Three Essays in Empirical Historical Economics

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Andrew James Thomas
Master of Science
Johns Hopkins University, 2015
Bachelor of Science
University of Minnesota, 2008

Director: Noel Johnson, Professor
Department of Economics

Summer Semester 2022
George Mason University
Fairfax, VA

Dedication

For Jordan.

Acknowledgments

I would like to thank my dissertation committee for their comments and encouragement. Dr. Mark Koyama and Dr. Jonathan Schulz served as wise committee members, and Dr. Noel Johnson, my Chair, provided valuable input and opportunities throughout the process.

I would also like to thank my parents for their support.

Table of Contents

List of Tables

List of Figures

Abstract

THREE ESSAYS IN EMPIRICAL HISTORICAL ECONOMICS

Andrew James Thomas, PhD

George Mason University, 2022

Dissertation Director: Noel Johnson

In recent years text analysis and geographic information systems (GIS) have become vital as sources of data for economic research (Gentzkow et al. 2019,Donaldson and Storeygard 2016). In this dissertation I use these tools in conjunction with economic theory to investigate several historical questions. In the first chapter, text matching tools are used in the construction of a database of early European printed materials to explain why some cities adopted the printing press earlier than others during the 15th century. In the second chapter, GIS tools are used to construct a panel of environmental controls to assess the impact of a land redistribution on agricultural productivity in Ethiopia. In the final chapter, both sets of tools are used to explore a database of English soldier names from the 14th and 15th centuries and assess the importance of social status and regional recruitment in the armies of the Hundred Years' War.

Chapter 1: The Impact of the Black Death on the Adoption of the Printing Press

Co-authored with Noel Johnson and Alexander Taylor

Abstract

The diffusion of the printing press across Europe in the late 15th century has been linked to increased growth and the later spread of the Reformation. But what factors explain which cities adopted earlier than others? The literature on technical change supports the idea that market size is an important factor in the diffusion of innovation. A challenge, however, is that market size is endogenous to many unobserved variables. Using scraped data from the Universal Short Title Catalogue we create a new database linking early European printed material to historical city population estimates. We leverage plausibly exogenous variation in mortality from the Black Death (1347–52) across European cities to estimate the causal impact of market size on early print adoption. We find that cities whose populations were more heavily impacted by the Black Death were less likely to be early adopters of the press. We also investigate patterns of adoption on both the extensive and intensive margins as well as specialization in subject areas up to 1600.

1.1 Introduction

The movable type printing press developed by Johannes Gutenberg around 1440 was one of the most important technologies of the last millennium. By dramatically reducing the cost of book production information could be reproduced and disseminated faster than ever before (Barbier, 2017; Eisenstein, 1980). Recent research has established a causal link between a city being an early adopter of the printing press and subsequent growth (Dittmar, 2011). The press has also been shown to have facilitated the spread of the Protestant Reformation (Rubin, 2014). An important question, then, is why did some cities adopt the printing press earlier than others?

In this paper we exploit a natural experiment, the massive population shock caused by the Black Death to the urban system of Europe between 1347 and 1352, to identify a causal relationship between market size and early press adoption. Historians of the printing press note that large cities were early adopters yet, to our knowledge, there is no study establishing a causal link between city size and early press adoption. In this paper we provide evidence that, at the city-level, mortality from the Black Death was plausibly random.[1] We then exploit this fact to estimate a three stage model. In the first stage, exogenous plague mortality is used to predict city growth between 1300 and 1400. In the second stage, predicted growth between 1300 and 1400 is used to estimate the level of a city's population in 1400. In the last stage, predicted city population in 1400 is used to predict press adoption between 1450 and 1500. Our identification strategy relies on our argument that city-level Black Death mortality rates were random and that, therefore, the level of a city's population in 1400 predicted using these rates will also be unrelated to any potential confounders that may also be correlated with press adoption.

In addition to exploiting plague mortality rates to predict city population, our identification strategy also includes controls for university presence or a bishopric—two of the main sources of both demand and supply of books according to the historical literature. We also

[1] The material in this paper on the exogeneity of Black Death mortality at the city-level is based on work in Jedwab et al. (2022*b*), Jedwab et al. (2022*a*), and Jedwab et al. (2019).

control for a city's trade potential in 1300, fifty years before the Black Death shock, using market access measures constructed from digitized maps of historical trade routes and city populations (Donaldson, 2018; Donaldson and Hornbeck, 2016). In addition, we will control for the potential agricultural productivity of land surrounding a city using data from GAEZ (Fischer et al., 2021). This is important since agricultural fecundity is well known to be a strong predictor of city growth during the pre-industrial period (Henderson et al., 2018; Jedwab et al., 2020). Finally, we will control for distance to Mainz since this is a popular instrument for press adoption (Dittmar, 2011; Rubin, 2014).[2]

It is well established in the literature on technical change that two important determinants of technological diffusion are the relative prices of inputs and market size (Acemoglu, 2002; Acemoglu and Linn, 2004). Theoretically, the Black Death could have affected press adoption through either of these channels. In the aftermath of the plague wages increased dramatically in Europe and didn't return to pre-plague levels in many places until 1600 (Jedwab et al., 2022*a*). It's possible that this increase in the relative price of labor incentivized the adoption of labor saving technologies, among them the press (Allen, 2011; Pamuk, 2007). Alternatively, the Black Death may have destroyed market potential in places where it was most severe (Campbell et al., 2015; Jedwab et al., 2022*b*).

We find strong evidence for the latter hypothesis—cities that lost more population because of the plague were also less likely to be early adopters of print technology. In our baseline estimates, a city with 10% lower population due to Black Death mortality in 1400 is predicted to be 3.5% less likely to have a press by 1500. When we extend the analysis to take into account the impact of the Black Death on the entire urban network in which a city is embedded, this point estimate increases in magnitude. We find a 10% decline in market access due to the plague led to a 4.5% decrease in probability of press adoption by 1500.

We also investigate the intensive margin of printing using data on the number of editions

[2]We do not make any claims in this paper as to validity of the distance to Mainz instrument. We simply control for it in many regressions to alleviate potential concerns that distance to Mainz may be correlated with Black Death mortality.

published in each city in each year between 1450 and 1600 scraped from the Universal Short Title Catalogue. When measured by number of unique editions printed, our model suggests that a city with 10% lower population due to the plague published 16% fewer editions between 1450 and 1500 and 35% fewer editions between 1450 and 1600. Finally, using data on the subject matter of unique editions printed, we find suggestive evidence that there was less specialization (less variety of subjects) published in cities impacted more by the Black Death.

1.2 Data

1.2.1 Black Death Mortality

The Black Death first appeared in Europe in the port of Messina, Sicily, in October of 1347. Over the next three years it was responsible for the death of approximately 40% of the population of Europe (see Appendix Figure A.1). It was a bacterial infection, caused by the bacterium *Yersinia Pestis.* As currently understood by scientists, the infection cycle was that fleas would become infected by the disease, which would block their esophagus making it impossible for them to feed. These unsatiated fleas would then continue to bite humans and other animals in an attempt to feed, in the process regurgitating the bacterium into humans. Within a week the bacteria would reach the lymph nodes and cause them to swell into painful buboes. Typically about two-thirds of those infected would die within ten days of first contact with the disease.

One of the only robust predictors of the virulence of the Black Death was how early in the pandemic a city was infected. This follows from the epidemiological fact that epidemic diseases tend to be more virulent when they first enter into a previously untouched population. Then, as people get infected and either die or recover with immunity, resistance to the disease emerges. As such, cities that were more connected to the point of origin of the Black Death in Europe, Messina, were more likely to experience high mortality rates. However, this did not necessarily mean geographic proximity to Messina drove mortality as trading ships, with rats carrying infected fleas, did not only go to Mediterranean destinations. For example, as can be seen in figure A.1, London was hit relatively early, around the same time as many towns on the Iberian pennisula.

The Black Death need not have entered Europe through Messina. Appendix figure A.2 shows alternate trading routes through which the plague could have entered Europe. We know the trade center of Astrakhan was hit in 1345 and the Genoese colony of Kaffa was infected in 1346. It was from Kaffa that Genoese galleys traveled to Messina, the starting point of the disease in Europe. However, the disease could just as easily gone through the popular trade-route from Astrakhan to Moscow, then to Novgorod, and have entered through the Baltic rather than the Mediterranean. Or, it could have easily entered via Prague in central Europe.

We use data on city-level Black Death mortality compiled by Jedwab et al. (2019). They collect the data from Christakos et al. (2005, 117–122) who compile the mortality estimates from a wide range of sources including ecclesiastical and parish records, testaments, tax records, court rolls, chroniclers' reports, donations to the church, financial transactions, mortality of famous people, letters, edicts, guild records, hospital records, cemeteries and tombstones. Jedwab et al. (2019) check these mortality data against other sources including Ziegler (1969), Russell (1972), Gottfried (1983), and Benedictow (2005).

The final Black Death mortality data set comprises 274 towns and cities spread across 16 modern European countries. Of these, 177 are percentage estimates. Forty nine are literary descriptions which Jedwab et al. (2019) rank according to their implied severity.[3] Nineteen cities give mortality of the clergy, which, when cross-referenced against the subset of these 19 for which non-clergy mortality is known, suggests that clergy died at about an 8% higher rate than the general population. We, therefore, adjust the clergy mortality estimates downward accordingly in the main data set. Twenty nine cities have desertion rates. Jedwab et al. (2019) follow Christakos et al. (2005) in lowering these rates for use in

[3] 5% for "spared"/"escaped", 10% for "partially spared"/"minimal", 20% for "low", 25% for "moderate", 50% for "high", 66% for "highly depopulated", and 80% for "decimated".

the main data as it is known that desertation rates were 1.2 times mortality on average.

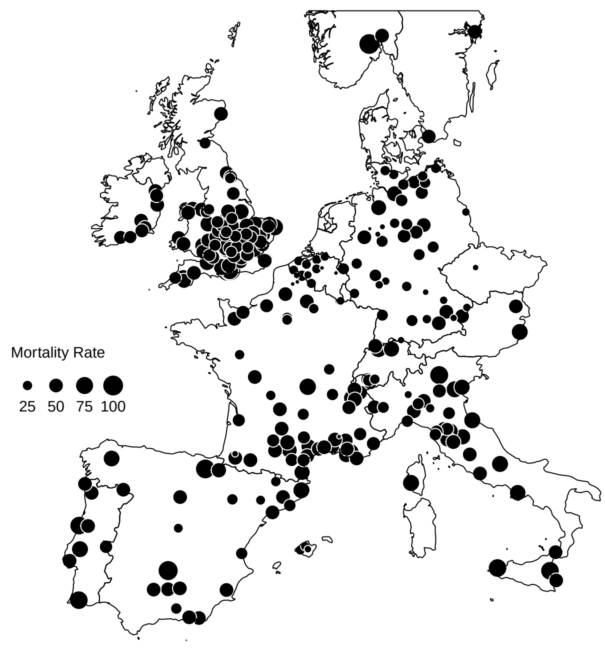

Figure 1.1: Black Death Mortality

Figure shows the 274 cities with known Black Death mortality rates between 1347 and 1352. Source Christakos et al. (2005).

What seems to have mattered for the city-level virulence of the Black Death is how early a city was infected, which had a significant random component. Figure 1.1 shows the distributions of mortality rates across Europe. English towns appear to have suffered significantly (though some are relatively unscathed) as did Northern Italy (though Milan was relatively untouched). It is difficult to identify any major geographic patterns in the mortality data.

The Black Death appears to be a comparatively pure population shock: infrastructure was left intact, only people were killed, it did not explicitly target a sub-group of the population (e.g. intellectuals or a particular ethnic group), and there was no government or international organization sponsored aid in the aftermath. As such, the city-level mortality rates constitute a potentially powerful natural experiment to test the impact of a demographic shock on technology adoption—namely the printing press.

1.2.2 City Populations

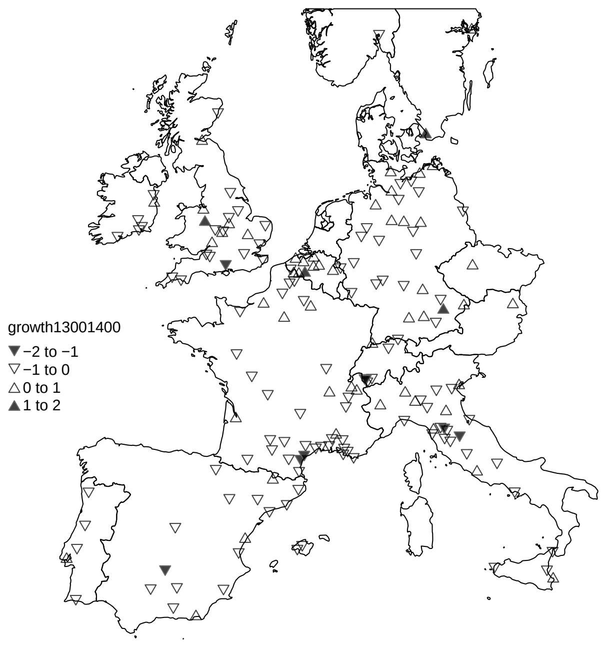

Figure 1.2: City Growth 1300–1400

Figure shows the 169 cities resulting from intersecting the mortality data with population data. Triangles indicate whether city grew or shrunk between 1300 and 1400.

Since we are interested in the demographic impact of the Black Death, we intersect the 274 mortality cities with the Bairoch city population data set (Bairoch, 1988).[4] The Bairoch data contains estimates of the population of 1,801 towns and cities in 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1750, 1800 and 1850. In order to be in the data set in a given year city population has to be 1,000 people or greater. In 1300 there are 457 cities that satisfy

[4]We supplement the Bairoch data with those from Chandler and Fox (2013). This allows us to have estimates for cities on the "edges" of our main data, which will be important to ensure that the market access variables we create aren't biased down on the populated edges (e.g. Eastern Europe as opposed to the Atlantic edge).

this criteria.

When we intersect the 274 mortality cities with the 457 Bairoch cities, there are 169 cities which form our baseline data set that we will attempt to match with data on year of printing press acquisition. Figure 1.2 shows these 169 cities. The city icons in the map indicate whether the city population shrank or grew between 1300 and 1400. As can be seen, the spatial distribution in city populations affected by the Black Death appears random.

We also collect data on a host of control variables proxying for *locational fundamentals*, *increasing returns*, and *institutions*. The manner in which these are created is described in Appendix Section A.2.

With our 169 city data set, we can investigate the broad patterns of the impact of the Black Death on city populations in both the short and long-runs. We estimate a series of regressions of the form:

$$\Delta \text{Pop}_{i,t} = \alpha + \beta \text{Mort}_{i,1347-52} + \varepsilon_{i,t}, \tag{1.1}$$

where $\Delta \text{Pop}_{i,t}$ is the percentage change in city i's population over some time period.[5] $\text{Mort}_{i,1347-52}$ is Black Death mortality for city i or mean mortality for the modern country in which city i is located.

In table 1.1 we report the β's from estimating specification 1.1 on various samples. In column 1 the dependent variable is city growth between 1300 and 1400. Assuming no city growth in the absence of the Black Death shock and zero recovery in the fifty years immediately following the shock, we would expect the coefficient to, mechanically, be equal to one. That is, a 10% mortality shock would be associated with a city shrinking in population by 10%. In fact, the coefficient is -0.0085 which suggests that a city experiencing 10% mortality from the Black Death would be only 8.5% smaller by 1400. In other words, demographic

[5]Since during our period of study there are both large positive and negative growth rates in city populations, we calculate percentage growth using the midpoint method.

Table 1.1: The Impact of the Black Death Shock

	(1) g13001400	(2) g13001400	(3) g13001600	(4) g12001300	(5) g11001200
mortality	-0.0085^{***}		-0.0024	0.0025	0.0010
	(0.0027)		(0.0034)	(0.0051)	(0.0026)
mean_mortality		-0.015^{***}			
		(0.0041)			
N	169	169	168	107	62

Robust standard errors in parentheses.
Mortality is measured from 0-100. Aggregate mortality is average modern-country mortality.
Standardized beta coefficient for (1) is -0.26 and for (3) is -0.05.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

recovery, on average, was very slow in the immediate decades following the Black Death.[6] This is an important fact to establish as our study is premised on the assumption that the demographic impact of the plague persisted long enough to disrupt printing press adoption more than 100 years after the shock.

In column 2 we take a different approach to measuring the short-run impact of the plague. We investigate whether there were spillover effects from the mortality shock. That is, did a city recover more slowly if neighboring cities were also hit more severely. As such we calculate the average mortality rate for the modern country in which city i is located and estimate specification 1.1 using this as the measure of mortality for city i. When we do this, the coefficient on mortality is approximately twice as big as when we measure mortality using just the own-city impact. The coefficient implies that 10% mortality for a region implies that a city in that region will shrink by 14.5% by 1400. This is even more evidence that the short-run impact of the plague persisted and recovery could possibly have been affected by how neighboring cities were impacted. We will exploit these potential spillovers in part of our analysis of print adoption below.

In specification 3 we estimate the impact of own-city mortality on own-city growth between 1300 and 1600. The coefficient of -0.0024 is both statistically and economically insignificant.

[6]This claim would be even stronger if there was some positive city growth between 1300 and the onset of the Black Death.

The standardized beta for the estimate in column 3 is -0.05 (compared to -0.26 for column 1). Our interpretation of this is, consistent with the account of historians, cities had, on average, recovered from the Black Death by 1600. Though, as described in Jedwab et al. (2022*a*) and Jedwab et al. (2022*b*) there was a significant amount of heterogeneity in recovery. Some cities recovered quite quickly and flourished relative to their pre-pandemic position whereas other cities faded away and never returned to their position in the European urban network.

1.2.3 The Black Death as a Random Shock

We adopt several strategies to establish that city-level Black Death mortality was largely random.

First, in table 1.1 in columns 4 and 5 we report placebo regressions in which we investigate whether Black Death mortality can explain city growth in the two centuries *previous* to it's arrival in Europe. Both estimated coefficients are small and statistically insignificant.

Second, we collect data on many variables which could potentially have been related to severity of the Black Death in different locations. We use these to test whether any city-level observables can explain Black Death mortality rates. The variables we construct fall into three categories (see Appendix Section A.2 for detailed descriptions):

1. **Local Physical Geography**:
    - Average temperature between 1500 and 1600
    - Elevation
    - Cereal suitability
    - Potato suitability
    - Pastoral suitability
    - Within 10km of sea

- Within 10km of river
- Longitude
- Latitude

2. **Local Economic Geography**
    - City population in 1300
    - Market access in 1300
    - Within 10km of major Roman road
    - Within 10km of any Roman road
    - Within 10km of major Roman road intersection
    - Within 10km of any Roman road intersection
    - Within 10km of medieval road (from Shepherd Atlas)
    - Within 10km of medieval intersection (from Shepherd Atlas)
    - Medieval fair location
    - Member of Hanseatic League
    - Roman aqueduct
    - University

3. **Local Institutions:**
    - Capital city
    - Representative institution in 1300
    - Years parliament met in 14th century
    - Distance to parliament
    - Distance to battles between 1300 and 1350

In appendix table A.3 we regress the variables in each of these categories on mortality. We also regress them on mortality as a group. These coefficients with confidence intervals are given in figure 1.3. In all regressions the only two variables that ever show up as (weakly) statistically significant are proximity to a river and latitude. The latter is consistent with the Black Death being more virulent in locations closer to Messina. The former is consistent with chance: we have twenty-six RHS variables, so one or two should be significant by chance. Furthermore in all the balance regressions the R-squared is never greater than 10%, suggesting that observables do a very poor job explaining Black Death mortality.

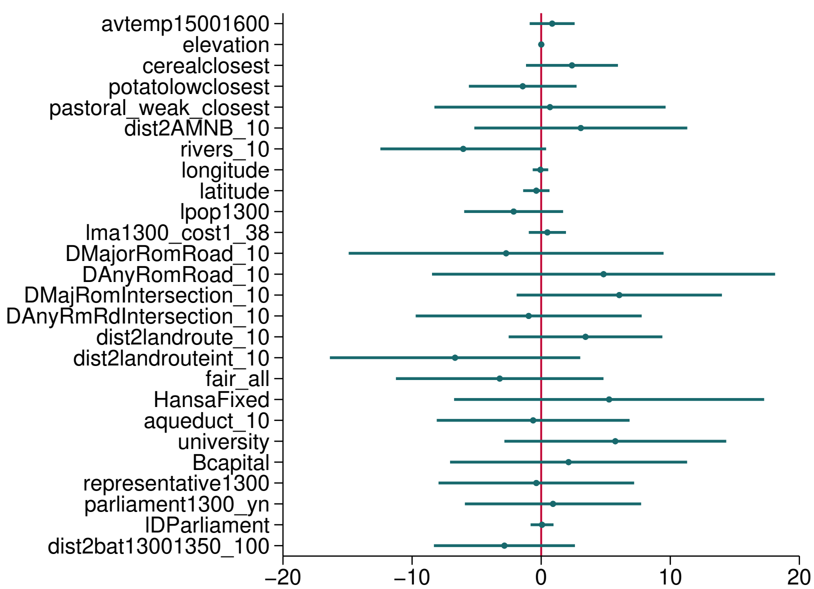

Figure 1.3: Balance Regression Coefficients

Our third identification strategy is to construct measures of market access for each city in 1300 based on historical maps of transportation routes and costs of using different travel technologies. Under some standard assumptions in modern trade models market access captures all of the direct and indirect costs of trade for a location (e.g., Donaldson and Hornbeck, 2016). The theoretical framework from which market access is derived accounts for the general equilibrium relationship between producers and consumers across the entire potential trade network for a city.

Several sources suggest that the Black Death traveled along trade routes (see, e.g., Boerner and Severgnini (2014)). If it arrived earlier in places that were better connected to trade networks, then this could result in higher mortality rates for places better suited for trade. Market access is also likely correlated with press adoption. By controlling for market access in 1300, fifty years before the Black Death, we are able to better focus on the variation in city/market size generated by Black Death mortality.

For city j market access is defined as:

$$MA_j = \sum_i \mathrm{N}_i \tau_{ji}^{-\sigma} \ , \tag{1.2}$$

Where N_i is the population of city i, τ_{ji} is the least cost cost of travel between city j and city i, and σ is a trade elasticity.

The most computationally challenging task in constructing the market access variable is to create a measure of the cost of travel τ between cities. To do this we begin by creating maps of Roman roads, medieval trade routes, major rivers, and seas.[7] Estimates from Boerner and Severgnini (2014) allow us to assign the cost of transporting goods by each of these routes (portage is assumed to be used when there is no better alternative).[8]

The trade elasticity, σ, measures the responsiveness of trade to changes in transportation costs between towns. Higher values reflect a more elastic response of trade to increases in transport costs. We follow Donaldson (2018) in using a value 3.8 for σ.[9]

[7] The data on river locations are from Nussli (2011). The data on Roman roads are from the *Digital Atlas of Roman and Medieval Civilizations*. It is available from: `https://darmc.harvard.edu`. We complement the Roman road data with data on medieval trade routes from digitized maps in Shepherd (1923). These data are especially helpful given that the Roman road coverage did not extend into the northeastern part of Europe.

[8] The normalized estimates for speed of travel we use from Boerner and Severgnini (2014) are: porters = 1; roads = 0.50; rivers = 0.50; seas = 0.13.

[9] The appropriate σ depends on context. For modern and developed economies, researchers tend to

We divide Europe into 5km x 5km grids and assign the lowest travel cost to each grid. We apply Djikstra's algorithm to determine the lowest cost of travel between all city pairs ($\approx$ 3,200,000) (van Etten, 2012). Appendix figure A.3 shows an example of the least cost travel path between Paris and Rome. Using the travel cost measures, we then create an index measuring market access for each city.[10]

Finally, we note that Jedwab et al. (2022*a*) instrument Black Death mortality using the month of arrival of the disease in a city (conditional on year) and find the impact of the disease on city growth in the short and long-runs is similar to running OLS regressions without controls.

1.2.4 The Printing Press

The movable type printing press was invented by Johannes Gutenberg in Germany around 1440, with the first print shop operating in Mainz by 1450. Though Gutenberg and his associates maintained a strong monopoly on movable type printing for several years, the press rapidly spread throughout Europe with presses present in 284 cities in Western Europe by 1500 (*Incunabula Short Title Catalogue*, 1998). Setting up and operating a press was expensive, requiring constant injections of funds and lucrative side contracts to support larger projects. The great Gutenberg Bible took two years to complete, and required both a financier and side contracts printing small but popular schoolbooks, pamphlets, and broadsheets to support the business in the meantime. Even by the end of the 15th century, the estimated cost of running a large printing house was potentially as high as 200 ducats a month (Pettegree, 2010). The last quarter of the 15th century saw book production on a near industrial scale, with the largest shops operating as many as 24 presses.

estimate higher values. For example, Eaton and Kortum (2002) use 8.28 for OECD trade flows in 1995. Donaldson and Hornbeck (2016) estimate an average $\sigma = 8.22$ for trade flows in the U.S. in the second half of the 19th century. By contrast, Donaldson (2018) estimates $\sigma = 3.8$ for colonial India. Kopsidis and Wolf (2012) assume $\sigma = 1$ for their study of Prussian trade during the Industrial Revolution. This is also the value assumed by many earlier studies of 'market potential' or 'market access' (Harris, 1954).

[10]Theory says market access should be logged (see, e.g., Donaldson and Hornbeck (2016)), which is why we use log market access in our regressions.

Trading cities such as Augsburg, Nuremberg, and Basel were early major centers of book production, as well as university cities like Cologne. Most early printers operated solely in their local market or at trade fairs, with only a handful of exceptions.[11] While some inter-city trade did exist, it was limited by high transportation costs over land and the relative fragility of early books (Barbier, 2017). Some sources place the allowance for transportation costs as high as 20% above the book price, giving local print shops a substantial advantage against sellers transporting books from other cities (Febvre and Martin, 1976). Even when inter-city trade did occur, it did not significantly affect the establishment of a press in a city. The bookstore established by Johann Fust in Paris dominated the city's book market for a mere 4 years before a press was established in the city, and its success was likely more due to opposition to the press by manuscript writers than genuine competitive superiority. Overall, the rapidly expanding market for books demanded a much higher quantity than inter-city booksellers could provide.

Beyond monetary and transaction costs, human capital also acted as a barrier to entry into the printing industry. Boerner et al. (2021) draw a connection between the upper-tail human capital necessary to assemble, maintain, and repair clocks and future adoption of the printing press. Emphasizing the role of technological agglomeration and the connectedness of seemingly separate technological innovations, they find that clock towns were 34-40 percentage points more likely to adopt the printing press by 1500. This is another channel through which the Black Death may have affected adoption of the press. If higher Black Death mortality hindered a city population's ability to accumulate human capital, the city would not have the sufficient skills and upper-tail human capital to reap the gains of technological agglomeration and adopt either clocks or the press, or both.

Most books spread to cities not by transportation but through reprinting by local print shops. This trade dynamic is undergirded by solid economic logic and, we argue, played a role in press adoption. While some books inevitably traveled beyond their print city,

[11] For example, Gutenberg's former partner and financier, Johann Fust, established a bookstore in Paris in 1466, hoping to find a new market for books printed back in Mainz. Fust died of the plague shortly after (see Pettegree, 2010, chap. 2).

high transportation costs prevented the systematic trade of books as described above. If a particular book showed sales potential (i.e. if printers detected demand for the book), local printers held substantial production advantages. If a city did not have a press and the market was sufficiently large, merchants with the relevant knowledge could reap large gains from establishing one and cornering the local market.

Literacy rates also likely influenced early book production. Data on literacy is uncertain at best, but some scholars have provided country-level estimates. Allen (2003) estimates that Belgium and the Netherlands had the highest literacy rates in 1500, only reaching 10%. Buringh and Van Zanden (2009) estimate slightly higher but similar rates for the second half of the 15th century, with the Netherlands and Italy having the highest rates at 17% and 15%, respectively, and Western Europe as a whole having a literacy rate of 12%.

