Center for Research in the Economics of Development | CRED

DEADLY ANCHOR: GENDER BIAS UNDER RUSSIAN COLONIZATION OF KAZAKHSTAN, 1898-1908

GANI ALDASHEV AND CATHERINE GUIRKINGER

FACULTÉS UNIVERSITAIRES NOTRE-DAME DE LA PAIX · NAMUR ·

UNIVERSITY OF NAMUR

WP 1111

DEPARTMENT OF ECONOMICS

WORKING PAPERS SERIES

Deadly Anchor: Gender Bias under Russian Colonization of Kazakhstan, 1898-1908*

Gani Aldashev†　　　　*Catherine Guirkinger*‡

July 29, 2011

Abstract

We study the impact of a large-scale economic crisis on gender equality, using historical data from Kazakhstan in the late 19th – early 20th century. We focus on sex ratios (number of women per man) in Kazakh nomadic population between 1898 and 1908, in the midst of large-scale Russian in-migration into Kazakhstan that caused a sharp exogenous increase in land pressure. The resulting severe economic crisis made the nomadic organization of the Kazakh economy unsustainable and forced most Kazakh households into sedentary agriculture. Using a large novel dataset constructed from Russian colonial expedition materials, we document a low and worsening sex ratio (in particular, among poor households) between 1898 and 1908. The theoretical hypothesis that garners most support is that of excess female mortality in poorer households (especially among adults), driven by gender discrimination within households under the increasing pressure for scarce food resources.

Keywords: sex ratio, excess female mortality, nomadic economy, Kazakhstan.

JEL Classification Codes: J16, O15, I10, Q56.

---

* We thank Zhuldyzbek Abylkhozhin, Jean-Marie Baland, Rakhym Beknazarov, Maristella Botticini, Stephan Klasen, Umit Otegenova, Jean-Philippe Platteau, Rohini Somanathan, and participants at the CRED workshop for useful suggestions, and the Slavonic Library of the University of Helsinki for providing the original Russian colonial expedition publications. Elena Shubina provided excellent research assistance. This research is partly financed by UAP-PAI Grant P06/7 (Belgian Science Policy).

† Corresponding author. Department of Economics and CRED, University of Namur (FUNDP), and ECARES, Université libre de Belgique. Email: gani.aldashev@fundp.ac.be.

‡ Department of Economics and CRED, University of Namur (FUNDP), Belgium. Email: catherine.guirkinger@fundp.ac.be.

1. Introduction

Despite enormous economic growth of per capita income over the last two centuries and great legislative progress towards promoting gender equality, discrimination against women remains pervasive throughout the world. In recognition of this fact, improving the well-being of women constitutes one of the Millennium Development Goals. Beside ethical considerations, the expectation is that a greater level of growth would result from enabling women to fully exploit their capabilities. Moreover, a host of development outcomes (e.g. child health, education, fertility decisions) are crucially linked to the welfare of mothers. Thus, two key questions for development economists are: Does economic growth entail gender equality? Can economic crises jeopardize the progress made towards this objective?

The basic measures of gender inequality are gender bias in mortality rates (excess female mortality) and the resulting number of “missing women”, i.e. the difference between the actual number of women in the population and the hypothetical number of women that would exist under gender-unbiased birth rates and access to vital resources. In economics, the pioneering work by Sen (1990), inspired by his analysis of Indian society, estimated the number of missing women worldwide as being roughly 100 million. Later work (Coale, 1991, Klasen and Wink, 2002) improved on Sen’s methodology and corrected the estimates as being around 60 to 90 million. More recently, Anderson and Ray (2010) examined proximate causes of this phenomenon, by decomposing the number of missing women by age and cause of death. They found that most missing women in India and China were among adults and that as a fraction of total female population, the number of missing women was highest in sub-Saharan Africa. Moreover, the authors argue that a comparable fraction of female population was missing in the United States in the early 20th century.

Looking at the mechanisms of gender discrimination, the first explanation advanced in economics is the so-called lifeboat argument (Stiglitz, 1976; see also Chapter 8 in Ray, 1998). This argument states that a household might find it optimal to concentrate a disproportional amount of its resources on a subset of its members, as the concentration of resources increases their return, which may be necessary for the future survival of all household members. This can result in women having less access to vital resources than their male counterparts. This would be the case if, for example, the capacity curve (linking the work capacity of an individual to his/her vital resources, i.e. food or income) were convex in its lower part. This argument is compatible with a unitary-household model. More recent theoretical explanations (see Section 3.2 of Bergstrom, 1997, for a detailed review) focus on bargaining models of intra-household resource allocation. In these models, a lower bargaining power of women as compared to that of men (for example, because of less favorable outside options), is associated with access to a disproportionately low share of household resources. At low levels of income, the unequal access to vital resources is more likely to harm women's health and to lead to excess female mortality. In both classes of models, a gender-neutral increase in household income would lead to a relatively larger increase in women's welfare (as compared to that of men).

However, empirically the correlation between women's welfare and household wealth is less clear. The studies of the relationship between resource scarcity and gender bias belong to two broad categories. The first group of papers concentrates on the comparative health outcomes of female vis-à-vis male children. Chapter 4 in Dreze and Sen (1989) discusses numerous descriptive studies finding that during economic hardship, poor households in less developed Asian countries give priority to male over female children for nutritional resources. Detailed econometric studies using Indian data (Behrman and Deolalikar, 1990; Rose, 1999) find that price increases (adversely) affect more the nutrition of girls as compared to that of boys and that positive weather shocks increase the ratio of the probability of girls' survival over that of boys' survival. Schultz (1985), DeTray (1988), and Alderman and Gertler (1997) find that investment into health and education of female children increases more than that of male children when household income rises. Bhalotra (2010) establishes that adverse aggregate income shocks in India result in substantial increase in (distress) labor supply of mothers in poor households, which translates into a large increase in infant mortality of girls (that of boys remains unaffected). Baird et al. (2011) find similar results using a large micro-level dataset for 59 developing countries. Contrarily, using data from Indonesia, Levine and Ames (2003) find that girls did not fare worse than boys during the economic crisis of 1997-1998. Similarly, Gertler et al. (2004) find, using Indonesian data from 1994-96, that loss of a parent (of either gender) does not affect (negatively) girls more than it affects boys.

While the majority of findings in this first group of studies indicate that economic hardship disproportionally affects female children, one cannot conclude that during economic crises *adult* women are affected more severely than their male counterparts. In fact, the second group of studies (mainly by economic historians) consistently finds that during famines (even those not related to violent military actions), men are more likely to die than women (see studies in Dyson and O Grada, 2002). The main reason seems to be physiological: women are more resistant to starvation than men. This suggests that that more important is the literal starvation during a famine, more likely it is that the gender imbalance in excess mortality is biased against men (Mokyr and O Grada, 2002). On the other hand, in non-famine periods, the picture seems to be reversed. For instance, using genealogical data from Germany in 1740-1860, Klasen (1998) finds that women die in greater numbers than men in months of the year that are associated with highest overall mortality and the most severe scarcity of vital resources.

To the best of our knowledge, no studies try to analyze the effect of economic hardship on women of *different* age groups in the *same* population. The main difficulty is related to data availability. Whereas the studies in the first group rely on datasets that have wealth of information concerning children, they rarely have sufficient information on adults. Contrarily, most of the historical studies have no or very few observations for young children. This makes it difficult to construct a complete picture of the facts regarding gender discrimination and excess female mortality, which, in turn, hampers the attempts to provide a valid theory that can explain both the cross-country and time-series facts.

The contribution of this paper is threefold. First, we provide an attempt to fill the gap noted above, by studying the effect of a long-run economic crisis on gender bias in different age groups, in the context of Kazakhstan under Russian Empire between 1898 and 1908. We do this by exploiting a unique dataset that we have constructed from the records of the Russian Imperial statistical expeditions in Kazakhstan, conducted in two waves (1896-1903 and 1906-1915), which we supplement with the data from the All-Russian Imperial Census of 1897. Using this dataset, we study sex ratios in the Kazakh population in the period when large-scale Russian peasant in-migration into Kazakhstan caused a sharp increase in land pressure and provoked a severe economic crisis among the nomadic Kazakh population. This crisis made the nomadic organization of the Kazakh economy unsustainable, and rapidly forced most Kazakh households into sedentary agriculture. Our main finding is that adult women were affected by the crisis more severely than female children. We document a low and worsening sex ratio (in particular, among poorer households) between 1898 and 1908, with most of the decline occurring in the group aged over 14 years old. Next, we consider several theoretical hypotheses to explain these patterns. The hypothesis that garners most support in our data and descriptive historical sources is that of differential mortality (biased against women) in poorer households, caused by gender discrimination in access to vital resources.

Second, we contribute to the debate on the cross-sectional analyses of gender bias and wealth. Generally, there has been found no evidence of a monotonic relationship between wealth and gender bias in mortality. Sen (1990) states that, comparing across Indian regions, worse sex ratios are found in more wealthy Indian states. This leads to a hypothesis that the relationship is U-shaped, i.e. that the gender bias is highest at the intermediate ranges of wealth distribution. Contrarily, using Kazakh data, we find a monotonic relationship: gender bias is worst at the lower end of the wealth distribution and is consistently better for higher-wealth households.

Third, we contribute to expanding the geographic scope of studies that look at gender bias and its economic determinants. Until now, most studies were focused on Eastern Asia (India, China, and Indonesia) and Western Europe. However, we believe that the geographic and temporal extent of the set of facts to be explained by a theory of gender bias should be much wider than it is now. Otherwise, there is a risk of developing explanations around some cultural factor(s) specific to a particular region of the world. In this concern, our study is important for two reasons: (i) it covers a part of the world for which currently there is very little data; and (ii) it analyzes a society that traditionally was based on nomadic pastoralism – a social structure that substantially differs from the sedentary cultures of Eastern Asia or Western Europe.

The rest of the paper is organized as follows. In Section 2, we describe the historical and institutional context from which our data and estimates come. This should help the reader to understand better the empirical results presented in Section 3. Section 4 analyzes alternative theoretical hypotheses in the light of our statistical findings. Section 5 discusses the broader implications of our findings and suggests avenues for future work.

2. Historical Context

2.1. Organization of Kazakh economy and households before Russian in-migration

Before the massive Russian in-migration in the late 19th – early 20th centuries, the economic organization of Kazakh society was mainly determined by the climatic and geographic characteristics of the land area that Kazakh tribes populated. Archeological research shows that until around 1500-1000 BC, the population of current-day Kazakhstan conducted mainly sedentary agriculture; however, starting from 1500 BC (and definitely by 1000 BC), the tribes that switched to nomadic pastoralism became dominant entities (*Kazakh Economy*, 1979: 33-34; Abuseitova et al., 2001: 22-23). The archeologists hypothesize that long-run natural desertification processes led to the formation of large areas in Central and Northern Kazakhstan that are now arid and semi-arid regions. This, in turn, increased the competitive advantage of nomadic pastoralism as compared to sedentary agriculture. The nomadic economy thus formed remained basically unchanged in its key characteristics (in particular, seasonal transhumance during the year) until the last third of the 19th century, when the large-scale Russian peasant migration into Kazakhstan started.

The fundamental characteristic of the nomadic Kazakh economy was seasonal transhumance, which consisted in changing physical location of the economic unit four times during the year, i.e. once in each natural season. Livestock (horses, sheep, goats, camels in some areas, and – in later periods – cattle) was both the principal asset and the main production input. The principal economic activity consisted of herding and animal husbandry. Regular back-and-forth moves from summer to winter pastures (with relatively shorter stays on autumn and spring stops) guaranteed the provision of fodder throughout the year. The steppe summer pastures provided abundant and high quality fodder during the warmer months but became inhabitable during harsh winters (with temperatures often falling below –35°C, accompanied by strong winds). Thus, during winters Kazakh nomads moved to areas with milder temperatures that were also better protected from winds. This implied that the distances between the winter and summer pastures were often large. Taizhanova (1995: 29) and Chermak (1899: 170) report that whereas in Northern Kazakhstan these distances were around

50-70 kilometers, in Central Kazakhstan the nomads often traveled up to 1000 kilometers (one-way) during transhumance. Kazakh nomads thus rationally adapted to the geography and the climate of the area, by weighing relative benefits and costs of transhumance. On the benefit side, the scarcity of good winter pastures (i.e. areas close to rivers, lakes, and hills) implied that traveling long distances in summer allowed a nomadic tribe to preserve the fodder of its winter pasture. On the cost side, the relatively flat landscape in most of the Central and Western Kazakhstan made long-distance transhumance easier. Figure **1a** shows the main transhumance routes on the territory of Kazakhstan (the tip and the start of the arrow indicate summer and winter pastures, respectively). Figure **1b** shows the positions of different seasonal pastures along a typical transhumance route in Central Kazakhstan.

Given the harsh climatic conditions and the lack of diversification in production, the nomadic economy was extremely fragile to external shocks (e.g. large variations in temperature, disease outbreaks among livestock). Tolybekov (1971: 541-542) reports that during the harsh winter of 1879-1880, in Irghiz and Turgay uezds (sub-regions) the loss of livestock corresponded approximately to 59 per cent of total livestock held by the nomads. Such shocks occurred regularly: the winters of 1850-51, 1855-56, 1879-80 and 1891-92 were those during which large-scale losses of livestock in Turgay oblast were reported (Tolybekov 1971: 542). Similarly, for Western Kazakhstan, Larin (1928) reports that in 45 years preceding his study (i.e. in the period 1882-1927), massive livestock loss caused by poor climatic conditions were registered in 7 winters (the so-called *jut* years). When such shocks hit the nomadic population and were local, Kazakh families had to count on the help of kinsmen that were geographically distant.

As Kazakh historians argue (Taizhanova 1995: 10-11), cooperation networks were organized on the basis of kin; thus, the notion of kin is central for understanding economic relationships among Kazakh nomads. The winter stops were organized around extended families (the so-called *aul-q'stau*), which typically consisted of several nuclear households (usually, closely related by kin) living together during winter. Each household (virtually all households were monogamous nuclear families) consisted of a married couple and their young children. Summer pastures, instead, were organized on the basis of larger kin organizations (the so-called *jazgy aul*, which broadly corresponds to communes), made of several extended families (again, mostly related by kin).

Property rights on land were defined both at the extended family and at the commune level. Winter stops were closed-access common property resources of extended families, whereas summer stops were common property resources of communes. These were also generally closed-access, but the access was less strictly enforced, given the relative abundance of summer pastures. Individual households had no property rights on land but had private property rights on livestock.

Women in Kazakh nomadic families supported a heavy workload, taking part both in herding activities and in the management of the household. Numerous historical sources state that women's economic role was extremely important, despite their relatively low social status (as compared to that of men). For instance, in a detailed analysis of customary law in the nomadic Kazakh society, Makoveckii (1886) writes:

> *"While severely limited, from the point of view of customary law, in terms of her proprietary and social rights, a Kazakh woman commands nevertheless an important role. The fact that her life is restricted to the boundary of aul [i.e. nomadic village] implies that all of the domestic economy and property lies in her hands. Whereas a Kazakh man spends most of the year on the horseback, in continuous moves, taking care of social affairs of the kin, volost [i.e. district], and starshinstvo [i.e. Russian administrative village], his wife remains the real head of the household and manages all of it, thus reducing her husband to the role of the nominal head" (p. 31).*

Zeland (1885) in his ethnographic study of Kazakhs is more cautious about the domestic leadership of a woman, but he also acknowledges that Kazakh women played a crucial economic role in the household:

> *"The status of men and women among Kazakhs is far from being equal. Clearly, the conditions of the nomadic life are not such that a wife is obliged to stay inside the house or hide her face, as among other Muslim people; nevertheless, she plays the role of the husband's servant… However, one cannot say that there is maltreatment of women, [because] men need women as workers" (p. 28).*

2.2. Russian in-migration: its causes, size, and consequences for Kazakh economy

The pre-1917 Russian migration into Kazakhstan started in the 17th century and continued until the October Revolution. It developed in two large waves, each of which had a specific (and temporally different) cause (Demko 1969). The cause of the first wave – the Cossack military migration and creation of cities and fortifications (approximately along the current-day Northern border of Kazakhstan) – was driven by the willingness of Russian Czars to defend the Southern Russian territories from incursion of nomadic tribes. This wave started with the construction of the city of Ural'sk in 1613 and ended approximately around 1850s with the completion of the so-called Defense Line, consisting of a chain of military fortifications from the Caspian Sea to Altai Mountains at the Eastern tip of Kazakhstan. Although this wave resulted in expropriation of important land areas from the Kazakh population (Sedelnikov, 1907, notes, for instance, that the Orenburg Cossacks occupied 7.5 million desyatinas, i.e. approximately 8.2 million hectares of land, that belonged to Kazakh tribes), it was relatively small in terms of in-migration of population and did not lead to fundamental structural changes in the Kazakh nomadic economy.

