in female education.

The paper is organized as follows: Section 2.2 discusses the state of debate regarding the discrimination of girls within the household and identification strategies used. Section 2.3 introduces Tajikistan and its cotton sector in more details. Section 2.4 describes data and section 2.5 discusses empirical strategy. Results are presented in section 2.6. Section 2.7 presents some robustness checks, section 2.8 concludes and section 2.9 is bibliography used. Appendix is in section 2.10.

2.2 Literature Review

2.2.1 Main factors driving girls' discrimination

This section provides key reasons to explain existing gender bias in parental allocation of household resources and proceeds with the overview of the empirical strategies used in the literature. In particular, patrilocality is one of the crucial determinant of the sex ration in

---

[3]School attendance is not a reliable measure in the context of Tajikistan, where it could be manipulated to produce better statistics.

developing countries (Ebenstein, 2013). For instance, sex ratio has been improved in Korea once a pension expansion made parents less dependant on children after retirement. Similarly, despite high expectation put on boys, it was found that the one who is the eldest carries the most weight of responsibility for parents and extended family, thus, he is also the most favored one in intrahousehold allocation of resources (Das Gupta, 1987, Mullatti, 1995, Ebenstein, 2013). In some cultures the eldest son is also important for religious reasons and is responsible for performing the last rites of parents, which is another reason to be attributed to the special position that he enjoys (Arnold et al, 1998, Jayachandran & Pande, 2013).

Kinship norms and social expectations are the key ingredients of gender discrimination. In the country with the most skewed sex ratios - India - when parents get old, they almost solely depend on their sons, since daughters get married and move to their new family, which wellbeing they are expected to fully contribute to. As a result, daughters have much lower perceived returns to education than sons, since after marriage they will not play any role in the wellbeing of their natal families. This situation is common to all countries with prevailing patriarchal conservative social norms (Das Gupta, 1987, Foster & Rosenweig, 1999). Another strong support towards girl's discrimination comes from the dowry system, where girls become a liability. Parents have an additional incentive to increase son's weight in the marriage market by investing more in them (Das Gupta et al, 2005, Ashraf et al, 2020).

Moreover, inheritance practice is another foundation for women discrimination in many developing countries (Mullatti, 1995, Ebenstein, 2013, Bose & Das, 2017). Despite legally inheritance rights of women and men have been equalized in various countries, social norms keep dominating people's behaviour. Sons and not daughters are meant to inherit family estate and to live with parents as a joint family (Bose & Das, 2017). This is entwined in marriage customs discussed above.

In general, the analysis of the roots of child gender bias can be formulated as a result from the diversification in returns from girls and boys in the future. This difference may appear from the dissimilarity of men and women in labour returns or from patriarchal family structure that is common for the vast majority of Asian countries including Tajikistan (Rose, 2000, Akramov Shreedhar, 2012). Furthermore, since mentioned returns are actualized long period after initial investment made, parental intrahousehold decisions depend on the ability to finance investments into children on the early step. Therefore, credit market is important factor while analyzing rural households in developing countries. The financial status of family is crucial since the households, especially, poorest ones, are constrained in their intrahousehold spending (Rose, 2000). Therefore, the potential increase in the income may have a positive effect on the reduction of gender bias in the household spending.

2.2.2 Discrimination in educational spending

In this subsection, I discuss several studies on the discrimination in educational intrahousehold spending towards girls. For instance, Masterson (2012) analyzed gender bias in education spending in Paraguay using MECOVI 2001 data and applying Working and Leser specification of the Engel curve. He found pro-male education spending bias with some variations across regions and age categories. He also pointed out that urban households have greater educational expenditures only due to the fact of higher income and opportunities. Implying similar estimation techniques Zimmermann (2012) attempted to detect gender discrimination among Indian children in the intrahousehold allocation of goods. Specifically, he concentrated on the evaluation of possible gender bias in education expenditures. His findings illustrated discrimination against girls in all regions of India. In particular, girls face discrimination starting from 10 years old with nearly perfect disadvantage in the quantity of educational spending in the age group of 15-19 years old. Similarly, Rose (2000) discovered that women in rural regions of India allocate more time for bringing up boys rather than girls. She also examined the impact of households being constrained in credit market or not on the strength on child gender bias. Financial opportunities of families played important role when household head made his/her decision regarding the quantity and target of investments. Furthermore, Lancaster et al. (2003) found considerable gender bias for children within the age group of 10-16 years in rural provinces of Bihar and Maharashtra. To be more precise, they found significant bias in educational attainment in favour of boys in rural provinces, while urban regions did not demonstrate the same problem. The possible explanation may be the higher level of income among urban residents that allow them to investment in boys and girls equally.

Controversial results were obtained by Himaz (2010). She explored the extent to which intra-household allocation of educational spending is different among boys and girls in Sri Lanka by implying demand analysis. Her finding showed gender bias considerably preferring girls at the 5-9 and 17-19 age categories in 1990-1991 and at the 5-9 and 14-16 age categories for the period of 1995-1996. She argued that this bias was a product of rational household decision, since wage returns on education were significantly greater for females than for males in the observed time span. Moreover, she stated that Sri Lanka is the country where women have greater status along with labour opportunities in contrary to the majority of Asian counterparts. This is also a possible factor of favouring girls in the base steps of educational process. Evidence from Kingdon (2005) was based on the analysis of gender bias in intrahousehold expenditures across Indian states. She used both household and individual level data and highlighted the relevance and higher reliability of individual level estimation in this particular study. She obtained different effects in both level groups of the existence of child gender bias among diverse age categories. However, age group of 15-19 years old girls experienced the considerable discrimination against

boys in both observations. Thus, within mentioned age frame Indian girls faced gender bias in educational spending within the household. Another research within the same regions was done by Azam & Kingdon (2013). They found weak evidence of gender bias in school enrollment, but considerable pro-male educational expenditures. Moreover, they revealed the significant influence of regional discrepancies in nature and presence of gender bias. Results also suggest the major impact of age categories on the existence of child gender bias in education. To be more precise, the higher is the age group of the child, the greater is the evidence of gender bias. Furthermore, Emerson & Souza (2008) found interesting results concerning household spending patterns in Brazil. Their findings depicted that mothers favour daughters and fathers favour sons in child labour participation decisions, however, both parents favour sons in educational attainment. They also controlled for household income and found that parents' income have no impact on education of sons and daughters, however, mother's income reduces the probability for daughter to be exposed to child labour.

Overall, the described studies agree on the presence of pro-male bias in intrahousehold educational spending. Yet, it is not clear if mother's income or increased labour participation improves daughters' outcomes. This analysis contributes to the literature by using uncommon triple difference identification strategy, which will be discussed in details in Section 2.5. In addition, the majority of studies proposed that the additional source of income may reduce gender bias in expenditures (Oosterbeck et al, 2008, Skoufias & Parker, 2011, Saavedra & Garcia, 2012). However, this paper argues that positive income shock may not necessarily lead to bias reduction in certain circumstances. There is the heterogeneity in incentives such as income, size of household, religion and location that affect parental motivation in gender discrimination. The aim of this paper is to evaluate how some of these incentives such as income and increase in perceived returns in the labour market may further exacerbate the problem of educational bias in spending.

2.3   Tajikistan and cotton

Tajikistan is one of the five Central Asian countries without any access to the sea. Its population comprises 9.475 mln in 2021. According to the International Monetary Fund, Tajikistan is the 172nd country among 188 overall in terms of GDP per capita. Despite agricultural sector paying the less, almost 70% of people are engaged into agriculture (Van Atta, 2009). Moreover, Tajikistan is the largest source of migrants to Russia, around $ 2.5 billion or 33 % of GDP is coming from remittances, which leads to male labour being frequently missing and women being the one economically active. Especially in the rural areas, females are highly present in the labour force (Hegland, 2010, Kan, 2018).
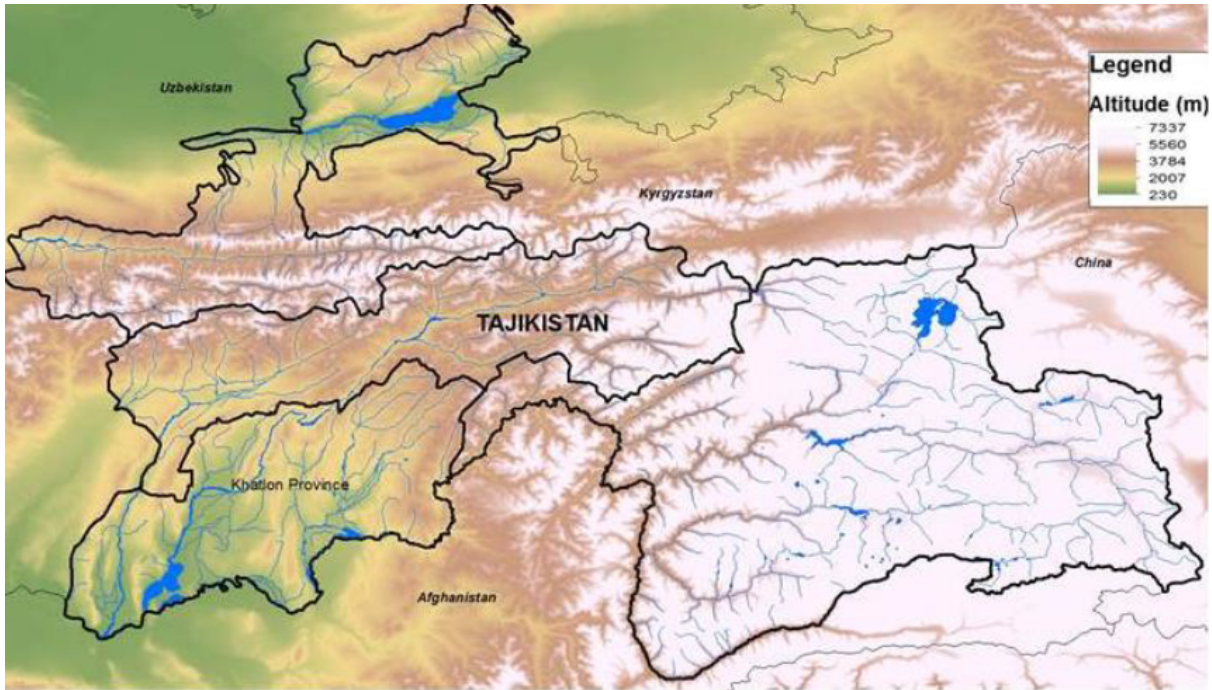
Figure 2.1 Geographical profile of Tajikistan: DIVA-GIS (2014)

The agricultural sector is one of the most important for Tajik economy despite 93% of Tajikistan's land being mountainous. Cotton, in particular, plays a crucial role in the lives of the majority of Tajik rural population. Its production is possible in certain regions due to the special climatic and geographic characteristics that makes the land suitable for cotton growing (Figure 2.2). Since the collapse of USSR, Tajikistan cotton sector faced considerable liberalisation and partial privatisation of cotton farms, which was based on lottery principle and through collective land tenure. More specifically, ginning sector was privatised along with financing and marketing of cotton, cotton farmlands were restructured and input prices were liberalized (Better Cotton Initiative, 2021).
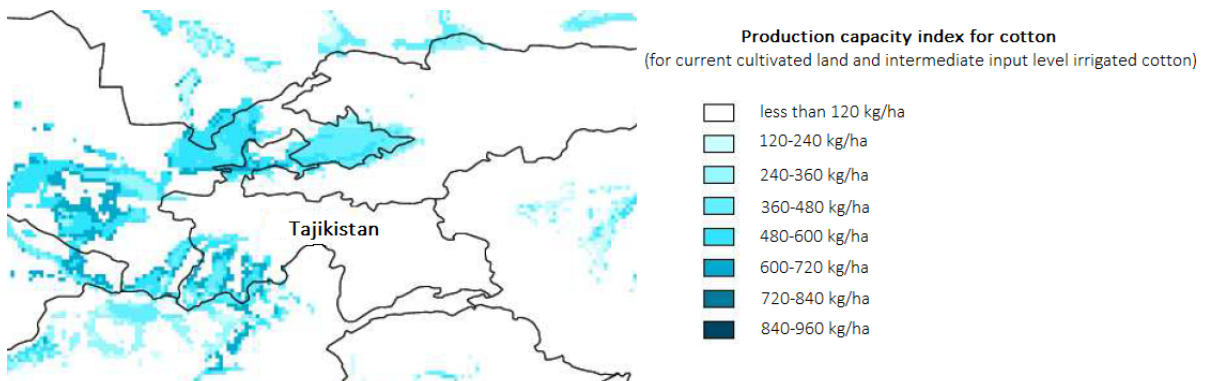
Figure 2.2 Regional variation of cotton production in Tajikistan: based on data from FAO-GAEZ

Cotton alongside aluminium comprises over 75% of Tajikistan exports. In the world market Tajikistan is a small exporter and its share comprises to only 2% of total world exports (Trading Economics, 2021). Therefore, cotton export price shock that happened in 2010/2011 season can be treated as exogenous. The price spiked as a result of severe drought in China, which was a world largest consumption and production country. This shock led China to double its cotton imports and causing world prices to increase more than twice (OECD-FAO Agricultural Outlook, 2020). According to Akramov & Shreedhar (2012), the surge in cotton prices in 2010/2011 promoted cotton production in Tajikistan to soar by nearly 40%, which changed the existing decreasing trend caused by lack of investment into irrigation and infrastructure. In 2011 the harvested cotton area increased by 30%.

In Soviet Tajikistan cotton production was in hands of big state-owned farms, so-called kolkhozes and sovkhozes. Seasonal workers were recruited regularly during harvest period. After the independence privatisation process has started for all state-owned farms. 77% of the cultivated land became private, while only 23% remained in the state ownership. The privatized land was distributed based on lottery to the small households and peasant (dekhkan, hereafter) farms. 22% went to the household plots (nearly 0.3 hectares) and 55% respectively to dekhkan farms. Dekhkan farms can be family owned or collectively owned, in other words, large private cooperative with more than 20 hectares of land (FAO, 2009). Collective dekhkan farms resemble state cooperatives of Soviet times, where individual farmers and workers are low paid on daily basis for specific activity or they can use a small share of the collective land for their needs. The farm manager makes all financial and production decisions. The group of shareholders has land right from 5 to 20 years which can not be inherited. On the other side, private individual or family dekhkan farms, smaller in size (2-20 hectares), have well specified inheritable rights for 99 years. Officially, all types of farms have free choice of the crop that they want to cultivate.

Student and child labour was widely used to pick up the cotton (ILRF, 2007). The same pattern remained now, since cotton picking is not mechanized. However, state farms are more rigorously monitored now in order to get rid of negative image associated with child labour that Tajikistan has been famous for (Umarov, 2011).

During the harvest season, additional workers are required for cotton picking, which is typically women and children. Some of the men is missing due to labour migration and also women are famous for their higher degree of agility. However, among all agricultural activities the lowest paid workers are in cotton industry (SOAS, 2010). Small family farms start hiring in the extended family and only afterwards, they go on the local market. This is one of the main reasons that children are involved into cotton picking process. At the same time this is a great working opportunity for local women. Women are mainly working within one community,

58

since they face travelling restrictions due to the conservative Islamic traditions in the country. Therefore, the labour mobility is not present in rural Tajikistan (Akramov & Shreedhar, 2012).

2.4 Data

This paper uses the data from Tajikistan Living Standards Survey (TLSS), which was conducted by the World Bank and UNICEF in 2007 and 2009 with a follow up in 2011, which was conducted by the Institute for East and Southeast European Studies. The information was collected during the cotton harvesting, which gives fair measure of labour market participation in cotton picking. TLSS 2007 is a representative sample of 4860 households from 270 primary sampling units (psu). TLSS 2009 and 2011 comprises 1503 households from 167 psu.

This analysis merges all three waves and uses the information on 1503 households from 167 psu to construct a panel dataset. It covers 16788 individuals with a series of socio-economic characteristics, including gender, age, occupation, ethnicity, education, income. The panel also includes 4217 children. Farms are separated into two types: private/family and state/kolkhoz (please refer to the Section 2.3). Combining the map of cotton suitability from FAO-GAEZ and geolocation of each psu from TLSS, I generate the Figure 2.3. Particularly, villages which are located in the areas with cotton suitability (blue areas) are treated as cotton villages, others are non-cotton correspondingly. Hence, in the Figure 2.3, green villages are classified as cotton, while red are non-cotton.
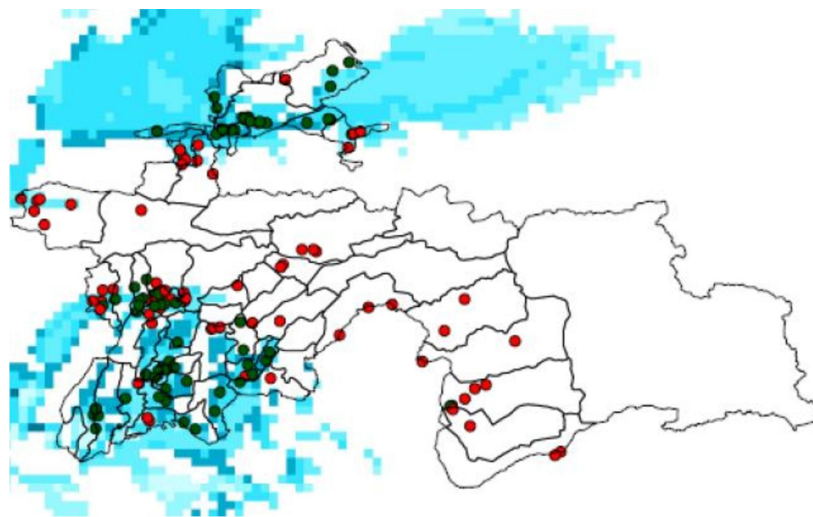
Figure 2.3 Suitability of cotton: data from FAO-GAEZ and TLSS

Notes: Blue areas - land, which agro-environmental factors make it suitable for cotton cultivation; red dots - villages in the areas without cotton suitability; green dots - villages in the areas which are suitable to grow cotton.

This exogenous variation in cotton suitability is one of the pillars of current identification strategy. Another pillar is exogenous variation in time. I will use the spike in cotton prices occurred between 2010 and 2011 taken from IndexMundi.com, which is a representative source on historical commodity price fluctuations (see Figure 2.4). Red lines depict the timing of the data used in this paper. It can be seen that the cotton price was the same in September 2007 and 2009, while September 2011 shows higher prices right after the spike.
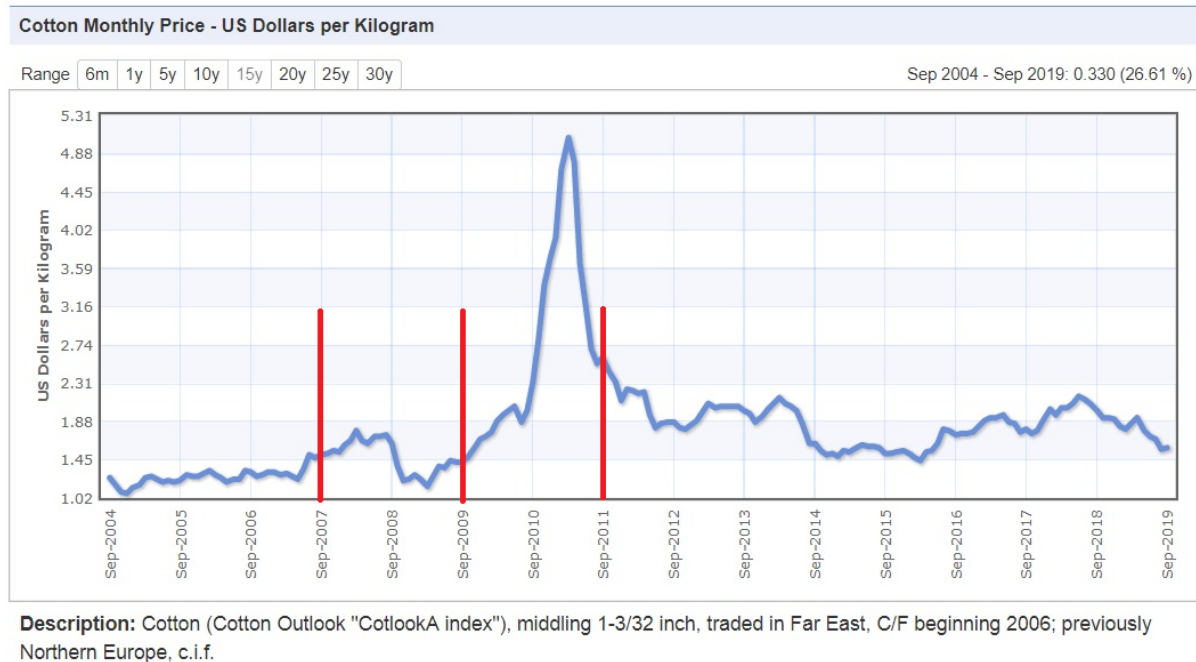
Figure 2.4 Cotton prices over time: IndexMundi.com

Tajikistan is a small exporter on the cotton world market, with a share of only 2%. The spike in prices was caused by a severe drought in China, which was an exogenous event for Tajikistan. I will exploit both variations (cotton suitability and price shock) to present the triple difference approach in the next section.

