Human capital formation in Europe at the regional level – implications for economic growth

Dissertation	Thèse pour
zur Erlangung des Doktorgrades	l'obtention du titre de Docteur
der Wirtschaftswissenschaft	en Sciences Économiques
im Rahmen einer	dans le cadre d'une
Cotutelle de thèse zwischen	cotutelle de thèse entre

Eberhard-Karls-Universität Tübingen

Wirtschafts- und Sozialwissenschaftliche Fakultät

&

Université de Strasbourg

Bureau d'Économie Théorique et Appliquée (UMR 7522 CNRS)

vorgelegt von présentée par

Ralph Hippe

Tübingen & Strasbourg

2013

Tag der mündlichen Prüfung / Jour de la soutenance:	30.10.2013
Dekane / Doyens:	Prof. Dr. rer. soc. Josef Schmid Universität Tübingen Prof. Thierry Burger-Helmchen Université de Strasbourg
Vorsitzender des Prüfungsausschusses / Président du jury:	Prof. Dr. Dr. h.c. Josef Molsberger Universität Tübingen
Gutachter / Directeurs de thèse:	Prof. Dr. Jörg Baten Universität Tübingen Prof. Claude Diebolt CNRS, Université de Strasbourg
Externe Gutachter / Rapporteurs externes:	Prof. Muriel Dal Pont Legrand Université des Sciences et Technologies de Lille Prof. Jean Luc Demeulemeester Université Libre de Bruxelles Prof. Laurent Heyberger Université de Technologie Belfort-Montbéliard

Acknowledgements

First, I would like to express my profound gratitude to my two supervisors at the University of Tuebingen and at the University of Strasbourg and the Bureau d'Économie Théorique et Appliquée, Professor Joerg Baten and Professor Claude Diebolt. It has been a great honour to work with and be advised by such whole-heartedly dedicated researchers. They have been inspiring mentors and their guidance has been essential throughout the work on this thesis. Their ideas, suggestions and insights have consistently been invaluable. In addition, they have provided me with the opportunity to gain teaching experience and present my work at numerous international conferences, summer schools and workshops. These occasions have allowed me to obtain important comments on my work and share ideas with many international scholars.

In addition, I had the exceptional opportunity to write the PhD thesis within the framework of a German-French "cotutelle de thèse". It is the first of its kind at the department in Tuebingen and without the active support of my two supervisors, even before the beginning of the project, it would never have been realised. In retrospect, the cotutelle has been an exciting adventure for me and it has enabled me to gain a broad range of professional and personal experiences that otherwise would have not been possible. These experiences have also been very stimulating for my research. Therefore, I am very grateful for this incomparable opportunity and thank all persons and institutions involved.

Furthermore, my colleagues have provided me with important advice, support and feedback to advance my academic work. In Tuebingen, I am grateful to Dr. Dominic Behle, Dr. Matthias Blum, Dácil Juif, Christina Mumme, Valeria Prayon, Dr. Mojgan Stegl, Dr. Yvonne Stolz, Franziska Tollnek, Dr. Linda Twrdek and Carolina Vicario. In Strasbourg, I would particularly like to thank Tim Eggebrecht, Faustine Perrin, Walliya Premchit, Naouel Rasouri-Kouider and Qiao Zhang. Many thanks also to the other doctoral students of the Ecole Doctorale Augustin Cournot. In addition, I had the pleasure to work with Professor Joël Petey, Professor Anne-France Delannay and Professor Juan-José Torreiro-Lareo during my time as a Research Assistant (ATER) at Sciences Po Strasbourg. I am particularly indebted to all of them for their tremendous support.

Financial assistance from the German-French University (DFH/UFA), the List Foundation and the University of Strasbourg is gratefully acknowledged. Part of my research is produced within the HI-POD project funded by the European Commission's 7th Research Framework Programme.

Finally, I am deeply grateful to my family and friends, who have provided invaluable support. Their continuous encouragement has been crucial throughout the development of this thesis.

