Three Essays on the Economics of Nomads and Settlers

Andrea Matranga

TESI DOCTORAL UPF / ANY 2017

DIRECTOR DE LA TESI

Hans-Joachim Voth
Departament Economia i Empresa

Dedicated to Ana and Mila. I knew I'd find you!

Acknowledgements

My greatest debt of gratitude is towards my advisor, Hans-Joachim Voth, for taking what was a very simple idea on a related matter, steering me towards a more interesting topic, and preventing from straying from the correct path. Further, he honed my logical, research, and exposition skills through countless conversations and emails, and at no point lost sight of my best interests.

Luigi Pascali, Davide Cantoni, Paola Giuliano and Noam Yuchtman have been a constant source of encouragement and guidance since I've met them, and also an example of the professional, intellectual, and human qualities which an academic should embody.

The papers that compose this thesis benefited greatly from being presented at the seminar series at Nuffield, UCLA Anderson, UC Irvine, UCSD, Berkeley, Harvard, LMU Munich, Santa Clara, Stanford, LSE, Warwick, Toulouse, NES, as well as the World Economic History Congress, World Clio, the Zeuthen Workshop, the Econometric Society World Congress, the UPF Jamboree, the UPF Summer Forum, and the Economic History Society conference. The attendees of the International Lunch at UPF deserve a special mention for having had to endure many intermediate stages of this project. The final result reflects their comments and guidance.

Oded Galor was kind enough to invite me to Brown for a semester, and many of the most interesting parts of the Neolithic project were performed, or at least initially conceived during that stay. It was also in the Brown Library that I started to be interested in the Mongol Empire, an interest which has so far spawned the third chapter of my thesis, and which I plan to continue to explore over the coming years. My Neolithic paper is, in too many ways to count, inspired by Stelios Michalopoulos's work with Quamrul Ashraf, and while I have never had the pleasure of meeting Quamrul, Stelios has been a recurring source of support and advice throughout the years. During my time at Brown I also benefited greatly from conversation with David Weil and Louis Putterman, as well as with their fantastic PhD students and fellow visiting scholars Francesco Cinnirella, Rawaa Harati, and Elena Esposito. I would also like to thank John Cherry to allow me to audit his archaeology graduate class, and all his students for their advice, in particular Tom Leppard.

My colleagues in the PhD programs at BGSE have supported my progress in every possible way. I am particularly indebted to Mrdjan Mladjan, Bruno Caprettini, Peter Koudijs, Francesco Amodio, Felipe Valencia, Jacopo Ponticelli, Vicky Fouka, Andrei Potlogea, Dmitry Khametshin, Dijana Zecirovic, Caterina Alacevic, Matt Delventhal, Hrvoje Stojic.

I am particularly grateful to Jared Diamond for first of all inspiring my work on the Neolithic Revolution, but also for reading the chapters on the Neolithic

Revolution, and for offering valuable advice.

To my academic colleagues at NES I owe a special debt of gratitude, in mentoring and guiding me through these early stages in my career, and sharing coffees, ideas, and good times.

Abstract

This thesis looks at the interplay of nomadism and settlement in two different periods of history. In the first two chapters I develop a theory for the Neolithic Revolution, the transition form nomadic hunting and gathering to settled agriculture. I argue that an exogenous increase in climate seasonality made our ancestors become sedentary in order to store food. Once sedentary, inventing agriculture was only a matter of time. In the first chapter I construct a model encapsulating this intuition, and in the second I test empirically the predictions of the model. In the third chapter, I instead argue that Russia introduced serfdom the 16th century in order defend against slave raids from their nomadic neighbors to the south. If labor had remained free, the population would have clustered around the most fertile areas, leaving less productive areas undefended, and thus vulnerable to the raids.

Resum

Aquesta tesi estudia la interacció entre nomadisme i sedentarisme en dos periodes diferents de la història. En els dos primers capítols desenvolupo una teoria sobre la Revolució Neolítica, la transició de la caça nòmada i la recol·lecció a l?agricultura sedentària. Defenso que un increment de l?impacte de l?estacionalitat climática va fer que els nostres avantpassats es tornessin sedentaris amb la finalitat de poder emmagatzemar menjar. Un cop sedentaris, la invenció de l?agricultura era només una qüestió de temps. En el primer capítol construeixo un model que captura aquesta intuició, en el segon provo les prediccions del model de manera empírica. En el tercer argumento que Rússia va adoptar la servitud durant el segle XVI per tal de defensar-se contra les ràtzies esclavistes dels seus veïns nomàdics del sud. Si no s?hagués introduït la servitud, i la mà d?obra hagués romàs lliure, la població s'hauria agrupat al voltant de les zones més fèrtils, deixant sense defensar zones menys productives i, per tant, vulnerables a les incursions.

Preface

This doctoral thesis consists of three essays at the intersection between economic history, long run growth and economic geography. They are linked by a common focus on the interaction between nomadism and settlement. The first two chapters analyze the first instance in which large numbers of humans became sedentary, while the last one looks at the one of the last cases in which nomads were competitive with settlers on the world stage.

In the first chapter, I propose a new theory for the Neolithic Revolution, i.e. the transition from nomadic hunting and gathering to settled agriculture – an event which transformed almost every facet of human life. While the existing literature has overwhelmingly focused on the agriculture part, I argue that becoming sedentary was a crucial intermediate step. I argue that an exogenous global increase in climate seasonality made nomadism a less effective risk mitigation strategy, compared to the preceding history of our species. As a result, many of our ancestors decided to become sedentary, so that they could smooth consumption by storing food. Once a population is living in the same location year round, agricultural techniques are much easier to develop, and also more necessary (given the limited extent of land a settled population can access). I construct a simple but adaptable model that encapsulates this intuition, and from it I derive several testable implications on the type of climate which should be associated with invention and early adoption, as well as on the likely topography of locations that are fast adopters, and on the expected evolution of health variables across the transition. I also sketch several possible extensions to the basic model.

In the second chapter, I test whether the data supports seasonality as the main cause for the invention agriculture. The first part of my analysis is performed at the global scale, and I show that a) agriculture is strongly correlated wit the date of adoption; b) agriculture predicts the locations where agriculture was independently invented; and c) agriculture predicts the speed at which agriculture diffused. For this last result I introduce a novel "frontier" empirical strategy which focuses exclusive on the expanding shell of locations that have at least one neighbor that has already adopted agriculture, but still hasn't adopted themselves. I then focus the analysis only on the most comprehensively studied part of the world, using a dataset of 765 archaeological sites from the Middle East and Europe, which allows me to sidestep many of the issue related to sampling bias in archaeological excavations. This regional sample gives the the same result as the previous global scale analysis: more seasonal locations adopted agriculture first. I then use an even more restricted dataset of locations from only in the Middle East, which all had access to similar climate and domesticable plants. Within this very homogenous context, I find that areas with more uncorrelated microclimates (proxied by variations in altitude) delayed their adoption of agriculture. This is consistent with the local populations being reluctant to become sedentary where geographic

heterogeneity meant the opportunity cost of abandoning nomadism was high. Finally, I present some evidence that the transition from hunting and gathering to agriculture was indeed associated with a reduction in the seasonality of food consumption.

The third chapter looks at nomads and settlers in a completely different context: that of Eastern Europe in the 15th to 18th Century, a period during which the settled farming population Russia was suffering nearly constant slave raids from the South. I argue that this threat was responsible for the adoption of serfdom in the 16th century. As these raids were conducted by nomadic horsemen without any kind of logistical support, it was impossible to defend against them with point fortifications such as castles or fortresses: the raids could simply bypass them and strike rural regions in the interior. Instead, the only way of mounting a successful defense was by building continuous lines of fortification blocking all avenues of approach. Given the limited state capacity for taxation and decentralized garrison administration, the only way to finance such an extended system of fortifications was through feudal land partitioning: each soldier was assigned land close to the section of wall he was responsible for, and the proceeds from the harvests would cover his expenses and serve as reward. Since the campaign season coincided with the agricultural season, the actual farming clearly had to be performed by somebody either hired farmhands, or tenant farmers. But in either case, landless peasants would have flocked to the most fertile areas, leaving more marginal lands underpopulated — and impossible to defend. I therefore argue that in the Russian case serfdom was introduce as a response to the presence of lands that were economically unimportant, but strategically crucial. Using data from a later population census, I show that areas along the lines of fortification had the highest prevalence of serfdom.

Contents

List of Figures

List of Tables

Chapter 1

THE ANT AND THE GRASSHOPPER: A SESONAL THEORY FOR THE ORIGINS OF AGRICULTURE

Why was agriculture invented? The long run advantages are clear: farming produced food surpluses that allowed population densities to rise, labor to specialize, and cities to be constructed. However, we still don't know what motivated the transition in the short run (Gremillion et al., 2014, Smith, 2014). After 200,000 years of hunting and gathering, agriculture was invented independently at least seven times, on different continents, within a 7,000 year period. Archeologists agree that independent inventions occurred at least in the Fertile Crescent, Subsaharan Africa, North and South China, the Andes, Mexico, and North America. Moreover, the first farmers were shorter and had more joint diseases, suggesting that they ate less than hunter gatherers and worked more (Cohen and Armelagos, 1984). Why would seven different human populations decide to adopt remarkably similar technologies, around the same time, and in spite of a lower standard of living?

In this Chapter, I propose a new theory for the Neolithic Revolution, construct a model capturing its intuition, and present a few stylized facts in support. I argue that the invention of agriculture was triggered by a large increase in climatic seasonality, which peaked approximately 12,000 years ago, shortly before the first evidence for agriculture appeared. This increase in seasonality was caused by well documented oscillations in the tilt of Earth's rotational axis, and other orbital parameters (Berger, 1992). The harsher winters, and drier summers, made it hard for hunter-gatherers to survive during part of the year. Some of the most affected populations responded by storing foods, which in turn forced them to abandon their

nomadic lifestyles, since they had to spend most of the year next to their necessarily stationary granaries, either stocking them, or drawing from them. While these communities were still hunter-gatherers, sedentarism and storage made it easier for them to adopt farming.

To guide the empirical analysis of Chapter 2, I develop a simple model that analyzes the incentives faced by hunter-gatherers relying on a resource base that varies across both space and time. I modify the standard Malthusian population dynamic by assuming that consumption seasonality reduces fertility. I find that a large increase in seasonality can cause agents to switch from nomadism to settlement, even if they still don't know how to farm. Despite consuming less on average, the ability to smooth consumption through storage more than repays this loss, meaning that the settlers are now better off both in the short and long run.

The theory suggests that more seasonal locations should receive agriculture sooner. My theory is supported by a wealth of archaeological evidence. In the Middle East, the Natufians, ancestors of the first farmers, lived for thousands of years as settled hunter-gatherers, intensively storing seasonally abundant wild foods (Kuijt, 2011). Even in historical times, hunter-gatherers exposed to seasonal conditions have responded by becoming sedentary and storing food for the scarce season. For example, Native Americans in the Pacific Northwest relied on highly abundant, but highly seasonal salmon runs, which they would trap en masse and smoke for the winter (Testart, 1982). While the complex life cycle of salmon placed them beyond their ability to domesticate, they nonetheless evolved societies which had most of the characteristics of farming villages, except for farming itself: permanent houses in which they lived in year long, elaborate material cultures, and social stratification.

Further, taking storage into account allows us to understand why agriculture was adopted in spite of the reduction in consumption per capita: the first settlers accepted a poorer average diet in exchange for the ability to smooth their consumption. Evidence from growth-arrest lines in their bones confirms that, while farmers ate less than hunter-gatherers on average, they suffered fewer episodes of acute starvation (Cohen and Armelagos, 1984).

The setting of the Neolithic Revolution is unique in that very similar technologies were developed multiple times by different groups. Unlike e.g. the Industrial Revolution, it is therefore possible to draw parallels between different adoptions and identify what all of them had in common. Many contributions have focused on changes in average climate. The Neolithic period started shortly after the end of the Late Pleistocene glaciation, which lasted from 110,000 to 12,000 years ago. This has led some researchers to hypothesize that either warmer weather made farming easier (Bowles and Choi, 2013) or drier conditions made hunting and gathering more difficult (Childe, 1935). Ashraf and Michalopoulos (2013) propose a variant on the climatic theme, and argue that intermediate levels of

inter-annual climate volatility led to the gradual accumulation of latent agricultural knowledge. The problem with these explanations is that they assume that farming was motivated by a desire for greater *average* food consumption. The fact that they ended up eating less suggests that greater food consumption is unlikely to be the motive.

Other contributions have focused on explaining the reduction in consumption per capita. This loss has been variously attributed to unforeseen population growth (Diamond, 1987), the need for defense (Rowthorn and Seabright, 2010), or expropriation by elites (Acemoglu and Robinson, 2012). While these may all have been contributing factors, they do not explain why agriculture was invented in particular places and at particular times. The key contribution of this paper lies in proposing a unified theory for the origins of agriculture, which can explain both of these puzzles: the geographic pattern of adoption and the resulting decrease in consumption per capita. The model I propose generates clear empirical predictions, which I test against the paleoclimatic record, the local topography of early adoption sites, and the evidence from the skeletons of the first farmers.

This paper also contributes to the vast and growing literature on the economic effects of climate and the environment, for which Dell et al. (2013) provide an extensive review. I argue that increased climatic seasonality presented a challenge to the established way of life of humans, who responded by adopting a novel life strategy — sedentary storage — to mitigate the negative consequences of this change in climate. This new lifestyle was already a big change, but it would be soon overshadowed by the incredible technological and social innovations that it facilitated: agriculture, stratified societies, and the accumulation of capital. As in Acemoglu et al. (2012), these finding remind us that when environmental factors force societies to invest in radically different technologies, the effect on the incentives to innovate are often more important than the immediate changes in lifestyle.

1.1 Literature review

A large multidisciplinary literature has tried to explain why humans started to farm. Early contributions (Darwin, 1868) focused on the greater abundance of food that agriculture allowed, but the decrease in standard of living suggests that this was not the primary reason. Climate change is arguably the only factor capable of explaining the simultaneous invention on different continents Richerson et al. (2001), and indeed agriculture was invented after the end of the last Ice Age. This suggested that warmer climates may have made farming more productive (Diamond, 1997, Bowles and Choi, 2013), or else drier conditions made hunting and gathering worse (Braidwood, 1960). For Dow et al. (2009), the Neolithic

revolution was the result of a large climatic reversal: first, improving climates allowed population density to rise, but a later return to near-glacial conditions forced hunter-gatherers to concentrate in the most productive environments. The problem with all these stories is that the last Ice Age lacked neither warm conditions, nor dry ones, nor climatic improvements followed by rapid reversals, and yet agriculture was not invented. Humans had inhabited areas with similar conditions for tens of thousands of years without any sign of progress towards agriculture.

Ashraf and Michalopoulos (2013) propose that intermediate levels of inter-annual volatility favored accumulation of latent agricultural knowledge. They use modern cross-sectional climate data to show that both very high and very low levels of year-on-year variation in temperatures appears to have delayed adoption. Their paper is in some ways similar to my own — both isolate a type of climate as crucial for agriculture and test their hypothesis using a variety of climate and adoption data. However, I focus on seasonality, rather than on inter-year volatility, and I argue that the crucial step was the decision to become sedentary and store food.

Other contributions have focused on the role of population growth. One possibility is that overexploitation decreased the productivity of hunting and gathering (Olsson, 2001, Smith, 1975). Locay (1989) proposed another channel: rising populations reduced the size of each band's territory and thus reduced the need for nomadism. Populations responded by becoming settled, which made farming much easier. As in the present paper, settlement is thus seen as an essential stepping stone towards the Neolithic. However, I argue that the loss of nomadic usefulness came from highly seasonal climate, which made all locations within migratory range similarly unproductive at the same time.

A large multidisciplinary research effort has investigated the long run impact of the invention of agriculture. Cohen and Armelagos (1984) documented a large and persistent decrease in a number of health measures. Diamond (1997) argued that populations that transitioned early gained an early technological lead, that largely predetermined which continents would eventually inflict colonialism, and which would suffer it. The switch to farming influenced our genes, by selecting for certain psychological and physiological traits which we still carry (Galor and Michalopoulos, 2012), (Galor and Moav, 2007). Crops that required plowing placed a premium on upper body strength, resulting in persistent differences in gender norms (Alesina et al., 2013). Indeed, cultivation of the same crops could result in very different social institutions, depending on the surrounding geography (Mayshar et al., 2013).Olsson and Paik (2013) suggest that continued farming gradually increased land productivity but eventually led to more autocratic societies.

My analysis suggests that our ancestors rejected an abundant but risky lifestyle, in exchange for one that had lower returns but was more stable. Risk aversion has

been proven to be a powerful motive for lifestyle decisions, especially in populations close to the subsistence limit. McCloskey (1991) showed how English farmers preferred to diversify their labor investment across scattered fields, even though this reduced their productivity. Acemoglu and Zilibotti (1997) argued that the presence of large risky projects slowed down technological progress. Tanaka (2010) examined farmer's utility functions in a series of field experiments in Vietnam and found that the inhabitants of poorer villages were more risk averse. In most of these contributions, risk-aversion is seen as an economically costly trait. I show that a desire for stability can also promote economic growth, if the risk mitigating strategies adopted happen to make innovation less costly.

In the basic Malthusian framework, populations should never be able to maintain consumption per capita significantly above subsistence. To explain how some societies can enjoy high incomes for extended periods, Galor and Weil (2000) proposed that continued population growth increased the rate of technological progress, motivating parents to have fewer children, with more human capital. This shift could have led to the proliferation of genetic traits that were complementary to economic growth (Galor and Moav, 2002). Alternatively, the death of a significant part of the population could force a shift to a production system that encouraged higher mortality (Voigtländer and Voth, 2013b), and lower fertility (Voigtländer and Voth, 2013a). Wu et al. (2017) show that incomes can remain above subsistence if agents derive utility also from non-food items, such as entertainment. I contribute to this literature by showing that a population equilibrium with high consumption per capita can also be caused by consumption seasonality.

A number of recent contributions have explored the effect of topographic relief on economic outcomes. Nunn and Puga (2012) showed that rugged areas in Africa were partially protected by slaving incursions. Michalopoulos (2012) documented the role of ruggedness in forming ethnolinguistic groups. Fenske (2014) noted that regions with more varied ecosystems have greater incentives to trade, and showed that the more successful African governments benefit from these conditions. My research contributes to this literature by showing that variations in altitude can have opposing effects depending on the scale at which they occur. In particular, they can create a variety of different microclimates within a compact region, affecting the usefulness of mobility.

Latitude correlates heavily with most measures of development. Explanations for this phenomenon have included unabashed racism (Montesquieu, 1748), thinner soils, more harmful parasites, ferocious diseases, unstable rainfall, and lack of coal deposits (Bloom et al., 1998). Acemoglu et al. (2002) maintain that the direct effect of these geographic differences is overshadowed by the institutional outcomes which they support. Easterly and Levine (2003) find support for this in a dataset linking GDP, institutions, the mortality of the first settlers, and several measures of natural resources. Since latitude and seasonality are highly correlated,

the findings of this paper suggest that part of the association between latitude and development outcomes might be due to the different amount of time humans have been performing agriculture at various distances from the equator.

1.2 Historical background

For the first 200,000 years of our species' existence, our ancestors relied exclusively on wild foods for survival. The hunting and gathering lifestyle sustained them from the plains of Africa, throughout their successive migrations. By 14,000 BP, humans had colonized all continents except Antarctica and hunted and gathered from the tropical rainforest to the arctic tundra. The incredible versatility of this lifestyle was partly due to nomadism. By constantly moving to temporarily more abundant areas, humans could survive even where no single location provided a reliable food supply. Hunter-gatherers managed to develop rich and unique cultures and technologies, adapted to the opportunities and requirements of their specific surroundings. These trends solidified approximately 60,000 years ago, when humans acquired behavioral modernity: they developed languages, made art, decorated their bodies, and buried their dead.

After this milestone, however, further progress had been comparatively modest. Our ancestors continued to refine their techniques, and to adapt them to changing environments, but the basic pattern remained unchanged. In particular, no population is known to have domesticated crops until about 12,000 years ago.

The Neolithic transition is now understood to have occurred gradually, starting from relatively minor actions – such as pulling up weeds, and culminating in highly complex endeavours – such as the excavation of massive irrigation channels. These activities changed the selective pressures operating on cultivated species, which soon evolved to take advantage of human assistance — they became domesticated (Harlan, 1992). This resulted in crops which were more productive, easier to harvest, and able to grow in a wider range of conditions.

The very earliest farmers belonged to the Pre-Pottery Neolithic B culture, which domesticated wheat and barley in the hills of the Fertile Crescent approximately 11,500 years ago (Belfer-Cohen and Bar-Yosef, 2002). Within seven thousand years, agriculture would be invented independently at least six more times, in the Andes, North and South China, Mexico, Eastern North America, and Sub-Saharan Africa (Purugganan and Fuller, 2009). Each of these locations had different climates and available plant species, and was inhabited by populations who had not been in contact for tens of thousands of years. Figure 1.1 shows the independent farming inventions and their dates.

Thanks to farming, the same amount of land could feed more stomachs. The

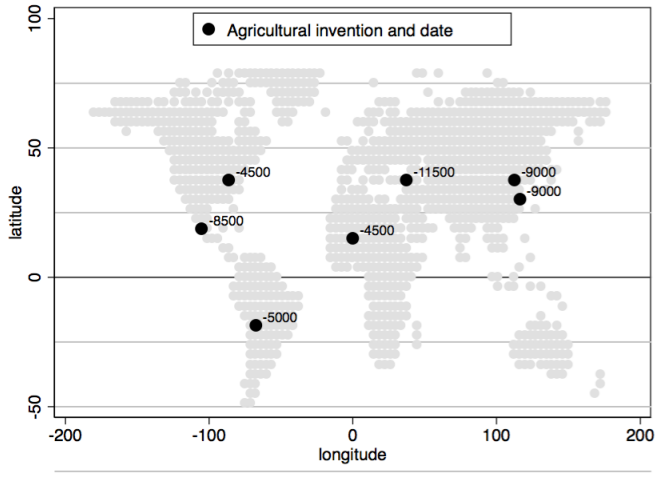

Figure 1.1: The locations where agriculture was invented and their respective dates in years before present.

increased population density led to the rise of the first cities, with specialized labor and centralized leadership. Agriculture spread rapidly to neighboring communities, through various combinations of inter-marriage, conquest, and imitation. Eventually, hunter-gatherers were relegated to a few isolated or inhospitable locations. This process of diffusion is largely responsible for the current distribution of ethnic groups, languages, and food staples (Ammerman and Cavalli-Sforza, 1984). Farmers were sedentary, and thus free to accumulate more personal possessions than nomads. Pottery, metalworking, and architecture were just some of the technologies that emerged as a result.

The lack of progress towards agriculture after achieving behavioral modernity was at least partly due to the nomadic lifestyle, typical of hunter-gatherers. Since successful farming requires constant interaction with the plants under cultivation, it was very difficult for a nomadic population to discover agricultural techniques. First, nomads would typically never witness the same individual plant growing throughout the year. They were thus less likely to understand how their actions affect plant growth. Second, even if they did find out how to cultivate certain plants, they would have found it hard to schedule their movements so as to be present when farm work needed to be done.

I argue that the rise of the Neolithic was ultimately caused by unprecedented climate seasonality. What caused these conditions? The patterns of climatic seasonality experience on Earth depend chiefly on the shape of Earth's orbit, as described by three parameters: axial tilt, eccentricity, and precession. During the Ice Age, the Earth's axis of rotation was less tilted, and its orbit was less elliptic.

Moreover, when the northern hemisphere was tilted towards the Sun, the planet was at its aphelion — the furthest point from the Sun along its orbit. As a result, the two effects partially canceled out, and climate was not very seasonal. Between 22,000 and 12,000 BP, changes in these parameters made global climate patterns become steadily more seasonal (see Figure 1.2). By 12,000 BP, sunlight seasonality in the northern hemisphere was higher than it had been at any time since our species had acquired behavioral modernity, 50,000 years prior. In the northern temperate zone (between $30°\mathrm{N}$ and $40°\mathrm{N}$), hunter-gatherers could gorge themselves during the hot rainy summers, but they risked starving in the harsh winters. Conversely, tropical areas enjoyed warm weather year round but often suffered from intensely seasonal rainfall. Between $15°$ and $20°$ on either side of the equator, vast areas would come to life during the wet season and then become barren during the dry one. In fact, all confirmed independent inventions of farming occurred within these two absolute latitude bands: the Middle East, Eastern North America, North China and South China all lay within the temperate zone of the Northern hemisphere, while Sub-Saharan Africa, the Andes, and Mexico are all within the tropical area of rainfall seasonality.

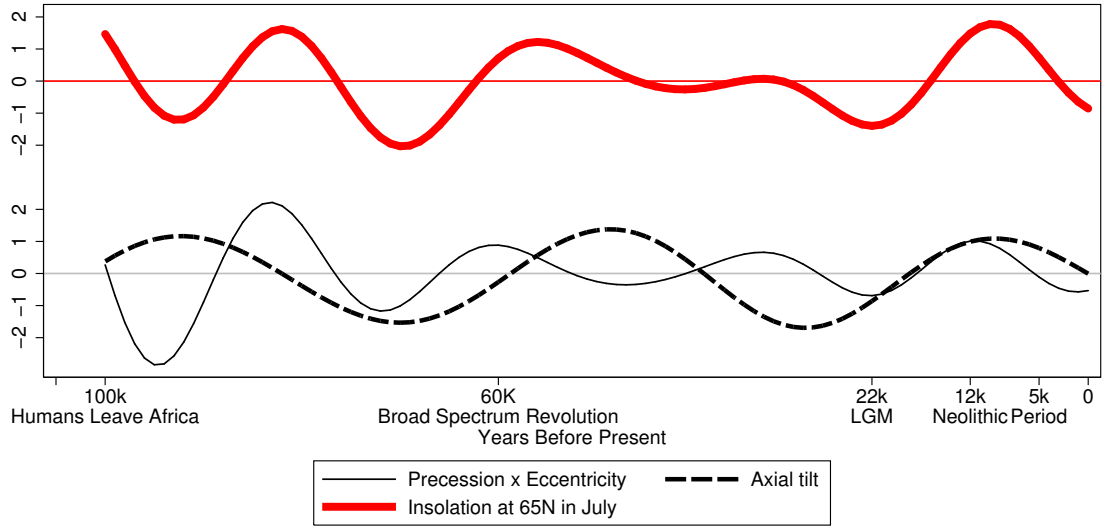

Figure 1.2: Three parameters combine to determine insolation seasonality in the northern hemisphere. During the Early Neolithic, these three cycles peaked simultaneously for the first time in over 100,000 years (black, I show the effects of axial tilt, and the combined effect of precession and eccentricity). As a result, the northern hemisphere was more seasonal then it had been at any point since humans left Africa. Data from Berger (1992). Seasonality conditions at 65° N (red) are indicative of those in the rest of northern hemisphere.

The change in seasonality was also responsible for the end of the last Ice Age. The warm summers caused ice to melt, while the cold winters actually inhibited snowfall. As a result, the glaciers which covered wide areas of the northern hemisphere retreated, raising global temperatures by 7 to 8° C. The spread of hunter-gatherers occurred against the backdrop of the Late Pleistocene glaciation (120,000 to 13,000 BP), during which average temperatures were up to $8\,^{\circ}\mathrm{C}$ lower than today. Since agriculture was invented shortly after the start of the current warm period (the Holocene) it is tempting to assume that agriculture was a response to change climate averages. Childe (1935) proposed that as the glaciation came to a close, drier conditions in the Fertile Crescent forced humans to concentrate in a limited number of oasis with a reliable supply of freshwater. These narrow confines would have provided the right incentives for agricultural adoption. Wright (1970) took the opposite tack, arguing that more *favorable* conditions at the end of the last Ice Age had allowed easily domesticable species such as wheat, barley and oats to colonize the Taurus-Zagros mountain arc, where agriculture would eventually emerge. While this explanation fits the evidence from the Middle East, it is unlikely that the global invention of agriculture was caused by changes in average climate. If the theory were true, we would expect farming to be developed in very warm locations. Instead, agriculture was invented in climates as different as those of Sub-Saharan Africa (hot and dry), Southern China (hot and wet), the Andes (cold and dry) and Eastern North America (cold and wet). While most of these locations did become warmer in the early Holocene, humans living elsewhere had experienced similarly pleasant conditions for tens of thousands of years.

1.3 Model

In this section, I model the incentives faced by a single band of hunter-gatherers, as it adapts its life strategy to a changing environment. First, I will present a simple static model in which population size is constant. I assume a pure endowment economy, in which the underlying resource base varies across space and time. I find that low seasonality makes the band choose nomadism, precluding the development of agriculture. However, a sufficiently large increase in seasonality will cause the band to prefer settlement, catalyzing the development of farming. When the band becomes sedentary, it loses access to some resources that could only be accessed nomadically, but the ability to smooth consumption through storage more than makes up for the loss in consumption per capita.

I then extend this basic intuition into a dynamic setting, in which population evolves endogenously. I modify the basic Malthusian setup by assuming that fertility is increasing in consumption per capita but decreasing in consumption sea-

sonality. Nomads are unable to perfectly smooth their consumption, resulting in lower net fertility, and higher consumption per capita in equilibrium. Settlers, in contrast, are able to perfectly smooth consumption through storage. Their stable diet ensures the maximum possible fertility, so that in equilibrium they have the lowest consumption per capita possible.

1.3.1 Setup

The unit agent of the model is a band, which has exclusive control over a specific territory. There are two locations in the territory of the band, the Hill and the Plain, and two months in the year, December and July. The Hill provides an endowment of $1 + \sigma$ in July, and $1 - \sigma$ in December, while the Plain provides no food in July and $1 - \sigma + \gamma$ units of food in the Winter. The parameter σ indicates the amount of climate seasonality in the region, while γ represents how much extra food is available in the Plain in December.

Table 1.1: Endowments of each location in each season

	July	December
Hill	$1 + \sigma$	$1 - \sigma$
Plain	0	$1 - \sigma + \gamma$

For example, we could imagine that the general area has a warm but dry summer but a cold and rainy winter. Hills are usually colder than the surrounding plain, but they receive more rainfall. Therefore, we would expect that in the summer, the hills will be hot and wet, plants will grow well, and food availability will be very high. In winter however, the hill is too cold and will provide much less food. In the plains, the lack of rainfall make food extremely hard to find in summer. In winter, the plains are warm enough and wet enough and temporarily provide more food than the hills. This general pattern can be adapted to model a variety of resource availability regimes.