2.5   Empirical framework

To bring a causal evidence of differential treatment in educational spending, this project exploits an approach called difference-in-differences-in-differences ("triple differences")[4] to bring a causal evidence to the differential treatment of children in educational spending.

In order to check whether price shock affected girls and boys differently, I construct Figure 2.5. It can be seen that there is a decrease in educational spending for girls in 2011, while spending for boys kept rising. The observation in 2011 is right after the positive increase in prices for cotton. This graph signals that it is possible to evaluate the difference for girls and boys after the price shock. However, DiD[5] can be used here only if this price shock was not area specific.
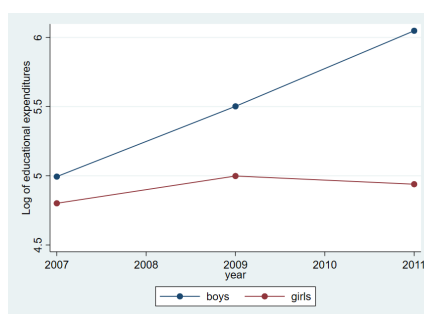
Figure 2.5 Gendered educational expenditures

Looking at the Figure 2.6 (a,b) it can be seen that the triple difference approach is the necessary option. This specification is needed, since the assumption that the rise in cotton prices was a common shock does not hold. It shows that the time shock was area specific making difference-in-difference insufficient in recovering the treatment effect. In particular, cotton areas were affected, but not non-cotton ones.[6] In addition, according to Gruber (1994), in case when there is a bias in two difference-in-difference estimators, taking the third difference will unbias the results since the bias will be differenced out.

In order to see the differences between cotton and non-cotton areas, I construct two graphs to test parallel trends assumption for gender differences after the prices shock. Figure 2.6 (a) shows that throughout the time the gap in educational expenditures was widening between boys and girls, however, in 2011 the trend for girls was reversed and the amount of expenditures dropped. I argue that this shift happened due to profitability of cotton farming and preference for girls to work in the field rather than study. Despite positive impact of cotton price spike on women labour force participation and associated income effect, it further exacerbate the existing bias. From Figure 2.6 (b) it is interesting to see that the difference between girls and boys is larger in

---

[4]details in Gruber (1994), Imbens & Wooldridge (2007), Cunningham (2021).

[5]Difference-in-difference

[6]I evaluate the period right after the shock occurred making it impossible for agriculturalist to adapt to changing prices and change the crop that they grow.
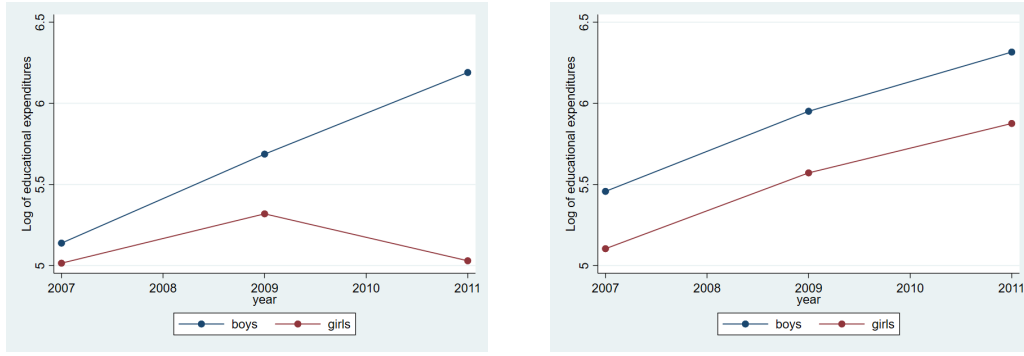
(a) In cotton areas   **(b)** In non-cotton areas

Figure 2.6 Gendered educational expenditures

non-cotton areas. While the gap in educational spending exists, there is no observed changes after the positive price shock of year 2011. Thus, the positive price shock was not a common one, but rather area specific. In this case, girls in non-cotton can be used as a control group, but the potential differences across area beyond the changes in price could be problematic and bias the results. Similarly, using difference-in-difference approach only within cotton areas could be problematic due to the factors affecting educational spending unrelated to the price shock. Despite the fact that there were not any reforms or changes in Tajikistan that could affect educational spending within households during the observed period, using triple difference helps to control aforementioned potentially confounding trends: (i) changes in educational spending of girls across areas and (ii) changes in educational spending for everyone in cotton areas[7]. The main specification is as following:

$$Ln(EducExp)_{i,h,t} = \alpha + \beta_1(\text{girl x post-year x cotton})_{i,h,t} + \beta_2(\text{girl x cotton})_{i,h} +$$
$$+ \beta_3(\text{post-year x cotton})_{h,t} + \beta_4(\text{girl x post-year})_{i,t} + \beta_5\text{girl}_i +$$
$$+ \beta_6\text{post-year}_t + \beta_7\text{cotton}_h + \gamma X_{i,t} + \sum_t \sigma_t Time_i + \sum_h \sigma_h Household_i + \varepsilon_{h,t}$$

where $Ln(EducExp)_{i,h,t}$ is a natural logarithm of intrahousehold expenditures on education of a child i in the household h over years t. $Girl_i$ is a dummy variable taking 1 if a child is girl below 18[8] and 0 otherwise. $Cotton_h$ is a dummy variable taking 1 if household is located in the village which is in the area that is suitable for cotton farming (refer to the Figure 2.3) and 0 otherwise. $Post-year_t$ is a dummy variable where 1 is year 2011 after the price shock and 0 is year 2007 and 2009 before the price shock.

The treatment effect is captured by β_1, which reports the effect of the cotton price shock on

---

[7]DDD starts with the time changes in averages for girls in cotton areas and nets out the change in means for girls in the control area and the change in means for boys in cotton areas.

[8]18 is the age of attaining legal adulthood. A legal adult is a person who has attained the age of majority and is therefore regarded as independent, self-sufficient, and responsible.

girl child in cotton villages compared to girls in non-cotton villages. In addition, I include individual characteristics $X_{i,t}$ such as gender, age, occupation, ethnicity, income, household size, hukumat, marital status, relation to the household's, head, living in urban area, education of parents, education of household's head, number of siblings, being first born, being the youngest son. Also I include household time fixed effects to control for unobserved time invariant heterogeneity. Standard errors are clustered at the household level[9]. I use cluster-adjusted standard errors in addition to the fixed effects to account for intra-cluster correlation which the fixed effects approach is not taking into account (Angrist & Pischke, 2009). Following the main specification, I split the sample between small family-owned farms and big state owned farms, while maintaining the triple difference approach and same set of controls. This is done in order to analyze the role of farm structure in gender bias in educational spending in cotton areas.

2.6   Results

In this section I present the main findings of this chapter. Table 2.1 presents the results using the main triple difference specification described in the previous section. In particular, it can be seen that girls are discriminated 13% more on average after controlling for time invariant characteristics. The price shock itself had a sizeable positive effect on intrahousehold educational spending in general, and especially for girls. However, the situation is reverse for cotton areas. There, expenditures on education are 18% lower on average. Moreover, after the price shock a girl in cotton areas receives 35% less spending on education than a boy of the same age. It can be seen that price shock had a negative impact on educational spending for girls in cotton areas. This is despite the established that positive income shock brings women empowerment and has positive effect on girls' human capital formation. In the next steps, I evaluate what are the potential mechanisms and factors that play a role in the observed discrimination.

---

[9]Using standard errors which are robust and clustered at district level and psu level do not change the results. Since the sample was selected by randomly sampling households from within the hukumats; and I am focusing on the educational spending in the overall population, I choose to cluster standard errors by household, since there are households in the population of interest beyond those seen in the sample.

VARIABLES	(1)	(2)	(3)
	Log of educational spending		
girl	-0.232***	-0.114***	-0.132***
	(0.0483)	(0.0385)	(0.0398)
post_year	0.796***	0.587***	0.940***
	(0.0595)	(0.0641)	(0.0876)
cotton	-0.168*	-0.185**	-0.185**
	(0.0920)	(0.0893)	(0.0889)
girl#post_year	0.0851	0.136*	0.142*
	(0.0843)	(0.0764)	(0.0787)
girl#cotton	0.202*	0.222**	0.284***
	(0.114)	(0.0921)	(0.0966)
post_year#cotton	-0.00335	0.196	0.187
	(0.125)	(0.124)	(0.135)
girl#post_year#cotton	-0.189	-0.347**	-0.354**
	(0.165)	(0.157)	(0.168)
Constant	5.423***	5.845***	6.088***
	(0.0410)	(0.211)	(0.267)
Controls	No	Yes	Yes
FE year and household	No	No	Yes
Observations	3,338	3,246	3,246
Number of households	996	996	996

Clustered standard errors in parentheses

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

Table 2.1 DDD Baseline Results for girls

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

VARIABLES	(1) Log of expenses: fees	(2) uniforms	(3) books	(4) meals	(5) repair	(6) other	(7) private tutoring
girl	-0.403***	-0.123***	-0.0868**	-0.439**	-0.0123	-0.281**	-0.00403
	(0.125)	(0.0312)	(0.0370)	(0.180)	(0.0327)	(0.134)	(0.0100)
post_year	0.677**	0.906***	0.842***	-0.0273	0.501***	0.238	-0.0547***
	(0.311)	(0.0629)	(0.0743)	(0.496)	(0.0835)	(0.384)	(0.0158)
girl#post_year	0.306	0.0865	0.0922	0.338	0.0633	0.275	0.0530**
	(0.250)	(0.0663)	(0.0790)	(0.299)	(0.0800)	(0.194)	(0.0222)
girl#cotton	-0.184	-0.0225	-0.114	-0.107	-0.0798	-0.352	0.0385
	(0.454)	(0.0709)	(0.0862)	(0.275)	(0.0865)	(0.589)	(0.0242)
post_year#cotton	1.600**	0.510***	0.0525	-0.537	0.344**	0.421**	0.0972***
	(0.799)	(0.151)	(0.155)	(0.856)	(0.150)	(0.149)	(0.0292)
girl#post_year#cotton	-3.268***	0.0301	0.167	-0.853	-0.0549	-0.199	-0.115***
	(0.817)	(0.168)	(0.171)	(1.504)	(0.148)	(0.629)	(0.0394)
Constant	3.725***	4.595***	2.884***	4.118***	2.245***	3.086***	0.966***
	(0.140)	(0.0287)	(0.0368)	(0.257)	(0.0331)	(0.221)	(0.0522)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year and household	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,081	2,805	2,663	689	2,300	562	3,489
Number of household	531	923	902	407	849	306	1,017

Clustered standard errors in parentheses

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

Table 2.2 DDD results for educational expenses in details

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

First of all, in Table 2.2 I use specific categories in order to understand what is the main source of spending of households. Overall, girls are discriminated in almost all categories including paying for school uniforms and meals. It can be seen that girls receive 40% less money to cover school fees and strikingly 43% less on school meals. Also they have 12% less spending on uniforms and 8% less on purchase of books. Regarding the discrimination in cotton areas, it can be seen that after the price shock, there was a significant decrease in school fees and private tutoring according to the columns 1 and 7 of Table 2.2. Both indicators support the hypothesis of lesser investment into girls in cotton areas after the spike in cotton prices. In particular, parents being unwilling to spend more money on private tutoring after controlling for income and living in an urban area signals that they could possibly be not interested in girls' educational quality, since it does not pay off in the future. In cotton areas, parents tend to spend 11% less on private tutoring. One of potential explanations is that they they could prefer daughters to contribute to household tasks rather than investing into their education.

| | (1) | (2) | (3) |
VARIABLES	Number of siblings more than 1	Number of brothers more than 1	Number of sisters more than 1
girl	-0.111**	-0.0969	-0.0278
	(0.0508)	(0.160)	(0.128)
post_year	1.113***	1.457**	1.328***
	(0.131)	(0.664)	(0.420)
girl#post_year	0.0192	0.575	0.523
	(0.142)	(0.512)	(0.524)
girl#cotton	0.334***	0.287	0.119
	(0.119)	(0.184)	(0.205)
post_year#cotton	0.327*	0.329	0.288
	(0.195)	(0.479)	(0.338)
girl#post_year#cotton	-0.497**	-0.677	-0.608
	(0.229)	(0.525)	(0.583)
Constant	4.982***	4.694***	4.754***
	(0.0593)	(0.334)	(0.206)
Observations	1,492	367	571
Number of households	426	136	245

Clustered standard errors in parentheses

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

Table 2.3 DDD results for educational expenses for different composition of siblings

Note: Standard errors are clustered at household level. Individual controls include being firstborn, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

Another potential effect could result from fertility pattern within the household. Girls are typically care givers, therefore, the more siblings a girl has, the more work inside the household she does. Parents may spend less on daughters' education since the time which they spent taking care of the siblings is a better pay off for the family. I evaluate this hypothesis in Table 2.3, which presents the results for different composition of siblings. Column 1 shows that having siblings makes the effect of discrimination stronger. In particular, parents spent 14% less on education of girls in cotton areas when cotton prices increased. There could be a gendered differentiation in sibling's composition, since parents may want a better care for their sons or other daughters may need less care. However, columns 2 and 3 of Table 2.3 show that there is no difference in gender composition of the siblings. It implies that the gender of siblings does not play an important role, but rather the quantity of other children. This can stem from the fact that the more children are there in the family, the harder it is to allocate intrahousehold spending equally to everyone. Moreover, it could be driven by the fact that older girls can take care of younger.

In Table 2.4 I check whether the birth order affects the parental decision to spend money on education for girls. Interestingly, the effect is not there across all dimensions except cotton areas after the price shock. The result is great in magnitude and implies that being the eldest daughter

is associated with taking care of the household, working more and marrying earlier. In particular, early marriage and presence of siblings make the eldest daughter the least probable candidate of parental investment in conservatives societies. Leaving the parental household after marriage and spending time to take care of siblings further exacerbate the bias in educational spending. Hence, parents do not have incentives to invest more into the education of the eldest daughter after positive income shock.

VARIABLES	(1) First born
girl	0.242
	(0.308)
post_year	0.601
	(0.486)
girl#post_year	0.712
	(0.603)
girl#cotton	-0.0446
	(0.386)
post_year#cotton	0.630
	(0.529)
girl#post_year#cotton	-1.464**
	(0.682)
Constant	5.066***
	(0.176)
Observations	753
Number of households	705
FE year and household	Yes

Clustered standard errors in parentheses

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

Table 2.4 DDD results for educational expenses for the sample of first born

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

Furthermore, it is important to further investigate the age effect, in order to check if the effect is mainly driven by older girls who are more capable of offering quality help in the household and in the field. In Table 2.5 I separate the effect between girls below and older than 12 years old. I chose the age of 12, because (i) girls in this age cohort are old enough to perform work in the field and (ii) this is an unofficial age in Tajikistan when children get engaged into cotton

67

harvesting (EJF, 2007). Table 2.5 shows that there is no effect among younger age cohorts, yet the effect exists for older girls.This results indicates that price spike affected the distribution of intra-family expenditures differently according to the age of the children. It is also interesting to note that there is general positive impact on girls below 12 in the last year, which is not driven by cotton areas.

VARIABLES	(1) Below 12	(2) After 12
girl	-0.0913	-0.107**
	(0.176)	(0.0457)
post_year	1.936***	1.011***
	(0.407)	(0.104)
girl#post_year	-0.496*	-0.0182
	(0.278)	(0.0869)
girl#cotton	0.930***	0.152
	(0.318)	(0.105)
post_year#cotton	0.0501	0.216
	(0.427)	(0.158)
girl#post_year#cotton	-0.384	-0.304*
	(0.591)	(0.173)
Constant	6.165***	5.487***
	(1.020)	(0.338)
Controls	Yes	Yes
FE year and household	Yes	Yes
Observations	466	2,410
Number of households	354	873

Clustered standard errors in parentheses

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

Table 2.5 DDD results for educational expenses for different age groups

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

Another crucial factor affecting intrahousehold allocation of resources is the parental background. In Table 2.6 I compare to which extent the education of mother and father affects educational expenditures. I account for better educated parents, since it has been shown that higher parental education is associated with better human capital outcomes for their children (Brown, 2006, Ashraf et al, 2020). However, results show that higher education still leads to discrimination and that the maternal education has around 2% smaller impact than paternal education on resources

allocation. This is a surprising finding, since maternal education commonly signals of better decision making and higher involvement into children's formation. One possible explanation could be found in Tajik culture, where traditions and social opinion are very important and mothers support the role of daughters as future wives to preserve familial honour (Qodir, 2012, Mahdavian, 2018). It is important to highlight that the discrimination in education against girls is significant without accounting for cotton price increase. Girls tend to receive 13-14% less spending than boys if parental education is high.

VARIABLES	(1) Education of mother is higher or more	(2) Education of father is higher or more
girl	-0.130***	-0.148***
	(0.0418)	(0.0452)
post_year	0.814***	0.815***
	(0.0790)	(0.0911)
girl#post_year	0.139*	0.146
	(0.0817)	(0.0907)
girl#cotton	0.308***	0.267**
	(0.108)	(0.117)
post_year#cotton	0.226	0.271
	(0.143)	(0.183)
girl#post_year#cotton	-0.334*	-0.353*
	(0.176)	(0.209)
Constant	5.777***	5.807***
	(0.251)	(0.279)
FE year and household	Yes	Yes
Observations	3,022	2,569
Number of households	953	843

Clustered standard errors in parentheses

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

Table 2.6 DDD results for educational expenses for education of parents

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

The aforementioned results demonstrate that the discrimination of girls in educational spending is widespread in Tajik society and does not depend on parental background. It is more severe

in the families with more children and for older girls. The spike in cotton prices in 2011 had a positive effect on educational spending, at the country-level, however, the situation was paradoxically reversed in cotton areas. It was also shown that being a girl in cotton areas means to be most discriminated. In particular, living in cotton areas is associated with lower educational spending in general.

Observing that the negative effect on girls' education comes from cotton areas, this analysis proposes a peculiarity of the culture of cotton to be the main mechanism: cotton is easy to pick up and women, including young girls, are the main workers during the harvest season. If households prefer to send daughters to work in the field rather than spending on her schooling, the unequal spending of education should only be observed in private farms. State farms focus on meeting government's goals in the volume of cotton that should be picked. They hire people to pay them on daily basis and they do not have incentives to discriminate based on gender.