Abstract

This thesis highlights the formation of human capital in the European regions and its implications for economic growth. It is characterised by its combined regional, long-term and European approach. To this end, I refer to Unified Growth Theory and New Economic Geography as the most important recent theoretical contributions and construct an unparalleled new and large database on regional human capital and other economic factors from numerous diverse sources. For the empirical analysis, spatial and GIS methods are employed in addition to standard econometric models. In this way, the thesis explores human capital formation in the regions of the European continent between 1790 and 2010. Moreover, it underlines the relationship between human capital proxies, the determinants of human capital and the long-run impact of human capital on economic growth.

Résumé (version courte)

La thèse traite le sujet de la formation du capital humain en Europe au niveau régional et ses implications sur la croissance économique. Elle est caractérisée par son approche à la fois régionale, à long terme et européenne. A cet effet, je me réfère à la Théorie de la Croissance Unifiée et à la Nouvelle Economie Géographique qui sont les contributions théoriques les plus importantes dans le domaine et je construis une nouvelle grande base de données sur la formation du capital humain et d'autres facteurs à partir de nombreuses sources diverses. Pour les analyses empiriques, des méthodes spatiales et SIG ont été employées en plus des modèles économétriques standards. Ainsi, la thèse explore la formation du capital humain dans les régions du continent européen entre 1790 et 2010. Par ailleurs, elle souligne la relation entre les indicateurs du capital humain ainsi que les déterminants du capital humain et les implications du capital humain sur la croissance économique à long terme.

Résumé (version longue)

La thèse traite le sujet de la formation du capital humain en Europe au niveau régional et ses implications sur la croissance économique. Il est généralement considéré que le capital humain joue un rôle critique pour la croissance économique. Ainsi, les théories innovatrices de croissance endogènes des années 1990 soulignent cette importance primordiale qu'il faut attacher au capital humain. Plus récemment, les modèles de la Théorie de la Croissance Unifiée (TCU) mettent en évidence l'impact décisif du capital humain pour l'explication de la croissance à long terme.

De surcroît, le capital humain et la croissance peuvent également être analysés dans l'espace. En fait, l'analyse des évolutions régionales est plus pertinente que celle des pays en raison du développement souvent inégal des régions au sein des pays. Cette conclusion a été soulignée notamment par la Nouvelle Economie Géographique (NEG). De cette manière, il convient d'étudier les inégalités régionales et leurs évolutions à travers le temps.

Ainsi, on a besoin des données au niveau régional et à long terme afin de valider les hypothèses de ces différentes théories. Cependant, mesurer le capital humain en général et au niveau régional, dans le passé en particulier, n'est pas un exercice facile. Il faut utiliser des indicateurs approximatifs qui n'ont pas toujours été disponibles selon le pays et l'époque jusqu'à maintenant. Pour cette raison, la thèse vise à combler cette lacune dans la littérature en mettant en œuvre une analyse au niveau régional à long terme dans toute l'Europe. De cette manière, elle élargie les connaissances existantes sur le capital humain en Europe dans les dimensions temporelles et spatiales.

A cet effet, différents indicateurs du capital humain sont utilisés. D'abord, la méthode d' « age heaping » mesure les capacités numériques (« numeracy ») dans le passé, c'est-à-dire les capacités à compter et à connaître son propre âge. Cette méthode analyse la

distribution de l'âge qui est déclaré, par exemple, par des individus lors des recensements de population. De cette façon, elle approxime le capital humain de base d'une région. En particulier, elle a l'avantage de permettre de suivre l'évolution du capital humain pour toute l'Europe plus tôt que d'autres indicateurs. En conséquence, les recherches effectuées pendant les dernières années ont établi les capacités numériques comme indicateur reconnu du capital humain. Par ailleurs, l'alphabétisation (à savoir la capacité de lire et d'écrire) et le niveau d'études sont utilisés pour mesurer le capital humain au XX^{e} et XXI^{e} siècle. C'est de cette manière que le capital humain est estimé dans les régions européennes entre 1790 et 2010.