Previous scholarship has demonstrated a positive impact of the printing press to a city's economic growth. Dittmar (2011) argues that cities that received a press in first 50 years received a significant growth advantage. He suggests that this was due localized impact of press, as inter-city book trade was small, and presents limited evidence that the circulation of educational printed materials (e.g. commercial arithmetics, merchant manuals) gave these cities a human capital advantage. For example, double-entry bookkeeping was first described and spread through printed materials in the late 15th century.

1.2.5 Data from the Universal Short Title Catalogue

We assemble a database linking the Black Death mortality and city population data to information on the spread of the printing press. We scraped all the data from the Universal Short Title Catalogue (USTC), a project collecting all known information on printed material from the first 200 years of the printing press in Europe.[12] Each book edition has a unique page on the site with specific information about the edition. To scrape these

[12] https://www.ustc.ac.uk

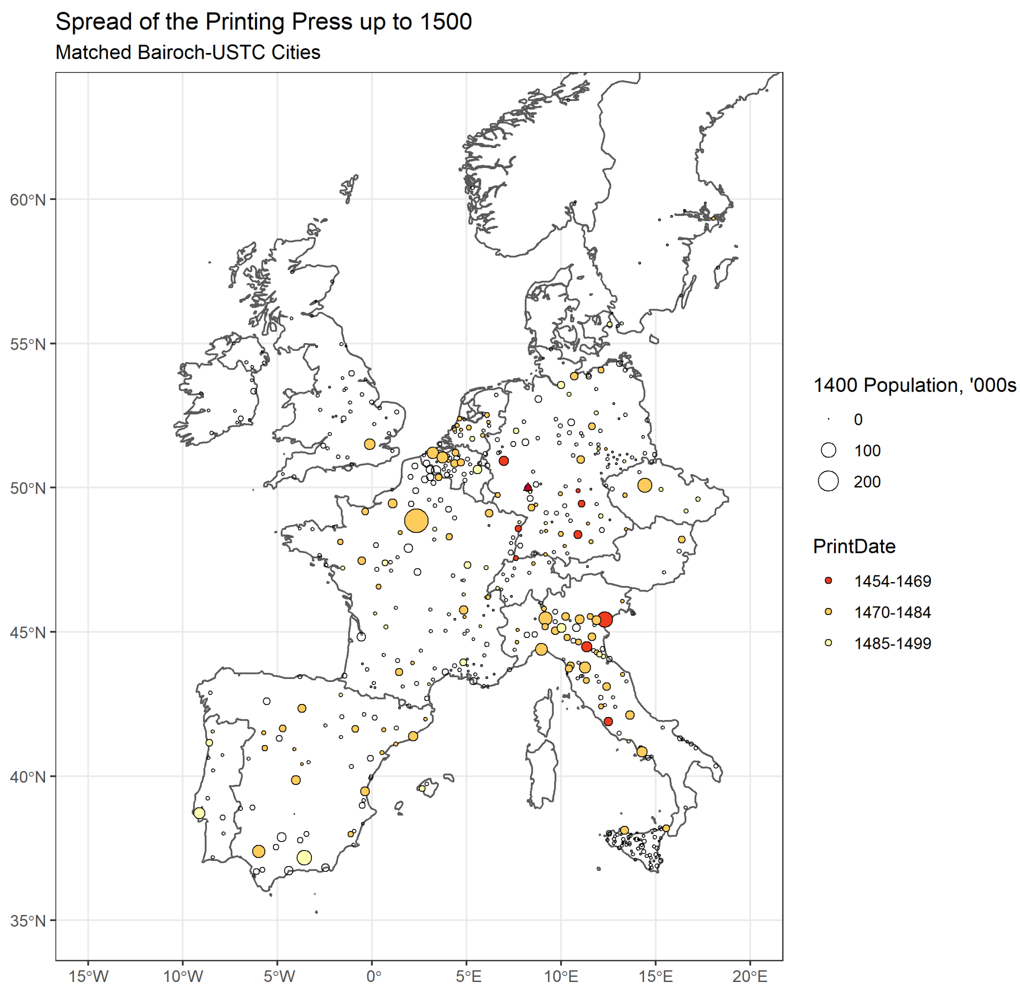

Figure 1.4: Spread of the printing press up to 1500

data, first the html code for each edition page was saved. Next, basic information about each edition was extracted from the saved html code, including print year, print location, imprint, subject, and author. This results in an initial dataset of 826,084 observations.

Some entries are incomplete, so we remove any observations lacking a place or year. This leave us with 712,982 observations, of which 343,989 occur up to 1600. Each record comes with an 'imprint', a transcription from the title page of a book which also typically contains a year, but a small number of records have a discrepancy between this imprint year and the year encoded in the metadata. We eliminate any records with a discrepancy greater than 10 years, giving us a final working sample of 343,660 records from 1450 to 1600. From this, 26,012 observations are in the early print adoption period of 1450 to 1500.

The chief challenge in matching the USTC data to the Bairoch data is the lack of regularized city names across the two datasets. USTC records contain a 'place' mentioned in the object imprint, but no further details. We utilized a combination of direct and fuzzy matching, along with hand coding, to link just under half of USTC places with our base data.

First, USTC data were joined directly by city name. There are a few instances of the same city name occurring in different countries in Bairoch; for example, Brest is both a city in France and in modern day Belarus. We verify the match from USTC by checking the language editions are printed in (in this case Russian), and remove the match to Brest, France.

However, many cities have alternate names or spellings, particularly when anglicizing languages with diacritical marks. We institute another round of matching of remaining places by allowing for some string distance between names. We utilize a text matching algorithm via the *fuzzyjoin* package in R, using the Optimal String Alignment distance metric, a variant of the Levenstein distance, that counts the number of insertions, deletions, and substitutions necessary to convert one string to another while allowing for transposition of adjacent characters. This returns some false matches, which we verify and remove.

Finally, we hand match any remaining places that used different names across data sources.

Chief examples are Prague (Praha), or Aix-en-Provence (Aix).

Of 1,312 unique placenames present in the USTC data, we find matches to 631 cities in our base data. We find 193 cities from 1450–1500, and 475 cities from 1450–1600.

One concern is that printers may be operating in places too small to be included in Bairoch. Looking at the remaining unmatched cities with entries before 1600, we find that most unmatched cities have only a single edition present in USTC, or otherwise comprise a small portion of the total records matched.

We match 98% of total records in the 1450–1500 period, and 96% of total records in the 1450–1600 period.

Table 1.2: USTC Matches

Period	1450-1500	1450-1600
Matched USTC Cities	193	475
Missing USTC Cities	68	340
% Cities Matched	0.74	0.58
Matched USTC Editions	25,494	329,658
Missing USTC Editions	518	14,002
% Editions Matched	0.98	0.96

Figure 1.4 represents the joined Bairoch-USTC data of 564 cities with an estimated population in 1400 along with their year of print adoption if prior to 1500. The origin of the printing press, Mainz, is represented by the red triangle. There is a clear pattern of larger cities gaining presses in the initial 50 year period, and smaller cities not adopting. During the first 15 years of the press its adoption appears regionally limited to southern Germany and a handful of Italian cities (Venice, Bologna, and Rome), but by just 1485 there is evidence of a press in all corners of Western Europe. From this sample, the 169 cities that also contain Black Death mortality are used below.

1.3 Estimation Strategy

We estimate a three stage model using David Roodman's Conditional Mixed Processes framework (Roodman, 2011).

$$\text{growth13001400}_i = \alpha + \beta \text{mortality}_i + \gamma X_i + \varepsilon_i \tag{1.3}$$

$$\text{lpop1400}_i = \delta + \eta \widehat{\text{growth13001400}} + \theta X_i + \zeta_i \tag{1.4}$$

$$\text{print}_i = \kappa + \mu \widehat{\text{lpop1400}} + \lambda X_i + \xi_i \tag{1.5}$$

Where i refers to city i and X is a vector of controls. We always control for university or bishopric presence in a town. In other specifications we also control for market access in 1300, potential cereal suitability, and least cost distance to Mainz.

We also construct a measure of print specialization using a Herfindahl-Hirschman Index. Editions in the USTC are classified as falling into one of thirty-eight categories (see Appendix figure A.6). Using the sample of records with both year and subject present, we calculate the share of editions in each subject per city per decade. The HHI is calculated by summing the squares of each of these shares and ranges from zero to one. An index value of one means there is only one subject being printed in the local market.

$$HHI = s_1^2 + s_2^2 + s_3^2 + \dots s_n^2 \tag{1.6}$$

where s_i is the share of each subject of all categorized editions per city for the time period.

1.4 Empirical Results

Table 1.3 presents the main results of our three-stage model. We run several specifications to ensure we control for variables that could influence our results. We control for bishopric or university presence in all regressions. Academics and members of the clergy were the most literate segments of the population. Including this variable controls for differences in literacy rates, which influence book demand and the likelihood a city adopts the printing press. In column (2) we add our control for market access in 1300. Market access could have facilitated the spread of the Black Death or the adoption of the press through higher levels of trade, travel, and other interactions between cities, making its inclusion important in ensuring Black Death mortality's role as an instrument. In column (3) we include the control for cereal suitability, which could play a role in estimating city size. Finally, in column (4) we add a control for distance to Mainz, a common instrument for the spread of the printing press used by both Dittmar (2011) and Rubin (2014).

In Stage 1, column (1) the fitted model predicts that a 10 percentage point increase in a city's Black Death mortality reduced population growth between 1300 and 1400 by 8.4 percentage points. A mortality shock closer to the mean 40% would on average reduce growth between 1300 and 1400 by 34%. In column (4) with all controls included the point estimate is reduced to -0.0054, suggesting a 40% mortality rate reduced city growth by 23%.

In stage 2 we use predicted growth based on Black Death mortality from stage 1 to predict log city size in 1400. The coefficient reported in column (1) is 1.77. Given that the standard deviation of growth is 0.49 and a standard deviation in log 1400 population is 0.95, this means that the model predicts that a city for which Black Death mortality lowered its growth by one standard deviation would have been about 90% of a standard deviation smaller in 1400. This effect increases in size to 1.4 standard deviations smaller if we use the estimates from column (4) which includes all controls.

In stage 3 we use predicted population in 1400 to predict printing press adoption by 1500.

Table 1.3: The Impact of Black Death Mortality on Press Adoption by 1500

	(1)	(2)	(3)	(4)
		*Stage*1		
dep. var.: growth13001400				
mortality	-0.0084^{***}	-0.0084^{***}	-0.0067^{**}	-0.00533^{*}
	(0.0028)	(0.0029)	(0.0031)	(0.0029)
dist_Mainz				-0.0100
				(0.0073)
		*Stage*2		
dep. var.: lpop1400				
predicted growth13001400	1.774^{***}	1.821^{***}	2.068^{**}	2.719^{*}
	(0.587)	(0.615)	(0.877)	(1.445)
dist_Mainz				0.0318
				(0.0289)
		*Stage*3		
dep. var.: USTC_print_1499				
lpop1400	0.354^{**}	0.336^{***}	0.363^{***}	0.342^{**}
	(0.139)	(0.126)	(0.138)	(0.150)
dist_Mainz				-0.0022
				(0.0066)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is a dummy variable equal to 1 if a city has a first appearance in the USTC by 1500. The estimate in column (1) suggests that having a 10% smaller population due to Black Death mortality in 1400 decreased the probability of a city being an early adopter of the press by about to 3.5%. This coefficient size is stable across specifications, with the estimate in column (4) showing that a 10% smaller population due to Black Death mortality decreased the probability of early press adoption by 3.4%. The results suggest that the Black Death predominantly affected press adoption through its impact on market size and potential. While the relative price of labor likely increased and incentivized press adoption, any effect is outweighed by the negative impact of market destruction. The high fixed costs of establishing a press were prohibitive to market entry in cities who lost many potential customers to the plague.

We note that in column (4) the coefficient on distance to Mainz, the city where Gutenberg invented the press and from which it diffused, is in effect zero in all three stages.

We now investigate the robustness of these results across different dependent variables in the third stage. Specifically, we want to know if the impact of the Black Death on city population was persistent enough to deter press adoption beyond 1500. We also investigate different ways of measuring press presence in a city. In particular, we can measure press presence on the intensive margin by looking either at the year of adoption or the number of editions printed. We also will investigate whether cities that shrank due to the Black Death published books on fewer topics than larger cities. Finally, we will look at whether the impact of the Black Death shock on the broader market access of a city had a bigger impact on press adoption than when we focus on just its impact on the population of the city alone.

Unless otherwise noted, the results for the first two stages of the following regressions are the same as in table 1.3 so, to avoid needless repetition, we only report the estimates from the third stage.

Table 1.4: The Impact of Predicted Population in 1400 on Year of Press Adoption

	(1)	(2)	(3)	(4)
		*Stage*3		
dep. var.: first_edition				
predicted lpop1400	-25.31^{*}	-24.03^{*}	-30.83^{**}	-25.45^{*}
	(13.60)	(12.58)	(14.04)	(13.87)
dist_Mainz				0.520
				(0.612)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: The Impact of Predicted Population in 1400 on Press Adoption by 1600

	(1)	(2)	(3)	(4)
		*Stage*3		
dep. var.: USTC_print_sample				
predicted lpop1400	0.344^{***}	0.320^{***}	0.280^{**}	0.289^{**}
	(0.122)	(0.112)	(0.125)	(0.134)
dist_Mainz				0.000913
				(0.00500)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

First, we assess the persistent impact of the Black Death on press adoption beyond the early printing period. Table 1.5 shows the same specifications as table 1.3, but with the timeline for a city adopting the printing press extended through 1600. The coefficient in column (1) closely resembles that in table 1.3, column (4). However, the estimated effect of a 10% increase in city size when including all controls is smaller, at 2.9%. The coefficient on distance to Mainz remains small and statistically insignificant. Even extending the timeline through the 16th century, we continue to find a negative, if slightly smaller, effect of Black Death mortality on press adoption.

In Table 1.4 we estimate the impact of predicted city population due to Black Death mortality on year of print adoption. This explores the impact of the Black Death on the intensive margin for those cities that eventually adopted the press in our time period. Accordingly, for these regressions we use the sample of press adopters up to 1600 (n=118) since there are only 73 adopters by 1500.[13] We find that increasing a city's size by one standard deviation (0.95) would lead it to adopt the press between 25 to 31 years earlier. This suggests that higher Black Death mortality did not only prevent some cities from adopting the press, but also delayed adoption is cities that eventually had a press by 1600. While never adopting the printing press in our time period is a more severe effect, this effect demonstrates that eventual adopters did not go unscathed by the destruction of the plague.

We also investigate the impact of population size on the intensity of printing as proxied by number of editions published over the relevant timespan. We argue that editions published is a sufficient proxy for printing intensity, even though we do not know the quantity of books printed of any edition. More editions being printed in a city could indicate either reprinting of existing books or runs of new books that had not yet been published locally. In the first case, new editions of existing books indicate high demand for those texts. Popular texts were frequently printed by multiple printers, resulting in more editions, and living authors would update their writings and have new editions printed. In the second case, a

[13]The sample size in the regression table is indicated as 169 since in stages one and two all 169 cities are used. However, in stage three the sample size is 118—the same as the number cities adopting the press by 1600.

Table 1.6: The Impact of Predicted Population in 1400 on Intensity of Printing, 1450-1500

	(1)	(2)	(3)	(4)
		Stage3		
dep. var.: ustc_total_1500				
predicted lpop1400	1.664***	1.607***	1.707***	1.388**
	(0.637)	(0.593)	(0.645)	(0.633)
dist_Mainz				−0.0323
				(0.0280)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: The Impact of Predicted Population in 1400 on Intensity of Printing, 1450-1600

	(1)	(2)	(3)	(4)
		Stage3		
dep. var.: ustc_total_1600				
predicted lpop1400	3.457***	3.301***	3.221***	2.561***
	(0.862)	(0.749)	(0.820)	(0.856)
dist_Mainz				−0.0668
				(0.0413)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

larger number of editions of different books indicates a diverse readership. While we do not explore the differences between these two cases here, neither poses problems for using the number of editions in a city as a proxy for printing intensity.

Tables 1.6 and 1.7 show that larger cities print more editions. The coefficients in these table are elasticities. In column (1) of table 1.6 the coefficient of 1.64 therefore suggests that a 10% decrease in population of a city due to the Black Death is associated with a 16% decrease in editions published. If we extend the sample out to cities that had adopted the press by 1600, this estimate increases to 35%.

Table 1.8: The Impact of Predicted Population in 1400 on Printing Specialization, 1450-1600

	(1)	(2)	(3)	(4)
		Stage3		
dep. var.: ustc_hhi_1600				
predicted lpop1400	-0.179^{**}	-0.175^{**}	-0.162^{**}	-0.150^{*}
	(0.0749)	(0.0742)	(0.0802)	(0.0869)
dist_Mainz				0.00113
				(0.00321)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

What about the effect on specialization? How does Adam Smith's dictum that 'the extent of specialization is limited by the extent of the market' hold up in the early print world? Using our measure of subject concentration (Herfindahl-Hirschman Index) we find that larger cities have more variety in the kinds of materials printed. In table 1.8 column (1) the estimated coefficient of -0.18 suggests that a city with 10% smaller population due to mortality from the Black Death had a Herfindahl Index 0.018 lower, which is about 10% of a standard deviation in the index.

Table 1.9: The Indirect Impact of Black Death (MA Shock) on Press Adoption

	(1)	(2)	(3)	(4)
		Stage1		
dep. var.: growth13001400				
lshock_ma_1300_1353	0.832***	0.857***	0.593**	0.410
	(0.249)	(0.252)	(0.261)	(0.278)
dist_Mainz				−0.00868
				(0.00912)
		Stage2		
dep. var.: lpop1400				
growth13001400	1.798***	1.962***	2.541**	4.918
	(0.574)	(0.610)	(1.019)	(3.033)
dist_Mainz				0.0666
				(0.0576)
		Stage3		
dep. var.: USTC_print_1499				
predicted lpop1400	0.445***	0.379***	0.434***	0.415***
	(0.138)	(0.114)	(0.129)	(0.114)
dist_Mainz				−0.00133
				(0.00681)
Bishopric or University	*Yes*	*Yes*	*Yes*	*Yes*
Market Access in 1300	*No*	*Yes*	*Yes*	*Yes*
Cereal Suitability	*No*	*No*	*Yes*	*Yes*
Distance to Mainz	*No*	*No*	*No*	*Yes*
N	169	169	169	169

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we investigate the indirect impact of the Black Death using the subsequent market access shock instead of mortality in the first stage. In order to compute the shock to market access of the Black Death we need estimates of plague mortality for all of the cities that are potential trade partners. This requires us to impute the mortality for those cities for which we don't have direct historical estimates. We do this using an inverse distance weighted imputation method where we calculate the optimal weight for distance using a leave one out cross-validation technique.

Table 1.9 shows that the impact of city size on print adoption is larger than when we consider the impact on own city population. The coefficient estimate in column (1) of 0.45 suggests that a decrease in trade network size due to Black Death of 10% led to a 4.5% less probability of press adoption for a city. This is about a 30% larger effect (0.45 vs 0.35) than when considering simply the impact of the Black Death on own city population.

1.5 Conclusion

The Black Death and the invention of the printing press were two of the most important events of the late medieval period. The former devastated the human population but arguably broke the feudal equilibrium, making it a critical point in the long run economic growth of Western Europe. The latter enabled the rise of mass media and the diffusion of knowledge at an unprecedented level, revolutionizing how information and ideas were spread, manipulated, and shared. Here we have explored how both events, while deeply important in isolation, are connected. While Black Death mortality may have feasibly impacted press adoption either positively or negatively, we found that it predominantly prevented or delayed a city's adoption of the printing press.

This has important implications for how we think about the relationship between disease environment and economic development. While previous work has demonstrated that the Black Death led to an increase in relative wages, incentivizing the adoption of less labor-intensive capital, our results show that a sufficiently large market is an absolute necessity for technology adoption to occur. While highly lethal disease will improve the wages of survivors through its impact on the supply of labor, the more intuitive outcome of economic destruction is still both dominant and likely. As shown here, the direct demographic impact of the Black Death was devastating to technological diffusion. Finally, our results speak to the literature on technological innovation in general—market size seems to matter. No matter how labor-saving a technology may be, its use requires sufficient demanders for the final product.

Chapter 2: Redistribution and Agricultural Productivity in Ethiopia

Abstract

Development in many poor countries depends upon increases in agricultural productivity. Institutions supporting property rights have been tied to productivity and growth through two main literatures. In one, low productivity persists due to insecurity of property. In the other, misallocation of agricultural land persists due to high exchange costs. I investigate the relevance of these theories to agricultural productivity in Ethiopia at the end of the 20th century using subnational variation in an unexpected land redistribution as a natural experiment. I link agricultural production data at the village and regional level with information on local soil suitability, yearly variation in rainfall, and market access. Using a difference-in-differences framework, I find no evidence of an impact from redistribution on agricultural yields or patterns of land investment.

2.1 Introduction

Over the past 200 years world agricultural output has increased by an order of magnitude, with post-World War II productivity increases driving the trend in the developed world (Federico, 2009). This productivity growth has been slow to reach many developing countries, and shrinking the agricultural productivity gap is key for their long run development (Gollin et al., 2002). One potential cause of this gap is institutional: the systems governing the allocation and security of productive land in such regions often rely upon non-market mechanisms such as communal and traditional tenure systems. A wide range of theory and research has focused on the importance of these channels for economic outcomes. One influential literature focuses on the security of property as the key to productivity enhancing investment (North and Weingast 1989, Acemoglu et al. 2001, Johnson et al. 2002, Acemoglu and Johnson 2005, Hornbeck 2010), while another focuses on the misallocation of productive inputs in the face of high transaction costs as a major source of low productivity (Restuccia and Rogerson 2008, Hsieh and Klenow 2009, Gollin et al. 2014).

In this chapter I investigate the importance of property rights institutions in the context of Ethiopia, a large developing country which in recent years has seen consistently high economic growth after decades of stagnation. An unexpected and regionally limited land redistribution in 1997 had exogenous elements that allow for a causal investigation into the degree that property rights institutions were a binding constraint to growth. I find no clear evidence that this combination of negative shock to tenure security and reallocation of agricultural land had an effect on cereal crop yields. This may suggest that the security and distribution of land were not the binding constraints to agricultural productivity growth in Ethiopia in the 1990s.

The structure of this paper is as follows. In Section 2, I describe the background literature on property rights and agricultural productivity, the relevant literature in the context of Ethiopia, and the broad contours of the 1997 redistribution process. In Section 3, I describe the relevant data and collection process. In Section 4, I present the empirical framework used to investigate potential channels. Section 5 presents results, and Section 6 concludes.

2.2 Background

2.2.1 Property Rights and Agriculture

Demsetz (1967) identified two key components of property rights that are relevant for economic efficiency: *use* rights and *exchange* rights. Use rights refers to the ability to access a good, and exchange rights refer to the ability to transfer the good to others. Transaction costs can occur for either of these: it is costly to enforce use rights, and mechanisms of exchange like markets also come with opportunity costs in their implementation.

Both of these channels have been investigated extensively in the historical context. North and Weingast (1989) made an influential case for the importance of security of property rights for long run development in analyzing the consequences of the English Glorious Revolution. They argued that a shift in political equilibrium away from the crown and towards Parliament allowed for a credible commitment that property rights would not be infringed, setting the stage for the later Industrial Revolution. Following this thread, Acemoglu et al. (2001) and Acemoglu and Robinson (2012) argue that an important cause of poor development in many areas of the world is due to extractive institutions, such as those set up by European colonizers, that do not respect the security of property.

An inefficient distribution of property rights, held up by transaction costs in exchange, is also a potentially important source of low productivity. A growing literature on misallocation, in the vein of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), finds sizable effects of inefficient resource allocation on productivity. When there are barriers to resource allocation inputs to production can remain under the control of less efficient firms, reducing overall total factor productivity.

Security of tenure has been studied extensively in the development literature. Under property rights institutions where a farmer expects to lose land, the returns to productivity enhancing investment may be quite low. Jacoby et al. (2002) for example, find that tenure insecurity reduced investment in fertilizer in Chinese farmers. Bellemare et al. (2018) presents evidence for the inverse: a positive shock to property rights security in Vietnam led to increased productivity and investment.

Extending the misallocation methodology to farms, a growing body of research finds that low efficiency is a result of poor allocations of land and farmers, whether from traditional tenure systems (Restuccia and Santaeulalia-Llopis, 2017), or from reallocation following land reform (Adamopoulos and Restuccia, 2020).

Which effect is more important in any given situation is theoretically ambiguous. Acemoglu and Johnson (2005) argues that overall, property securing institutions have been more important to economic growth than the contracting institutions that facilitate exchange. However, the particulars of time and place matter: Finley et al. (2021), for example, find that confiscatory land redistributions in the wake of the French Revolution led to increased agricultural productivity. In this case, inefficiencies resulting from transaction costs in exchange outweighed the threat of expropriation.

2.2.2 Agriculture and Land Rights in Ethiopia

The Ethiopian highlands have been a center of cereal agriculture in sub-Saharan Africa for centuries due to their unique climate and geography, as well as historical cultural linkages to the Near East that brought the Semitic ox-plow agricultural complex. European travelers to the highlands in the early modern era noted the remarkable abundance of productive cropland, alongside irrigation and terracing. (Budge 1928, McCann 1995). By the 1990s, however, Ethiopia had become virtually synonymous with famine. While this is partly a result of the high variability of the summer rains which farmers are dependent on, the political economy of land rights has also been cited as a cause.[1] Indeed, arguments over the access and distribution of land rights have been central to Ethiopian political development (Henze 2000, Joireman 2000, Zewde 2001).

Smallholder farming in the highlands is based around rain-fed plots operated by individual households. Unlike many other sub-Saharan societies, communal agricultural activities are extremely limited. Plowing the soil requires the combined effort of a plow operator (typically male) and a team of oxen, the latter being the scarcest input to production. Oxen possession is an important determinant of wealth, and local social institutions (markets, family exchange, even raiding) for the distribution of oxen services are widespread (McCann, 1995).

A number of studies have examined the impact of tenure security on development outcomes in Ethiopia, but most focus on the impact of land registration programs from the mid-2000s aimed at increasing tenure security. Deininger et al. (2008) and Deininger et al. (2011) find positive impacts of land certification in Amhara region, such as increased perceptions of tenure security, greater land investment, and increased the supply of land to the rental market.

One of the few studies to directly investigate the impact of a negative tenure security shock is Benin and Pender (2001). They use data from a community-level recall survey conducted in 98 villages across the Amhara Region in 2000 and find that villages that experienced the 1997 redistribution had a higher share of farmers using inputs such as fertilizer and herbicides. They suggest that this resulted from land being distributed to those more likely to use such inputs.

The challenge in these studies is that they examine outcomes only within Amhara region, often within a small geographic area. In being so limited, they neglect the possibility that farmers not exposed directly to redistribution may still behave differently operating in an

[1] E.g. Acemoglu and Robinson 2012 argues that the same state capacity that allowed Menelik II to repel the Italian invasion in 1896 and avoid colonization also inhibited long-run development through the threat of expropriation.

environment where redistribution is a possibility, versus those farmers living in areas where redistribution is uncommon. This is akin to the distinction between intent-to-treat and actual treatment in RCT studies. This is particularly important when trying to measure the impact of perceived tenure security: a farmer that did not have land taken in a particular redistribution may still fear another redistribution in the future and make choices accordingly. The particulars of the 1997 Amhara land redistribution allow an investigation that accounts for these factors.