The band has a log utility function defined over consumption per capita in each period

$$U = \log(c_J) + \log(c_D) \tag{1.1}$$

1.3.2 Static model

I first compare the outcomes from the two strategies in a static model, in which I assume that population size is fixed. If the band is nomadic, it will spend each month in whichever ecosystem is most abundant at the time. It will therefore choose to spend July on the Hill but will descend onto the Plain in December. Its mobility will allow it to smooth its consumption geographically but will prevent it from storing food. The settled band will instead settle in the Hill (which has the highest aggregate endowment), and it will be able to perfectly smooth its consumption through storage. However, it will no longer be able to access the resources of the Plain, so aggregate consumption will necessarily be lower.

Specifically, the Nomadic band will consume C_N, and the Settled band will consume C_S, where

$$C_N = \{1 + \sigma, 1 - \sigma + \gamma\} \tag{1.2}$$
$$C_S = \{1, 1\} \tag{1.3}$$

Each consumption profile shows first consumption July, and then consumption in December. Utilities from the two strategies are simply:

$$U(N) = \ln(1 + \sigma) + \ln(1 - \sigma + \gamma) \tag{1.4}$$
$$U(S) = 0 \tag{1.5}$$

The utility of the settlers is always zero, but that of the Nomads depends on the environmental parameters. A higher σ will lower nomadic utility, while a higher γ will increase it. These relationships are represented in Figures 1.3 and 1.4.

For the band to be indifferent between the two strategies, it must be true that:

$$\sigma = \frac{\gamma + \sqrt{4\gamma + \gamma^2}}{2} \tag{1.6}$$

The higher the level of γ is, the higher seasonality must be before the band is willing to switch to sedentism. From these results, we can therefore reach the following conclusions:

Proposition 1. *In the static model we find that:*

1. *If the climate is not very seasonal (high σ, and the band has access to uncorrelated ecosystems (high γ), nomadism will be optimal.*

2. *An increase in seasonality can cause settlement to become optimal.*

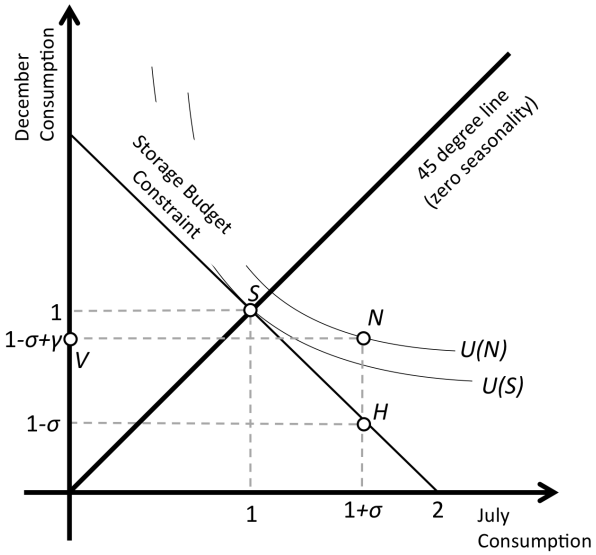

Figure 1.3: Circles H and V represent the endowments of Hill and Valley respectively. The Nomads are able to always reside in the best territory during each month, and therefore enjoys a consumption profile of N. The Settler can only harvest the resources of H but can smooth consumption costlessly. It will therefore equalize its consumption across periods and achieve a consumption profile of S. In this case, seasonality σ is low, and the usefulness of mobility γ is high. The band, therefore, has a higher utility if it remains nomadic.

3. *The higher γ is, the more seasonal climate must be before settlement becomes optimal.*

4. *Consumption per capita will be lower after the transition.*

1.3.3 Dynamic Model

I now add endogenous population growth to show that the instantaneous results of the static model also hold in the long run. The population dynamic of the band is determined by its consumption profile. Specifically, net individual fertility ϕ is a weighted average of consumption per capita in both months, with the weighting favoring consumption per capita in the scarcest period:

$$\phi = \alpha \max(c_J, c_D) + (1 - \alpha) \min(c_J, c_D) \tag{1.7}$$
$$0 < \alpha < 0.5$$

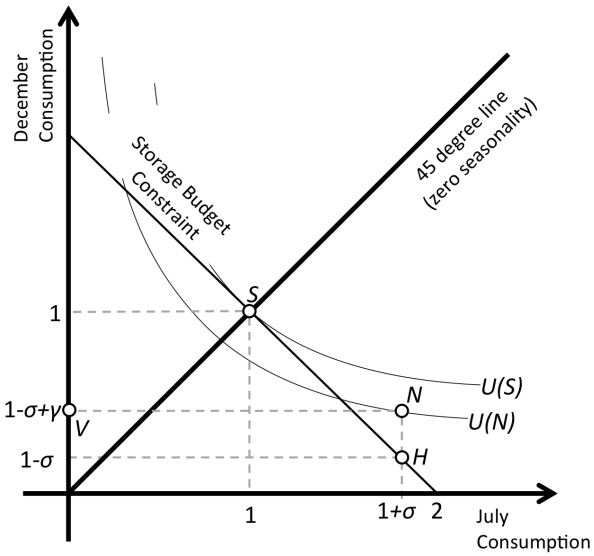

Figure 1.4: Now σ is higher, and γ is lower. A nomadic band would now be exposed to high consumption seasonality, so that utility is now higher if it switches to settlement. This is true despite settlement having a lower consumption per capita.

If α were equal to 0, then fertility would be equal to the minimum of consumption per capita in both months (the production process for children would have a Leontief form), while if α were equal to 0.5 fertility would only depend on average consumption per capita, and the entire model would collapse to the standard Malthusian case. I assume that the fertility dynamic lies somewhere in between these two extremes: higher average consumption per capita will increase fertility, but for any average consumption per capita, higher consumption seasonality will depress fertility (Almond and Mazumder, 2008). This dynamic could indifferently arise from either biological constraints on a population reproducing *ad libitum*, or else be the result of optimizing behavior by a population that has control over its fertility, and prefers more children when food supply is abundant and stable.

The first step is to calculate the equilibrium levels of population for each lifestyle. Population size will be stable if:

$$\begin{aligned} 1 &= \phi \\ 1 &= \alpha \frac{C_J}{P_N} + (1-\alpha)\frac{C_D}{P_N} \end{aligned} \tag{1.8}$$

Where C_X is aggregate consumption of the band in month X, and P_N is the population of the band. By substituting the appropriate values we find that the

equilibrium level of population for the two lifestyles will be:

$$P_N^\star = 1 - \sigma(1-2\alpha) + \gamma(1-\alpha) \tag{1.9}$$
$$P_S^\star = 1 \tag{1.10}$$

dividing the endowments by the equilibrium level of population, we can thus derive consumption per capita in the long run for both strategies in equilibrium:

$$c_N^\star = \left\{ \frac{1-\sigma+\gamma}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)}, \frac{1+\sigma}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)} \right\} \tag{1.11}$$
$$c_S^\star = \{1, 1\} \tag{1.12}$$

Settlers, irrespective of environmental parameters, will always consume one unit of food per capita, per month: their ability to smooth consumption ensures that the standard Malthusian result prevails. In contrast, Nomads suffer a population penalty due to the seasonality in their diet. This ensures that consumption per capita is an increasing function of their diet seasonality.

The consumption profiles for both strategies allow us to derive the respective equilibrium levels of utility:

$$U_N^\star = \log\left(\frac{1-\sigma+\gamma}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)}\right) + \log\left(\frac{1+\sigma}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)}\right) \tag{1.13}$$
$$U_S^\star = 0 \tag{1.14}$$

Nomadism will be optimal *in the long run* whenever $U_S^\star > U_N^\star$, leading to the long run threshold condition:

$$\sigma = \frac{1+\gamma(1-2\alpha+\alpha^2)-2\alpha}{1-2\alpha+\alpha^2} \tag{1.15}$$

The higher γ is, the higher σ must be for settlement to provide a higher utility than nomadism.

However, the long-run equilibrium outcomes of settlement could not be guessed by the populations that abandoned nomadism. For this adaptation to become widespread, it is important that settlement is also better than nomadism soon after the transition, i.e. before population size adjusts to the new equilibrium. The short run

$$c_S^- = \frac{C_S}{P_N^\star} c_S^- \quad = \left\{ \frac{1}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)}, \frac{1}{1-\sigma(1-2\alpha)+\gamma(1-\alpha)} \right\} \tag{1.16}$$

Settlement will increase utility in the short run if $c_S^- > c_N^\star$. This disequation is simply the condition for optimality derived for the static model, scaled by a constant (the equilibrium population size of nomads). Since preferences are homothetic, we know that the optimality condition will be the same as in Equation 1.6.

$$\sigma = \frac{\gamma + \sqrt{4\gamma + \gamma^2}}{2} \tag{1.17}$$

These results can be condensed in the following proposition, which parallels the statements of Proposition 1

Proposition 2. *In the dynamic model we find that:*

1. *If the climate is not very seasonal(high σ), and the band has access to uncorrelated ecosystems (high γ), nomadism will be optimal both in the short run and in the long run.*

2. *An increase in seasonality can cause settlement to be better than nomadism both in the short and long run.*

3. *The higher γ is, the more seasonal climate must be before settlement becomes optimal.*

4. *Consumption per capita will be lower after the transition and will remain lower even after population adjusts.*

1.3.4 Predictions

The result of the models generate a number of empirical predictions, which can be verified using the archaeological and paleoclimatic record for the invention and spread of agriculture.

1. If a nomadic band becomes settled, average consumption per capita will immediately decrease due to the loss of access to the December Refuge endowment, but consumption seasonality will disappear.

2. In the long run, average consumption per capita of the settlers will remain lower than during nomadism (since consumption seasonality no longer depresses fertility).

3. For any level of γ, a sufficiently large increase in seasonality can make settlement optimal both in the short run and in the long run.

4. The higher γ is, the higher σ will have to be before settlement becomes optimal.

Thus we would expect settlement to be adopted en masse where seasonality is high and correlated across locations. These are precisely the conditions that became common shortly before agriculture appeared.

1.3.5 Extensions

Endogenous agricultural progress

In the main model formulation we have abstracted from actually describing the process of agricultural adoption and diffusion, other than arguing that it would be a natural outcome of a population becoming sedentary under favorable conditions. The underlying reasoning is that if a sufficiently large number of human populations found it optimal to become sedentary even in the absence of a viable agricultural technology, some of them would inevitably be located in areas that favored the development of farming. In this subsection we will sketch a model describing just what those favorable conditions are, and what areas are likely to have them.

Plants are defined by possessing the ability to photosynthesize sugars starting from water, CO_2, and sunlight. These sugars are then used as both sources of chemical energy, and as building blocks for the long cellulose chains that make up most of the structure of plants. The ultimate goal of this process is the reproduction of the plant itself, most commonly through the production of seeds, but sometimes through vegetative reproduction. Humans can gain nourishment from plants in a variety of ways. Unfortunately, humans are unable to extract energy from cellulose itself, which composes the vast majority of the bulk of most plants. However, herbivores can digest it, and humans are very well equipped to both kill them and digest them.

The most straightforward way of extracting energy from plants directly is is by consuming the fruits, which are the only structure in plants that are made expressly for being eaten by animals — with the purpose of dispersing the seeds they contains. While this is in principle incredibly convenient, not all plants produce fruits edible by humans, and those that do are also visited by many other animals,

some of which are much better suited to this activity (for example by being better climbers, or being able to fly). Add that outside of the tropics most fruits are available in the same brief season, and that the vast majority is exceptionally hard to store for longer than a few weeks, and it is clear that wild fruits are for the most part a welcome treat, rather than a viable staple crop.

Another way of extracting energy through plants is by consuming sugary, starchy or oily parts of plants. These chemicals are much more energy dense than cellulose itself, and are produced when the plant wants to store energy in a compact and stable form, usually because it wants to gives its seeds an initial and amount of energy to ensure it can produce the first leaves, which will allow it to use sunlight to produce its own energy – this is the case of cereal crops. Alternatively, some plants produce these compounds to store energy to survive a period of dormancy and avoid inclement weather or episodic stresses – such is the case of the potato, the turnip, and the carrot.

Such energy rich structures are also the preferred targets of parasites, and over millions of years plants have developed a bewildering arrays of poisons, shells, spines, and camouflage to protect these high value targets. One of the crucial advantages of humans compared to other plant parasites was our ability to efficiently break down the mechanical protections with tools, and neutralize the poisons by appropriate processing techniques (chiefly through cooking, and repeated washings to leach out the toxic chemicals). These adaptations, together with a reasonably accommodating digestive system, allow us to consume a wide variety of plants. Indeed researchers estimate that approximately 7,000 of the worlds 320,000 vascular plants have been cultivated or gathered from the wild by humans for use as food, and about 30,000 have at least some parts which could be consumed by humans.

Nonetheless, in each ecosystem it a few plants have the right combination of abundance, proportion of mass which is edible, convenience for gathering and processing, and palatability, which made them the staple foods for hunter-gatherers. The purpose of agriculture is therefore to increase the abundance of these species, and to increase the productivity of each individual plant in terms of edible fraction of biomass. I will now specialize the model described in the main model section to describe a few ways in which humans could use agriculture to increase the food available to them and call attention to the conditions that would make it more easier to implement.

So far we assumed that a given level of food was exogenously available to the population inhabiting a given area. The first thing we will do is to specify a simple production function for the food producing plants:

$$F_{it} = \min(W_{it}, L_{it}) \tag{1.18}$$

where F s the amount of food produced, W is the amount of water, and L is the amount of light, i is the ecosystem and t is the season. The choice of the Leontief production function is founded in the biological observation that typically plants and other biological process are limited by the scarcest available resource, with very little substitution allowed (International Fertilizer Development Center, 1998) [1]. The baseline model implicitly assumed that there were fixed amounts of water and light available to food producing plants in each ecosystem, and in each season, thus determining the respective food endowments available.

Let us further assume that the population inhabiting the ecosystem has an agricultural technology which allows it to increase the amount of water and light available in fixed proportions by expending effort e, according to

$$\Delta W = \omega e \tag{1.19}$$

$$\Delta L = \lambda e \tag{1.20}$$

That is, each unit of effort generates ω extra units of water and λ extra units of light. For example, watering plants from a bucket would be an inefficient way of increasing the amount of water, (low ω, and zero λ), whiled digging an irrigation ditch would have high omega and zero lambda. Clearing away a tree which didn't produce edible fruits, would increase light by a lot, and also increase a bit the amount of water available.

We could say that effort reduces utility, but a more straightforward way of analyzing the cost-benefit relationship would be limit analysis to the energy cost of effort, and assuming that each unit of effort is equivalent to reducing the food available for consumption by one unit. A further assumption is that the effort performed in one season only gives results in the next, since plant takes time to grow and reach maturity. We will further assume that, at least initially, the actual effort that can be spent in agriculture is essentially experimental, and therefore negligibly small relative to both the amount of time spent hunting and gathering, and relative to the aggregate endowments of water and light.

Thus, the extra food available because of agricultural effort will be equal to:

$$F_{it} = \min(W_{it} + \omega e_{i,t-1}, L_{it}\lambda e_{i,t-1}) \tag{1.21}$$

Four results immediately stand out. First, regardless of the level of λ and ω, agricultural technology requires sedentism in order to give any benefit. Quite simply, a nomadic population will not be present to harvest the fruits of their labor, when they in fact become available. Second, no agriculture will ever be viable if both ω and λ are less then 1, since the returns will not cover the energy cost of effort.

[1] This is true mainly for each specific species. Different plants are optimized for different ratios of sunlight, water, nutrients, etc

Third, the net returns from agriculture depend also on the initial endowments of water and light. For example, if we imagine a population that has at its disposal a pure irrigation technology with $\omega > 1$, but $\lambda = 0$. If $L > W$, then this technology will increase the amount of the binding resource available to the the edible plants, and is therefore viable and farming will in fact increase the amount of food available. However, if $W > L$, then in fact plant growth is sunlight-limited, increasing the amount of water will not increase the amount of food available at all, and any effort expended in agriculture will be wasted. Fourth, it is important to keep in mind that even if a given population possesses an agricultural technology capable of increasing the amount of food available, the payoff might be too small to justify becoming sedentary, if the considerations discussed in the main model (seasonality and geographic heterogeneity) recommend remaining nomadic. However as the level of agricultural technology increases, the amount of seasonality necessary to make sedentarism optimal decreases. Naturally real plants need much more than water and light, but the analysis can be generalized to control of erosion, removal of excess water, parasite control, provision of nitrogen, etc.

To round off the analysis, I will now simply assume that there exists learning by doing in agricultural technology, which means that ω and λ depend on the cumulative amount of agricultural activity performed by populations that are in contact with each other. The stage is now set to provide a concise explanation for the observed pattern of Neolithic invention and spread.

Late Paleolithic people possessed in many cases a rudimentary agricultural technology, which was capable of providing a positive return on effort, but not so great as to justify abandoning nomadism as long as seasonality remained moderate. When seasonality increased, affected populations became sedentary in order to store food, and also were now in a position to take advantage of their primitive agricultural techniques. As they gained experience through generations, and especially after domesticated plant varieties ensued, the returns to agricultural effort increased. To the extent that such progress was transmissible, some of their neighbors would now find it optimal to become sedentary and farmer, even though they had not experience an increase in seasonality large enough to justify abandoning nomadism for storage alone. Nonetheless, the higher the seasonality a group was exposed to, the lower the level of agricultural technology that would make the combined agriculture-storage bundle worth switching.

Evolutionary Formulation

The model presented so far conforms to the standard economic assumption of a fixed utility function that agents maximize. While this choice simplifies comparison to the existing set of economic models for the Neolithic revolution, it is reasonable to assume that over the timespans discussed in this paper the distribu-

tion of utility functions both within and across different population groups is itself is subject to change, through both random mutation, and selective pressures — in short - through evolutionary forces.

In practical terms, by the late Paleolithic times, the most likely factor to cause the demise of a specific human population was conflict with other human groups. Therefore, we would expect that, over time, traits that made success in such conflicts more likely would dominate. While obviously there are innumerable such traits which affect success one way or another, two important and obvious drivers of defensive ability D are population size N, and food consumption per capita N(which will impact individual strength). Therefore:

$$D = d(N \uparrow, c \uparrow) \tag{1.22}$$

Since we are considering a simple endowment economy it should be clear that average food consumption is simply the average endowment divided by the population size, which is a function of seasonality σ. Therefore:

$$D = \mathrm{d}(N \uparrow, C/N \uparrow)D \qquad = \mathrm{d}(\mathrm{n}(\sigma) \uparrow, C/\mathrm{n}(\sigma) \uparrow) \tag{1.23}$$

Where $\mathrm{n}(\cdot)$ is the function linking population size to seasonality. In words, an increase in seasonality will tend to decrease defensive ability by reducing the population, but will also increase it by improving average consumption per capita. Whether in practice the effect is positive ore negative will depend on the precise shape of the defense function. In practice it is easy to imagine that the best use of a given amount of food is somewhat balanced: neither too many semi-starved troops, nor a single overfed champion subject to decreasing returns to extra food.

The important point is that levels of seasonality map to levels of defensive ability. Under different circumstances, the same utility function might lead a group to selecting a level of seasonality that achieves the optimal defensive ability, while in others it might not. Ultimately populations with an adaptive utility function will tend to win out on those that have a less adaptive one. For example, imagine two neighboring nomadic population inhabits a location which both have close to the optimal level of seasonality for defensive purposes, i.e. the seasonality reduces population just enough to reach the efficient tradeoff of population size and strength. If one of them were so risk averse as to decide to become sedentary under these conditions, it would move to a no seasonality equilibrium, in which it would have too large a population of very weak individuals. If relations ever soured between these neighbors, and push came to shove, we would imagine that the inefficiently risk averse population would have the wore of it. If seasonality were to increase greatly, then the situation would be reversed. The risk would be

that the excessive seasonality would reduce population to such an extent that the few group members left, regardless of how well fed they were, would be unable to beat off a horde of a settlers. Clearly the greater danger in this case would be of being too risk neutral, and therefore deciding to not become settlers. It should be clear that over time, the bulk of the population would be made up of individuals with intermediate levels of risk aversion. Figure 1.7 clarifies these relationships.

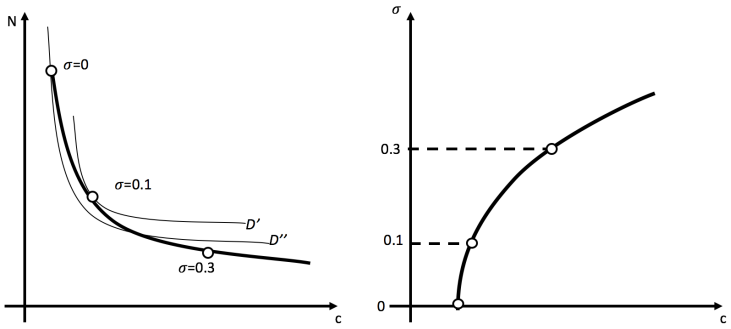

Figure 1.5: Left panel: the thick line shows how population N size AND consumption per capita c evolve as seasonality σ increases. When there is no seasonality, population is maximized, and consumption per capita is at it's minimum possible level. As seasonality increases, the population size decreases, but consumption per capita increases. Which combination offers the best defense? The lines D' and D'' show iso-defense lines, increasing towards the upper right corner. The maximum possible defense value is realized when $\sigma = 0.1$. More seasonality results in very strong individuals, but too few of them, while less seasonality results in a very large population of inefficiently weak individuals. Right Panel: whatever attitudes the population has towards risk, in the long run they have to be compatible with choosing $\sigma = 0.1$ (and the associated higher consumption per capita) over $\sigma = 0$, but $\sigma = 0$ over $\sigma = 0.3$.

Relaxing the endowment economy assumption

The baseline model assumed a pure endowment economy, that is, a specific amount of food was simply made available to whoever happened to be around at a given time. In practice, hunter gatherers have to work for their food like everyone else. However, I will now show that this is assumption is almost entirely transparent.

Let's assume that hunter gatherers can procure food through a Cobb Douglas function of their own labor, and the level of natural resources in the environment.

$$Y = N^{\alpha} R^{1-\alpha} \tag{1.24}$$

let us consider two cases: fixed population size, and endogenous population size.

If the population size is fixed at some level $N^{\star}$, then the result is trivial. Any level of resource will map directly to an outcome level of food:

It is therefore entirely equivalent to say that a given environment provides an endowment of E, or a resource level of $R = (EN^{-\alpha})^{\frac{1}{1-\alpha}}$

But of course, over the timeframes analyzed in this study population size can adjust. For once, the analysis of this case is actually simpler. Let's assume that the population growth rate is increasing in consumption per capita, and one unit of consumption per capita is necessary to maintain the population level constant. The amount of food produced will depend on the current population size, as well as on the level of resources, but in a population equilibrium it will be true that population will be equal to the actual amount of food gathered (because consumption per capita must be 1) , which itself will be equal to the resource level (because of the properties of the Cobb-Douglas). So

$$Y = N = R \tag{1.25}$$

Another potential criticism for the model would note that the amount of food available in one ecosystem is not exogenous, but depends on how many resources have been taken out of the ecosystem in the past. One simple way of modeling this is to first assume that the reproduction rate of the food species is equal to $g = (1 - n)n^{\star}$, where n is the current population of the food species and $n^{\star}$ is the carrying capacity of the environment, or the population level that will result in 0 population growth. The growth in levels will then be

$$g = n(1 - n) \tag{1.26}$$

which has the plot:

This implies that the optimal population level for the resource is $0.5N^{\star}$, as this is the population level that results in the highest possible sustainable harvest. Keeping the food species population at a lower level would result in either a smaller harvest, or a continuously dwindling stock of the food species, followed by a smaller harvest. If instead the stock of the food species were kept at a higher level, intraspecific competition for resources would ensure in any case a lower maximum sustainable harvest.

One interesting consequence is that under some parameterizations, the population will have to consciously reduce gathering effort to avoid overtaxing it's environments. For example, let's assume that the production function has equal weighting on labor and resources ($\alpha = 0.5$), and that the carrying capacity of the environment is 1. Then the optimal stock of the natural resource is 0.5, and the maximum sustainable harvest size is 0.25. But if the population of the hunter gatherers were 0.25 (that is, if it were in a theoretically sustainable population

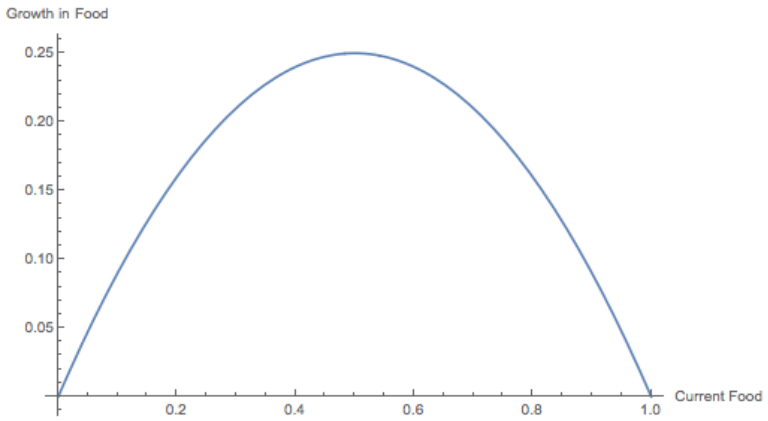

Figure 1.6: Maximum growth in the food species is obtained at half of carrying capacity. That is the maximum amount that can be harvested sustainably.

equilibrium), then the resulting harvest size will be 0.35..., which means both that a) consumption per capita is greater than 1, and therefore population will rise, and b) the naturally occurring harvest will be larger than the maximum sustainable harvest, which will lead eventually to ecosystem depletion.

Therefore, the only way hunter-gatherers will be able to maintain the maximum sustainable harvest (and population size), will be by consciously restricting the amount they harvest each year. Since Paleolithic populations cannot employ sophisticated ecosystem surveys, biological modeling, and quota systems (indeed, we struggle with this even today), we can expect that they might develop simpler behavioral approximations, such as evolving a high value for leisure, or considering certain hunting grounds or periods as off limits. This would reconcile well with the observation that hunter-gatherers had considerable leisure time, or the common complaint amongst missionaries and assorted colonialists that 'the natives were lazy'. In reality, any further activity would have run the risk of eventually precipitating environmental collapse. It is of course possible that such attitudes could persist after their usefulness has passed, though I am not testing this in this paper.

One ecosystem dominates the other

In the main model, we have assumed that one environment has more food in one season, and the other has more food in the other. Of course there are many situations in which one environment to naturally have more food than another in every season. Is nomadism useless under these rather common conditions? Not necessarily.

All that is necessary is to acknowledge that the band is under no obligation to

spend equal amounts of time in both ecosystems. For example, we can assume that there are still two ecosystems (Hill and Valley), but there are now four seasons in two groupings. The first one is Spring and Summer, and the second one is Fall, and Winter. Let's further assume that in the Hill there are 150 units of food for Spring and Summer *combined*, while in Fall and Winter there are 50 units *combined*. In the valley there are 0 units of food in Spring and Summer, and 40 in Fall and Winter. The idea of grouping season is that the population can (if it so chooses), harvest all of the 150 units of food over the two Spring in Summer, or just in one of the two. But if it harvests all of the food in Spring, there will be nothing left for the Summer. However the 50 units available for Fall and Winter are independent of Spring and Summer (different species are involved). Note that while the Hill in principle dominates the Valley in all seasons, a Nomad can still harvest more food by moving to the Hill, Harvesting 75 units in Spring, and 75 in Summer, then harvesting all 50 units in the Valley in the Fall, and finally moving to the Valley to harvest the 40 units of food available there. By doing so it will harvest 240 units of food over the course of the year, instead of the only 200 if it were sedentary.

This admittedly simple version of the model is simply designed to show that if resources remain harvestable for a period of time which is significant when compared to the speed of movement, regions that have an absolute disadvantage in all seasons can have a *comparative* advantage in a certain season, and Nomadism can still be an effective coping strategy.

Coping with inter-annual volatility

Up to this point, the only type of volatility I have discussed is climate seasonality, that is entirely predictable and takes place within a year. The fact that it is predictable, and the relatively high frequency of the oscillation meant that an effective storage technology could essentially eliminate this problem. However this was not the only type of resource variation faced by our ancestors, nor was it the only kind that storage could buffer (though less effectively than it could do for seasonality).

An obvious type of variation to think about is the random variation in the food availability from one year to the next. Such variations are most commonly linked to climate, but can also be influenced by fires, pests or epidemics in prey animals. While storage can help smooth this type of volatility, its efficiency in doing so is much lower than in the seasonality case since the length of time food must be stored is longer, which creates more chances for spoilage, and the magnitude of the change is in general unpredictable, so that food will sometimes be stored needlessly (if there are several good years in a row), and sometimes even a sound stockpile will not be enough (if there are several bad years in a row).

I will now present a simple simulation to help us think about these topics.

A group, with starting population n inhabits a territory over several periods, each of which represents a year. Each period the territory confers on the inhabitants a specific endowment of food, which is distributed uniformly on the the 0.5 to 1.5 interval. In the case without storage, the group consumes the period's endowment directly, and the population evolves according to

$$N_t = Min(C_t, N_t - 1\phi) \tag{1.27}$$

where N_t is the population at time t, C_t is conscumption at time t, and ϕ is a fertility parameter which is a little greater than one. The interpretation is straightforward. Each individual requires one unit of food to survive. If there is enough food for everybody, then the population increases by ϕ, while if there is not enough food the population decreases to the maximum level sustainable. We can iterate this formula backwards by one step, and we get

$$\begin{aligned} N_t &= \min(C_t, \min(C_{t-1}, N_{t-2}\phi)\phi) && (1.28)\\ &= \min(C_t, C_{t-1}\phi, N_{t-2}\phi^2) && (1.29)\\ &= \min(C_t, C_{t-1}\phi, C_{t-2}\phi^2, C_{t-3}\phi^3, C_{t-4}\phi^4, ...) && (1.30) \end{aligned}$$

Since ϕ^n balloons up to a very large value, and in this implementation there is a firm lower bound to the possible amount of food, we can safely ignore the terminal N_t, since it will never be binding in the limit if ϕ is greater than zero.

If on the other hand, the population is able to store food, then the degree of consumption smoothing will depend on the cost of storage, whether there exists a hard limit to the amount of food stored, and how much the population dislikes going hungry. However the basic comparative statics can be made clear by simply assuming a naive utility function. Specifically I will assume that the band only cares about consumption up to the point where nobody starves and current fertility is maximized, and derives no utility from further consumption. Under this simplistic, but illustrative example, the group can do no better than consume exactly what they need to attain their satiation point, and store all the rest.

The following graphs compare the results.

The restriction on fertility generates a sawtooth pattern. Famines kill off part of the population, which then takes time to rebound. In the interim, the survivors and their offspring enjoy higher consumption per capita than would otherwise be possible. Eventually the population grows enough, and a sufficiently unlucky climatic realization takes place, so that another famine again kills part of the population, and the entire cycle repeats.