Sur le plan de l'unité d'analyse, la thèse fait référence à la Nomenclature des Unités Territoriales Statistiques (NUTS) de l'Union Européenne. En conséquence, les indicateurs employés donnent la possibilité d'analyser l'évolution du capital humain au niveau régional (NUTS 2) et même, dans quelques pays, au niveau départemental (NUTS 3). Ainsi, l'approche régionale à long terme est effectuée en faisant le lien entre les régions historiques et actuelles. De plus, j'ai recours aux théories existantes dans le domaine (en particulier à la TCU et à la NEG) et à certaines méthodes spatiales et économétriques qui ont été peu utilisées dans de tels contextes jusqu'à maintenant.

La thèse est composée de plusieurs chapitres dont la plupart sont des articles scientifiques autonomes. De cette façon, après l'introduction et les remarques méthodologiques, elle souligne l'évolution de quelques indicateurs représentant une partie des facteurs de production en Europe. Même si les variables employées ne constituent que des notions des facteurs de production (la terre, le capital, le travail, le progrès technologique et le capital humain), elles mettent plus en perspective le facteur de l'analyse suivante : le capital humain.

Basée sur ces premières intuitions, la relation de deux indicateurs du capital humain est éclairée. Plus précisément, il s'agit des capacités numériques et de l'alphabétisation. Il s'avère que ces deux indicateurs sont bien liés historiquement et méthodologiquement. Ensuite, l'analyse économétrique emploie des données historiques européennes de la fin du XIX[e] siècle et des données des pays en voie de développement en Afrique, en Asie et en Amérique latine de la deuxième moitié du XX[e] siècle. En prenant en compte différentes variables de contrôle, les résultats mettent en évidence que l'alphabétisation et les capacités numériques sont étroitement liées.

Ainsi, il est possible d'avancer la recherche concernant le capital humain en utilisant les capacités numériques. Dans cet esprit, on suit l'évolution du capital humain entre 1790 et 1880. A cet effet, une nouvelle base de donnés portant sur les capacités numériques a été créée au niveau régional en Europe. Au total, cette base de données comprend plus de 570 régions dans 39 pays européens, permettant de souligner les différences régionales du capital humain. En fait, les pays les plus avancés se situaient au nord et au centre de l'Europe. De nombreux pays étaient marqués par des disparités régionales importantes. Cependant, ces différences se sont allégées au cours du XIX[e] siècle.

Ensuite, il est nécessaire de comprendre et d'expliquer ces inégalités européennes. Entre autres, la distribution des inégalités foncières a récemment attiré l'attention des chercheurs. Galor *et al.* (2009) ont développé une théorie liée à la TCU sur les répercussions négatives d'une distribution inégale des terres sur le capital humain. En conséquence, nous évaluons l'impact des inégalités foncières sur le capital humain en Europe autour de la fin du XIX[e] siècle. Pour ce faire, nous utilisons la base de données sur les capacités numériques en rajoutant des données sur les inégalités foncières et d'autres variables pertinentes. En effectuant des modèles de régression de type moindre carrés ordinaires (MCO) et en employant la technique économétrique d'analyse avec variables instrumentales (VI), les résultats économétriques confirment l'hypothèse de Galor *et al.* (2009) pour les régions européennes les moins industrialisées.

Etant donné que l'unité d'analyse est la région, il est également possible d'étudier plus précisément l'impact de la localisation géographique sur le capital humain d'une région. En d'autres termes, on peut analyser la distribution spatiale des régions européennes et l'existence potentielle des clusters par l'Analyse Exploratoire des Données Spatiales (AEDS). Plus particulièrement, l'étude utilise la statistique I de Moran, le graphique de Moran et les cartes de significativité de Moran en 1850 et 1930. Il s'avère qu'il y avait des clusters importants dans le passé. La proximité géographique apparaît jouer un rôle important et significatif pour la distribution spatiale du capital humain.