2.2.3 Amhara State Redistribution

The primary narrative account of the redistribution I draw on is Ege (1997). In 1997, Ethiopia was 6 years removed from the collapse of the Derg, a Marxist military junta that deposed the monarchy of Haile Selassie I (r.1930–1974) and ruled the country from 1974–1991. In 1991, the Ethiopian People's Revolutionary Democratic Front (EPRDF), a coalition of rebel groups headed by the Tigrayan People's Liberation Front (TPLF), defeated the Derg after years of conflict in the northern highlands and set up a transitional government. With the implementation of a new constitution in 1995, the country restructured its administrative organization along ethnolinguistic lines in a system of 'ethnic federalism'. This was partly in reaction to the perceived Amhara dominance over other ethnic groups present ever since Menelik II (r.1889–1913) brought the modern day boundaries of Ethiopia under centralized control in the 1890s. This reorganization led to some formerly unified historical states, such as Shewa in the center of the country, being divided across the new regions.

Part of this federalist system involved devolving land use policy to the new regional states, although the new national level government still maintained, as had the Derg, that all land ultimately belonged to the state. One of the factors in the Derg's loss of popular support had been the attempt beginning in 1985 to rationalize land use throughout the country through forced villagization (Kebbede 1992; see also Scott 1998, 247–52). While the Derg had enacted a substantial land reform program in 1975 that redistributed agricultural lands away from large elite landowners and to the farmers that worked on them, it had not made any efforts to prescribe land use.

In November 1996, less than two years into its existence, the Amhara regional parliament unexpectedly announced a land redistribution for the coming year. The initial announcements were vague on implementation, and farmer attitudes seem to have expected something similar to the style of redistribution present during the Derg: based on the notion of formal equality, with a certain degree of favoritism based on social relations of the implementers.

Instead, the process that unfolded was justified as an attack on former Derg officials as well as remaining pre-1974 elites (dubbed 'feudal remnants'). The *birokrasi* (bureaucracy) had been the Derg's contact with local population as the head of local Peasants Associations (PA), organizations tasked with implementing land redistributions and other state directives. As compensation, these farmers had been allowed to possess extra lands. The new EPRDF-allied government officials declared as part of the redistribution that *birokrasi* and 'feudal remnants' could now possess no more than 1 hectare of land, whereas other households were promised a minimum of 2 hectares, and could own up to 3 hectares.

Ege argues that punitive nature of the process and uncertainty at each stage created considerable confusion among farmers, weakening state legitimacy, and in some cases shattering expectations of what was possible. For example, another one of the central directives handed down to each PA was that houses built after 1994 were illegal and would be demolished. Traditionally such an investment on land was widely accepted as a legitimate way of strengthening ownership claims, in line with the findings of Besley (1995) on property rights enhancing investment.

The rather extreme demands of the new government represent a negative shock to tenure security that was not present in the other regions of Ethiopia, largely in the service of the political aim of impairing the legitimacy of remaining elements of the former government. At the same time, it is possible that lands were indeed inefficiently allocated previously: Acemoglu et al. (2014), for example, argue that civil society capture by local elites leads to poor development outcomes. Alternatively, the redistribution could have lead to an even less economically efficient distribution of lands.

2.3 Data

2.3.1 Ethiopian Rural Household Survey

The Ethiopian Rural Household Survey (ERHS) (Hoddinott and Yohannes, 2011) is a longitudinal household level dataset that has completed seven rounds: 1989, two 1994 rounds, 1995, 1997, 1999, 2004, and 2009. This rich data source reports crop area, production, and yield, as well as other household demographic characteristics and survey data, for 1477 households in 15 villages across 4 regions. These villages were chosen to span the diversity of farming systems in Ethiopia, with the majority located in the highlands. I use data from the 1994, 1995, 1997, 1999, and 2004 rounds. The initial 1989 round of the ERHS only covered 6 villages and was sufficiently different in design from the later years that direct comparisons are difficult, and a land certification program in the Amhara region aimed at increasing tenure security began after 2004.

One potential challenge is that the ERHS villages are limited in number and were not randomly selected. As seen in figure 2.1, most are located in a rough transect that parallels the East African Rift, likely for relative ease of access to researchers. Within each village, however, households were selected at random years before the redistribution took place. This should ease concerns about selection bias for households influencing results.

I limit my analysis to highland villages without perennial crops to keep the treatment and control groups as similar as possible. This eliminates 5 southern villages where agriculture was largely based on permanent crops like coffee, chat, and ensete (false banana), which are very uncommon in Amhara region and may have differing local property rights institutions.

Figure 2.2 shows the results from survey questions regarding changes in land holdings from 1999 and 2009. In 1999 about 30% of households in the Amhara villages had seen their

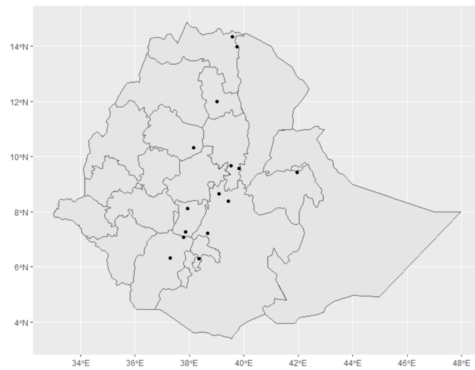

Figure 2.1: Locations of ERHS study villages

land shrink during the previous five years, the vast majority of which was due to state mandated land redistribution. This cause for land size change is almost entirely absent from households in the other regions. Ten years later, only a single Amhara household had seen their lands shrink for this reason.[2] The same pattern is present for increases in land size. This confirms that the government directed land redistribution in 1997 was almost entirely limited to the Amhara region, and that no further redistributions were taken later.

Figure 2.3 shows that there is variation in direct exposure to the redistribution even within the Amhara region. Two of the six villages in this region saw very little redistribution, with most occurring in the four villages on the outskirts of Debre Berhan as well as Yetmen. This can be explained by the fact that Dinki and Shumsheha had been under the control of EPRDF prior to the collapse of the Derg, and standard policy was for liberated villages to have their land redistributed (Ege, 1997). This further supports the notion that the primary motive of the redistribution was for shoring up EPRDF legitimacy. While this does bias the results, it should bias them towards zero as untreated individuals are being included in the treatment at the regional level.

The panel data also allow me to look at some specifics of how the redistribution was implemented. Table 2.1 presents the 1997 characteristics of the Debre Berhan villages. I group

[2]It is possible even this is a data error, as several households are coded in 2009 as having had their lands shrink/grow and stay the same size simultaneously.

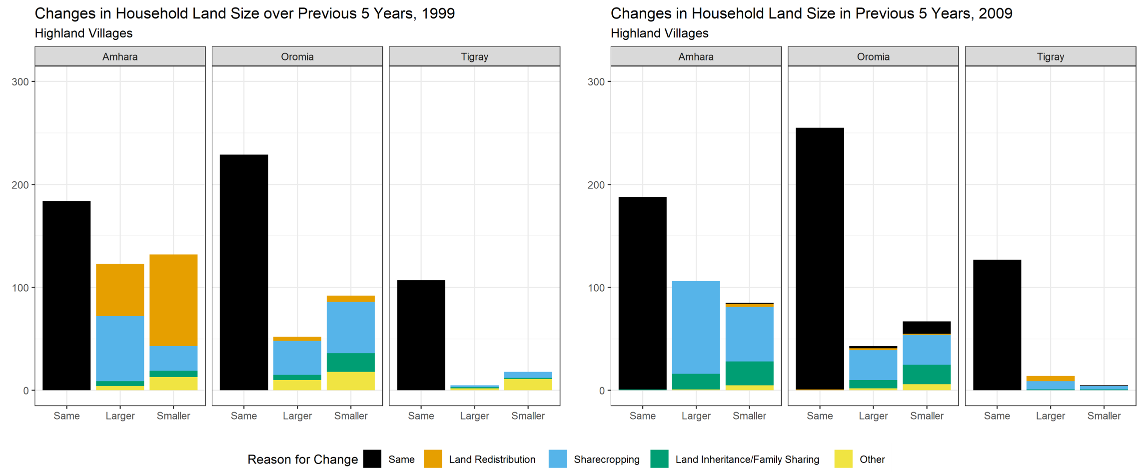

Figure 2.2: Land changes by region

households by whether they reported recently gaining or losing land due to government redistribution in the 1999 survey (all other possibilities are classified as 'Other'). These data were taken just before the implementation of the redistribution, and selection into the ERHS in 1994 should have no relationship to being involved the later redistribution.

Table 2.1: Debre Berhan Summary Statistics, 1997

	Given			Other			Taken			
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	Test
Land Size	246	3.28	1.99	294	3.27	1.81	414	3.27	1.81	F= 0.006
Household Size	246	5.66	2.04	306	5.20	1.99	426	5.92	2.32	F= 9.986***
Real Consumption per Capita	246	139.03	103.95	306	227.97	457.29	426	124.57	67.34	F= 14.666***
Number of Oxen	248	1.77	0.62	307	1.66	0.97	426	1.93	1.37	F= 5.528***

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

While the redistribution was justified as a means of taking undeserved lands from former *birokrasi* and feudal remnants, there appears to be no significant difference in land size just before the redistribution among those who lost land, gained land, or were unaffected. Household size is slightly larger in households involved either way in the redistribution, and those who had land taken had more oxen on average. This may support the narrative that

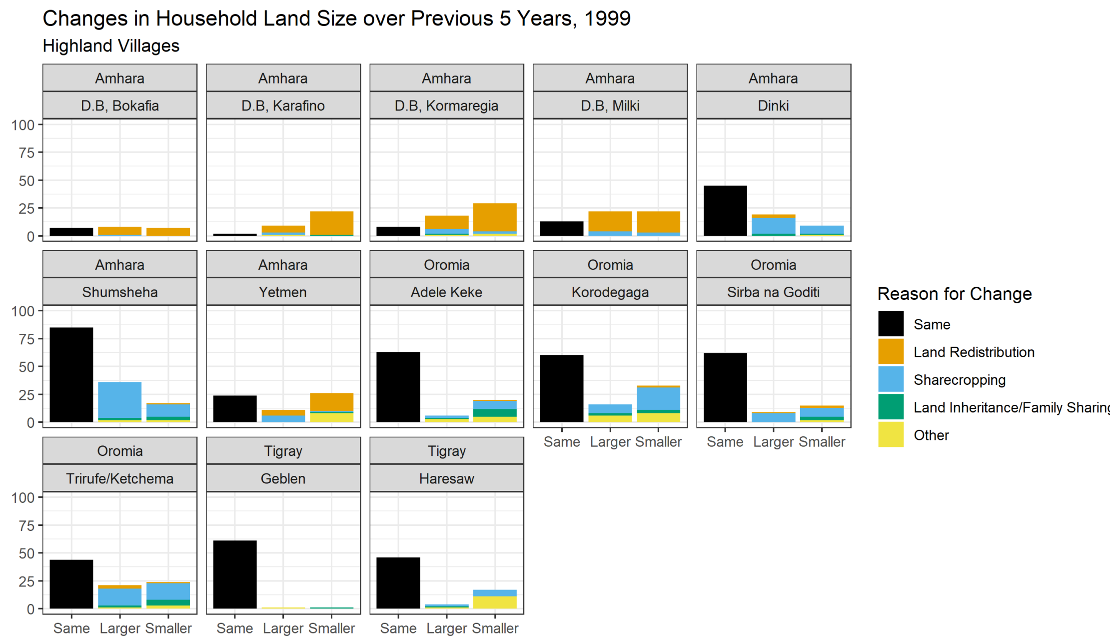

Figure 2.3: Land changes by village

those targeted were better off; however, real consumption per capita within the household suggests the opposite.

2.3.2 Agricultural Sample Survey

Given the limited scale of the ERHS, I also construct a regional level panel based on the Ethiopian Central Statistical Agency's (CSA) Agricultural Sample Survey (AgSS)[3]. This dataset provides national and regional estimates of rural agricultural production based on representative samples of households across the country. While the household level data are not publicly available, the CSA has made public the estimates for higher level adminstrative units. I digitize these reports for the years 1995–2000 to construct a panel of yields for the 6 largest cereal crops by production, and crop area under some kind of improvement (fertilizer, irrigation, improved seed, and pesticide). The AgSS was not conducted in 1993-94 as CSA resources were directed towards the national census, and there is another break

[3]Available at statsethiopia.gov.et

in the publicly available data post-2000. Additionally, some administrative reorganization occurred after 2001, such as the creation of new zones and shuffling of zones across regions, making direct comparison to earlier years infeasible.

The top level administrative regions in Ethiopia are regions, which are in turn composed of zones. For 1995–96, the CSA reports only for aggregations of zones, typically groups of 3–4. For 1997–2000, data are reported for each zone separately. To maintain consistent units of comparison for 1995–2000, I aggregate area and production by crop for 1997–2000 to construct equivalent Zone Aggregations (ZA) to the 1995–96 reports.

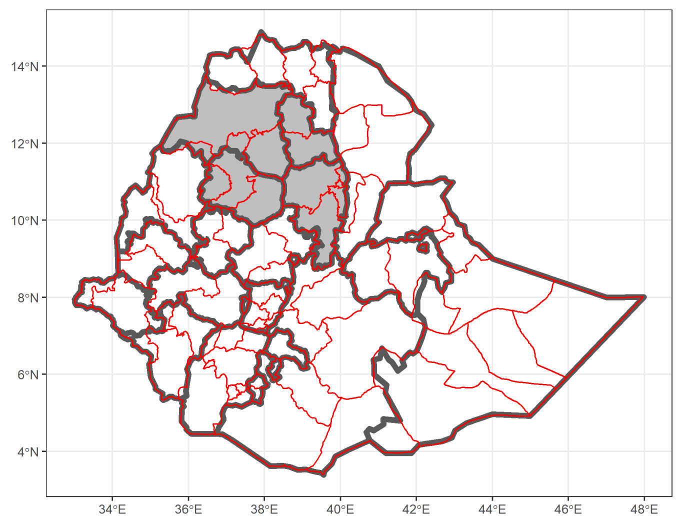

Figure 2.4: Construction of Zone Aggregates

Figure 2.4 shows the construction of the Zone Aggregates from. The region undergoing the government redistribution is shaded.

2.3.3 GIS Controls

Agriculture in Ethiopia was and still is dominated by rural smallholders highly dependent upon the natural environment. Controlling for potential environmental differences across observation units is crucial given the geographically limited aspect of the redistribution.

Rainfall is a particularly important confounder as it varies across both time and space. I construct a panel of controls at the regional and village level using GIS sources on soil suitability, rainfall deviation, nightlights, population density, and market access. For soil suitability, I use the FAO's Global Agro-Ecological Zones (GAEZ) 1960–2000 average measure for five of the main cereal crops present in the ERHS and CSA data: barley, maize, millet, sorghum, and wheat (FAO and IIASA, 2020). For teff, the endemic cereal crop that is a staple of highland farms, I substitute the suitability of wheat. Rainfall data is constructed using the Global Precipitation Climatology Centre's (GPCC) monthly precipitation product, which provides gridded rainfall estimates at the $.25°$ x $.25°$ level by year (Schneider et al., 2020). Nightlights serve as a proxy for existing economic development (Li et al., 2020). Gridded population density is first available in 2000 (CIESN, 2018). Jedwab and Storeygard (2021) have calculated market access for a $.1°$ x $.1°$ grid of Africa at ten year intervals from 1960–2010.

For each village in the ERHS, the average value for a buffer of 10 km radius around the location provided in the project documentation is extracted. This reflects the spread out nature of these rural communities, and ensures each village has at least one observation across each spatial input. For the CSA level data, the average value within each ZA is calculated for each control.

Table 2.2 presents summary statistics for the ERHS data with village level geographic controls. There are some important differences between the households in the treated Amhara region and the control villages in Tigray and Oromia regions. On average farmers in the Amhara villages have larger lands, about one less person per household, a greater number of oxen, and higher real per capita consumption. Summer rainfall levels are higher in the Amhara villages, but yearly deviations during the summer and harvest seasons do not appear to differ between treatment and control villages. The sample size for time invariant village level geographic features is likely too small to have enough power to detect meaningful differences between groups. While market access grew on average for all villages

Table 2.2: ERHS Summary Statistics

Variable	N	Control Mean	SD	N	Treatment Mean	SD	tstat	p.value
Household Level								
Land Size	2068	1.54	1.334	1692	2.232	1.603	-14.188	0
Hhold Size	2258	6.212	2.841	1830	5.226	2.255	12.376	0
No. Oxen	2206	0.785	1.078	1801	1.112	1.109	-9.399	0
Real Consumption per cap.	2247	84.935	79.939	1829	105.196	121.414	-6.136	0
Village Level Time Variant								
Summer rain (level)	30	134.133	35.152	20	194.308	56.636	-4.224	0.0002
Harvest rain (level)	30	46.817	27.625	20	47.61	32.052	-0.090	0.929
Summer rain (% dev.)	30	6.072	19.216	20	9.78	19.166	-0.665	0.510
Harvest rain (% dev.)	30	-5.454	31.625	20	-7.087	31.845	0.177	0.860
Nightlights	30	0.535	0.537	20	0.307	0.487	1.548	0.129
Village Level Time Invariant								
Market Access 1990	6	5.478	2.499	4	3.287	2.806	1.263	0.254
Market Access 2000	6	6.502	2.137	4	3.938	2.414	1.721	0.136
δ MA	6	1.024	0.79	4	0.651	0.426	0.965	0.363
GAEZ Cereals	6	5.457	1.32	4	4.746	0.526	1.186	0.274
Log Pop. Density 2000	6	5.126	0.721	4	4.891	0.538	0.590	0.572

between 1990 and 2000, the Amhara villages appear to have had slower growth during the decade. This could bias results downwards, as market access is associated with access to output enhancing investments such as fertilizer. See Appendix B for further detail on GIS controls.

2.4 Empirical Framework

Besley and Ghatak (2010) provide a basic model of a producer with stochastic output and a linear utility function. The optimal choice of labor for such a producer facing expropriation is:

$$e^* = \left[\frac{(1-\tau)A}{2}\right]^2 \tag{2.1}$$

where e is effort, A is output, and τ is the probability of expropriation. This is equivalent to a simple tax, and has the result that increases in *tau* decrease effort and output.

Measuring τ is difficult, and studies on tenure security have used a number of proxies.[4] Given the unexpected and regionally limited aspect of the 1997 land redistribution it should be the case that farmers within the Amhara region expected τ to be greater post-1997 than previously, while farmers in other regions were unaffected. A difference-in-differences framework allows me to isolate the effect of this increase in expected expropriation regardless of any pre-existing differences in τ across regions.

I use a simple two-way fixed effects framework of the form:

$$Y_{icgt} = \alpha_g + \alpha_t + \delta Treated + \zeta_c + X'_{ict}\beta + \epsilon \tag{2.2}$$

where Y_{icgt} is the yield for crop c from producer i in group g at time t. Fixed effects for membership in the treated group and the post-treatment time period are α_g and α_t, respectively. The coefficient of interest is the interaction of the group and time fixed effects δ, which captures the impact of the land redistribution on yields. I follow Gollin et al. (2021) in including a fixed effect ζ for each crop, as crops have differing yields and the crop mix varies across producers—this carries the implicit assumption that crops are not changing in relative productivity over time. The vector of controls X_{ict} includes factors that have both spatial and temporal variation over the study period that may affect agricultural productivity such as seasonal rainfall and changes in nightlight intensity. I also include controls for soil suitability by crop. If the increased risk of expropriation is the dominant effect of the redistribution then the sign on δ should be negative.

One important assumption for difference-in-differences is that there are parallel trends in the dependent variable prior to treatment (Huntington-Klein, 2021). Figure 2.5 shows the distribution of yields for treatment and control groups across crops in the ERHS sample. Yields in the Amhara region villages are lower during the pre-treatment period and move

[4] See Arnot et al. (2011) for an overview.

closer to the level of the control villages in the post-treatment period. The considerable year-to-year variation makes the assessment of parallel trends absent controls difficult, however. Differences in rainfall may be an important contributing factor to this variation.

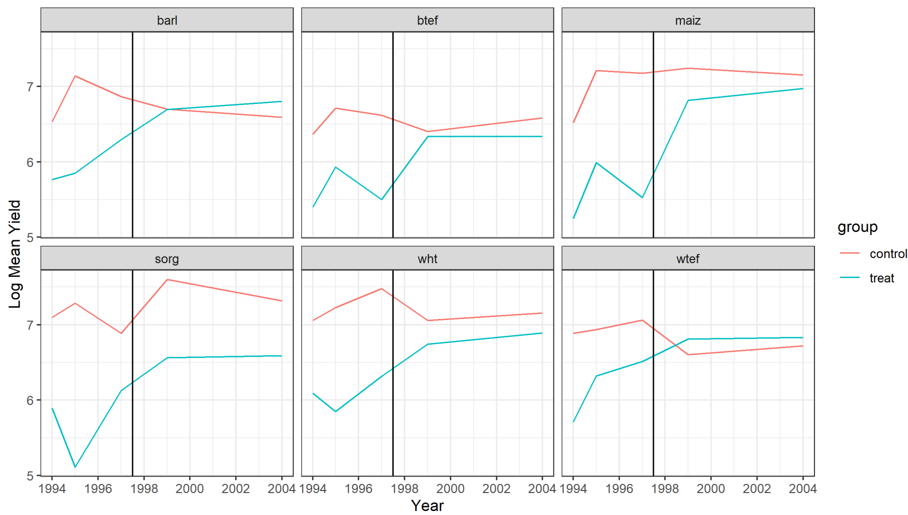

Figure 2.5: Log mean yields by crop over time, ERHS

2.5 Analysis

I investigate the effect of the 1997 land redistribution at two levels. First, I treat the land redistribution as an intent-to-treat policy effect at the regional level to try to capture any broader changes in farmer behavior due to the increase in tenure insecurity. Second, I hone in on the effect of those directly targeted by the redistribution in the Debre Berhan villages.

2.5.1 Intent-to-treat Policy Effect

I follow equation 2.2 in setting up a two-way fixed effects regression using presence in the Amhara region as the group fixed effect and the post-1997 redistribution period as the time fixed effect. The dependent variable is the log of cereal crop yield, and the coefficient of interest is the interaction between the group and time fixed effect 'Treated'. I leverage the panel nature of the data to use Driscoll-Kraay standard errors, which account for spatial and temporal dependence even when N is much larger than T (Driscoll and Kraay, 1998).

Table 2.3 shows the regression results for the ERHS data. While the sign on Treated is positive, the standard errors are large enough that this is not statistically significant. Adding controls moderates the effect somewhat from an implied 61% increase in yields due to treatment down to 47% with the full suite. This reflects the overall pattern present in figure 2.5, with some crop yields rising from low levels in 1994 and 1995 to match the yields present in the control villages.

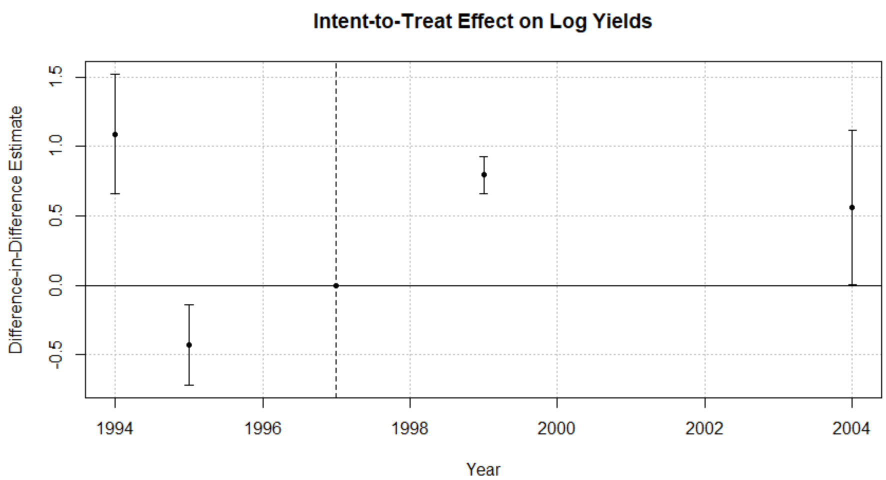

Figure 2.6: Intent-to-treat effect on log cereal yields, ERHS

To get a finer grained picture of the policy effect, figure 2.6 shows the treatment fixed effect interacted with year fixed effects. The period just prior to the redistribution is set as the reference point. In a typical difference-in-differences setting, the coefficients prior to the treatment taking effect should be zero: there should be no significant effect on the dependent variable from being in the treated group prior to the treatment taking place. Here the years prior are both significantly above and below zero. This likely reflects the high variability in the data seen in figure 2.5, and suggests that the included controls are insufficient. It may also reflect more fundamental problems with the source data.

Table 2.3: Intent-to-treat Effect on Log Yields, ERHS

Dependent Variable:		lyield	
Model:	(1)	(2)	(3)
Variables			
TreatedTRUE	0.6172	0.4848	0.4628
	(0.3211)	(0.3243)	(0.3977)
larea		-0.2959***	-0.3021***
		(0.0317)	(0.0386)
hhsize		0.0436**	0.0401**
		(0.0141)	(0.0107)
toxen		0.0292	0.0318
		(0.0541)	(0.0530)
gaez			0.4350
			(0.2123)
sumdev			0.0054
			(0.0111)
harvdev			0.0016
			(0.0052)
nightlights			-0.6656
			(0.6936)
Fixed-effects			
village	Yes	Yes	Yes
year	Yes	Yes	Yes
crop	Yes	Yes	Yes
idcode	Yes	Yes	Yes
Fit statistics			
Observations	8,252	8,072	8,072
R^2	0.42036	0.43169	0.43879
Within R^2	0.01183	0.03336	0.04544

Driscoll-Kraay (L=1) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2.4: Intent-to-treat Effect on Log Yields, CSA

Dependent Variable:		lyield	
Model:	(1)	(2)	(3)
Variables			
TreatedTRUE	0.0184	0.0163	0.0082
	(0.0412)	(0.0417)	(0.0412)
summer rain		-0.0004***	-0.0005***
		(9.86×10^{-5})	(8.72×10^{-5})
harvest rain		-0.0006	-9.68×10^{-5}
		(0.0004)	(0.0003)
nightlights		-0.0040	-0.0376
		(0.0722)	(0.0748)
gaez			0.1198***
			(0.0038)
Fixed-effects			
CropType	Yes	Yes	Yes
Region	Yes	Yes	Yes
time	Yes	Yes	Yes
Fit statistics			
Observations	552	552	552
R^2	0.68154	0.68777	0.72216
Within R^2	0.00026	0.01982	0.12776

Driscoll-Kraay (L=1) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2.4 repeats the regression using the regional level CSA data. Again, the sign on Treated is positive but not statistically significant. Interacting the treatment with year fixed effects in figure 2.7shows no response.

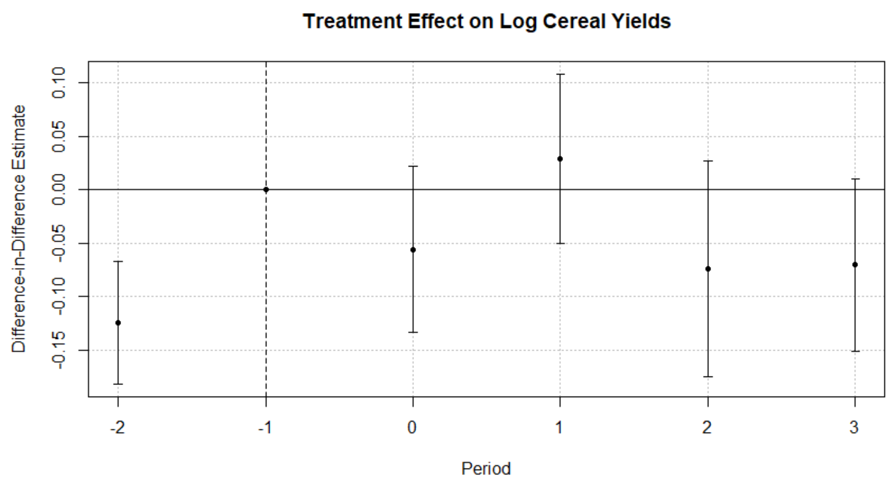

Figure 2.7: Intent-to-treat effect on log cereal yields, CSA

I also look at patterns of land investment in fertilizer, irrigation, pesticide, and improved seed uptake. Unfortunately the CSA data are inconsistent in providing estimates of these agricultural improvements at the zonal level, prohibiting the construction of a useful panel at the ZA level. Figure 2.8 shows the provided regional level estimates for land improvement in the four largest agricultural regions over time. There are no obvious common trends prior to the treatment and no apparent change in the Amhara region after 1997. While a sizeable minority of farmers use fertilizer most other improvements see very little uptake or change during the study period.