I suppose that since Tajik society is very conservative, a young girl would not be let to work with unknown people in the field, most of whom are hired seasonally and may come from other villages. In addition, state farms are monitored by the government and NGOs, therefore, they would avoid hiring children being aware of the negative image that child labour could bring to the country. In Table 2.7 I divide the sample between state owned farms and privately owned (refer to Section 2.3 for more details). It can be seen that the effect holds only among privately owned farms. Moreover, girls are being discriminated in general, but also after the spike in cotton prices if they live in the household who owns the land. On contrary, girls living in the households who participate in state farms do not have observed discrimination. It could stem from the fact that state farms include different organization of farms, cooperatives, collectives, commercials. Unfortunately, the present data do not allow me to differentiate further. Yet the results from Table 2.7 signals that girls are more discriminated if they have a chance to be engaged in cotton picking within their household.

VARIABLES	(1) Private farm	(2) State farm
girl_child	-0.0785**	0.0869
	(0.0353)	(0.0860)
post_year	1.701***	1.717***
	(0.105)	(0.106)
girl_child*post_year	-0.969***	0.968
	(0.0668)	(0.703)
private_cotton	1.351*	
	(0.734)	
girl_child*private_cotton	-1.979**	
	(0.801)	
post_year*private_cotton	-1.159	
	(0.908)	
girl_child*post_year*private_cotton	-1.857*	
	(0.962)	
kolkhoz_cotton		-1.497
		(1.145)
girl_child*kolkhoz_cotton		1.075
		(1.162)
post_year*kolkhoz_cotton		1.328
		(1.096)
girl_child*post_year*kolkhoz_cotton		0.982
		(0.987)
Constant	5.190***	5.192***
	(0.229)	(0.235)
Observations	3,337	3,337
FE year and household	Yes	Yes
Number of households	1,003	1,003

Clustered standard errors in parentheses

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

Table 2.7 Triple difference results

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

In order to further support this hypothesis, I used the question from TLSS regarding child labour. In particular, whether in this population point children are hired or not. Table 2.8 shows that in cotton villages where there is a record of hiring children, the results hold and the discrimination in education is around 45% for girls. While in villages where there is no indication of children working, the effect vanishes. This result supports our initial hypothesis that girls get mostly discriminated in the cotton areas, since it is considered a better investment to send them to work in the field rather than at school.

VARIABLES	(1) If in this population point children earn nothing/are not hired	(2) If in this population point children earns money if hired
girl	-0.0137	-0.0427
	(0.0883)	(0.0419)
post_year	1.032***	0.993***
	(0.192)	(0.0853)
girl#post_year	-0.243	0.0237
	(0.246)	(0.0915)
girl#cotton	-0.0210	0.344***
	(0.239)	(0.0894)
post_year#cotton	0.0571	0.306*
	(0.304)	(0.166)
girl#post_year#cotton	0.0207	-0.446**
	(0.325)	(0.178)
Constant	5.025***	5.053***
	(0.115)	(0.0372)
FE year and household	Yes	Yes
Number of households	100	788

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

Table 2.8 DDD results for working children

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

2.7   Robustness checks

In this section I will discuss some concerns that arise in regards to this analysis. First, girls typically stay in their families before they get married. Also, they do not have much of decision making power before they reach 21. Given the results in Table 2.5, I would like to check if the bias could be driven by older girls. For this purpose, I do not focus solely on children, but I run the same analysis for all girls below and including 21 years old. Table 2.9 shows that the bias is 5% higher for this age cohort. It is mainly coming from elder girls who have more capability to work in the field and be productive. It is also important to notice that general discrimination is around 10%, which is lower than 13% among girls-children. It could be explained by age in two possible ways. First, older girls are more empowered to disagree with the household head.

Second, they also most probably finished school and did not continue education, so there could be no educational expenditures associated with them.

VARIABLES	(1) Log of educational spending	(2) —	(3) Log of educational spending
girl	-0.0846*	-0.0654*	-0.0958**
	(0.0465)	(0.0396)	(0.0405)
post_year	0.870***	0.668***	1.029***
	(0.0653)	(0.0705)	(0.0974)
cotton	-0.126	-0.183**	-0.183**
	(0.0983)	(0.0926)	(0.0926)
girl#post_year	-0.0509	-0.0591	-0.0272
	(0.0800)	(0.0762)	(0.0791)
girl#cotton	0.0998	0.164*	0.234**
	(0.112)	(0.0901)	(0.0961)
post_year#cotton	-0.00831	0.253*	0.244
	(0.148)	(0.144)	(0.151)
girl#post_year#cotton	-0.0857	-0.346**	-0.408**
	(0.164)	(0.154)	(0.165)
Constant	5.328***	5.205***	5.403***
	(0.0395)	(0.225)	(0.300)
Controls	No	Yes	Yes
FE year and household	No	No	Yes
Observations	2,968	2,876	2,876
Number of households	930	930	930

Clustered standard errors in parentheses

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

Table 2.9 DDD Baseline Results for girls till 21 years old

Note: Standard errors are clustered at household level. Individual controls include being firstborn, eldest son, youngest son, number of brothers and sisters, ethnicity/religion, age, hhsize, marital status of parents, relation to head of household, living in urban area, income, number of children in the household

Another potential concern is that families, which have participated in the study, are coming from the better developed regions of the country, thus, they may be richer. In this case, controlling for income may not be sufficient to account for selection bias. Therefore, I select only those households which have higher income level than the average country level and apply the same specification as in the main part of the analysis. The results are shown in Table 2.10, where

no change is detected in the coefficient describing girls' discrimination in cotton areas after the positive price shock (it remains around 37 %). Girls are more discriminated (around 17%) on average, but after the price shock there was an increase in educational spending for girls, however, this effect is reversed in cotton areas.

VARIABLES	(1) Households with higher than average monthly income
girl_child	-0.174***
	(0.0415)
girl_child#cotton	-0.278***
	(0.0978)
post_year	0.942***
	(0.0681)
girl_child#post_year	0.158*
	(0.0915)
post_year#cotton	0.207
	(0.134)
girl_child#post_year#cotton	-0.377**
	(0.165)
Constant	5.161***
	(0.0349)
Observations	3,322
FE year and household	Yes
Number of hhid	1,000

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 2.10 Additional checks: income

Informal insurance mechanisms developed between households could nullify the impact of the price shock or could be driving the results. I subset the sample into households who are engaged into the system of mutual assistance (data also taken from TLSS). This is a form of insurance mechanism for agricultural households during adverse shocks. It may take both financial and in-kind forms. Table 2.11 shows that girls tend to be even more discriminated in these households (nearly 48%). One of the potential explanation could be stronger ties and collaboration between the households that are engaged into mutual assistance, which makes possible network for marriage match making. Thus, girls may face even stronger discrimination in access to education, since this is not an important factor for future successful marriage, while taking care of the household and being productive in the field will be considered as a good signal for potential in-laws (CAA Network, 2017). Overall, higher coefficient in the variable of interest signals that not accounting for insurance mechanisms is most likely resulting in upward bias, which does not interfere with the general conclusion of this study.

VARIABLES	(1) Households in the system of mutual assistance
girl_child	-0.512***
	(0.162)
girl_child#cotton	0.596***
	(0.189)
post_year	1.202***
	(0.198)
girl_child#post_year	0.301
	(0.207)
post_year#cotton	-0.0363
	(0.235)
girl_child#post_year#cotton	-0.483*
	(0.257)
Constant	5.009***
	(0.0918)
Observations	685
FE year and household	Yes
Number of hhid	187

Clustered standard errors in parentheses

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

Table 2.11 Additional checks: the system of mutual assistance

Moreover, in order to check that the composition of siblings in further details, I re-evaluate the estimated effect only in the households with only two children of both gender. In this setting, I expect to get strong discrimination against daughter in line with the mechanism proposed above. Table 2.12 totally supports this claim as discrimination is raising dramatically to 150%. This result demonstrates that in the case of pure trade-off between son and daughter, parents tend to strongly discriminate girl in terms of educational spending in cotton areas even after the positive price shock. Since the effect on girls regardless of household activity seems to be absent, it supports the claim that cotton areas are mostly driving the results. Hence, it seems that the cultivation of cotton exacerbate an existing intrahousehold gender bias in the traditional Tajik society.

VARIABLES	(1) Households with two children: girl and boy
girl_child	0.00617
	(0.117)
girl_child#cotton	0.350
	(0.338)
post_year	0.302
	(0.236)
girl_child#post_year	0.266
	(0.457)
post_year#cotton	0.423
	(0.513)
girl_child#post_year#cotton	-1.573**
	(0.783)
Constant	5.342***
	(1.023)
Observations	211
FE year and household	Yes
Number of hhid	75

Clustered standard errors in parentheses

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

Table 2.12 Additional checks: two children of both gender

The crucial factor that could play a role in the girls' discrimination is decision making within the household. However, TLSS does not provide an opportunity to construct an adequate measure for this purpose. The only relevant question was inquiring about who decides to spend and save in the household. After selecting woman's answers, the dataset size shrank to 161 individuals only, with incomplete answers. This sample did not allow to control for the effect before and after the price shock. Instead I checked whether women decision making makes a change for girls in general and in cotton areas particularly. Table 2.13 shows that results are not significant, but the direction of the effect is positive for girls on average and in cotton areas in particular. However, limited sample does not allow me to draw any conclusions and further research is required.

VARIABLES	(1) Households where decisions to spend and save are made by women
girl_child	0.176
	(0.343)
cotton	0.409
	(0.324)
girl_child#cotton	0.0878
	(0.449)
Constant	7.707***
	(1.102)
Controls	Yes
FE household	Yes
Observations	161

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 2.13 Additional checks

2.8 Conclusion

Human capital investment is a cornerstone of economic development and poverty reduction (Becker, 1994). Nevertheless, one of the most important human capital dimension - education - is hardly accessible due to the costs it entails for families. Constraints exist not only in direct form (e.g. fees, materials, transportation, uniforms, textbooks), but also in indirect form in terms of income that is lost with children who could instead be working. Also, parents of girls face an additional trade off between providing them with an adequate schooling and saving these expenses, since girls will get married and leave the parental household. Moreover, credit-constrained families are not able to smooth their consumption over time and the costs of children's education go even higher (Rose, 2000, Kingdon, 2005, Maccini & Yang 2009, Barcellos et al, 2014). In this case, given the prevalence of son preference across developing world, girls are even further limited in their access to education (Jayachandran, 2015, Evans et al, 2020).

This chapter explored the case of Tajikistan in order to causally detect the existence of gender bias in cotton areas of the country. Despite the fact that cotton harvesting increases options for women employment, it has negative implications on girls' education. By looking into educational spending rather than school attendance, this study shows that positive income shock along with engagement into agricultural activity is further deepening the negative effect on girls' education. For this, it uses a triple difference identification strategy relying on exogenous spikes

in cotton prices, spatial variation in suitability for cotton, and split sample analyses based on the exogenous allocation of household land contracts.

Results show that girls experience lower educational spending, which further dropped in cotton areas following the cotton price spikes. The bias is further characterized by showing that the eldest girl is the most discriminated, that the bias strength is proportional to the number of siblings, and that the education of the mother has no effect on it. Results are driven by small households holding their land privately, since they can rely on family members as a labour force. Hence, girls bring more profit to the household by working in the fields than by investing into their education. Furthermore, the household head is more interested in investing into sons who will inherit the land and take care of parents, further widening the gender gap in the long term. This study brings evidence that positive income shock may be a driver of gender discrimination if it is associated with increased labour force participation in the childhood. This is the case in patriarchal societies in developing countries where discrimination of girls is rooted in cultural norms and reinforced by the type of agricultural activity.

2.9 Bibliography

Akramov, K.T., Shreedhar, G., (2012). Economic Development, External Shocks, and Food Security in Tajikistan. IFPRI Discussion Paper 01163.

Alderman, H., King, E., (1998). Gender differences in parental investment in education, Structural Change and Economic Dynamics, 9, issue 4, p. 453-468

Angrist, J., Pischke, J., (2009). Mostly harmless econometrics. ISBN: 9780691120355. Princeton University Press.

Arnold, F., Kim, C., Roy, T.,K., (1998). Son preference, the family-building process and child mortality in India. Population Studies. 52(3):301–315

Asfaw, S., and Maggio, G., (2017). Gender, weather shocks and welfare: evidence from Malawi. The Journal of Development Studies 54(2), 271–291

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2020) Bride price and female education. Journal of Political Economy 128(2), 591–641

Azam, M., Kingdon, G., (2013). Are girls the fairer sex in India? Revisiting Intrahousehold allocation of education expenditure. World Development, Elsevier, vol. 42(C), pp. 143-164

Barcellos, S. H., Carvalho, L. S., Lleras-Muney, A. (2014). Child Gender and Parental Investments In India: Are Boys And Girls Treated Differently?. American economic journal. Applied economics, 6(1), 157–189

Becker, G., (1994). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, Third Edition. THE UNIVERSITY OF CHICAGO PRESS

Better Cotton Initiative, (2021). Tajikistan. [Online] Available from: https://bettercotton.org/where-is-better-cotton-grown/tajikistan/

Bose, N., Das, S. (2017). Women's Inheritance Rights, Household Allocation, and Gender Bias. The American Economic Review, 107(5), 150–153.

Brown, P. H. (2006). Parental Education and Investment in Children's Human Capital in Rural China. Economic Development and Cultural Change, 54(4), 759–789. https://doi.org/10.1086/503582

CAA Network, (2017). A narrative about care: how to be a woman in Tajikistan. [Online] Available from: https://www.caa-network.org/archives/11021

Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), (2005). Consideration of reports submitted by States Parties under Article 18 of the Convention on the Elimination of All Forms of Discrimination against Women. Combined initial, second and third periodic reports of States Parties. Tajikistan.[Online] Available from: http://www.un.org/womenwatch/daw/cedaw/reports.htm

Cunningham, S., (2021). Casual Inference: Mixtape. Yale University Press. ISBN: 9780300251685.[Online] Available from: https://mixtape.scunning.com/

Das Gupta, M., (1987). Selective Discrimination against Female Children in Rural Punjab, India. Population and Development Review 13(1)

Das Gupta, M., (2005). Explaining Asia's 'missing women': A new look at the data. Population and Development Review. 31(3):529–535.

Deaton, A. (1989). Looking for boy–girl discrimination in household expenditure data, The World Bank Economic Review,3(1), pp.1–15.

Duflo, E., (2000). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. NBER Working Papers 7860, National Bureau of Economic Research, Inc.

Duflo, E. (2012). Women empowerment and economic development, Journal of Economic Literature, 50(4), 1051–1079

Duflo, E., Dupas, P.,Kremer, M., (2012). Education, HIV, and Early Fertility: Experimental Evidence from Kenya. Stanford University Press

Ebenstein, A. (2013). Patrilocality and missing women. Working Paper, Jerusalem: Hebrew University

EJF, (2007). The Children behind Our Cotton. Environmental Justice Foundation, London, UK. ISBN No. 1-904523-13-7 [Online] Available from: https://ejfoundation.org/resources/downloads/The-Children-behind-Our-Cotton-FINAL_small.pdf

Emerson, P., Souza, A., (2008). Birth Order, Child Labor, and School Attendance in Brazil.World Development Vol. 36, No. 9, pp. 1647–1664

Eurasianet, (2020). Tajikistan: Coronavirus taking heavy toll on remittances [Online] Available from: https://eurasianet.org/tajikistan-coronavirus-taking-heavy-toll-on-remittances#:~:text=Tajikistan's%20economy%20is%20perilously%20reliant,gross%20domestic%20product%2C%20or%20GDP

Evans, M., Lindauer, M., and Farrell, M. (2020). A pandemic within a pandemic—intimate partner violence during covid-19. New Engl. J. Med. 383, 2302–2304

Food and Agricultural Organization (FAO) (2009). FAO crop and food security assessment mission to Tajikistan. Special report 2009. [Online] Available from: http://www.fao.org/docrep/012/ak339e/ak339e00.htm

Foster, A. D., Rosenzweig, M. R. (2004). Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000. Economic Development and Cultural Change, 52(3), 509–542.

Gibson, J., McKenzie, D., (2010). The development impact of a best practice seasonal worker policy.WorldBankPolicy Research working paper ; no. WPS 5488

Gibson, J., Rozelle, S., (2004). Is it Better to be a Boy? A Disaggregated Outlay Equivalent Analysis of Gender Bias in Papua New Guinea. Journal of Development Studies, 40(4):115-136.

Gong, X., van Soest, A., Zhang, P., (2005). Sexual Bias and Household Consumption: A Semiparametric Analysis of Engel Curves in Rural China. Journal of Applied Econometrics, 20: 509-527.

Grogan, L. (2007). Patrilocality and human capital accumulation: Evidence from Central Asia, Economics o fTransition, 15(4), pp.685–705.

Gruber, J. (1994). The incidence of mandated maternity benefits. American Economic Review, 84(3):622–641.

Hallman, K., (2000). Mother-farther resource control, marriage payments, and girl-boy health in rural Bangladesh. IFPRI FCND Discussion Paper No.93

Hegland, M. E. (2010). Tajik Male Labour Migration and Women Left Behind, Anthropology of the Middle East, 5(2), 16-35. [Online] Available from: https://www.berghahnjournals.com/view/journals/ame/5/2/ame050203.xml

Hildebrandt, N., McKenzie, D. (2005). The effects of migration on child health in Mexico. Economia, 6(1), pp. 257–289.

Himaz, R., (2010). Intrahousehold Allocation of Education Expenditure: The Case of Sri Lanka, Economic Development and Cultural Change, University of Chicago Press, vol. 58(2), pp. 231-258, 01

Ichimura, H. (1993). Semi-parametric least squares (SLS) and weighted SLS estimation of single index models. Journal of Econometrics 58: 71-120.

ILO, (2011). Mainstreaming child labour concerns in education sector plans and programmes, International Labour Organization, International Programme on the Elimination of Child Labour (IPEC) - Geneva.

Imbens, G. W. and Wooldridge, J. M. (2007). What's new in econometrics? lecture 10 difference-in-differences estimation. NBER Summer Institute. [Online] Available from: www.nber.org/WNE/Slides7\T1\textendash31\T1\textendash07/slides_10_diffindiffs.pdf

IndexMundi, (2021). Cotton Monthly Index. [Online] Available from: https://www.indexmundi.com/commodities/?commodity=cotton&months=60&currency=eur

International Monetary Fund, (2011). Poverty reduction strategy of the Republic of Tajikistan for 2010-2012: Annual progress report for 2010. Washington, D.C.: International Monetary Fund.

Jayachandran, S., (2015). The Roots of Gender Inequality in Developing Countries. Annual Review of Economics 2015 7:1, 63-88

Jayachandran, S., Kuziemko, I., (2011). Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. Quaterly Journal of Economics. 126:1485–538

Jayachandran, S., Pande, R. (2013). Choice not genes: Probable cause for the India-Africa child height gap. Economic and Political Weekly, 48(34), 77-79.

Kan, S., (2018). Labor Force Participation of Women Left behind in Tajikistan. GLM|LIC Working Paper No. 40, Institute of Labor Economics

Khanna, P., Kimmel, Z., Karkara, R., (2016). Convention on the elimination of all forms of discrimination against women for youth. UN Women. Plan 50-50 by 2030. [Online] Available from: https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2016/cedaw-for-youth.pdf?la=en&vs=657

Kingdon, G. (2009). The Economic Benefits of Education in India, Journal of the Lucknow Management Association, 4th Annual Issue, November.

Koc, I. and Onan, I (2004). International migrants' remittances and welfare status of the left-behind families in Turkey. International Migration Review, 38(1), pp. 78–112.