Ensuite, on peut également considérer l'aspect de l'espace sous un angle différent. La théorie de la NEG souligne l'importance de l'accès au marché pour le développement économique régional. Dans un papier récent, Redding et Schott (2003) rajoutent le facteur du capital humain au modèle de base de la NEG. Leur modèle théorique révèle que l'éloignement des grands marchés commerciaux constitue un frein à l'accumulation du capital humain des individus. De cette manière, les régions avec peu d'accès au marché ont un niveau de capital humain plus bas que celles qui ont un bon accès au marché. Cette hypothèse de la « pénalité de l'éloignement » pourrait donc expliquer les inégalités nationales et régionales du capital humain. Elle a déjà été confirmée par des analyses empiriques portant sur les régions européennes aujourd'hui. En conséquence, il paraît important d'explorer cette pénalité dans le long terme en faisant référence aux années 1850 et 1930 et en utilisant les capacités numériques et l'alphabétisation. Les résultats indiquent que l'accès au marché a une influence positive et significative sur le capital humain dans les modèles de régressions de type MCO, Tobit et VI. Ainsi, ils mettent en évidence l'hypothèse de la « pénalité de l'éloignement » à long terme.

Jusqu'ici, la thèse considère la formation du capital humain à partir de la fin du XVIII^{e} siècle jusqu'à la première moitié du XX^{e} siècle. Il convient alors d'élargir le champ d'étude afin de faire le lien avec le présent. A cet effet, la base de données a été complétée afin de montrer les inégalités régionales en 1850, 1900, 1930, 1960, 2000 et 2010. Pour cette raison, les capacités numériques, l'alphabétisation et le niveau d'études sont employés comme indicateurs du capital humain. L'analyse utilise des boîtes à moustaches pour montrer les différences régionales. De plus, deux mesures d'inégalités (à savoir le coefficient de variation et le coefficient de Gini) soulignent les variations au sein des pays. Il s'avère que les inégalités régionales caractérisent beaucoup de pays européens à travers le temps. Ces résultats mettent encore une fois en évidence l'importance de l'analyse régionale.

Finalement, il est possible d'exploiter la nouvelle base de données pour traiter la question de l'impact du capital humain sur l'innovation et la croissance économique à long terme. Donc, les données historiques du capital humain pour 1850, 1900, 1930 et 1960 sont mises en relation avec des indicateurs actuels de l'innovation (soit les demandes de brevets par million d'habitants déposées à l'Office Européen des Brevets) et de la croissance économique (le PIB par habitant). Prenant en compte plusieurs variables de contrôle, il ressort que l'historique du capital humain d'une région est lié de manière significative à la richesse économique d'aujourd'hui.

En somme, la thèse souligne d'abord que les capacités numériques (mesurées par la méthode d' « age heaping ») sont bien liées à d'autres indicateurs comme l'alphabétisation. Ensuite, en utilisant les différents indicateurs du capital humain de base, il s'avère que les niveaux du capital humain ont augmenté au fil du temps. Cependant, beaucoup de pays sont caractérisés par des inégalités régionales importantes. Celles-ci sont en partie plus grandes que les différences entre les pays, mettant en évidence la pertinence de l'approche régionale. Par ailleurs, la distribution inégale des terres et la localisation géographique d'une région sont des facteurs expliquant ces inégalités du capital humain. Enfin, le capital humain apparaît comme facteur important pour le développement économique d'une région.

Ainsi, il faut que les décideurs politiques mettent l'accent sur l'augmentation du capital humain afin de générer de la croissance économique. De plus, il est nécessaire de créer des environnements qui sont favorables à la formation du capital humain, particulièrement dans les régions périphériques. L'amélioration des infrastructures peut y contribuer d'une façon importante. Par ailleurs, les inégalités régionales soulignent la nécessité de prendre en compte les spécificités régionales dans le processus d'élaboration des politiques.

Dans le futur, les recherches devraient davantage se diriger vers l'exploration du niveau régional à long terme en Europe. La combinaison de différents indicateurs existants et la création de nouveaux indicateurs du capital humain permettrait d'améliorer l'approximation du capital humain. Le même raisonnement s'applique également aux approches théoriques qui devraient encore mieux éclairer le rôle du capital humain dans le processus de la croissance économique.