The second wave started in the 1880s and had as its main cause the abolition of serfdom in Russia in 1861 (Galiev et al. 2009: 223; Demko 1969: 52). Subsequently, the landless peasants started to move in large numbers into the European part of Russia, thus creating substantial tensions in and around large cities. The solution that the Czarist administration adopted was the 1889 law which offered these peasants land "for free", in the amount of 15 desyatinas (approximately 16.4 ha) per household, in the Asian part of the Russian Empire (Olcott 1995: 87). Note that from the legal point of view, Kazakhstan was a protectorate of the Russian Empire. In his study of Russian colonization of Kazakhstan, Demko (1969) states:

> *"By 1900, even members of the intelligentsia and influential men in government considered resettlement in the East to be the best solution to the peasant land problem" (p. 57).*

This triggered a large-scale peasant migration from the European part of Russia into Central Asia, with the bulk of this flow moving into Western, Northern, and – later – Central and South-Eastern Kazakhstan. According to Russian historian Ivan Popov, "[Russian] peasants ran from their beggarly allotments, famines, hunger, and social disorder" (cited by Demko 1969: 55).

Table 1 illustrates the size of Russian in-migration relative to the size of Kazakh population of the four regions in the West and the North of Kazakhstan. The growth of Kazakh population in the period 1897-1916 was relatively low in all the four regions, whereas that of Russians was massive. For instance, in Turgay oblast, the population of Russians increased from 35000 people in 1897 to over 300 000 in 1916. The change was also huge in terms of the fraction of the total population. For example, whereas Russians made about one-third of the total population in Akmolinsk oblast in 1897, by 1916 they were already making almost 60 per cent of the total population.

Russian intelligentsia of the colonial period held to the positive "white-man's-burden" view on the effect of Russian migration on Kazakh population. In their writings, the change in lifestyle and economic organization is described as being fundamentally beneficial for Kazakhs. For instance, Lobysevich (1871) states:

> *"Kirghiz[1] [Kazakh] steppe – given its correct exploitation – is the richest source for the State; however, for this, two conditions are necessary: full guarantee of the well-being of the Kirghiz [Kazakh] people and its Russification… It is absolutely fundamental to introduce [among Kazakhs] the various concepts about sedentary lifestyle, agriculture, and the living conditions of a Russian person… It is advised to require and induce Kirghizs [Kazakhs] to sedentarize" (p. 273-274).*

Some thirty years later, Vladimirskii (1902) writes:

> *"The essence of evolution of Kirghiz [Kazakh] economy lies in the continuous intensification and assimilation to the forms of sedentary lifestyle… Russian colonization … speeds up the natural process of reduction of pastures [of Kazakhs]… It creates [for Kazakhs] new sources of revenue and new occupations, encourages the processes of exchange in the Steppe, transforming its in-kind form into the cash economy" (p. 22-24).*

Despite some positive impact that Russian migration brought to Kazakhstan (agricultural technology transfer for crop cultivation, modernization of education and health facilities), our quantitative findings in Section 3 indicate that the above rosy view ignores the fact that the

[1] In Russian documents before 1917, the native population of Kazakhstan and Kyrgyzstan is denoted with under the same name of "Kirghizs".

reduction of pastures triggered a struggle for survival and conservation of nomadic life among Kazakhs – a struggle that eventually failed. This reduction also had profound negative consequences on the Kazakh society.

2.3. Crisis in Kazakh nomadic economy and the forced sedentarization

The fundamental reason why the second wave of Russian migration caused a crisis of the Kazakh nomadic economy was the increased pressure on land. Russian migrants occupied land that was considered "free" (or unoccupied) by Russian administration – as typical of a sedentary bureaucracy towards the territories of nomads – and this considerably limited the grazing land available for the pastoralist Kazakhs. A substantial fraction of the occupied land was the most scarce winter stops, on which kin-level property rights were carefully regulated among Kazakhs. Moreover, the occupied land often covered the transhumance routes between winter and summer pastures, thus obligating the nomads to change their long-established routes and lengthening (sometimes substantially) the time devoted to transhumance. Figure 2 shows the variation in the territory covered by Russian peasant settlements. From these figures, one sees clearly how the peasant settlements progressed from North towards South in barely fifteen years.

The detailed account how this crisis evolved is given in the 1907 book by T. Sedelnikov, a Russian political thinker who lived in Kazakhstan in the period of sedentarization. In his book entitled *The fight for land in the Kazakh steppe (Bor'ba za zemlu v kazahskoi stepi*), he describes that given a massive increase in land pressure, the only alternative that Kazakh nomads faced was to switch to sedentary agriculture. He writes:

> *"Reduction in pastures led to increasing death of livestock in winter, and this forced weaker and poorer tribes to re-consider their future: given that the previous form of the economy could not provide their subsistence, they had to look for another one, that better corresponds to the new situation… And now these tribes sedentarize in the north to live there for the entire year …" (p. 23).*

Virtually all the tribes (and households) tried to hold on to the nomadic lifestyle as far as they could. Under increasing pressure on land, this meant that stronger and more numerous tribes tried first to occupy the land of weaker ones. The nomads of the weaker tribes thus suffered a double pressure: from Russian peasants and from the stronger Kazakh tribes. Thus, these weaker tribes were the first ones to switch to sedentary agriculture.

The Russian colonial administration calculated the amount of land considered as sufficient for Kazakhs. Anything above this bureaucratically determined need was considered "excess land," which could then be confiscated and passed to incoming peasants for agricultural development (Olcott 1995: 87-88). However, the amount of land that Russian land surveyors considered necessary to feed 24 domestic animals (without consideration of soil quality or water access) was clearly insufficient for a Kazakh household's subsistence. As a result, ever increasing quantities of traditionally nomadic pasture and migration land was set aside for peasant settlement. In 1909, the final bulwark against land confiscation was removed, when a new law ruled that: "Previously designated [Kazakh] structures for household needs or temporary shelter do not serve as barriers to seizure." (Martin 2001: 73).

Since, in a nomadic economy, given the natural shocks (especially in winter), the 24 heads of cattle was clearly insufficient for survival as nomads – the only alternative was to adopt sedentary lifestyle. Thus, Martin (2001) notes,

> *"By the end of the nineteenth and beginning of the twentieth centuries, observers noted increased tension between rich and poor over their mutual land claims. Competition over land pitted Kazakh against Kazakh, nomad against semi-nomad or settled Kazakh, in a struggle for survival that was more intense than in any previous era. But these struggles over land rights were waged within a colonial system that provided nomads and former nomads the opportunity to find new ways to ensure their subsistence, even as it changed their lives in fundamental ways." (pp. 65-66)*

The calculation by Olcott (1995: 98) shows that before Russian in-migration, an average Kazakh household needed about 150 heads of livestock which required 150 desyatinas of land under pastoralism and at least 30 desyatinas animals were stalled all winter. Obviously, the comparison with the above numbers shows that sharply increasing land pressure left the Kazakhs with the only option: to convert to sedentary agriculture. We now turn to the evolution of sex ratios among Kazakhs in this period of a deep economic and social crisis.

3. Missing Kazakh women: statistical evidence

3.1. The data

Our main data source is the unique statistical materials of two waves of Russian colonial expeditions (Shcherbina 1903a,b; Khvosortanskij 1912; Khvosortanskij 1914). In order to regulate the peasant migration flows, Russian colonial administration financed a first expedition in summer of 1896. A prominent Russian statistician, F.A. Shcherbina, headed this expedition. It covered 12 uezds (second-largest administrative units, or sub-regions) in 3 oblasts (the largest administrative units, or regions) in Western, Northern, and Central Kazakhstan, and overall took seven years to complete (the last of the 12 uezds was studied in 1903). The outcome of this expedition was very detailed datasets at the level of extended families. *De facto* it was an extremely detailed agricultural census (i.e. virtually all households existing on the territory of these 12 uezds were covered). The main aim of this expedition was to calculate how much land could be expropriated from the Kazakh population if it were converted from nomadic to sedentary way of life.

Despite the conclusion that several millions of hectares could be “freed” as the result of sedentarization, the Czarist administration found this figure still unsatisfactory, and the second wave of expeditions was financed, starting from 1907 (and finishing in 1915). This second wave covered 21 uezds, including the original 12 uezds covered by the first-wave expedition.

Given the political motivation behind these studies, one could question the reliability of the data collected during the expeditions. Fortunately, several sources confirm – using both qualitative and quantitative arguments – the attention devoted by the expedition administration and data collectors to data accuracy and the resulting high quality of the dataset. First of all, two prominent Russian statisticians – Rumyantsev (1910) and Kaufman (1907) – critically assessed the data collected by Shcherbina expedition. The first author stated that the classification of households by livestock wealth was partially incorrect, whereas the second questioned the potential under-declaration of livestock wealth by Kazakhs and pointed out occasional mistakes in the calculation of agricultural land use by Kazakhs. However, both conclude that, overall, the data collected by the expeditions was of very high quality and correctly reflected the socio-economic situation of the area covered by the expeditions. Second, prominent Kazakh historians (e.g. Shahmatov (1964), Tolybekov (1971)) note that the Shcherbina expedition materials are in line with the qualitative evidence on principal socio-economic characteristics of Kazakhstan in the period under study. Finally, Volkova (1982, 1983) conducted a full-fledged quantitative analysis in which she studied the correlation of ten principal variables from the Shcherbina expedition data (at the uezd level) with the same variables coming from administrative records (registered in 1893). She found that the correlation between variables from the two datasets was very high, which confirms quantitatively the high quality of the Shcherbina dataset.

Our secondary data source is the All-Russian National Censuses of Population, conducted in 1897 and 1926 (*First General Census*, 1905; *All-Union Census*, 1928). These censuses cover a larger geographic area, but essentially contain only the demographic information (i.e. all the information on economic behavior and social organization of households is absent).

Statistical materials of the expeditions were published as books in Russia between 1897 and 1916 in several volumes. For these publications, the household level information was aggregated at different levels (extended family, commune, group of households…). These publications now are considered as rare books. We were able to access four volumes (two for the first wave, and two for the second), available at the Slavonic Library of the University of Helsinki, and have the data inserted in spreadsheets by our research assistants.

The data that we use for this paper comes from two North-Western uezds (Aktyubinsk and Kustanay), in particular from the so-called combinatory tables (Tables C in the original publications). In these tables, cumulative numbers are given, at each volost (administrative units below uezd, i.e. district) level, for household units separated according to wealth (measured in livestock wealth) and principal economic activity category. In other words, an observation in the original dataset is an aggregate of households that belong to a given category. In the first expedition a category is defined by the volost to which a household belongs and the number of horses it owns (0, 1, 2-5, 6-10, …, more than 100). In the second expedition, categories are finer as households are also grouped according to their participation in the labor market. The four main labor market categories are: "Households that have members hired out in agriculture", "Households that have members working as craftsmen", "Households that are labor-autarchic" (i.e. neither hiring in nor hiring out), and "Households that hire in agricultural labor".

3.2. Kazakh population and sex ratios by age categories in 1898 and 1908

The two *uezds* under scrutiny exhibit low population growth and highly biased sex ratios both in 1898 and in 1908. Table 2 reports, for the entire Kazakh population in these *uezds* and for three different age categories (over 14, below 12, and below 1 year old): population size, the sex ratio *sensu stricto* (we use the ratio women to men) and the proportion of males, together with the 95% confidence interval.[2]

The total Kazakh population in the two *uezds* increased from 214 690 individuals in 1898 to 228 214 individuals in 1908. This corresponds to an average annual growth rate of 0.6 per cent. Compared to the average growth rate of 1.6 per cent in the European part of the Russian Empire, this figure is very low. Kazakh demographers attribute such a low rate to a combination of high infant mortality rate and a highly biased sex ratio (in favor of men) in the fertile age group (Asylbekov and Zharkenova, 2001: 9).

The sex ratio in the total population is 0.8725 (women per one man) in 1898. This ratio declines further to 0.8573 in 1908.[3] We use the Model Stable Populations Tables constructed by Coale et al. (1983) that uses demographic data from Europe in the late 19th and early 20th centuries. The authors group the countries from which data are available into four areas

[2] We prefer using the proportion of males for statistical analysis, because – contrary to sex ratios sensu stricto - it is symmetrical (a decrease of 10 in the number of women will increase the proportion by the same amount that an increase of 10 in the number of women would decrease it) and it follows a well-behaved distribution. Assuming that the sex of an individual is a random draw from a Bernoulli distribution, the proportion of males (or females) follows a binomial distribution that can be approximated by a normal distribution if the sample size is large enough. Hardy (2002) discusses the problems related to the use of the sex ratios sensu stricto in statistical analysis.

[3] In order to verify whether these low and declining sex ratios are dramatic but geographically concentrated episodes (i.e. in some parts of the area under study) or whether we are looking at large-scale changes occurring everywhere in the Kazakh society, we constructed Figure 3. It reports the sex ratios in 1908 for each volost in the two uezds as a function of the corresponding sex ratios in 1898. All but one volost lie below the 45° line: the sex ratios have declined basically everywhere across the period 1898-1908. Therefore, the overall drop in the sex ratio is relatively evenly geographically distributed.

(West, North, East and South). For each area, the Model Stable Populations Tables provide the age distribution in a stable population for different level of mortality and gross reproduction rates (or population growth).[4] Klasen (1998) discusses the mortality patterns in these tables in the context of high mortality environments and argues that the four regions exhibit excess female mortality, with the problem being generally less acute in the North table. We thus choose the North table to compute our first benchmark sex ratios.[5] As a second benchmark and for robustness checks, we use the East tables as it corresponds to the geographical area closest to Kazakhstan. [6]

To compute benchmark sex ratios, we then need to pin down three parameters: a level of mortality, a gross reproduction rate and a sex ratio at birth. We follow Klasen's (1998) study of Germany for the slightly earlier period and choose a high mortality environment with a life expectancy at birth of 30 years for women (level 5 in the Model Tables). For the gross reproduction rate, we choose a level of 2.5, which corresponds to a growth rate for the population of about 5%. We performed sensitivity analysis and computed sex ratios and the implied numbers of missing women for a very wide range of gross reproduction rate (from 2 – a rate that implies a negative population growth rate - to 4 which is a rate that implies a population growth rate of more than 17%). The corresponding change in the total number of missing women is small, i.e. our analysis is not sensitive to the assumption about the gross reproduction rate. Finally, the choice of an appropriate biological sex ratio at birth is more delicate and has greater consequences on our results. The difficulty is that unbiased sex ratios at birth vary substantially by ethnicity. Data from the United States show that there are around 1.03 male births for one female birth among African American compared to 1.07 for Chinese (Anderson and Ray, 2010). As we have no precise estimate for Kazakh people, we decide to use two different sex ratios at birth for our computations: the first is the median sex ratio at birth for all ethnic groups (1.059 male per female birth or a sex ratio of 0.944) and the second

---

[4] The gross reproduction rate is defined as the average number of daughters that would be born to a woman if she survived at least to the age of 29.

[5] While the North table have the lowest overall level of excess female mortality, this is not true for the youngest age category. We discuss this point when comparing the number of missing women obtained with the two benchmarks.

[6] Data for the North tables stem from Norway, Sweden and Iceland and for the East tables from Germany, Austria, Czechoslovakia and Northern Italy.

is the sex ratio at birth for Chinese (0.935). The later implies very conservative estimates of the number of missing women.

Table 3 reports the benchmark sex ratios obtained for the overall population and for the under-12 years old and above-14 years old categories.[7] A comparison of the proportions of men by age category in the Kazakh data (Table 2) with the benchmark proportions reveal that for all age categories, all benchmark proportions are outside the 95% confidence intervals of the proportion of male in the Kazakh population. Thus, the Kazakh sex ratios are abnormally biased against women. Moreover, the confidence intervals in 1898 and 1908 do not overlap. This implies that the proportion of men significantly increased over the 1898-1908 period. In other words, the already biased sex ratio worsens over this period.