Lancaster, G., Maitra, P., and Ray, R. (2008), Household Expenditure Patterns and Gender Bias: Evidence from Selected Indian States, Oxford Development Studies, 36(2), 133-157

Lokshin, M., Bontch-Osmolovski, M. and Glinskaya, E. (2010) Work-related migration and poverty reduction in Nepal. Review of Development Economics, 14(2), pp. 323–332.

Lopez, J. H., Fajnzylber, P. and Acosta, P. (2007). 'The impact of remittances on poverty and human capital: Evidence from Latin American household surveys', Policy Research Working Paper Series No. 4247, Washington, DC: The World Bank.

Lopez-Cordova, E. (2005). Globalization, migration and development: the role of Mexican migrant remittances. Economia, 6(1), pp. 217–256.

Maccini, S., and Yang, Y., (2009). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. American Economic Review, 99 (3): 1006-26.

Mahdavian, E. (2018). Gendered Nostalgia: Tajik Traditional Dance and the Logic of Nationalism. Asian Theatre Journal, 35(2), 329–353. [Online] Available from: https://www.jstor.org/stable/26533717

Mansuri, G. (2006). 'Migration, school attainment and child labor: Evidence from rural Pakistan', World Bank Policy Research Working Paper No. 3945, Washington, DC: The World Bank.

Mansuri, G. (2007). Does Work Migration Spur Investment in Origin Communities? Entrepreneurship, Schooling, and Child Health in Rural Pakistan.International Migration, Economic Development and Policy. Washington, DC:Palgrave Macmillan: World Bank,99-140.

Masterson, T., (2012).An empirical analysis of gender bias in education spending in Paraguay. World Development, Elsevier, vol. 40(3), pp. 583-593

McKenzie, D. and Rapoport, H. (2007) Network effects and the dynamics of migration and inequality: theory and evidence from Mexico. Journal of Development Economics, 84(1), pp. 1–24.

McKenzie, D.,Rapoport, H., (2011). Can migration reduce educational attainment? Evidence from Mexico, Journal of Population Economics, Vol. 24(4), pp. 1331-1358.

McMohan, W.W. (2002), Education and Development: Measuring the Social Benefits, Oxford University Press.

Micklewright, J. (1999). 'Education, inequality and transition', Economics of Transition, 7(2), pp.343–376.

Mullatti, L. (1995). Families in India: Beliefs and realities. Journal of Comparative Family Studies, 26(1), 11–25

Murphy, S., Belmonte, W., Nelson, J., (2009).Investing in girls' education. An opportunity for corporate leadership. Corporate Social Responsibility Initiative: Harvard Kennedy School.

Nguyen, T.,Purnamasari, R. (2011): Impacts of International Migration and Remittances on Child Outcomes and Labor Supply in Indonesia, World Bank Policy Research Working Paper No. 5591.

Northrop, D. (2001). Subaltern dialogues: Subversion and resistance in Soviet Uzbek family law, Slavic Review,60(1), pp.115–39

OECD-FAO Agricultural Outlook, (2020). OECD-FAO Agricultural Outlook. Chapter 10. [Online] Available from: https://www.oecd-ilibrary.org/sites/630a9f76-en/index.html?itemId= /content/component/630a9f76-en

Oosterbeek, H., Ponce, J., Schady, N., (2008). The impact of cash transfers on school enrollment : evidence from Ecuador. Policy Research Working Paper Series 4645, The World Bank.

Parpiev, Z., Yusupov, K., Yusupov, N., (2012). Outlay equivalence analysis of child gender bias in household expenditure data.The Economics of Transition. The European Bank for Reconstruction and Development, vol. 20(3), pages 549-567

Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance. The Quarterly Journal of Economics, 123(3), 1251–1285.

Qodir, H., (2012). Education Hampers Marriage for Rural Tajik Women. IWPR. Institute for War & Peace Reporting. [Online] Available from: https://iwpr.net/global-voices/education-hampers-marriage-rural-tajik-women

Rose, E., (2000). Gender Bias, Credit Constraints and Time Allocation in Rural India.The Economic Journal, 110(465):738-758

Saavedra, J., Garcia, S., (2012). Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries.RAND Corporation. Working Paper 921-1

Schultz, P. T. (2002), Why Governments Should Invest More to Educate Girls, World Development, 30(2), 207-225.

Sen, A., (2003). Missing women — revisited.British Medical Journal,327(7427),pp.1297–1298.

Sen, A., Sengupta, S., (1983). Malnutrition of Rural Children and the Sex Bias. Economic and Political Weekly;Annual Number (18)

Silverman B. 1986. Density Estimation for Statistics and Data Analysis.Chapman and Hall: London

Skoufias, E., Parker, S., (2011). Conditional Cash Transfers and their impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico. International Food Policy Research Institute. IFPRI FCND Discussion Paper No. 123

SOAS, (2010). WHAT HAS CHANGED? Progress in eliminating the use of forced child labour in the cotton harvests of Uzbekistan and Tajikistan. Centre for Contemporary Central Asia and the Caucasus. [Online] Available from: https://www.soas.ac.uk/cccac/centres-publications/file64329.pdf

Subramanian, S, Deaton, A., (1991). Gender Effects in Indian Consumption Patterns. Sarvekshana 14 (4): 1-12

Trading Economics, (2021). Tajikistan Exports. [Online] Available from: https://tradingeconomics.com/tajikistan/exports

Ueyama, M., (2007).Income Growth and Gender Bias in Childhood Mortalityin Developing Countries.IFPRI Discussion Paper 00739.

Umarov, Kh., (2011). Monitoring of child labour use in the cotton fields in Tajikistan. Dushanbe : International Organization for Migration. [Online] Available from: https://www.loc.gov/item/2014363404/

UN, (2020). Achieve gender equality and empower all women and girls. [Online] Available from: https://unstats.un.org/sdgs/report/2020/goal-05/

UNESCO Institute for Statistics, (2021). eAtlas of Gender Inequality in Education. [Online] Available from: https://en.unesco.org/news/

Van Atta, D. (2009). "White Gold" or Fool's Gold? The Political Economy of Cotton in Tajikistan. Problems of Post-Communism, 56(2), p. 17–35.

Xue, M., (2019). High-Value Work and the Rise of Women: The Cotton Revolution and Gender Equality in China. Revise and Resubmit, Review of Economic Studies

Yount, K., (2003b). Gender bias in the allocation of curative health care in Minia, Egypt. Population Research andPolicy Review 22 (3): 267-29

Zimmermann, L., (2012). Reconsidering gender bias in intrahousehold allocation in India. Journal of Development Studies, vol. 48(1), pp. 151-163

2.10   Appendix

The intuition of the current analysis on parental preferences is based on the combined model of Chiappori (1992) and Pasqua (2005). In this section I will present a simplified version of it. Consider the household with father-husband (m), mother-wife (f) and two children: boy (b) and girl (g), living in 3 periods: 0 - individuals are children and they live with their parents; 1 - people get married and have children, they decide their consumption, labor supply and investment in children's education; 2 - individuals are old and they do not work, they survive with their savings and with care/transfers from their children. Moreover, there are two types of goods: private and public. Parental preferences are represented by unique utility function: $U_{HH} = U_{HH}(c_1^m, c_1^f, c_2^m, c_2^f, x_1^b, x_1^g)$, where c_t^i - consumption of goods for parent i=m,f in period t=1,2; and x_t^j - investment in child j's education (j=b,g). Let's assume that labour supply is inelastic and equal to one. Therefore, parental income = sum of non-labour income I^i and labour income $\bar{w}^i(1 + \eta^j \bar{x}_0^i)$ where $\bar{w}^i$ - wage of the non-educated worker, $\bar{x}_0^i$ - level of i's education, η^j - rate of return to education.

Given the nature of the model parents decide how much to invest in their children's education given. There is a minimum cost of education p_1^j, while returns to education which is equal to η^j [0,1] are different for boys and girls. This model also captures existing social norm according to

85

which adults give a fraction λ^j [0,1] of their income to old parents, which again different for boys and girls. Thus, the household's maximization problem is:

$$\max_{c_1^m, c_1^f, c_2^m, c_2^f, x_1^b, x_1^g} U_{HH} = U_{HH}(c_1^m, c_1^f, c_2^m, c_2^f, x_1^b, x_1^g)$$

$$\text{s.t.} p_1^m c_1^m + p_1^f c_1^f + p_2^m (1+r)^{-1} c_2^m + p_1^b x_1^b + p_1^g x_1^g$$

$$= (1-\lambda^b)((\bar{w}^m(1+\eta^b \bar{x}_0^m)) + I^m) + (1-\lambda^g)((\bar{w}^f(1+\eta^g \bar{x}_0^f)) + I^f)$$

$$+ (1+r)^{-1}(\lambda^b(\bar{w}^m(1+\eta^b x_1^b) + I^b) + \lambda^g(\bar{w}^f(1+\eta^g x_1^g) + I^g))$$

Since this model is applied for the case of Tajikistan, the traditional values are dominant, as a result both parents prefer to invest into son. The solution becomes:

$$\frac{x_1^g}{x_1^b} = \left(\frac{p_1^b - \lambda^b \bar{w}^m \eta^b}{p_1^g - \lambda^g \bar{w}^f \eta^g}\right)\left(\frac{\epsilon}{\delta}\right)$$

It can be seen that the difference between level of education between boys and girls depends on the relationship between rate of return and the fraction of their income for old parents. Meaning that the lower is the rate of return and the higher the share of income dedicated to parents, which is typical case for the countries with strong patriarchal values, the bigger will be an educational gap between sons and daughters. Since Tajikistan is a country with prevailing traditional values, it can be expected from parents to invest more into sons than daughters when it comes to education.

3 "Alphabet beyond the role of a communication tool"

Abstract

This paper explores the consequences of the language reform beyond the role of a national building tool. In particular it focuses on the change of the alphabet, which has not been systematically tested yet. This study intends to provide a new evidence on the role that the alphabet plays in building national sentiments, but also affecting educational and occupational choices. I focus on the case of Uzbekistan which experienced the most frequent change of alphabet systems within the last century. This paper aims to uncover the effect of exposure to the Latin alphabet on informational literacy, educational and occupational choices. The analysis is built on the difference between the cohorts who got affected by the alphabet reform of 1993 and were the first to get the complete secondary education by using Latin alphabet. I also discuss the observed trends in foreign and domestic policy of the country that motivated the change of the alphabet and its potential consequences. The comparison between the cohorts prior and after the change indicated that the level of literacy and informational literacy are higher among children who studied by the means of Latin alphabet. In addition, the exposure to the Latin alphabet is associated with more people choosing to work in state administration, while the number of young entrepreneurs tends to decrease. In the case of Uzbekistan, the change of alphabet had an impact on occupational choices, educational outcomes as well as people's ability to perceive and digest important information. These findings signal that language reforms tend to have wider impact than intended political target.[1]

JEL classification: Z13, J24, O10

Keywords: language, alphabet, national identity, literacy, occupation

"Language of the nation is the best what the nation has."

<div align="right">*Joseph Brodsky*</div>

3.1   Introduction

Biologists and psychologists have long emphasized the centrality of the language in shaping cognitive development (Moore et al, 2014, Lupyan, 2016). Economists followed the discussion and showed that language also acts as a repository for cultural values (Chen, 2013, Sutter et al, 2015, Guin, 2017). Moreover, words and images are linked in the mind of people. Hence, language - not only as its sounds, also as its written symbols – is a powerful tool of cultural identification (Chen et al, 2006). On top of that, language forms a basis for ethnic, regional, national or international identity. The concept of nationhood resonates with an adoption of national language and its alphabet around which the diverse ethnic communities can unite. In France, for instance, the forceful adoption of French as the national language significantly reduced the import and value of the ten-plus regional dialects. As a result, France could boast of a true national culture; nationhood had been secured thanks to a unifying language (Blanc & Kubo, 2021). Likewise, the adoption of Kiswahili in Tanzania served as a securing force strengthening the nationhood's feeling. While in Kenya, the concept of nationhood remains evasive, partially due to the ambivalent status of Kiswahili (Lakoff & Johnson, 1980, Jennings, 2007).

It is crucial to emphasize that culture and language are intertwined like the two-sides of the same coin. Language gives full expression to people's values and norms, and since values and norms are dynamic by nature, language has to be in tandem with cultural transformations. Technological, political, economic and social innovations require the language to enrich its lexicon and representation to capture the new realities. Therefore, any policy intervention targeting the language almost always aims at changing one of the mentioned aspects of nation's well-being.

These properties of language make its reform a very potent political tool. However, in certain circumstances it implies not only huge budgetary costs, but also entails a gap between generations, loss of access to the accumulated knowledge base and slows down the development of the language (Jalilov, 2021). Despite having a long lasting effect, these changes are very often implemented quickly. Hence, it creates a favourable setting to study the evolution of socio-cultural norms in the short run, which are still poorly understood (Bursztyn et al, 2017, Giuliano & Nunn, 2019). Due to the link between words, mental representations and people's behaviour being very close, I hypothesize that the language reform such as change of the alphabet will have a wider influence than intended strengthening of national identity. This study aims to contribute to the literature on socio-economic consequences of rapid cultural changes by evaluating whether alphabet's change affects informational literacy, educational and occupational choices.

This paper focuses on the case of Uzbekistan which had the most frequent alphabet change during one century (Uzman, 2010). The country officially switched to Latin alphabet in 1993, however, media,

government and people are predominantly using Cyrillic (Alimdjanov, 2019). Identity in Uzbekistan has taken many forms as a consequence of the multi-ethnic state, a history of conquests, rise and fall of the communist regime. The inclination of Uzbek government to strengthen the authority of Uzbek identity with the help of Latin alphabet has a series of negative consequences both for the development of the language and access to the information (Juraev, 2016).

Recent studies have shown the effectiveness of several nation-building tools, such as sport (Depretis-Chauvin et al, 2018), propaganda (Blouin & Mukand, 2019), population resettlement programs (Bazzi et al, 2019) and education (Bandiera et al, 2018, Alesina et al, 2019, Bazzi et al, 2020). However, change of the alphabet as a nation-building tool has not been systematically tested yet. This project provides a new evidence on the role that the alphabet plays beyond a simple communication tool. The findings of this study are far stretching and relevant for the other countries. For instance, it could be applicable to the case of Serbia, where nationalists insist on using the official Cyrillic alphabet, while most people use Latin (Jovanovic, 2018). Similarly, Tatarstan, one of the largest Russian regions, is considering to adopt Latin alphabet for Tatar language in order to strengthen Tatar ethnic identity over broader Russian Federation notion (Bartolomä, 2016).

In order to shed light on the impact of the alphabet change, this paper focuses on the differences between the cohorts who got affected by the reform of 1993[2] and were the first to receive their complete secondary education using Latin alphabet. The respective cohorts were differentiated by the month of birth, since a child should reach 7 years old by the end of September[3] (Angrist & Krueger, 1991, Fouka, 2016). Those who were born in 1989 before October, were eligible to start school in 1996 when the complete transition to teaching in Latin alphabet was made. By comparing existing trends between those cohorts, I find that the level of literacy and informational literacy are higher among the people who were exposed to the Latin script. Also, the knowledge of Latin alphabet is associated with more people choosing to work in state administration, while the number of entrepreneurs tends to decrease. However, the general level of tertiary education of the affected cohort also tends to be lower. These findings show that language reforms, in particular, alphabet change, tend to have an impact beyond their intended political target. They may also affect people's ability to perceive and process important information and the formation of certain labour market trends.

The paper is organized as following: Section 3.2 discusses the existing literature on the role of language in various socio-economic settings. Section 3.3 introduces language reform in Uzbekistan in more details. Section 3.4 presents potential consequences of the alphabet change. Section 3.5 describes data and section 3.6 discusses methodology that has been used. Results are presented in section 3.7. Section 3.8 concludes and section 3.9 lists the bibliography.

---

[2]The reform was announced in 1993, but it was adopted only in 1996, since more time was required to adapt educational materials. The reform meant that official alphabet in use is Latin and school education should be conducted by the means of this alphabet.

[3]The 2nd of September is an official day when the school starts in Uzbekistan.

3.2 Literature Review

Different studies have attempted to evaluate the socio-economic consequences of language change, education and its relation to the national identity. In particular, Fouka (2016) sought to establish the impact of assimilation policy in the United States and language policy restrictions on integration and identification with the host country. His paper employed a difference in difference approach by incorporating grouped data of school-age children and old age cohorts by exploiting the effect of German language banning in the United States elementary schools. The policy attracted a backlash due to promoting a heightened sense of cultural identity among the minority groups in the United States. The findings suggest that the affected cohorts were less likely to participate in World War II. Also, the studied cohort were more likely to marry within their ethnic groups, with a decided choice of German names for their offsprings (Fouka, 2016).

Other studies also observed that the identity formation process played a pivotal role in the development, general perception and the capacity of public bodies to carry out interventions. A study by Clots-Figueras & Masella (2013) sought to determine the policy change impact of introducing the Catalan language alongside Spanish in schools (Clots-Figueras & Masella, 2013). Similar to the approach in this paper, Clots-Figueras & Masella used the years of exposure to the new education policy after its implementation in 1983. Their results showed that education policy changes affected political preferences and attitudes towards the organization of the state. Individuals born in the 1970s were more likely to undergo partial exposure to the policy, while those born in later years were more likely to experience total exposure to the system. Similar effects were common among non-Catalan natives. Thus, social and political attitudes depended on individuals' exposure to the new policy.

In addition to the alternation of socio-economic outcomes, language policies also arise from intergroup association and integration. For instance, Bazzi et al. (2019) studied the population resettlement program in Indonesia and its impact on intergroup contact. In 1980, the national government promoted a resettlement program by relocating over 2 million ethnically diverse migrants. Based on this case analysis, the research established a greater integration in fragmented communities that exhibited smaller groups (Bazzi et al., 2019). The fragmentation was evident in the individuals' languages at home, intermarriage patterns, and children's names. While the above statement held in fractionalized societies, polarized communities exhibited a somewhat different attribute. Particularly, polarized communities with large groups showed an increase in ethnic attachment and a subsequent decline in integration. Overall, Bazzi et al (2019) concluded that residual segregation dampened the associated effects of public goods and social capital.

Another important dimension of how language policy affects nation building is education. Mass schooling is fundamental in enhancing nation-building. However, a subsequent work of Bazzi et al. (2020) established negative impacts of competition within the education sector in Indonesia. Primarily, religious schooling promoted system schooling and catered for a large student population. Indonesia's Islamic schools, however, adversely responded to the state's attempt to establish over 600,000 public elementary

schooling centers. The study used novel data on construction and curriculum to confirm the short- and long-run impacts of the shifts and education dynamics on exposed cohorts. First, religious education increased in Islamic schools. The system also entered new markets with greater competence to compete against the government's development of elementary schools. While the Islamic state response enhanced greater religiosity and education, the action presented adverse effects on secular national unity. Therefore, the emergent competition between the two entities undermined the nation-building impacts of mass schooling. On the other hand, there are some papers showing the positive relationship between education, in particular, civic schooling, and nation-building. This thread of research evidence further indicates that America was the earliest most learned country globally, having achieved an enrollment rate of 50 percent and a more significant divergence from other countries (Bandiera et al., 2018). Primarily, America introduced civic laws to instill civic values to culturally change migrants' mindset in the age of mass migration during the 1850s. An analysis of cross-country data suggests that migrants exhibited a lower demand for America's schooling pre-compulsion (Bandiera et al., 2018). However, it also showed that civic education has a pivotal role in enhancing nation-building. First, the values provided a foundation for state-building by underpinning democratic institution roles. Thus, the policy improved performance from all the stakeholders. Also, civic values enable children and adults alike to make informed decisions. For instance, the policy instilled discipline that allows children to deal with issues regarding instability, reducing the costs of social interaction, coordination, and information exchange. Based on the policy implementation, residents in the United States exhibited an increased personal responsibility by taking actions to enhance social welfare. Therefore, socialization played a role in curriculum design due to its impact on promoting national unity (Bandiera et al., 2018). More importantly, language choices and cultural practices widely impact identity development in the education of minority students. An interview conducted on Latinas and other ethnic groups sought to establish the above statement. It evaluated the immediate effects of language, cultural identity, and loss change on individuals' lives. Further, the study assessed English as a second language on individuals' social development and identity (Patricia, 2012). The findings suggest a significant relationship between the variables under study. For instance, there existed a pervasive impact of language identity on language choices among minority students. Further, these choices impacted career paths by limiting the choices and regions one was willing to work with based on language barriers. Overall, language choice, policies, and values significantly impact national unity, building, and growth.