2.5.2 Direct Treatment Effect

To measure the impact of the expropriation shock more directly, I implement the same difference-in-differences model as in equation 2.2 using a subset of the data containing the four periurban villages in the Debre Berhan area. Based on answers from the 1999 questionnaire these villages all saw extensive redistribution. For these regressions the treatment effect is the interaction of the post-1997 period with whether a farmer reported having lost (or gained) land in the 1997 distribution. All four villages share the same location characteristics so geographic controls are removed.

First I take reported the log of land size as the dependent variable to see whether the redistribution actually resulted in treated farmers having less land. Table 2.5 shows that treated households saw their lands decrease by 25% in the post-redistribution period, whereas those households that were given land saw it increase by about 8.6%. Figure 2.9 shows the interacted treatment coefficients for each year relative to just before the redistribution took place. Land sizes appeared to have been higher prior to treatment and are reduced in the period afterwards, with some moderation in the longer run.

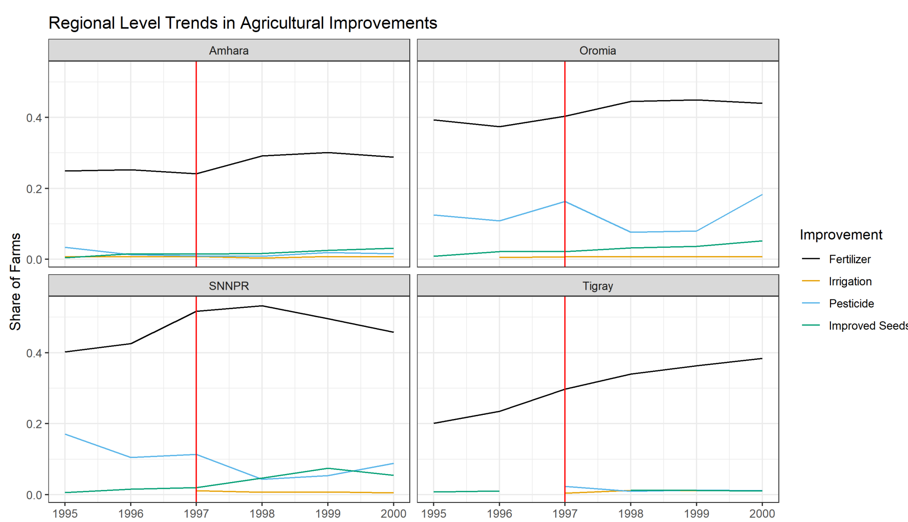

Figure 2.8: Share of farmers using land improvements by region

Table 2.5: Treatment effect on land size

Dependent Variable:	lland	
Model:	(1)	(2)
Variables		
TreatedTRUE	-0.2538***	
	(0.0333)	
GivenTRUE		0.0863**
		(0.0232)
Fixed-effects		
taken	Yes	
time	Yes	Yes
village	Yes	Yes
idcode	Yes	Yes
given		Yes
Fit statistics		
Observations	768	768
R^2	0.36057	0.35004
Within R^2	0.01777	0.00160

Driscoll-Kraay (L=1) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2.6: Treatment effect on crop yield

Dependent Variable:	lyield	
Model:	(1)	(2)
Variables		
TreatedTRUE	-0.0651	-0.0810
	(0.0550)	(0.0473)
hhsize		-0.0037
		(0.0241)
toxen		0.0736**
		(0.0216)
lland		-0.1015
		(0.0643)
Fixed-effects		
taken	Yes	Yes
time	Yes	Yes
crop	Yes	Yes
village	Yes	Yes
idcode	Yes	Yes
Fit statistics		
Observations	1,383	1,347
R^2	0.42605	0.43420
Within R^2	0.00033	0.00917

Driscoll-Kraay (L=1) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

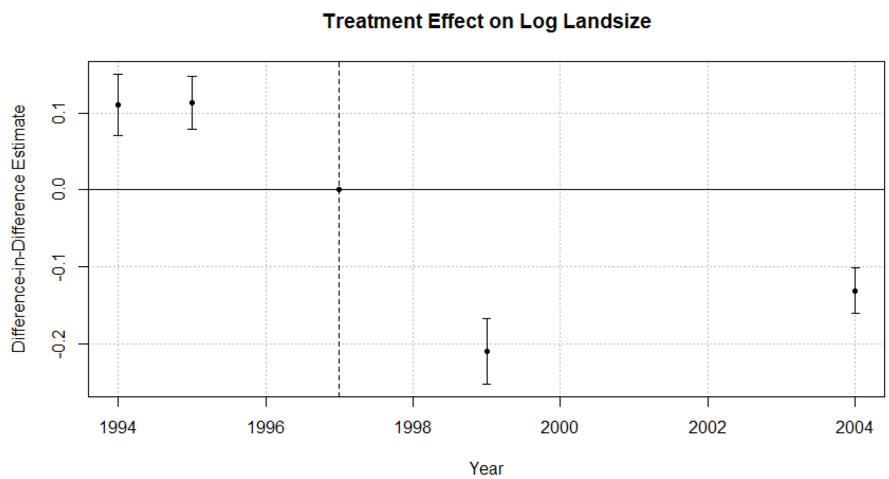

Figure 2.9: Treatment effect on landsize

In table 2.6 I take the log of crop yield as the dependent variable to examine the direct effect of land confiscation on crop yields. Here the sign is negative, but again statistically insignificant.

Given these results, one reason the redistribution may not have aggregate effects on Amhara farmers is that is was not very impactful on average for those directly targeted. This may suggest that transaction costs are low enough in some regard that land reassignment has little effect other than transferring some wealth between households. As noted earlier in section 2.2.2, oxen are a critical input to plow agriculture, are generally scarcer than land, and local market institutions governing the distribution of oxen services have a rich history. If an oxen rich household were to lose some lands, they may still recoup some of that loss by lending out the oxen that will still be needed to plow those fields, or via the ox-poor households lending land back to those with the scarce factors of production. Table 2.1 showed that households who had land taken had more oxen on average.

Additionally, recall from figure 2.2 that although government redistribution ceased after 1997, the ERHS households still frequently experienced land changes in 2009 via sharecropping, either lending out land or renting it. The government redistribution in 1997 may have temporarily substituted for/crowded out sharecropping that would have taken place regardless.

Another potential element at play is that farmers had the ability to locally affect the redistribution process, despite centralized directives. Ege (1997) describes some villages where the *birokrasi* were valued members of society, and farmers there acted to mitigate the damage done by the redistribution either through social pressure on the elected committee members, or through subsequent illegal land sales back to *birokrasi*. In this case, local property rights may have been *de facto* secure despite the government's attempts at *de jure* reassignment.

Despite local adaptations to reduce exchange costs, it could still be the case that a general environment of low tenure security in Ethiopia was an important factor in agricultural stagnation in the 19th and 20th centuries. The few years experience of liberalized EPRDF land policy at the study time may simply have not been long enough for any meaningful increase in perceived security in non-Amhara regions.

2.6 Conclusion

Overall my results are inconclusive. Crop yields in the Amhara region seem to have been higher in the post-redistribution period than previously, but there is considerable year-to-year variation that makes a precise estimation difficult. More direct measures of the impact of redistribution show a slight negative effect on yields, but again this effect is not statistically significant. This discrepency between the intent-to-treat effect and actual treatment effect may suggest that there were other important changes in the Amhara region post-1997 that are unaccounted for. One such factor may be prices for both crops and agricultural inputs.

Another root cause of these results may be data quality. The underlying data of both the ERHS and CSA are built upon farmer recall, and such self reported data for land size has been shown to differ from that collected by technologies such as GPS (Carletto et al., 2015). As yields are constructed from estimates of both land area and output volume, high variability in both likely interacts to create unreliable estimates.

Despite these challenges, this episode deserves further study as it relates to the intersection of political economy and development in a very poor country. New advances in difference-in-differences methods, such as synthetic controls, may be better able to leverage the limited data available to researchers.

Chapter 3: Late Medieval English Soldiers

Abstract

The development of European state capacity after 1500 has been linked to a Military Revolution in the 16th century that saw dramatic increases in the costs of warfare, in part due to the need to recruit masses of professional infantry. Yet some historians have argued that the roots of this process lay earlier in the Hundred Years' War, a long running conflict between the royals of England and France. In this chapter, I use tools from the field of text analysis to analyze a database of over a quarter million soldier names from the years 1369–1453. Using soldier surnames as a proxy for social background, I find that the participation of soldiers of higher social status decreased over the length of the conflict. I also find suggestive evidence that the links between soldiers and their recruiting captains grew increasingly impersonal compared with the earlier medieval era.

3.1 Introduction

A longstanding thread in economic history is the importance of state capacity for long-run economic development (Johnson and Koyama, 2017). By most measures, European states only began to grow in capacity post-1500 (Dincecco, 2015). One influential argument for why growth occurred at this time was the need to collect tax revenue to fund warfare that was growing increasingly expensive due to the Military Revolution (Gennaioli and Voth, 2015). During the period 1500–1800, a combination of growing army sizes, increasing professionalization, and technological advances such as gunpowder, made state survival dependent upon mobilizing increasingly large sums from the populace (Parker, 1996). Thus the Military Revolution likely played a large role in the formation of the first modern centralized states.

However, historians have debated the nature, timing, and importance of this Military Revolution since the introduction of the thesis in Roberts (1956).[1]. In the view of Rogers (1993), the Military Revolution is best understood as a series of incremental and cumulative changes, and the typical focus on post-1500 developments ignores an earlier Infantry Revolution occurring in the two centuries before. In this telling the Hundred Years' War (1337–1453), a period of extended conflict between the royals of England and France, drove important advances in the social technologies of organization and recruitment that set the stage for later mass conscription after the advent of gunpower. England in particular had developed a decentralized system of recruitment, funded by centralized payment via the royal Exchequer, that was able to mobilize armies on a much greater scale than the centuries prior.

How this system operated, and who participated in it, are still topics of active study in late medieval history. Luckily for researchers, the record keeping needed to pay these growing numbers of soldiers has left behind a wealth of evidence. The Soldier in Later Medieval

[1] See Rogers (1995) for an overview of the debate

England project[2] (Bell et al., 2013) has assembled a database of all existing English military documents from the period 1369–1453, covering the latter half of the Hundred Years' War. Such databases allow for new kinds of social science analysis using the tools from big data to shed light on the development of the state and centralized control over violence.

In this chapter, I use tools from text analysis to investigate the claim that the English soldiery was growing increasingly professionalized over the course of the Hundred Years' War. First, I use evidence from the distribution of elite surnames to examine changes in the social background of those participating in the conflict. I find that soldiers of relatively elite backgrounds were present less often over time as patterns of military activity changed from short term raids to long term garrisoning. Second, I compare the distribution of soldier surnames among the retinues of the great landed magnates, the dukes and earls that held regional power and were crucial managing the recruitment of military force for the king. Using methods from latent semantic analysis and population genetics, I find little evidence that the great captains were geographically limited in their recruitment reach. This supports the existence of national level military community and a relatively developed market for specialists in violence.

This chapter relates to a number of growing literatures. The use of names as a source of social information has been used widely across economics and economic history (Fryer and Levitt 2004, Clark 2014, Cook et al. 2014, Anderson and Bentzen 2022). This chapter also relates to the literature on how elites mobilize individuals to participate in conflict (Gates 2002, Dippel and Heblich 2021, Rogall 2021). It builds upon a body of work by historians investigating the armies of late medieval England via large scale assembly and linkage of military records (Ayton 1999, Simpkin 2008, Bell et al. 2011, Bell et al. 2013). The closest paper in spirit to this is Bai et al. (2021), which uses casualty lists from the Taipeng Rebellion (1850–1864) to investigate network-induced mobilization of soldiers by regional elites during and how that mobilization affected political power after the war.

[2] Available at medievalsoldier.org

The rest of this chapter proceeds as follows: In section 2, I give an overview of changes in the English army leading up to and through the Hundred Years' War. Section 3 describes the data from the Soldier in Later Medieval England project. Section 4 presents the analysis of social background and regional recruitment through surnames. Section 5 concludes.

3.2 Background

3.2.1 Late Medieval England

Neither England nor France prior to 1500 should be considered anything like a modern state. In the framework of North et al. (2009) they were 'natural states', groups of colluding elites able to use their ability to control violence to extract economic rents from the social peace dividend. The king was not a divine autocrat, but merely the largest player in a dominant coalition of regional magnates—the large baron landowners able to tax their populations and raise up armies. In England this system arose from the Norman Conquest of 1066, when the Duke of Normandy, William the Conquerer, invaded from northern France and imposed a feudal monarchy. Throughout the next several centuries, conflict would continuously erupt not only between the royal houses over control of the territories of modern day France, but between kings and their barons. By the 14th century an equilibrium of power sharing between the king and the barons, via Parliament, had been established allowing for nascent rule of law and taxation with consent (Koyama, 2016).

These struggles were accompanied by changes in the industrial organization of military force. Rulers were constrained by low tax capacity, so in the early feudal system force was intially mobilized by social and political bonds. During the 12th century, under the reigns of Henry I (r.1100–1135) and Stephen (r.1135–1154) a system of *scutage* had developed whereby reluctant barons could pay a fine rather than render military service. The crown then used these fees to hire mercenaries to fight instead (Crouch 2007).

However, 1270–1327 saw a 'remilitarization of the gentry' in which traditions of service among gentry families formed as a result of a period of intense military activity (Simpkin

2008 ,Simpkin 2011). King Edward I (r.1272–1307) organized many campaigns across four theaters: Wales, Gascony, Flanders, and most intensely Scotland, which saw near continuous war from 1296 to 1328 (Prestwich, 2005). These campaigns saw extensive and repeated soldiering among the landholding community in service as mounted cavalry who were typically not paid by the crown but served in fulfillment of feudal obligations, as in earlier eras (Prestwich, 2005). Men were led in retinues by the captain who recruited them, typically members of the peerage or other political elites. Simpkin (2011) finds that retinue leaders at this time would repeatedly return to the same localities and recruit the same men, and these recruiters had stronger links to some areas than others.

Fighting alongside these mounted aristocrats were archers and light infantry recruited by *commissions of array*: orders from the sovereign to local gentry and administrators at the local *ville* to provide the desired number of foot soldiers. These soldiers were typically ill-trained and ill-equipped, and served mainly as a human shield for the cavalry as they readied for charges. The armies of Edward I were then a hybrid of mounted aristocrats supported by large blocks of peasant infantry (Prestwich, 1999).

3.2.2 Edward III and Military Innovation

The reign of Edward III (1337–1377) is widely seen as a key turning point for the English military structure (Simpkin, 2008). Two major innovations occurred: 1) the Crown would pay directly for military service, rather than invoke feudal obligation; 2) mixed retinues of mounted men-at-arms and archers would increasingly be recruited together, rather than separately via magnate retinue and commissions of array.

The nominal wage rates set by Edward III would persist throughout the Hundred Years' War: 12 pence a day for men-at-arms, and 6 pence for archers; those serving exclusively on foot (typically only in some garrisons), would sometimes see wages about 33% lower (Bell et al., 2013). In annual terms, a man-at-arms would expect to earn about the same income as a small manor for a year of service. An archer, at half the wage rate, would earn just over

£9 over a year, which compared favorably with the £10 needed to support the lifestyle of a *genteel* man. Other workers in England saw lesser but rising wages throughout the 15th century. For comparison, a carpenter would have earned around 4 pence a day, but also would rarely work for a full year (Dyer, 2002). Overall, compensation for military service compared quite favorably with other options for those of non-aristocratic background.

The shift in retinue recruitment seems to have proceeded quickly. In 1338 around 25% of archers were integrated members of retinues; by 1359 this share had risen to nearly 70% (Prestwich, 2005). This mixed retinue represented the major tactical innovation of Edward III's forces: rather than engage in combat on horseback, men-at-arms and archers would use their mounts to move into tactical position, then dismount for the actual fighting, which took the form of a defensive line of men-at-arms flanked on both sides by archers. Volleys from the longbowmen would weaken enemy cavalry charges, leaving short work for the men-at-arms, who wielded a variety of melee weapons such as lances, swords, or halberds (Prestwich, 2005). These tactics required training, coordination, and discipline on the part of soldiers.

This transformation in the organization of military force was part of a general shift in English political economy. By the early 14th century, the feudal system of heritable obligations was being replaced by a 'bastard feudalism' (McFarlane, 1944) in which service to a lord was organized by contract and compensated by monetary payment[3]. The great landed magnates built up large 'affinities', men retained to the household via indentured contract[4] to serve in both administrative and military capacities. These retainers would then serve as the core of armies recruited for the crown's military adventures, but also served important roles in managing the administration of the magnates' estates.

In terms of military service, Ayton (2009) argues that the initial transition to payment for service led to little change in the kind of men who showed up for service. This is also

[3]See North et al. (2009, 77-109) for a description of the operation of bastard feudalism and land ownership

[4]So called because the contract would be written twice on the same sheet and then separated along a 'toothed' line so that the two halves could be uniquely matched in the case of a dispute of terms.

emphasized by (Walker, 1990) who notes that an indenture of retainer merely formalized and recorded the unwritten rules governing a relationship between lord and retainer. However, historians have suspected changes in the goals and conduct of war over time led to broader changes in the soldiery (Bell et al., 2013). The Hundred Years' War contains a prime example of such a shift in war aims. During the first half the main goal of the English army was not to directly engage with French forces, but to ransack the French countryside. These raids, or *chevauchees*, were conducted in pursuit of plunder, and to undermine confidence in French leadership (Allmand, 2001). After the invasion of Normandy from 1415, and particularly after the Treaty of Troyes in 1420, however, English policy shifted towards long-run occupation of territory in Normandy. This shift has been suspected to have made warfare less amenable to those with lands and manors to attend to back in England as garrisoning required long-terms commitments, unlike the limited duration of the *chevauchee*.

3.2.3 Recruitment

Exactly how suppliers and demanders of military force were matched under the indentured retinue is still somewhat unclear. Clusters of knights from a particular locality would serve together regularly, even if not always with the same captain (Ayton, 2009). Captains and their recruiters would reach 'beyond household personnel and men bound by the terms of indenture of retinue to take advantage of established networks of tenants, neighbours, family and friends...' (Ayton, 2009). Early on, the recruiting grounds of the leading captains were related to the distribution of their estates: a typical knight in 1345 raised up of half of their retinue from among the gentry families and tenantry close to their lands (Morgan, 1987).

On the other hand, evidence suggests London was growing as a centralized recruitment point throughout the war. Based on letters of protection, legal documents filed for the safekeeping of assets before a soldier went on campaign, Bell et al. (2013) have speculated that overall military service was a national effort, rather than regionally based. The share of letters of protection naming London as a soldier's place of origin grew from an average of 14.8% in 1369–1389 to 26.5% in the years 1415–1453. This suggests that recruiting agents grew to rely upon centralized recruitment in focal points such as London rather than returning to the same regional areas Bell et al. (2013, 216–241). One practical reason for this is that soldiers headed for France were not paid until mustering just before departing at from port in southeast England. Any soldiers traveling from far away would have to travel for effectively free.

Historians have described the men-at-arms as a social melting pot 'containing on one hand reprentatives of families of ancient lineage... and on the other, men of obscure origins whose names and reputations rested solely on their ability to wield lance and sword effectively in the service of the crown' (Bell et al., 2013). In the sections that follow, I attempt to determine the degree to which this was true for the men-at-arms as well as longbow archers.

Table 3.1: Timeline of Major Events during the Hundred Years War

Phase	Year	Selected Events
Edwardian Phase		
	1337	Hundred Years' War begins when Philip VI of France disputes Edward III's claims to Aquitaine in France
	1346	Longbow archers defeat French cavalry at Crecy; English capture the port of Calais
	1356	Battle of Poitiers: King John II of France captured for ransom
	1360	Treaty of Bretigny: Edward III's right to French lands recognized
Caroline Phase		
	1369	Charles V of France resumes hostilities; much of England's continental territory is lost; military activity focused on raiding the French countryside
	1389	Richard II of England and Charles VI agree to the Truce of Leulinghem
Lancastrian Phase		
	1415	Henry V resumes war and invades Normandy, English archers again secure victory at the Battle of Agincourt
	1429	Tide turns against English with loss at Orleans; England slowly loses territory over the next twenty years
	1453	Battle of Castillon, England loses last mainland territory apart from Calais

3.3 Data

The Soldier in Later Medieval England project has transcribed over 250,000 names from military documents ranging from 1369-1453, which covers the latter two of the three traditional phases of the Hundred Years' War.

A typical entry in the database contains a soldier's name and rank, along with their captain, the service they were engaged in, and the date and source of the document. Occasionally, entries are supplemented with a soldier's origin and additional levels of commanding officers.

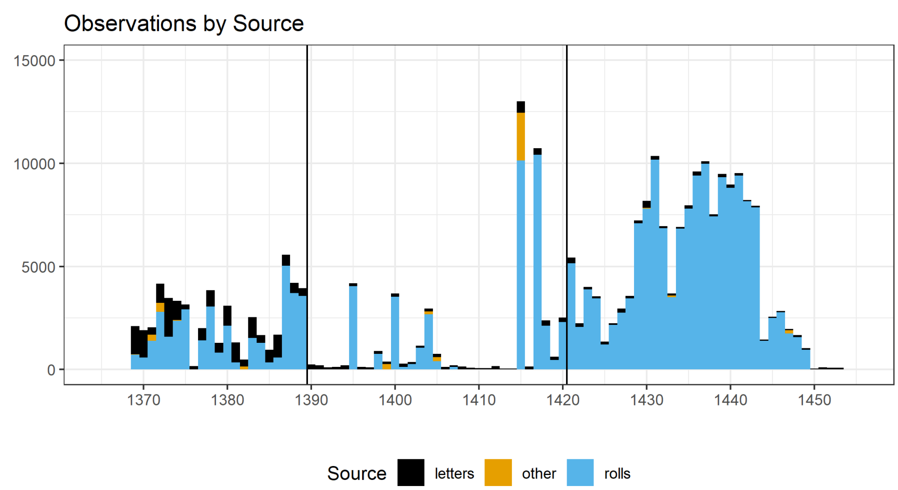

Figure 3.1: Records by Source

Figure 3.1 shows that entries in the database come from two main document types. One, letters of protection/attorney, were legal documents filed for the protection of property while the owner was away on military campaign for extended periods of time. These represent just under 10% of all entries, concentrated in the early years. The other primary source are muster or retinue rolls, which comprise 89% of entries. These served as the documentation for the indenture system. The remaining sub-2% of entries are a few scattered sources: one large example is the 'sick list' from Agincourt, which noted all the soldiers sent back to England after contracting dysentery during the campaign.

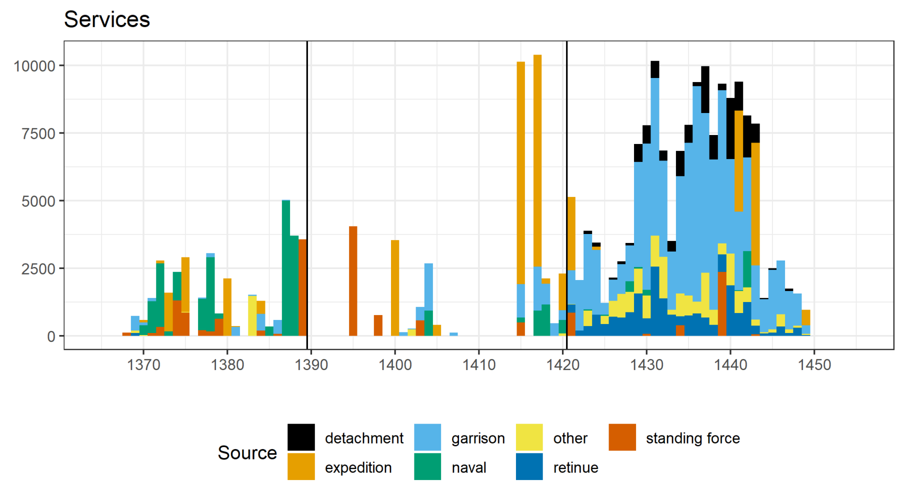

Figure 3.2: Records by Service

The surviving muster rolls record a wide range of activity over the eighty years. In Figure 3.2, I group the noted service in the documents as falling into 7 major categories: expeditions, naval service, standing force, garrisons, retinues, and other. The mix of activity changes significantly across the recorded periods. Several expeditions to France are recorded, along with a large expedition to Scotland in 1400, and a small expedition to Wales in 1405. Naval service is most prevalent in the early period, though present across the sample. A number of standing forces across multiple theaters are recorded. Most dramatic is the increase in garrisons associated with the invasion of Normandy intiated with the 1415 and 1417 expeditions. These garrisons not only defended towns captured by the English, but were also expected to serve in the field to support sieges or other engagements via detachments. Retinue records are either the personal or official retinues associated with captains in the course of the war in France. All other services are categorized as other: these are a mix of varied escorts, ordinance companies in sieges, field service in Normandy where origins are unclear, and otherwise unspecified service.

Figure 3.3 demonstrates changes in the ranks of soldiers over time. There are two primary categories: the man-at-arms, and the archer. These represent 27% and 64% of entries over the entire sample, respectively. Note the change in the ratio of men-at-arms to archers. At some point after 1390, royal policy shifted from requesting equal numbers of men-at-arms and archers in retinues, to three archers for every man-at-arms. This change could reflect some combination of budget constraints and increasing difficulty in recruiting men-at-arms over the course of the war (Bell et al., 2013). Just 1% of labelled records contain some other rank, such as crossbowman or hobelar (light mounted cavalry), and 9% of entries lack any information on rank; these missing ranks are overwhelmingly from letters of protection.

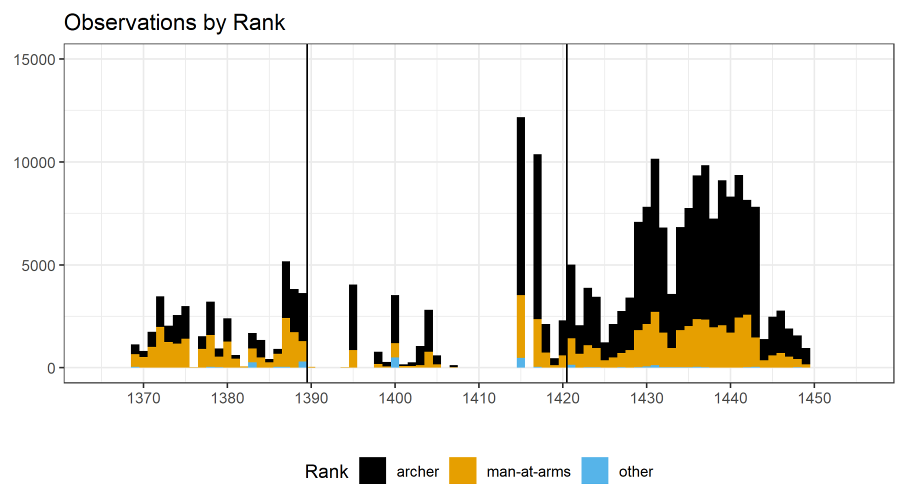

Figure 3.3: Records by Rank

One question is how complete this record is. Curry (1992) estimates that for the garrisons of Lancastrian Normandy perhaps 50% of the source material survives. The sizable campaigns into Scotland of 1385 and 1400 are underrepresented, as are the smaller Irish campaigns of the 1390s. Overall the database should be considered as a random sample. [5]

[5]See Bell et al. (2013, 271–274) for total yearly campaign estimates for data coverage period.