The addition of storage is essentially equivalent to reducing the degree of volatility. The outcomes is population size is on average higher, and consumption

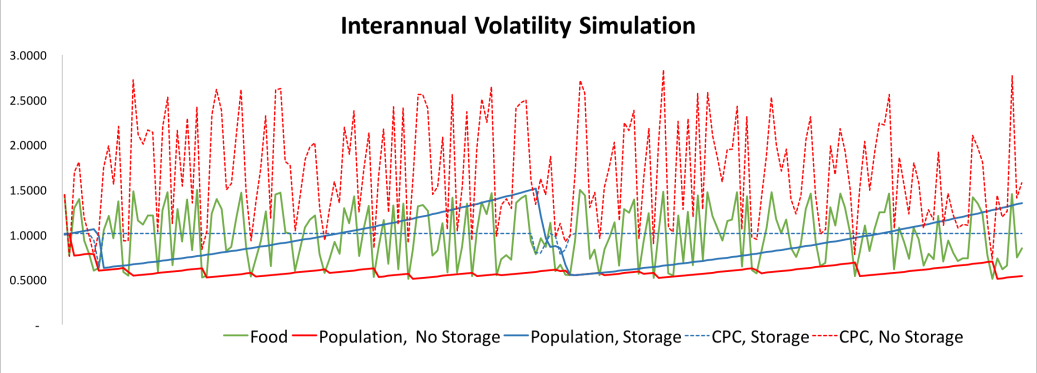

Figure 1.7: The amount of food available in each period is uniformly distributed between 0.5 and 1.5, expressed in the amount of food each individual needs to eat in each period to survive. In the case without storage, the population grows at 1.01% per year if Food is greater than current population. If Food is lower than current population, just enough people die so that the rest can survive. If storage is introduced, the group saves all food in excess of current needs, and draws down it's reserves when there is insufficient food. Note how population is on average lower without storage, due to the more frequent famines. Conversely, consumption per capita is higher when storage is absent, due to the same amount of food being spread amongst less mouths on average.

per capita therefore lower. We therefore see that while storage is less effective at controlling inter-annual volatility than predictable seasonality, the character of the results is essentially the same.

1.4 Conclusion

In this chapter, I have presented a new theory for the Neolithic Revolution: an exogenous increase in consumption seasonality made reduced the ability of nomads to smooth their consumption, and made storage necessary. Since storage essentially requires settlement, an unprecedented number of people were now in a position to observe plant growth throughout the entire year, and therefore were excellently placed to develop farming techniques.

This theory is able to account for the main observable stylized facts, namely: the reduction in consumption per capita, the clustered timing of the multiple invention of agriculture, and the broad distribution of latitudes in which the transition occurred. Further it also generates specific predictions on the climatic conditions that should characterize the independent inventors, as well as the fastest adopters. Another prediction of the model is that locations that have a lot geographic heterogeneity accessible to nomads, should delay adoption. The model also makes

predictions as to the health patterns we expect to observe across the transition. Nomadic hunter gatherers should be tall, but show evidence of regular famines, while settled farmers should be short, but appear to have eaten regularly. In the next Chapter, we will be testing the predictions of the model empirically.

The theory has the ulterior advantage of being compatible with a range of other existing theories for the Neolithic. For example, while it has a strong hypothesis for the incentives that made settlement necessary, and farming desirable, it is entirely agnostic as to how exactly specific human groups would discover and improve particular farming techniques. E.g., It doesn't matter whether a particular group found out that seeds it had thrown out in the garbage were sprouting, or whether ancestral botanical knowledge of plant biology finally found a situation in which it could be applied to farming: both of these paths would be infinitely more probable once a population had become sedentary.

Similarly, the theory is compatible with other more "macro" theories for the Neolithic. It could very well be that, amongst seasonal places, those with the easiest species to domesticate, or those who's elites had the largest appetite for expropriated harvest, would be amongst the most enthusiastic adopters of farming. Such theories may be acting in parallel to my own – or not. The advantage of seasonality is that it can explain many different facets of the Neolithic which up to now had required different theories.

Chapter 2

SEASONALITY, STORAGE, AND AGRICULTURE: EXAMINING THE EVIDENCE

2.1 Introduction

The previous chapter has shown that introducing seasonality can explain the main stylized facts of the Neolithic Revolution as a global phenomenon in a single parsimonious theory. Specifically, the model I developed showed that it was perfectly rational for a population exposed to a large spike in seasonality to abandon their previous nomadic lifestyle, and become sedentary in order to store food. Just as importantly, it also showed that it was perfectly possible for such a decision to be rational even if it resulted in a reduction in consumption per capita. In this chapter, I will test the model empirically, by first replicating the most advanced approaches taken in the literature on the Neolithic revolution, and showing that seasonality is a better candidate as an explanatory variable, and then introducing two entirely novel empirical tests, which are highly specific for the seasonality hypothesis.

The first part of my analysis is performed at the global scale, using a panel dataset of reconstructed climates, covering the entire world for the past 22,000 years, using agriculture adoption data from Putterman and Trainor (2006), and invention data from Purugganan and Fuller (2009) I will first use a simple linear regression to show that seasonality explains the basic where an when of the adoption of agriculture on the global scale. I do not attribute any causal value to this analysis, but simply use it to show that seasonality is a good statistical predictor of when we observe agriculture to appear in the archaeological record. I will then disaggregate this result between the probability of independently discovering agriculture, and the speed with which it is then adopted. A preview of my results

can be seen in Figure 2.1.

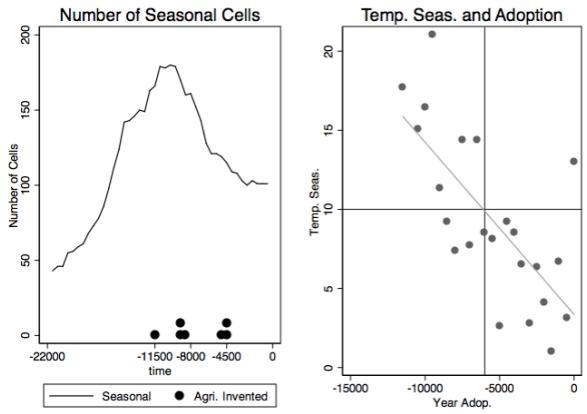

Figure 2.1: Left panel: climate became more seasonal shortly before agriculture was invented multiple times. Right panel: binned scatterplot of temperature seasonality and adoption; early adopters tend to be highly seasonal, and vice versa.

The invention of agriculture is a tantalizing historical event for scholars of technology. On the one hand, it is the most prominent instance in which a crucial technological advance did not originate from a single time and place, but was instead developed in parallel multiple times. This means that unlike the steam engine or the printing press, it is possible to compare the various places that invented agriculture, to check whether they all had a certain characteristic or not. Unfortunately, depending who you ask and precisely how you word the question, the number of independent inventions is between seven and two dozen. This necessarily limits the number of controls that can realistically be included in the analysis. However, I find that seasonality performs about as well as could be expected of any regressor under these trying circumstances.

I then turn my attention to the process of diffusion through which farming came to dominate food production on all continents inhabited by humans. As discussed in Chapter 1, storage and farming were effectively a bundled set of technologies, that were either adopted together or not at all. The more seasonal the local climate was, the more valuable the storage component of the bundle was, which would have led the local population to switch sooner, even if the available agricultural technology was still relatively poor. Where climate was not seasonal, farming would be adopted only when farming techniques had advanced enough to make the switch worthwhile on their own.

A naive implementation would simply regress the year of adoption on the level of climate seasonality, controlling for the distance to the closest location of invention. The problem with that approach is that it would introduces a lot of

noise due to the varying speed of adoption of the "in between" locations. For example, for the case of agriculture starting in the Fertile Crescent, it would be unfair to expect Afghanistan and Greece to adopt at the same time (even if local conditions were identical), since Greece can be reached across the fertile northern shore of the Mediterranean, while to reach Afghanistan one must cross several hundred kilometers of desert. To use an analogy, it would be like evaluating two runners that are both fourth in a relay race by their finishing time, even though one runner was preceded by world class runners, while the other was preceded by middle school team mates. Clearly a fairer way to evaluate them would be by measuring only the time they took to run their own segment. I avoid this problem by using an expanding frontier approach, in which I compare the time it takes different locations to adopt agriculture *once at least one of their neighbors has already adopted themselves*. I find that the more seasonal a given location, the faster the local population adopted farming.

The statistical relationship between climate seasonality and agricultural adoption is significant and robust but could be unrelated to the incentives to store food. For example, a short growth season might favor the evolution of plants that are exceptionally easy to cultivate (Diamond, 1997). To help separate these two channels, I look at a subsample of sites that had the same seasonality and domesticable species but differed in the opportunities they offered to a nomadic band. Some sites were close to large changes in elevation, meaning that nomads could migrate seasonally to areas with uncorrelated resource shocks. Other sites were surrounded by areas of similar altitude to their own, making such migrations pointless. Consistent with my theory, I find that adding a 1000m mountain within 50km of a given site (i.e. out of reach of a settled band, but easily accessible to nomads) delays adoption by 500 years. Crucially, I show that this is effect is only present for variation in altitude that occurs over distances that are useful to nomadic populations. If the altitude variation occurs within 5 or 10 km (i.e., variation that could be accessed even by a sedentary population) the effect actually goes in the other direction.

Finally, I show that what evidence we have for actual consumption seasonality across the transition also supports an important role for storage in the Neolithic Revolution. Specifically I use data from Harris Lines, which are formed in bone when a growing child is subject to a period of growth arrest, for example due to a period of famine, or disease. While the sample is heavily biased towards North American sites, the clearly shows that Hunter Gatherers tended to have many more growth arrest episodes during childhood and adolescence, and further that such episodes tended to occur at regular annual intervals. While sedentary farmers were in fact shorter, the almost complete absence of Harris Lines suggests that they were at least able to smooth their consumption, exactly as my model predicts.

2.2 Data

My analysis requires information on where and when agriculture was invented independently, the dates in which it reached particular areas, and information on the climate prevalent at the time.

2.2.1 The invention and spread of agriculture

Data on the invention of agriculture comes from two main sources: direct archaeological evidence of domesticated plants or farming implements, which are typically dated by ^{14}C; and DNA sequencing of large populations of modern crops, which are then compared to modern wild plants to determine the locations with the closest match, and the time elapsed since the last common ancestor (and hence the approximate time and place of domestication). (Purugganan and Fuller, 2009) synthesize evidence from these two distinct lines of research, and distinguish between generally accepted primary (i.e. independent domestications centers) and potentially important secondary domestication centers.

The previous dataset has information on the time and place of domestication but does not track the gradual spread of the Neolithic to neighboring areas. Putterman and Trainor (2006) provides data on the earliest date for which there is evidence of agriculture for 160 countries. This dataset compiles for each country the year for which agriculture first appears in the archaeological record. Note that while the Purugganan and Fuller dataset is compiled mainly from genetic evidence (the number of generations which separate modern crops from their wild cousins), the Putterman dataset is based entirely on archaeological reports. As such, the dates are not always in perfect agreement. To harmonize the two datasets, I assign to individual cells whichever adoption date is earliest: that of the country it belongs to, or that of any domestication area it may be a part of.

While the Putterman dataset enables me to track the spread of agriculture on a global scale, the use of countries as a unit of analysis limits my ability to examine diffusion at the regional level. To obtain finer-grained data, I employ the data collected by Pinhasi et al. (2005), giving the dates for the first evidence of agriculture in 765 different archaeological sites in Western Eurasia. These sites chronicle the spread of the middle eastern set of crops (mainly barley and various types of wheat), which were domesticated in the so-called fertile crescent and diffused into Europe at an average speed of approximately one kilometer per year. The location of each archaeological site was checked against the literature, and the exact coordinates adjusted where necessary.

2.2.2 Climate data

My main source for climate data is the TraCE Dataset He (2011), which uses the CCSM5 model to simulate global climatic conditions for the entire planet for the last 22,000 years. The model employs the orbital parameters of Earth, the extent of the glaciers in each hemisphere, the concentrations of various greenhouse gases, as well as changes to sea level. The model outputs average temperature and precipitation totals for each trimester, for 3.75x3.75 degree cells, at a yearly frequency. I aggregate the time dimension of the dataset into 44 periods of 500 years each. This data allows me to analyze the invention and spread of agriculture using climate conditions contemporaneous to the Neolithic rather than to proxy using modern datasets.

The TraCE data has the advantage of providing insight into past climates, but for regional-scale analysis, its spatial resolution is marginal. To complement the Pinhasi dataset on European adoption dates, I instead use present climate data from the BIOCLIM project (Hijmans et al., 2005), which is representative of average conditions between 1950 and 2000, and is available at 10km resolution. From this dataset, I employ Mean Temperature, Mean Precipitation, Average Temperature of Coldest Quarter, Average Temperature of Hottest Quarter, Average Precipitation of Driest Quarter, and Average Precipitation of Wettest Quarter. The use of present data is potentially problematic, especially when comparing outcomes which are distant in space or time. In this case, the analysis is limited geographically to Western Eurasia, and chronologically to the period after the end of the Ice Age. Together, these constraints allow us to tentatively assume that ordinal relationships are largely preserved (i.e. if Denmark is colder than Lebanon in the present, it is very likely that it was also colder in 8,000 BC).

2.2.3 Other data sources

The altitude data comes from the Shuttle Radar Topography Mission (SRTM), as described in Farr et al. (2007). For part of the analysis, I limit the dataset to the subset of archaeological sites which had access to barley, emmer wheat or einkorn wheat. I use the maps from Harlan (1998), from page 94 and onwards.

2.2.4 Variable construction

The model predicts that agriculture will be adopted when nomadic hunter-gatherers have to suffer through periods of seasonal scarcity. This will tend to happen when a given region experiences high seasonality in temperatures, precipitation, or both. Under these conditions, plant growth will be vigorous during part of the year, but virtually absent in another.

The response of plants to temperature is not linear. In particular, no photosynthesis can occur once groundwater freezes, meaning that below 0°C, further decreases in temperature have little effect. At first sight, a location where winter is 40°C colder than summer might appear to be highly seasonal. But if this oscillation occurs between -10°C and -50°C, in practice there will never be any food, and resource seasonality will effectively be zero.

To avoid counting such a location as seasonal, I concentrate on the temperature range above 0 °C , that is:

$$TempSeas = \max(Temp.Warmest, 0) - \max(Temp.Coldest, 0)$$

That is, I first censor the average temperatures of each quarter at zero degrees Celsius, and then take the difference between the two. The principle behind this measure is the same used by several commonly used measure of agricultural suitability, which also censor temperature variation below a specified limit. For example Growth Degree Days are calculated by first taking the maximum between the temperature of each day and a baseline value, and then summing all of the results. The baseline varies depending on the species being analyzed, but is always above 0° Celsius. The measure I employ will therefore be approximately proportional to the difference in Growing Degree Days experienced in different seasons.

For precipitation, I use the amount of precipitation during the wettest month, minus the level during the driest, divided by mean precipitation.

$$PrecipSeas = \frac{Precip.Wettest - Precip.Driest}{MeanPrecip.}$$

It would prove useful in the analysis to have a single measure reflecting both types of seasonality. Combining these two variables is problematic: water and temperature affect the food availability in complex ways. In the absence of an obvious candidate which can be calculated directly with the data at hand, I define the following Seasonality Index:

$$\text{SeasIndex} = \max(Quantile(TempSeas), Quantile(PrecipSeas))$$

That is, for each cell and period, I transform the two seasonality measures into quantiles (1000 categories). The seasonality index is equal to whichever of the two measures has the highest score. For example, if a location has a Seasonality Index of 900, it must either have more temperature seasonality than 90% of the cell-period observations, or more precipitation than 90% of the cell-period observations. I choose the minimum rather than the average because plant growth is limited mainly by the least abundant factor. For example, Sub-Saharan Africa is

never cold, but the presence of a long dry season is sufficient to make food supply highly seasonal.

I proxy for the average food supply by using climatic averages. Mean Temperature is the average temperature in degrees Celsius across the four seasons. Similarly, Mean Precipitation is the the average amount of rainfall in the four seasons, measured in mm per day.

2.3 Results

The goal of this section is to show that climatic seasonality was the main driver of the multiple invention of agriculture. First, I check whether the areas in the world where agriculture was invented where unusually seasonal, and I find that in all seven, a warm and moist season alternated with either very harsh winters, or a very dry season. Second, I show that farming spread faster in highly seasonal locations. Third, I estimate the combined effect of invention and spread on the timing of adoption, and find that one extra standard deviation of temperature seasonality is associated with adopting agriculture 1,500 years earlier. I replicate the most important steps of this analysis on a higher resolution regional dataset for Western Eurasia, which confirms the earlier findings.

The preceding establishes a strong and robust link between climate seasonality and the adoption of agriculture, but it does not identify the channel. For example, Diamond proposed that the invention of agriculture was caused by the availability of plants that were easy to domesticate, such as large seeded grasses. Did a short growth season favor the evolution of such plants? To avoid this threat to identification, I concentrate on a subsample consisting entirely of highly seasonal locations, but with heterogeneity in the ability of nomads to leverage their mobility.

Further verification for the model's findings come from the paleopathological record of the Neolithic. Analysis of skeletal remains shows that consumption per capita decreased after the invention of farming, but the absence of growth-arrest line confirms that consumption seasonality decreased as well.

2.3.1 Global-scale analysis

The climate data consists of $48 \times 96 \times 22,000$ observations (Latitude$\times$ Longitude$\times$ Years). My first step is to contract the dataset along the time dimension by averaging the climatic variables by 500 year periods. The resulting dataset has $48 \times 96 \times 44$ observations, each representing the conditions present in a specific latitude and longitude, during a specific period. I drop all observations that are covered by water, and Antarctica, leaving 1036 cells.

To this dataset, I merge my data on agricultural invention, by generating a dummy that takes the value of 1 if agriculture was invented in a particular place and time and 0 otherwise. This variable is based on the Purugannan and Fuller data. I also generate another dummy – based on the Putterman and Trainor data on agricultural adoption – which takes the value of 1 if agriculture had already been adopted in a particular time and place (regardless of whether it was invented locally or adopted by neighbors).

I will begin by presenting some summary statistics for the Neolithic Revolution. I collapse the data to a cross-section, by averaging all values of each variable for a given location, through time. YearAdop is the date of the earliest evidence for agriculture in a given country, expressed in years before present. The very first farmers appeared 11,500 years ago, while some locations are still populated by hunter gatherers today (e.g. Greenland). The average location on Earth started farming 4,500 years ago, had an average temperature of $2.5\,^{\circ}\mathrm{C}$, received 1.8mm/day of rainfall (approximately 650mm/year), had a temperature seasonality of $9\,^{\circ}\mathrm{C}$, a precipitation seasonality of 1.3, and a seasonality index of 625 (out of 1000).

	mean	sd	min	max
Year Adop.	-4500.00	2500.43	-11500.00	0.00
Temp. Seas	8.85	7.26	0.00	28.98
Precip. Seas	1.35	0.67	0.16	3.58
Temp. Mean	2.49	17.44	-33.98	27.64
Precip. Mean	1.80	1.63	0.02	10.40
Seas. Index	625.13	225.53	84.37	993.60
Observations	1036			

Table 2.1: Summary statistics for the adoption cross-section dataset.

How well does my story fit the basic features of the data? Figure 2.2 shows how many cells were seasonal during each period of the last 22,000 years. A location is considered seasonal if it has a Seasonality Index above 925. Seasonal locations were rare during the Ice Age, but became increasingly common in the lead up to the adoption of agriculture, more than tripling in frequency. This trend was driven by the simultaneous peaks in the three orbital parameters influencing seasonality (as discussed in Section 1.2). Figure 2.3 shows how six out of seven of the independent inventions occurred precisely in these areas, or in very close proximity. The outlier is Mexico, where drylands with highly seasonal rainfall coexist in close proximity with tropical rain forests on the other side of the mountains. The spatial resolution of the climate dataset is marginal for these conditions, as it necessarily average rainfall figures that vary tremendously on the ground. Today, Oaxaca state (where Central American agriculture originated) has an extremely seasonal precipitation pattern, with virtually all rainfall occurring during half the

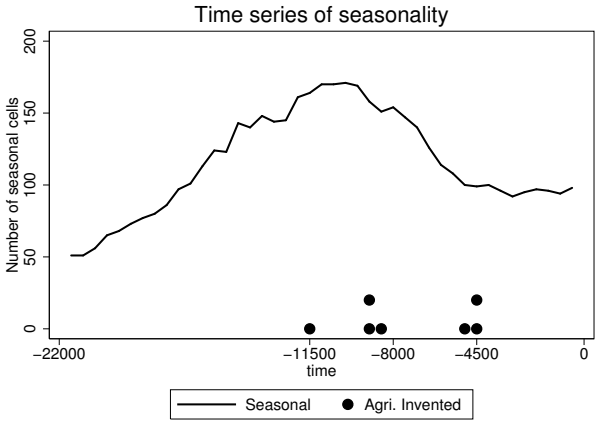

Figure 2.2: The number of cells with seasonal climates (Seasonality Index > 925), through time. The black dots mark the timing of the independent adoptions. At the start of the Neolithic period, there were more than three times as many seasonal locations as during the Ice Age. This was primarily driven by the changes in orbital parameters described in Figure 1.2.

year.

Independent invention

I will first check whether higher seasonality made invention more likely. I examine this prediction in the global context, by using the data on independent domestications from (Purugganan and Fuller, 2009) and the panel of climate data from He (2011). Each observation is one 3.75x3.75 degree cell, during a specific 500-year period, and I drop each location after it adopts agriculture. The basic specification is:

$$I_{it} = \alpha + \beta_1 \mathbf{T}_{it} + \beta_2 \mathbf{P}_{it} + \gamma C_{it} + \epsilon_{it} \tag{2.1}$$

$$\tag{2.2}$$

Where I_{it} is a dummy for whether agriculture was invented in cell i at time t, α is a constant, T_{it} is temperature seasonality, P_{it} is precipitation seasonality, and C_{it} is a vector of controls. The adoption dummy I_{it} is 0 for all locations and periods, except for seven 1s representing the times and places where agriculture was invented. As each location invents agriculture or adopts it from neighbors, I drop it from the panel.

I use logistic regression to estimate the model and present the results in Table 2.2. In column (1), the only explanatory variables are the two individual season-

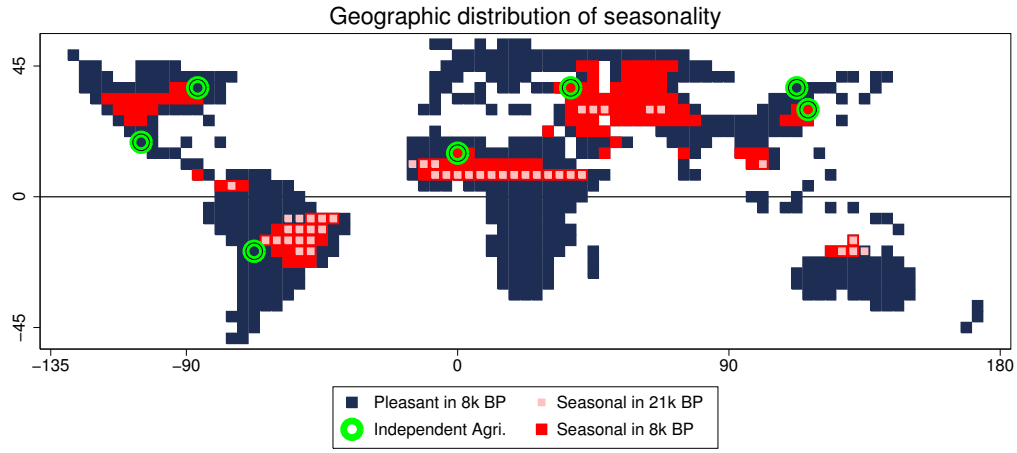

Figure 2.3: The map shows the global distribution of seasonal locations. Pink cells were already seasonal in 21k BP. Cells that were seasonal in 8,000 BP, are in red. Dark blue cells are hospitable in 8,000 BP (average temperature > 0 and annual precipitation $>$ 100mm). Locations that were not hospitable in 8,000 BP are omitted. Most of the areas where agriculture was invented had recently become extremely seasonal.

ality measures. The coefficient on temperature seasonality is positive and statistically significant, while precipitation seasonality is not significant. In column (2), I add controls for mean temperature, mean precipitation, and absolute latitude. The coefficient on both types of seasonality increases, and the coefficient temperature seasonality remains significant. The same pattern holds in column (3), where I include a New World dummy, and quadratic terms for absolute latitude and the two climatic averages. In column (4), I add controls for the modern level of temperature and precipitation seasonality. This confirms that the effect comes from climate conditions present at the time and not through correlation with present conditions. Finally, column (5) shows that the Seasonality Index is also a good predictor of independent invention. Very similar result are obtained using the Rare Events Logit estimation described by King and Zeng (2001), by clustering standard errors at the location level, or if different measures of seasonality are used. These results are in line with the predictions of the model: the places that invented agriculture were all extremely seasonal.

	Dependent variable: invention dummy				
	(1)	(2)	(3)	(4)	(5)
	Basic	Controls	Controls2	ModernSeas	SI
Neol7					
Temp. Seas.	0.197***	0.188***	0.232**		
	(0.051)	(0.063)	(0.106)		
Precip. Seas.	0.676	0.683	0.015		
	(0.633)	(0.679)	(1.339)		
Seas. Index				8.525**	6.571*
				(4.021)	(3.879)
Temp. Mean	0.046	0.050	0.028	0.053	0.091
	(0.050)	(0.125)	(0.129)	(0.038)	(0.149)
Precip. Mean	0.846***	1.639***	1.591**	0.812***	1.036
	(0.216)	(0.625)	(0.713)	(0.301)	(0.713)
Abs Lat	0.051	0.128	0.128	0.083	0.206***
	(0.034)	(0.088)	(0.101)	(0.050)	(0.065)
Temp. Seas. Today			-0.055		
			(0.207)		
Precip. Seas. Today			0.819		
			(1.265)		
Seas. Index Today					-0.280
					(2.021)
Extra Controls	No	Yes	Yes	No	Yes
p	0.00	0.00	0.00	0.00	0.00
N	38533	38533	38533	38533	38533

Standard errors in parentheses

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

Table 2.2: The effect of climate on adoption. Dependent variable is a dummy which is 1 if agriculture was invented in a particular cell and period, and 0 otherwise. Each location is dropped from sample after they adopt agriculture. Logistic regression on climate variables and controls.

Spread of farming

I now turn my attention to the process of agricultural diffusion, which saw farming grow from a handful of isolated outposts to becoming the dominant lifestyle on Earth. For this part of the analysis, I construct a dataset consisting only of locations that are likely to receive agriculture soon. Specifically, from the full panel, I keep only observations that have hospitable climates [1], haven't already adopted agriculture, and have neighbors that are already farming. This sample represents the population which is "at risk" of adopting agriculture from neighbors.

The basic specification is:

[1] A location is considered hospitable if it has average temperatures above 0 ° C, and more than 100mm of rain a year.

$$A_{it} = \alpha + \beta_1 \mathbf{T}_{it} + \beta_2 \mathbf{P}_{it} + \gamma C_{it} + \epsilon_{it} \tag{2.3}$$

Each observation represents a specific cell i, during a specific period t. I keep only observations which are on the agricultural frontier: cells that still haven't adopted agriculture themselves, even though at least one of their neighbors already has. The dummy variable A_{it} is coded as 1 if agriculture was first adopted in location i at time t and 0 in all other periods. This model is estimated using the logistic estimator (first tree columns of Table 2.3, and then with the linear probability model (last three columns). In both cases, I find that seasonality is associated with a higher probability of adopting agriculture from neighbors. Clustering residuals at the level of 123 geographic neighborhoods preserve the significance of temperature seasonality and the seasonality index, but precipitation seasonality becomes less significant.

	Dependent variable: adoption dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear	Linear Geog.Cluster	LinearSI	Logit	Logit+ Geog.Cluster	LogitSI
main						
Temp. Seas.	0.005**	0.005*		0.027**	0.027*	
	(0.002)	(0.003)		(0.011)	(0.015)	
Precip. Seas.	0.035*	0.035		0.174*	0.174	
	(0.019)	(0.029)		(0.092)	(0.144)	
Seas. Index			0.168*			0.861*
			(0.096)			(0.506)
Temp. Mean	-0.007***	-0.007*	-0.007***	-0.032***	-0.032*	-0.034***
	(0.002)	(0.004)	(0.003)	(0.010)	(0.017)	(0.012)
Precip. Mean	0.023***	0.023	0.017	0.113***	0.113	0.086
	(0.008)	(0.015)	(0.012)	(0.038)	(0.071)	(0.058)
Observations	1735	1735	1735	1735	1735	1735

Standard errors in parentheses

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

Table 2.3: Effect of climate seasonality on spread of agriculture. The sample is composed only of location-period combinations on the Neolithic frontier (at least one of their neighbors is already farming, but they are not). The dependent value is a dummy for whether agriculture was adopted. Regression of adoption dummy on climatic variables. Model 1 is Logit with robust s.e., models 2 and 3 Logit with geographic clustering. Model 4, linear probability with robust s.e., models 5 and 6 linear probability with geographic clustering.

I also estimate a continuous time duration model with Weibul distribution and plot the resulting survival curves for various climate types (Figure 2.4). The more seasonal a location was, the sooner the locals would adopt agriculture from farming neighbors. For example, 2,000 years after being exposed to agriculture, a

location with zero temperature seasonality still has a 40% change of being occupied by hunter-gatherers. An otherwise equivalent location with a temperature seasonality of 25 C has only a 20% chance. Very similar results are obtained for precipitation seasonality. In the Appendix, I show that these results also hold when using a parametric survival model.

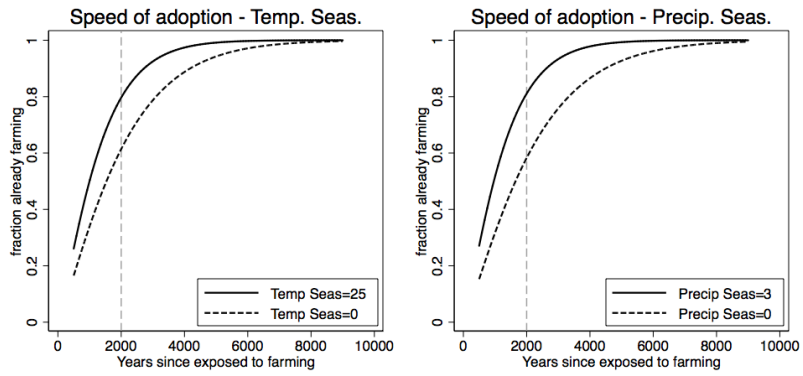

Figure 2.4: Fraction of locations expected to already farm, after a given number of years of being exposed to farming neighbors. Solid lines: high seasonality locations. Dashed lines: unseasonal locations. Left panel: temperature seasonality. Right panel: precipitation seasonality.