Some studies explored how language reforms interacted with schooling and their contribution to developing individual identity. For example, Bracons et al. (2008) conducted a survey in Catalonia and Basque analyzing compulsory language educational policies and their contribution to identity formation. Using regression analysis, the survey realized that while the compulsory language policy implemented in Catalonia shows a significant effect on identity, the non-compulsory policy adopted by Basque does not have any effect on identity. In addition, the surveyors identify the effect of law using a difference in differences approach to analyze the impact of cohort variation in both languages of instruction and years of education. Moreover, the survey findings depict that the citizens who attained the Catalan compulsory system have stronger feelings than their counterparts in Basque, who achieved a non-compulsory system.

Based on the regression analysis, the compulsory rate in either system dictates the differences between the two reforms. The surveyors concluded that a non-compulsory system did not interfere with the intergenerational transmission of identity in Basque since parents do not choose the language of instruction, and weaker parental influence exists. On the contrary, critical systems affect identity formation in Catalonia.

Similarly, Cantoni et al. (2015) conducted a survey in China on the impact of school curricula on student's political attitudes from 2004 to 2010. The survey depicts that the new academic curriculum in China contributed to changed views on the country's governance, change in views on democracy, and a criticism towards the country's free markets. Further, the survey measures the political beliefs and attitudes of four students from 29 Chinese provinces who went to high school between 2006 and 2009. The researchers applied a generalized difference in differences framework to test how students' political attitudes who underwent the new curriculum differed from those of their counterparts who underwent the old curriculum. To derive hypotheses for testing, the surveyors examined government documents containing the goals of the reform and compared the old versions of textbooks to new textbooks both qualitatively and quantitatively. Conversely, the survey reveals that the government of China wished to affect the attitude of students on governance, political institutions, economic institutions, Chinese identity, and the views towards China's environment. Regression analysis helped depict that the new curriculum positively impacted the students' attitude, enabling them to develop trust in the country's broad range of government officials, perceive the political institutions as democratic, and become more skeptical of free markets.

It can be seen that education and school curriculum are strong tools of the nation building process (Cantoni et al, 2015, Bandiera et al, 2018, Bazzi et al, 2020). Therefore, when the language reform targeting to strengthen nation identity is adopted through the educational system, it may have more powerful impact beyond the initial political aim. For example, it may affect occupation choices. For instance, Biavaschi et al. (2013) sought to explore the economic payoff of the name Americanization of immigrants in the labor market. They found out that there exists a relationship between individual identity and labor market success in modern America. Their findings showed that immigrants who Americanized their names attained more occupation-based earnings than their counterparts who failed to Americanize their names. Although, Americanization is not a particular language reform, it has a common ground with language and alphabet as a part of a cultural identification. Therefore, it is promising to look at the labour market outcomes of the people affected by language policies.

To date, the existing literature was mainly focusing on the impact of language reforms on national identity building. This paper aims to show that the consequences could be further reaching and affect other dimensions of economic development such as occupational choices, informational literacy and the probability of choosing academic education. In addition, this paper focuses not on a particular language change, but on the switch from one alphabet to another, which is a novel research target of the national identity building.

3.3 The alphabet reform in Uzbekistan

3.3.1 Peculiarity of the Uzbek language and history of alphabet changes

The language that was spoken by Uzbek ancestors before the Russian Empire revolution of 1917 was called Sart. It was based on the Chagatai language which was wide spread in Central Asia and was influenced by Uighur in the past. Kipchak and Karluk dialects used during 19th and early 20th centures were also embedded in the creation of the modern Uzbek language (Allworth, 1990). In 1921 the first regional congress discussing the alphabet reform was held in Tashkent and was followed by the 1st Turkic Congress in Baku, Azerbaijan, in 1926. There it was decided that all Turkic languages should adopt new Latin alphabet - Yanalif, which literally means "new alphabet" and forgo Arabic or Cyrillic scripts. The transition took place 3 years later, in 1929. However, in the spite of growing power of pan-Turkism, Soviet government decided to translate Turkic languages into Cyrillic script. Thus, in 1940 Uzbek language adopted Cyrillic writing symbols (Center for Economic Research, 2013). At the time of the collapse of the USSR, in 1991, Turkic republics participated in the conference that was held in Istanbul, where they decided to switch to the Latin script that will be as close as possible to Turkish. At that period Turkey was an attractive Muslim country with secular values and a government keeping religion aside (Beyond Transition, 2006). Since the collapse of the USSR created a threat of growing Islamic movements in Central Asian region, Uzbekistan among other countries saw Turkish model as an alternative to get separated from Moscow influence and to avoid islamisation. Switching to the Latin script would promote national patriotic rhetoric's of the newly established elite and focus on creation of a new generation separated from Soviet values (CABAR, 2021). As a result, on 2 September 1993 the law "On the introduction of the Uzbek alphabet based on the Latin script" was adopted. Schools should be completely ready and start teaching in new script from 1996. However, deteriorating relationship with Turkey and suppression of the opposition led to the absence of motivation to make the reform going beyond educational and governmental sectors (CABAR, 2021). Moreover, Russia remained a major trade partner of Uzbekistan and a big political player in the region with a trade turnover comprising US $5.6 billion US dollars in 2020 (State Statistics Committee, 2020). The process of adoption of Latin alphabet stagnated and was not supported by administrative means. It was widely accepted since it gave freedom to the different segments of the population to use Uzbek language as they would prefer. As a consequence, Uzbekistan became a country where two alphabets (Latin and Cyrillic) coexist for one language, but with various application purposes and various types of users. The adoption of the law in schools has led to further separation between the young generation, who was educated in the Latin, and the old generation, who strongly preferred Cyrillic alphabet (Alimdjanov, 2019, CABAR, 2021).

3.3.2 Associated consequences of the alphabet reform

The adoption of Latin alphabet at the earliest stages of independence prompted many challenges for government, society and language itself. The strenuous situation around the Uzbek language arises from the fact that high ranking government officials still do not have a profound knowledge of the language and

failed to switch to Latin alphabet themselves. Most of Uzbek elite leaders got their education and formed as officials during USSR times. After Uzbekistan acquired independence, the perception that the quality of education being substantially better in Russian combined with the fact that Uzbek elite speaks Russian - both made Uzbek to be associated with the language of the mass (Alimdjanov, 2019, CABAR, 2021).

The reform was announced in 1993, however, it was frozen till 1996 when the first cohort had the full school curriculum in Latin alphabet. Despite full adoption of the law within educational institutions across the country, Uzbek did not become a language of power and politics. Wealthy class preferred to speak Russian, while low and middle classes were using Uzbek full of dialects and introducing English and Russian words in it. The papers' workflow, record and correspondence management are officially held in Uzbek, while high ranked executes and parliament members actively use Russian. The reason is not only that their careers were built in Soviet administration, but also the general stagnation of the Uzbek language. For example, the administrative Uzbek was not developed over time, which means that it is not well adapted for a legal and political usage (Alimdjanov, 2019). During the observed time span[4] laws were still being created in Russian. Russian achieved its niche as a language of well-educated privileged people and a communication tool between ethnicities. Local intelligentsia[5] and elite created a subculture of modernized and europeanized part of the population, who mostly spoke Russian. The spoken language became an indicator of progress. to the point where the mere ability to speak Uzbek has become a sign of underdevelopment. The nation has lost Soviet Uzbek literary language. The lack of development and the poor quality of textbooks did not allow Uzbek to flourish and take its intended central part in the life of a newly created country (Alimdjanov, 2019).

The adoption of Latin alphabet caused mixed reaction from the public (Sayidov & Umansky, 1998). Initially, it was associated with strengthening pan-Turkic[6] ideology and an attempt to discriminate Russian-speaking minority. Cyrillic alphabet was a symbol of the common Soviet past and strong ties with Russia. Therefore, Russian speakers perceive the reform as a threat of increasing hatred towards them. However, a fraction of Uzbek ethnic population considered being part of the USSR as a continuation of Russian expansion and they welcomed any attempts to strengthen national identity. As a result, the reform caused ethnic fragmentation (Center for Economic Research, 2013). Moments of intolerance to Russian, islamization of the region and more favorable conditions in Russia resulted in migration of Russian speaking population. It led to the 50% decrease after USSR's collapse and currently comprises around 4.7 million people (Shustov, 2016). This became a problematic issue for transition economies, since highly qualified people, especially in medicine, education and science, were the ones migrating (Beyond Transition, 2006).

---

[4]From 1993, when the reform was announced, to 2013 when the data were gathered.

[5]Group of well-educated people involved in the rigorous mental labour that provides criticism, guide in forming the culture and politics of their society (Williams, 1983).

[6]"Pan-Turkism is the name given to the idea of uniting all Turkic-speaking peoples of the Caucasus, the Volga-Ural region, the Crimea, Western and Central Asia under the aegis of a greater Turkish state" - Pekesen, 2019.

In additional to the ethnic division, the reform bisect the society based on affiliation to the certain generation. The annual report by Center for Economic Research (2013) stated that there were mainly two groups of people: those older than 30, who learned in Cyrillic and are slow to read and almost unable to write in Latin, and those below 30, who learned in Latin, but do not have proper access to the information, since most of it still is in Cyrillic. According to Alimdjanov (2019), at least 10 million received school education in Latin alphabet, which comprises nearly 30% of the total population (43% of 15-64 years old) - the most active part. It can be seen that there is a division into two large groups with different preferences towards alphabet in use. Despite the perception that Latin alphabet should mainly benefit young group of the population, it is far from being true. This reform made the majority of young population to have a scarce access to the literature and knowledge that has been publishing in Cyrillic from 1940. During this time nearly eighty thousand books were published including famous 14 volume Uzbek Soviet Encyclopedia (CABAR, 2021).

The quantity of educational materials written in Latin is very scarce and the quality is rather low. For example, the libraries of the Academy of Sciences of Uzbekistan have over 5 million copies, predominantly in Cyrillic. The same situation is observed in other libraries of the country. Media, advertisement companies and government officials keep using Cyrillic to disseminate the information (The Open Asia, 2017, Schweitzer, 2020). Therefore, if young people have a poor command of the Cyrillic alphabet, they do not have an opportunity to reach fiction, scientific and classic literature. Poor access to the different types of literature affects not only the quality of education received by young generation, but also the personal growth and identity formation. For instance, if they learn about Uzbekistan history only from the textbooks at schools without any access to the primary materials, they could not have an objective vision of it. In addition, they do not have an access to the literature that was/is associated with anti-establishment views, criticism towards government. The education of young generations can be restricted only to mass media and the course that the government has chosen. Given poor access to information, the adoption of Latin alphabet also created distortions on the labour market, especially in public sectors like education, health and governmental jobs (Beyond Transition, 2006). In addition, Cyrillic alphabet is deeply associated with Soviet legacy and Russian political influence. It is attributed to the foreign culture, which do not coincide with traditions of Uzbekistan (Otamuradov, 2016). This perception is segregating the society into a generation of Soviet legacy and a generation of Independence, who may believe that Latin alphabet represents a more independent Uzbek society and identity, and it is also more practical in use (Schweitzer, 2020).

The adoption of the Latin alphabet was mainly inspired by political interests and the idea of national independence from Soviet legacy. Instead of strengthening publishing and printing facilities, consolidating cultural heritage, forming common rules for Uzbek language before switching to the Latin alphabet, the law divided population in halves. In its attempt to get rid from the Soviet past, independent Uzbekistan just reinforced old way of decision-making and strong-arm administration.

3.4   Data

This section introduces the data that have been used in this paper to demonstrate to which extent the cohort who got an education in Latin alphabet is different to the ones who did not in terms of informational literacy, educational and occupational choices. The data came from CALISS 2013 (Central Asia Labour Immigration Skills Survey) conducted in Uzbekistan. It is a nationally representative household dataset collected by the World Bank and GIZ[7]. It covers nearly 1500 households in 14797 mahalla committees across 14 regions of Uzbekistan. The total proportion of urban population comprises approximately 37% and rural – 63%. More details about the data collection can be found in the next subsection.

3.4.1   Data collection

The survey on labour, skills and migration in Uzbekistan was conducted between the 30th of July and the 6th of September of 2013. The sample consists of 1,500 households with 8,622 individuals, representative at the oblast (region) and urban/rural level[8].

	Strata	Urban Population	%	Sample	Rural Population	%	Sample
1	Karakalpakstan	764 100	49.1%	30	790 700	50.9%	31
2	Andijan	684 600	30.0%	27	1 601 200	70.0%	63
3	Buhara	449 600	30.4%	18	1 028 700	69.6%	40
4	Jizzakh	307 000	30.0%	12	717 900	70.0%	28
5	Kashkadarya	575 600	25.0%	23	1 722 400	75.0%	67
6	Navoi	320 400	39.9%	12	482 300	60.1%	19
7	Namangan	758 800	37.6%	30	1 260 900	62.4%	49
8	Samarkand	736 900	26.3%	29	2 060 900	73.7%	81
9	Surhandarya	360 600	19.6%	14	1 477 700	80.4%	58
10	Syrdarya	211 900	31.7%	8	455 900	68.3%	18
11	Tashkent	963 200	39.7%	38	1 463 000	60.3%	57
12	Fergana	799 000	28.8%	31	1 978 500	71.2%	78
13	Horezm	321 200	23.1%	12	1 072 200	76.9%	42
14	Tashkent city	2 157 800	100.0%	85			
	TOTAL in Uzbekistan	9 410 700	36,9%	369	16 112 300	63.1%	631

Table 3.1 Proportionate stratification by urban/rural population: CALISS, 2013

The sample is grouped into 75 PSUs[9], which are geographical areas of a walk-able size. Each PSU was then grouped according to the population size of each 27 urban and rural region. Afterwards, weighted proportionally to its size, each PSU is randomly chosen within each region. Each PSU comprises of

---

[7]Deutsche Gesellschaft für Internationale Zusammenarbeit.

[8]The sampling strategy is an adaptation of the sampling methodology commonly used in Uzbekistan.

[9]Primary sampling unit.

96

20 households, which are chosen using a geographical sampling procedure. This procedure consists of generating a random point using a numbered grid over a map. From this starting point within the PSU, one out of every 5 households is interviewed, following a systematic route designed for each PSU. The total number of either refusals or absences noted after 3 attempts amounts to 1,067 households. Each missing and refusal was replaced with another household by extending the geographical sampling procedure within the PSU[10].

At the household level, there were two parts of the questionnaire targeting two different categories of people within the household. First of all, the most knowledgeable person of the household[11] was asked the main part of the questionnaire, which includes questions regarding each household member for their education, health spending and labour and migration. This main part also includes a complete household expenditure module, questions about remittances, government transfers, financial services, subjective poverty and questions about the housing conditions. The second part of the questionnaire was asked to a randomly chosen adult in the age from 15 to 64 who is not currently a migrant, using a random number table (Kish grid) to ensure the randomness of the selection[12]. In this part there were questions regarding labour and work expectations, migration and preparation for migration, language skills, and technical skill training. It also included a self-assessment of technical skills and knowledge, as well as non-cognitive, and a cognitive test. From this last sector of non-cognitive and cognitive tests, the measures for literacy and informational literacy were taken.

Finally, control visits by independent consultants ensured the quality of the data collection exercise. A total of 156 home visits and 225 phone calls, with at least 3 controls in each PSU allowed an early detection of various issues and lead to a successful resolution of the identified problems.

3.4.2   Construction of literacy and informational literacy indices

The literacy and informational literacy indices were part of the construction of cognitive skills evaluation by Tubbs, C., Bahry, L., Audy, R., (2013). Data come from a 34-item survey module designed for use by the World Bank to assess five different "cognitive" skills. In this paper I solely focus on literacy and informational literacy, which are part of the domain-specific skills (Boekarts, 1997). Within literacy, these concepts include: (1) semantics, assessed using seven items, with five items assessing respondents' familiarity with vocabulary, one item testing understanding of a national idiom, and one item measuring comprehension of the meaning of a complex sentence; (2) reading comprehension, assessed by asking respondents to read a 257-word non-technical narrative text and then answering five questions about the text; and (3) information comprehension, assessed using four items based on instructions for taking a medicine and reading a timetable describing inter-city bus schedules. Based on the aforementioned concepts, the questionnaire focused on technical skills, such as reading and writing, to measure literacy

---

[10]4 PSUs out of 75 were replaced for various accessibility reasons, and neighbouring PSUs showing similar characteristics were selected to replace them.

[11]It is determined by the age and the level of decision making in the household.

[12]Sometimes, the same person responded to both sections. Also when it was not possible to reach the person selected after 3 attempts, another person was selected using the same random procedure.

level. Informational literacy was measured by assessing cognitive skills: language, text comprehension, table comprehension, publicity and graph comprehension. The alphabet used was Latin. The exact questions are provided in the Appendix.

3.5   Methodology

In this section I present the approaches I used to evaluate the differences between cohorts affected and non-affected by the adoption of Latin alphabet. First, I present a graphical representation of descriptive trends over time to show the effect of the law on occupational choices of the associated cohorts. Then I differentiate these cohorts by the quarter of birth to show the results for literacy, informational literacy and willingness to get tertiary education. Differentiation by the quarter of birth stems from the fact that children who were born after October 1988 and before October 1989 were eligible to go to school in 1996, when the law came into effect. Following Angrist & Krueger (1991), I could compare the differences between children based on the quarter of birth.

Following Stock et al (2002), Angrist & Pischke (2015) regarding inadequate IVs, Angrist & Imbens (1995), Stephen & Yang (2016) regarding sample heterogeneity associated with school law, and in particular, Hoogerheide & Dijk (2006), Deaton (2009), Deaton & Cartwright (2018) who disputed the empirical strategy of Angrist & Krueger (1991), I do not use the time of birth as an instrument in this analysis, instead, I use it to identify treated individuals, who got exposed to the language reform. Small sample size issues, problem of weak instrument and potential violation of monotonicity[13] are among potential concerns when using 2SLS with quarter of birth. Therefore, in this analysis my identification strategy is a linear model with age and district fixed effects[14] to evaluate the effect of being treated[15] on literacy and informational literacy following Clots-Figueras & Masella (2013) and Fouka (2016).

$$Y_{i,d,t} = \beta_1 Treated_i + \beta_2 X_{i,d,t} + \sum_i \sigma_i Age_i + \sum_d \sigma_d District_i + \varepsilon_{i,t} \qquad (3.1)$$

In this specification $Y_{i,d,t}$ is a measure of literacy and informational literacy of an individual i in district d at time t; $Treated_i$ depicts an individual being treated (0,1); $X_{i,d,t}$ is a set of individual socio-economic characteristics; Age_i corresponds to the age fixed effects and $District_i$ corresponds to the district fixed effects. Standard errors are clustered at district level.