3.4 Analysis

Using this database I am able to investigate aspects of the professionalization of the English soldiery during the Hundred Years' War. I implement tools from the field of text analysis to regularize and categorize the surnames of soldiers in the muster rolls. First, I code whether each soldier or captain surname falls into one of several surname categories associated with different social status. I obtain the simple sum of the number of occurences of each surname per year per category, and divide by the total number of entries present in that year to get the share of each surname class. Based on previous historical research, I expect that 1) surnames will be associated with social status, and 2) the share of relatively elite surnames will be decreasing over time.

Second, I compare the surname distributions of soldiers who served under different regional magnates. If regional recruitment was more important than national, then magnates whose lands were closer together should see more similar surname distributions. If recruitment occurred on a more national level, then surname distributions should be similar regardless.

3.4.1 Surnames

The origins of English surnames can be grouped into four categories: 1) surnames of place; 2) surnames of relationship; 3) surnames of occupation or office; 4) nicknames (Reaney and Wilson, 1991). Place surnames can be further divided into *locative* and *topographical* surnames (Mckinley, 2014). In work on long-run social mobility in England Clark (2014) and Clark and Cummins (2014) use the relative frequencies of Norman placenames, locative surnames, and artisan surnames across elite institutions such as universities and the broader populace to find that social status has remained highly heritable since at least the 14th century.

Norman placenames originate from the Norman conquest of 1066, when William the Conquerer left his base of power in Normandy and took control of the Anglo-Saxon polities across the English Channel. Those who accompanied William and were installed as regional lords typically brought their Norman placenames with them[6]. These individuals and their descendents continued to be important players in the development of the English state, and would be among the most elite members of English society.

Similarly, locative surnames are those originating from the towns and villages across England. Since in the medieval era most people rarely moved, identifying oneself by a town of origin would only make sense for those who had left their place of origin. McClure (1979) argues that these locative surnames were most common among manorial lords, free tenants, merchants, tradesmen, and clergy. One challenge in identifying locative surnames is that there are thousands of potential places. Clark and Cummins (2014) utilizes the regularized endings of many towns: -ton, -dun, -ham, etc.

Artisan surnames are a subset of occupational surnames, focused on those specializing in a skilled trade (see Appendix C). Many of these surnames are familiar today: Smith, Baker, Cook, etc. While nowhere near as elite as the previous two categories, Clark suggests such artisans may have represented the 40th to 60th percentile of earnings, as they could earn wages double that of unskilled workers[7].

One important caveat is that these arguments presume that surnames were inherited across generations, and not simply adopted *ad hoc.* Evidence from lists of feudal landholders in the 12th and 13th centuries shows that a large fraction had heritable locative surnames (Mckinley, 2014). For non-elites, contemporary evidence is less clear cut. Clark and Cummins (2014) argues, based on differences in occupational surnames and listed occupation in poll tax documents in the late 14th century, that English surnames became heritable around 1250–1300.

The primary challenge in classifying recorded soldier surnames is that spelling was not regularized at this point in time, and so any given surname may appear as one of several

[6]Examples can be found in registers from the Domesday Book (1086)

[7]See Dyer (1989)

spellings. Smith, for example, is also just as likely transcribed as Smithe, Smyth, or Smythe. In text analysis, *tokenizing* is the process of regularizing words or phrases for later multi-dimensional analysis. The usual tokenizing strategies, such as stemming or matching to a list of common mispellings, are not helpful here so I create my own tokenizing program. I use the *fuzzyjoin* packing in R to match soldier surnames to the surnames in each category list based on the string distance, a count of the number of single-character edits needed to transform one surname to another. There are numerous measure of string distance; I use the Optimal String Alignment (OSA), a variation of the common Levenshtein distance, as OSA allows for accidental transpositions between adjacent letters without counting as a step. This is useful in a large dataset transcribed from written documents where mistyping may be common. When different surnames are close (e.g. Baker-Barker) this procedure can result in false matches, so I manually check matches and eliminate any that are incorrect. Finally, I hand match any major spelling variations of the same surname that have too great a string distance, e.g. Taylor and Taillour.

A point of concern is that without comparison to a reference group, it is possible that any trends in surnames could reflect other secular trends in society. One counterpoint based on Clark (2007) is that in a Malthusian world, higher status surnames should actually *increase* in the general populace over time. However, it remains possible that non-English soldiers were recruited. Private mercenary companies were utilized throughout Europe during this time, with some regions contributing specialized soldiers such as crossbowmen. While historians have noted that the English fought the Hundred Years' War overwhelmingly with their own soldiers, the garrisons of Lancastrian Normandy were known to hire local French Norman soldiers (Bell et al., 2013). From 1429 on English policy placed limits on the number of French soldiers that could serve in a garrison, but enforcement seems to have been inconsistent (Curry, 1992). Given the wide scale of garrison activity during the last thirty years of the war, it is possible that a sufficient number of these locals could give the appearance of the false trend.

I take two steps to address this. First, I categorize some of the most common remaining surnames to serve as a placebo. For additional occupations, I match to a list of service occupations typically associated with official duties in an elite household. For personal names, I match with common appearance based names. For patronyms, I match with the suffix -son (eliminating other matches such as mason). Topographic surnames are those associated with generic local land features, such as the village green, a brook, or woods. Other than the possible case of official occupations, none of these other categories should expect to see any particular correlation with social class. Second, I leverage the surnames of 1,708 soldiers identified as French or Norman in the muster rolls. While this is small, I match the surnames of these individuals across the entire sample to create a category of imputed French surnames. Note that these surnames are distinct from the Norman placenames associated with elite families in England.

Across all 8 categories about 36% of total surnames are categorized. Many of the unmatched surnames are either of ambiguous origin (eg Halle is a common surname, but is associated both with someone living in the manorial hall as well as living near coastal salt flats), placenames without locative suffixes, or variations on patronyms with a first name. For this to bias my results, it would have to be the case that uncategorized surnames were associated with social status within each category; this seems unlikely.

Figure 3.4 shows the volume of imputed French soldiers compared with all soldier records per year. Because of the historical Norman origins of many English, there is a small background level of soldiers identified as French throughout the sample. As expected, there is an increase in the number of these soldiers during the Lancastrian occupation of Normandy. For the moment, I remove these soldiers from the calculation of surname shares so as not to bias the denominator.

Figure 3.5 verifies that locative and Norman surnames are effective proxies for high social status. Panel A compares surname shares across the two main document sources. It would be expected that letters of protection and attorney would skew towards elite representation,

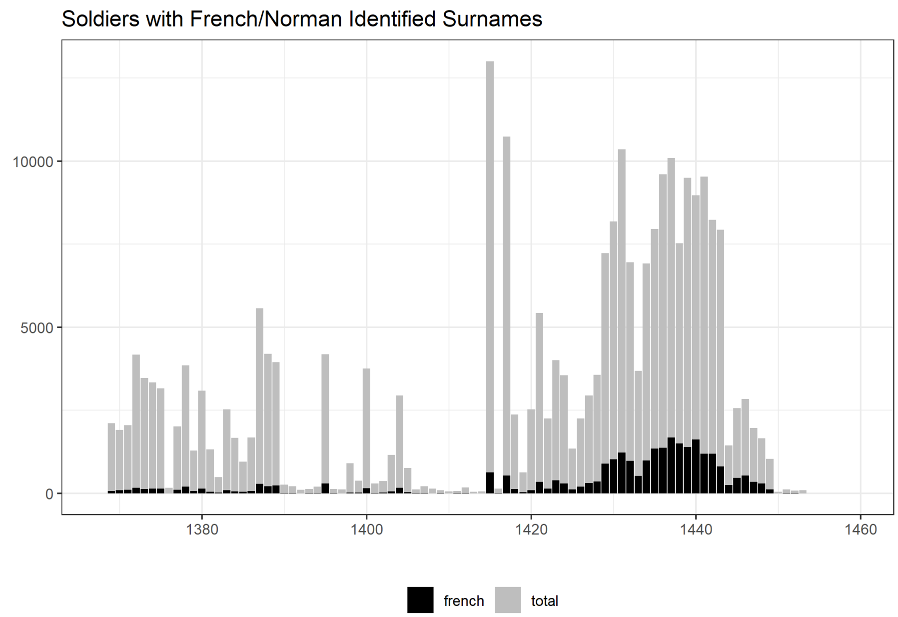

Figure 3.4: Soldiers with imputed French surnames

as only those with property to protect would be filing. Artisan surnames, associated with middling social status, see the reverse pattern. This also demonstrates the overall downward trend in elite surnames over time across both sources, despite varying levels of coverage. Panel B, showing the surname shares of captains and soldiers, also reflects the same pattern. Note that Norman surnames have a much higher level of represention among captains, in line with the importance of Norman descendants to the political order. Panel C again shows the same pattern, with higher status men-at-arms seeing greater shares of locative and Norman surnames than archers, and the reverse for artisan surnames. As further support, the apparent outlier points in 1398 for Norman surnames are members of the king's bodyguard retinue, which could reasonably be expected to be of high status.

Figure 3.6 shows the distribution of the placebo surnames. Personal and topographic surnames show no patterns of difference between any of the elite categories. Surnames associated with official occupations and -son patronyms do see similar patterns in levels as with artisan surnames across all panels, suggesting they are less elite. A noteworthy trend

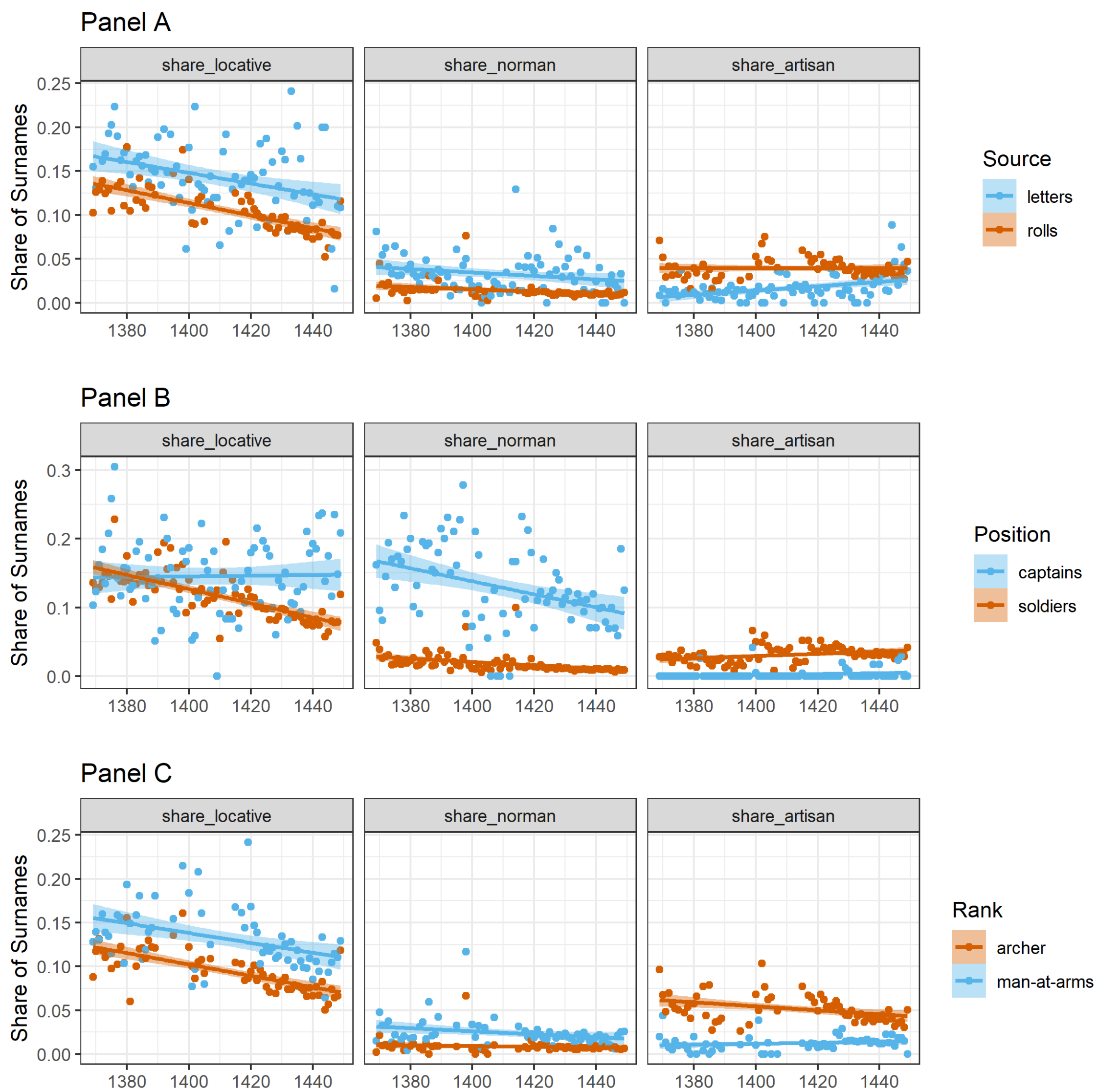

Figure 3.5: Validating elite surnames

in Panel B is the increase in -son patronyms among soldiers, and office related surnames among captains. Compare this to the near absence of artisan occupations among captains seen in figure 3.5 panel B.

In addition to the patterns in levels in figures 3.5 and 3.6, there are also trends over time in the shares of surnames. In particular, it appears that the elite coded locative and Norman

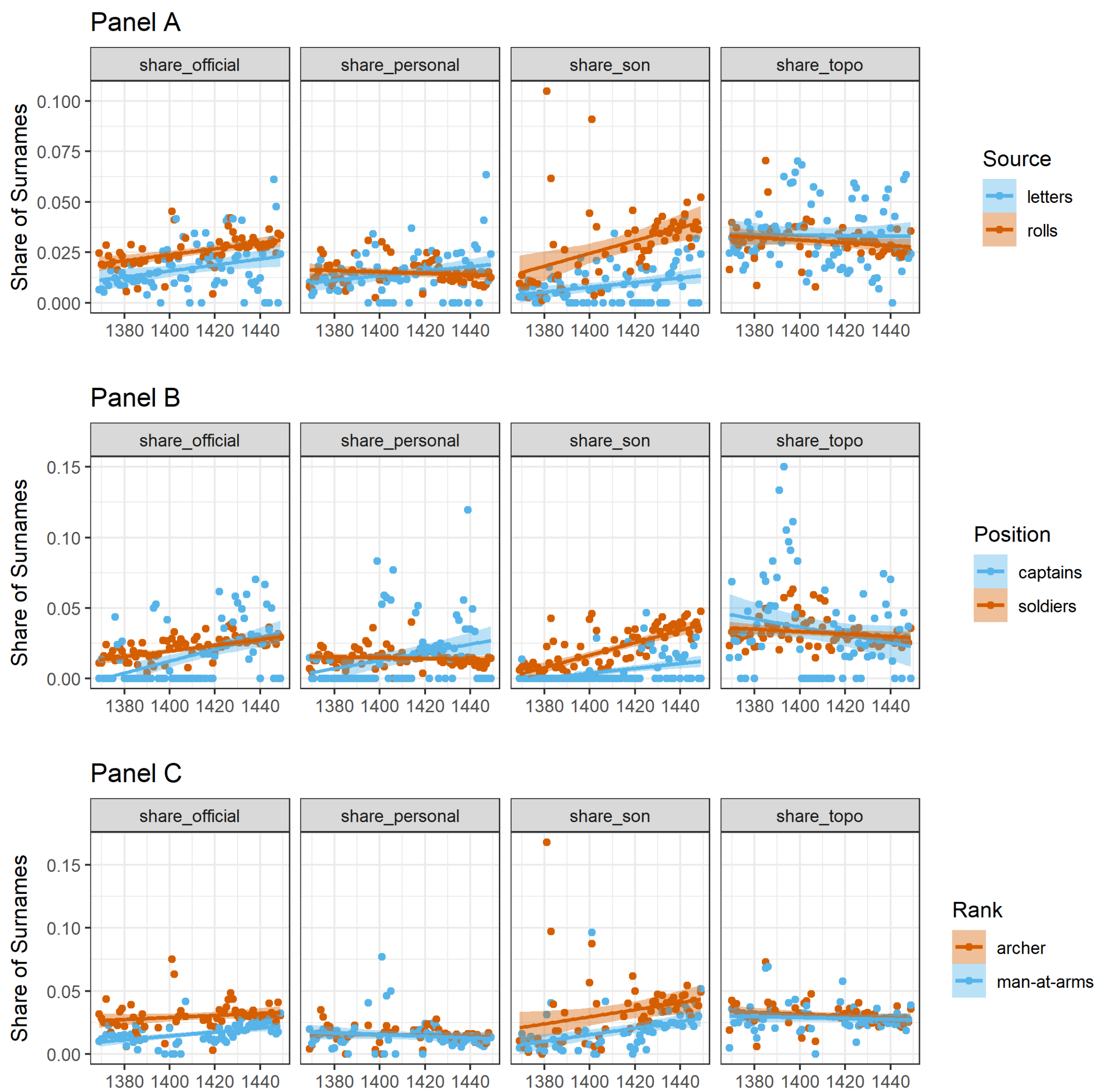

Figure 3.6: Validating non-elite surnames

surnames are decreasing over time. To assess the significance of these trends, I run a simple regression of the form:

$$share_{surname} = \alpha + \beta time + \epsilon \tag{3.1}$$

I use only data from the rolls, as they are the most comprehensive. The range is 1369 to 1449 for a total of 194,467 names. I remove all imputed French soldiers from the regression calculations so as to examine trends only within English soldiers.

Table 3.2: Decreasing elite surnames over time

	Dependent variable:						
				share			
	locative	norman	artisan	official	personal	topo	son
Constant	13.763***	1.909***	3.677***	1.544***	1.483***	3.138***	1.274**
	(0.443)	(0.239)	(0.301)	(0.156)	(0.157)	(0.215)	(0.435)
time	−0.062***	−0.013*	−0.005	0.005	−0.003	−0.004	0.039***
	(0.009)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)	(0.009)
Observations	62	62	62	62	62	62	62
R^2	0.444	0.102	0.012	0.047	0.011	0.012	0.250
Resid. Std. Err. (df = 60)	1.788	0.966	1.214	0.630	0.634	0.867	1.755

Note: All values in percentage points *p<0.05; **p<0.01; ***p<0.001

Table 3.2 shows that the downward trend in locative and Norman surnames is statistically significant. Over the 80 years of the sample, this represents a decrease of about 30% (4.8 percentage points) in the share of locative soldiers, and just under 50% (0.96 percentage points) in Norman soldiers. Artisan, official, personal, and topographic surnames see no linear trend. Also interesting is the increase in share of -son patronyms.

While this trend is present over the entire pool, recall from figure 3.2 that the kinds of services soldiers were engaged in changed, particularly after 1420. This different composition of activity may be driving these results so I repeat the surname share calculation for each kind of service, focusing on locative surnames. I exclude detachments as these are subsets of garrisons.

Figure 3.7 shows surnames shares broken down by service of the soldier. The number of observations for some activities are fairly low, so apparent trends should be interpreted with caution. In Panel A, locative surnames appear to have downward trends in garrisons, naval

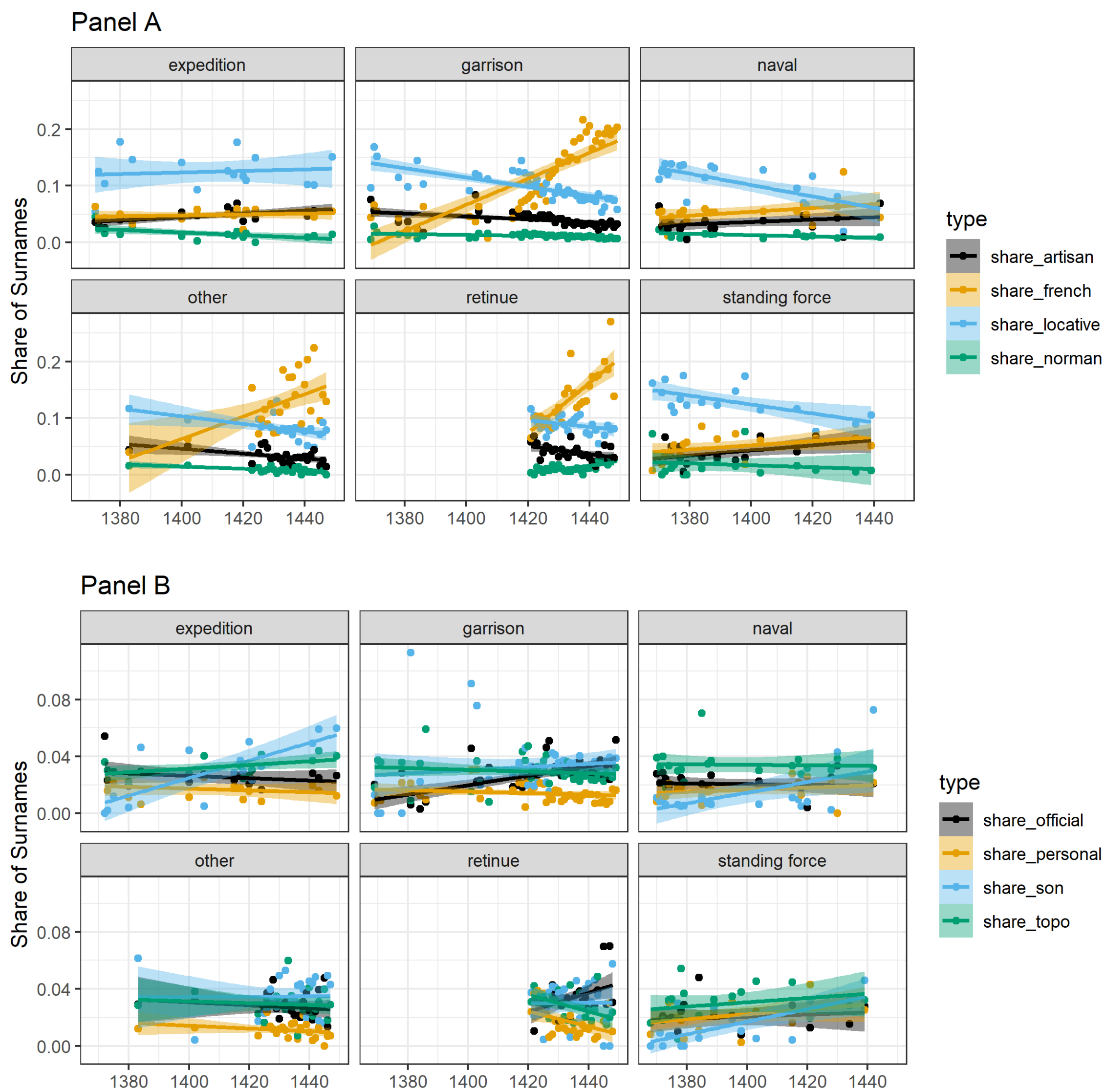

Figure 3.7: Share of surnames by service over time

activity, standing forces, yet maintain roughly the same share across expeditions. Note the dramatic increase in imputed French soldiers in garrison, retinue, and other (most often field service) activity. The same services breakdown using the placebo surnames is shown in Panel B. As previously there is no apparent trend outside of the increase in names with a -son patronym, most dramatically in expeditions.

Table 3.3: Locative surnames by service

	Dependent variable:				
	Share locative surnames				
	expedition	garrison	retinue	naval	other
Constant	12.922***	11.998***	13.462***	12.825***	11.565***
	(1.459)	(0.904)	(1.537)	(1.271)	(1.275)
time	−0.003	−0.052**	−0.071**	−0.083*	−0.049*
	(0.033)	(0.016)	(0.023)	(0.031)	(0.021)
Observations	18	49	30	23	34
R^2	0.0004	0.174	0.245	0.257	0.147
Residual Std. Error	3.600 (df = 16)	2.728 (df = 47)	1.856 (df = 28)	3.669 (df = 21)	2.803 (df = 32)
Note: All values in percentage points					*p<0.05; **p<0.01; ***p<0.001

Again fitting a simple regression model of surname share on year, table 3.3 shows that the decrease in share of locative surnames appears significant across all categories other than expeditions, which also have the least number of observations. The low number of observations across each category should make interpretation of these results cautious other than the general downward trend.

To sum up, surnames reveal important differences in social background. Elite coded surnames are associated with higher status positions in the English military and the share of these surnames is decreasing over time. Local Normans appear to have been important to the defense of Lancastrian Normandy, though whether these soldiers themselves were local elites is unknown. An increase in the share of soldiers with a -son patronym is also apparent. Given the relative patterns in figure 3.6, it seems these soldiers were of lower social background. This evidence supports the conclusions of historians that warfare during this was becoming more dependent upon individuals serving for economic reasons.

3.4.2 Regional Recruitment and Military Community

To investigate whether captains relied upon their own lands for recruitment or a growing centralized pool, I use tools from text analysis and population genetics to estimate the distance in surname space among the retinues of powerful magnates. I then test whether there is a correlation with geographic distance to their land holdings.

Historical evidence suggests lords at least sometimes recruited soldiers from their domains. For example, Edward III's son Edward the Black Prince, as earl of Chester, 'stimulated the participation of Cheshiremen in his service in Aquitaine' (Morgan, 1987). Likewise, the duchy of Lancaster estates were a prime recruiting ground for Henry IV during his Scottish campaign in 1400 (Curry et al 2010). Yet as mentioned in 3.2.3, there is also evidence that there were centralized recruitment focal points where men from around the country would gather. Overall, the judgement of Bell et al. (2013) is that during this period recruitment was becoming decoupled from formal lordship and local community networks.

A robust link between geographic distance and the distribution of surname frequencies has been found both regionally Scapoli et al. 2007, Cheshire et al. 2011) and within a number of countries (Rodriguez-Larralde et al. 1998,Rodriguez-Larralde et al. 2011,Barrai et al. 1996,Barrai et al. 1999,Barrai et al. 2002). The distribution of surnames within Britain has been examined in the modern day and historically (Sokal et al. 1992, Cheshire et al. 2010, Longley et al. 2011), with distinct regional surname clusters present. Most relevant, Parkin (2015) finds regional patterns of surname distribution in England using the poll tax returns of 1377, 1380, and 1381. Even without knowing the precise distribution of surnames, the fact that regionalism existed can be used to shed light on the question of the dynamism and distribution of the military community. The further apart the lands of magnates were the more different the surname distribution of their population should be, and *vice versa*. If magnates were largely dependent upon their own lands for soldiers, then the surname distributions of estates closer to each other should be more similar than those further away.

For this comparison some measure of similarity between different groups of surnames is needed. Research into regional surname structure uses metrics derived from population genetics, with surnames frequencies substituting for gene frequencies in a population. The two most popular are Lasker distance and Nei distance Cheshire (2014). Both are based on the coefficient of isonymy, defined between localities I and J as

$$I_{ij} = \sum_k p_{ki} p_{kj}$$

where p_{ki} and p_{kj} are the relative frequencies of surname k in groups I and J.

Lasker distance (Lasker, 1977) is a simple transformation of the coefficient of isonymy

$$L = -\log(I_{ij})$$

Nei distance (Nei, 1972) is related to the Lasker distance, but downweights the coefficient of isonymy between groups by the isonymy within each group

$$N_d = -\log(I_{ij}/\sqrt{(I_{ii} I_{jj})})$$

Each of these has slightly different assumptions about the underlying distribution of surnames. One caveat in using these is that they assume common surnames are monophyletic (have the same origin), when it is understood that many English occupational surnames are polyphyletic (multiple independent originations). For the purposes of this research, however, this distinction should have little impact on results.