En conclusion, le capital humain est un facteur important dont l'importance ne devrait pas cesser d'augmenter au XXI^e siècle.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

GENERAL ACRONYMS, ABBREVIATIONS AND SYMBOLS

2SLS	Two-stage least squares
A	Level of technology
ABCC	A linearly transformed Whipple Index (named after A'Hearn, Baten, Crayen and Clark)
ALL	Adult Literacy and Life skills Survey
Benelux	Belgium, Netherlands and Luxembourg
BICSE	Board on International Comparative Studies in Education
BRIC	Brazil, Russia, India and China
CV	Coefficient of variation
d	Distance
E(I)	Expected value of Moran's I
EFP	European Fertility Project
EFTA	European Free Trade Agreement
EMP	European Marriage Pattern
EPO	European Patent Office
ERDF	European Regional Development Funds
ESDA	Exploratory Spatial Data Analysis
ESF	European Social Fund
EU	European Union
FE	Fixed Effects
FIMS	First International Mathematics Study
FISS	First International Science Study
GDP	Gross Domestic Product

GENERAL ACRONYMS, ABBREVIATIONS AND SYMBOLS

GIS	Geographic Information System
GMV	Galor, Moav and Vollrath
GNP	Gross National Product
H	Human capital
ha	Hectares
HH	High-high cluster
HL	High-low cluster
I	Moran's I
IAEP	International Assessment of Educational Progress
IALS	International Adult Literacy Survey
IEA	International Association for the Evaluation of Educational Achievement
ILO	International Labour Organization
INSPIRE	Infrastructure for Spatial Information in the European Community
IPUMS	Integrated Public Use Microdata Series
IR	Industrial Revolution
IRS	Increasing Returns to Scale
ISCED	International Standard Classification of Education
IV	Instrumental Variable
K	Capital
k	Number of neighbours
L	Labour
LAU	Local Area Unit

GENERAL ACRONYMS, ABBREVIATIONS AND SYMBOLS

LDC	Least Developed Country
LH	Low-high cluster
LIML	Limited Information Maximum Likelihood
LISA	Local Indicator of Spatial Association
LL	Low-low cluster
MA	Market Access
max	Maximum
min	Minimum
μ	Mean value
n (or: N, obs.)	Number of observations
NEG	New Economic Geography
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-Operation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PPS	Purchasing Power Standard
R&D	Research and Development
rw	Reading and writing ability
S_0	Scaling factor

GENERAL ACRONYMS, ABBREVIATIONS AND SYMBOLS

SA	Supplier Access
σ (or: sd, Std. Dev.)	Standard deviation
SIMS	Second International Mathematics Study
SISS	Second International Science Study
SITES	Second Information Technology in Education Study
t	Time period
TALIS	Teaching and Learning International Survey
TFP	Total Factor Productivity
TIMMS	Third International Mathematics and Science Study; Trends in International Mathematics and Science Study
UGT	Unified Growth Theory
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
w	Element of the spatial weight matrix
WI	Whipple Index
WIPO	World Intellectual Property Organization
Wz	Spatially lagged vector
Y	Output
yrs.	Years
z	Vector

COUNTRY ABBREVIATIONS

AL	Albania
AM	Armenia
AT	Austria
AZ	Azerbaijan
BA	Bosnia-Herzegovina
BE	Belgium
BG	Bulgaria
BO	Bolivia
BR	Brazil
BY	Belarus
CH	Switzerland
CL	Chile
CO	Colombia
CY	Cyprus
CZ	Czech Republic
DE	Germany
DK	Denmark
EC	Ecuador
EE	Estonia
ES	Spain
FI	Finland
FR	France
GE	Georgia

COUNTRY ABBREVIATIONS

GR	Greece
HR	Croatia
HU	Hungary
IE	Republic of Ireland
IN	India
IS	Iceland
IT	Italy
KE	Kenya
LI	Liechtenstein
LT	Lithuania
LU	Luxembourg
LV	Latvia
MD	Moldova
ME	Montenegro
MK	FYROM (Former Yugoslav Republic of Macedonia)
MT	Malta
MX	Mexico
NL	Netherlands
NO	Norway
OT	Ottoman Empire
PA	Panama
PL	Poland
PT	Portugal

COUNTRY ABBREVIATIONS

RO	Romania
RU	Russia
SE	Sweden
SI	Slovenia
SK	Slovakia
SR	Serbia
TR	Turkey
TZ	Tanzania
UA	Ukraine
UK	United Kingdom
US	United States
USSR	Union of Soviet Socialist Republics
YU	Yugoslavia

The most valuable of all capital is that invested in human beings

(Alfred Marshall 1890, Principles of Economics, VI.IV.11).