Impact of seasonality on date of adoption

The next step of my analysis is to estimate the cumulative effect of climate seasonality on the timing of the Neolithic. Figure 2.5 shows binned scatterplots of date of adoption against measures of seasonality. The early adopters were unremarkable in their average climates but were clearly highly seasonal.

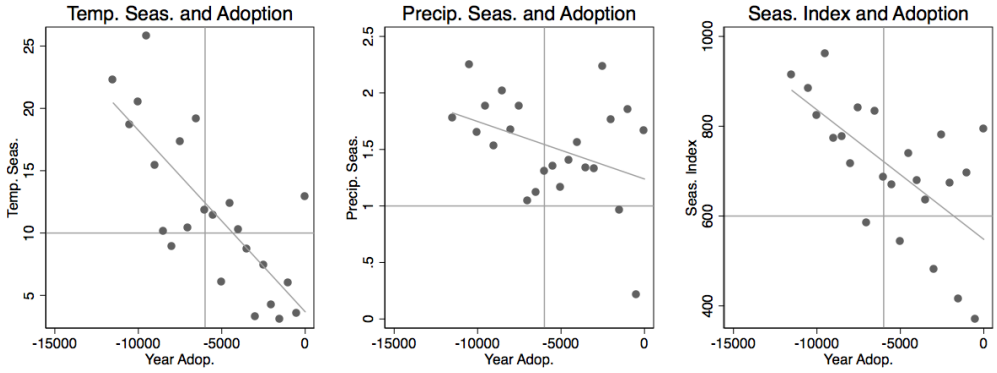

Figure 2.5: Binned scatterplots of different forms of climate seasonality vs the date of adoption. Locations exposed to more seasonal climates adopted agriculture ahead of more stable climates.

For this part of the analysis, I collapse the data into a cross-section, where the dependent variable is the date of adoption, and each explanatory variable is given the value it had when agriculture was adopted in that location. The basic specification is:

$$Y_i = \alpha + \beta_1 \mathbf{T}_i + \beta_2 \mathbf{S}_i + \gamma [C]_i + \epsilon_i \tag{2.4}$$

Where Y_i is the date in which cell i adopted agriculture, in years Before Present (i.e. ten thousand years ago is represented as -10,000).

The results of this analysis are presented in Table 2.4. Both Temperature and Precipitation Seasonality are associated with earlier adoption of agriculture, across a wide range of specifications. The effect is large and statistically significant for both factors, as well as for the combined Seasonality Index. Column (1) reports the direct effect of temperature and precipitation seasonality on adoption, without controls. The point estimate suggests that one extra standard deviation of Temperature Difference will result in agriculture appearing approximately 1,000 years earlier than would otherwise have been the case. One extra standard deviation of rainfall seasonality will instead result in adopting agriculture 300 year earlier. Column (2) inserts basic geographic controls (climatic means and absolute latitude). These help discriminate the seasonality story from the most obvious correlates. When these controls are included, the point estimates of the effect of both types of seasonality increase, to 1,500 and 400 years respectively. Column (3) adds controls for the squares of climatic means and latitude, as well a dummy for the New World, and clusters the standard errors. The results are very similar to those from column (1). Column (4) removes all the controls except for mean

temperature and mean precipitation and instead uses fixed effects for 123 geographic regions taken from an evenly spaced grid. This approach removes most of the variation in the sample, and results in weaker (but still significant) point estimates. Column (5) and Column (6) substitute temperature and precipitation seasonality with the Seasonality Index and replicate the first two columns. One extra standard deviation of the index is associated with adopting agriculture between 1,000 and 1,250 years earlier.

	Dependent variable: year of adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
	Basic	Controls	Controls2	GeoFE	SI	SI+Controls
Temp. Seas	-131.1***	-222.5***	-143.8***	-51.6***		
	(10.1)	(13.4)	(38.4)	(17.5)		
Precip. Seas	-152.2	-529.4***	-936.5***	-435.3***		
	(110.4)	(131.1)	(249.2)	(112.3)		
Seas. Index					-3.3***	-5.1***
					(0.3)	(0.4)
Temp. Mean		107.3***	71.5**	9.5		42.7***
		(15.9)	(29.6)	(15.8)		(15.2)
Precip. Mean		-464.3***	90.0	-51.1		-257.2***
		(71.2)	(235.8)	(113.6)		(72.4)
Abs Lat		46.3***	207.6***	3.4		4.7
		(13.6)	(64.9)	(15.3)		(12.6)
Extra Controls	No	No	Yes	Yes	No	No
Geographic FE	No	No	No	Yes	No	No
r2	0.15	0.24	0.40	0.87	0.09	0.12
p	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses

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

Table 2.4: Effect of seasonality on the date of adoption (both invention and adoption from neighbors). Linear regression of date of adoption on time-averaged climatic variables for each cell. Column 3: clustering for 123 geographic neighborhoods. All other columns: robust standard errors.

It is worth noting that, while the measures of seasonality preserve their significance throughout the various specifications, the same cannot be said for the measures of climatic averages. This confirms the predictive weakness of linking agriculture to the end of the Ice Age. The results are similarly strong using a spatial lag model and Conley's geographically adjusted standard errors. The results from these robustness checks are presented in the Appendix.

2.3.2 Results from the Western Eurasia dataset

The preceding analysis has established that seasonality can account for a significant fraction of the variation in the date of agricultural adoption observed in the

world sample, and the effect can be observed both in the selection of places that originally invented farming, as well as in the speed with which these new techniques spread throughout the globe.

However, the data employed present certain limitations in geographic resolution that cannot be overcome easily. The methodology used to construct the climate dataset does not take into account small-to-medium scale topography, which has a large effect on the realized climate outcomes. Also the dependent variable (agricultural adoption) was coded with a single value for each state, which creates issues when dealing with large countries. In any case, different regions around the world have been excavated to different degrees, leaving the possibility that agriculture was adopted in e.g. the Amazon or Sub-Saharian Africa at a much earlier date than is currently known.

To verify the findings of the global-scale analysis in a setting free from these particular shortcomings, I now look at the spread of agriculture from the Middle East into Europe. These regions have been at the center of concentrated study for well over a century, and are undoubtedly the most researched case of agricultural invention and expansion.

Specifically, Pinhasi et al. (2005) have collected a dataset of 765 archeological sites for which the date of earliest agriculture has been established through ^{14}C dating. The resolution of the TraCE climate dataset is far too low to be useful on this scale, so I substitute the BIOCLIM data of Hijmans et al. (2005), which is representative of average climatic conditions from 1950-2000, but has the advantage of being available at 10km resolution.

As Figure 2.6 shows, the earliest agriculture in this sample occurred in a wide arc joining the Eastern Mediterranean to the Persian Gulf. In fact, this area is currently believed to have been the earliest case of plant domestication anywhere in the world. From the flanks of the Zagros and Tauros mountains, farmers and their crops spread out onto the plains of Mesopotamia, and westwards across the Bosphorus, into the Balkans, and in two parallel thrusts into the northern European plains and the central and western Mediterranean.

Since agriculture was invented only once within this region, systematic statistical techniques are clearly inappropriate. However, the so-called Fertile Crescent is in fact not *particularly* fertile. Many locations on the Northern shore of the Mediterranean enjoy similar conditions of high average temperatures and adequate rainfall. What seems to set the area apart is the fact that it is simultaneously a pleasant environment and a seasonal one. Thus, the Western Eurasian story of invention conforms to the general pattern observed globally in which the most seasonal locations adopted agriculture sooner.

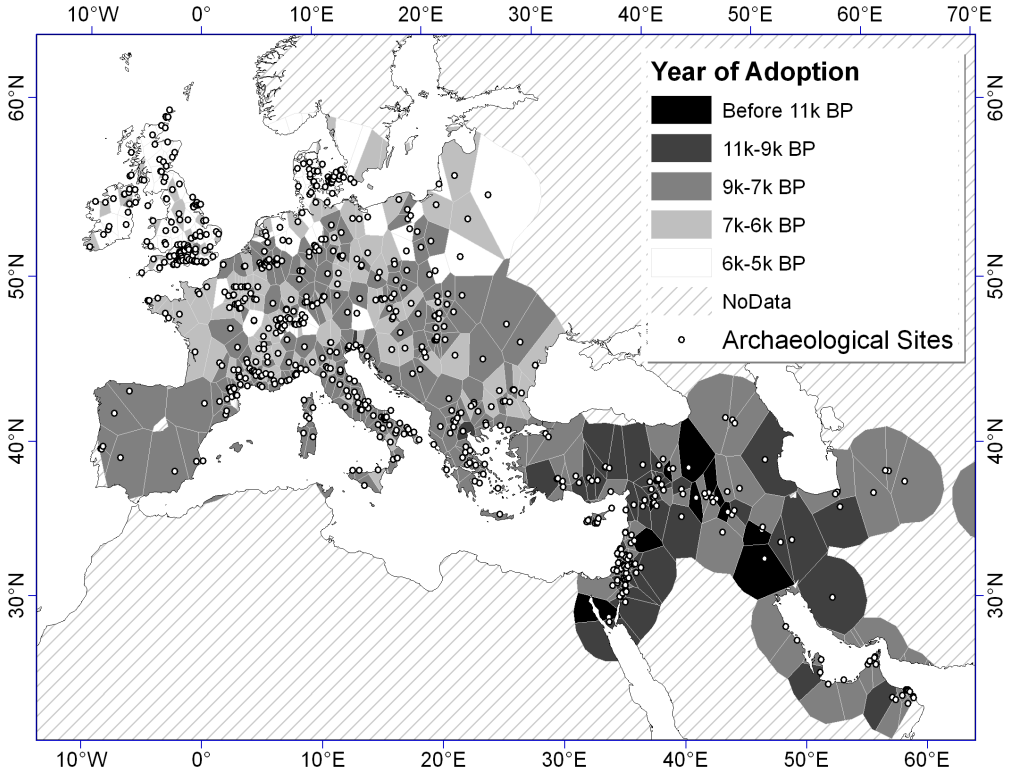

Figure 2.6: The Pinhasi et al. (2005) dataset provides ^{14}C dates for the onset of agriculture in 765 locations, chronicling the spread of agriculture from the Middle East into Europe.

	mean	sd	min	max
Year Adop.	-7218	1424	-12811	-5140
Temp. Seas.	15.2	3.2	6.9	25.1
Precip. Seas.	.23	.18	.038	.72
Temp. Mean	12.0	4.7	4.4	30.2
Precip. Mean	1.84	.73	.04	4.77
Observations	765			

Table 2.5: Summary statistics for the Western Eurasian dataset.

This relationship is also apparent from the analysis of the raw data on the diffusion of farming techniques through the archaeological sites in the sample and their date of adoption. As the scatterplots in Figure 2.7 show, the locations which adopted early had high seasonality of temperature and precipitation, while locations with stable climates adopted agriculture much later.

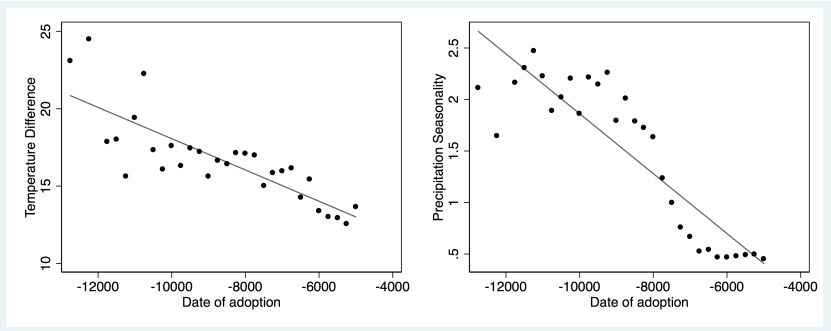

Figure 2.7: Binned scatterplot of climate seasonality and adoption dates. More seasonal locations adopted earlier, while less seasonal climates adopted later.

The basic specification is the same that for the basic linear model of Subsection 2.3.1:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 P_i + \gamma C_i + \epsilon_i \tag{2.5}$$

Where Y_i is the year in which archaeological site i adopted agriculture, T_i is temperature seasonality, P_i is precipitation seasonality, and C_i is a vector of controls. The results are presented in Table 2.6, which once again shows how high seasonality is a strong predictor of early adoption, even when controlling for distance to the locations where agriculture originated, altitude, distance to the coast, and the usual controls from the previous regressions.

Column (1) shows the direct effect of temperature and rainfall seasonality on the date of adoption. One extra standard deviation of temperature seasonality results in agriculture being adopted about 400 years earlier, while an equivalent change in rainfall seasonality is associated with adopting agriculture approximately 900 years later. These two variables alone account for over 60% of the variance in date of adoption observed in the sample. In Column (2), I add controls for climatic averages which slightly increases my point estimate for temperature seasonality, while reducing the one for precipitation seasonality. Column (3) adds controls for latitude, altitude, and distance from the Fertile Crescent (where agriculture started, for this dataset). In Column (4), I add a control for distance from the coast, and Column (5) concludes by adding quadratic terms for the climatic means. As more controls are added, the magnitude of the estimated coefficients falls, but all retain statistical and economic significance, as well as the correct sign.

	(1) Basic	(2) +Means	(3) +Geo	(4) +Geo2	(5) +Mean2
Temp. Seas.	-136.8***	-148.6***	-72.80***	-75.06***	-46.22**
	(12.25)	(13.11)	(20.15)	(22.59)	(23.48)
Precip. Seas.	-5102.7***	-3711.5***	-2042.6***	-2060.2***	-2028.4***
	(226.4)	(350.7)	(346.8)	(355.3)	(387.3)
Temp. Mean		-74.19***	19.76	21.00	-195.8***
		(14.73)	(20.02)	(22.32)	(43.26)
Precip. Mean		-90.87	-124.1**	-123.4**	239.3
		(68.41)	(61.74)	(62.47)	(245.5)
Dist Coast				5.703	-32.77
				(26.62)	(28.02)
Temp Mean 2					7.068***
					(1.375)
Precip Mean 2					-71.14
					(48.93)
GeoControls	No	No	Yes	Yes	Yes
Observations	765	765	765	765	765
R^2	0.610	0.627	0.692	0.692	0.706

Standard errors in parentheses

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

Table 2.6: Climate seasonality and adoption in the Western Eurasia dataset, linear model, robust standard errors.

2.3.3 Geographic heterogeneity

The analysis conducted so far has established that seasonality is strongly associated with the adoption of agriculture. These findings agree with the results from the model previously developed, and suggest that the farming was invented in locations where the incentive to store food was high.

However, the association between seasonality and agriculture could also be due to the availability of easily domesticable plants, in the spirit of Diamond (1997). Plants have adapted to highly seasonal environments react by conducting their own forms of storage, either by storing energy in their roots, or by producing large amounts of seeds during the short growth season. Both of these adaptations create plants that are easier to cultivate, and that are in some sense pre-adapted to domestication. It is therefore possible that agriculture was first developed in highly seasonal locations not because of the incentives to store available food, but because these conditions were the only ones in which suitable plants thrived. Once these plants had been domesticated, it is only natural that the spread should have been faster in locations with similar climates, thus providing a potentially plausible explanation for the observed pattern of invention, and spread.

While these factors could have further assisted the development of agriculture, I can show that the nomadism-storage tradeoff retains independent explanatory power. To this end, I focus on those areas of the Middle East where cereals are known to have grown wild, i.e. areas that had very similar endowments of domesticable species. All of these locations are extremely seasonal, so both temperature and precipitation seasonality lose their explanatory power. The model shows that settled agriculture should be adopted earlier where mobility is less useful — i.e. where all locations in practical migratory range lack food at the same time.

To test this prediction empirically, I first limit the analysis to the subset of locations from the Pinhasi et al. (2005) dataset that are within a specified radius of known concentration of wild cereals. I then construct a series of proxies, each measuring the range in altitudes present within a specified distance from the location under observation. Areas with different altitudes will experience different temperature and precipitation regimes, are likely to have slopes with different exposures to the sun, and will generally possess a wide variety of microclimates. In short, it is highly unlikely that areas at widely differing altitudes will suffer the type of perfectly correlated seasonal food shocks that makes nomadism pointless.

The behavior of the band will differ based on the scale on which these variations occur. If great altitude variability can be found within a small distance – say, 5km – then the band will be able to access this variation from a single location, and we expect settlement to actually occur faster than if no variation had been present. Altitude heterogeneity at larger radii ($\approx$ 50km) will instead lie beyond the grasp of the settler but will be easily accessible to the nomad. Locations with such a topography will create an incentive to remain nomadic. Eventually, at very large distances, the uncorrelated food sources will be beyond the migratory ability of even the most mobile nomads, and therefore irrelevant. Table 2.7 presents the summary statistics for the sites in the Pinhasi dataset that are within 100km of known concentrations of wild cereals. Note that all of these places are quite seasonal.

In Figure 2.8, I show the locations in the Pinhasi dataset that are close to known concentrations of wild cereals. I will use four sites in particular to illustrate how topography affects the incentives to remain nomadic or transition to settled storage. These are all within a 250km-radius circle at the border of Iraq, Syria and Turkey, and all had access to the same domesticable species. However, they differ greatly in local topography, as shown in Figure 2.9. Location (1) is Jerf el Ahmar, which lies on the banks of the Euphrates river, in the middle of a flat plain. Location (2) is Qermez Dere, on the southern flanks of a steep mountain, surrounded by an extensive and homogeneous plain. Location (3) is Girikiacian, which lies on a flat stretch of land close to some mountains. Finally, location (4) is Gawra, which is right next to some reasonably tall mountains, but has some truly impressive peaks around 40kms away. For each archaeological site, I plotted

	(1)			
	mean	sd	min	max
Years Ago	-9520	1336	-12811	-7276
r(5)	366.7	297.8	16	1330
r(50)	1485.3	666.4	99	3108
Temp. Seas.	18.1	4.12	11.4	24.7
Precip. Seas.	.54	.10	.21	.67
Temp. Mean	17.9	3.3	8.1	24.1
Precip. Mean	1.03	.60	.10	3.26
Latitude	34.2	3.01	29.5	41.4
Longitude	37.9	4.25	26.11	49.63
Altitude	487.2	523.5	-405	2376
Dist Coast	1.80	1.58	0	5.86
Observations	101			

Table 2.7: Summary statistics for the subsample of the Western Eurasian dataset which had access to wild cereals.

a line originating at the site's location, in the direction of the greatest changes in altitude.

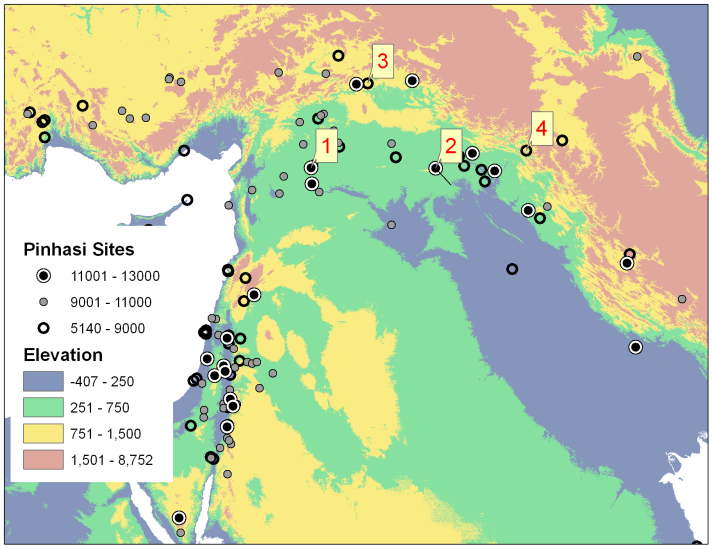

Figure 2.8: The map shows the Neolithic sites in the Middle East from the Pinhasi dataset that are within 100km of known concentrations of wild cereals. The sample is divided in locations that adopted before 11,000 years ago, between 11,000 and 9,000 years ago, and after 9,000 years ago. The four example sites discussed in Figures 2.9 and 2.10 are highlighted.

In Figure 2.10, I show elevation profiles taken along these lines, allowing us to better appreciate the differences in local topography. Locations (1) and (3) both

have only moderate changes in altitude within 5km of the site, but the land around (1) is flat in all directions for at least another 100km, while (3) has significant peaks within the assumed nomadic radius of 50km. In contrast, Locations (2) and (4) both have large changes in elevation within their immediate neighborhood, but (2) is surrounded by a flat plain, while (4) has even larger mountains within the migratory radius of nomads.

As predicted by the theory, locations (1) and (2) – which had little to loose from abandoning nomadism – were amongst the first locations to adopt farming, while locations (3) and (4) – where the opportunity cost of abandoning nomadism was high – adopted only more than 2,000 years later. The local topography was not crucial: the areas within 5km of the two early adopters look very different from each other. What mattered was that the prospective settlers could find a location from which they could access the same variety of ecosystems which they could exploit as nomads.

This pattern is not specific to these four locations but is found generally within the Middle-Eastern sample. As Figure 2.11 shows, the early adopter of agriculture have a significantly lower $r(50)$, compared to late adopters with similar levels of $r(5)$. In particular, note that the seven locations with the highest $r(50)$ all adopted agriculture very late.

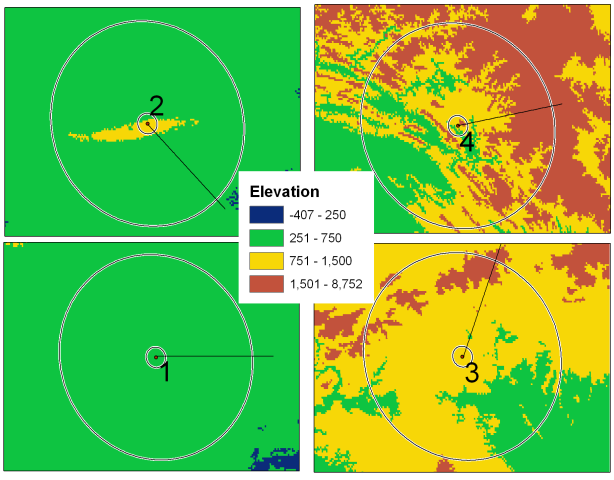

Figure 2.9: The four graphs show the local topography for the four examples sites, shown in Figure 2.8. The small circles have a 5km radius and are indicative of the area that could be accessed by a settled community occupying the site. The large circles are 50km in radius and shows the area that would have been available to a nomadic band.

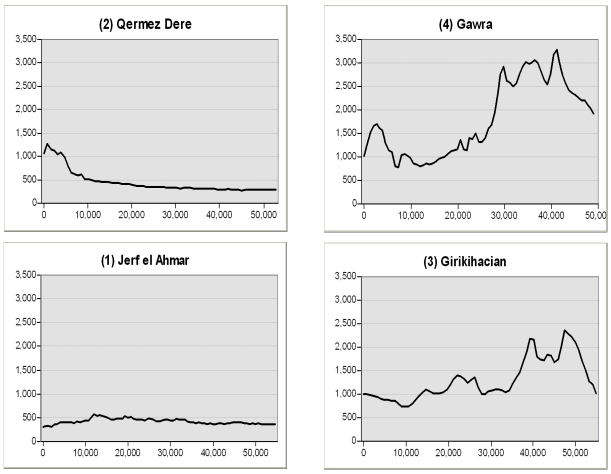

Figure 2.10: The four graphs show altitude profiles for the four lines shown in Figure 2.9. (1) has virtually no altitude variation in the local area. (2) Has a lot of variation close by, but nothing in the wider area. (3) has little variation close by, but a lot in the wider area. (4) has a lot of variation close by, but even more variation within the local area. Locations (1) and (2) adopted early, while locations (3) and (4) adopted later on.

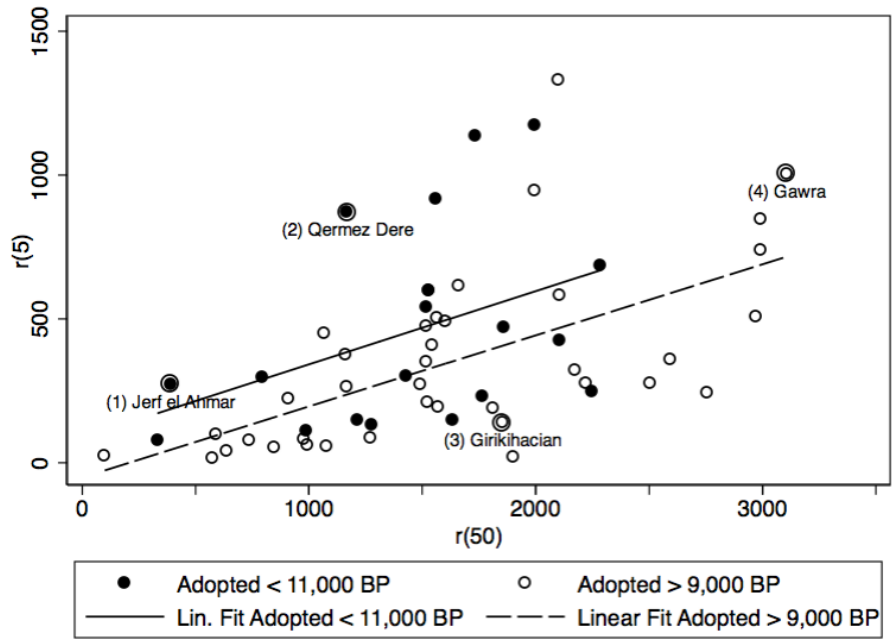

Figure 2.11: The graph shows how, irrespective of the altitude range available to settlers (the $r(5)$), locations with a lot of altitude range available to nomads (the $r(50)$) adopted agriculture later than those with a low $r(50)$. The examples presented in Figure 2.9 are highlighted and labeled, and follow the general pattern.

I now investigate these relationships systematically using linear regression. The basic specification is:

$$Y_i = \alpha + \beta_1 r(5) + \beta_2 r(50) + \gamma C_i + \epsilon_i \tag{2.6}$$

Where Y_i is the year in which agriculture was adopted in archaeological site i, $r(5)$ is the range of elevations present within 5km of the site, $r(50)$ is the range of elevations present within 50km of the site, and C_i is a vector of controls. The model predicts that farming will be adopted first where nomadism does not materially improve the variety of ecosystems the band can access, i.e. where $r(50)$ is low, and $r(5)$ is high. The model is estimated through a straightforward linear specification, and the results are presented in Table 2.8.

Column (1) shows the direct effect of $r(5)$ and $r(50)$ on adoption. The sample is limited to sites which are within 250km of known dense cereals. Altitude variety within settled range (5km) led to earlier adoption of farming. Conversely, altitude variety which could be exploited by nomads (i.e. located 5 to 50km away) resulted in later adoption. The measured effect is large and statistically significant. Adding a 1000m mountain within 50km of a given site delayed adoption by approximately

500 years. In column (2), I restrict the analysis to sites within 100km of known wild cereal distributions. Concentrating on the core areas increases the magnitude and significance of the coefficients. Column (3) keeps the 100km restriction and adds controls for climatic seasonality, average climate, altitude, latitude, distance from the Neolithic epicenter, and distance from the coast. In this highly homogeneous environment, the coefficients on climatic variables are not significant, but those on the altitude ranges are effectively unchanged. In column(4), I add a control for $r(200)$. I find that if variations in altitude happened outside of comfortable nomadic radii they are no longer predictive of date of adoption. Finally, I substitute my measures for sedentary-radius and nomadic-radius altitude variety with two smoothed versions: $r(5:8)$, which is the average of $r(3)$,$r(5)$ and $r(8)$; and $r(50:100)$, the average of $r(50)$, $r(75)$, and $r(100)$. Column (5) shows that, while these measures are less predictive, the magnitudes of the coefficients is not affected, and that of $r(50:r100)$ is statistically significant.

	Dependant variable: date of adoption				
	(1)	(2)	(3)	(4)	(5)
	<200km	<100km	Clim. Means	r(200)	Smooth Meas.
r(5)	-0.772*	-0.990**	-0.986*	-0.970*	
	(0.414)	(0.496)	(0.580)	(0.579)	
r(50)	0.414**	0.517**	0.587**	0.540*	
	(0.179)	(0.221)	(0.267)	(0.306)	
r(3:8)					-0.858
					(0.597)
r(50:100)					0.500*
					(0.254)
r(200)				0.111	
				(0.266)	
Temp. Seas.			-161.6	-158.0	-144.5
			(114.1)	(116.4)	(116.1)
Precip. Seas.			737.9	471.2	-442.4
			(4268.1)	(4417.6)	(4040.5)
Controls	No	No	Yes	Yes	Yes
Observations	129	101	101	101	101
R^2	0.037	0.051	0.110	0.111	0.101

Standard errors in parentheses

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

Table 2.8: Effect of local topography on the timing of agricultural adoption. Linear regression of year of adoption of agriculture on the range of altitude within various radii. More variation in altitude within 50km (greater opportunity cost of abandoning nomadism) delayed the adoption of agriculture.

2.4 Consumption seasonality and human health

The model predicts that the transition from nomadic hunting and gathering to sedentary agriculture should be associated with a lower average food consump-

tion but much greater stability. In this section, I will detail how chronic malnourishment and acute starvation differ in their effects on the human body, and how the evidence from the Neolithic Revolution compares to the the welfare outcomes predicted by the model.

Healthy adults carry fat reserves, the body's primary long-run energy reserves, which generally allow them to survive periods of acute malnourishment. These are complemented by the body's energy conservation strategies, such as reducing body temperature, decreasing fidgeting and unnecessary movement, and generally lowering the basal metabolism (Keys et al., 1950). Unless starvation is prolonged, lost weight can be regained when conditions improve, and the individual need not suffer significant long term consequences. However, fat reserves can only last for so long. Eventually, if the body is unable to reduce its energy requirements to fit the available resources, death by starvation will ensue.

As discussed in the introduction, in most of the locations for which data exist, consumption per capita decreased when farming replaced hunting and gathering. Achieved adult height is one of the most commonly used proxies for health, and as Figure 2.12 shows, this parameter declined drastically as agriculture became the dominant lifestyle (Cohen and Armelagos, 1984). Similar declines in health are evident from a host of other indicators, such as measures of skeletal robustness, tooth wear, joint diseases due to overwork, and evidence of disease and infection. These are the findings that prompted Diamond to title his famous article "the worst mistake in the history of the human race" (Diamond, 1987).

It should be noted that the height decrease was unlikely to be entirely due to the transition from a more meat-based diet of hunter-gatherers to a cereal-based diet during the Neolithic. In many cases, late Paleolithic communities were already highly dependent on the plants that were eventually cultivated and domesticated, and most of the early farmers were still hunting significant amounts of game from their surroundings (Humphrey et al., 2014). Further, in some cases (e.g. the Natufian in the Middle East), height was seen to decrease as soon as the population became sedentary and started to store food, even though cereals were still not a dietary staple.