I supplement these distance measures with a similarity metric used in latent semantic analysis. Cosine similarity is defined as

$$\cos(\theta) = \frac{I * J}{||I||||J||}$$

and is the the angle between two vectors in k-dimensional space, ranging from 0 (completely orthogonal) and 1 (parallel). Cosine similarity is often used to compare texts of different sizes, as the frequency of appearance of a token doesn't affect cosine similarity even as it affects the multidimensional vector length.

I use the *quanteda* package in R to calculate the relative frequency of each tokenized surname under a duke or earl as well as the cosine similarity, and then manually program the calculation for the Lasker and Nei distance. I remove any imputed French soldier names, as their increased presence in garrisons noted above would likely skew results.

To find the geographic distance, I use the Nussli (2011) map for the year 1400 and match the captains identified as dukes or earls in with their titled region (e.g. Edmund of Langely, duke of York, is matched with York). Using the *sf* package in R, I obtain the centroid of each locality and get the pairwise straight line distance with every other locality.

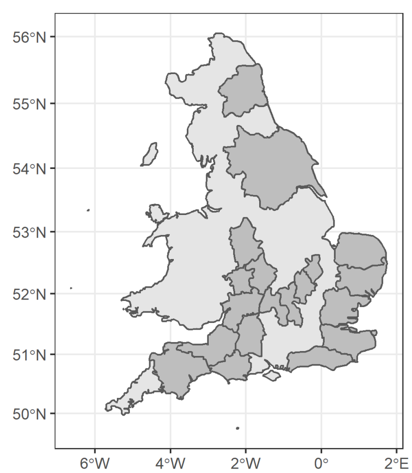

Figure 3.8: Magnate localities

Figure 3.8 shows the duchys and earldoms associated with magnate captains present in the data in dark grey. This covers most of the main broad regional divisions in England, and cuts across the surname regions found in Kandt et al. (2016). A list of the captains and their titled regions can be found in table C.1.

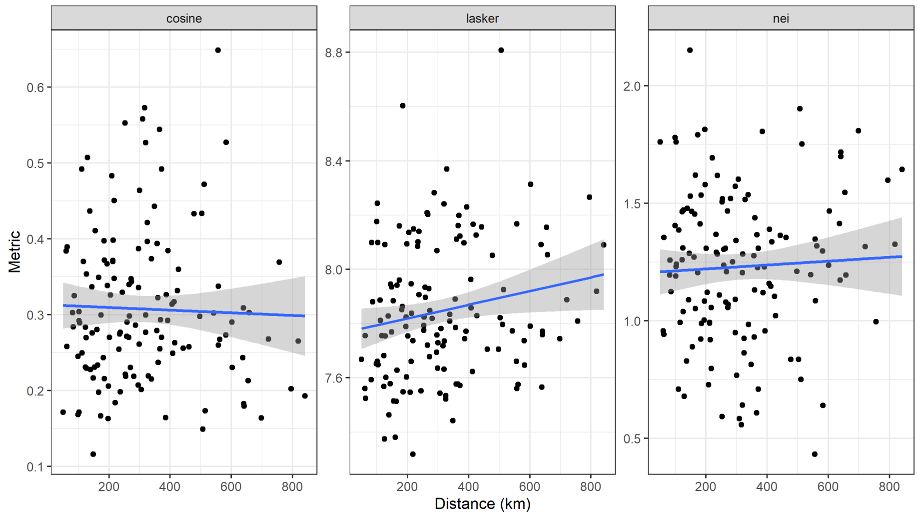

Figure 3.9: Surname distance vs geographic distance

Figure 3.9 shows the relationship between geographic distance and the different surname distance metrics. Cosine similarity is bounded between 0 to 1, in which pairs of vectors with more similar surname mixes will have higher values. In contrast, the Lasker and Nei distances will be larger the more dissimilar the surname mix between two vectors is. Each point represents the pairwise comparison of the surname pools of soldiers serving under a titled regional magnate, with the geographic distance between the two localities on the x-axis. One immediately visible limitation is the relative dearth of localities greater than 500 km apart.

A simple linear regression of each metric on geographic distance in table 3.4 shows no relationship for cosine similarity and Nei distance. The Lasker distance, however, does have a weak positive relationship, which would be consistent with a continuity of regional recruitment.

To get at the source of this possible link to regional recruitment, I separate the sample between men-at-arms and archers and redo the distance metric computation. Figure 3.10 shows there are clear differences in levels between the two groups: archers consistently have surname pools more similar to each other than do the men-at-arms. Rerunning the same

Table 3.4: Surname distance measures

	Dependent variable:		
	lasker	nei	cosine
	(1)	(2)	(3)
distance	0.0003*	0.0001	−0.00002
	(0.0001)	(0.0002)	(0.00005)
Observations	136	136	136
R^2	0.033	0.002	0.001
Residual Std. Error (df = 134)	0.251	0.323	0.102
Note:	*p<0.05; **p<0.01; ***p<0.001		

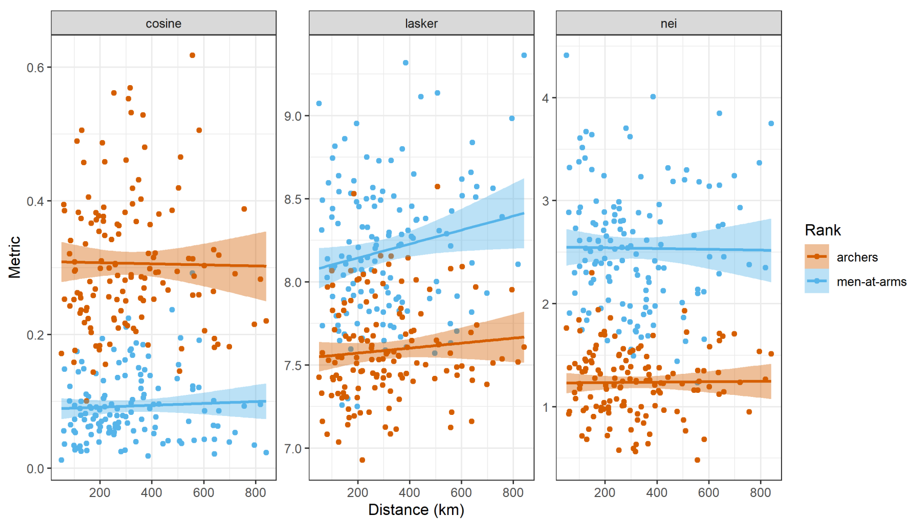

Figure 3.10: Surname distance vs geographic distance by rank

regression as above, table 3.5 shows that the connection between geographic distance and Lasker distance is moderated, and statistical significance for both groups is lost.

Table 3.5: Surname distance measures by rank

	Dependent variable:	
	lasker	
	Men-at-Arms	Archers
distance	0.0004	0.0002
	(0.0002)	(0.0001)
Observations	136	136
R^2	0.035	0.008
Residual Std. Error (df = 134)	0.407	0.299
Note:	*p<0.05; **p<0.01; ***p<0.001	

In general, there is no apparent connection between geographic distance and surname distance, which would be consistent with the idea that captains were less reliant on their regional lands for recruitment. A relatively flat distribution of surnames distances would be in line with Bell et al's (2013) suggestion that towns like London served as a recruitment market matching the supply and demand for military force. There are some hints in the Lasker distance that men-at-arms see more dissimilarity with distance, however the given relative sparsity of longer geographic distance comparisons this could just be noise.

One potential weakness of this approach is that some important duchys and earldoms were shuffled around to the king's supporters as management of the dominant coalition. For example, the Neville family had its origins in Durham in northern England, yet Richard Neville (1400–1460) replaced (via marriage) the House of Montagu as Earl of Salisbury with its lands in sourthern England. This would suggest that the improper geographic units had been used in finding physical distance. As a measure of robustness, I repeat the above exercise, this time matching magnates whose families had ancestral lands in particular regions (see table C.2).

Figure 3.11 shows that there may indeed be a link between geographic distance and surname

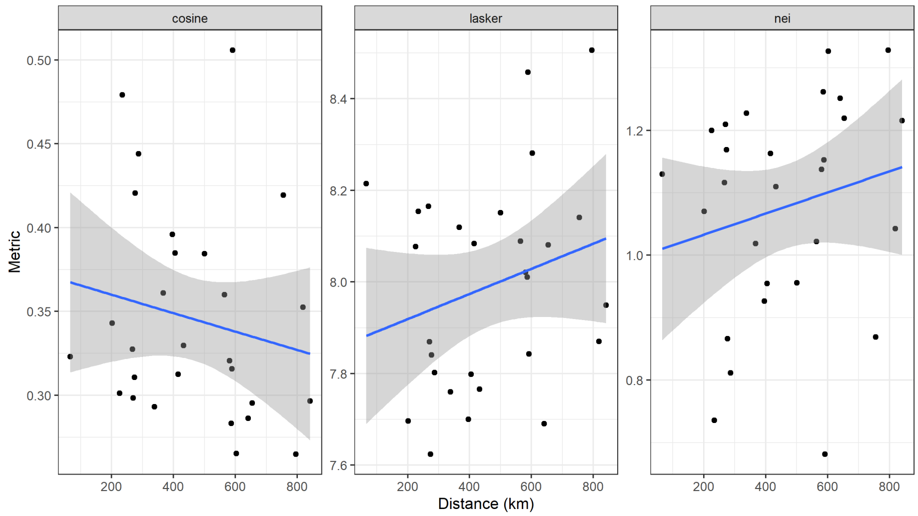

Figure 3.11: Surname distance by family

distance, as both the Lasker and Nei distances are positively related and the cosine similarity is negatively related. The sample here is much smaller however, and table 3.6 shows that these correlations are not significant. This is an area that deserves further study.

Table 3.6: Surname distance measures, families

	Dependent variable:		
	lasker	nei	cosine
	(1)	(2)	(3)
distance	0.0003	0.0002	−0.0001
	(0.0002)	(0.0002)	(0.0001)
Observations	28	28	28
R^2	0.062	0.042	0.033
Residual Std. Error (df = 26)	0.227	0.172	0.063
Note:	*p<0.05; **p<0.01; ***p<0.001		

The distance metrics also allow for a final comparison of soldiers using hierarchical agglomerative clustering. This is an unsupervised method that links each soldier pool to its closest neighbor in surname space. Pools that are more similar to one another than to others form clusters. These clusters are found using a pairwise distance metric between two vectors (in this case, Lasker distance). I follow other surname studies in using an unweighted average linkage clustering (UPGMA) algorithm, a technique that splits the difference between two other popular methods. This is performed using the default *stats* package in R.

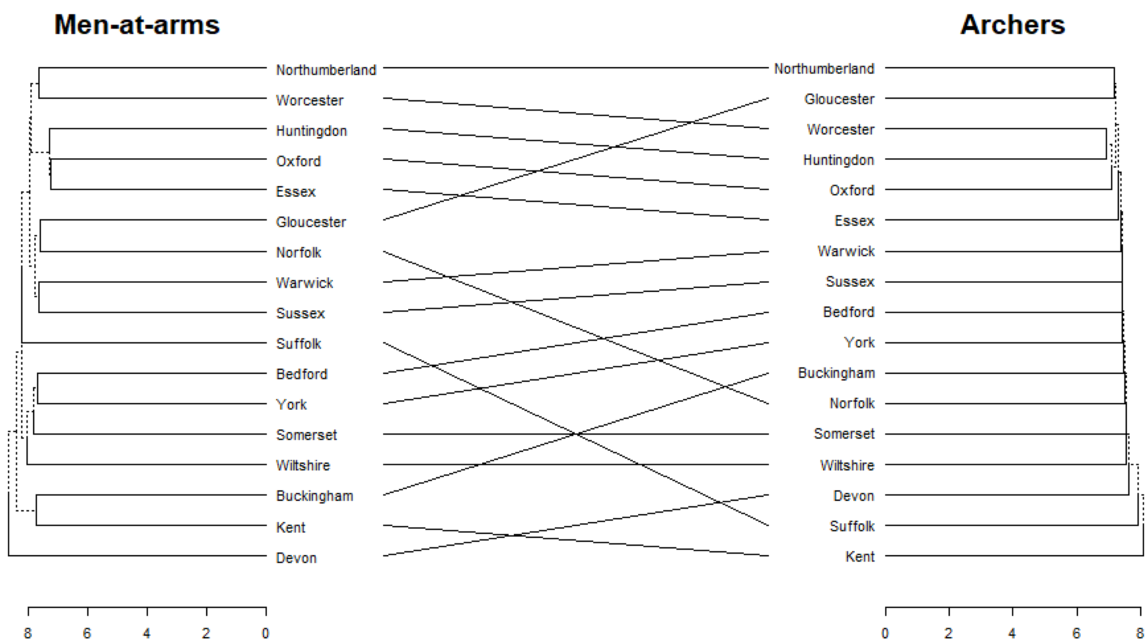

Figure 3.12: Surname clustering by rank

Figure 3.12 compares the results of hierarchical clustering using the *dendextend* package in R. The width of the horizontal line segment joining each branch to the larger tree represents the distance in surname space between elements. In the case of men-at-arms, the dendrogram shows that the soldiers employed by the earls of Northumberland and Worcester are more similar to one another than either is to the cluster containing Huntingdon, Oxford, and Essex. Men-at-arms show stronger clustering than archers, as evidenced by the slightly deeper tree. The pattern present for archers, with each group simply nested inside the next, is consistent with weak differentiation among elements. The lines connecting each dendrogram element also show that there is no common clusters between the two ranks of soldiers.

Altogether the evidence from surname distances suggests that while regional or social ties may have still played a small role in the recruitment of men-at-arms, for archers (and by extension, most soldiers), service with a magnate was likely more consistent with a common recruitment pool.

3.5 Conclusion

In this chapter I used evidence from the Soldier in Later Medieval England project to accomplish three tasks. First, I verified that certain surnames types are an effective proxy for social status. Second, I found that soldiers with elite backgrounds were decreasing in participation over time as the English crown shifted away from raids and towards garrisoning. Finally, I investigated the evidence for a national level military community that served as a common pool for recruitment. Initial evidence seems to support this historical hypothesis but further work is needed.

In principle, the analysis presented here can be built upon to incorporate more of the rich historical information available on many members of the peerage and upper gentry who, while not at the levels of dukes and earls, were still important players in the recuitment and management of retinues. Including the locations of these lands and soldiers will allow for greater clarity in measures of surname distance. Another major boon would be the incorporation of the Poll Tax records of 1377, 1379, and 1381, which would give needed verification to the link between geographic distance and the various surname distance metrics.

In contrast with the armies of the early 14th century, the men employed by the English crown in Lancastrian Normandy were increasingly composed of non-elites. These men appear to have come from all corners of England, and were likely to choose to serve as employment, not out of social obligation. The flexibility this indenture system allowed in organizing soldiers at scale also proved to be its undoing. Soon after being defeated in France, conflict erupted between the affinity groups of the large magnates in a struggle for control of the monarchy. This period of civil conflict, the Wars of the Roses, would last until 1485 and the rise of the Tudor dynasty. One of reform items was the banning of the retainer system, aside from the monarchy itself. This set the stage for the long run centralization of military power within the monarchy, and the eventual birth of the fiscal-military state in the late 17th century.

Appendix A: Printing Press

A.1 Figures and Tables

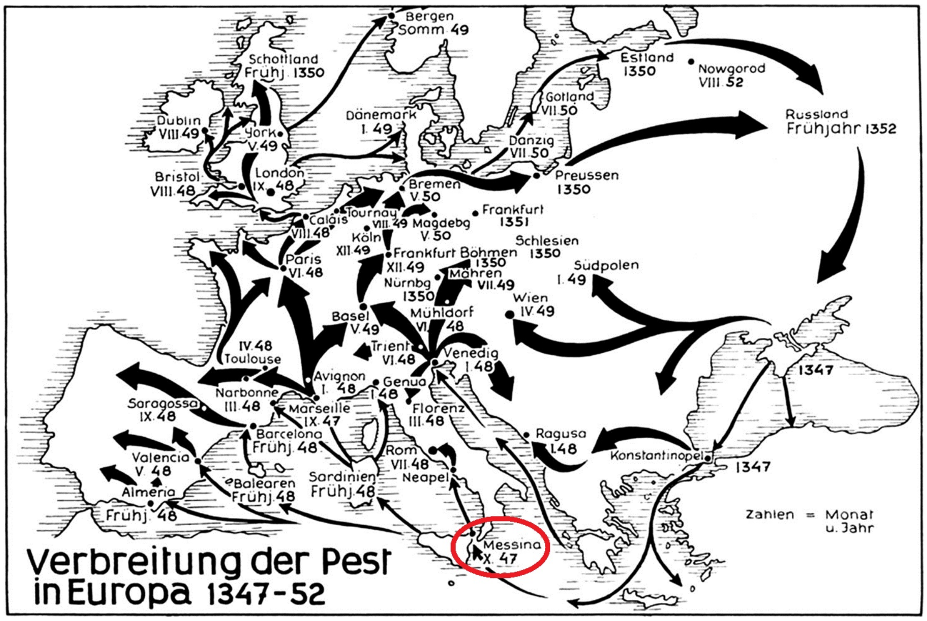

Figure A.1: Spread of the Black Death

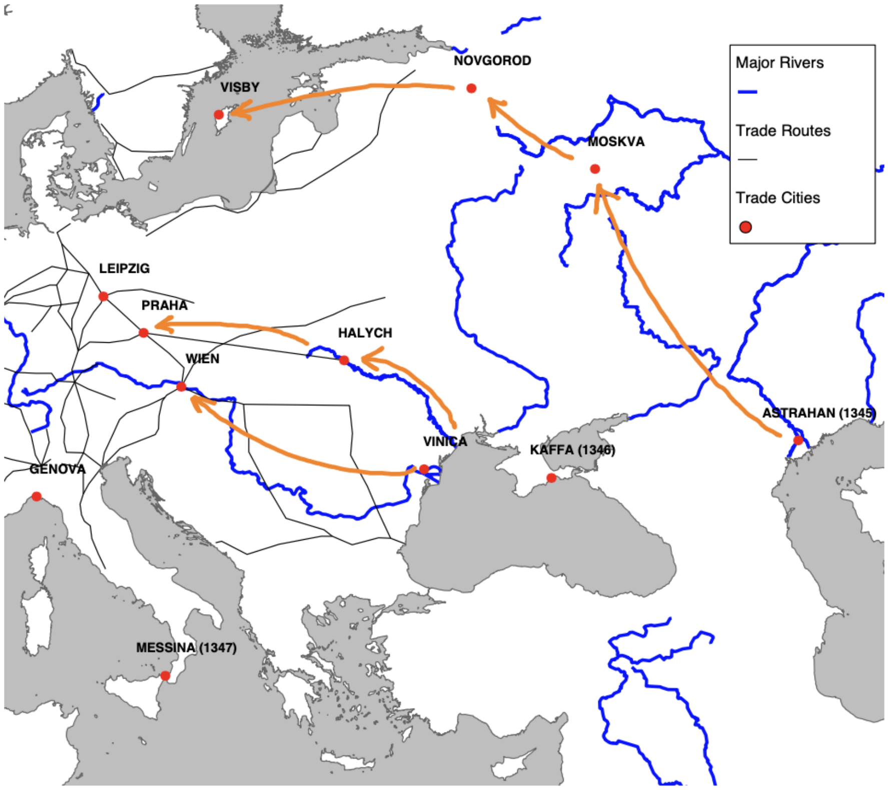

Figure A.2: Potential Routes of Entry for Black Death

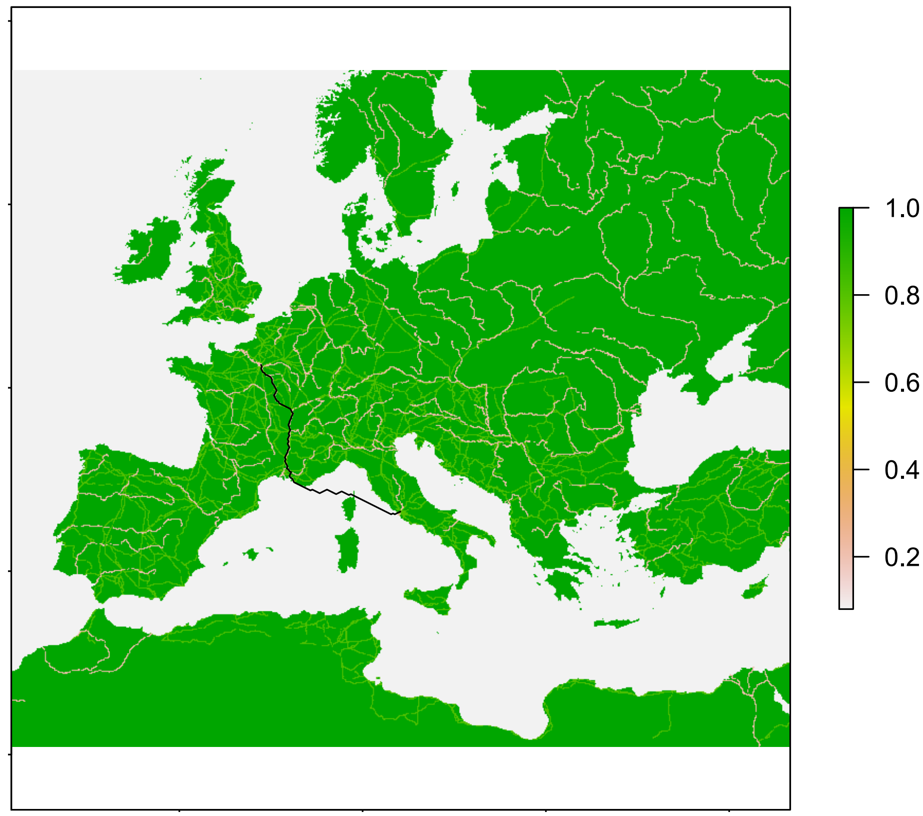

Figure A.3: Least Cost Travel Path from Paris to Rome

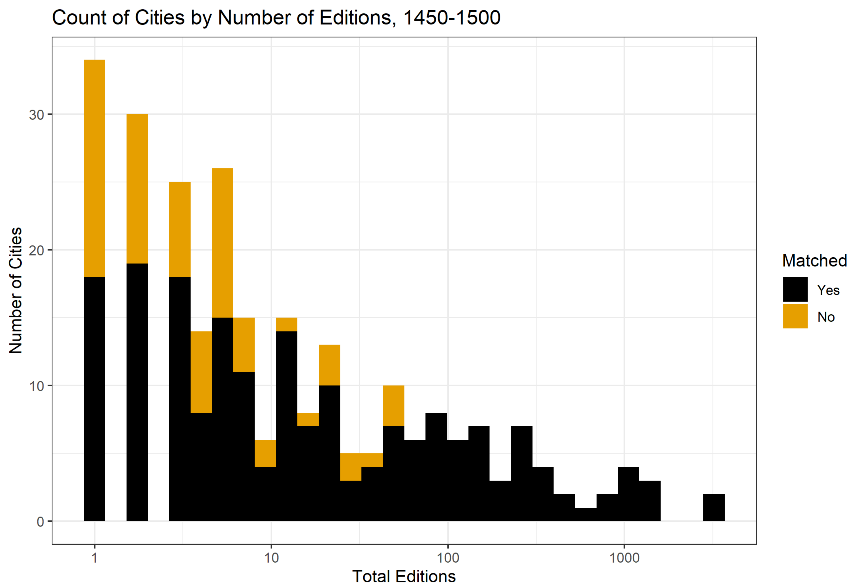

Figure A.4: USTC Editions Matched by City, 1450–1500

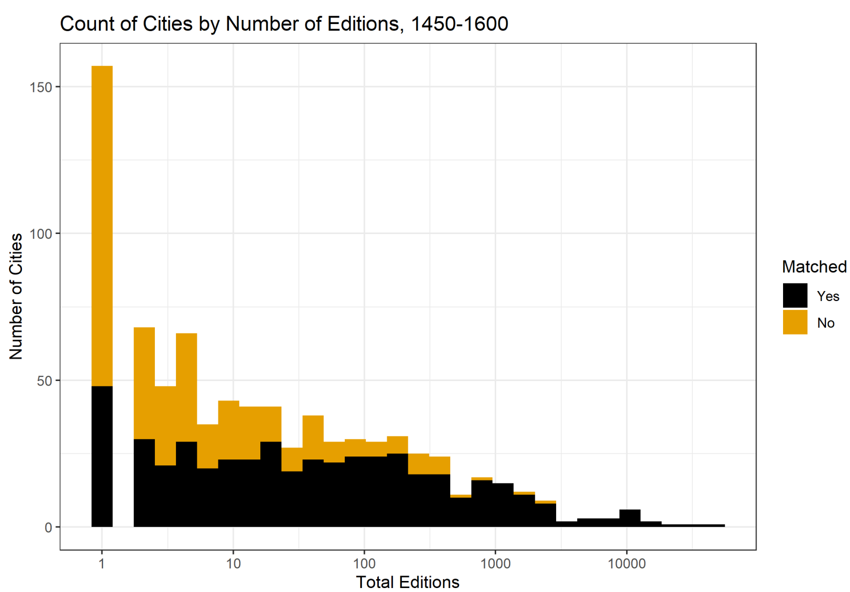

Figure A.5: USTC Editions Matched by City, 1450–1600

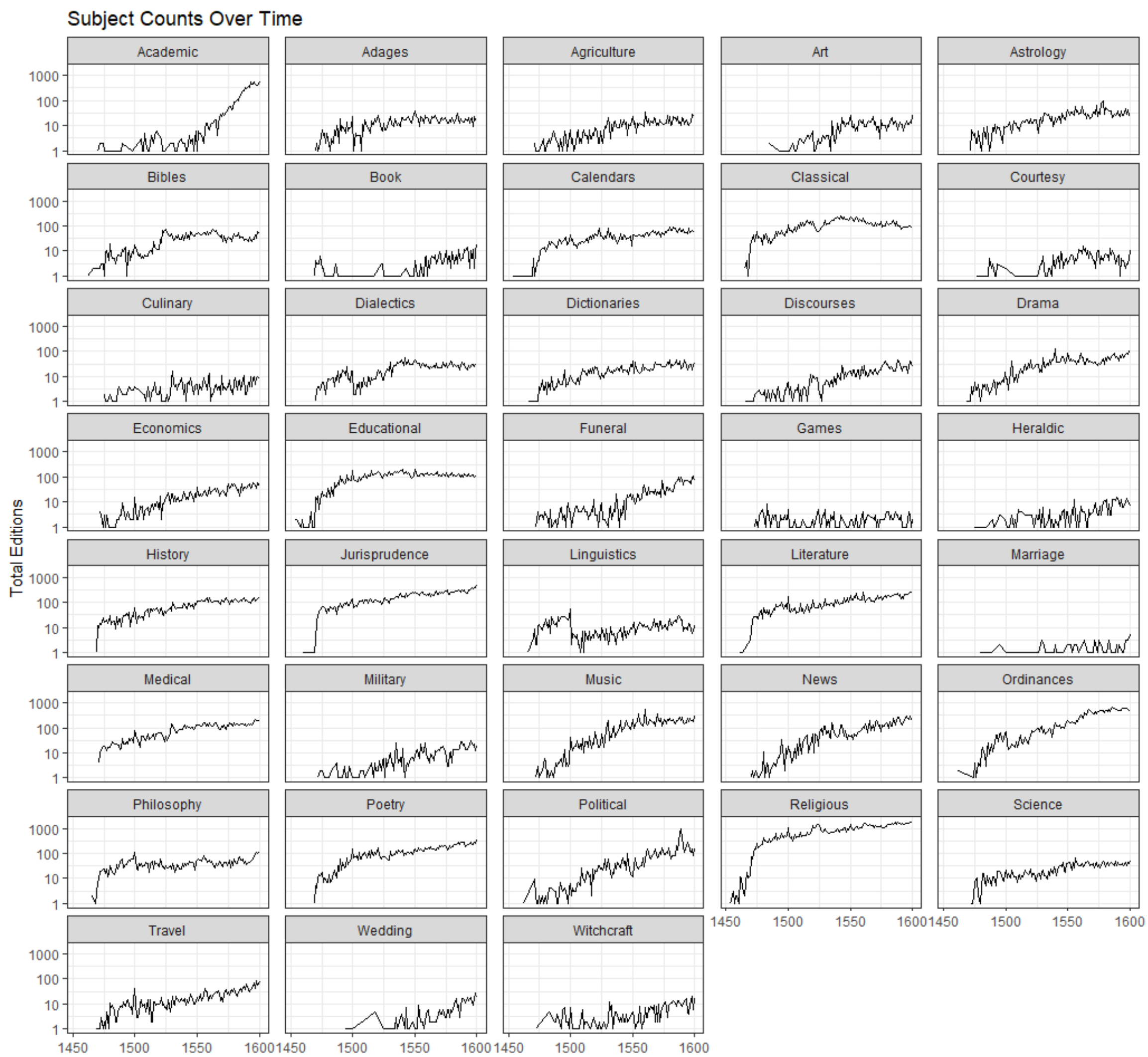

Figure A.6: USTC Subject Counts Over Time

Table A.1: Descriptive Statistics for Variables in Mortality Balance Table

	mean	sd	min	max
mortality	40.14	16.59	0.00	93.00
avtemp15001600	18.05	3.21	9.33	24.41
elevation	148.81	227.56	1.00	1999.00
cerealclosest	4.87	1.26	2.00	8.00
potatolowclosest	5.70	1.06	3.00	8.00
pastoral_weak_closest	0.90	0.30	0.00	1.00
dist2AMNB_10	0.23	0.42	0.00	1.00
rivers_10	0.32	0.47	0.00	1.00
longitude	4.65	6.25	-9.14	18.06
latitude	47.11	5.08	36.83	59.91
lpop1300	2.33	1.13	0.00	5.43
lma1300_cost1_38	-0.27	2.49	-4.78	5.97
DMajorRomRoad_10	0.65	0.48	0.00	1.00
DAnyRomRoad_10	0.72	0.45	0.00	1.00
DMajRomIntersection_10	0.36	0.48	0.00	1.00
DAnyRmRdIntersection_10	0.48	0.50	0.00	1.00
dist2landroute_10	0.35	0.48	0.00	1.00
dist2landrouteint_10	0.10	0.30	0.00	1.00
fair_all	0.21	0.41	0.00	1.00
HansaFixed	0.10	0.30	0.00	1.00
aqueduct_10	0.13	0.34	0.00	1.00
university	0.08	0.28	0.00	1.00
Bcapital	0.12	0.33	0.00	1.00
representative1300	0.27	0.44	0.00	1.00
parliament1300_yn	0.55	0.50	0.00	1.00
lDParliament	4.34	3.25	-11.83	6.74
dist2bat13001350_100	0.36	0.48	0.00	1.00
N	165			