Looking across age categories, in both years, the sex ratios in the Kazakh population are monotonically decreasing in age. In the youngest age group, the sex ratios in 1898 and 1908 are 0.947 and 1.059, respectively. The sex ratios for children under 12 are worse (0.941 and 0.924), whereas the worst ratios are for the population aged 14 and above, with less than 830 women per 1000 men in both years. This worsening of sex ratios over age is in stark contrast with the age profile of benchmark sex ratios presented in Table 3. In fact, Coale (1991) notes that in all European populations since the mid 19th century, male mortality at all age has been greater than female mortality: while "biologically" slightly more boys are born than girls, sex ratios are improving with age.

Based on the benchmark sex ratio, we have computed an estimate of the number of missing women in the Kazakh population. Table 4 reports the number of missing women by age category and year of census, as the number of women that should be added to the population in order to reach the benchmark – holding constant the number of men. The number of missing women depends on the benchmark used, especially in terms of the choice of sex ratio at birth. When we use a conservative estimate of this parameter, the overall percentage of missing women in the female population decreases from 18.9% to 17.6% in 1898 (Model North). The difference resulting from relying on the East instead of the North Model Stable

[7] For the below 12 sex ratio we actually use the below 10 sex ratio readily computable from the table. The below 15 sex ratio, also readily computable is similar (1.058), we are thus confident that the below 12 would be very similar to the below 10.

Population Table is less substantial overall, but more pronounced for the below 12 age group (for 1898, the North and East benchmarks suggest respectively 6% and 9% of missing women). Demographers recognize the relatively high proportion of men among the young age groups in Northern Europe in the 19th century and explain it by the presence of a tuberculosis epidemic that lead to higher mortality rates among girls than among boys (Coale et. al, 1983) . This gender contrast is usually explained by lower levels of nutrition among young girls. As we have no evidence for the presence of this type of epidemic in Kazakhstan over our period, we rely on the East benchmark for the rest of the discussion (and choose the median sex ratio at birth as our reference).

The sheer size of the missing-women phenomenon in the Kazakh population is daunting. The observed sex ratios translate into roughly 18300 missing women in 1898 and 21500 in 1908. Thus the stock of missing women represents about 17% of the total female population in 1898 and 19% in 1908. The break-down by age categories shows again that it is among the adult population that the problem is the most acute. Missing women above 14 represent 24% of the above-14 female population in 1898 and 25% in 1908. Supposing that these figures are the result of excess female mortality (as we argue in Section 4), they suggest that an additional 25% of women aged 14 and above would have been alive if the excess female mortality in Kazakhstan were no greater than in Western European countries at that time.

More insights are gained by examining distributions of sex ratios in the population, thereby exploiting the fact that demographic information is available by household category, where a category is defined by wealth (measured by livestock in adult horse equivalent) owned by the household and the district the household is living in.[8] Those groups have different sizes, with a median of 240 individuals, a minimum of 10 and a maximum of 5105 in 1898. To take this feature into account, we weight the data points proportionally to the size of population in the group when constructing kernel densities of the proportion of men by category. Figures 4 to 11 compare the distribution of the proportion of male in our data to the distribution of hypothetical proportions based on benchmark sex ratios (we use the East benchmarks for both birth sex ratios). These hypothetical proportions are generated by assuming that the number of

---

[8] In 1908, in addition, the grouping is based on the household participation to the labor market in 1908. To generate comparable distributions across year, we aggregate the 1908 data by wealth and districts, so as to have the same structure as in 1898.

men in each category is drawn from a binomial distribution with a mean equal to the benchmark proportion.[9] Figures 4 and 5 present the results for the overall proportion of male in 1898 and in 1908, respectively. Figures 6 and 7 present the distributions for the under-12 years old population and Figures 8 and 9 for the above-14 years old population.

The distributions of the proportions of men in the data in both years and in all age categories are clearly located at the right of the benchmark distributions. The shift of distribution is particularly striking for the above-14 category: there is nearly no common support between the benchmark and the observed distribution, especially in 1908. This confirms the worsening of the situation between the two census years and suggests that compared to western populations of the same period, nearly all wealth / district categories of the Kazakh population exhibit a much larger proportion of men.

For the youngest age group (below 2), we perform a similar analysis and use the benchmarks provided by the East tables for the age category 0 to 1 for two different levels of sex ratios at birth. Figures 10 and 11 report the benchmark distributions for 1898 and 1908 respectively, along with the observed distribution in our data. These figures suggest three facts. First, in both years the proportion of male in the youngest age category is remarkably close to the hypothetical distributions. There is therefore no evidence of excess female mortality in the youngest age category. Second, the widening and flattening of the distribution across years is consistent with a strong drop in fertility or a remarkable gender neutral increase in infant mortality. The flattening is related to the much lower number of individuals in the youngest age category in the second year.[10] Note that the composition of the youngest age groups is different across two years: in 1898 the youngest group consists of children of age 1 and younger, while in 1908 it is restricted to children strictly less than 1 year old. Thus, part of the sharp decrease of the size of the youngest population is an artifact of the change in definition

---

[9] If the gender composition of each group would be the same as the gender composition of the European population from the time used as a benchmark, the number of men in a group of size *X* would follow a binomial distribution *(X, p)* where *p* is the benchmark proportion of men.

[10] In smaller groups, proportion of male are more widely distributed: if, for instance, there are only three individuals in one category, a male proportion of one is far more likely than if there are 30 individuals.

of the youngest age category.[11,12] This dramatic change in population dynamics over one decade confirms that the Kazakh population was undergoing a major crisis. Finally, the comparison of the real to the hypothetical distribution reinforces our confidence in the quality of the data and speaks against a systematic undercounting of women (see also Section 4.1).

4. Missing Kazakh women: competing explanations and mechanisms

What can explain the patterns described above? There are three main candidate explanations: (1) Misreporting or systematic undercounting of women; (2) Net migration biased by gender; (3) Differential mortality.

4.1. Systematic undercounting

The first possibility is a systematic misreporting of women in the surveys conducted by the expedition members. Given that culturally the role of women in the Kazakh society was inferior to that of men, normally the survey respondent would be a senior male member of the household. The strong virilocal and exogamy norms might also imply that the female children in the family are considered as the future members of another extended family. Given this, the respondents might have omitted to mention some of the female children when asked about the number of children by gender. If this hypothesis were correct, our statistics on sex ratios would be biased downwards.

Given the high quality of the dataset (as attested by the sources cited in Section 3.1), it is unlikely that such systematic misreporting took place and was not noted by the data collectors. We found no discussion of such potential data problems anywhere in the introductory sections of the expedition publications, whereas for other variables – e.g.

[11] In a stable population with constant fecundity, the size of the age category 0-1 would be greater than one-half of the size of the age category 0-2 (because of infant mortality). Instead, here the group of children under 1 in 1908 represents less than one quarter of the group aged 1 and below in 1898.

[12] Ideally, we would like to look at the change in the absolute number of women of childbearing age, in order to compute the change in the fecundity rate. Unfortunately, our data is not sufficiently disaggregated by age.

livestock wealth – the expedition administration explicitly mentioned the difficulties and potential mis-measurement problems in several occasions (Volkova 1988: 178-179).

Furthermore, two characteristics of the patterns presented above speak against this hypothesis. First, the age profile of sex ratios is difficult to reconcile with systematic undercounting of women. Indeed, it seems less likely that men would omit to mention adult women rather than young girls. In such case (and under the systematic undercounting of female children), we should observe *better* sex ratios in the group above-14 than in the under-12 age group. However, we observe exactly the opposite: the sex ratios in the above-14 group are worse than in the group under-12. Furthermore, the distributions of sex ratios among infants presented above are remarkably close to theoretical distribution, suggesting that for this category, counting was accurate. Finally, the drop in sex ratios over time is difficult to attribute to misreporting, because this would imply that misreporting worsened over time.

4.2. Massive female out-migration

The second possibility is that there is an important out-migration within the period under study, especially for women. This includes two forms: geographical population displacement that is biased towards women and inter-marriages with Russians (which would imply that young Kazakh women move to live with Russian and quit Kazakh households; thus, they would not be counted in the expedition data). The historical evidence speaks against the first possibility. Contrarily, there was some regional out-migration of young men towards the mining areas of Eastern Kazakhstan, given the labor-intensive technology that was used in the mines (Abuseitova et al. 2001: 416-418). Moreover, the Russian empire censuses of 1897 and 1926 indicate that there were extremely few ethnic Kazakhs living outside the territory of Kazakhstan. Finally, in the neighboring regions (for which we have less detailed information) and, generally, overall in Kazakhstan, the sex ratios are very similar to those in our area under study.

Theoretically, if Russian migration were heavily male-biased, if Russian men married with Kazakh women, and thus many Kazakh women moved to live with Russians, then these women would not appear in our dataset. It is unlikely that such phenomenon explains the low sex ratio that we document. First of all, the Russian migration was principally the migration of families. Demko (1969: 93) notes that 47.4 % of Russian migrants were women. Second, the inter-ethnic marriages were extremely rare in Kazakhstan until the post-WWII period (Carrere d'Encausse, 1959). This can be explained by the huge linguistic, cultural, and religious differences across the two ethnicities. Finally, the analysis of the 1926 census data also helps to discard this hypothesis. The 1926 census provides information about the Russian language skills of inhabitants, as well as their native language. It reports extremely few women of Kazakh origin who were able to speak Russian, suggesting that intermarriages were very rare. We can thus confidently discard this explanation.

4.3. Excess female mortality

This leaves us with the third possibility: women die more frequently than men, more so at adult age, and the differential mortality gets stronger between 1898 and 1908. There are two main potential mechanisms behind differential mortality. The first is biological, i.e. gender-differentiated biological or medical factors. The second is behavioral, i.e. gender-biased resource allocation. There is no historical evidence of gender-biased disease incidence in the period under study and, overall, little support for the first mechanism in the literature. As mentioned above, Coale (1991) notes that in all the European countries from the middle of the 19th century until now, male mortality rates have been higher at every age, conditional on the relatively unbiased access to nutrition and health conditions (and this over the range of life expectancy from 35 to 80 years).

A major cause of mortality for women may have been maternal mortality which, at first sight, appears orthogonal to discrimination in resource allocation. However, discrepancies in levels of mortality of women of child bearing age across population having access to the same medical technology are largely explained by differences in nutrition levels. This argument is developed by Ransel (1991) in his study of infant care in the Russian Empire, where he explores differences in infant mortality and women survival across ethnic groups within Russia over our period of interest. He notes that while infant mortality rates are smaller among Muslim ethnic groups, these rates are negatively correlated to women's mortality in childbearing years (contrarily to non-Muslim groups, where women's mortality rates are positively correlated with infant mortality rates). He argues that this is related to Muslim mothers having to breastfeed their children until the age of two without access to adequate nutrition, which led to depletion of their physical forces and provoked serious health problems, especially after giving birth to several children. Finally, for 18-19th century Germany, Klasen (1998) convincingly shows that despite a high rate of maternal mortality, it can account only for a small portion of "the extraordinary survival disadvantage of women."

The unequal access to resource hypothesis is even more likely when resources were scarce. Next, we turn to the evidence that suggests a very strong correlation between household-level wealth and the sex ratio.

4.4. Wealth and sex ratios

We exploit the two cross-sections of data to highlight the correlation between sex ratios and wealth. A first indicator of wealth is the number of horses that a household owns. Table 5 and Table 6 present the proportion of men (along with the 95% confidence interval) by wealth category, for 1898 and 1908, respectively. The last line of Table 6 provides some evidence about the correlation between our measure of wealth (horse-ownership) and average cash expenditure per person (in each horse-ownership category). The last horse-ownership category (accounting for about 5% of the population) individuals spend 5 times more than those in the first two categories (that account for about 10% of the population).

In both years, overall sex ratios, above-14 sex ratios and below-12 sex ratios increase monotonically with wealth (the only exception is for the below-12 sex ratio in the wealthiest category in 1908). For 1898, it is only for the two wealthiest categories, accounting for less than 4% of the population that the overall proportion of male in the population is not significantly greater than the benchmark ratio of 0.492 (East Table, median birth ratio). In 1908, the situation is worse across wealth levels and it is only for the last category, representing less than 5% of the population that overall proportion of male is not significantly greater than 0.492. Below-12 proportions of male are lower than above-14 but significantly larger than the benchmark of 0.494 except for the three wealthiest category in 1898 (9.8% of the population), and for the poorest in 1908. This last fact is driven by the low size of that category (114 individuals younger than 12), which leads to a very wide confidence interval. The increase in the proportion of men in the below-12 category throughout the wealth spectrum across survey years appears particularly worrying. The picture is similar for the above-14 age category. It is only in the three wealthiest categories in 1898 that the proportion of men above 14 is not significantly greater than the benchmark of 0.491. Given the small total size of the population of below-1 children, the confidence intervals on the proportion of males for this age group are extremely wide.

Looking at the change over the two periods, we observe a worsening of the sex ratios across the board despite the overall increase in wealth (as measured by livestock ownership in adult horse equivalent) and the shift of the wealth distribution to the right. Figure 12 illustrates the distribution of population by wealth for the two survey years and the change in sex ratio. The simultaneous increase in wealth and decrease in sex ratios may appear puzzling at the first sight. However, one should remember that the sex ratio captures the differential survival rates of men and women across their lifetime (and thus depends on the economic conditions over a relatively long time span), while horse ownership reflects the current-year economic situation. When the economic situation improves, it is only after a certain time lag that the population sex ratio adjusts correspondingly.

Moreover, in a nomadic pastoralist economy, year-to-year fluctuations in livestock may be very large, with particularly dramatic consequences for the households at the bottom of the wealth distribution. As Tolybekov (1971) writes:

> *"[During jut] many pastoralists in some one-two months almost entirely lost their wealth. Sometimes even the wealthier households become the middle-class, or – occasionally – the poor families. The less wealthy households of middle-class and poor Kazakhs became destitute. The mass of people, having lost its main production tool and the only source of subsistence – livestock – had to face famine and death." (p. 541).*

Thus, the observed increase in horse ownership does not necessarily reflect an increase in the permanent income. In fact, the winter of 1897-1898 was particularly harsh, i.e. the so-called *jut* year with substantial livestock deaths in winter (Tolybekov 1971: 79).

The detailed data on food consumption collected by the expeditions enables us to dig deeper into the correlation between sex ratios and the resource availability. Kazakhs consumed three broad types of food: meat, milk and grain. For each type of food we have information about the average quantity available in each category of households. For meat, we know the type and number of animals slaughtered over the last 12 months (separately during winter and in other seasons.) For milk, we can estimate milk production based on information about the number of cows, ewes, goats, camels and mares that gave birth over the last 24 months. Finally, for grain, we know the quantity consumed over the last 12 months.[13] Using realistic assumptions about the nutritional value of these food items, we estimate calorie availability per equivalent adult for each group of households (recall that a unit of observation is a group of households living in the same district and belonging to the same wealth category).[14] It is important to note that while it is relatively easy to evaluate the nutritional value of grain, the estimation of the calories available from meat and milk is more complicated, for two reasons. First, the productivity of animals – both in terms of meat and milk production – depends heavily on the animal breed and nutrition. Given that we have imprecise information on animal characteristics, we follow historical and agronomic sources that provide average productivity for well-fed animals. Second, the nutritional values of milk and meat themselves depend on animal breed and nutrition, introducing an additional source of noise in our estimations. Available figures are, again, based on product from well-fed animals. Overall, our estimates of food availability are likely to be biased upwards, especially for poor households, who had less well-fed animals.

On average, we estimate that across the sample in both years about 3600 kcal are available per equivalent adult per day. About one fourth of the population had less than 2600 kcal available per day. Table 7 provides the exact descriptive statistics, broken by the type of food. While the average figures appear high, it hides substantial variation across the sample, as illustrated by Figures 13 and 14 that show the distributions of calorie available per adult equivalent (per day) for the two years of study. Furthermore, energy needs for Kazakhs in 1898-1908 where substantially higher than nowadays, given the harsh climatic conditions that they faced and the important amount of hard physical work that they had to do. Experimental studies of nutritional needs in cold environments reveal that an active adult sleeping in a tent and experiencing outdoor temperature below -25°C need on average 57 kcal per kg body weight per day. A Kazakh adult weighted on average 65 kg at the end of the 19th century (Zeland

[13] Given that there are minor differences across the two study years in terms of the type of information available, we analyze the data from two years separately.