These observations are in agreement with the welfare implications of the model, which predicted that average consumption should decrease as soon as a population becomes sedentary and starts to store, and should thereafter remain relatively constant, even as farming is adopted.

Measuring consumption seasonality is more difficult: height overwhelmingly reflects the *average* nutritional status an individual experienced through childhood, while *volatility* in food intake is only marginally recorded. Acute starvation episodes in children can in fact pause skeletal growth entirely, but if sufficient nutrition is provided thereafter, the child will experience faster than normal growth. This catch-up growth will generally result in the child rejoining its original growth

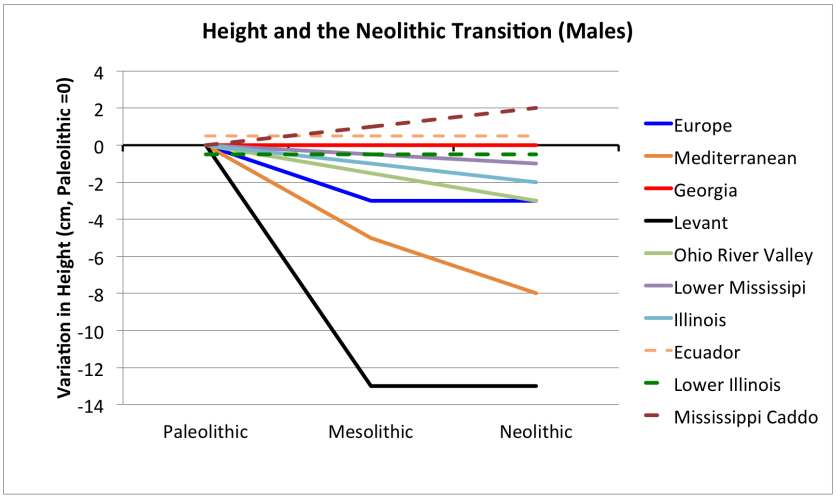

Figure 2.12: Achieved adult height across the Neolithic sequences reported in Cohen and Armelagos (1984). Each line represents the progression in observed heights in one location, expressed as a difference from its value during the Paleolithic (nomadic hunting and gathering). The sedentary farmers (Neolithic) were clearly shorter than their nomadic ancestors. In the cases for which independent data were independently recorded for the Mesolithic (settled hunter-gatherer) phase, the decrease in standard of living can be seen to have predated the Neolithic.

curve and achieving virtually the same adult height as if the starvation episode had not occurred (Williams, 1981). Similar considerations hold for other skeletal disease markers, which also tend to show accumulation of stress factors over time (e.g. tooth wear and joint disease inform us of the average grittiness of food and the amount of labor expended in procuring it, rather than the time pattern of these factors). Thus, the most commonly used health markers are woefully inappropriate for assessing the degree of seasonality in consumption.

However, catch-up growth leaves telltale signs along the length of the bones themselves. Long bones (such as those of the leg) grow from their end outwards. If a growth-arrest episode is ended by a rapid return to favorable conditions, the body will deposit a layer of spongy bone in the normally hollow interior. These layers, called Harris lines, will form a permanent record of the number of growth disruption suffered by an individual before the end of adolescence (Harris, 1933). Harris lines can be examined by sectioning the bone lengthwise, or non-destructively through x-rays (see Figure 2.13).

In most locations where Harris lines were counted before and after the transition, they were found to be numerous during the nomadic-hunting and gathering stage, while comparatively rare during the farming Neolithic. Cohen and Armelagos (1984) report Harris line counts for seven pairs of pre- and post-transition

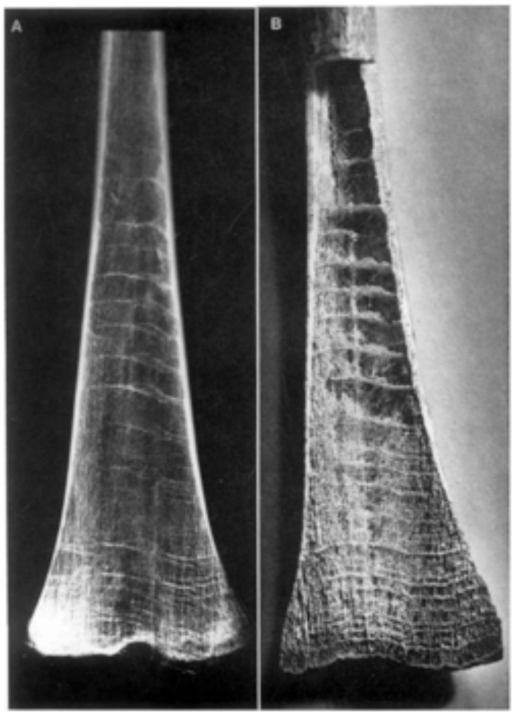

Figure 2.13: Example of Harris lines in an Inuit adult. The regular spacing of the Harris lines show that each winter, food intake would drop low enough to arrest bone growth. Each spring, the arrival of migratory species would rapidly increase food intake, a catch-up growth spurt would occur, and a line for more calcified bone would be deposited (whiter in the x-rays). Such a regular pattern is extremely unlikely to occur due to illnesses. Source: Lobdell (1984)

groups and find marked decreases in five, no significant movement in one case, and a slight increase in the last. For example, nomadic hunter-gatherers in the Central Ohio Valley were 165cm tall on average and had an average of eleven Harris Lines each. When they started to farm, they became about three centimeters shorter but had only four lines on average.

The evidence from Harris lines, together with that from height suggests that hunter gatherers ate well on average, but were forced to starve during part of the year.

2.5 Conclusion

What caused the Neolithic Revolution? I examine the invention and early spread of agriculture and find that increased climatic seasonality was the most likely trigger. Using data on both invention and adoption, I find that higher seasonality made the invention of agriculture more likely, and the spread of farming faster. The channel I propose – increased incentives for storage – explains why farmers accepted a decrease in the standard of living. This interpretation is also supported by the data on the local topography of early sites and the absence of growth arrest lines in their bones.

This paper also helps explain the technological advantage historically enjoyed by the northern hemisphere. Today, New Zealand, Australia, South Africa and Argentina have very similar climates to some of the areas where agriculture originated. Why didn't they invent agriculture? The shock to seasonality that triggered the invention of farming only happened in the northern hemisphere Berger (1992). As a result, these areas never experienced the extreme seasonality that was common where agriculture was invented.

The intuition of the model is relevant to a wide range of settings. Many human societies are subject to seasonal resource availability. If such conditions coexist with inefficient storage technologies, the local inhabitants would experience the same fertility-reducing fasting suffered by hunter-gatherers. The model predicts that such a society would have a lower population density but higher consumption per capita.

Appendix: econometric robustness

Though seven locations show strong evidence of having independently invented agriculture, at least seventeen more are believed to have been important domestication centers (Purugganan and Fuller, 2009). Almost certainly, some of these centers also invented agriculture independently, but archaeologists disagree over which ones. The small number of sites that are universally accepted as independent originators of agriculture, leads to a highly skewed distribution of the dependent variable in the panel of agricultural invention: 38,853 zeros to only seven ones. I address this limitation in two ways: first I repeat the analysis of Table 2.2, using the Rare Events Logit model proposed by King and Zeng (2001). This is shown in the first four columns of Table 2.9. Then, I repeat the analysis of Columns (2) and (3) but using the sample of 24 domestication centers rather than only the seven confirmed adoptions. The inclusion of locations of uncertain invention weakens the power of the analysis considerably, but the signs are preserved and the coefficient on temperature seasonality is significant.

	Dependent variable: adoption dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
	Basic	Controls	Controls2	SI	Neol24	Neol24 SI
Temp. Seas.	0.118***	0.174***	0.199***		0.0898*	
	(0.0443)	(0.0515)	(0.0630)		(0.0462)	
Precip. Seas.	0.263	0.641	0.454		0.0852	
	(0.532)	(0.633)	(0.679)		(0.479)	
Seas. Index				7.219*		2.415
				(4.021)		(1.841)
Temp. Mean		0.0338	-0.133	0.0336	0.0515	0.0542
		(0.0500)	(0.125)	(0.0382)	(0.0446)	(0.0388)
Precip. Mean		0.822***	1.162*	0.784***	0.479**	0.498**
		(0.216)	(0.625)	(0.301)	(0.237)	(0.214)
Abs Lat		0.0487	0.0685	0.0699	0.00912	0.0255
		(0.0344)	(0.0878)	(0.0504)	(0.0409)	(0.0366)
Extra Controls	No	No	Yes	No	No	No
N	38533.00	38533.00	38533.00	38533.00	38533.00	38533.00

Standard errors in parentheses

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

Table 2.9: The effect of climate on invention. Dependent variable is a dummy, which is 1 if agriculture was invented in a particular cell and period and 0 otherwise. Each location is dropped from the sample after they adopt agriculture. All columns: Rare Events Logistic regression on climate variables and controls. Columns 5 and 6: using the 24 possible Neolithic sites instead of the 7 certain ones.

Next, I will explore the robustness of my analysis of the spread of agriculture to changes in the econometric specification. To this end, I collapse the Neolithic

Frontier dataset to a cross-section in which each observation is one location that adopted agriculture from a neighbor. The dependent variable is the number of years that elapsed from when they were first exposed to farming and when they started to farm themselves. For each cell, I assign the average of the values of each explanatory variable during the period the location spent in the frontier. The effect is estimated using a parametric survival model with Weibul distributions, and the results are presented in Figure 2.10.

Temperature and precipitation seasonality both hasten the adoption of agriculture. Increasing temperature seasonality by one standard deviation results in agriculture being adopted 250 years earlier, while doing the same for precipitation seasonality is associated with adopting 200 years earlier. This is equivalent to saying that one extra standard deviation of climatic seasonality made agriculture advance approximately 0.5 km/year faster.

	Dependent variable: no. of periods until adoption				
	(1)	(2)	(3)	(4)	(5)
	Seasonality	Controls	Controls2	Index	Index+Controls2
Temp. Seas.	-33.600***	-36.305***	-17.660		
	(8.335)	(11.015)	(16.856)		
Precip Seas	-22.771	-271.235***	-307.552**		
	(80.707)	(104.015)	(130.880)		
Seas. Index				-1.416***	-1.008*
				(0.478)	(0.581)
Temp. Mean		38.271***	4.223	39.189***	10.740
		(11.272)	(44.643)	(9.358)	(44.619)
Precip. Mean		-151.651***	-159.218	-124.245**	-119.856
		(56.568)	(137.292)	(53.962)	(113.156)
Abs Lat			-56.459*		-65.099**
			(32.189)		(29.837)
GeoControls	No	No	Yes	No	Yes
Climate^2	No	No	Yes	No	Yes
Observations	530	530	530	530	530

Standard errors in parentheses

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

Table 2.10: The effect of climate on the spread of agriculture. The dependent variable counts how long each location waited before adopting agriculture, after first being exposed to it. Each location is dropped from sample after they adopt agriculture. All columns: robust standard errors. The more seasonal the climate, the less the locals waited before becoming farmers.

Finally, I check whether the regressions of year of adoption on seasonality are robust to taking into account spatial correlation. Table 2.11 contrasts the results from three approaches. The first two columns show the results with simple robust standard errors. Columns (3) and (4) show the results for the spatial lag

model. Columns (5) and (6) use Conley spatial standard errors. The coefficients on temperature seasonality are weaker when spatial lags are added to the model, but overall the estimates are remarkably consistent and significant.

	Dependent variable: year of adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
	Basic	Controls	Basic Spat.Lag	Controls Spat. Lag	Basic Conley	Controls Conley
main						
Temp. Seas	-222.5***	-143.8***	-42.4***	-45.5***	-222.5***	-143.8***
	(13.4)	(38.4)	(11.1)	(14.1)	(24.7)	(29.0)
Precip. Seas	-529.4***	-936.5***	-347.1***	-469.2***	-529.4**	-936.5***
	(131.1)	(249.2)	(94.2)	(104.6)	(245.5)	(243.4)
Temp. Mean	107.3***	71.5**	-21.7**	-22.7**	107.3***	71.5***
	(15.9)	(29.6)	(10.6)	(10.5)	(33.0)	(26.3)
Precip. Mean	-464.3***	90.0	-414.1***	-103.6	-464.3***	90.0
	(71.2)	(235.8)	(50.5)	(112.2)	(122.3)	(231.9)
Abs Lat	46.3***	207.6***	-40.3***	29.8	46.3*	207.6***
	(13.6)	(64.9)	(9.3)	(19.2)	(27.8)	(44.4)
Extra Controls	No	Yes	No	Yes	No	Yes
r2	0.24	0.40			0.82	0.86
p	0.00	0.00	0.00	0.00	0.82	0.86

Standard errors in parentheses

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

Table 2.11: Regression of date of adoption of climate seasonality. Columns (1) and (2): robust standard errors. Columns (3) and (4): spatial lag model. Columns (5) and (6) Conley spatial standard errors.

Chapter 3

ALL ALONG THE WATCHTOWER: TATAR SLAVE RAIDS AND THE INTRODUCTION OF SERFDOM IN RUSSIA

3.1 Introduction

Until approximately 1550, the Russian lands were overwhelmingly farmed by free tenant labor (Klyuchevsky, 1911). The Russian peasant did owe rent to his landlord, but he was free to move at any time, and as a result his rent was determined largely by market forces. These freedoms were gradually removed over the course of the following century, so that by 1650 the peasant was bonded to his landowner, and essentially at his complete mercy (Khodarkovsky, 2002). This tragic loss of freedom is all the more puzzling given that most of Europe was experiencing the opposite trend, abolishing the worst strictures of feudal societies, and letting the peasants relocate at will.

The traditional explanation for the enserfment of the Russian peasant was provided by (Domar, 1970), who argued that since land was abundant in Russia, free agricultural labor would have resulted in very high equilibrium wages. Landowners therefore had very strong incentives to collude and keep the labor force enserfed. In the West on the other hand, population density was much higher, so that wages were in any case reasonably close to the subsistence limit. Under these conditions, it made no sense to spend a lot of time and effort controlling peasants and restricting their movement, and it was easier and cheaper to simply pay them

the fair market wage.

The problem with this theory (as Domar himself pointed out), was that it would predict a return to serfdom in Western Europe after the Black Death essentially halved its population, while instead the process of serf enfranchisement continued, indeed in many cases accelerated. Domar therefore proposed that the effect of population density on the conditions of the farming population was fundamentally ambiguous: on the one hand, a lower population density increases the equilibrium wage that agricultural labor would receive if it were free, but on the other it also increased the incentives of the landed class to dramatically restrict their freedom.

I propose a new theory for the introduction of serfdom in the Russian lands, which generates unambiguous predictions on which types of societies should adopt serfdom, and which should allow free movement of agricultural labor. Central to my theory is the fact that the decision of laborers to locate in different areas can impose important defense externalities on their fellow citizens. In the Russian case, the primary threat came from Tatar raiders to the South, who would regularly scour the countryside in cavalry raids up to 80,000 men strong (Khodarkovsky, 2002). Their primary objective was capturing slaves, which were then mainly sold onwards to the Ottoman empire. Since the raiders had superior mobility and no logistical tail, it was impossible to block these raids by building fortresses on the main lines of communication – the raiders could simply bypass such point defenses. An efficient way to defend against such raids was by adopting a cordon defense (Clausewitz and Howerd P., 2007), which in the Russian case meant building – and manning – a continuous series of ramparts and palisades stretching over 1000 km.

The fundamental theory advanced by this paper is that while this system could potentially secure an enormous area, it was unfortunately incompatible with the free movement of labor. Crucial to the security of this system was its spatial continuity: the Tatars would certainly find any gaps and exploit them, putting in jeopardy the entire population of the state. Inevitably, parts of the defense line would need to cross areas that were less fertile than others, where the marginal productivity at any given population density would necessarily be lower. If peasants decided to leave for more fertile areas – which paid higher wages – the allocative efficiency of the economy would increase, but a dangerous gap would have been opened in the defenses. For the labor of those peasant was necessary to ensure that the local land owners/soldiers had the means to afford horses and arms, as well as report for duty on the wall for months at a time. A theoretical alternative would have been to let peasants reallocate as they pleased, and then tax them (or the landowners), so as to pay salaries to a dedicated military class. Indeed this was the system towards which nearly every Czar from Peter I onwards tried to laboriously maneuver Russia. But in the 16th century, no European country had

the state capacity to efficiently tax a country as agrarian and dispersed as Russia.

To support my theory, I first analyze the Russian defense system, and show how the military and productive spheres interacted within the region. Then I draw on records of serfdom from the 19th century to show that the regions with the highest prevalence of serfdom were those that had to support fortification lines. Specifically, areas that were part of the active defense line in the 16th century – when the army was organized along strictly feudal lines – maintained the highest percentage of private serfdom, while areas that were part of the 17th century defense lines – when the army was undergoing a series of centralizing reforms – maintained a very high percentage of state serfs (a more liberal land tenure regime).

The possibility exists that the correlation between serfdom and the defense lines may be due to some other underlying variable, which has a causal effect on both. For example, certain areas might be more fertile, making both their defense important, and the use of serfs profitable. To avoid possible endogeneity concerns, I construct a novel measure for the suitability for fortifications, which combines the direction of the main nomadic threat, their most frequent target, and the network of rivers which determined both the best axis of attack, and the cheapest fortification construction costs. I then show that areas that were more suitable for fortifications still have more night lights today. This is because the defense line was punctuated by fortified towns, which then persisted, becoming the cities of today. This finding highlights the extent to which the Czar was willing to compromise the economic efficiency – in this case, the placement of towns – to achieve defense goals, and that this particular compromise was only enacted on the southern front, the direction of the nomadic threat.

This research improves on the existing literature along several axis. First of all, I introduce a new theory for the introduction of serfdom, that differs from the existing theories by assuming that – aside from distributional considerations between social classes – rulers had strong preferences over the actual geographic location of their agricultural labor force, and that these preferences were related to the need to efficiently defend the borders. Essentially, I argue that whenever the free movement of labor would result in a geographic distribution of the population which severely complicates the defense of the state, rulers will have a powerful incentive to limit the mobility of the population, potentially by enacting serfdom. Another way of viewing this is by saying that in deciding to move away from low-wage (but strategically important) districts, workers are neglecting the negative security externality that they are imposing on the rest of the population. In this interpretation, accepting a ruler which imposes serfdom is a crude coordination device in the face of an enemy more fearsome than indentured work.

I also show how this theory can help understand the differential impact of the introduction of gunpowder on serfdom between Eastern and Western Europe. In

the West, medieval warfare was based on extended cavalry raids (DeVries, 2003). While soldiers lived as settled landowners during peacetime, for the purpose of warfare they became effectively *pro tempore* nomads, scouring the land untrammeled by supply lines or fixed bases of operations. While the European topography did not require continuous lines of fortifications, defending effectively against such a threat still required some soldiers (and their support population) to live in some areas where it was not economically efficient to do so. When artillery was introduced, cavalry armies were robbed of their operational mobility advantage, and in particular became dependent on roads. Thus, strong fortifications along the main trade routes were sufficient to create impermeable frontiers, even though the space in between was technically passable. As a result, Western European states concentrated resources in improving their urban centers, and could afford to let the rural population allocate according to productive efficiency. Along Russia's southern border however, the lack of natural obstacles and the proximity of the steppe meant that cavalry remained competitive long after the introduction of gunpowder, and the Russian state had to control the disposition of peasants to insure the lines of defense could be manned and supplied.

From a methodological standpoint, this paper is the first to calculate both the axes of attack, and the optimal fortification lines to defend against a specific, directional threat, rather than using generic measures of vulnerability such as ruggedness (as used by Nunn and Puga (2012). This method can be adapted and expanded to modern military threats in a variety of setting, both historical and contemporary. For example, many conflicts in Africa is centered around the use of "technicals" (commercially available 4x4 trucks with a heavy weapon bolted on the back). These vehicles are mostly employed for raids and armed reconnaissance, essentially fulfilling light cavalry missions.

3.2 Literature Review

Domar (1970) is arguably the seminal paper for the discussion of Russian serfdom within economics. As discussed in the introduction, Domar argues that the fundamental driver of transitions in and out of serfdom are changed to the population density, either due to fertility/mortality, or due to the acquisition and loss of new lands. Domar's model was essentially a formalization of Klyuchevsky's 19th century work. A number of recent works have sought to test the Domar model empirically. Duleep (2012) showed that the rigidity of India's caste system covaries across both space and time with the local land/labor ratio. Fenske (2014) found a similar relationship held amongst ethnic groups in Nigeria.

During the '70s and '80s, Marxist historians such as Robert Brenner contested the assumption that demographics were behind the end of serfdom, instead argu-

ing for the primacy of micro-class relations: the interaction between the political capacities of peasants and landowners in different regions. This thesis was itself attacked by Postan and Hatcher (1978), who instead argued for the primacy -with due caveats- of the interaction of Malthusian and economic factors. Specifically they argued that the drop in population caused by the Black Death increased the bargaining power of the labor force, which succeeded in securing more rights. A recent addition to this literature is Peters (2016), which argues that serfdom was maintained where rulers had few alternatives for raising funds, and provides evidence of this using a dataset of serfdom prevalence in Europe. Dari-Mattiacci (2013) proposed that serfdom was less likely for task where effort was harder to observe, and showed that both in Africa and in Ancient Rome, slaves with hard to observer occupations were more likely to be manumitted.

Kaufmann and Pape (1999) argued that the abolition of slavery was first and foremost a moral action, which followed decades of patient and persistent persuasion by a cadre of often religiously motivated campaigners, who slowly but surely forged a sufficiently wide alliance of interests to finally abolish the odious institution. However, while the history of anti-slavery campaigning is clearly crucial for understanding the who and how of abolition, it cannot explain why such campaigners racked success upon success for almost 500 years with very little backsliding to speak of. Haskell (1985) also argued that the abolition of slavery was first and foremost a moral shift, but argues that this shift was caused by the diffusion of a capitalist morality. Indirectly he therefore ties the spread of free labor to the progressive opening of the global trade routes during the age of discover.

Acemoglu and Wolitzky (2011) provide a generalized model of forced labor, which they frame as a principal agent model in which coercion allows the landowner to impose contracts which agents would not accept of their own volition, but which still must be compatible with keeping the serf alive. This formulation allows the authors to develop a number of empirical predictions such as that more productive producers should use more coercion with their workers, which can result in lower welfare for them, despite simultaneously receiving high compensation as well.

Importantly, all of these models consider only the sphere of production as relevant for determining whether labor will be free or not. In contrast, I note that the decision to enserf peasants also has important implications for the defense of the realm, and that serfdom might be adopted even when economically efficient, if it is crucial for keeping enemies out.

Parallel to the research effort on the causes of coerced labor, a growing literature has analyzed its long run effects. Dennison (2011) analyzed archival sources to reconstruct the demography, economy and sociology of a specific feudal estate in Russia. Markevich and Zhuravskaya (2017) looked at how agricultural,

trade, and industry all gained considerably after the abolition of serfdom. Buggle and Nafziger (2015) examined the long shadow cast by serfdom on even recent economic outcomes, and their empirical analysis found suggested that the main transmission mechanisms were low provision of public goods, lower urbanization, and delayed industrialization.

3.3 Historical Background

	-1250	1250	1500	1550	1600	1650	1700	1750	1800
Important Czars	Various Principalities Kiev Dominant	Tatar Yoke	Ivan III	Ivan IV	Boris Godunov	Alexis I	Peter I	Catherine II	Alexander I
Serfdom	Mostly Free labor	Mostly Free	Yuri's Day	Forbidden Years	Fixed Years	Fixed Years = Forever	Peasant Revolts, Serfs can Enlist	Secularization	Three Day Corvee
Defense Lines	Some local defense lines against Crimean raiders	None	Oka Line	Zasechnaya Cherta	Doubling of Zasechnaya	Belgorod Line	Izium Line	Siberian Lines	Battles with Prussia
Warfare	Squabbling between principalities	Mongol Invasion	Standoff on Ugra	Many Crimean Raids	Fire of Moscow, Molodi	Battles with Poland and Chinese empire	Wars with Sweden	Vs Turkey and Prussia	First Mobilization of Serfs
Military Organization	Feudal	Feudal	First use of Musketeers, More Artillery in Towns	First Regular Army, Opricihina, Heavy Weapons Admin	Military Statute, Foreign Formation Regiments	Service People Can't Become Serfs, Feudal Order Abolished	First Conscription, No More Hereditary Lands	No More Mandatory Service, Cossack state Liquidated	Essentially European

Figure 3.1: Timeline of relevant Sovereigns, Labour Arrangements, Fortified Lines, Conflicts, and Military organization.

From the 9th century onwards, the East Slavic tribes had been tied in a loose confederation of principalities known as Kievan Rus (Klyuchevsky, 1911). At its height in the 11th century, Rus stretched from the Baltic to the Black Sea, and from the Vistula to the Volga. Its main source of wealth was its control of the river trade routes through which furs and other forest products flowed south from Scandinavia to the Byzantine empire. The capital was in Kiev, which controlled trade along the Dniepr river, and other important centers were Novgorod in the northwest, and Vladimir in the Northeast. As invaders from the East continued to put pressure on Constantinople, the trade network began to brake down, and Rus fell into decline. Peripheral principalities fought for greater independence from the center, and the confederation descended into a series of armed conflicts of varying intensity.

It was against this backdrop that the Mongol invaders burst onto the scene. In 1237 a Mongol army under Batu Khan invaded Kievan Rus (Saunders, 2001). They first thoroughly ransacked the Vladimir Suzdal region, and cowed Novogorod into submission. Then, in 1240 they sacked Kiev, brought under their control the entire region. This led to two and a half centuries of the so-called Tatar Yoke, where all the Russian principalities were forced to pay heavy tribute to the Mongols, and occasionally participate in their military campaigns.

It is important to understand that the Mongols did not occupy Russia in the same sense that the Germans occupied France. By and large, the areas under their control were left to be governed by their traditional power structure, with no permanent military presence. So the problem faced by the Russians was not how to kick the Mongols out, but rather how to avoid retribution if they ever decided to stop paying tribute.

Amid this general background, Moscow slowly but steadily rose in importance. The statelet arose in 1283, from the division of the lands of Alexander Nevsky between his sons (Khodarkovsky, 2002). Daniel, being the youngest, received the least desirable principality. By vigorously courting favor with the Mongols, shrewd political maneuvering, and occasional open conflict, the tiny principality was able to secure its safety from both Tatar raids and neighboring Russian principalities, eventually obtaining the right from the Mongols to collect the tributes owed from the other principalities, before forwarding them to the Khans to the South. This right allowed Moscow to grow richer and more important, so that over the next two centuries it was able to annex its main rivals: Novgorod, Vladimir and Tver.

Having secured their position of supremacy over the other Northern Russian principalities, the next logical step for the Moscow Princes was to shake off their submission to the Mongols. After an initially promising, but ultimately doomed attempt by Dmitry Donskoy in 1380, the turning point came in 1480, when the forces of the Golden Horde, angered by Ivan III's repeated refusal to pay tribute, marched on Muscovy with 100,000 troops, their Khan Ahmat at their head (Khodarkovsky, 2002). They had secured an alliance from the Polish king Casimir IV, who was to attack Muscovy simultaneously from the west. Ivan's reaction was two-fold. First, he convinced the Crimean Khanate, which by that point had broken off from the Golden Horde, to raid southern Poland-Lithuania, tying down the Polish army and preventing them from launching their western attack. Second, he divided his forces all along the Oka river, reinforcing the garrisons at Kolomna and Serpuchkov.

Ahmat's scouts reported that the Oka was strongly held, so the horde tried to bypass the defenders by detouring to the west, in hopes of crossing the lightly defended Ugra river. However, Ivan had already sent forces commanded by his son and brother to Kaluga, the fortress at the confluence of the Ugra and Oka rivers.

From this strategic position, they could easily cover the fords Ahmat was planning to use, and his advanced guard had to abandon their attempts to force a crossing in the face of stiff musketry fire, and wait for reinforcements. Over the next several days, both armies took positions on the opposite riverbanks. Neither side was sufficiently confident to attempt an opposed river crossing, so the expected battle turned into a waiting game. Ahmat had a far larger army, but Ivan had musketry and artillery units in good defensive positions, which minimized their exposure to the Horde's archers. In early November, both sides lost their nerve almost simultaneously. Ivan was worried that the river would freeze and enable Ahmat to cross easily, and therefore retreated towards more defensible ground. Ahmat was finding it hard to feed his army, and suspected the retrograde movement was a trap. In the end both armies essentially ran away from each other, with neither side giving chase. On their way back to the Caspian steppe, the horde plundered a few Lithuanian towns as retaliation for Casimir's failure to provide the required support.

This tactical draw, which later became known as the Great Standoff on the Ugra, was in fact a resounding strategic victory for Muscovy. Ivan III had now been refusing payment to the Golden Horde for a decade, and the Khan had been unable to punish him for it. Not only had he failed to reach Moscow, but he was even incapable of undermining its economic base by ravaging its countryside (Dennis, 1985). This defensive campaign taught Ivan and all future Moscow rulers three important lessons. First, it showed that defeating the nomads did not require a pitched battle. Given the short campaign season, making an invading army waste time was almost as good as defeating it (in modern warfare this is called a "mission kill"). Second, it was crucial that the defending armies present a continuous line of defense along all possible avenues of approach. This would prevent the nomads from exploiting their superior mobility to outflank single fortresses. Third, the defense line had to be sited so as to cover not only the main population centers, but also the farmland from which the towns – and the fortification line itself – drew their support. These three principles would guide Russian defensive policy agains nomadic raids until all steppe areas were eventually annexed, some three centuries later.

By the time of the Great Stand on the Ugra, the Golden Horde was already declining due to internal strife. The inconclusive campaign accelerated this trend, and by 1502 their capital of Sarai had been sacked by the Crimeans, who then became the dominant nomadic force in Ukraine and modern southern Russia (Khodarkovsky, 2002). Freed from their common enemy, the alliance with Muscovy promptly broke down, and the Crimeans started a series of raids that lasted until their eventual annexation in 1783. It was completely unrealistic for the Crimeans to hope to submit Moscow as the Mongols had done. However, they could plunder their lands and sack their cities, which was the next best thing.

Figure 3.2: The standoff on the Ugra River of 1480, as depicted in a 16th century manuscript. Note that while the two armies appear to be equipped very similarly, only the Russian (on the left) have artillery and arquebuses.