Table A.2: Descriptive Statistics for Printing Press Adoption Regressions

	mean	sd	min	max
USTC_print_sample_1499	0.62	0.49	0.00	1.00
USTC_print_sample	1.00	0.00	1.00	1.00
mortality	38.79	15.97	5.00	80.00
growth13001400	-0.11	0.49	-1.75	1.79
lpop1400	2.68	0.95	0.69	5.62
bishop	0.64	0.48	0.00	1.00
university	0.12	0.32	0.00	1.00
lma1400_cost1_38	-0.43	2.32	-4.36	7.05
cereal25k	4.48	0.87	1.29	6.91
dist_Mainz	15.40	6.40	0.00	32.80
lshock_ma_1300_1353	-0.50	0.19	-1.07	-0.06
first_edition	1498.05	31.56	1453.00	1600.00
ustc_total_1500	2.36	2.41	0.00	8.07
ustc_total_1600	5.27	2.53	0.00	10.75
ustc_hhi_1600	0.48	0.24	0.17	1.00
print_sample	0.61	0.49	0.00	1.00
N	118			

Table A.3: Balance Table

	(1) mortality	(2) mortality	(3) mortality	(4) mortality
avtemp15001600	0.205			0.851
	(0.656)			(0.881)
elevation	0.00465			0.00893
	(0.00712)			(0.00801)
cerealclosest	0.926			2.384
	(1.559)			(1.801)
potatolowclosest	0.232			−1.430
	(1.893)			(2.110)
pastoral_weak_closest	1.089			0.683
	(4.235)			(4.528)
dist2AMNB_10	5.362*			3.068
	(3.169)			(4.171)
rivers_10	−4.677*			−6.037*
	(2.650)			(3.247)
longitude	−0.122			−0.0554
	(0.210)			(0.304)
latitude	−0.785*			−0.377
	(0.422)			(0.516)
lpop1300		−0.894		−2.133
		(1.377)		(1.938)
lma1300_cost1_38		0.570		0.480
		(0.631)		(0.729)
DMajorRomRoad_10		−5.001		−2.715
		(7.797)		(6.167)
DAnyRomRoad_10		8.059		4.832
		(8.401)		(6.720)
DMajRomIntersection_10		4.034		6.048
		(4.124)		(4.021)
DAnyRmRdIntersection_10		−1.403		−0.968
		(4.524)		(4.427)
dist2landroute_10		2.208		3.434
		(3.035)		(3.010)
dist2landrouteint_10		−5.047		−6.670
		(4.811)		(4.904)
fair_all		−5.657		−3.213
		(3.593)		(4.066)
HansaFixed		1.183		5.262
		(4.806)		(6.073)
aqueduct_10		1.864		−0.622
		(3.643)		(3.777)
university		6.722		5.742
		(4.186)		(4.345)
Bcapital			4.276	2.122
			(4.300)	(4.643)
representative1300			−5.292*	−0.376
			(3.155)	(3.832)
parliament1300_yn			3.040	0.916
			(2.673)	(3.452)
lDParliament			0.707	0.0676
			(0.450)	(0.449)
dist2bat13001350_100			−4.191	−2.848
			(2.614)	(2.762)
N	169	165	169	165
adj. R^2	0.099	0.009	0.036	0.086

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Control Variables and Other Variables

Average Temperature 1500-1600. We use temperature data from Luterbacher et al. (2004). They reconstruct seasonal European temperatures (celsius degrees) since 1500 using proxy data from ice cores, tree rings, and written records. The data cover 0.5° x 0.5° grids which is approximately 50km x 50km at European latitudes. The data extend from 25° W to 40° E and 35° N to 70° N which includes all of the cities in our full sample. We extract the growing season (summer) temperature for each of our cities during the 16th century as this is the closest century to the Black Death period for which we have data. No comparable data exist for earlier centuries.

Elevation. City elevation data come from Jarvis et al. (2008) which is available at `http://srtm.csi.cgiar.org`. These data report elevation in meters with a spatial resolution between 1 and 3 arc-seconds. Where there are missing data we have supplemented it using Wikipedia.

Cereal Suitability. Our soil suitability data are from the FAO Global Agro-Ecological Zones (GAEZ) dataset as described in Fischer et al. (2002). We use these in preference to Ramankutty et al. (2002) as the latter does not have full coverage for all of western Europe. We use the GAEZ's cereal suitability data assuming low inputs and rain-fed irrigation. We extract the average soil suitability within 10 km radius circles around each city. Overall, cereal suitability is scaled from 1-9 where 1 is best, 8 is unsuitable and 9 is water (seas and oceans are treated as missing values). In some regressions, we invert the measure so that positive values are associated with higher cereal suitability.

Potato Suitability. The potato suitabliity numbers are constructed using the Global Agro-Ecological Zones (GAEZ) data. We specifically use the data on white potatoes grown under conditions using low inputs and rain-fed irrigation for the baseline period 1961-1990. The raster file for the data along with support documentation are available for download from: `http://www.fao.org/nr/gaez/newsevents/detail/en/c/141573/`. These data are constructed in two stages. First the Food and Agriculture Organization (FAO) compiles information on the nutrients, soil, irrigation, and climatic conditions under which the potato grows best. Then the FAO compiles data on the physical environment for the entire world at a resolution of 5 arc minutes x 5 arc minutes ($\approx$ 10 x 10 Km). These characteristics include soil type, slope, average water availability, humidity, temperature, wind speed, etc. Then these two types of data are combined in order to create a value for "potential suitability for potato cultivation" for each raster cell. These values run from 1 to 9, where 1 is most suitable, 8 is least suitable, and 9 is water (or impossible to cultivate). See Monteduro (2012) for more details on the construction of the suitability raster. We use the GAEZ data to construct our city-level measures of potato suitability by extracting the average value of the raster cells within a 10 km radius of each city.

Pastoral Suitability. We control for the potential suitability of a region surrounding a city for pastoral farming with a variable measuring grazing suitability. This variable come from Erb et al. (2007) who create land use measures at a resolution of 5 arc minute cells ($\approx$ 10 km X 10 km). They record how land is used in each cell in 2000. The five categories they code for are: cropland, grazing, forest, urban, and areas without land use. Their grazing category is calculated as a residual after accounting for the percentage of area taken up by the other four uses. As part of this analysis they also generate a variable measuring the suitability of each cell for grazing (as opposed to actual present-day use). The suitability measure is created by first separating grazing land into three categories based on cover: 'high suitability of cultivated and managed areas, medium suitability of grazing land found under tree cover, and low suitability if shrub cover or sparse vegetation is detected in remote sensing' (Erb et al., 2007, 199). They then further subdivide the first two of these categories into areas with a net primary productivity of Carbon per meter squared is greater than 200 grams and those in which it is less than 200 grams. This results in five categories which they regroup into four categories with 1 = most suitable and 4 = least suitable. There is a fifth category which is 'no grazing' which we re-code as 5. We then create a dummy equal to 1 if the cell is most or moderately suitable. Finally, we extract the average suitability of the region around a city for grazing using circles of 10 km's.

Coasts and Rivers. We create variables to measure distances to the coast and major rivers using ArcGIS. We base these distances on the 1300 shape file downloaded from Nussli (2011). We then create two dummies for whether each city is within 10 km from the coast or a river.

Market Access. Market access for city i in 1300/1353 is defined as $MA_i = \Sigma_j \frac{P_j}{D_{ij}^{\sigma}}$, with P_j being the population of the other 1801 - 1 = 1800 cities $j \neq i$ in 1300/1353, D_{ij} the travel time between city i and city j, and $\sigma = 3.8$. To obtain the travel times, we use the gdistance package in R to compute the least cost travel paths via four transportation modes — by sea, by river, by road and by walk — with the transportation speeds from Boerner and Severgnini (2014).[1] More on the gdistance package is available for download at this link: `https://cran.r-project.org/web/packages/gdistance/vignettes/gdistance1.pdf`. The predicted population of each city in the aftermath of the Black Death (1353) is constructed as = pop. in 1300 x (100-mortality)/100.

Roman Roads. Data on Roman roads is provided by the *Digital Atlas of Roman and Medieval Civilizations*. We use this shape file to create two distances: (1) distance to all Roman roads and (2) distance to 'major' Roman roads. Since major settlements often formed along intersections of the road network, we also create variables for distances to Roman road intersection. We then create four dummies if the city is within 10 from any Roman road, a major Roman road, any Roman road intersection, or a major Roman road intersection.

Medieval Trade Routes. We use Shepherd (1923) to create a map of major medieval land trade routes. We create a GIS file that allows us to measure the distance to major medieval land trade routes or the intersection of two of them. We then create dummy variables that take the value of 1 if a city is within 10 kilometers of a trade route or an intersection of two of them.

[1] Normalizing the speed to porters to 1, this assigns a travel cost of 0.5 to roads and rivers and 0.18 to seas.

Market Fairs. We obtain data on the location of important medieval fairs from two sources. The main source is Shepherd (1923). The second source we use is the *Digital Atlas of Roman and Medieval Civilizations*. The original source for this information is: Ditchburn, David and MacLean, Simon (eds.) 2007, Atlas of Medieval Europe, 2nd edn, London and New York, p. 158. We drop fairs that they cannot be matched with cities in the Bairoch dataset.

Hanseatic League. We document whether or not a city was a member of the Hanseatic League. We do this by matching where possible the city data with available lists of cities which belonged to the League. We include only cities which were members of the League and do not include cities with Hansa trading posts or communities. The source we use is from Dollinger (1970) and is the most comprehensive list of Hanseatic cities available. Unfortunately, Dollinger does not provide details on when each city became a member of the Hanseatic league. However Wikipedia provides information on a subset of Hanseatic cities. Using this data, we estimate that approximately 75% of these cities were likely members of the league prior to the time of the Black Death, thus giving us confidence that our Hansa dummy mostly captures pre-plague conditions.[2]

Aqueducts. We use GIS to create a shape file for whether or not a town was within 10 km from a Roman aqueduct using the map provided by Talbert (2000) as well as information from two Wikipedia webpages: `https://en.wikipedia.org/wiki/List_of_aqueducts_in_the_Roman_Empire` and `https://fr.wikipedia.org/wiki/Liste_des_aqueducs_romains`.

Medieval Universities. Bosker et al. (2013) provides data on the presence of medieval universities for European cities with populations greater than 10,000 (at some point between 800 and 1800). We consulted Wikipedia and other sources to find evidence of medieval universities with smaller populations. There are five medieval universities missing from the list in Bosker et al. (2013): Angers, Greifswald, Ingolstadt, Tuebingen, and Uppsala. However, as none of these were established prior to the Black Death we do not include them

[2] Data available on request.

in our analysis.

State Capital in 1300. We use the data provided by Bosker et al. (2013) who collect data on capital cities from McEvedy and Jones (1978).

Representative Body in 1300. Bosker et al. (2013) provide information on the existence of communes for a subset of the cities in the Bairoch dataset. Bosker et al. (2013) create a variable"commune" that takes a value of 1 if there is indication of the presence of a local urban participative organization that decided on local urban affairs. Stasavage (2014) provides data on 169 cities that were autonomous at some point between 1000 and 1800. We use the variable for 1300-1400. Stasavage (2014) defines autonomous cities in the following terms:

> 'I have defined an "autonomous city" as being one in which there is clear evidence that such institutions of self-governance existed, and in addition there is also clear evidence of exercise of prerogatives in at least one of the policy areas referred to above. In the absence of such evidence the default is to code a city as non-autonomous (6).'

As Stasavage (2014) notes, his definition of city autonomy is stricter than the definition of commune used by Bosker et al. (2013). We create a dummy equal to one if the city is a commune in the Bosker et al. (2013) data set or a self-governing city according to Stasavage (2014).

Parliamentary Activity and Distance to Parliament 1300-1400. Our data on parliamentary activity is from van Zanden et al. (2012). This measures the number of times that Parliaments met at a regional level in 1300–1400. We create a dummy variable based on whether or not a town is in a region/country which had above the median number of parliamentary meetings. We also obtain a list of whether the parliaments were held for each region/country. We then use GIS to compute for each city the minimal Euclidean distance to a parliament.

Battles. As our main source we use Wikipedia's list of all battles that took place between 1300 and 1600. `https://en.wikipedia.org/wiki/List_of_battles_1301-1800`. This is a highly reliable source for the most important battles of the period. We are not concerned about sample selection here as Wikipedia's coverage of European history is extensive; battles not listed on Wikipedia are likely to have been extremely small. For each battle we assign geo-coordinates based on either the location of the battle or the location of the nearest city mentioned in the entry. Note that we exclude naval battles.

Appendix B: Agricultural Productivity in Ethiopia

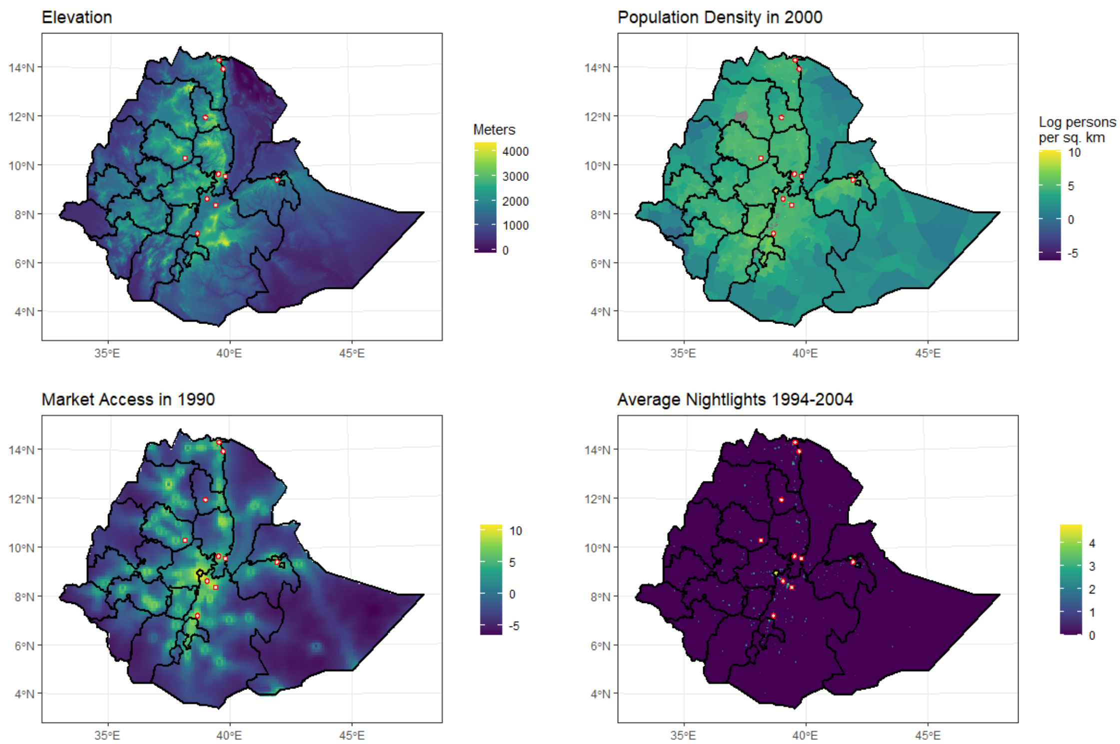

Figure B.1: GIS Controls

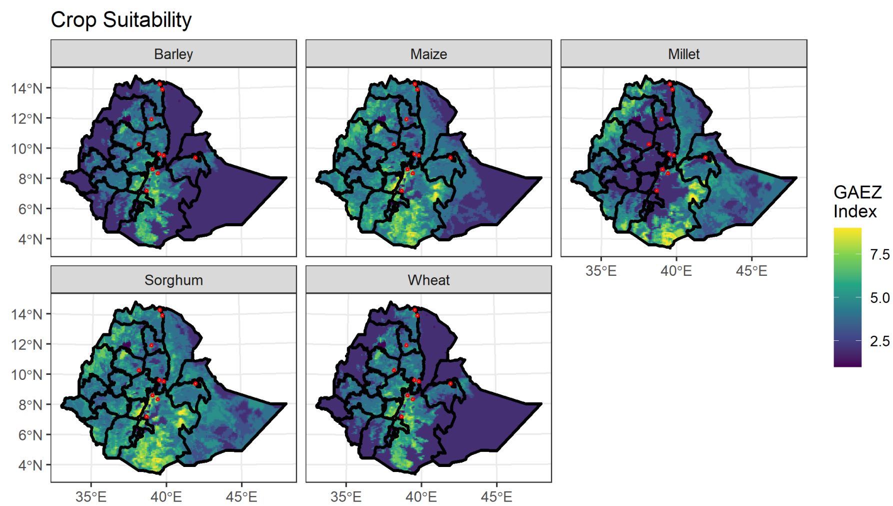

Figure B.2: Soil Suitability

Note: GAEZ index reverse coded such that higher values are more suitable

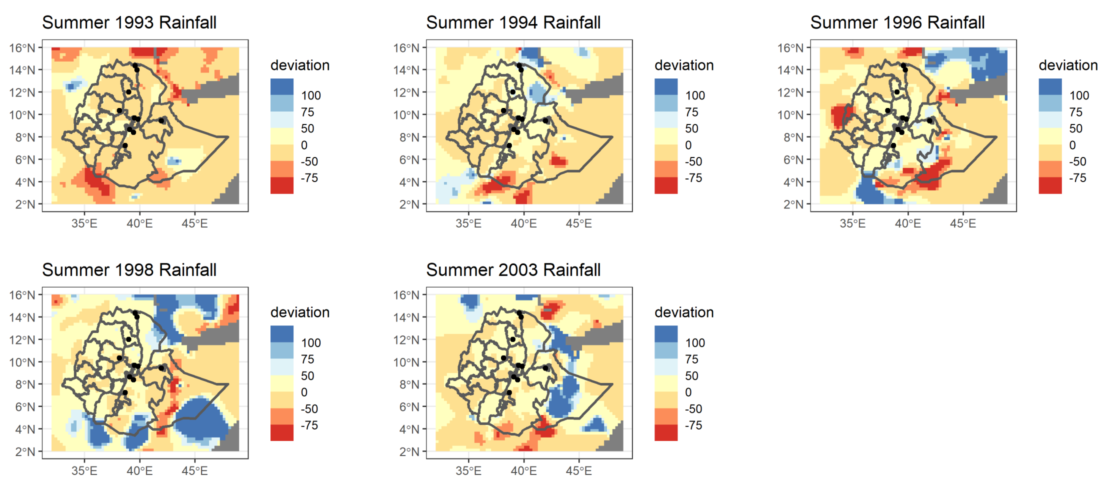

Figure B.3: Spatial variation in rainfall

Appendix C: Soldiers Appendix

Locative Suffixes: ton(n)(e), tun(n)(e), don(n)(e), ham(m)(e), land(d)(e), bur(r)(ie/y), ber(r)(ie/y)

Base for Norman names: Baignard, Belcamp, Beauchamp, Beners, Burdet, Busly, Cailly, Caron, Colville, Corbet, Albamarla, Damerel, Darcy, Courcy, Vere, Giffard, Lacy, Malet, Mandeville, Montgomery, Molbrai, Neville, Percy, Pomeroy, Sackville, Sai, Sancto Claro, Taillebois, Tournai, Venbales, Villare, Fitz

Base for artisan names:

- Building trades: Smith/Smythe, Ferrour, Carpenter, Wright, Faber, Mason, Thatcher, Plumber, Glazier, Painter, Sawyer, Slater, Tyler
- Farming trades: Carter, Shepherd, Coward, Plowman, Thresher
- Textile trades: Taylor, Webber, Weaver, Walker, Fuller, Barker, Tanner, Lister, Dyer, Dexter, Skinner, Glover
- Food trades: Baker, Baxter, Butcher, Cook, Brewer, Salter, Miller, Milner, Spicer

Base for official occupational names: Clerk, Chamberlain, Butler/Botiler, Marshall, Parker, Porter, Bailly, Spencer, Steward, Reve

Base for personal names: Brown, Red, White, Hore, Grey

Base for topographic names: Grene, More, Wode, Brome, Mede, Bois, Broke

Table C.1: Magnate Regions

Region	Captain	N
Bedford	John of Lancaster (1389 - 1435) duke of Bedford	2857
Buckingham	Stafford, Humphrey (1402 - 1460) duke of Buckingham	596
Devon	Courtenay, Edward ‘the blind earl‘ (c. 1357 - 1419) earl of Devon	662
Devon	Courtenay, Hugh (1389 - 1422) earl of Devon	724
Dorset	Beaufort, Edmund (1406 - 1455) count of Mortain, earl of Dorset, duke of Somerset	2244
Dorset	Beaufort, John (1371 - 1410) Earl of Somerset, Marquess of Dorset	125
Dorset	Beaufort, Thomas (c. 1377 - 1426) earl of Dorset, duke of Exeter	1479
Essex	Bourchier, Henry (1408 - 1483) count of Eu, earl of Essex	378
Gloucester	Humphrey of Lancaster (1390 - 1447) duke of Gloucester	1443
Gloucester	Thomas of Woodstock (d. 1397) duke of Gloucester	140
Huntingdon	Holland, John (1395 - 1447) earl of Huntingdon, duke of Exeter	1159
Kent	Grey, Edmund (1416 - 1489) Lord Grey of Ruthin, earl of Kent	198
Kent	Neville, William (1405 - 1463) Lord Fauconberg, earl of Kent	3677
Norfolk	Mowbray, John (1392 - 1432) earl-marshal, duke of Norfolk	662
Norfolk	Mowbray, Thomas (1366 - 1399) duke of Norfolk	1430
Northumberland	Percy, Henry (1393 - 1455) earl of Northumberland	369
Northumberland	Percy, Henry (d. 1408) earl of Northumberland	473
Oxford	Vere, John de (1408 - 1462) earl of Oxford	294
Oxford	Vere, Richard de (1385 - 1417) earl of Oxford	219
Somerset	Beaufort, bastard of the duke of Somerset	45
Somerset	Beaufort, Edmund (1406 - 1455) count of Mortain, earl of Dorset, duke of Somerset	2244
Somerset	Beaufort, John (1371 - 1410) Earl of Somerset, Marquess of Dorset	125
Somerset	Beaufort, John (1403 - 1444) duke of Somerset	6107
Somerset	Beaufort, Thomas (c. 1377 - 1426) earl of Dorset, duke of Exeter	1479
Suffolk	Pole, Michael de la (1367/ - 1415) Earl of Suffolk	329
Suffolk	Pole, Michael de la (1395 - 1415) earl of Suffolk	14
Suffolk	Pole, Michael de la (c. 1330 - 1389) earl of Suffolk	277
Suffolk	Pole, William de la (1396 - 1450) duke of Suffolk	1982
Sussex	FitzAlan, John (1408 - 1435) earl of Arundel, Lord Maltravers	1090
Sussex	FitzAlan, Richard (1346 - 1397) earl of Arundel	1249
Sussex	FitzAlan, Thomas (1381 - 1415) earl of Arundel	549
Warwick	Beauchamp, Richard (1382 - 1439) earl of Warwick	2676
Warwick	Warwick, Thomas de Beauchamp, earl of	426
Wiltshire	Montagu, John, Sir bastard of Salisbury	437
Wiltshire	Montagu, Thomas (1388 - 1428) earl of Salisbury	1259
Wiltshire	Montagu, William (1328 - 1397) earl of Salisbury	246
Wiltshire	Neville, Richard (1400 - 1460) earl of Salisbury	375
Worcester	Beauchamp of Bergavenny, Richard (d. 1422) earl of Worcester, Lord Bergavenny	366
Worcester	Percy, Thomas (1343 - 1403) Earl of Worcester	776
York	York, Edmund of Langley (1341 - 1402) duke of York	419
York	York, Edward of (1373 - 1415) duke of York	522
York	York, Richard of (1411 - 1460) duke of York	2955

Table C.2: Magnate Families

Captain	Family	Region	N
Courtenay, Edward ‘the blind earl‘ (c. 1357 - 1419) earl of Devon	Courtenay	Devon	635
Courtenay, Edward, Sir (1388 - 1418)	Courtenay	Devon	116
Courtenay, Hugh (1389 - 1422) earl of Devon	Courtenay	Devon	664
Courtenay, Philip de, Sir	Courtenay	Devon	210
Courtenay, Richard (d. 1415) bishop of Norwich	Courtenay	Devon	5
Courtenay, William	Courtenay	Devon	6
Neville, John, Sir	Neville	Durham	200
Neville, John, Sir (d. 1420)	Neville	Durham	160
Neville, Ralph, Sir	Neville	Durham	48
Neville, Richard (1400 - 1460) earl of Salisbury	Neville	Durham	337
Neville, Robert	Neville	Durham	5
Neville, Thomas (1377 - 1407) Lord Furnivall	Neville	Durham	480
Neville, William (1405 - 1463) Lord Fauconberg, earl of Kent	Neville	Durham	2741
Neville, William (1405 - 1463) Lord Fauconberg, earl of Kent et al.	Neville	Durham	178
Neville, William de, Sir	Neville	Durham	788
Humphrey of Lancaster (1390 - 1447) duke of Gloucester	Lancaster	Lancaster	1313
John of Lancaster (1389 - 1435) duke of Bedford	Lancaster	Lancaster	2419
John of Lancaster (1389 - 1435) duke of Bedford et al.	Lancaster	Lancaster	7
Thomas of Lancaster (1387 - 1421) duke of Clarence	Lancaster	Lancaster	965
Mowbray, John (1392 - 1432) earl-marshal, duke of Norfolk	Mowbray	Lincoln	618
Mowbray, Thomas (1366 - 1399) duke of Norfolk	Mowbray	Lincoln	1307
Percy, Henry (1393 - 1455) earl of Northumberland	Percy	Northumberland	360
Percy, Henry (d. 1408) earl of Northumberland	Percy	Northumberland	456
Percy, Henry ‘Hotspur‘ (d. 1403)	Percy	Northumberland	88
Percy, Thomas (1343 - 1403) Earl of Worcester	Percy	Northumberland	658
Percy, Thomas, junior, Sir	Percy	Northumberland	111
Percy, Thomas, Sir	Percy	Northumberland	3
Vere, Aubrey de, Sir	Vere	Oxford	244
Vere, John de (1408 - 1462) earl of Oxford	Vere	Oxford	279
Vere, Richard de (1385 - 1417) earl of Oxford	Vere	Oxford	209
Vere, Robert de, Sir (c. 1410 - 1461)	Vere	Oxford	42
Pole, Edmund de la, Sir	Delapole	Suffolk	2
Pole, John de la, Sir	Delapole	Suffolk	296
Pole, Michael de la (1367/ - 1415) Earl of Suffolk	Delapole	Suffolk	300
Pole, Michael de la (1395 - 1415) earl of Suffolk	Delapole	Suffolk	13
Pole, Michael de la (c. 1330 - 1389) earl of Suffolk	Delapole	Suffolk	194
Pole, William de la (1396 - 1450) duke of Suffolk	Delapole	Suffolk	1609
FitzAlan, John (1385 - 1421) Lord Maltravers	Fitzalan	Sussex	174
FitzAlan, John (1408 - 1435) earl of Arundel, Lord Maltravers	Fitzalan	Sussex	888
FitzAlan, Richard (1346 - 1397) earl of Arundel	Fitzalan	Sussex	1180
FitzAlan, Thomas (1381 - 1415) earl of Arundel	Fitzalan	Sussex	522

Bibliography

Acemoglu, Daron (2002), 'Directed Technical Change', *Review of Economic Studies* **69**(4), 781–809.