1. Introduction

1.1 Human capital, economic growth and regional analysis

Human capital is currently considered to play a crucial role in economic development. For example, the European Commission stresses that "[i]nvestment in education and training for skills development is essential to boost growth and competitiveness" (European Commission 2012a, p. 2). Although the concept of human capital has been developed over several centuries, it only achieved its current importance in both theory and empirics in recent decades. In particular, innovative endogenous growth models (e.g., Lucas 1988, Romer 1990) have placed new emphasis on human capital in the creation of economic growth. These theoretical underpinnings have given rise to a number of important contributions in the literature on the relationship between human capital and economic growth.

The scope of such research has been further expanded by the consideration of economic growth in the long run. This focus marks a new and important step because "our understanding of the contemporary world is limited and incomplete in the absence of a historical perspective" (Galor in Snowdon 2008, p. 120). The advances in this area have been driven by Unified Growth Theory (UGT; e.g., Galor 2005a, Galor 2012). Unified Growth Theory highlights the preponderant role of human capital in the long-run growth process. Its models vitally contribute to the understanding and explanation of the long-term transitions in economic development that have led to an explosion of economic growth in modern industrialised countries since the Industrial Revolution. This take-off brought about a substantial divergence in worldwide economic development over the last 200 years (Pomeranz 2000). While the gap in per capita GDP between the most advanced and the least advanced world regions was only 3 to 1 in 1820, it has increased to approximately 18 to 1 today (Galor 2005a). Therefore, a better grasp of this phenomenon is essential to comprehend not only the past but also the present and the future.

In this way, it is clear that the notion of human capital is very important in economic theory. However, human capital needs to be measured quantitatively to validate theoretical models. To this end, human capital has been measured in a range of ways. For example, literacy rates and educational attainment are two standard proxies in the literature. These and other proxies have their limitations, particularly in historical studies. Evidence on human capital formation before the 20th century remains relatively scarce. A new measure was recently developed that considerably increases the availability of such evidence. This proxy is numeracy as proxied by the age heaping method (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010a). In simple terms, the age heaping method takes advantage of certain heaping patterns in the distribution of ages in historical censuses or other documents. Thus, it estimates the basic numerical capacities of a population. In this way, the method has also enabled researchers to overcome one major hurdle in the quantification of human capital: the lack of numeracy studies. For example, as Houston stated at the turn of the last century, "quantitative historical studies of numeracy are absent" (Houston 1999, p. 385). The age heaping method fills this gap in the literature and enables human capital studies to extend further back in time because evidence on this indicator is available even in Roman times and beyond. This characteristic of the age heaping method makes it particularly appropriate for the evaluation of long-term growth processes as highlighted by Unified Growth Theory.

In addition, there has been another major strand in the economics literature over the last two decades that does not focus on the time aspect of economic development but more on its geographic and spatial distribution. This literature has been marked by New Economic Geography (NEG) models. NEG aims to provide explanations for the existence and evolution of economic concentration in space. This concentration leads to regional clusters of economic development and to convergence and divergence processes. Accordingly, NEG underlines the key aspect of regional analysis in economics. Therefore, its founder, Nobel laureate Paul Krugman, emphasises, "one of the best ways to understand how the international economy works is to start by looking at what happens inside nations" (Krugman 1991b, p. 3). Similarly, other prominent authors such as Porter emphasise that "the relevant economic area is smaller than many nations" (Porter 1994, p. 38). Thus, the use of regions as the unit of analysis allows researchers to clarify the economic growth process within nations. This aspect is particularly important because there are often substantial regional differences within countries, in many cases larger than those between countries. As a consequence, adopting a regional perspective may allow for an analysis of these differences within nations and explanations of how present-day regional economic inequalities emerged.