The Crimean peninsula was an ideal base of operation for Tatar slave raids. While it may looks small next to Russia and the Ukraine, at 27,000 sqkm it is slightly larger than Sicily. It has a mild climate, which ensured an abundance of winter pasture for up to several hundred thousand horses. The steppe areas of Southern Ukraine the north were even more extensive, and while they were largely impracticable in winter, they could support millions of horses from spring to fall. Crimea was separated from the mainland by an isthmus only 8km wide, which gave almost the same defensive advantages of being on an island, while allowing the movement of men and their horses (both for the normal transhumance, and for raiding expeditions) without need for a navy. The coastline had several deepwater ports, which provided excellent markets for the export of slaves to the Ottoman empire (much of it mediated by Genoese merchants). Finally, its position between the Dniepr and Don gave it ready access to the watershed in between the two rivers, which allowed them to move north without having to cross any significant geographic obstacles (Klyuchevsky, 1911).

The raids could range from simple family groups sending a few dozen riders to pick off farmers who had strayed to far from their settlements, all the way up to well coordinated operations consisting of up to 100,000 riders advancing in multiple columns. For these larger raids, a common strategic plan would see three or four columns of 20-30,000 advancing along parallel lines. Whichever column encountered the main Russian force would establish and maintain contact, essentially trying to waste the defenders time. Meanwhile the other two or three columns would strike deep into the enemy farmland, fanning out in groups of several hundred to capture as many slaves as possible (Klyuchevsky, 1911). Then the columns, swelled in size by their captives, and weighed down by their loot, would start the difficult trek home, usually under constant attack by Muscovite forces which tried to free the captives. The Crimean Tatars launched large scale raids in 1535, 1539, 1542, 1544, 1552, 1553, 1555, 1556, 1560, 1563-65, 1568-74, 1576 and 1580, or 21 years out of a 45 year period.

These raids followed a series of trails, or Shlyas, the most famous of which was the Muravsky Trail. These routes followed for the most part the watersheds between the big rivers, though they would occasionally cross them at convenient fords if it avoided a long detour, or a strongly defended area. Each route should not be understood a single road, but rather as a braid of roughly parallel, but intersecting trails, amongst which traders, troops, and animals selected the most appropriate one based on which ford was open, which streams were flooding, and which plains had good pasture.

The main objective of the raids was the taking of slaves, that were partly used locally, but mainly and sold on to the Ottoman Empire through the intermediation of Genoese traders in Kaffa (today Theodosia). Slaves were a versatile, easily marketable resource, not least because Moscow had enacted a specific tax to pay

for the ransom of captives who's families were unable to pay the price. Estimates of the total number of captives taken vary widely, with one source claiming that 3,000,000 slaves were taken from all Slavic lands from the 14th to the 17th century, while another account is for 150,000-200,000 between 1600 and 1650 alone (Davies, 2004). Whatever the precise number, these flows were enormous when compared to the total population for Russia in 1600 which was around 13 million, the vast majority of which was either incapable of surviving a trek of 1000 km, or else lived too far North to be under any threat from raiding parties. For able bodied men and women along the southern steppe frontier, raids parties must have been an ever present and terrifying danger.

The initial Mongol campaigns and the nearly continuous slave raids had depopulated a very vast area around the Crimea. Within 500km of the Black Sea, virtually all settlements had either been torched or abandoned. This vast area,,comprising some of the best black earth farmland in Russia, was known at the time simply as the "Wild Fields" and was used only for summer pasture by the Nomads, or traversed by small groups of enterprising hunters and trappers.

To defend against this extreme level of threat, the Russians tried to refine and systematize the lessons learned on the bank of the Ugra. The river fortresses were improved, and their garrisons expanded. While this afforded a measure of security, the Princes of Moscow (and later the Russian Czars) faced a fundamental logistical asymmetry working against them. While they could continue to parry the blows as best they could, there was simply no way for them to strike back at the Crimeans directly. Perekop was 1000 km from the Oka river, or at least 50 days of marching for infantry. As the country thawed from South to North, the nomads could just about reach Moscow by starting early in the campaign season, but the Muscovites had to necessarily wait longer before starting out.

Further, reaching Perekop required passing through the Wild Fields, where no food could be bought or requisitioned, given that they were nearly sterilized of any farming communities. Again, the nomads could drive herds of sheep or horses along with them to sustain them during their approach march, through what was essentially an immense pasture, but the cereal-based agriculture of the Moscow state provided no such quantities of livestock. Without the ability to strike back against the Nomadic economic infrastructure, it was only a matter of time before a plague, war, or revolt would sufficiently weaken the defenses on the Oka, and then the Crimeans could again invade Muscovy and cripple it, perhaps forever.

Since mounting an expedition against Perekop directly was out of question, the obvious solution was gradual expansion. Moscow would go on to establish further defense lines to the South, consolidate them over a few decades, and repeat the process. Over a few centuries, this would bring them within striking range of Perekop, and allow the annexation of Crimea. Meanwhile, it would expand the area controlled by Moscow, and make financing the defenses easier. While this

strategy was viable, it faced the complication that the Oka was the only river cutting the entire theater of operation from East to West. The other major rivers of southern European Russia all flow from North to South, and while some of their tributaries do flow in the right direction, inevitably a gap 40 or 50 km wide separates their headwaters. Further, most of these tributaries have smaller catchment basins than the Oka, and are correspondingly shallower, narrower, and slower.

To overcome this difficulty, the Czars essentially built their own versions of the Chinese Great wall, though on a smaller scale (Klyuchevsky, 1911). The first such work was approximately 300 km long, and was built 30 to 50 km south of the existing Oka riverbank lines. Where it could, it followed any small watercourse that happened to flow in the right direction, but otherwise it used a mixture of earthen ramparts, areas intentionally left forested or newly planted, palisades, and abatis (a continuous line of felled trees in the direction of the enemy, with their crowns forming an interlocking mass of branches). Along the line, a series of forts of various sizes were built to house the garrisons and control strategic passages. The construction was completed in 1566, and on this same pattern further fortifications were built further to the south, and farther to the east.

Over this same period, the life of Russian peasants underwent a fundamental transformation. Serfdom and slavery were relatively common in the Oka-Volga triangle before the 12th century, but the local landowners gradually implemented a policy of attracting farmers to their largely unsettled land by promising them hereditary use of land at moderate fees, as well as wide personal freedoms and self-government (Khodarkovsky, 2002). Many of the immigrants were former serfs from Germany or the Ukraine. These privileges had initially been restricted to the newcomers, but overtime they were largely extended to the native population as well. After all, just like the newcomers, they too could run away and obtain better conditions somewhere else.

Overall, this transition towards freer agricultural labor could be said to have had essentially no losers, except perhaps the landowners in the areas from which these migratory streams originated. The newcomer received underpriced lands, and greater personal freedom. The landowner received low, but positive rents from lands that had previously remained uncultivated. And the native population received greater economic and personal freedoms, due to their raised awareness of what they could gain by emigrating. Quite simply, the size of the pie had grown larger, so everybody could enjoy a larger slice.

This trend reversed in the second half of the 15th Century, when the Czars enacted a series of policies which chipped away at the acquired freedoms of the peasants. During the first stages of this process, the new regulation mostly restricted the ability of the peasants to abandon the lands they rented before they had paid all debts to their landlords, as well as any other outstanding service obligations. Ivan III law reform of 1497 restricted the peasant's freedom to change

landlord to the two weeks before and after Saint Yuri's day, or November 29th, when the agricultural year was over and all dues were accounted for. From then on, the major steps all occurred at roughly 50 year intervals. The 1550 law code reform of Ivan IV confirmed this limitation, and also stated that peasants were required to pay an extra fee to their landlord to be allowed to move. In 1581 the Czar declared a temporary ban on peasants changing landlords (the "Forbidden Years"), which was then made permanent in 1597 by his successor Boris Godunov. However, the statute of limitation on runaway serfs was five years, and Russia at this stage was a large country with a permissive rule of law along its frontiers, so enterprising serfs could still go on the lam for half a decade and acquire freedom. This loophole was eventually closed by Czar Alexis I in his 1649 law code reform, which eliminated the status of limitation on runaway slaves, and further punished anybody who knowingly harbored them.

3.3.1 Gunpowder and the Introduction of Serfdom in Russia

The Russian system of anti-tatar defenses based on uninterrupted lines was entirely dependent on the vast superiority of Russia in terms of firearms. Both artillery and muskets infantry were distributed along the line, and both were under direct supervision of the Czar (the cavalry forces were instead typically minor gentry commanded by noblemen). These weapons were extremely powerful in prepared positions. Inside a fortification, the soldiers were almost entirely safe from arrows. The static positions could be supplied with far more ammunition than could be carried on campaign. And the lines were invariably protected by natural or artificial obstacles, which made surprise attacks impossible, and force any assault to spend a lot of time at the optimal range for engagement from the line.

Despite these benefits, a cordon defense is in itself extremely risky. Defending the Oka-Ugra line required stationing troops over a perimeter of nearly 400km, or almost three weeks' march. Such a system ran a very real risk of spreading the forces so thinly that any determined Tatar attack would inevitably crash through. An example will clarify this difficulty: let's imagine that Moscow has 50,000 troops to man the defense cordon, while the Tatars are invading with 50,000 [1]. In general the attacker should have more troops than the defenders to have any hope of victory, so one would naively expect the Russian troops to easily repel the attacker. But in fact the average troop strength along the line is only 125 men per km. If the Tatars detached 10,000 men to form five feints of 2,000 troops all along

[1] These were in fact close to the total troop numbers that were available for both sides in normal times (Filjushkin, 2008). Special mobilizations could double this figure, but only for limited periods

the Oka, and concentrated the main force behind only one of these feints, they would attack at the chosen spot with 42,000 troops. How many would the Russian have? Even assuming that they were somehow able to obtain 24 hours' notice of where the real crossing would be attempted, and were thus able to gather all troops within 20km of the chosen spot, they would only be able to gather 5,000 troops to resist the assault. This would give the Tatars a better than 8 to 1 superiority in men. If equipment was similar on both sides, these are excellent odds for conducting a forced river crossing. In practice however, the Tatars would quickly realize that an enormous stretch of river on either side was completely undefended, and could therefore keep the defenders occupied in one location with 10,000 men, while the majority of the forces was sent to cross unopposed.

On the other hand, if the Russian defenders could count on artillery and musketry fire, even a few thousand troops could create a murderous crossfire over any fording site in the area, essentially ensuring a prohibitive loss rate for any Tatars attempting a forced river crossing. It should be remembered that the Tatar raiders were not fighting to defend their homeland, or for deep seated ideological reasons. Slave raiding was a business, and while the participants accepted that there was some unavoidable risk of dying, none of them could be expected to willingly sacrifice himself for the greater good. Had the Tatars possessed similar amounts of artillery, they could have poured fire on the defenders from their own river bank, suppressing the Russian troops long enough for the assault elements to affect the crossing with comparatively light casualties. But dragging artillery for 800km would have inevitably slowed their advance, and ruined any chance of maintaining the element of surprise.

Therefore it was only when firearms became relatively common in the late 15th century that a cordon defense became possible. However, this system still required a very large number of soldiers to report for duty on time each year, in a precisely arrangement, and ready to fight to the death against an enemy with complete manpower superiority. How could such a force be sourced? Today, the problem would be solved by imposing a proportional (or indeed progressive) income tax on all citizens, and by using part of the revenue to pay for the salaries and equipment of a dedicated professional army, which could then be ordered to take positions along the chosen lines of defense. The needs of the soldiers would be provided by a dedicated logistical service, that would buy food from farms close to the line and also pay for its transportation where it was needed.

However, this modern option would have been entirely impracticable in 15th or even 18th century Russia. First of all, the economy of the country was overwhelmingly based on subsistence agriculture, and it is exceptionally difficult to raise revenues from produce that for the most part never reaches even the local town market. Secondly, even if sufficient revenues could on average be raised, it would have been almost impossible for the central government to guarantee pay-

ment of the soldiers in the face of the occasional famine. Thirdly, no European power of the time had yet developed the logistic efficiency to supply tens of thousands of troops dispersed over hundreds of kilometers. And fourthly even if the troops could have been stationed and supplied where needed, they would have been unlikely to put up a spirited defense.

A feudal system provided the obvious answer to these problems. Having decided which soldiers should have been stationed in which section of the fortifications, it was a comparatively simple affair to assign them lands in the general vicinity of their duty station. The only "tax" required of these soldiers was their reporting for duty when requested, something that was easy to unambiguously observe (Klyuchevsky, 1911). Except in extreme cases, the soldier was responsible for smoothing his own income stream in the face of bad harvests, and in any case most of his equipment was durable. He was required to bring enough food to support him throughout the season from his own farm, obviating the need for logistical complexity. And finally, since both his land and his family would normally be quite close to his duty station, he could be expected to be naturally inclined to put up a spirited defense.

In short, the feudal system provided the Moscow government with a system that was fairly profligate in land use (it would ordinarily provide less than one soldier per square kilometer), but incredibly economical in terms of both coin and administrative capacity. The only thing needed to ensure the success of the scheme was that the soldiers be able to derive the income needed for their support, without the need to actually be present for most of the campaign season. Unfortunately, this meant that the cordon defense system, as implement by Muscovy first and Russia later, was in compatible with the free movement of labor.

The path of the fortification line was decided mainly in order to minimize the total length needed to block all avenues of attack, and so as to exploit the natural obstacle present. It was only natural that certain section of it would be more productive than others, either due to differences in soil fertility itself, or because of differences in the access to market for agricultural surplus. Naturally those soldier-landowners located in more desirable lands would be willing to offer higher wages than those with comparatively worse land holdings. In a situation in which labor was scarce, the unlucky landowners with undesirable plots would not be able to find anyone willing to work their lands at the wages they offered. It would thus be impossible for them to report for duty on the lines of fortification, dangerously weakening a section of the frontier.

The obvious solution to this problem was for Moscow to make it impossible for peasants to move from one place to another, thus ensuring that all along the defense line landowners could in fact derive enough income to take their stations.

3.3.2 Geography of serfdom and defense lines

In this preliminary analysis, I will show that serfdom is closely correlated with the defense lines from a geographic standpoint. To this end I use the tracks of the defense lines from (Nossov, 2006), and the data on Serfdom from Markevich and Zhuravskaya (2017). There were four main lines of defense used by Russia in their defense against Crimea. The first was the Bereg (russian for "shore") line, built directly on the Oka river – which is unique in cutting the theater of operations from east to west. The next line – the Zasechnaya Cherta, or Tula line– was built approximately 100km south of the Oka, using whatever water courses happened to be convenient, but also cutting across open ground for considerable stretches. It was completed in 1566. The third line was built almost 500km south of the Tula Line, and was called the Belogorod line. It was 800km long, and completed around 1650. Thirty years later Russia erected the 500km long Izium Line, closing a triangle of land between the Seversky Donets and the Oskol rivers, with the third side formed by the Belgorod line Itself. This last line was now within 150km of the black sea, and provided the final springboard needed for definitively boxing in the Crimean raids. The peninsula itself however would not fall until the close of the 18th century.

It is important to understand that defending these lines was not a matter of simply posting men on them uniformly and waiting for the Tatars to do their worst. Early warning of raid size and axis of advance was essential, and to this end an extensive network of patrol routes and guard posts was stationed up to 600 km south of the defense line proper, as detailed in Figure 3.3 (adapted from Margolin (1948)). The operational complexity of the defense plan is far greater than anything in use in Western Europe at the time, which goes some way towards explaining the need for centralized administration in Russia. To ensure this complex and dispersed war machine could feed itself, it was crucial that the agricultural labor force could be relied upon to remain where it was needed. Since Moscow lacked the state capacity to raise sufficient inducements to motivate them to do so, its only alternative was to force them.

Figure 3.4 overlays the defense line on a map showing the fraction of the rural population which in 1859 was composed of private serfs, i.e. peasants which were considered the private property of the landowner. This was the most repressive form of serfdom widely practiced in Russia, and as shown by the map is associated mainly with the Oka and Tula lines, which were the main line of defense during the 15th, 16th and early 17th centuries, when most of the garrison was composed of *pomeshchik*, soldiers who were given lands to support themselves, and were in return required to provide service during approximately half of the campaign season.

Figure 3.5 shows the equivalent pattern for State Serfs, a category that was

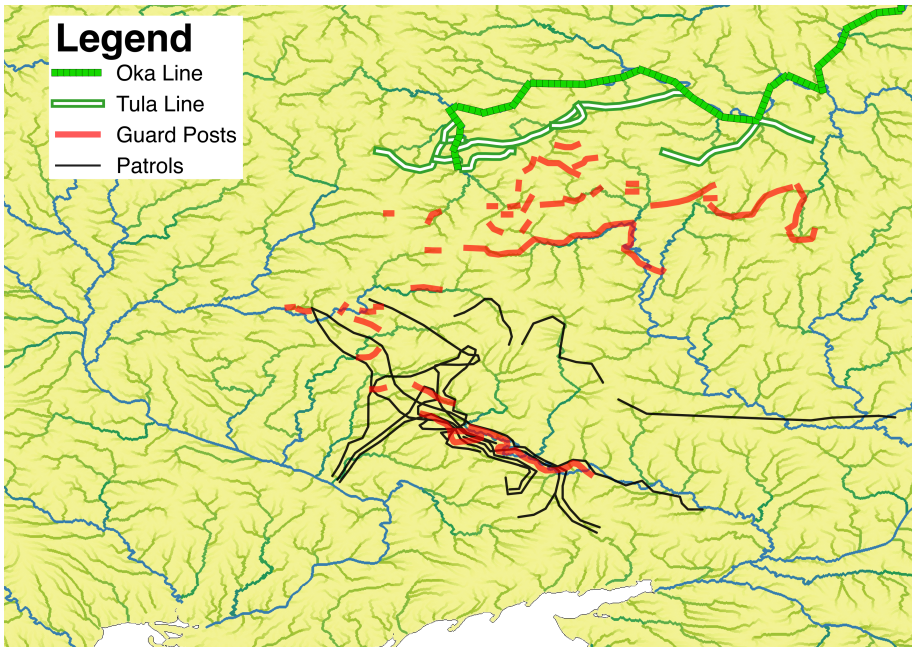

Figure 3.3: An extract of the defense plan for the Russian lands in the late 16th century. The Tula line was the main line of the resistance, with the Oka as the reserve line, but the system include to continuos lines of fixed observation posts on the Visnaya Sosna (north) and Seversky Donets (south), as well as scattered guard posts at various important fords and land bridges. Besides these fixed posts, roving patrols also crossed over the operations area. Both the guard posts and the patrol lines were supported by several advanced fortresses (not shown) usually sited well away from the main invasion paths. Between the two line of guardposts, a band of steppe 200 km wide would be burned to make it difficult for Tatars to pasture their horses.

created by Peter the Great. Though their precise rights varied through time, state serfs in general paid taxes to the state directly, and were in principle free to change residence and occupation. Their land plots were on average around twice as large as those of private serfs, and their standard of living was correspondingly higher. As is clear in the figure, state serfs were concentrated along the Belgorod and Izium defense lines, which were the active frontier when the army became more centralized.

3.3.3 Gunpowder and the Demise of Serfdom in Western Europe

Of course gunpowder was not introduced only in Russia, so why was the effect in western Europe diametrically opposite? To answer this question we must first

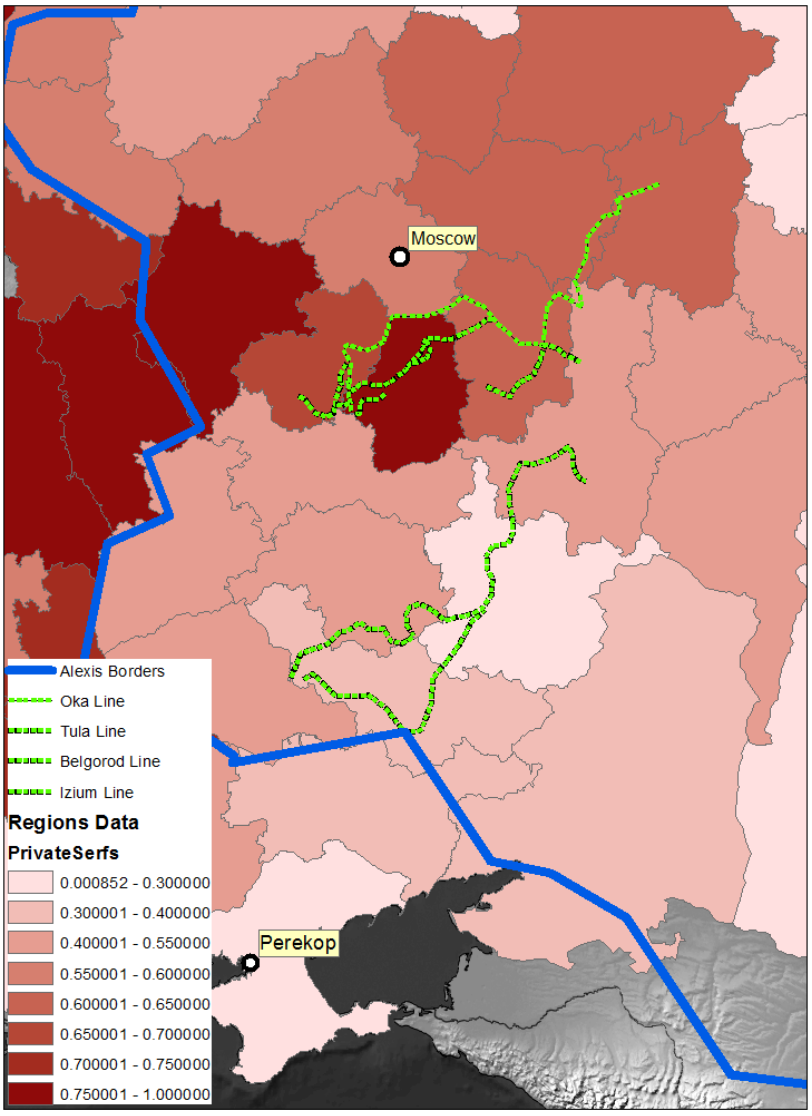

Figure 3.4: The defense lines overlaid over the prevalence for private serfs. Despite the inevitable diffusion during the intervening years, private serf ownership remains concentrated along the defense line which was active during the 16th century, when the Russian army was still overwhelmingly feudal.

characterize how the local military practices differed from those in Russia.

The dominant form of warfare during most of the Middle Ages was the cavalry raid, or chevauchée, as practiced extensively during e.g. the Hundred Years war. Essentially, a group of knights numbering from a few hundred to several thousands

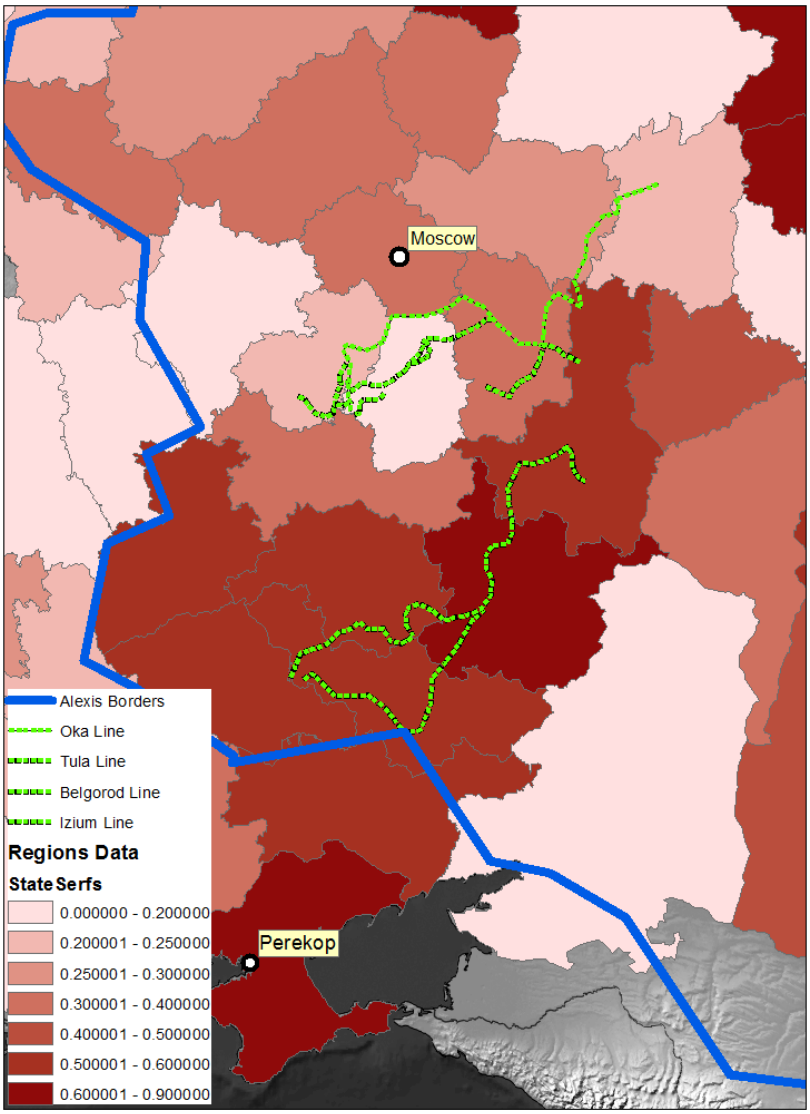

Figure 3.5: The defense lines overlaid over the prevalence of State serf in 1859. State serf ownership remains concentrated along the defense lines which were active during the 17th,and 18th century, when the Russian army (and state) were being centralized over time

would set off for the enemy's land, burning his crops, eating his food. and stealing his valuables (Allmand, 1988). If they encountered a castle that was too strongly held to be taken, the raiding party would simply bypass it and continue in search of easier pickings. If they received news of a stronger group of knights moving to

intercept them, the invaders would try to avoid them, or if that proved impossible retreat to their own territory. Since neither commander would willingly concede battle under unfavorable circumstances, it was only a very skilled defender which would succeed in cutting a raiding party off from all possible lines of retreat, or otherwise luring him into an ambush. Battles were the exception, rather than the rule, and most warfare aimed at destroying - or preserving - the economic basis that made keeping forces in the field possible.

Overall, this type of warfare is broadly similar to the type of raids endured by Muscovy after the 1480s. knights, which led a settled existence during peacetime, would essentially become ”honorary nomads”: for the duration of the campaign, adopting on the whole rather similar strategies and logistic approach. The main difference lay in the armament: while the steppe nomads relied on horse archers as their key force, the western European middle ages were dominated by lancer cavalry. This difference can be easily explained by the different geography in which these force operated. On the flat grasslands of the East, a horse archer could see and hit targets at extreme ranges, and the open terrain meant that he could always escape if superior forces threatened a close quarter battle (Saunders, 2001). On the rolling, forested terrain common in Western Europe, cavalry could not survive unless it was able to sustain close combat.

Another key difference lay in the number of troops involved in both attack and defense. The open steppe allowed enormous numbers of horses to be raised, and also made it easy for raiders to be concentrated from vast areas. In Europe, most of the land was either tilled or left as forest cover, and of the remaining pastures, the vast majority was used for food or draft animals, leaving comparatively small amounts available for war horses. The lack of open pastures also made it difficult for mounted armies to travel long distances, restricting the area from which a cavalry raid could be assembled.

Despite these differences, the underlying problem faced by those trying to defend was similar. Even though it was technically possible to construct point fortifications able to resist for some length of time against even overwhelming odds, such strongpoints were easily bypassed by the enemy, who could go on to plunder less hardened targets. The problem was obviously more felt in border areas, where specific land entities called Marches were created. Since military technology on both sides was essentially symmetric, it would have been suicidal for either side to distribute their forces piecemeal along an exposed border, no matter how well fortified. These frontier areas thus took the shape of a series of castles in depth, generally about one day’s march from each other. The areas directly invaded could take shelter within their fortifications, while the knight in the castles not directly invested could mass their forces and come to the relief of their besieged comrades. Taken together, the various fortifications of the March would shield a less fortified, and more productive interior.

While the architectural means of defense differed (a band of isolated forts, rather than a continuously fortified line), the basic problem remained that all avenues towards the interior needed to be blocked by castles and their garrisons. This meant that some areas needed to be defended, regardless of how unattractive they might be to potential settlers. It is possible that the need of retaining a farming population in these undesirable areas may have provided one of the motives for imposing serfdom on the farmers in these areas, and that the custom spread.

Let us now consider how gunpowder changed this situation. The traditional viewpoint is that the introduction of firearms brought about the demise of cavalry by enabling the lowly infantryman to dispatch the armored knight (Roberts, 1956). A more recent interpretation noted that artillery and muskets led to a new style of fortification, and new siege techniques, which fundamentally changed the dynamics of warfare (Van Creveld, 1991). Both of these intuitions can help us adapt the approach delineated for Russia to the case of Western Europe.

The first element is that gunpowder did indeed reduce the effectiveness of cavalry: besides the greater penetration of armor by firearms, the need of modern armies cumbersome artillery and ammunition trains everywhere severely restricted the operational mobility of armies. This robbed cavalry of its chief advantage, the ability to accept or decline battle at will. Cavalry raids now became much riskier, since it was much more difficult to evade and return home if superior forces were encountered (unless the raiding party was willing to abandon its artillery, which was not economically viable in the long run).

The second aspect is that the new style of artillery forts were not only time-consuming to capture, but also impossible to ignore or bypass. In the medieval period, it was perfectly possible for a raiding parry to sustain itself for months or even years, entirely by foraging from the countryside. After all, almost all necessary supplies were durable (horses, armor, swords), available from farms (food, shelter, tools, firewood), or could be manufactured or repaired while on campaign from commonly available materials (arrows, lances, saddles, horseshoes).

Gunpowder altered this equilibrium, by introducing a substance that was both indispensable for continuing a campaign, relatively heavy/bulky, and which could not be easily sourced in the required quantities from anything but a military installation. An army on campaign could typically only carry enough gunpowder for one major engagement, after which it would be essentially helpless until it either resupplied by cumbersome and vulnerable convoy, or it retreated to its base of operations. Neither of this options was compatible with the type of fluid, free-wheeling campaign which was so common during the middle ages. If resupply convoys were necessary, it would be reckless to bypass any enemy fortification, from which the garrison could later sally forth to intercept the much needed ammunition resupplies. Similarly, if the army had to return to its jump-off point after each major engagement (or string of skirmishes), the depth of the possible

penetration would necessarily be limited.

Essentially, military operations from that point onwards would have to more or less satisfy ”convexity”. Deep penetrations of any kind would be extremely dangerous, and all operations would have to follow established lines of communications such as roads or rivers. Under these conditions, it was completely unnecessary to ensure that remote areas were as well defended as the main roads, and since the latter locations were very desirable for trade and manufacturing purposes, they were already very well populated and needed no coercion to attract more still. For countries in this condition, it was better to set the peasants free and let them flock to where they were most useful.