Acemoglu, Daron and James A. Robinson (2012), *Why nations fail: the origins of power, prosperity and poverty*, 1st ed edn, Crown Publishers, New York.

Acemoglu, Daron and Joshua Linn (2004), 'Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry', *The Quarterly Journal of Economics* **119**(3), 1049–1090.

Acemoglu, Daron and Simon Johnson (2005), 'Unbundling Institutions', *Journal of Political Economy* **113**(5), 949–995.

Acemoglu, Daron, Simon Johnson and James A. Robinson (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review* **91**(5), 1369–1401.

Acemoglu, Daron, Tristan Reed and James A. Robinson (2014), 'Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone', *Journal of Political Economy* **122**(2), 319–368.

Adamopoulos, Tasso and Diego Restuccia (2020), 'Land Reform and Productivity: A Quantitative Analysis with Micro Data', *American Economic Journal: Macroeconomics* **12**(3), 1–39.

Allen, Robert C (2003), 'Progress and poverty in early modern europe', *The Economic History Review* **56**(3), 403–443.

Allen, Robert C (2011), ‘Why the industrial revolution was british: commerce, induced invention, and the scientific revolution 1’, *The Economic History Review* **64**(2), 357–384.

Allmand, C. T. (2001), *The hundred years war: England and France at war, c. 1300-c. 1450 (revised)*, Cambridge medieval textbooks, Cambridge University Press, Cambridge ; New York.

Anderson, Lars and Jeanet Bentzen (2022), In the Name of God! Religiosity and the Transition to Modern Growth, Technical Report DP16938, Center for Economic Policy Research.

Arnot, C. D., M. K. Luckert and P. C. Boxall (2011), ‘What Is Tenure Security? Conceptual Implications for Empirical Analysis’, *Land Economics* **87**(2), 297–311.

Ayton, Andrew (1999), *Knights and warhorses: Military service and the English aristocracy under Edward III*, Boydell & Brewer.

Ayton, Andrew (2009), Armies and Military Communities in Fourteenth-Century England, *in* P.Coss and C.Tyerman, eds, ‘Soldiers, Nobles and Gentlemen’, ned - new edition edn, Essays in Honour of Maurice Keen, Boydell & Brewer, pp. 215–239.

Bai, Ying, Ruixue Jia and Jiaojiao Yang (2021), The Nexus of Elites and War Mobilization, Technical Report w28667, National Bureau of Economic Research.

Bairoch, Paul (1988), *Cites and Economic Development, from the dawn of history to the present*, University of Chicago Press, Chicago. translated by Christopher Braider.

Barbier, Frédéric (2017), *Gutenberg’s Europe: The book and the invention of western modernity*, John Wiley & Sons.

Barrai, Italo, A. Rodriguez-Larralde, E. Mamolini and C. Scapoli (1999), ‘Isonymy and isolation by distance in Italy’, *Human biology* pp. 947–961. ISBN: 0018-7143 Publisher: JSTOR.

Barrai, Italo, Alvaro Rodriguez-Larralde, Franz Manni and Chiara Scapoli (2002), 'Isonymy and isolation by distance in the Netherlands', *Human Biology* pp. 263–283. ISBN: 0018-7143 Publisher: JSTOR.

Barrai, Italo, Chiara Scapoli, María Beretta, Carlo Nesti, Elisabetta Mamolini and A. Rodriguez-Larralde (1996), 'Isonymy and the genetic structure of Switzerland I. The distributions of surnames', *Annals of human biology* **23**(6), 431–455. ISBN: 0301-4460 Publisher: Taylor & Francis.

Bell, Adrian R., Anne Curry, Andy King and David Simpkin (2013), *The soldier in later medieval England*, Oxford University Press, Oxford ; New York, NY.

Bell, Adrian R., Anne Curry, Andy King, David Simpkin and Adam Chapman, eds (2011), *The soldier experience in the fourteenth century*, Warfare in history, Boydell Press, Woodbridge, Suffolk, UK ; New York.

Bellemare, Marc F., Kenn Chua, Julieth Santamaria and Khoa Vu (2018), Tenurial security and agricultural investment: Evidence from Vietnam, Working Paper 2018/76, WIDER Working Paper.

Benedictow, Ole J. (2005), *The Black Death 1346–1353: The Complete History*, The Boydell Press, Woodbridge.

Benin, S. and J. Pender (2001), 'Impacts of land redistribution on land management and productivity in the Ethiopian highlands', *Land Degradation & Development* **12**(6), 555–568.

Besley, Timothy (1995), 'Property Rights and Investment Incentives: Theory and Evidence from Ghana', *Journal of Political Economy* **103**(5), 903–937.

Besley, Timothy and Maitreesh Ghatak (2010), Property rights and economic development, *in* 'Handbook of development economics', Vol. 5, Elsevier, pp. 4525–4595.

Boerner, Lars and Battista Severgnini (2014), Epidemic trade. LSE Economic History Working Paper No. 212/2014.

Boerner, Lars, Jared Rubin and Battista Severgnini (2021), 'A time to print, a time to reform', *European Economic Review* **138**, 103826.

Bosker, Maarten, Eltjo Buringh and Jan Luiten van Zanden (2013), 'From Baghdad to London: unravelling urban development in Europe and the Arab world 800-1800', *Review of Economics and Statistics* **95**(4), 1418–1437.

Budge, E. A. Wallis (1928), *A History of Ethiopia: Nubia and Abyssinia*, Vol. 1 and 2, Meuthen, London.

Buringh, Eltjo and Jan Luiten Van Zanden (2009), 'Charting the "Rise of the West": Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries', *The Journal of Economic History* **69**(02), 409.

Campbell, Bruce MS, Alexander Klein, Mark Overton and Bas van Leeuwen (2015), *British economic growth, 1270–1870*, Cambridge University Press.

Carletto, Calogero, Sydney Gourlay and Paul Winters (2015), 'From Guesstimates to GP-Stimates: Land Area Measurement and Implications for Agricultural Analysis', *Journal of African Economies* **24**(5), 593–628.

Chandler, Tertius and Gerald Fox (2013), *3000 years of urban growth*, Elsevier.

Cheshire, James (2014), 'Analysing surnames as geographic data', *Journal of Anthropological Sciences* (92), 1–20.

Cheshire, James A., Paul A. Longley and Alex D. Singleton (2010), 'The Surname Regions of Great Britain', *Journal of Maps* **6**(1), 401–409. Publisher: Taylor & Francis.

Cheshire, James, Pablo Mateos and Paul A. Longley (2011), 'Delineating Europe's Cultural Regions: Population Structure and Surname Clustering', *Human Biology* **83**(5), 573–598.

Christakos, George, Ricardo A Olea, Marc L Serre, Lin-Lin Wang and Hwa-Lung Yu (2005), *Interdisciplinary public health reasoning and epidemic modelling: the case of black death*, Springer.

CIESN (2018), 'Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets'. Publisher: Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Clark, Gregory (2007), *A farewell to alms: a brief economic history of the world*, The Princeton economic history of the western world, Princeton University Press, Princeton. OCLC: ocn123029809.

Clark, Gregory (2014), *The son also rises: surnames and the history of social mobility*, Princeton economic history of the Western world, Princeton University Press, Princeton, New Jersey.

Clark, Gregory and Neil Cummins (2014), 'Surnames and Social Mobility in England, 1170–2012', *Human Nature* **25**(4), 517–537.

Cook, Lisa D., Trevon D. Logan and John M. Parman (2014), 'Distinctively black names in the American past', *Explorations in Economic History* **53**, 64–82.

Curry, Anne (1992), 'The nationality of men-at-arms serving in English armies in Normandy and the pays de conquête, 1415-1450: a preliminary survey', *Reading Medieval Studies* **XVIII**, 135–163.

Deininger, Klaus, Daniel Ayalew Ali, Stein Holden and Jaap Zevenbergen (2008), 'Rural Land Certification in Ethiopia: Process, Initial Impact, and Implications for Other African Countries', *World Development* **36**(10), 1786–1812.

Deininger, Klaus, Daniel Ayalew Ali and Tekie Alemu (2011), 'Impacts of Land Certification on Tenure Security, Investment, and Land Market Participation: Evidence from Ethiopia', *Land Economics* **87**(2), 312–334.

Demsetz, Harold (1967), 'Toward a Theory of Property Rights', *American Economic Review* **57**(2), 347.

Dincecco, Mark (2015), 'The Rise of Effective States in Europe', *The Journal of Economic History; Santa Clara* **75**(3), 901–918.

Dippel, Christian and Stephan Heblich (2021), 'Leadership in Social Movements: Evidence from the "Forty-Eighters" in the Civil War', *American Economic Review* **111**(2), 472–505.

Dittmar, Jeremiah E. (2011), 'Information Technology and Economic Change: The Impact of The Printing Press', *The Quarterly Journal of Economics* **126**(3), 1133–1172.

Dollinger, Philippe (1970), *The German Hansa*, MacMillan, London. translated and edited by D.S. Ault and S.H. Steinberg.

Donaldson, Dave (2018), 'Railroads of the raj: Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Donaldson, Dave and Adam Storeygard (2016), 'The view from above: Applications of satellite data in economics', *Journal of Economic Perspectives* **30**(4), 171–98. ISBN: 0895-3309.

Donaldson, Dave and Richard Hornbeck (2016), 'Railroads and american economic growth: A "market access" approach', *The Quarterly Journal of Economics* **131**(2), 799–858.

Driscoll, John C. and Aart C. Kraay (1998), 'Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data', *The Review of Economics and Statistics* **80**(4), 549–560. Publisher: The MIT Press.

Dyer, Christopher (1989), *Standards of living in the later Middle Ages: social change in England, c. 1200-1520*, Cambridge medieval textbooks, Cambridge University Press, Cambridge [England] ; New York.

Dyer, Christopher (2002), *Making a living in the Middle Ages: the people of Britain 850-1520*, The new economic history of Britain, Yale University Press, New Haven, Conn.

Eaton, Jonathan and Samuel Kortum (2002), 'Technology, geography, and trade', *Econometrica* **70**(5), 1741–1779.

Ege, S. (Norges Teknisk-Naturvitenskapelige Universitet (1997), 'The promised land: The Amhara land redistribution of 1997', *SMU-Rapport* .

Eisenstein, Elizabeth L (1980), *The printing press as an agent of change*, Vol. 1, Cambridge University Press.

Erb, Karl-Heinz, Veronika Gaube, Fridolin Krausmann, Chritoph Plutzar, Alberte Bondeau and Helmut Haberl (2007), 'A comprehensive global 5 min resolution land-use data set for the year 2000 consistent with national census data', *Journal of Land Use Science* **2**(3), 191–224.

FAO and IIASA (2020), *Global Agro-Ecological Zones (GAEZ v4) - Data Portal user's guide*, FAO and IIASA.

Febvre, L and HJ Martin (1976), *The Coming of the Book: The Impact of Printing 1450-1800*, Verso.

Federico, Giovanni (2009), *Feeding the world: an economic history of agriculture ; 1800-2000*, The Princeton economic history of the western world, 2. printing, 1. paperback printing edn, Princeton Univ. Press, Princeton, N.J.

Finley, Theresa, Raphaël Franck and Noel D. Johnson (2021), 'The Effects of Land Redistribution: Evidence from the French Revolution', *The Journal of Law and Economics* **64**(2), 233–267.

Fischer, Günther, Freddy O Nachtergaele, H van Velthuizen, F Chiozza, G Francheschini, M Henry, D Muchoney and S Tramberend (2021), 'Global agro-ecological zones (gaez v4)-model documentation'.

Fischer, Gunther, Harrij van Nelthuizen, Mahendra Shah and Freddy Nachtergaele (2002), *Global Agro-Ecological Assessment for Agriculture in the 21st Century: Methodology and Results*, Food and Agriculture Organization of the United Nations, Rome.

Fryer, R. G. and S. D. Levitt (2004), 'The Causes and Consequences of Distinctively Black Names', *The Quarterly Journal of Economics* **119**(3), 767–805.

Gates, Scott (2002), 'Recruitment and Allegiance: The Microfoundations of Rebellion', *Journal of Conflict Resolution* **46**(1), 111–130. Publisher: SAGE Publications Inc.

Gennaioli, Nicola and Hans-Joachim Voth (2015), 'State Capacity and Military Conflict', *The Review of Economic Studies* **82**(4), 1409–1448. Publisher: Oxford Academic.

Gentzkow, Matthew, Bryan Kelly and Matt Taddy (2019), 'Text as Data', *Journal of Economic Literature* **57**(3), 535–574.

Gollin, Douglas, Casper Worm Hansen and Asger Mose Wingender (2021), 'Two Blades of Grass: The Impact of the Green Revolution', *Journal of Political Economy* **129**(8), 2344–2384. Publisher: The University of Chicago Press.

Gollin, Douglas, David Lagakos and Michael E. Waugh (2014), 'Agricultural Productivity Differences across Countries', *American Economic Review* **104**(5), 165–170.

Gollin, Douglas, Stephen Parente and Richard Rogerson (2002), 'The Role of Agriculture in Development', *American Economic Review* **92**(2), 160–164.

Gottfried, Robert S. (1983), *The Black Death: Natural and Human Disaster in Medieval Europe*, The Free Press, New York.

Harris, Chauncy D (1954), 'The market as a factor in the localization of industry in the United States', *Annals of the Association of American Geographers* **44**(4), 315–348.

Henderson, J Vernon, Tim Squires, Adam Storeygard and David Weil (2018), 'The global distribution of economic activity: nature, history, and the role of trade', *The Quarterly Journal of Economics* **133**(1), 357–406.

Henze, Paul B. (2000), *Layers of Time: A History of Ethiopia*, Springer. Google-Books-ID: 3VYBDgAAQBAJ.

Hoddinott, John and Yisehac Yohannes (2011), 'Ethiopian Rural Household Surveys (ERHS), 1989-2009'. Type: dataset.

Hornbeck, Richard (2010), 'Barbed Wire: Property Rights and Agricultural Development', *The Quarterly Journal of Economics* **125**(2), 767–810.

Hsieh, Chang-Tai and Peter J. Klenow (2009), 'Misallocation and Manufacturing TFP in China and India', *The Quarterly Journal of Economics* **124**(4), 1403–1448.

Huntington-Klein, Nick (2021), *The effect: An introduction to research design and causality*, Chapman and Hall/CRC.

Incunabula Short Title Catalogue (1998), London: British Library.

Jacoby, Hanan G., Guo Li and Scott Rozelle (2002), 'Hazards of Expropriation: Tenure Insecurity and Investment in Rural China', *American Economic Review* **92**(5), 1420–1447.

Jarvis, A., H.I. Reuter, A. Nelson and E. Guevara (2008), Hole-filled seamless srtm data v4. International Centre for Tropical Agriculture (CIAT).

Jedwab, Remi, Noel D Johnson and Mark Koyama (2019), 'Negative shocks and mass persecutions: evidence from the black death', *Journal of Economic Growth* **24**(4), 345–395.

Jedwab, Remi, Noel D Johnson and Mark Koyama (2020), 'Medieval cities through the lens of urban economics', *Regional Science and Urban Economics* p. 103598.

Jedwab, Remi, Noel D. Johnson and Mark Koyama (2022*a*), 'The Economic Impact of the Black Death', *Journal of Economic Literature* **60**(1), 132–178.

Jedwab, Remi, Noel Johnson and Mark Koyama (2022*b*), Pandemics and cities: Evidence from the black death, Technical report.

Jedwab, Rémi and Adam Storeygard (2021), 'The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010', *Journal of the European Economic Association* p. jvab027.

Johnson, Noel D. and Mark Koyama (2017), 'States and economic growth: Capacity and constraints', *Explorations in Economic History* **64**, 1–20.

Johnson, Simon, John McMillan and Christopher Woodruff (2002), 'Property Rights and Finance', *American Economic Review* **92**(5), 1335–1356.

Joireman, Sandra Fullerton (2000), *Property rights & political development in Ethiopia & Eritrea, 1941-74*, Eastern African studies, J. Currey ; Ohio University Press, Oxford : Athens.

Kandt, Jens, James A. Cheshire and Paul A. Longley (2016), 'Regional surnames and genetic structure in Great Britain', *Transactions of the Institute of British Geographers* **41**(4), 554–569. ISBN: 0020-2754 Publisher: Wiley Online Library.

Kebbede, Girma (1992), *The state and development in Ethiopia*, Humanities Press, Atlantic Highlands, N.J.

Kopsidis, Michael and Nikolaus Wolf (2012), 'Agricultural productivity across Prussia during the Industrial Revolution: a Thünen perspective', *The Journal of Economic History* **72**(03), 634–670.

Koyama, Mark (2016), 'The long transition from a natural state to a liberal economic order', *International Review of Law and Economics* **47**, 29–39.

Lasker, Gabriel W. (1977), 'A Coefficient of Relationship by Isonymy: A Method for Estimating the Genetic Relationship Between Populations', *Human Biology* **49**(3), 489–493. Publisher: Wayne State University Press.

Li, Xuecao, Yuyu Zhou, Min Zhao and Xia Zhao (2020), 'A harmonized global nighttime light dataset 1992–2018', *Scientific Data* **7**(1), 168. Number: 1 Publisher: Nature Publishing Group.

Longley, Paul A., James A. Cheshire and Pablo Mateos (2011), 'Creating a regional geography of Britain through the spatial analysis of surnames', *Geoforum* **42**(4), 506–516.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean and H. Wanner (2004), 'European seasonal and annual temperature variability, trends, and extremes since 1500', *Science* **303**, 1499–1503.

McCann, James (1995), *People of the plow: an agricultural history of Ethiopia, 1800-1990*, University of Wisconsin Press, Madison, Wis.

McClure, Peter (1979), 'Patterns of Migration in the Late Middle Ages: The Evidence of English Place-Name Surnames', *The Economic History Review* **32**(2), 167–182. Publisher: [Economic History Society, Wiley].

McEvedy, Colin and Richard Jones (1978), *Atlas of World Population History*, Facts of File, New York.

McFarlane, Kenneth Bruce (1944), 'Parliament and 'bastard feudalism'', *Transactions of the Royal Historical Society* **26**, 53–79.

Mckinley, Richard (2014), *A History of British Surnames*, Routledge, London.

Monteduro, Patrizia (2012), Agro-climatically attainable yield for low input level rain-fed white potato for baseline period 1961-1990, Technical report, FAO/IIASA.

Morgan, Philip (1987), *War and society in medieval Cheshire, 1277-1403*, number 3rd ser., v. 34 *in* 'Remains, historical and literary, connected with the palatine counties of Lancaster and Chester', Published for the Chetham Society by Manchester University Press, Manchester, UK.

Nei, Masatoshi (1972), 'Genetic distance between populations', *The American Naturalist* **106**(949), 283–292. ISBN: 0003-0147 Publisher: University of Chicago Press.

North, Douglass C. and Barry R. Weingast (1989), 'Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England', *The Journal of Economic History* **49**(4), 803–832.

North, Douglass C., John Joseph Wallis and Barry R. Weingast (2009), *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*, Cambridge University Press.

Nussli, Christos (2011). URL: http://www.euratlas.com/about.html.

Pamuk, Şevket (2007), 'The black death and the origins of the 'great divergence'across europe, 1300–1600', *European Review of Economic History* **11**(3), 289–317.

Parker, Geoffrey (1996), *The military revolution: military innovation and the rise of the West, 1500-1800*, 2nd ed edn, Cambridge University Press, Cambridge ; New York.

Parkin, Harry (2015), 'The fourteenth-century poll tax returns and the study of English surname distribution', *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **48**(1), 1–12. ISBN: 0161-5440 Publisher: Taylor & Francis.

Pettegree, Andrew (2010), *The Book in the Renaissance*, Yale University Press.

Prestwich, Michael (1999), *Armies and warfare in the Middle Ages: the English experience*, Yale University Press.

Prestwich, Michael (2005), *Plantagenet England, 1225-1360*, The new Oxford history of England, Clarendon Press ; Oxford University Press, Oxford : New York.

Ramankutty, N.J., J. Foley, J. Norman and K. McSweeney (2002), 'The global distribution of cultivable lands: Current patters and sensitivity to possible climate change', *Global Ecology and Biogeography* **11**, 377–392.

Reaney, P. H and R. M Wilson (1991), *A Dictionary Of English Surnames.* OCLC: 1125436286.

Restuccia, Diego and Raul Santaeulalia-Llopis (2017), Land Misallocation and Productivity, Working Paper 23128, National Bureau of Economic Research.

Restuccia, Diego and Richard Rogerson (2008), 'Policy distortions and aggregate productivity with heterogeneous establishments', *Review of Economic Dynamics* **11**(4), 707–720.

Roberts, Michael (1956), *The military revolution, 1560-1660: an inaugural lecture delivered before the Queen's University of Belfast*, Vol. 2, M. Boyd.

Rodriguez-Larralde, A., I. Barrai, C. Nesti, Elisabetta Mamolini and Chiara Scapoli (1998), 'Isonymy and Isolation by Distance in Germany', *Human Biology* **70**(6), 1041–1056. Publisher: Wayne State University Press.

Rodriguez-Larralde, Alvaro, José Dipierri, Emma Alfaro Gomez, Chiara Scapoli, Elisabetta Mamolini, Germano Salvatorelli, Sonia De Lorenzi, Alberto Carrieri and Italo Barrai (2011), 'Surnames in Bolivia: A study of the population of Bolivia through isonymy', *American journal of physical anthropology* **144**(2), 177–184. ISBN: 0002-9483 Publisher: Wiley Online Library.

Rogall, Thorsten (2021), 'Mobilizing the Masses for Genocide', *American Economic Review* **111**(1), 41–72.

Rogers, Clifford J. (1993), 'The military revolutions of the Hundred Years' War', *The Journal of Military History* **57**(2), 241.

Rogers, Clifford J., ed. (1995), *The military revolution debate: readings on the military transformation of early modern Europe*, History and warfare, Westview Press, Boulder.

Roodman, David (2011), 'Fitting fully observed recursive mixed-process models with cmp', *The Stata Journal* **11**(2), 159–206.

Rubin, Jared (2014), 'Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation', *The Review of Economics and Statistics* **96**(2), 270–286.

Russell, Josiah Cox (1972), *Medieval Regions and their Cities*, David & Charles, Newton Abbot, Devon.

Scapoli, Chiara, Elisabetta Mamolini, Alberto Carrieri, Alvaro Rodriguez-Larralde and Italo Barrai (2007), 'Surnames in Western Europe: A comparison of the subcontinental populations through isonymy', *Theoretical Population Biology* **71**(1), 37–48.

Schneider, Udo, Andreas Becker, Peter Finger, Elke Rustemeier and Markus Ziese (2020), 'GPCC Full Data Monthly Version 2020 at 0.25 degrees: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data: Gridded Monthly Totals'. Artwork Size: 5 MB - 300 MB per gzip compressed NetCDF file Medium: gzip compressed NetCDF Pages: 5 MB - 300 MB per gzip compressed NetCDF file Version Number: 2020 Type: dataset.

Scott, James C. (1998), *Seeing like a state: how certain schemes to improve the human condition have failed*, Yale agrarian studies, Yale University Press, New Haven.

Shepherd, William R. (1923), *Historical Atlas*, Henry Holt and Company, New York.

Simpkin, David (2008), *The English aristocracy at war: from the Welsh wars of Edward I to the Battle of Bannockburn*, Warfare in history, Boydell Press, Woodbridge, UK ; Rochester, NY. OCLC: ocn220001749.

Simpkin, David (2011), Total war in the Middle Ages? The contribution of English landed society to the Wars of Edward I and Edward II, *in* A. R.Bell, A.Curry, A.King, D.Simpkin and A.Chapman, eds, 'The soldier experience in the fourteenth century', Warfare in history, Boydell Press, Woodbridge, Suffolk, UK ; New York, pp. 61–94.

Sokal, R.R., R.M. Harding, G.W. Lasker and C.G.N. Mascie-Taylor (1992), 'A spatial analysis of 100 surnames in England and Wales', *Annals of Human Biology* **19**(5), 445–476. Publisher: Taylor & Francis.

Stasavage, David (2014), 'Was Weber right? the role of urban autonomy in Europe's rise', *American Political Science Review* **108**, 337–354.

Talbert, Richard J. A, ed. (2000), *Barrington Atlas of the Greek and Roman World*, Princeton University Press, Princeton.

van Etten, Jacob (2012), 'gdistance: distances and routes on geographical grids. r package version 1.1-4'.

Walker, Simon (1990), *The Lancastrian affinity: 1361-1399*, Oxford historical monographs, Clarendon Press ; Oxford University Press, Oxford [England] : New York.

Zanden, Jan Luiten Van, Eltjo Buringh and Maarten Bosker (2012), 'The rise and decline of European parliaments, 1188–1789', *Economic History Review* **65**(3), 835–861.

Zewde, Bahru (2001), *A history of modern Ethiopia: 1855 - 1991*, Eastern African studies, 2. ed., reprinted edn, Currey [u.a.], Oxford. OCLC: 554054711.

Ziegler, Phillip (1969), *The Black Death*, Penguin Books.

Curriculum Vitae

Andrew Thomas graduated from Stevens High School, Rapid City, South Dakota in 2004. He received his Bachelor of Science from the University of Minnesota, Twin Cities, in 2008. He received his Master of Science in Applied Economics from Johns Hopkins University in 2015.