[14] To express the population size of each group in adult equivalent, we use the recommended dietary allowances (RDA) for 1989. The exact weights and details about calorie calculations are available upon request. We do not report them here, to economize on space.

1885), which amount to 3705 kcal per day. This suggests that our estimated average energy available falls below the average daily winter-period need. It is also clear that a substantial proportion of the population had inadequate food availability.

To investigate whether gender bias is correlated with calories available, we estimate the following econometric model (separately for the two waves of data):

$$\Pi_{ai}=\alpha_a \, cal_i+\beta_a' \, X_i+ \varepsilon_{ai} \qquad (1)$$

where Π_{ai} is the proportion of men in the age category a in the group i, cal_i is the number of calories available per adult equivalent in the group i, X_i is a vector of control variables including district fixed effects and, depending on the specifications, wealth category fixed effects, average household size and average area cultivated by households in group i, and ε_{ai} is an error term. We estimate the model separately for the 0-12 age category and for above-14 age category. We use two estimation methods: (i) ordinary least squares, and (ii) generalized linear model (GLM), which takes into account the fractional nature of the dependent variable.[15] Observations are weighted by the size of the population (of given age) in the group.

Table 9 reports the results of our estimations. Rows R1 to R14 report the coefficients on the calorie variable for different models estimated on the 1898 data set, whereas rows R15 to R28 report the results for the 1908 data set. The dependent variable for the estimation reported in rows R1 to R7 and R15 to R21 is the proportion of men in the above-14 age category, while the dependent variable in the other rows is the proportion of men in the below-12 category. Let's look first at the estimations corresponding to the above-14 age category. The results are extremely robust across specifications; the calories available per person are negatively correlated with the proportion of men, whether wealth-category fixed effects are included or not (R1/R2 and R15/R16), whether the 5% of the population with the highest calorie availability is excluded or not (R3 and R18), and whether or not we allow for a non-linear effect of calorie availability (R4 and R19). The statistical significance of the coefficient on the square term suggests a concave relationship between calorie availability and proportion of

[15] We use the strategy proposed by Papke and Wooldridge (1996) to handle proportion models with zeros or ones. Formally, we assume that the expected value of the proportion of men is:
$E(\Pi_{ai}) = G(\alpha_a \, cal_i+\beta_a' \, X_i+ \varepsilon_{ai})$, where G is a logistic function. To estimate the parameters, we use Bernoulli quasi-maximum likelihood estimators recommended by these authors.

men. When we use a GLM estimation method rather than OLS, the results are overall unchanged. The effects of household size and land area available per person differ across the two years. In 1898, only the household size is significant: larger household size is correlated with a lower proportion of adult men. Instead, in 1908, only land area per person is significant: a larger farm size is associated with a lower proportion of males. We come back to these results later in the discussion.

The results are less consistent across specifications for the below 12 age category. In 1898, there is no correlation between food availability and proportion of men, once we exclude the 5% of the population with the highest calorie consumption. This suggests that the correlation is driven by extreme values. Furthermore, the results in rows R11 and R13 suggest that there is no concavity in the relationship. Conversely, in 1908 the relationship appears concave and the correlation is significant when the sample is trimmed from extreme value. We are thus confident that there is a negative relationship between calories available and the proportion of men in the below-12 population in 1908; however, the results for 1898 are somewhat weaker.

To interpret the size of the coefficient, consider the results in row R3 (that concerns 95% of the population, i.e. excluding the 5% with the highest calorie consumption). The coefficient suggests that, holding constant the size of the male population, an increase in 1000 kcal available (per day) per adult equivalent would translate approximately into additional 9000 women, i.e. an increase of 14.6%, relative to the size of female population within a given household category.[16] The same computation for 1908 yields an increase of 16.4% in the size of female population.

So far, the evidence shows that the relative scarcity of resources is highly correlated with sex ratios. Intuitively, when the competition for resources intensifies, the less powerful elements in the society are more likely to be unable to satisfy their basic vital needs. We consider this argument more in detail in the next sub-section.

[16] We take the difference between the number of women in a given age category and the number of women that would be necessary to decrease the proportion of men by 0.034, holding the overall number of men constant.

4.5. Uncovering the mechanism behind differential mortality: competition over scarce resources

In this section, we build a very simple model of food allocation in a nuclear household. Our objective is to construct a basic framework that will help us to interpret the data patterns documented in the previous sections, via simple microeconomic mechanisms.

Consider a nuclear household composed of one parent (father) and two children (son and daughter). The father has access to a production technology (either nomadic or sedentary) and decides on the allocation of the total amount of food resources produced under this technology between his two children. The father is altruistic towards his children but has a moderate intrinsic son preference (e.g. for cultural reasons). His utility function is
$U(c_s, c_d)$, where c_s and c_d denote food consumption (measured in calories) of son and daughter, respectively, and the marginal rate of substitution between the c_s and c_d at $c_s = c_d$ is larger than 1 in absolute value.

An alternative formulation can be that of a selfish father who allocates food among his children to maximize the expected future return from the son (who will stay with him in the next period, given the patrilocal norm) and the daughter (who will get married to a son from another family in the next period, and will thus bring brideprice to her father). Qualitatively, the result of the model under such formulation would look similar to the model we develop here.

The life expectancy of a child depends positively on his/her calories intake. However, this mapping might differ by gender. In other words, denoting with p and q the probability that son and daughter, respectively, survive beyond a given age a, the functions $p_a = p(c_s)$ and $q_a = q(c_d)$ can be different.

Graphically, this decision and the resulting allocation look as depicted in Figure 15. The upper right quadrant represents the budget constraint of the household, $m = c_s + c_d$, and the set of indifference curves of the father. The solution of the maximization problem of the father (under three different levels of household wealth) gives the allocations X, Y, and Z. The lower right quadrant depicts the survival function for the son (once he becomes adult), $p_a =$

$p(c_s)$, whereas the upper left quadrant depicts the survival function of the daughter (once she becomes adult), $q_a = q(c_d)$. The allocations *X*, *Y*, and *Z* result, through the mappings $p(c_s)$ and $q(c_d)$ in the pairs of survival probabilities (p_a, q_a), denoted in the lower left quadrant of Figure 15 with the uppercase letters ***X***, ***Y***, and ***Z***.

Next, we consider the hypothetical equal-calories allocations under the three levels of wealth. In this allocations, the son and the daughter receive the same amount of food (calories). However, given that their survival probability functions can differ (for biological reasons), these allocations will not result in equal probabilities of survival for the two children, once they turn adults. The resulting survival probability pairs (p_a, q_a) are denoted in the lower left quadrant with the lowercase letters ***x***, ***y***, and ***z***.

This reasoning allows to construct the triangles *x0X*, *y0Y*, and *z0Z*. The economic meaning of these triangles is as follows. Consider the trianle *x0X*. Point *x* represents the survival probabilities for the son and the daughter under the unbiased food treatment. Let's denote these probabilities with p_x and q_x, respectively. Point *X* stands for the survival probabilities resulting under the equilibrium choice of the father. Let's denote these probabilities with p_X and q_X, respectively. Let the number of boys and girls born in the population be *B* and *G* and let's assume that our household is representative and population is sufficiently large for the law of large numbers to apply. Then, the sex ratio in the population below age *a*, under unbiased food treatment would be $(G/B)(q_x/ p_x)$, whereas the equilibrium sex ratio is $(G/B)(q_X/ p_X)$. Dividing the latter expression by the former, we obtain a measure of *the gap between the equilibrium and unbiased-treatment sex ratios*, $(q_X/ p_X)/(q_x/ p_x)$. It is easy to show that this measure is monotonically increasing with the angle *x0X*. Therefore, the economic interpretation of the angles *x0X*, *y0Y*, and *z0Z* is that as the resource constraint of the households in the population is relaxed, the gap between the resulting sex ratio and the unbiased-treatment sex ratio shrinks. In other words, the sex ratio improves and the gender bias gets smaller when a larger quantity of vital resources becomes available.

While we refer to a father and his children, the framework can also depict the behavior of a family head allocating resources between male and women dependents of any age, or even of a husband deciding upon his and his wife's consumption. In any case, this simple framework illustrates how, when resources get scarce, they get concentrated on the "preferred" or "more productive” members of the household (those from which the decision maker gets the highest return).

We can now exploit the information about household participation in the labor market. Our results suggests, in line with the simple model sketched above, that women are less numerous precisely in the households were the returns to women labor are likely to be lowest. The 1908 data classify households into four categories:[17]

- Households in which (some) members are hired out as agricultural workers (and no one works as craftsman);
- Households with some members working as craftsmen and no agricultural worker is hired in;
- Household with no member hired out or working as craftsmen and no agricultural worker hired in;
- Households that hire in agricultural workers.

Table 8 reports sex ratios by age and household types. Sex ratios, expenditure per person and area cultivated per person are strongly correlated with household participation in the labor market. The worst situation is that of a household with members hired out as agricultural workers while the best is for those with that employ workers. Households autarchic in terms of agricultural labor (with or without craftsmen) are in an intermediate situation. When we estimate the calorie regressions presented above (Eq. 1) and add controls for the labor market participation, the labor market participation categories have a significant impact on the proportion of men in both age groups: households with members working out exhibit higher level of bias in the proportion of men than those autarchic or hiring in.[18] Importantly, only men participated in the labor market, while many livestock-related tasks (e.g. feeding the animals) were performed by women. This implies that in households where a large part of the revenue is obtained from wage labor, men may contribute relatively more to the total household revenue and the returns from calorie intake may be greater for male household members. This may explain why, controlling for the availability of resources, discrimination

[17] There exist two further categories whose description we were unable to find in the description of the expedition materials. 91% of the population belongs to one of the four categories presented above.

[18] We do not report the results of these regressions here, for space constraints.

against adult women and young girls was more acute in households where men worked for a wage.[19]

There is a potential concurrent factor contributing to the wealth gradient in the sex ratio. Women born in poorer households could have moved into the wealthier ones. This explanation does not work for under-12 population. While there is some anecdotic evidence that wealthier families employed girls from the poorer families for household chores, these girls continued to live with their parents (thus, they were counted as members of the poorer households). It is, however, possible that girls from poorer backgrounds married up. This would reinforce the wealth gradient. Even in the relatively rich categories, the biased sex ratio in the under-12 group implies a shortage of women of the marriage age. The marriage market in the Kazakh society in the period under study was based on brideprices.[20] Thus, the long side of the marriage market (i.e. grooms) was the buyer side. Therefore, the richer grooms could outbid the poorer ones, leading to the upward mobility of girls. This can account for a part of the wealth gradient in the above-14 age group but not in the below-12 group. Furthermore, it cannot, of course, explain the highly biased sex ratio (in this age group) in the overall population.

The framework above suggests that women were actively discriminated against through the concentration of scarce resources on male members of the household. An alternative hypothesis is that women were more numerous in poorer households, and, while within households resources were shared equally, women overall had access to fewer resources than men. Intuitively, if parents continue to have children until a son is born, young girls belong - on average – to larger families than young boys. In such families, there are fewer resources available per child. Even in the absence of discrimination against young girls once they are born, the ex-ante preference for boys may thus explain differential mortality in young children

[19] In these households women may also have had a more limited control of the household budget and may have obtained less in a bargaining game over resources.

[20] The parents of the groom paid the parents of the bride. Secondary sources (Malyshev 1902: 45-50; Makoveckii 1886: 5-6) indicate that traditionally the value of brideprices was substantial even among the poorer strata of the Kazakh society (the lowest amount of brideprice was 7-9 large domestic animals).

(Jensen, 2003).[21] The rough evidence reported in Tables 4 and 5 lends no support for this hypothesis, given that average household size and average number of children increase over the wealth spectrum. Furthermore, in the calorie regressions household size has either no effect or a significant negative effect on the proportion of men in the household.

To summarize, the most plausible hypothesis to explain the very low sex ratios and the wealth bias in the data is that women in Kazakh society during the Russian peasant colonization in the late 19th – early 20th century were actively discriminated against. Kazakh men ate more adequately than women; thus men probably were more resistant to infectious diseases. There is some evidence that such diseases were primary causes of mortality in Kazakhstan. For instance, the malaria outbreaks before 1917 in the Kazakh population led to prevalence rates of 16-47 per cent, with the mortality rates between 10 and 30 per cent among the sick (Sharmanov 1980: 4). Public health resources were extremely poor: in 1913, in Kazakhstan there were barely 7 medical workers per 100 000 people, and the provision of medical services was strongly biased towards the wealthy families living in cities, where such families had access to Russian colonial hospitals (Sharmanov 1980: 4-5). Moreover, some historians (Taizhanova, 1995) report that epidemic outbreaks during our period of study were related to the transition from nomadic pastoralism to sedentary economy (which implied a higher population density and, therefore, a faster spread of contagious diseases).[22]

5. Conclusion

In this paper, we have studied the sex ratios in Kazakh population between 1898 and 1908, i.e. in the period when large-scale Russian in-migration into Kazakhstan caused a sharp increase in land pressure and a severe economic crisis. This crisis made the nomadic organization of the Kazakh economy unsustainable, and forced most Kazakh households into

---

[21] However, Rosenblum (2010) provides empirical evidence against this hypothesis in the context of India. She finds that a higher proportion of girls correlates with more unequal treatment.

[22] A major cause of mortality for women may have been maternal mortality. While it is plausible that poor women are more likely to die giving birth than richer ones because of lower level of nutrition, Klasen (1998) convincingly shows that in the 18-19th century Germany, despite a high rate of maternal mortality, maternal mortality can account only for a small portion of “the extraordinary survival disadvantage of women.”

sedentary agriculture. Using a unique novel dataset constructed from Russian colonial expedition materials, we document a low and worsening sex ratio (in particular, among poor households) between 1898 and 1908. We consider several theoretical hypotheses to explain these data patterns. The hypothesis that gains most support in the data and descriptive historical sources is that of differential mortality (biased against women) in poorer households, based on some form of discrimination (e.g. in calorie intake) by gender within households.

There is no doubt that the massive arrival of Russian and the induced rapid sedentarization of Kazakhs brought much distress among the Kazakh population and that the transition into sedentary agriculture has been traumatic. During the period under study, several harsh winters reinforced the problem, leaving many Kazakhs with few or no animals and severe difficulties in procuring vital resources. Women seem to have been the first victims during this hardship. One fascinating question is whether the bad sex ratios we document are the outcome of a temporary subsistence crisis or whether it is also related to the changed economic role of women in the new system. Clearly, the transition into sedentary agriculture deeply modified the position of women within households. Secondary evidence suggests that the involvement of women in field cultivation was very low. It is beyond the scope of this paper to tackle this question. As we pursue our analysis of historical data, we hope to be able to relate the evolution of key demographic variables such as sex ratios and fecundity to the change from nomadic life to sedentary agriculture and the subsequent change of the role of women in Kazakh society. We plan to compare regions where the Russian in-migration had relatively little influence on the traditional economy to the regions that we study in this paper. We also plan to use the 1926 census data to examine the evolution of sex ratios as the Kazakh population settles into agriculture.

References

Abuseitova, M., Abylkhozhin, Zh., Klyashtornyj, S., Masanov, N., Sultanov, T., and Khazanov, A. 2001. *History of Kazakhstan and Central Asia* (in Russian, *Istoriya Kazahstana i Central'noi Azii*). Almaty, Kazakhstan: Daik-Press.

Alderman, H., Gertler, P. 1997. Family resources and gender differences in human capital investments: The demand for children's medical care in Pakistan. In Haddad, L., Hoddinott, J., Alderman, H. (eds.), *Intrahousehold Resource Allocation in Developing Countries*. Baltimore, MD: The Johns Hopkins University Press.

Anderson, S., Ray, D. 2010. Missing women: Age and disease. *Review of Economic Studies* **77**: 1262–1300.

Asylbekov, M.-A., Zharkenova, A. 2001. *The Population of Kazakhstan in Late 19th – Early 20th Centuries (1897-1917)* (in Russian, *Naselenie Kazahstana v konce XIX – nachale XX vv. (1897-1917 gg.)*). Almaty, Kazakhstan: Orkeniet Press.

Baird, S., Friedman, J., Schady, N. 2011. Aggregate income shocks and infant mortality in the developing world. *Review of Economics and Statistics* **93**: 847-856.

Behrman, J., Deolalikar, A. 1990. The intrahousehold demand for nutrients in rural South India: individual estimates, fixed effects and permanent income. *Journal of Human Resources* **25**: 665-696.