Next, I evaluate the impact of exposure to Latin alphabet on labour market outcomes. For this reason, I use a multinomial logit model with fixed effects to evaluate the impact on the different types of education and fields of work. I change the model to logit given the nature of dependant variable. In particular, education comprises of four potential probabilities: less than secondary, general secondary, technical

---

[13]There is a concern that some parents may delay schooling for their children according to Aliprantis (2007). However, I was able to verify in my sample that everyone who was born in a certain month went to school in the respective year.

[14]It is important to account for sample heterogeneity and regional variation highlighted by Hoogerheide & Dijk (2006). It will account for average differences in the outcome variable across age and districts.

[15]This means being exposed to the Latin alphabet throughout the secondary education attainment.

secondary, tertiary[16].

$$\Pr(\text{Education} = 1 \mid \text{Less, General, Technical, Tertiary}) = \frac{\exp(\beta_0 + \beta_1 X_{i,d,t})}{1 + \exp(\beta_0 + \beta_1 X_{i,d,t})} \quad (3.2)$$

Similarly, I use the same specification for occupational choices. Now the dependant variable is field of work, which includes: entrepreneurship, family business or state jobs.

$$\Pr(\text{Job} = 1 \mid \text{Entrepreneur, Family, State}) = \frac{\exp(\beta_0 + \beta_1 X_{i,d,t})}{1 + \exp(\beta_0 + \beta_1 X_{i,d,t})} \quad (3.3)$$

Multinomial logit model follows the linearity assumption of the main specification. It is mainly required, because the dependant variable (both education and occupational choice) is measured at the nominal level. Important to emphasize that there is no outliers and no individual belonging to two different categories.

3.6 Results

This section describes the existing trends in the data concerning the occupational choices and acquired skills between the cohorts who were affected by the Latin alphabet reform in education and those who were not.

3.6.1 Descriptive trends

In this subsection I present descriptive graphs for various age cohorts in order to evaluate trends in education and occupation before and after adoption of Latin alphabet for teaching at schools. Those who were born in 1989 before October, were eligible to start school in 1996 when the complete transition to teaching in Latin alphabet was made[17].
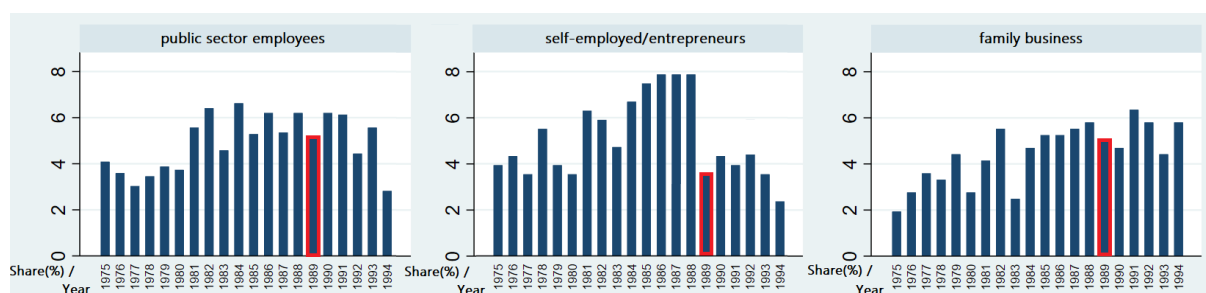

Figure 3.1 The share of employees across different sectors of the labour market by year of birth

Figure 3.1 shows the share of employees by year of birth across three sectors of the labour market: public, family business or self-employment. It can be seen that there was a noticeable drop in the amount of

---

[16]$X_{i,d,t}$ is a set of individual socio-economic characteristics.

[17]The last observed year of birth is 1994, since this person is 19 years old in 2013 when the data is collected. This age is the age of completion of the full secondary education for this person.

people choosing entrepreneurship among 1989-cohort. This could be partially explained by the fact that this was the first cohort taught completely by using Latin alphabet, which made them suitable for jobs requiring writing and reading in Latin, such as state administration and business correspondence. Mechanically, this hypothesis should have led to an increase of employment rate in other sectors. This is not observed in Figure 3.1, however, public sector in this case includes not only state administration, but also education, health, tax legislation among others. So the effect could be covered by the incomplete coverage of these employment categories. As for the involvement into family business no change was observed as expected.
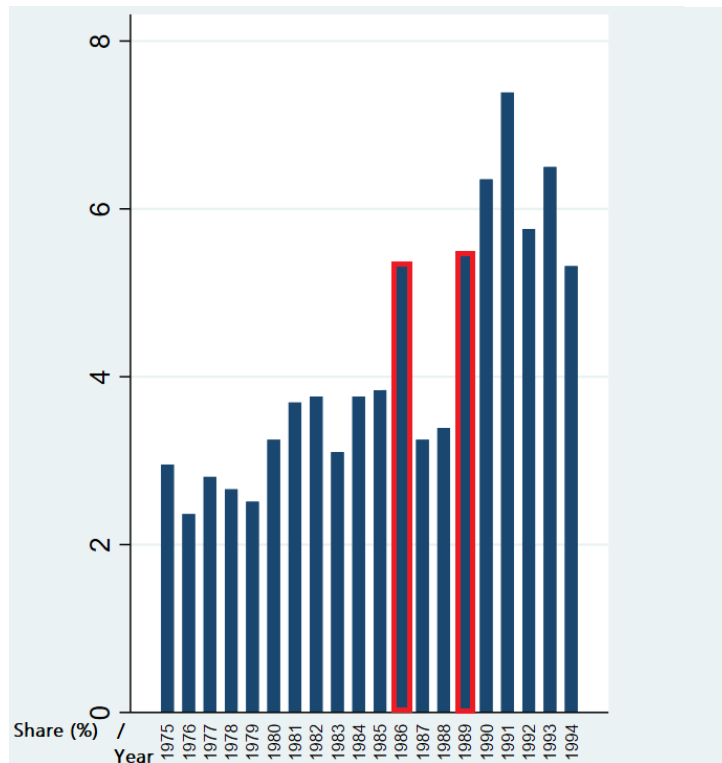
Figure 3.2 The share of employees working in state administration by year of birth

In addition, these data allow me to zoom in state administration job. Figure 3.2 shows that there was an increase for 1986-cohort and 1989-cohort and it remained consistently high. The 1986-cohort is the one who went to school when the change of the alphabet was announced. As a result, they may have thought Latin alphabet is important for the future. However, since the transition happened only in 1996 after announcement of the reform in 1993, the switch to Latin alphabet might have lost credibility in the eyes of the two subsequent cohorts. This could have affect their intention to learn Latin and to access the state administration jobs. When the reform was adopted in the educational sector in 1996, starting from the 1989-cohort children got secondary education fully in Latin which could explain observed increase in state administration jobs. Overall, it can be observed that there is a notable difference between affected and non-affected cohorts by the year of birth across job sectors.

Next, I would like to visualize the response to the law within birth cohort. For that reason, I subdivided years into quarters and exploit the difference in school admission based on the quarter of birth. In Uzbekistan the child should be at least 7 year old before the end of September to be allowed to start studies in the respective year. Since the full adoption of the Latin alphabet at schools happened starting from September 2, 1996, only children born in the 4th quarter of 1988 and the 1st to 3rd quarters of 1989 were eligible to enter school that year. Since people born in the same year are supposed to be similar to each other keeping other characteristics constant, differentiating by the birth quarter shows the potential impact of the reform on people's outcomes.
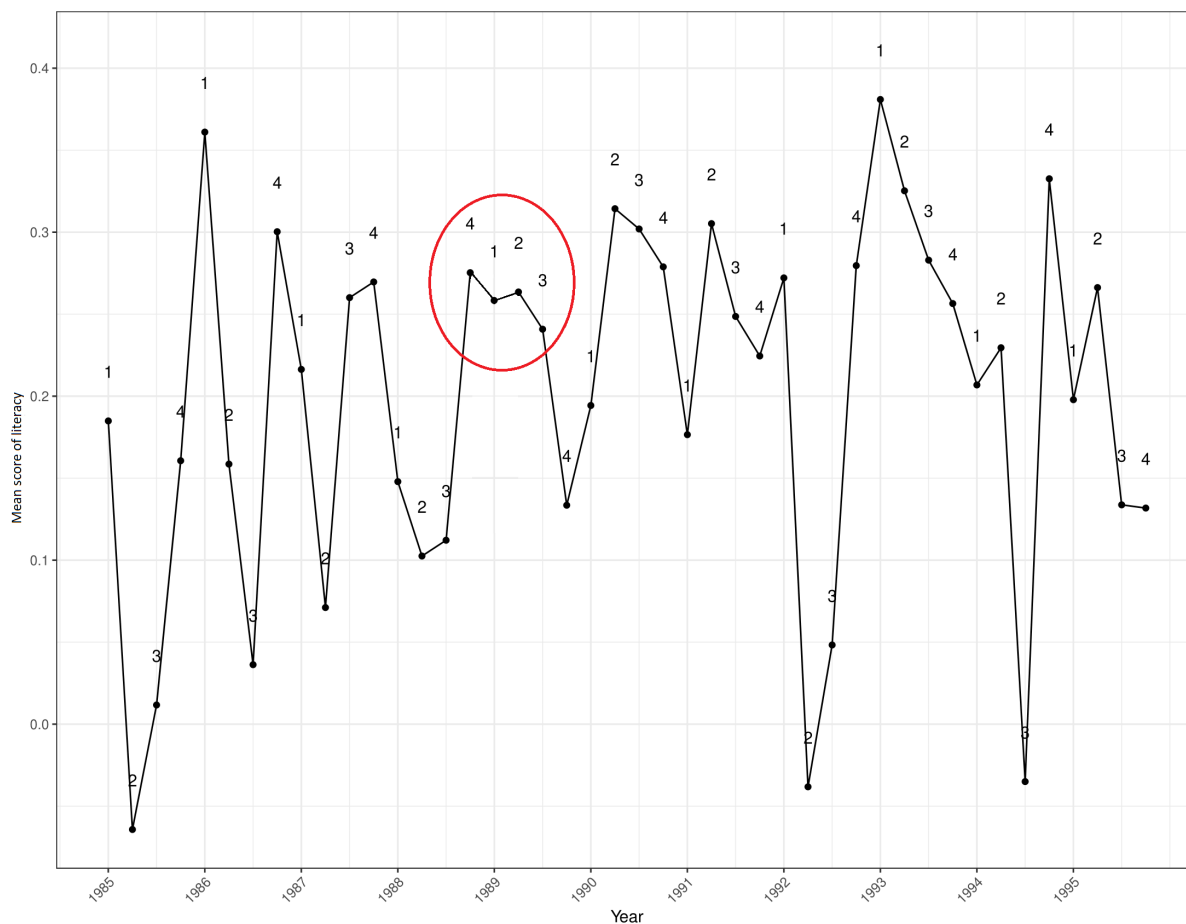

Figure 3.3 Mean score of literacy level by the quarter of birth

Figure 3.3 illustrates the mean level of literacy by the quarter of birth. It can be seen that there is a significant increase for the 4th quarter of 1988 and for the 1st, 2nd, 3rd quarters of 1989, however, there is a notable drop for the subsequent 4th cohort of 1989. It shows that the reform affected people by increasing their literacy levels. Those people who were born the same year, but in the 4th quarter, and people who were born last year show considerably lower level of literacy. This pattern signals that the alphabet reform improved the average literacy level.
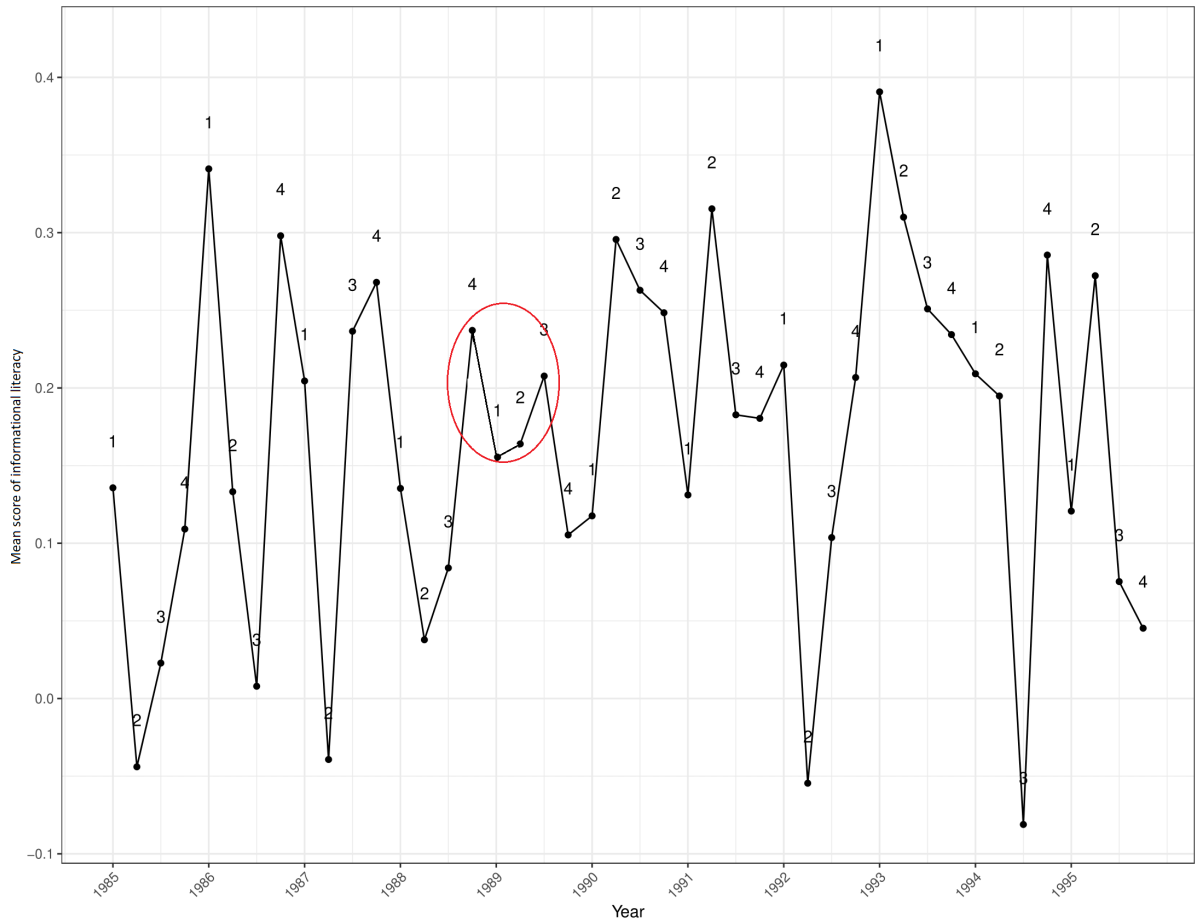
Figure 3.4 Mean score of informational literacy level by the quarter of birth

It is important to look directly at informational literacy, which could reflect whether learning Latin alphabet helps to understand better government announcements, information disclosed to the public during the elections, general advertisements on the streets. From the Figure 3.4 it can be seen that the pattern is similar to the literacy levels. Those people who were born in the last quarter of 1988 and in the first three of 1989 demonstrate higher informational literacy and the average trend keeps increasing. It could support the claim that adoption of Latin alphabet for instruction at schools helped to increase information literacy for the part of population and make them more active in the political realm of the country.
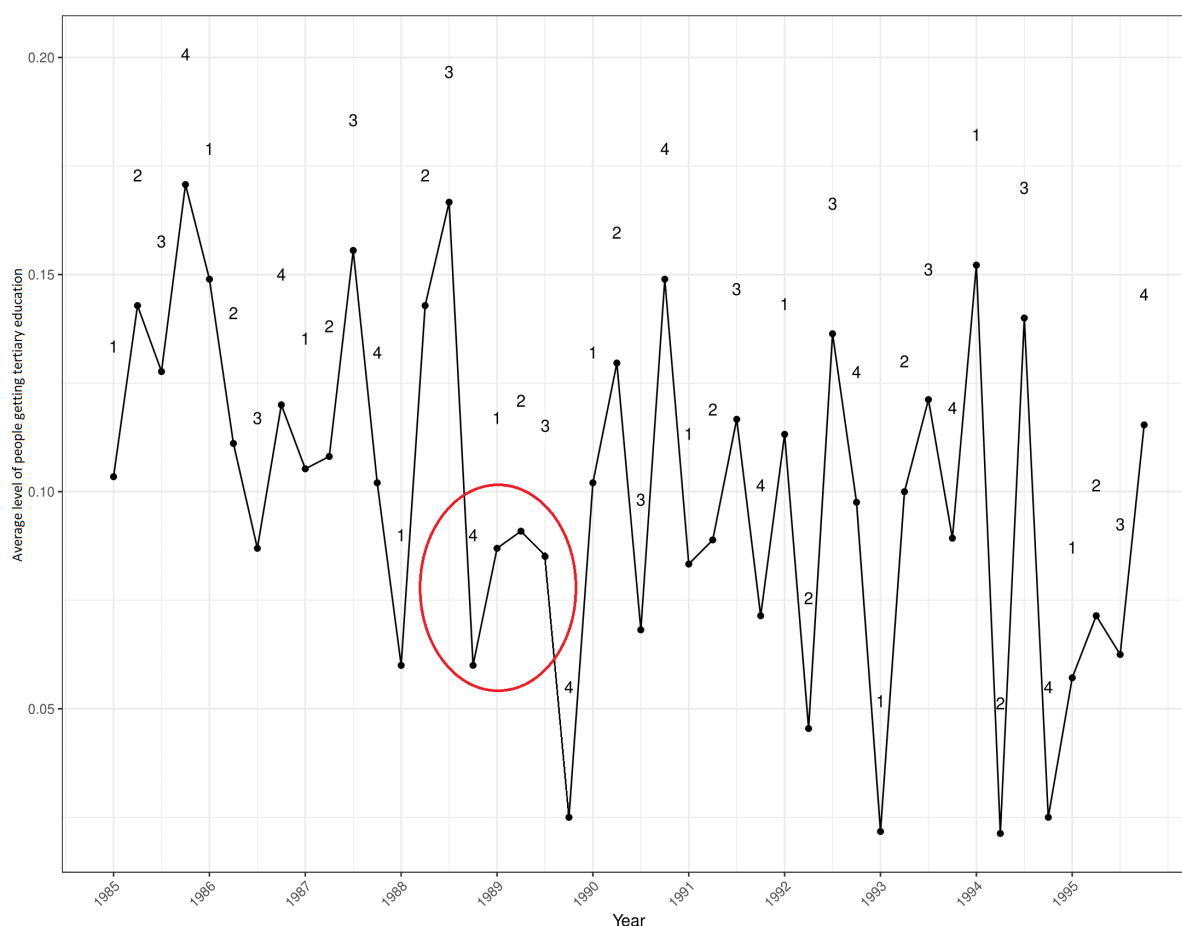
Figure 3.5 Share of people getting tertiary level of education by the quarter of birth

Similarly, I verified whether there was any effect of the law on the willingness of people to continue their education after secondary level. From the Figure 3.5 it can be seen that there was a considerable drop in the average level of tertiary education for the affected cohort. It could have two possible explanations. First, the knowledge of Latin alphabet could have facilitated the entrance to the junior job positions in state administration which do not require high education. Second, university studies could be associated with the work beyond the program and exploiting materials in addition to the textbooks which are vastly available in Cyrillic.

3.6.2 Regression results

In this section, I discuss the regression results following the specification in Section 3.5. First, Table 3.2 shows the results for equation 3.1, particularly, for literacy and informational literacy. It can be seen that the treated cohort - those who were born after September 1989 - has a higher level of literacy (3.6%) and informational literacy (4.1%). Since they obtained schooling using Latin alphabet, they are probably more capable of writing and reading in it using Uzbek language than the non-exposed generation. These skills are one of the most important criteria in assessing literacy level. Similarly, informational literacy hugely depends on the ability to read in Latin alphabet, which is the main alphabet for any governmental

communication - including publishing laws and political updates- and for advertisement on the streets. Therefore, it can be expected that treated cohort also exhibits higher level of informational literacy while being taught to use Latin alphabet in the basic education.