3.3.4 Raiding Today

The fundamental mechanism for my theory relies on a discrepancy between the economic and strategic significance of areas under the control of a given polity, combined with the impossibility of mobilizing resources produce in one region, for the defense of another. Since the introduction of gunpowder in the early modern period, this discrepancy has been steadily narrowing, under the pressure from both the increasing reliance on wheeled transportation for warfare, which requires the good roads which are only present in economically advanced regions, and the increasing portability of weapons systems and their supporting logistical units. For example during WWII, millions of American workers were able to participate in the European and Pacific struggles without ever leaving their hometowns. While ferrying tens of thousands tanks and airplanes across the Atlantic was no small feat, it was several orders of magnitude simpler than transporting the tens of millions of workers that participated in their construction, had they been needed to participate as infantry.

As a result of these converging trends, nearly all states now rely on taxation to raise funding for specialized soldiers, which are then deployed wherever they are most needed. Such a system has no use for the type of organization employed by feudal states. However, non-state actors often face similar constraints in mobilizing resources from one area for use in another. For example, if a guerrilla relies on friendly or intimidated farmers for shelter and food in a given area of the country, it might be impossible for them to mount sustained operations in a different region. Under these circumstances it is possible that such a faction would want to restrict labor outmigration from the area they control, which could lead to arrangements that are similar in effect, if not in name, to those described by this paper.

Raiding tactics have also made a comeback, particularly in the series of ongoing conflicts in the Middle East (Ignatius David, 2015). In Iraq, ISIS has graduated from executing primarily terrorist attacks, to waging a campaign consisting of a

series of limited-scale military attacks conducted by company or battalion-sized formations (150 to 600 fighters). A typical attack on a state-controlled village defended by an army outpost might see one or two suicide-VBIED (Vehicle Borne Improvised Explosive Devices, typically pickup trucks with improvised armor packed with explosives) driven against the base and detonated so as to disable the perimeter defenses. Infantry would then exploit the confusion to inflict maximum casualties against government troops within and around the base, and potentially also on civilians unsympathetic to their cause. If this attack had been conducted by a conventional army (perhaps using a well-timed airstrike instead of a VBIED), they would follow up on such success by moving in a second and third echelon to mop up remaining defenders, and consolidate control of the village. To have any hope of withstanding the inevitable counterattack, they would need to prepare field fortifications, and they would also need open lines of communication towards their existing strongholds, to ensure that reinforcements and supplies could get through.

ISIS forces would instead deviate from this playbook by quickly withdrawing and dispersing the attacking force, which would typically have suffered only minor casualties. Such attacks would be repeated at random intervals, until the government forces would recognize that their position was untenable and vacate the village. Civilians hostile to ISIS would at this point have no choice but leave their homes and become refugees, paving the ground for ISIS to enter the village at their leisure, and take control with minimal casualties or use of combat resources. Crucial to such tactics is the mobility provided by 4x4 pickup trucks, both for use as VBIED platforms, and to deliver – and just as importantly, withdraw– the infantry forces used in the followup attack. Like the cavalry mounts of Tatar slave raiders and medieval armies, they allow ISIS operational and strategic mobility independent of the road network, which is crucial to avoid detection and maintain the element of surprise. As in those earlier conflicts, ISIS raids make use of land that has little economic value - the desert - to attack population centers, which in the Middle East are strung mainly along watercourses.

How far can this analogy be carried? From the point of view of the Iraqi state, a return to a feudal society is unlikely. In Medieval Europe and Early Modern Russia, the strategically crucial land was economically unattractive, but far from sterile. It was perfectly possible for a family farming even suboptimal farmland could in fact support themselves, and provide the required surplus to their lord, so restricting their mobility was necessary mainly to prevent them from accessing *better* alternatives in cities or more fertile lands. In the Iraqi context, the strategically crucial land is mostly desert, and could not support meaningful population centers regardless of the level of enforcement. In any case the Iraqi state relies mainly on conventional military forces under centralized control, which do not depend primarily on locally available resources. A more limited goal for restrict-

ing population mobility might be to avoid the exodus of anti-ISIS civilians from frontline towns, which is typically one of the goals of ISIS raids, and usually a precondition for their eventual occupation. But since the raids generally result in the Iraqi army units being recalled, or abandoning their positions, it is unclear how such a ban on civilian movements could be enforced.

From the point of view of the ISIS, the situation is more complex. On the one hand, they rely on the local civilian population for both material needs, and to provide human shield against the superior firepower of the Iraqi state and their wester allies. They would therefore have stronger incentives to control the geographic distribution of the population under their control. On the other hand their main enemies are conventional armed forces, with their reliance on roads to conduct offensive operations. This largely eliminates the discrepancy between the actual, and their preferred distribution of civilian in areas they control. Nonetheless, the exodus of refugees from the frontline areas has hurt them a lot more than it has hurt the Iraqi state, by depriving them of forced contributions, human shields, and a pool of potential recruits with no alternative employment opportunities to speak of. As a result ISIS has reportedly resorted to a policy of shooting on site civilians that try to leave the areas under their control without authorization by ISIS commanders. While there is no data to asses whether this has already had economic effects above and beyond the general destruction caused by living in a war zone, the long run incentives of such an arrangement appear clear.

3.4 Model Sketch

3.4.1 Domar Model

The Domar Model assumes that:

1. The state is trying to maximize the amount of revenues it can extract

2. There is a fixed cost in taxing individuals, so that in practice it is cheaper to extract said revenue from the landed gentry, than from individual peasants. Therefore, only landowners can be taxed.

3. The marginal product of labor is decreasing.

4. There exists a fixed cost to keeping an individual peasant enserfed.

5. Under free labor movement, each peasant earns his marginal product of labor. Under serfdom, each peasant will be paid his subsistence wage.

The clear result of this combination of assumptions is that when labor is scarce, the marginal product of each peasant is high, and as a result landowner prefer to enserf the peasant, and pocket the difference between his marginal product and the sum of his subsistence wage and the cost of keeping him enserfed. On the other hand, if labor is abundant, the marginal product of labor is close to the subsistence wage, and it is better for landowner to allow him to be free and simply pay him his low wages.

3.5 Model

3.5.1 Assumptions

Russia is composed of two regions, Road (r) and Back Country (b). In each region, output is produced by combining Land (K) and Labor (L), using an equal weighted Cobb-Douglas production function:

$$Y_x = \sqrt{K_x L_x} \tag{3.1}$$

The endowment of land is equal in the two regions so $K_b = K_r = 1$, as is the initial distribution of Labor $L_b = L_r = 1$. However, the Road region is better connected to market, and therefore obtains a better price for its goods, $p_r = \delta p_b$ with $\delta > 1$.

Tatars have at their disposal a raiding force of fixed force T, with which they attempt to raid the Russian lands. If they succeed they receive a positive payoff, if they fail they receive a negative payoff, and if they don't attack at all they receive 0.

Russia can defend itself using its military strength in each region, which is endogenously determined based on both locally raised forces A_b, A_r (for Army in Road and Back Country respectively), which are equal to the local population, as well as centrally allocated artillery G (for Guns). To simplify notation, I assume that α is the fraction of the total population $A = 2$ which resides in the Back Country, while β, is the faction of the (exogenous) total amount of Guns available which is allocated by the state to defending the Back Country. These factors are themselves aggregated using an equal weighting Cobb-Douglas function, so that:

$$D_b = \sqrt{(\alpha)(\beta G)} = \sqrt{\alpha\beta G} \tag{3.2}$$

$$D_r = \sqrt{(1-\alpha)(1-\beta)G} \tag{3.3}$$

The Tatars can choose where to attack, and have perfect information on the level of defensive forces present in each region. Given the vast distances of Russia's steppe border, any forces the Russians might have in the other region are

completely irrelevant. Therefore, in practice the Tatars only need to match the defensive potential of the worst defended region. We will allow for one region being easier to defend than the other, through controlling a parameter γ, which when equal to one implies symmetric defense potential.

$$D = \min\{\gamma D_b, D_r\} \tag{3.4}$$

Where D is the final defensive potential of Russia as a whole.

3.5.2 Maximizing defense

Let's first assume that the population allocation α is fixed, and Russia wants to allocate its artillery (i.e. pick a β) so as to maximize this defensive potential. To maximize D, it must be true that $\gamma D_b = D_r$. Therefore we can calculate what is the optimal $\beta^\star$

$$\gamma\sqrt{(\alpha\beta G)} = \sqrt{(1-\alpha)(1-\beta)G} \Rightarrow \gamma\sqrt{\alpha\beta} = \sqrt{(1-\alpha)(1-\beta)} \Rightarrow$$
$$\beta^\star = \frac{1}{1+\left(\gamma\sqrt{\frac{\alpha}{1-\alpha}}\right)^2} \tag{3.5}$$

Plugging in, we find that the maximal level of defense attainable will be:

$$D^\star = \gamma\sqrt{\frac{\alpha G}{1+\frac{\alpha}{1-\alpha}\gamma^2}} \tag{3.6}$$

Let's first look at the simpler case in which $\gamma = 1$ (i.e. when both the Road and Back Country are equally defendable),

$$\beta^\star = 1-\alpha \tag{3.7}$$
$$D^\star = \sqrt{G(\alpha-\alpha^2)} \tag{3.8}$$

Equation 3.7 tells us that the Russian state will try to compensate or any imbalance in α by allocating artillery where population is lacking, while Equation 3.8 tells us that the maximum amount of defense potential will ensue when population is equally distributed $\alpha = 0.5$, while progressively greater imbalances will reduce the maximum possible amount of defense. The slightly more complicated equations 3.5 and 3.6 tell equivalent stories, but are simply skewed in the direction of providing more defensive resources to the least defensible region.

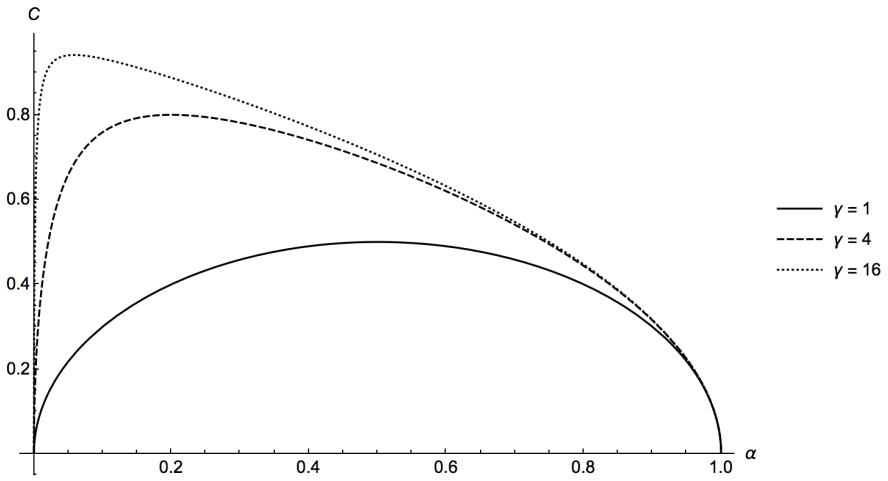

Figure 3.6: The figure shows how the maximum possible effective defense level changes as α (the population distribution) changes. The figure assumes that Russia is allocating its artillery optimally. As γ, (the defense multiplier of the Back Country) increases, the optimal α is more skewed towards having more population in the Road region.

The results of the model are thus entirely intuitive. The easier it is to defend the Back Country compared to the Road, the more the Czar will wish that the available population is available at the Road for defensive purposes. The further the population distribution is from the optimal, the more the Czar will allocate his artillery to the relatively unpopulated region, in order to compensate.

3.5.3 Production and Migration

We now analyze how changes in the price differential between regions translate to differences in population. If the labor force is free (f) to move as it wants, then:

$$MPL_{bf} = MPL_{rf} \Rightarrow L_r = \sqrt[0.5]{\delta} L_b = \frac{\delta^2}{1+\delta^2} L \tag{3.9}$$

Therefore, if the population was initially equally distributed, the net migration towards Road would be.

$$\Delta L = L_r - L\frac{1}{2} = \frac{\sqrt[t]{\delta} - 1}{2 + 2\sqrt[t]{\delta}} L \tag{3.10}$$

We can therefore write α as a function of δ

$$\alpha = 1 - \frac{\delta^2}{1+\delta^2} \tag{3.11}$$

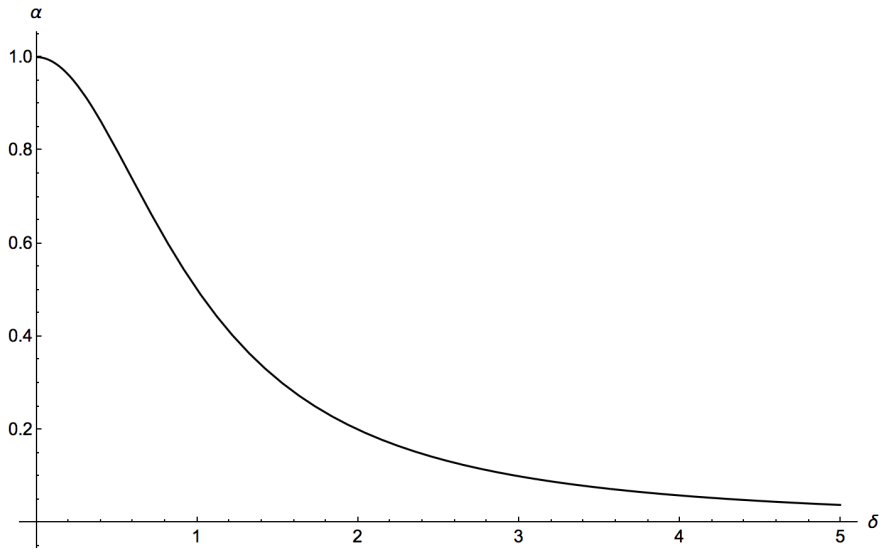

Figure 3.7: Equation 3.11 is easy to interpret. If labor is free to move, as the productivity advantage of the road increases, a greater fraction of the population finds it optimal to move there.

Since $L = 2$ and each region has one unit of land, total output:

$$Y_{bf} = \sqrt{\frac{2}{1+\delta^2}} \tag{3.12}$$

$$Y_{rf} = \sqrt{\frac{2\delta^2}{1+\delta^2}} \tag{3.13}$$

$$GDP_f = \sqrt{2(1+\delta^2)} \tag{3.14}$$

Based on this, the rental price of land (normalizing the price of output in the Back Country $p_b = 1$) will be

$$R_{bf} = MPK_b = \frac{1}{2}\sqrt{\frac{2}{1+\delta^2}} \tag{3.15}$$

$$R_{rf} = MPK_r = \frac{\delta}{2}\sqrt{\frac{2\delta^2}{1+\delta^2}} \tag{3.16}$$

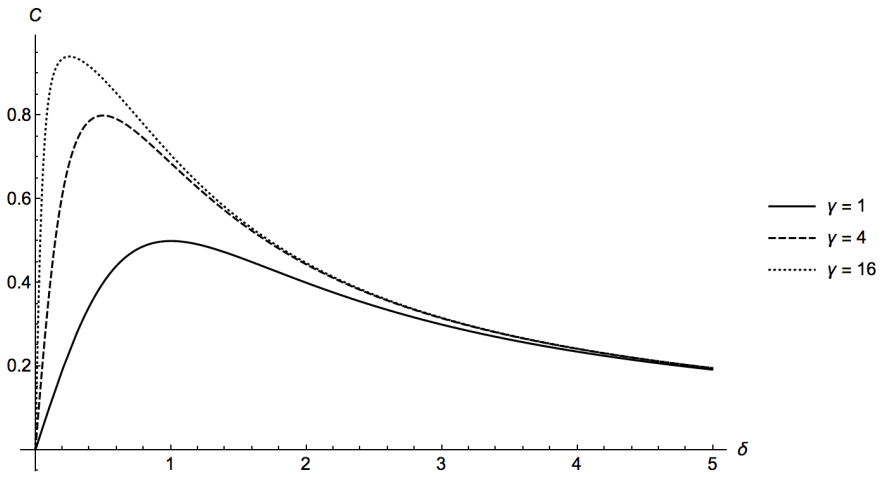

Figure 3.8: If $\gamma = 1$ (i.e. if both regions are equally easy to defend), then the maximum level of defense will result when the productivity of both regions is the same δ. If one region is easier to defend then the other, than the highest defense will result when the other region is the most productive. For example, if the Road was very hard to defend compared to the Backcountry, then the ruler should hope that the soil of the Road region is also more productive, so that more people naturally want to live there.

Wages will equal to MPL:

$$w_{bf} = w_{rf} = \sqrt{\frac{1+\delta^2}{8}} \tag{3.17}$$

Given the amount of population in each region, we can calculate the resulting defense, with optimal allocation of the artillery.

$$D_f = \gamma\sqrt{\frac{2G\delta^2}{(\gamma^2+\delta^2)(1+\delta^2)}} \tag{3.18}$$

On the other hand, if the Czar enacts Serfdom (s), then he can force the labor force to remain equally split between the two regions ($\alpha = 0.5$). in this case:

$$Y_r = Y_b = 1 \tag{3.19}$$

$$\text{GDP}_s = 1 + \delta \tag{3.20}$$

Since workers cannot move, the landowners are going to be able to pay them only the subsistence wage w_l:

$$w_b = w_r = w_l$$

The surplus for the owners will now be:

$$\Pi_{bs} = p_b Y_b - w_b L_b \tag{3.21}$$
$$= 1 - w_l \tag{3.22}$$
$$\Pi_{rs} = p_r Y_r - w_r L_r \tag{3.23}$$
$$= \delta - w_l \tag{3.24}$$

Again, given that we know the defense function, and the population of each region is known, we can calculate the actual resulting defense level.

$$D_s = \gamma \sqrt{\frac{G}{1+\gamma^2}} \tag{3.25}$$

3.5.4 Model Results

From the preceding assumptions and intermediate results, we can derive the following statements (see Figure 3.9 for economic analysis):

1. Serfdom is expensive for GDP:

$$GDP_f - GDP_s = \sqrt{2(1-\delta^2)} - (1+\delta) \tag{3.26}$$

   When $\delta = 1$ serfdom is costless, but for any other value, enacting serfdom reduces aggregate GDP.

2. If Serfdom is enacted, landowners in both regions will benefit, but the highest percentage benefit will go to the Backcountry Landowners. From the point of view of the Road Landowners, serfdom allows them to pay their existing workers less, but makes it impossible to attract as much labor as they could profitably use.

3. Serfdom is bad for workers. Of course, introducing serfdom results in them making only the subsistence wage, rather than the much higher market wage for their services.

4. Militarily, freedom can weaken the defensive posture of the state. As Figure 3.10 shows, the maximum defense potential occurs for a specific value of the population distribution α. The ideal α depends on the defense differential parameter γ, while the realized α will depend on the productivity differential δ. If the ideal and realized α differ significantly, it is possible that serfdom can significantly improve the security of the state.

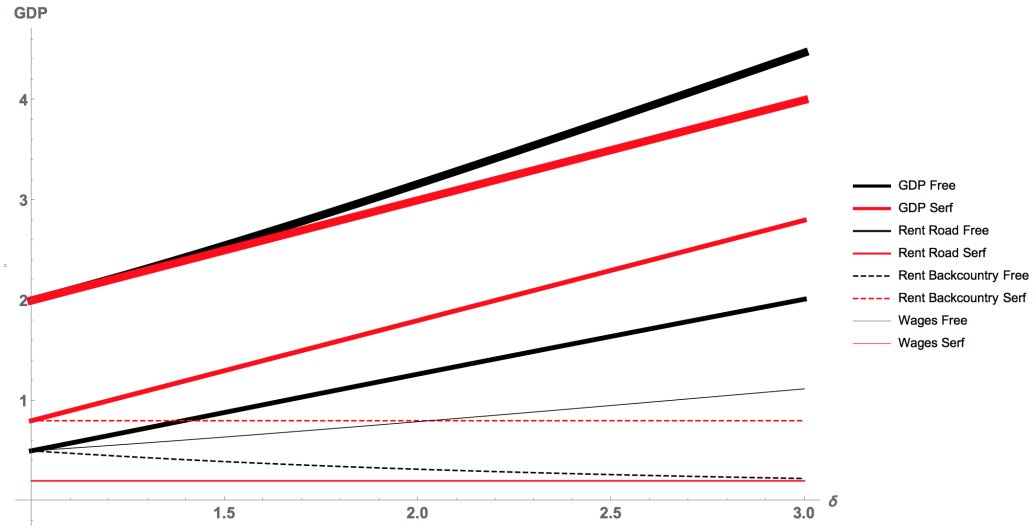

Figure 3.9: As long as $\delta > 1$, under Serfdom: 1)GDP will be lower. 2) Land Rents will be higher in both the Road and Backcountry regions, but the percentage difference will be greater in the latter. 3) Wages will be lower in both regions.

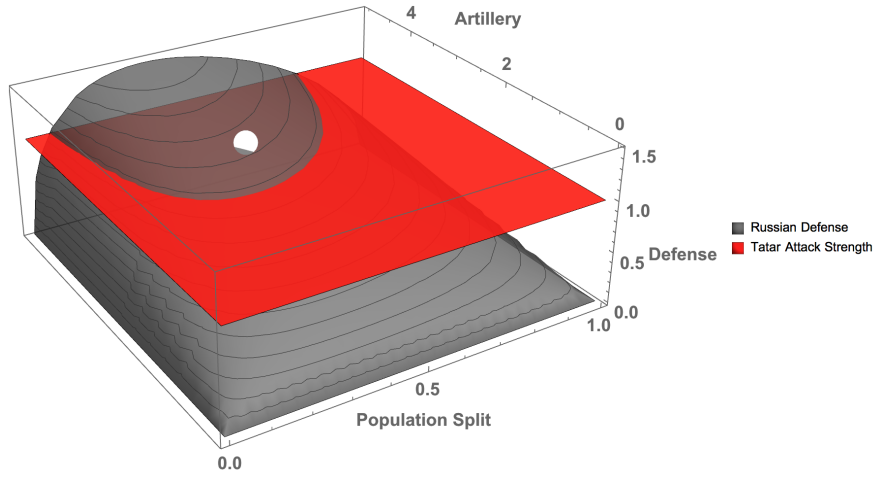

Figure 3.10: Defense potential is highest when the population is equally divided between regions ($\alpha = 0.5$), and the amount of artillery available G is large. For example, with $G = 8$, Russia would be able to beat back the Tatar attack if the population was equally divided ($C > 1.2$, the level of Tatar attack assumed in this picture.) , but not if there was an imbalance of e.g. 0.8

3.5.5 Accounting for the differential adoption of Serfdom

Figure 3.11 summarizes the story of Serfdom in Western Eurasia. The assumption is that the Road region has an output price three times as large as the Back Country $\delta = 3$, and that the population is initially equally distributed between regions. If the Czar allowed labor to allocate freely, this would result in $\alpha = 0.9$. This severe population imbalance complicates defense, particularly if the enemy's attack is not particularly impeded by the lack of roads in the Back Country (that is, if γ is low). On the other hand, γ is high, then enemies rely greatly on roads for advancing, so that defending the Back Country is easy even if very few people actually live there.

In the present model, the choice between serfdom and freedom is chiefly a choice over population distributions. If γ is below a certain value (thick dotted line in the Figure 3.11), then it is crucial that a sufficiently large population resides in the more economically disadvantaged regions (Backcountry), and Serfdom will result in the highest security for the nation. If γ is above the threshold, then free movement of labor will instead be preferable.

In the late Middle Ages, warfare all over Europe was still largely conducted as a series of reciprocal cavalry raids. Groups of knights from a few hundred to a few thousand strong would set out into enemy territory, pillaging villages and weakly defended manors. Since these raids relied entirely on the countryside for resupply, roads were not particularly necessary. If the enemy placed a castle in order to deny use of a road, it was a simple and routine matter for the raiders to detour across country for however long was needed to avoid the position. This was true in France, Spain, or Lombardy (F_1 in the Figure), and it was just as true in Russia (R_1), where the open steppe provided virtually limitless attacking options for cavalry, at least until the large forests of the upper Volga basin were encountered. Thus in both Western and Eastern Europe, Serfdom provided the best defense against they typical military threats. Why the difference in outcomes then?

Because of its proximity to the steppe, Russia was exposed to raids of a size no Western European country had to contend with, and in the initial stages was completely unable to defend itself, under either serfdom or free labor (in the figure, this is shown by their being in the green area). Instead, they had to pay a large tribute to the Mongols/Tatars each year, or suffer terrible consequences [2]. Under these circumstances, it was better for Russia to keep their labor force free,

[2] As citep klyu vol I, page 267 states: "[...]the Mongolian yoke had long ago relieved the princes and their retinues from the obligation of guarding the far south-eastern regions ? the regions which once had served the southern princes as a training ground for their warriors. Indeed, even after the great battle of Kulikovo had taken place, it con tinued to be tribute rather than troops that had to be dispatched thither

which would at least maximize national income and make their tribute payments easier to achieve. While Western European nations faced the same combination of γ and G, their distance from the steppe meant that they did not have to face a nomadic threat, but only invasion from their essentially symmetric neighbors. In this relatively benign threat environment, it was perfectly feasible for them to defend themselves as best they could, which in their case meant controlling their population distribution through serfdom.

How does the introduction of gunpowder weapons change things? There are two main effects: G will increase, and γ *may* increase. First of all, the production of artillery and gunpowder was almost universally under state control. Therefore G, the forces controlled centrally, will by definition expand as each state adopts gunpowder. In the West, any military advantage will be largely offset by the fact that each state's enemies are also adopting gunpowder. In Russia, on the other hand, the Tatar raiders resisted adopting gunpowder weapons, since the distances were too vast to cover during the limited campaign season with artillery in tow. Therefore, artillery will give Russia a decisive advantage in defending their borders, and this is what allowed them to overthrow the Tatar Yoke. But the second effect depends on what their man enemies do. In Western Europe, all of the potential enemies adopted artillery, which in practice meant that major military campaigns had to follow, and maintain continuos control over roads. If they didn't, even a battlefield victory could turn to disaster when the army proved impossible to resupply with ammunition from the logistical base.

In the context of the model, this means that γ increases a lot (Back Country areas can be defended effectively by small forces), which puts Western Europe at F_2, well within the parameter space where free labor is optimal. In Russia on the other hand, the Tatars did not adopt artillery, which meant that they could still attack virtually anywhere along the frontier. Therefore the effective γ faced by Russia alongs it's southern frontier did *not* increase, and the country ended up at point R_2. This explains why gunpowder was associated with the rise of serfdom in Russia, but with its end in the West.

c

3.6 Empirics

3.6.1 Concept

The most direct way to test my proposed theory would be by using variation in the introduction of serfdom in areas that were or were not part of the defense line. Unfortunately, the laws introducing serfdom were mostly passed at the national level, and while they were probably variations in enforcement and eligibility at

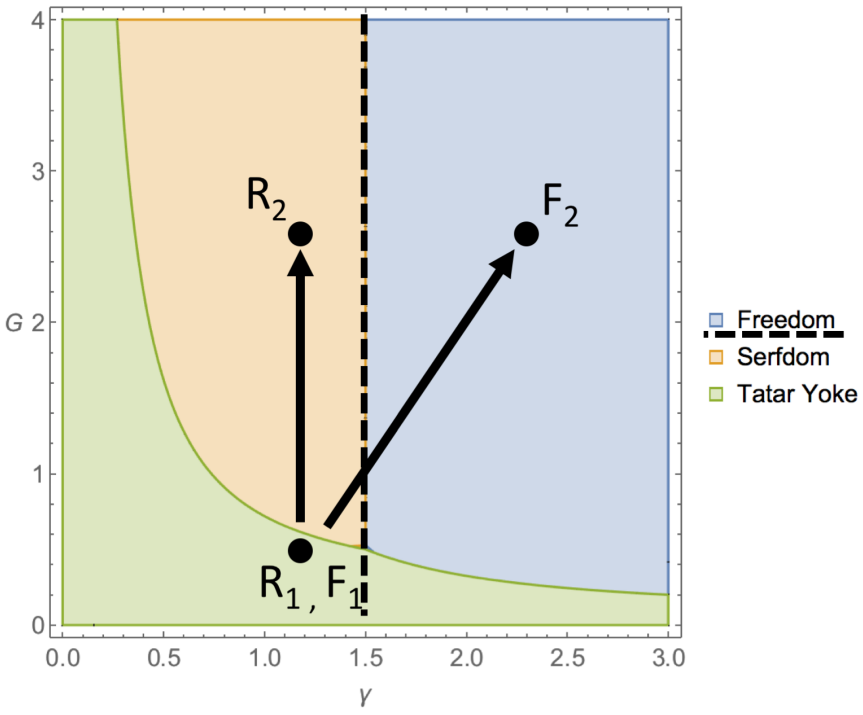

Figure 3.11: The parameter space of G and γ divided into areas where Serfdom and Freedom are optimal (divided by the dashed line). The green area at the bottom and left denotes parameter combinations where a Tatar raid would be able to defeat even a state adopting the locally optimal strategy (assuming it is within raiding range of the Steppe). Initially both Russia and Western Europe occupy the point R_1, F_1, though Russia is within raiding range and Western Europe is not

the local level, these have not been preserved sufficiently to permit econometric analysis.

A more indirect route would involve using panel population data to check whether areas along the defense lines were becoming depopulated before serfdom was enacted, and whether this trend reversed after the serfdom was enacted. Again the earliest census data for Russia sadly begins approximately a century after the implementation of serfdom, frustrating the possibility of conducting a diff-in-diff analysis.

Given that the two most direct test for my theory were impracticable, I decided to use a somewhat indirect test. The idea is that if Moscow was willing to sacrifice economic efficiency of the altar of national defense in deciding how to allocate its

rural farming population, we should see a similar pattern in its policy towards its urban centers. Specifically, we should see cities emerge not where they were most useful for economic reasons, but rather where they were best sited to aid the defense of the country.

Indeed the vast majority of the major population centers between Moscow and the Crimea were originally founded as fortresses against Tatar raids. As Muscovy and later Russia expanded southward, the military function of these settlement subsided, but they persisted as urban centers for their surrounding areas. If my theory is correct, we would expect modern day economic activity in the area to be clustered around the most favorable locations for defense, rather than the best locations for trade or agriculture.

The easiest way to test this hypothesis would be by using the actual historical location of the various defensive lines built by Moscow to defend its lands. However, this approach would run into serious endogeneity problems, since it could well be that the defense lines were sited so as to defend land that was particularly valuable from an economic standpoint. To avoid this risk, I have designed an algorithm (detailed below) that uses reasonable assumptions to construct a series of optimal defensive lines taking into account only the ease of defense, and completely ignoring trade or agricultural concerns.

3.6.2 Data

I first collected the geographic data I needed, namely the flow accumulation raster for the area in question from the Hydrosheds project Lehner, B., Verdin, K., Jarvis (2008), and the potential yield for barley (the most commonly cultivated grain in the area at the time), and for pasture grasses (which is a proxy for the ease with which the tatar raiders could pasture their horses). Both agricultural suitabilities are from the FAO GAEZ project version 3.0.