Bergstrom, T. 1997. A survey of the theories of the family. Chapter 2 in Rosenzweig, M., Stark, O. *Handbook of Population and Family Economics*. Elsevier, North-Holland.

Bhalotra, S. 2010. Fatal fluctuations? Cyclicality in infant mortality in India. *Journal of Development Economics* **93**: 7-19.

Carrere d'Encausse, H. 1959. Realities and limits of the Russian presence in Central Asia.

American Behavioral Scientist **3**: 17-23.

Chermak, L. 1899. *Sedentary Agriculturer Kirghizs [Kazakhs] on the Chu River* (in Russian, *Osedlye kirgizy-zemledel'cy na r.Chu*). Omsk, Russia. Available at: http://www.lib.okno.ru/ebook/13100436.pdf

Coale A. 1991. Excess Female Mortality and the Balance of the Sexes in the Population: An Estimate of the Number of "Missing Females". *Population and Development Review* **17**: 517-523.

Coale A., P. Demeny and B. Vaughan. 1983. *Regional model life tables and stable populations.* Academic Press, University of Michigan.

Demko, G. 1969. *The Russian Colonization of Kazakhstan, 1896-1916*. Bloomington, IN: Indiana University Press.

DeTray, D. 1988. Government policy, household behavior, and the distribution of schooling: A case study of Malaysia. *Research in Population Economics* **6**: 303-336.

Dreze, J., Sen, A. 1989. *Hunger and Public Action*. Oxford, UK: Clarendon Press.

Dyson, T., O Grada, C., eds. 2002. *Famine Demography: Perspectives from the Past and the Present*. Oxford, UK: Oxford University Press.

Galiev, V., Mazhitov, S., Shaimerdenova, K., Abdirov, M., Idrisov, R., Volkova, T., and Valikhanov, E. 2009. *History of Colonization of Kazakhstan in 1820-1860* (in Russian, *Istoriya kolonizacii Kazahstana v 20-60-h godah XIX veka*). Almaty, Kazakhstan: Mektep.

Gertler, P., Levine, D., Ames, M. 2004. Schooling and parental death. *Review of Economics and Statistics* **86**: 211-225.

Hardy, I.C.W. 2002. *Sex Ratios: Concepts and Research Methods.* Cambridge, UK: Cambridge University Press.

Jensen, R. 2003. Equal Treatment, Unequal Outcomes? Generating Sex Inequality through Fertility Behavior. IDF working paper 3030.

Kaufman, A. 1907. *On the Problem of Genesis of Russian Land Commune* (in Russian, *K voprosu o proishozhdenii russkoi zemledel'cheskoi obshchiny*). Moscow, Russia.

*Kazakh Economy at the Turn of the Century: XIX-XX (*in Russian, *Kazahskoe hozyaistvo na rubezhe XIX-XX vekov).* Almaty, Kazakhstan: Gylym Press, 1979.

Khvorostanskij, P. 1912. *Materials on Kirghiz [Kazakh] Land Use, Collected and Developed by the Statistical Party of Turgay-Uralsk Migration Region. Aktyubinsk Uezd* (in Russian, *Materialy po kirgizskomu zemlepol'zovaniju, sobrannye i razrabotannye statisticheskoi partiej Turgajsko-Ural'skogo pereselencheskogo rajona. Aktjubinskij uezd*). Orenburg, Russia: Izd.Peresel.Org.

Khvorostanskij, P. 1914. *Materials on Kirghiz [Kazakh] Land Use, Collected and Developed by the Statistical Party of Turgay-Uralsk Migration Region. Kustanay Uezd* (in Russian, *Materialy po kirgizskomu zemlepol'zovaniju, sobrannye i razrabotannye statisticheskoi partiej Turgajsko-Ural'skogo pereselencheskogo rajona. Kustanajskij uezd*). Orenburg, Russia: Izd.Peresel.Org.

Klasen, S. 1998. Marriage, bargaining, and intrahousehold resource allocation: Excess female mortality among adults during early German development, 1740-1860. *Journal of Economic History* **58**: 432-467.

Klasen, S., Wink, C. 2002. A turning point in gender bias in mortality? An update on the number of missing women. *Population and Development Review* **28**: 285–312.

Larin, I. 1928. Donguz-Dau valley and its role as fodder base for the Tabyn tribe in Adai uezd (in Russian, *Urochishche Donguz-Dau i ego znachenie kak kormovoi bazy v zhizni roda tabyn Adaevskogo uezda*). *The Economy of Kazakhstan* (in Russian, *Narodnoe hozyajstvo*

Kazahstana) **1**: 69.

Levine, D., Ames, M. 2003. Gender bias and the Indonesian financial crisis: Were girls hit hardest? CIDER Working Paper No. C03-130.

Lobysevich, F. 1871. Turgay oblast and its organization (in Russian, *Turgayskaya oblast i ee ustrojstvo*), *The Military Digest* (in Russian, *Voennyj sbornik*) **27**: 271-282.

Makoveckii, P. 1886. *Materials for the Study of Legal Custom of Kirghizs [Kazakhs]* (in Russian, *Materialy k izucheniju juridicheskih obychaev kirghiz*). Omsk, Russia. Available at: http://www.lib.okno.ru/ebook/11100405.pdf

Malyshev, N. 1902. *Traditional Family Law of Kirghizs [Kazakhs]* (in Russian, *Obychnoe semeinoe pravo kirghiz*). Yaroslavl, Russia: Tip.Gub.Pravl.

Martin, V. 2001. Nomadic land claims in the colonized Kazakh steppe. *Difference and Integration* **1**: 65-74.

Mokyr, J., O Grada, C. 2002. What do people die of during famines: the Great Irish Famine in comparative perspective. *European Review of Economic History* **6**: 339-363.

Olcott, M.B. 1995. *The Kazakhs*. 2nd ed. Stanford, CA: Hoover Institution Press.

Ray, D. 1998. *Development Economics*. Princeton, NJ: Princeton University Press.

Ransel, D. 1991. Infant care cultures in the Russian Empire. In: B. Evans Clements, Alpern Engel B., Worobec C. (Eds.) *Russia's Women: Accommodation, Resistance, Transformation.* University of California Press.

Rose, E. 1999. Consumption smoothing and excess female mortality in rural India. *Review of Economics and Statistics* **81**: 41-49.

Rosenblum, D. 2010. The Effect of Fertility Decisions on Excess Female Mortality in India. Working Paper.

Rumyantsev, P. 1910. *Kirghiz [Kazakh] People in Past and Present* (in Russian, *Kirgizskij narod v proshlom i nastoyashchem*). St.Petersburg, Russia.

Schultz, T. 1985. School expenditure and enrollments, 1960-1980: The effects of income, prices, and population. In Johnson, D., Lee, R. (eds.), *Population Growth and Economic Development*. Madison, WI: University of Wisconsin Press.

Shcherbina, F. 1903a. *Materials on Kirghiz [Kazakh] Land Use, Collected and Developed by the Expedition for the Analysis of Steppe Regions. Aktyubinsk Uezd* (in Russian, *Materialy po kirgizskomu zemlepol'zovaniju, sobrannye i razrabotannye ekspediciej po issledovaniju stepnyh oblastej. Aktjubinskij uezd*). Voronezh, Russia.

Shcherbina, F. 1903b. *Materials on Kirghiz [Kazakh] Land Use, Collected and Developed by the Expedition for the Analysis of Steppe Regions. Kustanay Uezd* (in Russian, *Materialy po kirgizskomu zemlepol'zovaniju, sobrannye i razrabotannye ekspediciej po issledovaniju stepnyh oblastej. Kustanajskij uezd*). Voronezh, Russia.

Sedelnikov, T. 1907. *Land Conflicts in Kazakh Steppe* (in Russian, *Bor'ba za zemlyu v kazahskoi stepi*). Saint-Petersburg, Russia: Delo.

Sen, A. 1990. More than 100 million women are missing. *The New York Review of Books* **37**: 20.

Shahmatov, V.F. 1964. *Kazakh Nomadic Pastoral Commune: Questions of Its Genesis, Evolution, and Dissolution* (in Russian: *Kazahskaja pastbishchno-kochevaya obshchina: Voprosy obrazovanija, evolucii i razlozhenija*). Almaty, Kazakhstan: Izdatel'stvo Akadmii Nauk Kazahskoi SSR.

Sharmanov, T. 1980. *The Development of Public Health in Kazakhstan* (in Russian, *Razvitie zdravoohranenija v Kazahstane*). Almaty, Kazakhstan: Znanie.

Stiglitz, J. 1976. The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs. *Oxford Economic Papers* **28**: 185-207.

Taizhanova, G. (Ed.). 1995. *The Kazakhs* (in Russian, *Kazahi*). Almaty, Kazakhstan: Kazakhstan Press.

Tolybekov, S.E. 1971. *The Nomadic Society of Kazakhs from the XVII to the Early XX Centuries: A Political-Economic Analysis* (in Russian, *Kochevoe obshchestvo kazahov v XVII – nachale XX veka: Politiko-ekonomicheskii analiz*). Almaty, Kazakhstan: Izdatel'stvo "Nauka" Kazahskoi SSR.

Vladimirskii, V. 1902. On the problem of transition of Kirghizs [Kazakhs] into sedentary lifestyle (in Russian, *K voprosu o perehode kirghiz v osedloe sostoyanie*). *Proceedings of the Western Siberian Section of Imperial Russian Geographic Society, 1898-1901* (in Russian, *Otchety o deyatel'nosti Zapadno-Sibirskogo otdela Imperatorskogo Russkogo Geograficheskogo Obschestva za 1898-1901 gody*): 22-27. Available at: http://www.lib.okno.ru/ebook/14100408.pdf

Volkova T.P. 1982. *The Materials on Kirghiz (Kazakh) Land Use, Collected and Developed by the Expedition for the Analysis of Steppe Oblasts: A Historiographic Analysis* (in Russian, *Materialy to kirghizskomu (kazahskomy) zemlepol'zovaniju, sobrannye i razrabotannye ekspediciej po issledovaniju stepnyh oblastei: Istochnikovedcheskoe issledovanie*). PhD (kandidat) dissertation, Moscow State University.

Volkova, T.P. 1983. Some problems of historiographic analysis of the Materials of the F.A. Shcherbina expedition (in Russian, *Nekotorye problemy istochnikovedcheskogo izuchenija materialov ekspedicii F.A. Shcherbiny*), in B.A. Tulepbayev (ed.), *Questions of Historiography of Kazakhstan* (in Russian, *Voprosy istoriografii Kazahstana*), Almaty, Kazakhstan: Izdatel'stvo "Nauka" Kazahskoi SSR.

Volkova, T. P. 1988. On the analysis and the use of the data from pre-1917 budgetary statistics of Kazakh households (in Russian, *K izucheniju i ispol'zovaniju dannyh dorevoljucionnoi statistiki kazahskih hozyajstv*), in: Alimbaev. N, Volkova, T. (eds.) *Problems of Historiography and Historical Source Analysis of Kazakhstan: Pre-1917 Period* (in Russian, *Voprosy istoriografii i istochnikovedenija Kazahstana: dorevoljucionnyj period*). Almaty, Kazakhstan: Nauka.

Zeland, N. 1885. The Kirghizs (in Russian, *Kirghizy*). *Notes of the Western Siberian Section of Imperial Russian Geographic Society* (in Russian, *Zapiski Zapadno-Sibirskogo otdela Imperatorskogo Russkogo Geograficheskogo Obschestva*) **8**. Available at: http://www.lib.okno.ru/ebook/13100423.pdf.