Moreover, higher level of education and the ability to read in Russian language increases literacy levels. Interestingly, the self allocation of people to a young generation has a strong positive impact on the level of literacy. This is in line with the fact that young generation being more exposed to the Latin alphabet and also more skilled with fast processing of new information.

VARIABLES	(1) Informational Literacy	(2) Literacy
treated	0.0413**	0.0362**
	(0.0168)	(0.0166)
urban	0.0189	0.0269
	(0.0781)	(0.0730)
gender	-0.00223	-0.00385
	(0.0378)	(0.0369)
young generation	0.808***	0.587***
	(0.216)	(0.219)
education	0.115***	0.0933***
	(0.0272)	(0.0228)
employed	0.0357	0.0247
	(0.0488)	(0.0457)
marital status	-0.00757*	-0.00586
	(0.00386)	(0.00391)
nationality	0.0148	0.0119
	(0.00903)	(0.0102)
mothertongue being Uzbek	0.00200	-8.00e-05
	(0.00579)	(0.00639)
good knowledge of Russian	0.0127	0.0194
	(0.0193)	(0.0195)
ability to read in Russian	0.0857**	0.0623*
	(0.0348)	(0.0352)
hhsize	-0.00584	-0.00405
	(0.00986)	(0.00936)
wealth level	0.00104	-0.00194
	(0.0186)	(0.0173)
Constant	-0.608**	-0.491*
	(0.243)	(0.262)
FE age and district	Yes	Yes
Observations	8,622	8,622
R-squared	0.254	0.227

Standard errors clustered at district level in parentheses

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

Table 3.2 Regression results for the literacy and informational literacy outcomes

Subsequently, I evaluate whether there was any effect on educational level obtained by different cohorts (refer to Equations 3.2 and 3.3). For this purpose I use Mlogit to assess the impact of being treated (being taught in Latin alphabet) on the probability of getting less than secondary, general secondary, technical secondary and tertiary education. Figure 3.6 shows the results following the application of Equation 3.2. It demonstrates that there is a higher probability for the treated cohort to have less than secondary education, but much lower probability of tertiary education. The main reason could be that Latin alphabet restricts an access to the scientific and classic literature beyond the scope of textbooks and school program, while this is essential for university studies and academic career.
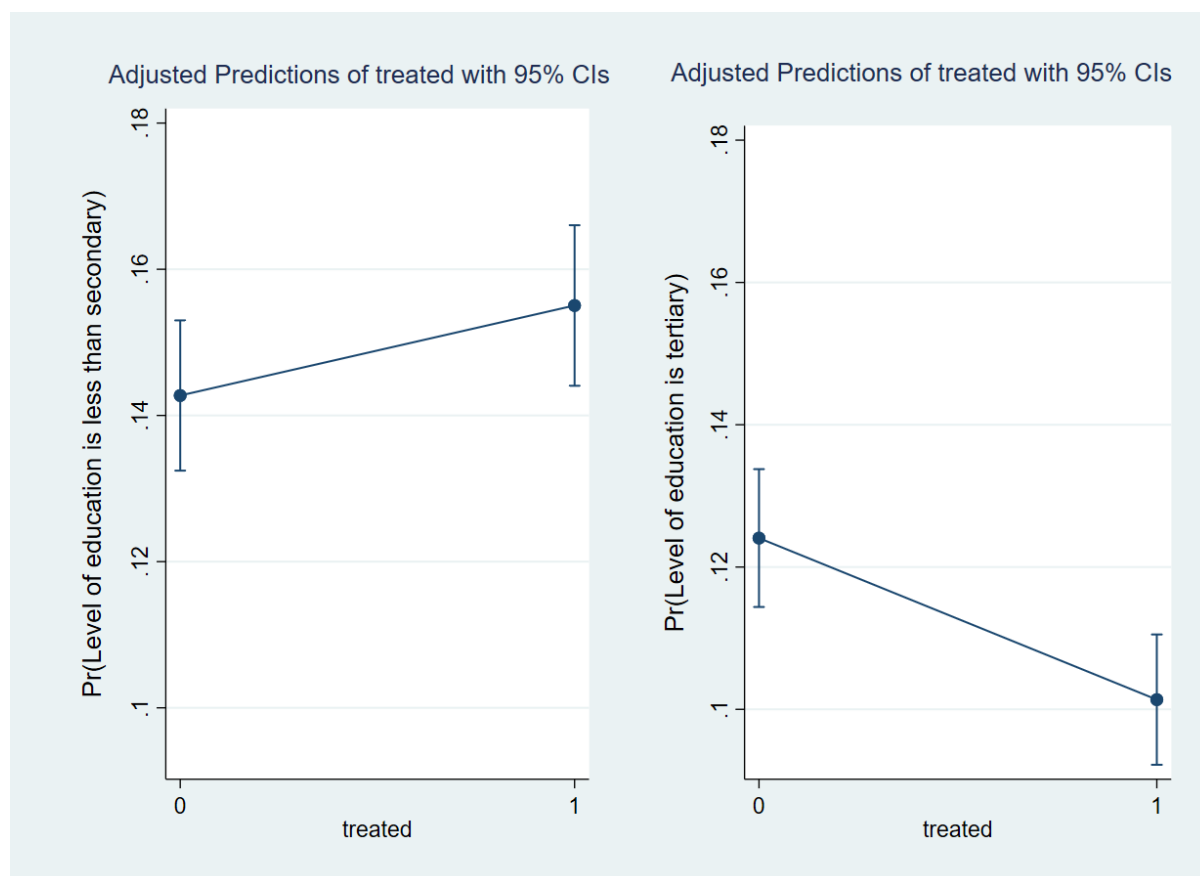

Figure 3.6 Mlogit results for less than secondary and tertiary educational levels

Figure 3.7 depicts the results for general and technical secondary levels. The picture is quite similar to the situation in Figure 3.6. Technical program becomes more appealing for those who want to start working instead of attending the university. This effect may partially be driven by the fact that academic studies require access to the materials beyond the course of the respective program, thus, vastly available in Cyrillic. Moreover, fluent usage of Latin alphabet in writing and reading makes it possible to start earlier public career or get access to the secretary level jobs. Latin alphabet is required for business correspondence, governmental acts and related documentation, which makes treated cohort attractive for local office holders. Typically, those people do not want to give away their positions and instead of learning themselves, they hire secretaries for paperwork activities.
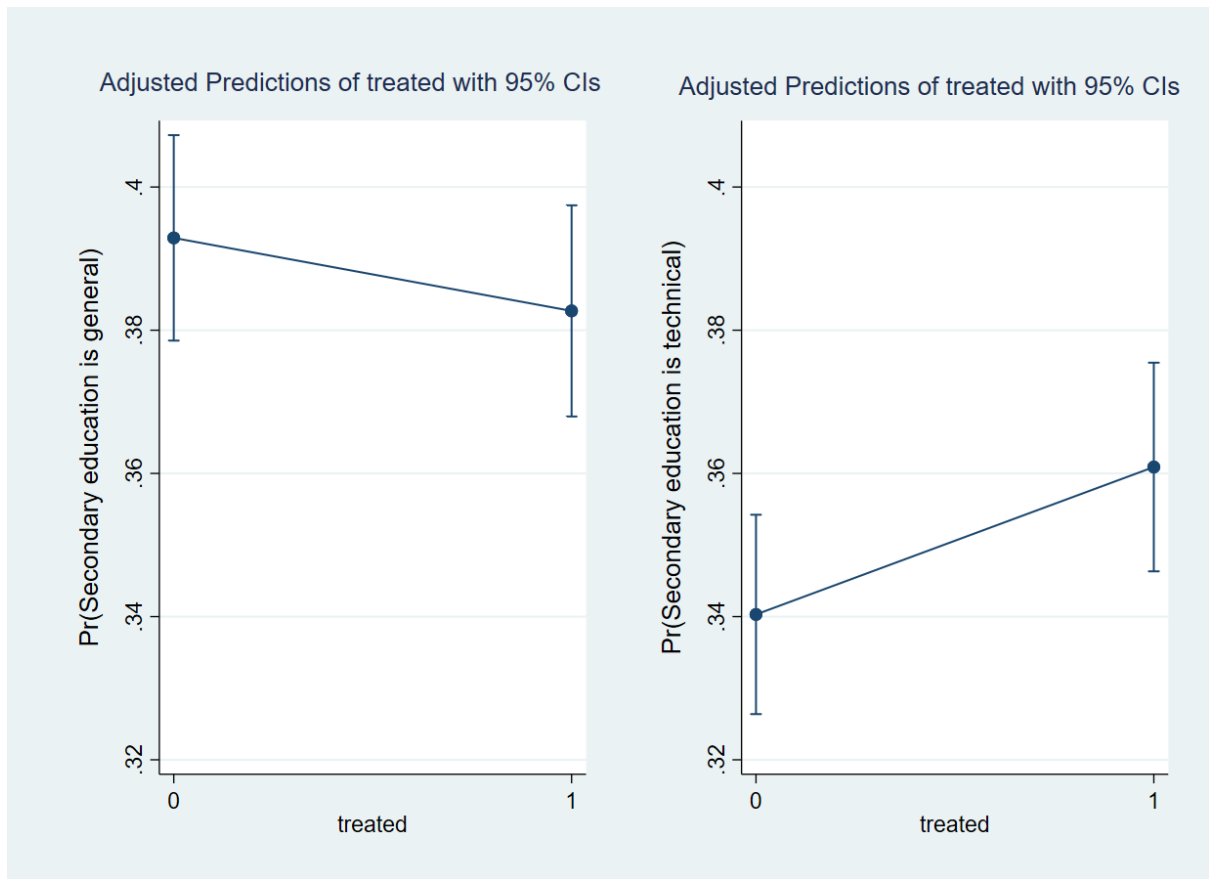
Figure 3.7 Mlogit results for general and technical secondary educational levels

Regarding the further job choice, Figure 3.8 follows the results for Equation 3.3. In particular, it exhibits the comparison of probabilities to choose occupation in state sector or family business or being engaged into entrepreneurship. The observed trend is interesting to explore. It can be seen that the probability of being self-employed decreased by nearly 2% for treated cohorts and involvement into family business by roughly 4%. This trend is supported by significant increase in the probability to get state jobs (by around 7%). It could be explained by the fact that Latin alphabet is required for government jobs, so those positions become more attractive for treated cohorts.
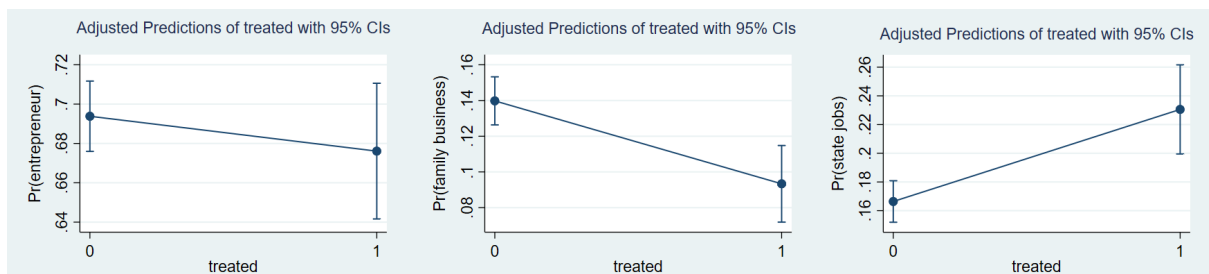
Figure 3.8 Mlogit results for different job types

3.7 Conclusion

Alphabet reforms always had a strong political meaning in Central Asian region, whether it was escaping from Arabic legacy or an attempt to unite Turkic nations within the USSR. The political motivation was even more powerful after most countries got independence in 1991. In Uzbekistan the alphabet reform of 1993 was interconnected with the foreign political course of moving away from the Soviet past. Since the reform was not well tailored for the needs of the language and society, it became a bone of contention between young and old generations.

This paper focuses on the change of the alphabet as a novel approach to strengthen national identity of the population. In particular, it examined the effects of exposure to the Latin alphabet beyond the intended political role. I evaluate how people affected by the switch to Latin alphabet differ in terms of average levels of literacy and informational literacy, educational and occupational choices. The differentiation is based on the quarter of birth since only those who turned 7 years old by the end of September 1996 were allowed to enter the school. By comparing existing trends between those cohorts, results have shown that the treated cohort exhibits higher levels of literacy and informational literacy (see Table 3.2). Moreover, the children who got their complete secondary education in Latin alphabet tend to choose state administration for the career, while the choice of entrepreneurship, which is so common for Uzbekistan, significantly decreased. One of the reasons could be the existing gap between generations in the usage of Latin script. The indifference to switching from Cyrillic among ruling elite created entry level positions for young workers skilled in Latin alphabet. Another interesting trend that is observed in this study concerns the general level of tertiary education of the affected cohort, which ended up being lower. Also getting technical secondary education became popular within people exposed to Latin alphabet. After technical colleges young people can start working without getting a degree level education. Therefore, the main reason for those patterns could be young specialists' expectations of the demand for their skills in state administration and their willingness to start the career as earlier as possible.

To conclude, the findings of this paper emphasize that language reform may have far-stretching consequences in socio-economic life of the country than initially planned national identity tool. In particular, it may affect educational and occupational choices of people. Moreover, the exposure to Latin script equipped young generation with better skills by increasing their informational literacy. Having high level of informational literacy is important not only for successful development in the fast changing technological world, but also for the creation of well-functioning society.

Potential extension of this analysis could include Uzbek communities living close to the border of Uzbekistan, but in neighbouring countries, which did not have the language reform and still use Cyrillic alphabet. Regression discontinuity design could be a worthy approach to apply in this case.

3.8 Bibliography

Akerlof, GA, Kranton, RE, (2000). Economics and identity. Quarterly Journal of Economics, 115(3):715–753

Aliprantis, D., (2007). A Note on Why Quarter of Birth is Not a Valid Instrument for Educational Attainment. MPRA Paper 5168.

Alimdjanov, B., (2019). Uzbekistan: Why Uzbek Language Has Not Become a Language of Politics and Science? [Online] Available from: https://cabar.asia/en/uzbekistan-why-uzbek-language-has-not-become-a-language-of-politics-and-science

Allworth, Edward A. (1990). The Modern Uzbeks: From the Fourteenth Century to the Present: A Cultural History. Hoover Institution Press. pp. 229–230.

Angrist, Joshua D., and Guido W. Imbens. 1995. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. Journal of the American Statistical Association 90: 431–42.

Angrist, Joshua D., and Joern-Steffen Pischke. 2015. Mastering 'Metrics: The Path from Cause to Effect, Princeton. Princeton and Woodstock: Princeton University Press.

Assouad, L., (2020). Charismatic Leaders and Nation Building. Paris School of Economics Working Paper 2020-38. [Online] Available from: https://halshs.archives-ouvertes.fr/halshs-02873520v2/document

Bandiera, O., Mohnen, M., Rasul, I. (2018). Nation-building Through Compulsory Schooling during the Age of Mass Migration. [Online] Available from: https://doi.org/https://doi.org/10.1111/ecoj.12624

Bannon, A, Miguel E, Posner D, (2004). Sources of ethnic identification in Africa. Afrobarometer working paper no. 44

Bartholomä, Ruth. "9. The Construction of the Tatar Nation in the Debate About the Introduction of Latin Script in the Republic of Tatarstan". Language Change in Central Asia, edited by Elise S. Ahn and Juldyz Smagulova, Berlin, Boston: De Gruyter Mouton, 2016, pp. 169-196

Bazzi, Samuel and Hilmy, Masyhur and Marx, Benjamin, Islam, and the State: Religious Education in the Age of Mass Schooling (May 2020). NBER Working Paper No. w27073, Available at SSRN: https://ssrn.com/abstract=3593989

Bazzi, Samuel, Gaduh, Arya and Rothenberg, Alexander and Wong, Maisy, Unity in Diversity? How Intergroup Contact Can Foster Nation Building (March 2019). NBER Working Paper No. w25683, Available at SSRN: https://ssrn.com/abstract=3360829

Blanc, G., Kubo, M., (2021). Schools, Language, and Nations: Evidence From a Natural Experiment in France. mimeo

Beyond Transition, (2006). Beyond Transition. [Online] Available from: http://documents1.worldbank.org/curated/en/961331468016849996/pdf/396430Pls0Repl1r1December2006optmzd.pdf

Biavaschi, C., Giulietti, C., Siddique, Z. (2013). The Economic Payoff of Name Americanization. Journal of Labor Economics. [Online] Available from: https://www.journals.uchicago.edu/doi/abs/10.1086/692531?journalCode=jole

Bisin A, Patacchini E, Verdier T, Zenou Y, (2006). 'Bend it like Beckam': identity, socialization and assimilation. CEPR discussion paper no. 5662

Bisin A, Patacchini E, Verdier T, Zenou Y, (2008). Are Muslim immigrants different in terms of cultural integration? Journal of European Economic Association, 6(2–3):445–456

Bisin A, Verdier T, (2000). The economic of cultural transmission and the dynamics of preferences. Journal of Economic Theory, 97(2):298–319

Bracons, O., Font, J., Figueras, I., Massela, P. (2008). Compulsory Language Education Policies and Identity Formation. Journal of The European Economic Association. Retrieved 20 September 2021, [Online] Available from: https://www.jstor.org/stable/40282653

CABAR: Central Asian Bureau for Analytical Reporting, (2021). Boosted Transition of Uzbek Language to Latin alphabet: Consequences for the Population. [Online] Available from: https://cabar.asia/en/boosted-transition-of-uzbek-language-to-latin-alphabet-consequences-for-the-population

Cantoni, D., Chen, Y., Yang, D., Yuchtman, N., Zhang, Y. (2017). Curriculum and Ideology. Journal Of Political Economy, [Online] Available from: https://doi.org/https://doi.org/10.1086/690951

Caselli F, Coleman W, (2006). On the theory of ethnic conflict. NBER working paper no. W12125

Center for Economic Research, (2013). Analytic report 2013/03: Urbanization in Central Asia.
Chen, M Keith, (2013), "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." American Economic Review, 103 (2): 690-731.

Chen Y, Li SX, (2009). Group identity and social preferences. American Economic Review, 99(1):431–457

Chen, S, Geluykens, R, and Choi, C, (2006), The importance of Language in Global Teams: A linguistic perspective, Management International Review, 46 (6), pp 679-695.

Clots-Figueras, I., Masella, P., (2013). Education, Language and Identity, The Economic Journal, Volume 123, Issue 570, Pages F332–F357

Constant A, Gataullina, L., Zimmermann, KF., (2008) Ethnosizing immigrants. Journal of Economic Behavior and Organization, 69(3):274–287

Constant A, Kahanec, M., Zimmermann, KF., (2009) Attitudes towards immigrants, other integration barriers, and their veracity. International Journal of Manpower (Emerald Group Publishing). 30(1/2):5–14

Deaton, Angus. 2009. Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development. NBER Working Paper Series; No. 14690. Cambridge: NBER.

Deaton, Angus, and Nancy Cartwright. 2018. Understanding and Misunderstanding Randomized Controlled Trials. Social Science Medicine 210: 2–21.

Georgiadis A, Manning A (2009) One nation under a groove? Multiculturalism in Britain. CEPR discussion paper no. 994

Fouka, V. (2016). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. Retrieved 17 September 2021, [Online] Available from: https://siepr.stanford.edu/sites/default/files/publications/591wp_0_6.pdf

Hoogerheide, Lennart, and Herman K. van Dijk, (2006). A Reconsideration of the Angrist-Krueger Analysis on Returns to Education. Econometric Institute Report EI 2006-15. Rotterdam: Econometric Institute.