My next step was to collect historical data on the most commonly employed invasion routes, and the precise location of the fortification lines. These came from Nossov (2006).

I also used the maps for the prevalence of serfdom in 1861 from Markevich and Zhuravskaya (2017).

My final step was to download the city light data from NOAA. I use city lights data because it allows me to abstract from the possibility of undereporting of economic activity in the official statistics, and particularly for the possibility of differential bias between the modern day countries involved (Russia, Ukraine, Belorussia, Lithuania).

I also construct rasters for latitude and the distance from Moscow.

3.6.3 Variable construction

Invasion routes

The first step is to construct the optimal invasion paths. A naive view would be that raiders would always take the shortest path, much like a trader would, but of course if this was the case, the Russians could have defended by concentrating their army along that one path. To avoid this simple countermeasures, raiders actually played the mixed strategy equilibrium of attacking along different paths in different years, usually by conducting several simultaneous feints along paths not being used to mask their true intentions. However, given the short length of the campaign season in Russia, the raiders still had to keep a somewhat direct route towards Moscow, or risk being caught by winter before having reached their objectives.

To find out the optimal invasion routes, I used the following procedure:

1. First I calculated a cost raster, approximating the cost for a nomadic raiding party. To do this, I took the flow accumulation value [3] for each cell.

2. Due to the way river networks naturally branch, this raster as extremely high right skewness (most cells have very low values, while the terminal part of large rivers have extremely high values). To correct this problem, I took the square root of the flow accumulation, which is approximately proportional to the width of the resulting river (beyond a few feet, the difficulty of crossing a river does not depend on its width).

3. I then calculated the lowest cost path between Perekop and Moscow, using the cost map resulting from step 2

4. Starting from the cost map calculated in step 2, I added a penalty of 200 to all cells that were within 15 km of the optimal path calculated at step 3.

5. I then iterated over steps 3 and 4, using the updated cost map to calculate a n-best path, and adding a new penalty to cells within 15 km of the result.

Steps 3-5 were repeated 4 time, thus obtaining the four fastest paths between Moscow and Perekop. As shown by Figure 3.12 there is excellent agreement between the first three paths and the three main historical invasion routes used by nomads.

[3] The flow accumulation value is the size of the basin drained through a particular cell, and is proportional to the amount of water that would flow through a particular cell, if all cells received the same amount of rainfall, and there were now evaporation or subsurface water flow.

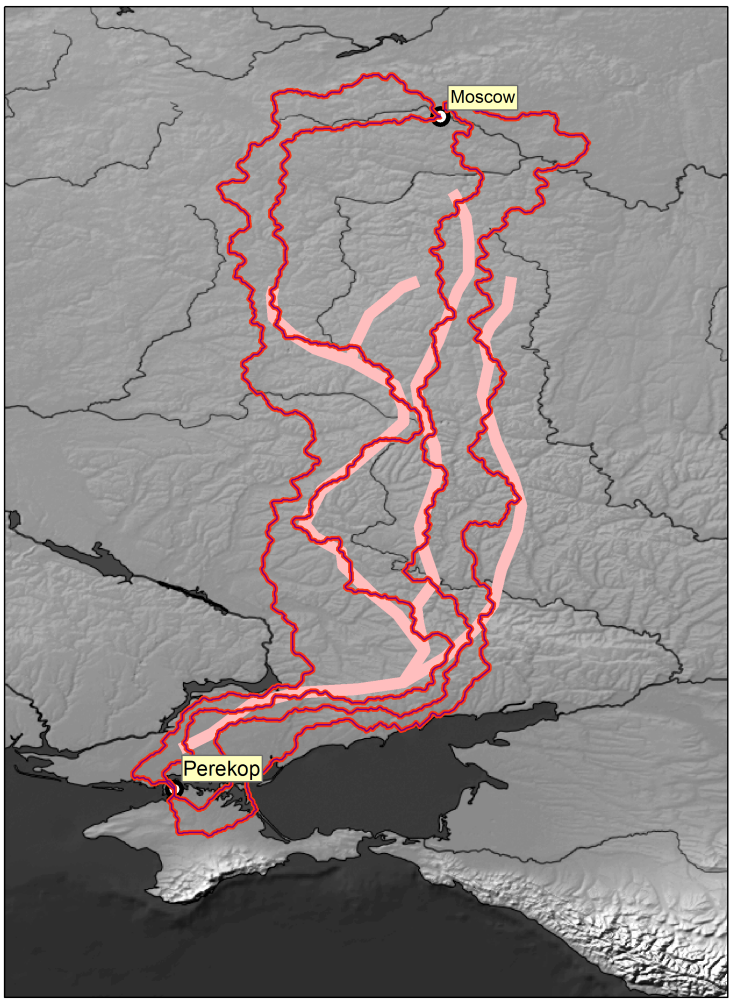

Figure 3.12: Pink thick lines: historical invasion routes. Red lines: calculated invasion routes.

Defense Lines

The next step was to calculate the optimal defense lines. The objective of any linear barrier is to block all possible avenues of approach to a given objective. In the case of a country like Russia with a limited campaign season and very large distances, this can be refined to ”blocking all possible avenues of approach

that take less time than is available to an attacker". To find the cheapest possible barriers I used the following algorithm.

1. Find all possible paths between Moscow and Perekop that take less than some specified maximum (the invasion envelope).

2. Calculate the construction cost as

$$\max(500 - \text{Movement cost}, 0)$$

3. Find the cheapest path (according to the construction cost map) between the western boundary of the invasion envelope, and its western boundaty.

4. Add a penalty of 200 to the construction cost of all cells that are within 50km of the optimal fortification line.

5. Iterate over steps 2-3, so as to generate the N best fortification lines to defend against nomadic raids.

The accuracy of this algorithm is less than in the case of the paths (see Figure 3.13, but it still does a reasonably good job of identifying convenient ways of using rivers to block the advance of nomadic raids.

3.6.4 Results

Summary statistics

The unit of analysis is the district (raion), equivalent approximately to a US county.

Table 3.1 summarizes the dataset of raions in the Perekop-Moscow invasion area.

	(1)				
	count	mean	sd	min	max
LnLights	344	1.189099	1.331214	-2.168142	5.018408
Dist. OptFort	344	.7236854	.7661776	0	3.892349
Dist. OptPath	344	.3607458	.3879302	0	1.970094
Grass Yield	336	497.1918	93.24023	326.6667	712.6667
Barley Yield	336	2475.748	153.1957	2092.645	2798.56
Dist. Moscow	344	147.2518	50.50512	28.00458	241.2706
Latitude	344	50.33925	2.835091	44.4891	54.9868
River Size	342	97.54745	126.2327	4	974.8318
River Size^2	342	25403.6	110180	16	950297

Table 3.1: Summary statistics

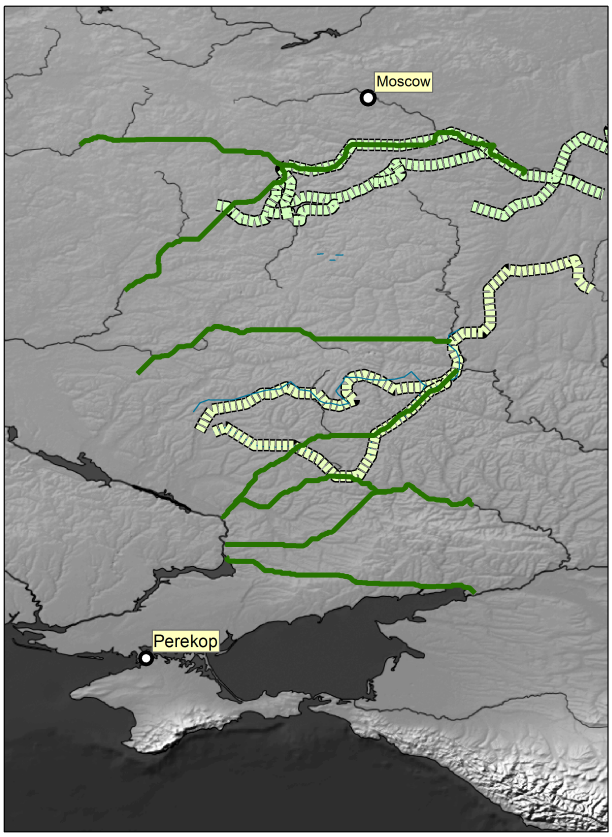

Figure 3.13: Thick dashed lines: historical defense lines. Thin solid lines: calculated optimal defense lines.

Table 3.2 shows the relevant table of naked correlations.

	(1)							
	LnLights	Dist. OptFort	Dist. OptPath	Grass Yield	Barley Yield	Dist. Moscow	Latitude	River Size
LnLights	1							
Dist. OptFort	-0.0842	1						
Dist. OptPath	0.0674	0.178***	1					
Grass Yield	-0.205***	-0.0988*	0.161***	1				
Barley Yield	0.0895	-0.143***	-0.0833	0.550***	1			
Dist. Moscow	0.212***	0.379***	-0.0305	-0.815***	-0.485***	1		
Latitude	-0.192***	-0.454***	0.0193	0.763***	0.472***	-0.972***	1	
River Size	-0.00584	-0.0583	-0.0294	-0.0600	-0.141**	0.0244	0.0284	1

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Summary statistics for the adoption cross-section dataset.

Main Results

The main specification is

$$\text{LnLights}_i = \beta_0 + \beta_1 \text{DistOptFort}_i + \beta_2 \text{DistOptPath}_i + \gamma \text{Controls} + \epsilon_i$$

where LnLights_i is the level of night lights measured in rayon i, DistOptFort is the distance from the closest optimally calculated fortification line, DIstOptPath is the distance from the optimally calculated invasion path, and Controls denotes of vector of control variables. The basic results are highlighted in figure 3.14, which shows how being closer to a calculated fortification line is associated with having more night lights, with a clearer effect once controls are included. Being closer to an optimal invasion route has virtually no effect without controls, and has actually a negative effect once controls are included.

The regression results are shown in Table 3.3. The distance from the optimal fortification lines is a weak predictor of night lights when used without controls, but increases in both magnitude and significance as more controls are added. When all controls are added, a one standard deviation increase in the distance from a fortification line results in a decrease in the number of night lights of approximately 0.2 standard deviations.

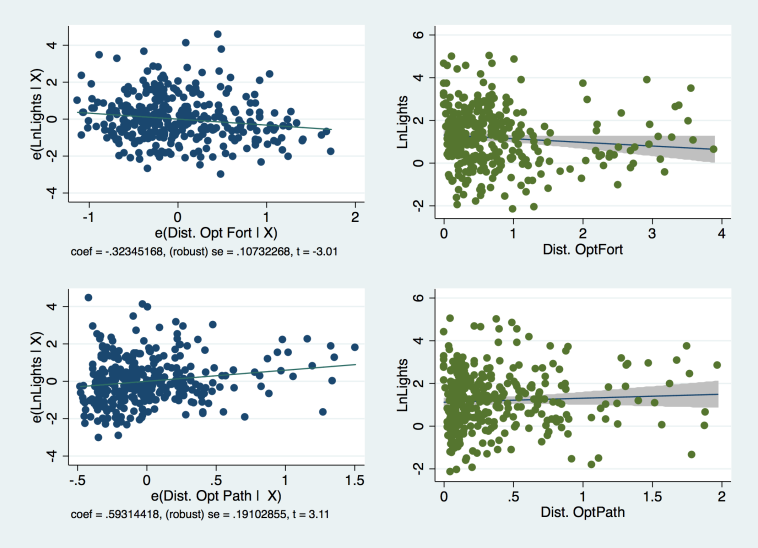

Figure 3.14: The effect on log night lights of being close to an optimal fortification line (top row) and an optimal invasion route (bottom row). The left column shows partial plots with linear fit, controlling for the yield for pasture grasses, barley, distance from Moscow, latitude, size of local rivers and the square of the size of local rivers. The right column shows unconditional scatterplots with linear fit.

	(1) LnLights	(2) LnLights	(3) LnLights	(4) LnLights	(5) LnLights	(6) LnLights	(7) LnLights	(8) LnLights	(9) LnLights
Dist. OptFort	-0.098*	-0.119**	-0.096*	-0.144**	-0.130**	-0.234***	-0.213***	-0.179***	
	(0.069)	(0.031)	(0.077)	(0.022)	(0.026)	(0.000)	(0.000)	(0.003)	
Dist. OptPath		0.084	0.090	0.128**	0.177***	0.170***	0.173***	0.167***	0.146***
		(0.173)	(0.152)	(0.021)	(0.001)	(0.002)	(0.002)	(0.002)	(0.007)
Barley Yield			0.084		0.316***	0.322***	0.329***	0.333***	0.339***
			(0.128)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grass Yield				-0.238***	-0.417***	-0.193**	-0.199**	-0.190**	-0.269***
				(0.000)	(0.000)	(0.037)	(0.031)	(0.039)	(0.003)
Dist. Moscow						0.122	0.123	0.212	0.326
						(0.651)	(0.661)	(0.450)	(0.242)
Latitude						-0.181	-0.176	-0.118	0.121
						(0.495)	(0.526)	(0.672)	(0.643)
River Size							0.023	0.303**	0.388***
							(0.731)	(0.033)	(0.006)
River Size^2								-0.303**	-0.400***
								(0.039)	(0.006)
Observations	344	344	336	336	336	336	334	334	334
R^2	0.010	0.016	0.021	0.069	0.135	0.158	0.160	0.171	0.152

Standardized beta coefficients; p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Regression results within the Perekop to Moscow area of operations. All columns show beta coefficients.

3.6.5 Placebo test: the Smolensk road

The above discussion leaves the possibility that a third unobserved factor may be correlated with both my measure of fortification optimality and the prevalence of night lights today, and that this statistical relationship might be spuriously driving my result. For example, the fortification optimality method I employ will tend to pass through places where the headwaters of two rivers systems are in close proximity, since this minimizes the gap requiring heavy engineering work. However, such areas will also be favored by river traders wanting to portage goods between the two river systems i.e. wanting to establish a portage. It is therefore conceivable that the results shown above are due simply to modern cities being located on the portages between the basins of the Don and the Dniepr.

To avoid this threat to identification, I construct a placebo test by repeating the analysis just performed, but for attacks on Moscow coming from Vilnius rather than Perekop. This was the axis of attack favored by virtually all Western invaders of Russia, including the Poles, Swedes, French and Germans. The key choke point was the gap between the Dvina, which flows North into the Baltic, and the Dnepr, which flows south into the black sea. The resulting land bridge was guarded by the fortress cities of Vitebsk on the Dvina, and Orsha and Smolensk on the Dnepr.

While the potential for trade was even greater than between the Dnepr and Don (the route provided the most economical transportation between the Baltic Sea and the Ottoman Empire), I find that the effect of fortification suitability on night lights in this area is almost exactly zero, as shown both by the scatterplots in Figure 3.15, and the regressions of Table 3.4. This effectively eliminates the trade explanation, while being easy to explain in terms of defense: attacks from the west were invariably undertaken by armies of settled, farming nations, which relied on cumbersome logistical tails, and were thus restricted to using the few available roads. Such attacks were therefore best parried by building fortresses along these roads. It was only against the Tatar raids that a cordon defense was indispensable, and it is therefore only to the south of Moscow that fortification suitability has an effect on present day city lights.

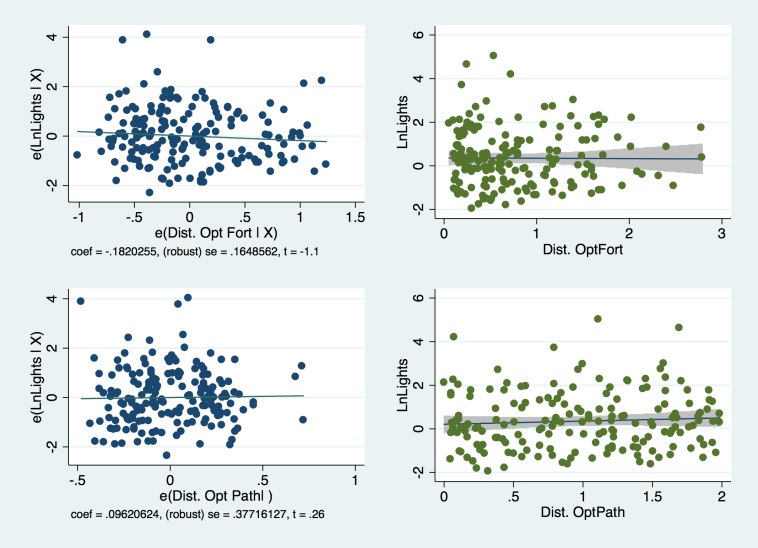

Figure 3.15: The effect on log night lights of being close to an optimal fortification line (top row) and an optimal invasion route (bottom row), within the Vilnius to Moscow operations theater. The left column shows partial plots with linear fit, controlling for the yield for pasture grasses, barley, distance from Moscow, latitude, size of local rivers and the square of the size of local rivers. The right column shows unconditional scatterplots with linear fit.

	(1) LnLights	(2) LnLights	(3) LnLights	(4) LnLights	(5) LnLights	(6) LnLights	(7) LnLights	(8) LnLights	(9) LnLights
Dist. OptFort	-0.008	-0.032	-0.039	-0.067	-0.060	-0.130*	-0.090	-0.084	
	(0.909)	(0.666)	(0.599)	(0.347)	(0.409)	(0.098)	(0.237)	(0.271)	
Dist. OptPath		0.077	0.138	0.018	-0.071	0.316**	0.093	0.044	-0.014
		(0.325)	(0.135)	(0.815)	(0.466)	(0.043)	(0.578)	(0.799)	(0.929)
Barley Yield			-0.120		0.150	0.201*	0.156	0.168	0.182*
			(0.166)		(0.151)	(0.059)	(0.143)	(0.112)	(0.084)
Grass Yield				-0.359***	-0.434***	-0.214	-0.456***	-0.480***	-0.477***
				(0.000)	(0.000)	(0.108)	(0.003)	(0.002)	(0.002)
Dist. Moscow						0.164	0.224**	0.215**	0.170*
						(0.124)	(0.029)	(0.031)	(0.062)
Latitude						0.571***	0.339**	0.303*	0.261
						(0.000)	(0.048)	(0.083)	(0.112)
River Size							0.229***	-0.077	-0.076
							(0.002)	(0.771)	(0.772)
River Size^2								0.327	0.343
								(0.225)	(0.197)
Observations	179	179	178	178	178	178	178	178	178
R^2	0.000	0.005	0.016	0.128	0.140	0.197	0.231	0.239	0.234

Standardized beta coefficients; p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Regression results within the Vilnius to Moscow area of operations. All columns show beta coefficients.

3.7 Conclusions

This paper has improved on the existing literature in three main ways.

First, it proposed a new explanation for the introduction of serfdom in Russia, by arguing that the restriction to the freedom of movement was fundamentally caused by an imbalance between the distribution of population that would result from the free movement of labor, and that whcih was necessary to ensure that the state could be defended. While I make no claim that this was the only reason for enacting serfdom in the world, or even within Russia, the theory opens a significant new axis of interpretation on one of the most historically important social constructs.

Secondly, I adapt this approach to explain why serfdom in most of Western Europe was abandoned. I argue that while the supporting economy of medieval Europe was of course very different from that of the tatars, the overall strategy for warfare was in fact similar, focused largely on wide-ranging raids deep inside enemy controlled territory. When gunpowder made it necessary for armies to be resupplied regularly, warfare became more positional, and strengthening cities on the main lines of communication became the most efficient way of defending a state. It therefore became unnecessary to maintain a widely dispersed feudal cavalry force to block all possible approaches.

Thirdly, the paper uses geo spatial methods to calculate high resolution measures of both the optimal attack route between two points, and more importantly the optimal defense lines to block those attacks. Currently, the standard tools used to measure exposure to military threats are ruggedness (Nunn and Puga, 2012), and geographic distance from particular points. The approach used in this paper can provide high frequency variation in contexts where the existing methods would lack empirical power (flat areas, or those with only one threat origin, correlated with other variables.

Bibliography

Acemoglu, D., Aghion, P., Bursztyn, L., and Hemous, D. (2012). The Environment and Directed Technical Change. *American Economic Review*.

Acemoglu, D., Johnson, S., and Robinson, J. (2002). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5):1369–1401.

Acemoglu, D. and Robinson, J. (2012). *Why Nations fail: the Origins of Power, Prosperity, and Poverty*. Crown, New York, 1st ed edition.

Acemoglu, D. and Wolitzky, A. (2011). The Economics of Labor Coercion. *Econometrica*, 79(2):555–600.

Acemoglu, D. and Zilibotti, F. (1997). Was Prometheus unbound by chance? Risk, diversification, and growth. *Journal of Political Economy*, 105(4):709–751.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Allmand, C. T. (1988). *The hundred years war : England and France at war, c. 1300-c. 1450*. Cambridge University Press.

Almond, D. and Mazumder, B. (2008). Health Capital and the Prenatal Environment: The Effect of Maternal Fasting During Pregnancy.

Ammerman, A. J. and Cavalli-Sforza, L. L. (1984). *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press.

Ashraf, Q. and Michalopoulos, S. (2013). Climatic Fluctuations and the Diffusion of Agriculture. *Review of Economics and Statistics*.

Belfer-Cohen, A. and Bar-Yosef, O. (2002). Early sedentism in the near east. In *Life in Neolithic Farming Communities*, pages 19–38. Springer US.

Berger, A. (1992). Orbital variations and insolation database. Technical report, IGBP PAGES/World Data Center for Paleoclimatology.

Bloom, D., Sachs, J., Collier, P., and Udry, C. (1998). Geography, demography, and economic growth in Africa. *Brookings Papers on Economic Activity*.

Bowles, S. and Choi, J. (2013). Coevolution of farming and private property during the early Holocene. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22):8830–8835.

Braidwood, R. J. (1960). The Agricultural Revolution.

Buggle, J. C. and Nafziger, S. (2015). Long-Run Consequences of Labor Coercion: Evidence from Russian Serfdom *.

Childe, V. G. (1935). *New light on the most ancient East*. Kegan Paul, Trench, Trubner.

Clausewitz, C. V. and Howerd P., M. . P. (2007). *On War*. Princeton University Press, Princeton, NJ.

Cohen, M. N. and Armelagos, G. J. (1984). *Paleopathology at the Origins of Agriculture*. University Press of Florida, Orlando.

Dari-Mattiacci, G. (2013). Slavery and Information. *The Journal of Economic History*, 73(01):79–116.

Darwin, C. (1868). *The variation of animals and plants under domestication*, volume 2. John Murray, London.

Davies, B. L. (2004). *State power and community in early modern Russia : the case of Kozlov, 1635-1649*. Palgrave Macmillan.

Dell, M., Jones, B., and Olken, B. (2013). What do we learn from the weather? The new climate-economy literature.

Dennis, G. T. (1985). *Three Byzantine military treatises*. Dumbarton Oaks, Research Library and Collection.

Dennison, T. K. T. K. (2011). *The institutional framework of Russian serfdom*.

DeVries, K. (2003). *Joan of Arc : a military leader*. Sutton.

Diamond, J. (1987). The Worst Mistake in the History of the Human Race. *Discover*, 8(5):64–66.

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Co., 1 edition.

Domar, E. D. (1970). The Causes of Slavery or Serfdom: A Hypothesis.

Dow, G., Reed, C., and Olewiler, N. (2009). Climate reversals and the transition to agriculture. *Journal of Economic Growth*.

Duleep, H. O. (2012). The Labor/Land Ratio and India's Caste System.

Easterly, W. and Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics*.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., and Others (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2).

Fenske, J. (2014). Ecology, trade and states in pre-colonial africa. *Journal of the European Economic Association*, 12(3):612–640.

Filjushkin, A. (2008). *Ivan the Terrible : a military history*. Frontline Books, Filjushkin2008.

Galor, O. and Michalopoulos, S. (2012). Evolution and the growth process: Natural selection of entrepreneurial traits. *Journal of Economic Theory*, 147(2):759–780.

Galor, O. and Moav, O. (2002). Natural Selection and the Origin of Economic Growth. *The Quarterly Journal of Economics*, 117(4):1133–1191.

Galor, O. and Moav, O. (2007). The Neolithic Revolution and Contemporary Variations in Life Expectancy. *Working Papers*.

Galor, O. and Weil, D. (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90(4):806–828.

Gremillion, K. J., Barton, L., and Piperno, D. R. (2014). Particularism and the retreat from theory in the archaeology of agricultural origins. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17):6171–7.

Harlan, J. (1998). *The living fields: our agricultural heritage*. Cambridge University Press, Cambridge.

Harlan, J. R. (1992). Origins and Processes of Domestication. In Chapman, G. P., editor, *Grass evolution and domestication*, volume 159, page 175. Cambridge University Press, Cambridge.

Harris, H. (1933). Bone Growth in Health and Disease: The Biological Principles Underlying the Clinical, Radiological, and Histological Diagnosis of Perversions of. *Journal of the American Medical Association*, 101(27):2143.

He, F. (2011). *Simulating Transient Climate Evolution of the Last Deglatiation with CCSM3*. PhD thesis, University of Wisconsin - Madison.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *International Journal of Climatology*, 25(15):1965–1978.

Humphrey, L. T., De Groote, I., Morales, J., Barton, N., Collcutt, S., Bronk Ramsey, C., and Bouzouggar, A. (2014). Earliest evidence for caries and exploitation of starchy plant foods in Pleistocene hunter-gatherers from Morocco. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3):954–9.

Ignatius David (2015). How ISIS Spread in the Middle East - And How to Stop IT. *The Atlantic*.

International Fertilizer Development Center (1998). *Fertilizer manual*. Kluwer Academic in cooperation with.

Kaufmann, C. D. and Pape, R. A. (1999). Explaining Costly International Moral Action: Britain's Sixty-year Campaign Against the Atlantic Slave Trade. *International Organization*, 53(04):631–668.

Keys, A., Brožek, J., Henschel, A., Mickelsen, O., and Taylor, H. (1950). *The biology of human starvation.(2 vols)*. Univ. of Minnesota Press.

Khodarkovsky, M. (2002). *Russia's steppe frontier : the making of a colonial empire, 1500-1800*. Indiana University Press.

King, G. and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9.

Klyuchevsky, V. (1911). *A history of Russia,*. J.M. Dent & Sons Ltd.;E.P. Dutton & Co., London; New York.

Kuijt, I. (2011). Home is where we keep our food: The origins of agriculture and late Pre-Pottery Neolithic food storage. *Paleorient*.

Lehner, B., Verdin, K., Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. :. *Eos, Transactions*, 89(10):93–94.

Lobdell, J. (1984). Harris Lines: Markers of Nutrition and Disease at Prehistoric Utqiagvik Village. *Arctic Anthropology*.

Locay, L. (1989). From hunting and gathering to agriculture. *Economic Development and Cultural Change*.

Margolin, S. L. (1948). *Vooruzhenie streletskogo voiska (Weapons of the Strel'tsy forces)*. Trudy Gosudarstvennogo Istoricheskago Muzei, (Works of the State Historical Museum),.

Markevich, A. and Zhuravskaya, E. (2017). The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire. *Working Papers*.

Mayshar, J., Moav, O., and Neeman, Z. (2013). Geography, transparency and institutions.

McCloskey, D. N. (1991). The Prudent Peasant: New Findings on Open Fields. *The Journal of Economic History*, 51(2):pp. 343–355.

Michalopoulos, S. (2012). The Origins of Ethnolinguistic Diversity. *American Economic Review*, 102(4):1509–39.

Montesquieu, B. D. (1748). The Spirit of the Laws. *Trans. by T. Nugent. New York (Colonial Press)*.

Nossov, K. K. (2006). *Russian fortresses 1480-1682*. Osprey.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*.

Olsson, O. (2001). The rise of Neolithic agriculture. *rapport nr.: Working Papers in Economics*.

Olsson, O. and Paik, C. (2013). A Western Reversal since the Neolithic? The long-run impact of early agriculture.

Peters, M. (2016). War Financing and the Re-imposition of Serfdom after the Black Death.

Pinhasi, R., Fort, J., and Ammerman, A. J. (2005). Tracing the origin and spread of agriculture in Europe. *PLoS Biology*, 3(12):e410.

Postan, M. and Hatcher, J. (1978). AGRARIAN CLASS STRUCTURE AND ECONOMIC DEVELOPMENT IN PRE-INDUSTRIAL EUROPE: POPULATION AND CLASS RELATIONS IN FEUDAL SOCIETY. *Past and Present*, 78(1):24–37.

Purugganan, M. D. and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature*, 457(7231):843–848.

Putterman, L. and Trainor, C. (2006). Agricultural Transition Year Country Data Set.

Richerson, P. J., Boyd, R., and Bettinger, R. L. (2001). Was Agriculture Impossible During the Pleistocene but Mandatory During the Holocene? A Climate Change Hypothesis. *American Antiquity*, 66(3):387–411.

Roberts, M. (1956). *The military revolution, 1560-1660 an inaugural lecture delivered before the Queen's University of Belfast.* M. Boyd, [Belfast].

Rowthorn, R. and Seabright, P. (2010). Property Rights, Warfare and the Neolithic Transition.

Saunders, J. J. (2001). *The history of the Mongol conquests*. University of Pennsylvania Press.

Smith, B. D. (2014). Failure of optimal foraging theory to appeal to researchers working on the origins of agriculture worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28):E2829.

Smith, V. (1975). The primitive hunter culture, Pleistocene extinction, and the rise of agriculture. *The Journal of Political Economy*.

Tanaka, T. (2010). Risk and time preferences: linking experimental and household survey data from Vietnam. *American Economic Review*, 100(1):557–571.

Testart, A. (1982). The Significance of Food Storage Among Hunter-Gatherers: Residence Patterns, Population Densities, and Social Inequalities [and Comments and Reply]. *Current anthropology*, 23(5):523–537.

Van Creveld, M. (1991). *The transformation of war*.

Voigtländer, N. and Voth, H.-J. (2013a). How the West "Invented" Fertility Restriction. *American Economic Review*, 103(6):2227–2264.

Voigtländer, N. and Voth, H.-J. (2013b). The three horsemen of riches: Plague, war, and urbanization in early modern Europe. *The Review of Economic Studies*, 80(2):774–811.

Williams, J. P. (1981). Catch-up growth. *Journal of embryology and experimental morphology*, 65 Suppl:89–101.

Wright, H. E. (1970). Environmental Changes and the Origin of Agriculture in the near East. *BioScience*, 20(4):210–212.

Wu, L., Dutta, R., Levine, D., and Papageorge, N. (2017). Entertaining Malthus: Bread, Circuses and Economic Growth. *Economic Inquiry*.