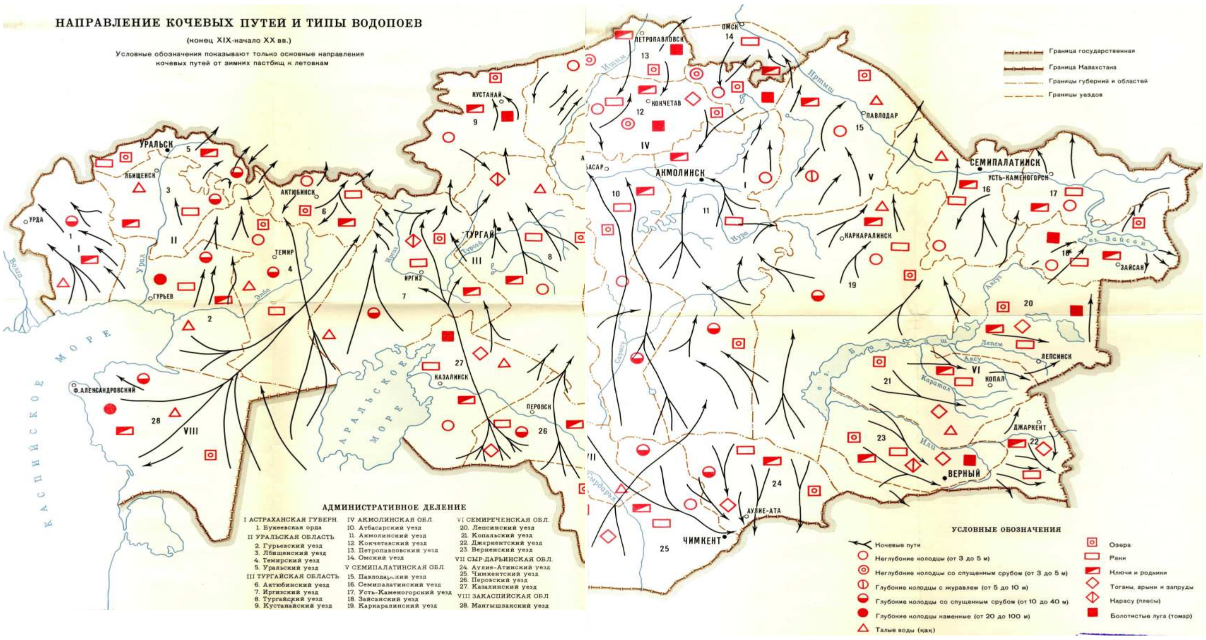

Figure 1a. Transhumance routes in Kazakhstan before sedentarization

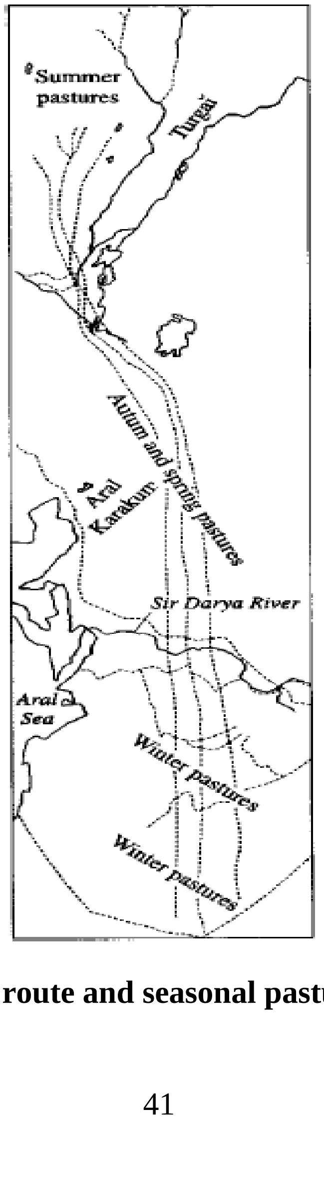

Figure 1b. A transhumance route and seasonal pastures in Central Kazakhstan

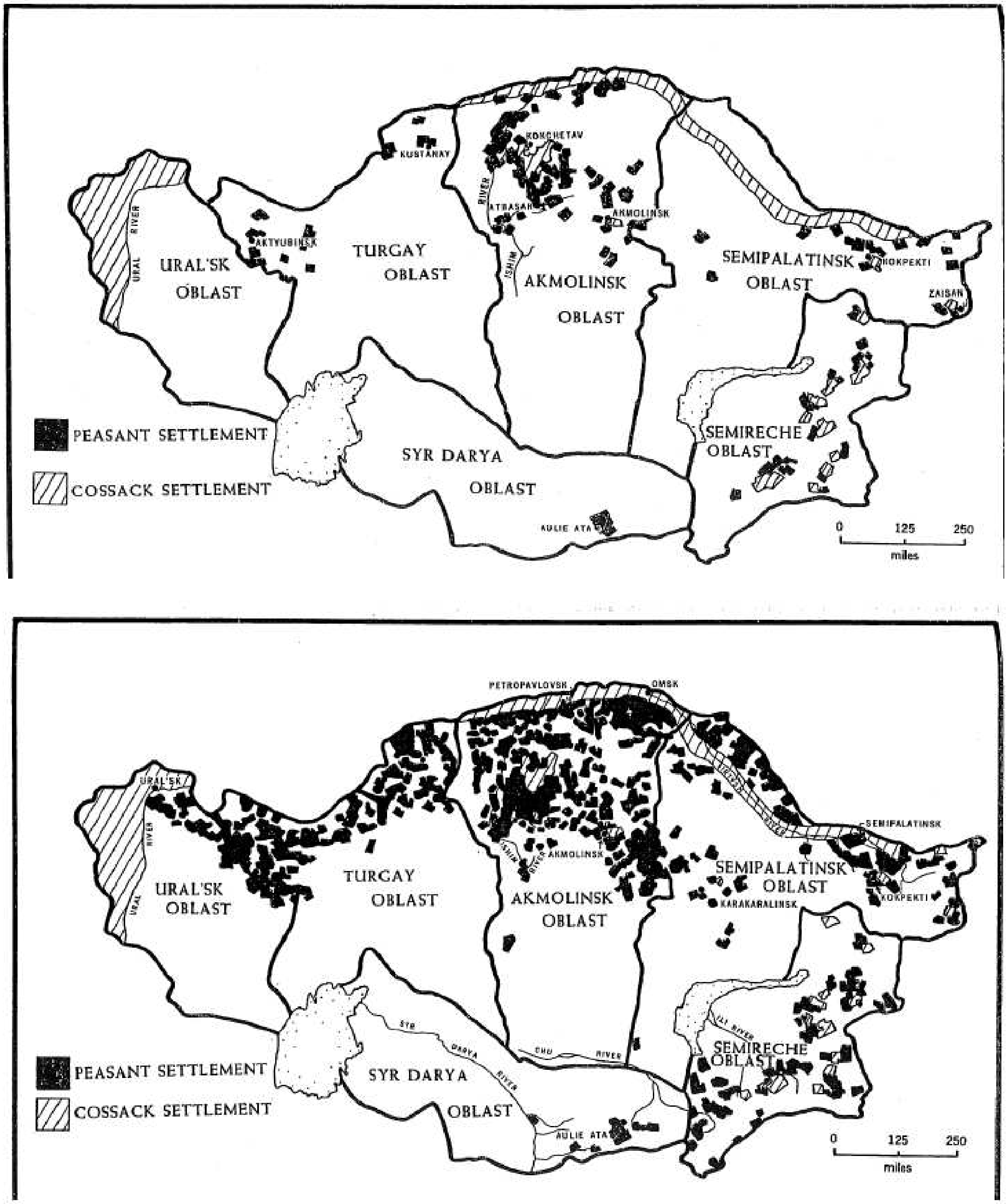

Figure 2. Settlement pattern of Russians in Kazakhstan, 1900 – 1915

Source: Demko (1969, Figures III-13 and III-14)

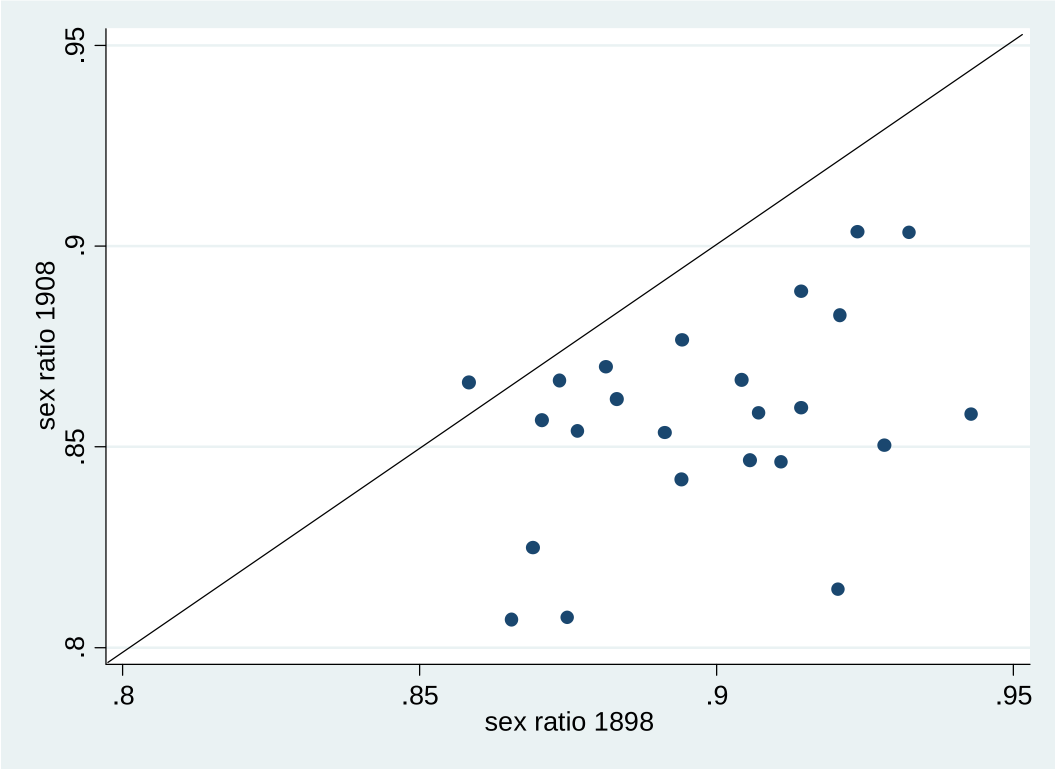

Figure 3. Overall sex ratios in 1908 as a function of sex ratios in 1898 by district (volost)

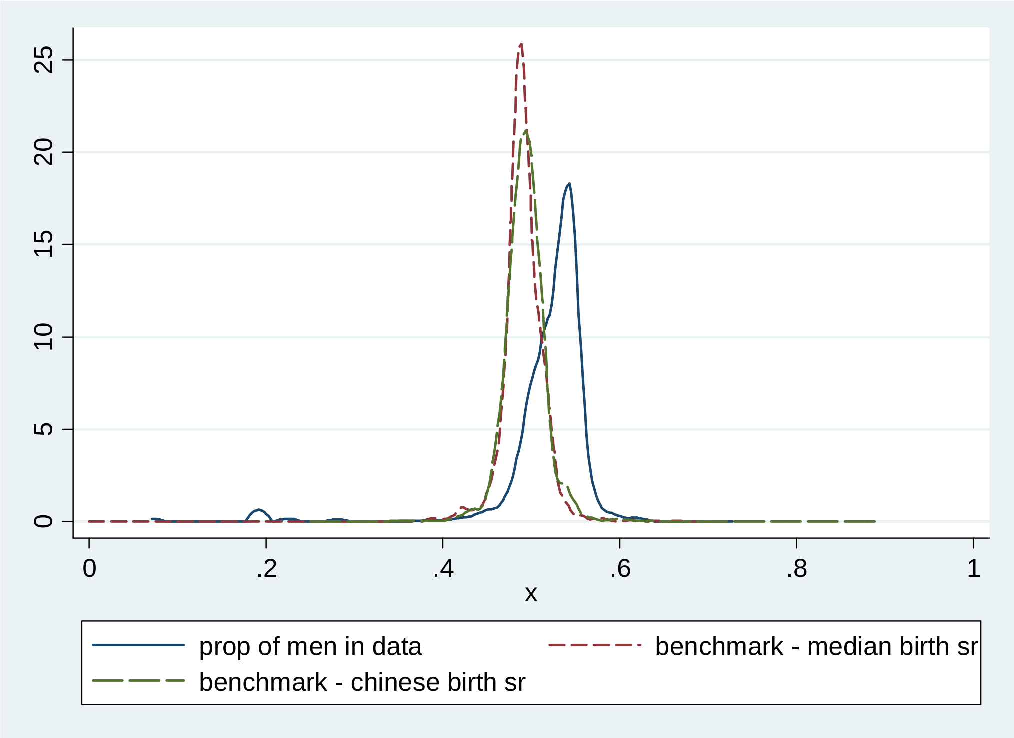

Figure 4. Distribution of proportions of men in sample (by wealth category) compared to benchmarks (East Model Stable Population Table) in 1898

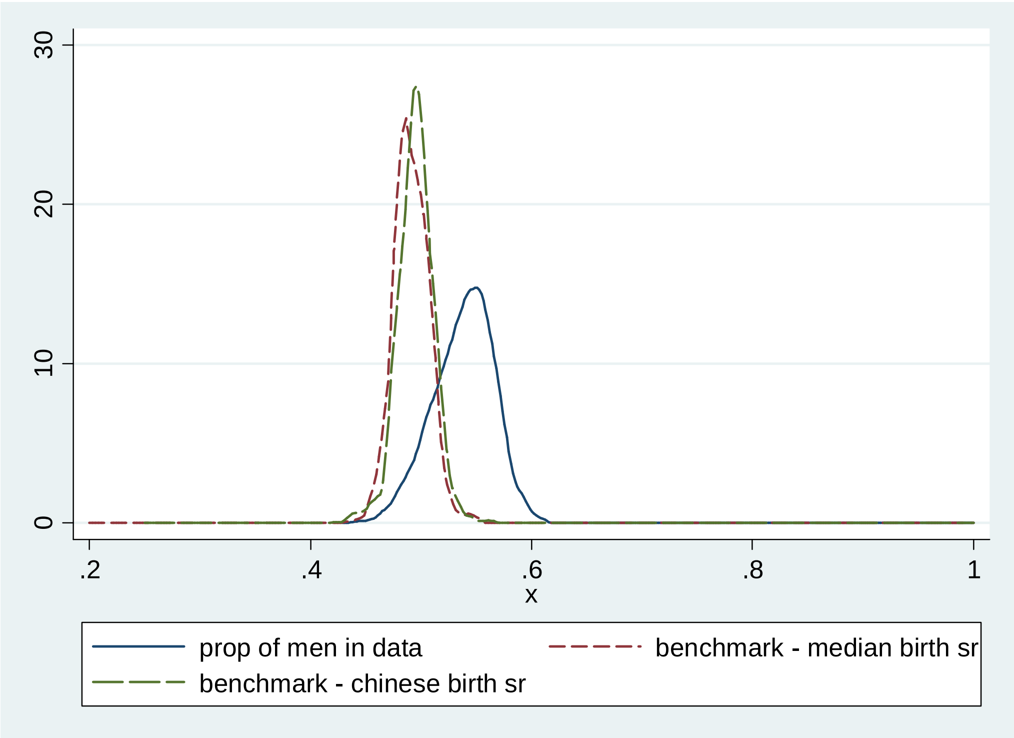

Figure 5. Distribution of proportions of men in sample (by wealth category) compared to benchmark (East Model Stable Population Table) in 1908

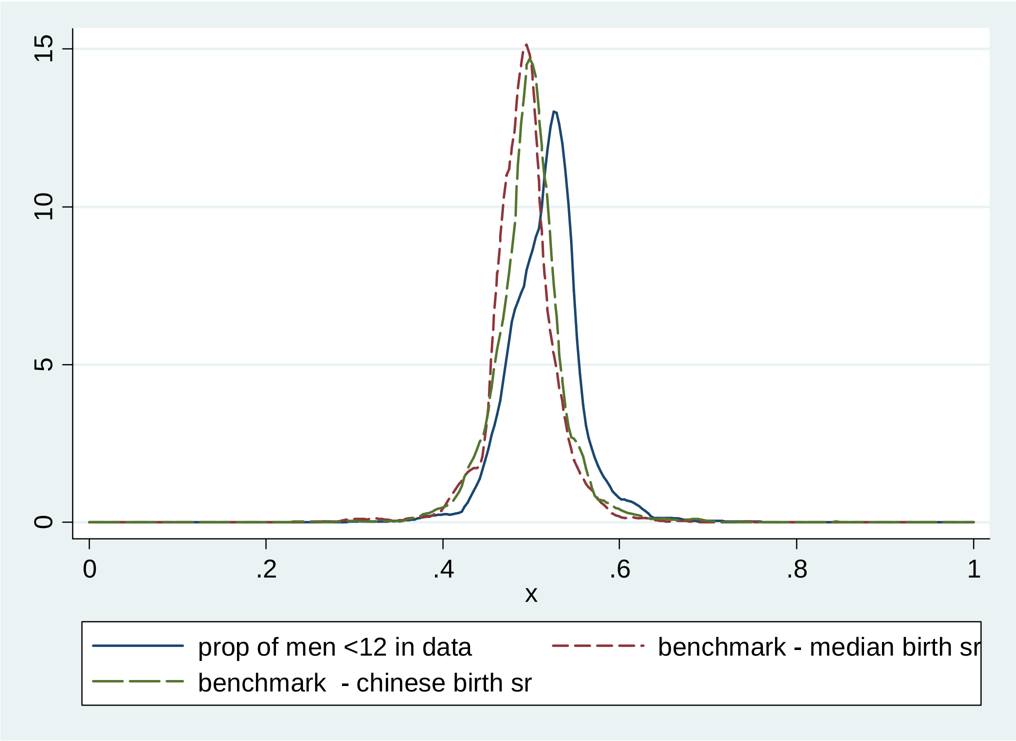

Figure 6. Distribution of proportions of men below 12 (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1898

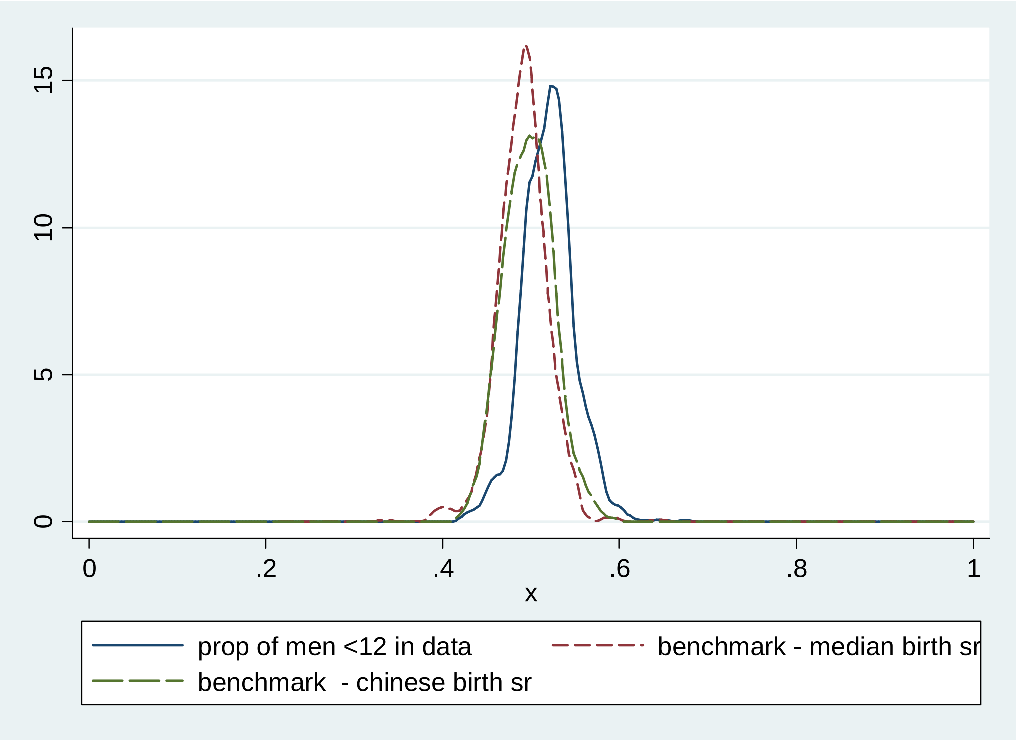

Figure 7. Distribution of proportions of men below 12 (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1908