Jalilov, K., (2021). "Dispute alphabet and reconciliation alphabet". Gazeta.uz. [Online] Available from: https://www.gazeta.uz/ru/2021/03/26/alphabet/

Jennings, M. (2007). "A Very Real War": Popular Participation in Development in Tanzania During the 1950s 1960s. The International Journal of African Historical Studies, 40(1), 71–95. http://www.jstor.org/stable/40034791

Jovanović, S.M., (2018). Assertive discourse and folk linguistics: Serbian nationalist discourse about the cyrillic script in the 21st century. Lang Policy 17, 611–631

Juraev, A., (2016). Standard Uzbek Language, mimeo

Lupyan, G., (2016). How Language Programs the Mind. Topics in Cognitive Science. Volume 8, Issue 2

Manning A, Roy S, (2009). Culture clash or culture club? National identity in Britain. Economic Journal. 120(542):F72–F100

Masella, P., (2013). National Identity and Ethnic diversity. Journal of Population Economics, 26:437–454

Moore, P., Lorenzo, F., Casal, S., (2014). The Effects of Content and Language Integrated Learning in European Education: Key Findings from the Andalusian Bilingual Sections Evaluation Project. Applied Linguistics 31(3):418-442

Otamuratov, S., (2016). Questions to the process of latinization. [Online] Available from: https://www.gazeta.uz/ru/2016/10/24/latin/?auth_finish=1#!

Patricia, S. (2012). Understanding the Interconnectedness between Language Choices, Cultural Identity Construction and School Practices in the Life of a Latina Educator1. Gist Education and Learning Research Journal. Retrieved 17 September 2021, [Online] Available from: https://files.eric.ed.gov/fulltext/EJ1062590.pdf

Pekesen, B., (2019). Pan-Turkism. European History Online. [Online] Available from: http://ieg-ego.eu/en/threads/transnational-movements-and-organisations/international-organisations-and-congresses/pan-ideologies/berna-pekesen-pan-turkism

Sayidov, A., Umansky, Y. (1998). The Factor of Polyethnicity in Uzbekistan, Security Challenges, Human Rights and Development Potential. Medzinárodné Otázky, 7(3), 50–71. [Online] Available from: http://www.jstor.org/stable/44961046

Schweitzer, M., (2020). "Alphabet Transition in Uzbekistan: Political Implications and Influences on Uzbek Identity". CAP Paper No.231 [Online] Available from: https://www.centralasiaprogram.org/wp-content/uploads/2020/05/CAP-Paper-231.pdf

Shustov, A., (2016). Shrinking Russian World. Independent Newspaper [Online] Available from: https://www.ng.ru/dipkurer/2016-02-01/11_asia.html

Stephens, Melvin, Jr., and Dou-Yan Yang. 2014. Compulsory Education and the Benefits of Schooling. American Economic Review 104: 1777–92.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. Journal of Business Economic Statistics 20: 518–29.

The Open Asia, (2017). Experience of Uzbekistan: 6 facts about transition to the latin alphabet. [Online] Available from: https://theopenasia.net/ru/post/opyt-uzbekistana-6-faktov-o-perekhode-strany-na-latinitsu

The State Committee of the Republic of Uzbekistan on Statistics, (2020). "With which countries does Uzbekistan have the highest foreign trade turnover in 2020?", Press service. [Online] Available from: https://stat.uz/en/press-center/news-of-committee/7603-2020-yilda-o-zbekistonning-qaysi-davlatlar-bilan-tashqi-savdo-aylanmasi-yuqori-bo-ldi-3

Uzman, M. (2010). Romanisation in Uzbekistan Past and Present. Journal of the Royal Asiatic Society, 20(1), 49–60. [Online] Available from: http://www.jstor.org/stable/27756123

Williams, R. (1983). Keywords: A Vocabulary of Culture and Society. Rev. Ed., p. 170

World Bank Data Bank, (2020). Data by Country. [Online] Available from: https://data.worldbank.org/

3.9   Appendix

Appendix A shows the questionnaire on literacy (8) and informational literacy (13-18) translated to English.

8. Technical skills: reading and writing

[Read instructions aloud:]

These questions are general questions about daily activities over the past 12 months. Not all the questions will apply to you, but I have to ask them all. When I speak of reading and writing that can be on paper or on computers.

1	**Do you read anything, including very short notes or instructions that are only a few sentences long?**	1. Yes 2. No *[>> 4]*	
2	**Thinking of the past 12 months, as a normal part of your life or work, have you read:** 1. Yes 2. No	a. E-mails	
		b. Forms	
		c. Bills or financial statements	
		d. Instruction manuals/ operating manuals	
		e. Newspapers, magazines, or books	
		f. Other, Specify:	
3	**Thinking about all the things you have read over the past 12 months, what is the size of the longest document that you have read?**	1. One page or less 2. 2 to 5 pages 3. 6 to 10 pages 4. 11 to 25 pages 5. More than 25 pages 99=D/K	
4	**Over the past 12 months, have you filled out bills or forms (such as government forms, application forms, etc)?**	1. Yes 2. No	
5	**Thinking of the past 12 months, have you written anything (else) including very short lists, notes or instructions that are only a few sentences long?**	1. Yes 2. No *[>> 7]*	
6	**Thinking about all the things you normally write, what is the size of the longest document that you wrote in the past 12 months?**	1. One page or less 2. 2 to 5 pages 3. 6 to 10 pages 4. 11 to 25 pages 5. More than 25 pages 99. Don't know	
7	**Thinking of your activities over the past 12 months, have you done any of the following:** 1. Yes 2. No	a. Measure or estimate sizes, weights, distances, etc.	
		b. Calculate prices or costs	
		c. Use or calculate fractions, decimals or percentages	
		d. Perform any other multiplication or division	
		e. Use more advanced math, such as algebra, geometry, trigonometry, etc.	
		f. Other math, specify:	
8	**Has a lack of reading and writing skills in [OFFICIAL LANGUAGE] ever kept you from getting a job, a promotion, or a pay rise, or held you back from advancing your business/own account activity?**	1. Yes 2. No 3. Never applied for a job 99. Don't know	

13 Cognitive Skills: Language

Question 5. 1. Which of these words is not closely related to the word building?
1) building
2) building
3) Scaffold
4) House

Question 5. 2. What does the expression is rare toes?
1) has a congenital disability
2) spending without measure
3) a very generous man
4) saves

Question 5. 3. Which is the opposite of shallow?
1) surface
2) smooth
3) thick
4) deep

Question 5. 4. Which of these words is similar in meaning to the word run?
1) go
2) go
3) run
4) go

Question 5. 5. Which is the opposite of strong?
1) Low
2) less
3) Wholesale
4) Less

Question 5. 6. Which of these words is similar in meaning to the word joy?
1) close
2) happy
3) Links
4) youtube

Question 5. 7. That sentence, which it has the closest respect to the following sentence: Child that the teacher grabbed his hand, smiling happily.
1) The child took his hand teacher.
2) The child and the teacher took each other's hands.
3) The teacher took the child by the hand.
4) The child handed his teacher's hand.

14. Cognitive Skills: Text Comprehension A

Table 9.6: Text Comprehension A

		Write selected choice 77 = No answer
1	**What decision did the old carpenter take?** 1) To send his family abroad 2) To start a business selling smaller and cheaper houses 3) To retire and spend time with his family 4) To go on a vacation in Europe to get some rest	
2	**What did his boss ask when the carpenter told about his decision?** 1) To leave work faster so unemployed builders can have jobs 2) To stay a little longer to make one last house as a personal favour 3) To stay a little longer to transmit his experience to younger carpenters 4) To remain as an advisor on material quality control	
3	**How did the old carpenter work as he was reluctantly accepting the request of his boss?** 1) He worked as usual and performed his tasks with professionalism 2) He worked intensely, finished on time and the work was high quality 3) He used excessively expensive materials and added luxurious details to the home 4) He worked halfheartedly and carelessly	
4	**What did the boss say when the old carpenter finished the work?** 1) He told the carpenter the house was for him because he is the best employee 2) He thanked him for his work and gave him a salary raise 3) He reprimanded the old carpenter for the ugly house 4) He said to the old carpenter the house was the best he ever saw	
5	**How did the old carpenter react to the boss's decision when the house was finished?** 1) He did not believe it because nothing is truly free 2) He blamed himself for not having worked harder and making the house better had he known 3) He said he could not accept a gift that tarnishes his reputation as a carpenter 4) He refused the gift and finally retired	

Story Card : The house in which you might live

After many years of hard work, an old carpenter decides it's time for him to retire, so he can spend more time with his family. When he informs his boss about his decision, his boss asks as a personal favour: to stay a little longer and build one last house before leaving.

The old carpenter agreed reluctantly. As he started the work on the new house, it was obvious his heart was not into the work anymore. He worked slowly, cutting corners, and would sit and long for hours instead of doing the work. It was obvious the house was not elaborate and the materials he chose were low quality. Even from afar one could see the measurements were eyeballed and the cuts were not straight at all.

The house was quite ugly.

The day the old carpenter finally finished building the house, the boss came to see the house. Upon opening the door of the new home, the boss turns to the old carpenter, hands him the key and says:

- "This house is for you. It is my gift to my best employee!"

The carpenter could not believe his ears. It never occurred to him this new house he was building would become his house. He was a little resentful and blamed himself for not working harder. Had he known before, he would have put much more effort into building the house. He could have made it much better and much more beautiful than this…

Story Question Card

1. What decision did the old carpenter take?
1) To send his family abroad
2) To start a business selling smaller and cheaper houses
3) To retire and spend time with his family
4) To go on a vacation in Europe to get some rest

2. What did his boss ask when the carpenter told about his decision?
1) To leave work faster so unemployed builders can have jobs
2) To stay a little longer to make one last house as a personal favour
3) To stay a little longer to transmit his experience to younger carpenters
4) To remain as an advisor on material quality control

3. How did the old carpenter work as he was reluctantly accepting the request of his boss?
1) He worked as usual and performed his tasks with professionalism
2) He worked intensely, finished on time and the work was high quality
3) He used excessively expensive materials and added luxurious details to the home
4) He worked halfheartedly and carelessly

4. What did the boss say when the old carpenter finished the work?
1) He told the carpenter the house was for him because he is the best employee
2) He thanked him for his work and gave him a salary raise
3) He reprimanded the old carpenter for the ugly house
4) He said to the old carpenter the house was the best he ever saw

5. How did the old carpenter react to the boss's decision when the house was finished?
1) He did not believe it because nothing is truly free
2) He blamed himself for not having worked harder and not making the house better
3) He said he could not accept a gift that tarnishes his reputation as a carpenter
4) He refused the gift and finally retired

15. Cognitive Skills: Text Comprehension B

Procedure

    1- *Hand over the medicine instruction card*

    2- *Read aloud the following instructions:*

This is the label of a bottle of pills. Read it carefully. You will then be asked some questions about the instruction label. While answering the questions, you will be able to view the instruction label. Let me know when you are ready to answer the questions.

    1- *Wait for the respondent to finish reading the text*

    2- *Read aloud the questions and answer choices, and record the selected answer:*

Table 9.7: Text Comprehension B

		Write selected choice **99 = No answer**
1	**What is the maximum number of tablets that can be taken by adults in one day?** 1) 8 2) 12 3) 4 4) Not specified	
2	**For how long can the tablets be used by children over 6 years of age before consulting a doctor?** 1) Indefinitely 2) 3 days 3) The tablets should only be used with a doctor's prescription 4) Not specified	

Medicine Instructions Card

PARATSETAM TABLETS (Paracetamol)

Dosage:

Adults – Take 2 tablets every 4 hours as needed. Do not take at intervals shorter than four hours and do not exceed 8 tablets in a 24 hour period.

Children (6 to 12 years) – Take 1/2 to 1 tablet every 4 hours as needed. Do not use more than 4 tablets in 24 hours. Do not use for more than three days without consulting a doctor. Do not give to children under 6 years of age, except with a doctor's prescription. If symptoms persist, consult your doctor.

Do not exceed the indicated dose. Exceeding the recommended dose can damage the liver. If you accidentally take more than the recommended dose, seek medical attention immediately even if you feel well.

Questions:

1. What is the maximum number of tablets that can be taken by adults in one day?
1) 8
2) 12
3) 4
4) Not specified

2. For how long can the tablets be used by children over 6 years of age before consulting a doctor?
1) Indefinitely
2) 3 days
3) The tablets should only be used with a doctor's prescription
4) Not specified

16. Cognitive Skills: Table Comprehension

Procedure

1- *Hand over the bus schedule card*

2- *Read aloud the following instructions:*

This is the schedule of buses departing from the central bus station in Sofia. Look at it carefully. I will ask you some questions about the schedule. While answering the questions, you will be able to see the schedule. Let me know when you are ready to answer the questions.

3- *Wait for the respondent to finish reading the table*

4- *Read aloud the questions and answer choices, and record the selected answer:*

Table 9.8: Table Comprehension

		Write selected choice 99 = No answer
1	**Which one of the following cities has a bus departing for it between 14:00 and 15:15 hours?** 1) Andijan 2) Gulistan 3) Ferghana 4) Samarkand	
2	**At what time does the last bus leave for Varna?** 1) 09:30 2) 20:10 3) 22:30 4) 19:15	

Bus Schedule Card

Bus schedule departing from Sofia Central Bus Station:

City	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
							Departure time									
Andijan				11:20		13:20										
Ferghana				11:00	12:00	13:30		15:20	16:50		18:40		20:10		22:00	
Jizak		09:30					14:30		16:50			19:15				
Bukhara				11:50			14:20			17:00		19:50			22:30	
Margilan			10:15				14:15					19:20		21:50		
Syrdarya		09:20							16:45							
Karshi				11:20										21:30		
Samarkand	08:30					13:50								21:10		
Gulistan				11:10		13:10		15:10		17:10		19:10				

Questions:

1. Which of the following cities have a bus departing for them between 14:00 and 15:15 hours?
   1) Andijan
   2) Gulistan
   3) Ferghana
   4) Samarkand

1) 09:30
2) 20:10
3) 22:30
4) 19:15

2. At what time does the last bus leave for Varna?

17. Cognitive Skills: Publicity Comprehension

Procedure

1- Hand over the water publicity card

2- Read aloud the following instructions:

This is a picture of the brochure for the sale of bottled water. Examine it carefully. I will ask you some questions about it. While answering the questions, you will be able to see the publicity. Let me know when you are ready to answer the questions.

3- Wait for the respondent to finish studying the publicity.

4- Read aloud the questions and answer choices, and record the selected answer:

Table 9.9: Publicity Comprehension

		Write selected choice 99 = No answer
1	What is the price for a bottle of the package during the promotion? 1) 250 sums 2) 300 sums 3) 350 sums 4) 375 sums	
2	How much would have cost three packages before the promotion? 1) 1130 sums 2) 4750 sums 3) 6250 sums 4) 6750 sums	
3	By how many cents is the price of the package reduced during the promotion? 1) 250 sums 2) 400 sums 3) 450 sums 4) 350 sums	

Publicity Card

ON THE CARD: REDUCED PRICE 1800 sums; REGULAR PRICE 2250 sums

What is the price for a bottle of the package during the promotion?

1) 250 sums
2) 300 sums
3) 350 sums
4) 375 sums

2. How much would have cost three packages before the promotion?

1) 1130 sums
2) 4750 sums
3) 6250 sums
4) 6750 sums

3. By how many cents is the price of the package reduced during the promotion?

5) 250 sums
6) 400 sums
7) 450 sums
8) 350 sums

18. Cognitive Skills: Graph Comprehension

Procedure

5- *Hand over the population of Bulgaria card*

6- *Read aloud the following instructions:*

This is a graph showing the population of Bulgaria according to official censuses conducted between the 1900s and today. Examine the graph carefully. I will ask you some questions about it. While answering the questions, you will be able to see the graph. Let me know when you are ready to answer the questions.

7- *Wait for the respondent to finish getting familiar with the population graph.*

8- *Read aloud the questions and answer choices, and record the selected answer:*

Table 9.10: Graph Comprehension

		Write selected choice 99 = No answer
1	**According to the graph, in which year was the population of Bulgaria the smallest?** 1) 1985 2) 2011 3) 1900 4) 1920	
2	**According to the graph, what was the population count of Bulgaria on 01.12.1965?** 1) 7 613 709 2) 5 478 741 3) 8 487 317 4) 8 227 866	
3	**According to the graph, the population in Bulgaria:** 1) Continuously decreases 2) Increases to a point and then decreases 3) Decreases to a point and then increases 4) Steadily increased	

Population of Bulgaria Card

Population of Bulgaria, 1900-2011
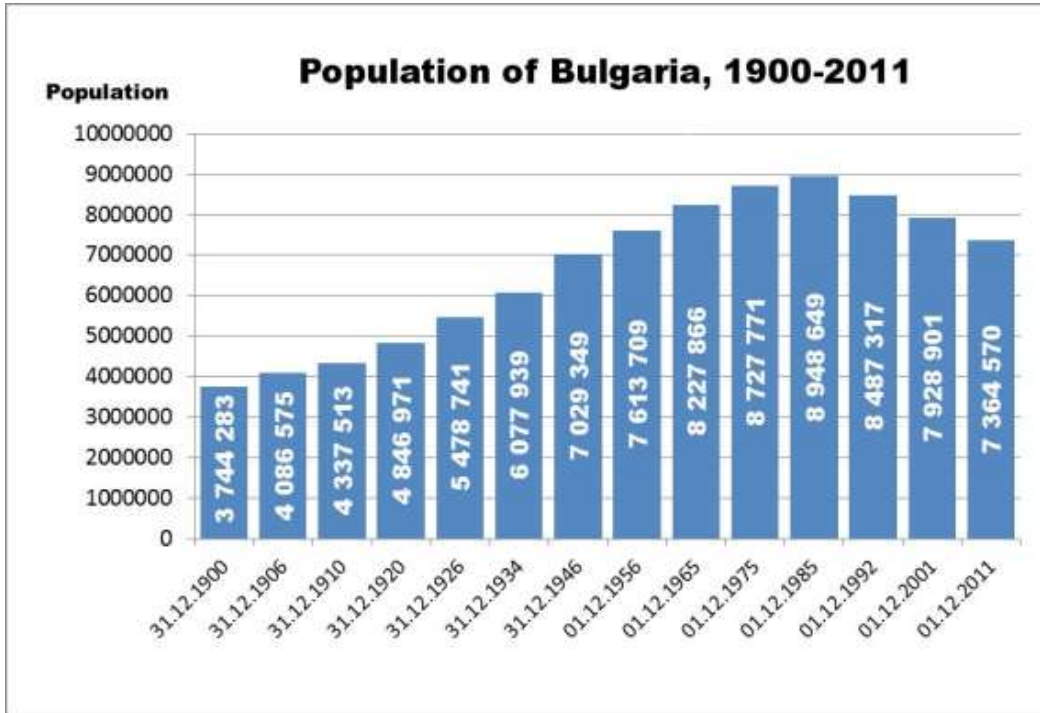
1. According to the graph, in which year was the population of Bulgaria the smallest?
    5) 1985
    6) 2011
    7) 1900
    8) 1920

2. According to the graph, what was the population count of Bulgaria on 01/12/1965?
    5) 7 613 709
    6) 5 478 741
    7) 8 487 317
    8) 8 227 866

3. According to the graph, the population in Bulgaria:
    5) Continuously decreases
    6) Increases to a point and then decreases
    7) Decreases to a point and then increases
    8) Steadily increased