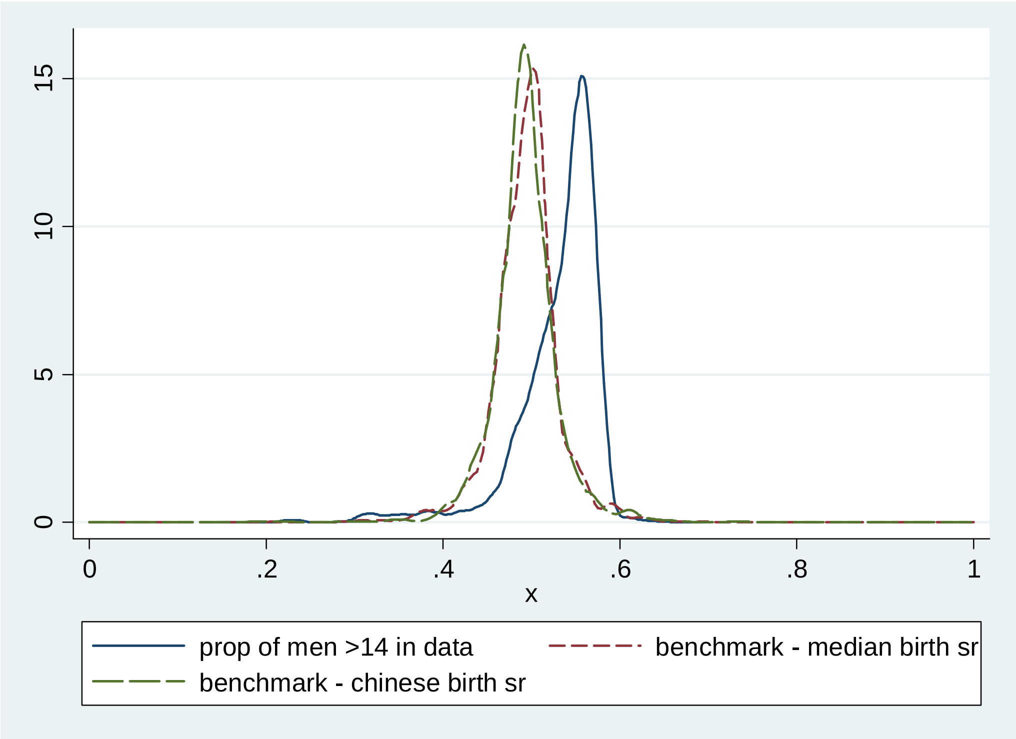

Figure 8. Distribution of proportions of men above 14 (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1898

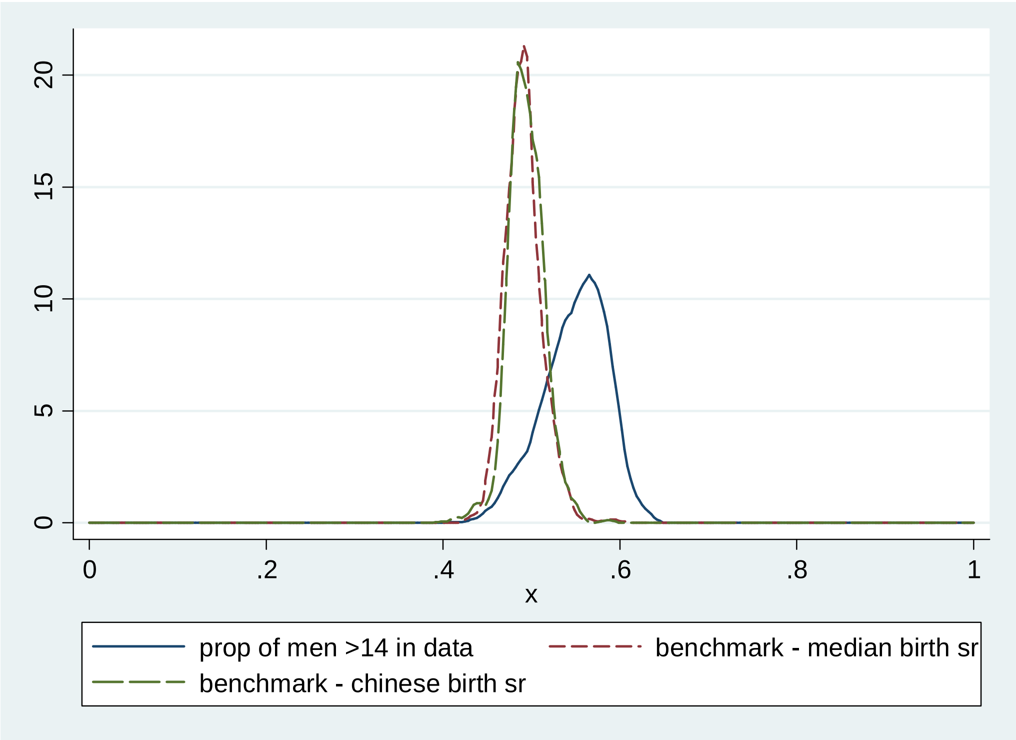

Figure 9. Distribution of proportions of men above 14 (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1908

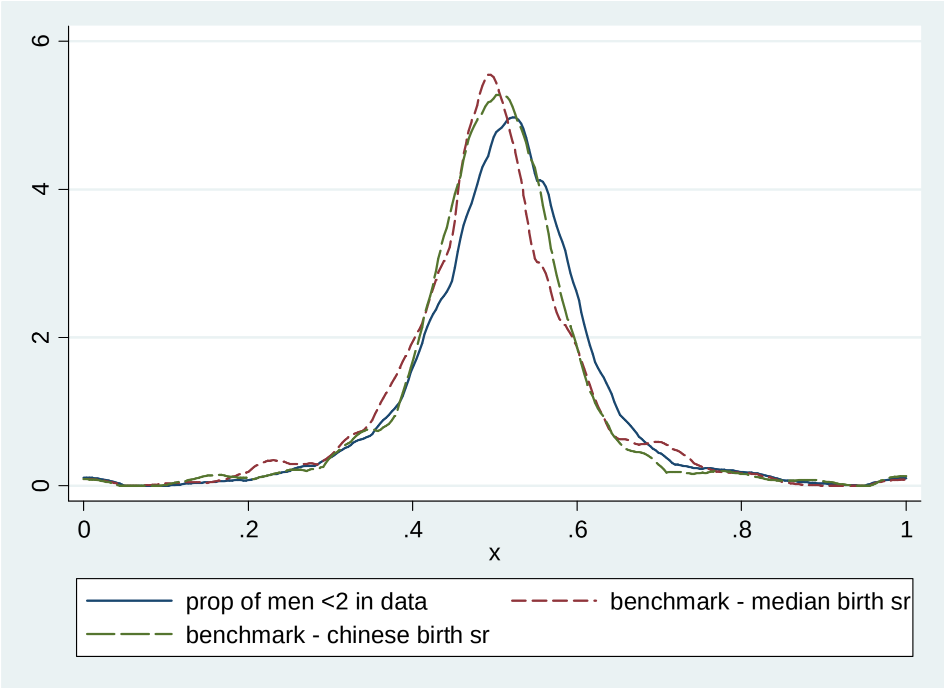

Figure 10. Distribution of proportions of men below one (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1898

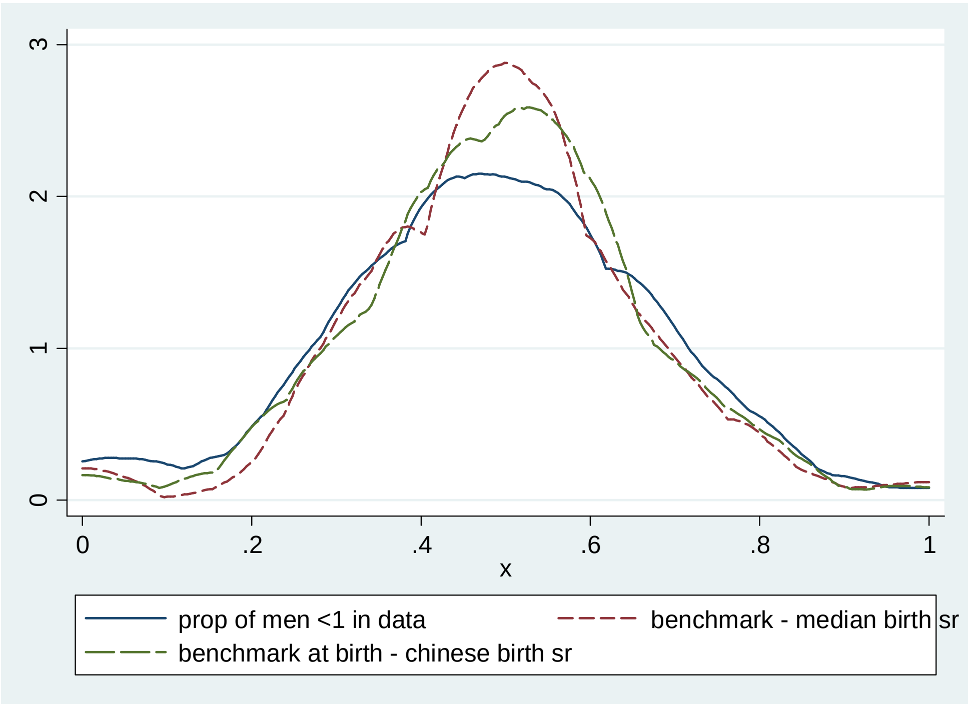

Figure 11. Distribution of proportions of men below one (by wealth category) in sample compared to benchmark (East Model Stable Population Table) in 1908

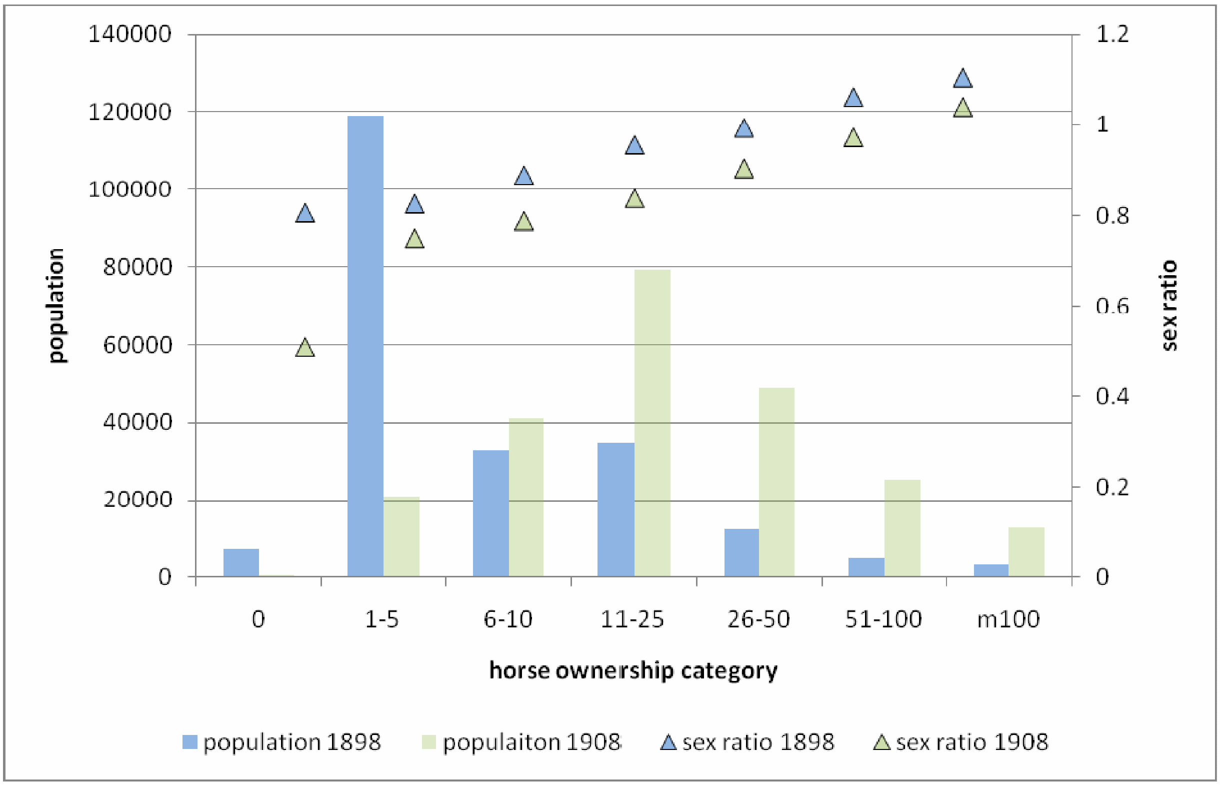

Figure 12. Wealth distribution and decline in sex ratios, 1898-1908

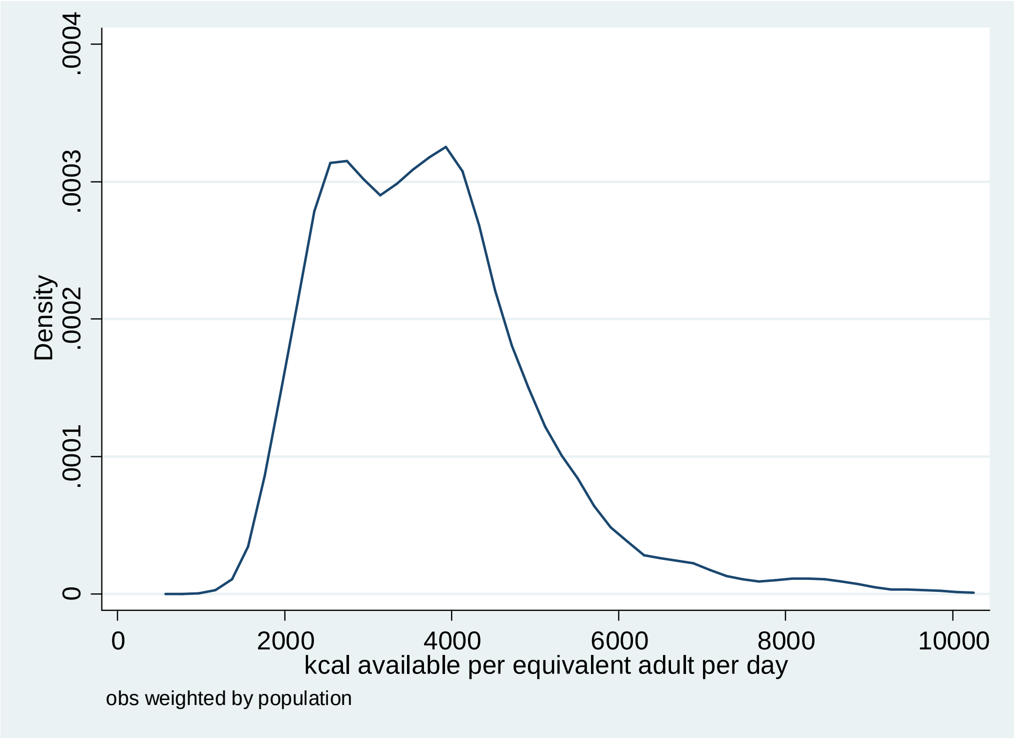

Figure 13. Kernel density estimate of calorie availability in 1898

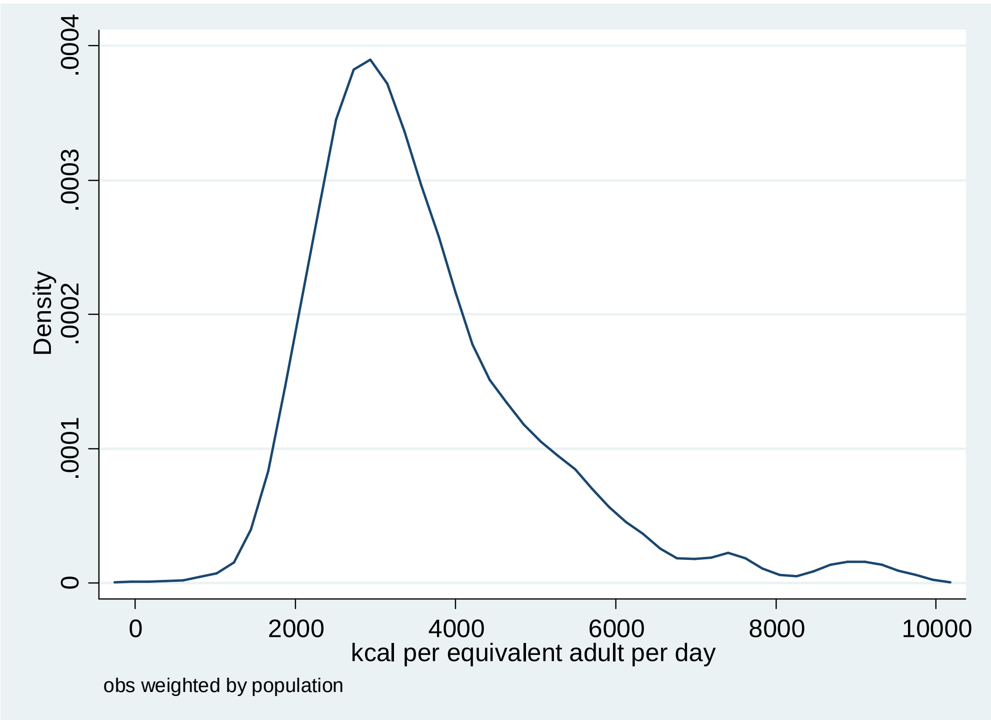

Figure 14. Kernel density estimate of calorie availability in 1908

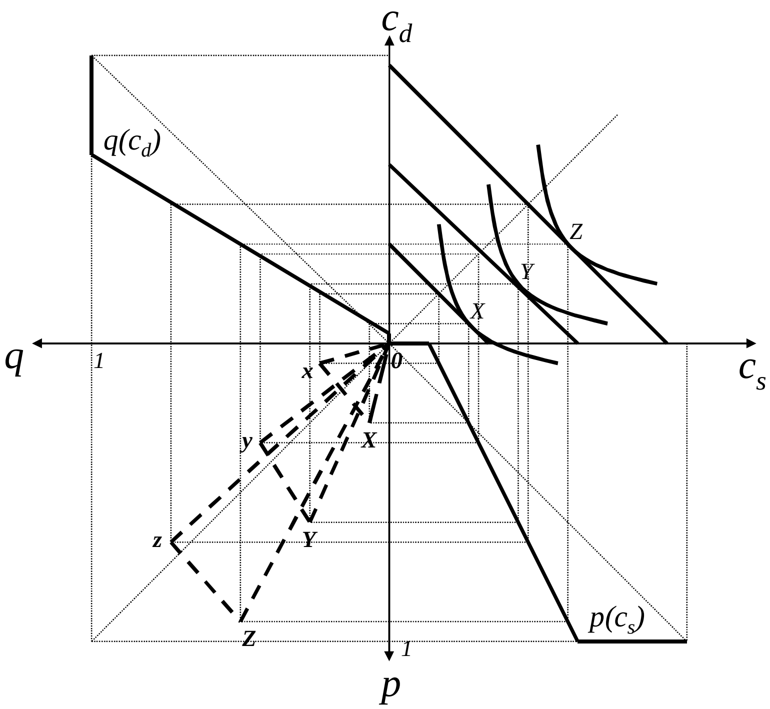

Figure 15. Children's food consumption, survival probabilities and sex ratios

Table 1. Population by ethnicity in selected regions of Kazakhstan, 1897 – 1916 (in thousands)

	1897		1905		1916		Annual growth rate in %, 1897-1916	
Oblast	Kaz.	Rus.	Kaz.	Rus.	Kaz.	Rus.	Kaz.	Rus.
Uralsk	460	164	477	268	480	278	0.224	2.817
Turgay	411	35	440	120	507	305	1.111	12.069
Akmolinsk	427	229	488	374	527	765	1.114	6.554
Semipalatinsk	605	68	669	82	665	200	0.499	5.842

Source: Demko (1969, various tables); our calculations.

Table 2. Population, sex ratios and confidence intervals by age categories, 1898 – 1908

	1898	1908	Definition/precisions
Population	214690	228614	Total population in both uezds
Sex ratio	0.8725	0.8573	#women/#men
Proportion male	0.5340	0.5384	#men/population
Confidence interval	[0.5320,0.5361]	[0.5364,0.5404]	for the proportion of male
Population >=14	135764	143990	# individuals aged 14 and above (>=14)
Sex ratio>=14	0.8287	0.8209	#women>=14 /#men>=14
Proportion male>=14	0.5468	0.5201	#men>=14/population>=14
Confidence interval	[0.5442,0.5494]	[0.5175,0.5227]	for the proportion of male
Population <12	68141	73116	# individuals aged 11 and below (<12)
Sex ratio <12	0.9408	0.9218	#women<12 /#men<12
Proportion male <12	0.5153	0.5204	#men<12/population<12
Confidence interval	[0.5115,0.5190]	[0.5168,0.5240]	for the proportion of male
Population =<1	8123	1831	for 1898 : # individuals aged 1 and below (<=1) for 1908: # individuals below 1 (<1)
Sex ratio =<1	0.947	0.988	for 1898 : # women <=1/# men <=1 for 1908: # women <1/# men <1
Proportion male=<1	0.5136	0.5029	for 1898 : #men<=1/population<=1 for 1908: #men<1/population<1
Confidence interval	[0.5027,0.5245]	[0.4800,0.5258]	for the proportion of male

Table 3. Benchmark sex ratios (SR) and proportion of men

	Model North		Model East	
	Median SR at birth	**Chinese SR at birth**	**Median SR at birth**	**Chinese SR at birth**
Benchmark sex ratio (all)	1.0372	1.0265	1.0321	1.0215
Benchmark sex ratio (<12)	0.9973	0.9871	1.0254	1.0148
Benchmark sex ratio (>14)	1.0582	1.0473	1.0355	1.0249
Benchmark proportion of men (all)	0.4910	0.4935	0.4921	0.4947
Benchmark proportion of men (<12)	0.5007	0.5032	0.4937	0.4963
Benchmark proportion of men (>14)	0.4859	0.4884	0.4913	0.4934

Source : Model Stable Population Table, Coale and Demeny (Mortality: 5, GRR:2.5). The median and Chinese sex ratios at birth are 0.944 and 0.935 respectively.

Table 4. Missing women in the Kazakh population in 1898 and 1908 for various levels of benchmark sex ratios

	Model North		Model East	
	Median SR at birth	**Chinese SR at birth**	**Median SR at birth**	**Chinese SR at birth**
1898				
Number of missing women	18859	17636	18284	17068
Number of missing women (<12)	1992	1632	2977	2607
Number of missing women (>14)	17030	16222	15345	14554
Missing / women pop	0.189	0.176	0.183	0.171
Missing / women pop <12	0.060	0.049	0.090	0.079
Missing / women pop >14	0.277	0.264	0.249	0.237
1908				
Number of missing women	22092	20782	21477	20173
Number of missing women (<12)	2812	2422	3879	3478
Number of missing women (>14)	18797	17937	17002	16160
Missing / women pop	0.210	0.197	0.204	0.191
Missing / women pop <12	0.080	0.069	0.110	0.099
Missing / women pop >14	0.290	0.276	0.262	0.249

Table 5. Population and sex ratios by age category and wealth in 1898

	0 horse	1-5 horses	6-10 horses	11-25 horses	26-50 horses	51-100 horses	>100 horses
Population	7492	118899	32575	34670	12487	5099	3468
Sex ratio	0.8053	0.8251	0.8888	0.9560	0.9935	1.0619	1.1044
Proportion male	0.5539	0.5479	0.5294	0.5112	0.5016	0.4850	0.4752
Confidence interval	[0.5427, 0.5652]	[0.5451, 0.5507]	[0.5240, 0.5348]	[0.5060, 0.5165]	[0.4928, 0.5104]	[0.4713, 0.4990]	[0.4586, 0.4918]
Population >=14	4611	76265	20441	21594	7696	3100	2057
Sex ratio>=14	0.7660	0.7689	0.8444	0.9369	0.9922	1.1204	1.1517
Proportion male>=14	0.5663	0.5653	0.5422	0.5163	0.5019	0.4716	0.4648
Confidence interval	[0.5519, 0.5806]	[0.5618, 0.5688]	[0.5354, 0.5490]	[0.5096, 0.5230]	[0.4901, 0.5131]	[0.4540, 0.4892]	[0.4432, 0.4863]
Population <12	2470	36832	10476	11257	4160	1729	1218
Sex ratio <12	0.8643	0.9267	0.9513	0.9666	0.9914	0.9540	1.0237
Proportion male <12	0.5364	0.5190	0.5125	0.5085	0.5022	0.5118	0.4941
Confidence interval	[0.5167, 0.5561]	[0.5140, 0.5241]	[0.5029, 0.5221]	[0.4992, 0.5177]	[0.4870, 0.5174]	[0.4882, 0.5353]	[0.4660, 0.5222]
Population =<1	239	4236	1230	1415	558	250	195
Sex ratio =<1	0.7704	0.9431	0.9680	0.9252	1.0440	0.9380	1.0526
Proportion male=<1	0.5649	0.5146	0.5081	0.5194	0.4892	0.5160	0.4872
Confidence interval	[0.5020, 0.6277]	[0.4995, 0.5296]	[0.4802, 0.5361]	[0.4934, 0.5455]	[0.4477, 0.5307]	[0.4540, 0.5780]	[0.4170, 0.5573]
Population	4.23	5.25	6.21	6.95	7.79	8.60	10.17
Sex ratio	1.39	1.63	2.00	2.26	2.60	2.92	3.57
Household size	4.23	5.25	6.21	6.95	7.79	8.60	10.17
Household size <12	1.39	1.63	2.00	2.26	2.60	2.92	3.57
Area cultivated per person (desyatinas)	0.40	0.66	0.59	0.77	1.07	1.55	2.69

Table 6. Population and sex ratios by age category and wealth in 1908

	0 horse	1-5 horses	6-10 horses	11-25 horses	26-50 horses	51-100 horses	>100 horses
Population	494	20869	41020	79158	48886	25370	12784
Sex ratio	0.5107	0.7505	0.7886	0.8376	0.9041	0.9720	1.0392
Proportion male	0.6619	0.5713	0.5591	0.5442	0.5252	0.5071	0.4904
Confidence interval	[0.6202, 0.7037]	[0.5646, 0.5780]	[0.5543, 0.5639]	[0.5407, 0.5477]	[0.5208, 0.5296]	[0.5009, 0.5132]	[0.4817, 0.4990]
Population >=14	363	13716	26146	50092	30327	15539	7731
Sex ratio>=14	0.4405	0.6985	0.7306	0.7874	0.8857	0.9808	1.1035
Proportion male>=14	0.6942	0.5887	0.5778	0.5595	0.5303	0.5049	0.4754
Confidence interval	[0.6468, 0.7416]	[0.5805, 0.5970]	[0.5718, 0.5838]	[0.5551, 0.5638]	[0.5247, 0.5359]	[0.4970, 0.5127]	[0.4643, 0.4865]
Population <12	114	6181	12822	25121	16087	8487	4360
Sex ratio <12	0.7538	0.8811	0.9061	0.9314	0.9316	0.9456	0.9257
Proportion male <12	0.5702	0.5316	0.5246	0.5178	0.5177	0.5140	0.5193
Confidence interval	[0.4793, 0.6611]	[0.5192, 0.5440]	[0.5160, 0.5332]	[0.5116, 0.5240]	[0.5010, 0.5254]	[0.5033, 0.5246]	[0.5045, 0.5341]
Population =<1	4	182	375	606	436	220	134
Sex ratio =<1	1.0000	1.2195	1.4351	0.9675	1.0185	0.8966	0.8611
Proportion male=<1	0.5000	0.4505	0.4107	0.5083	0.4954	0.5273	0.5373
Confidence interval		[0.3783, 0.5228]	[0.3609, 0.4604]	[0.4684, 0.5481]	[0.4485, 0.5423]	[0.4613, 0.5932]	[0.4530, 0.6217]
Population	2.92	4.49	5.4	6.37	7.51	8.27	9.67
Sex ratio	0.67	1.33	1.69	2.02	2.47	2.77	3.30
Household size	2.92	4.49	5.40	6.37	7.51	8.27	9.67
Household size <12	0.67	1.33	1.69	2.02	2.47	2.77	3.30
Area cultivated per person (desyatinas)	0.29	0.70	0.98	1.17	1.42	1.84	2.72
Expenditure per person	15.3	26.9	30.6	34.2	45.1	65.5	130.0

Table 7. Descriptive statistics of calorie availability per average adult (per day) (observations weighted by population)

	1898				1908			
	Mean	S.D.	Pc 25	Median	Mean	S.D.	Pc 25	Median
Kcal GRAIN	2269	804	1653	2148	1858	920	1495	1712
Kcal MILK	1013	788	637	784	1280	1026	678	959
Kcal MEAT	350	267	179	328	401	266	231	323
Kcal TOTAL	3632	1690	2597	3440	3540	1847	2466	3037

Table 8. Population and sex ratios by age category and household type in 1908

	HHs with workers hired out	HHs with craftsmen	Labor-autarchic HHs	HHs with workers hired in
Population	32172	40997	68268	67490
Sex ratio	0.6347	0.7827	0.8609	1.0241
Proportion male	0.6117	0.5609	0.5374	0.4940
Confidence interval	[0.6064,0.6170]	[0.5561,0.5657]	[0.5336,0.5411]	[0.4903,0.4978]
Population >=14	22304	26495	42888	40156
Sex ratio>=14	0.5462	0.7283	0.8222	1.0761
Proportion male>=14	0.6468	0.5786	0.5488	0.4817
Confidence interval	[0.6405,0.6530]	[0.5727,0.5845]	[0.5441,0.5535]	[0.4768,0.4866]
Population <12	8458	12662	21859	23677
Sex ratio <12	0.8880	0.8997	0.9297	0.9407
Proportion male <12	0.5297	0.5264	0.5182	0.5153
Confidence interval	[0.5190,0.5403]	[0.5177,0.5351]	[0.5116,0.5248]	[0.5089,0.5216]
Population <1	255	437	535	581
Sex ratio <1	1.1429	1.2296	1.1063	0.8864
Proportion male<1	0.4667	0.4485	0.4748	0.5301
Confidence interval	[0.4054,0.5279]	[0.4019,0.4951]	[0.4325,0.5171]	[0.4895,0.5707]
Household size	6.03	6.57	6.26	7.16
Household size <12	1.59	2.03	2.00	2.51
Expenditure per person	26.76	33.01	31.76	72.97
Area cultivated per person (desyatinas)	0.72	0.99	1.19	2.10

Table 9. Regression of the proportion of men on calorie availability and controls

	Dep. variable	Model	Regressors						Sample
			1000*kcal/c	$(1000*kcal)^2$	HH size	Area/cap	Wealth FE	District FE	
R1	prop men >14	OLS	-0.0067***				Yes	yes	all 1898
R2	prop men >14	OLS	-0.0165***				No	yes	all 1898
R3	prop men >14	OLS	-0.0340***				No	yes	trim top 5%
R4	prop men >14	OLS	-0.0380***	0.0014***			No	yes	all 1898
R5	prop men >14	GLM	-0.0678***				No	yes	all 1898
R6	prop men >14	GLM	-0.1536***	0.0057***			No	yes	all 1898
R7	prop men >14	GLM	-0.1206***	0.0045***	-0.0291***	0.0028	No	yes	all 1898
R8	prop men <=12	OLS	-0.0046***				Yes	yes	all 1898
R9	prop men <=12	OLS	-0.0028**				No	yes	all 1898
R10	prop men <=12	OLS	-0.0033				No	yes	trim top 5%
R11	prop men <=12	OLS	-0.0017	-0.0001			No	yes	all 1898
R12	prop men <=12	GLM	-0.0000**				No	yes	all 1898
R13	prop men <=12	GLM	-0.0066	-0.0003			No	yes	all 1898
R14	prop men <=12	GLM	-0.012	0.000	0.0054	-0.0074	No	yes	all 1898
R15	prop men >14	OLS	-0.0212**				Yes	yes	all 1898
R16	prop men >14	OLS	-0.0191***				No	yes	all 1908
R17	prop men >14	OLS	-0.0381***				No	yes	trim top 5%
R18	prop men >14	OLS	-0.0508***	0.0026***			No	yes	all 1908
R19	prop men >14	GLM	-0.0775***				No	yes	all 1908
R20	prop men >14	GLM	-0.1016***				No	yes	all 1908
R21	prop men >14	GLM	-0.0646***		0.0157	-0.0515*	No	yes	all 1908
R22	prop men <=12	OLS	-0.0006				Yes	yes	all 1908
R23	prop men <=12	OLS	-0.0014				No	yes	all 1908
R24	prop men <=12	OLS	-0.0037**				No	yes	trim top 5%
R25	prop men <=12	OLS	-0.0076**	0.0005**			No	yes	all 1908
R26	prop men <=12	GLM	-0.0056				No	yes	all 1908
R27	prop men <=12	GLM	-0.0302**	0.0021**			No	yes	all 1908
R28	prop men <=12	GLM	-0.0410***	0.0026***	0.001	0.0129	No	yes	all 1908

Note: In specifications R20 and R21 the squared (calorie availability) term has not been added to the regression because it creates the problems of covergence with GLM.