Europe represents a prime example because the regional economic differences in Europe are striking. For example, the EU "generates 43 % of its economic output in just 14 % of its territory" (European Commission 2008, p. 2). Furthermore, considering the European regions is particularly appropriate in this context because the recent renewed interest in economic geography is linked to the need to understand the potential consequences of increased European integration (Fujita *et al.* 1999).

Therefore, the importance of analysing human capital *both* in the long term and at the regional level in Europe is evident. Unfortunately, although an increasing number of studies has been conducted on regional education in Europe, "[t]he geography of education, especially at [the] subnational level, is [still] a huge black box" (Rodríguez-Pose and Tselios 2011, p. 358). This statement is even more valid in the long run. As a result, it is essential to further investigate this black box to analyse regional human capital formation and the process of economic growth.

1.2 Aim and contribution of the thesis

Thus, the aim of this thesis is to highlight the formation of human capital in the European regions and its implications for economic growth. It is particularly characterised by its regional, long-term and European focus. In this way, it emphasises the formation of human capital in the long run (i.e., 1790-2010) and its geographical distribution at the subnational level in a European perspective. Thus, it broadens the existing evidence in both the spatial and temporal dimensions.

To this end, both theory and empirics have been considered. In particular, Unified Growth Theory and New Economic Geography are referred to as the most important recent contributions in economic theory on human capital, economic growth and regional development. For the empirical analyses, the available regional data on human capital at a European scale were insufficient or non-existent. Therefore, a new and unparalleled large database has been constructed on regional human capital formation and other factors in Europe. Note that Europe is understood here as a more global concept than the European Union. Therefore, the entire European continent is considered, from the Atlantic (Portugal) to the Urals (Russia) and from Scandinavia to the Mediterranean. This required the collection of numerous sources and the analysis of a hundreds of regions from many different countries and throughout time.

Moreover, a correspondence between the historical and current regions was developed to permit the comparison of historical and current data. This was done by applying the Nomenclature of Territorial Units for Statistics (NUTS) classification developed by the European Union to historical regions. In other words, this thesis proposes an integrated analysis of regional development in the long run.

Initially, this database concerned numeracy as measured by the innovative age heaping method. Long-term regional analyses at the European level have been lacking in the corresponding literature. Therefore, the present study explicitly addresses this missing piece of evidence in the long-term formation of human capital in Europe. As this thesis was developed, this already large database on numeracy at the regional level was enlarged through the inclusion of other indicators. The first alternative indicator to numeracy, and possibly the most standard indicator for measuring past human capital, is literacy. The ability to read and write as a proxy for literacy is particularly appropriate due to the methodological proximity to the calculated numeracy values. For more recent periods, this literacy proxy no longer has any real meaning because contemporary European populations can generally be considered literate. Currently, educational attainment is a standard means of measuring human capital. For this reason, this third indicator of human capital has been used for the current period.

Nevertheless, the scope of this thesis extends beyond the mere measurement of human capital. First, the thesis analyses some of the factors that may explain the different levels of human capital within and between countries. In particular, it considers land inequality, geographical proximity and market access because they are important elements in the underlying economic growth theories. Further databases have been newly constructed or adapted to this end. In addition to econometric models, GIS and spatial data analysis software and methods from (economic) geography and spatial econometrics were employed. Some of the econometric methods and the theoretical underpinnings have only seldom been used in such long-term perspectives, despite their potential applications being substantial (e.g., Exploratory Spatial Data Analysis (ESDA)). Therefore, the thesis highlights some of the possibilities of using state-of-the-art spatial or geographical methods in long-term analyses.

Finally, not only the determinants of human capital but also the impact of human capital on current economic outcomes are important issues. The economic growth theories noted above postulate that human capital plays a major role in determining innovation and economic growth. Therefore, the thesis analyses the effects of human capital on innovation and economic growth in the long run.

1.3 Outline of the thesis

The thesis is divided into several chapters. As this is a cumulative thesis, most of these chapters (i.e., chapters 3 to 10) are autonomous research papers.

This introductory chapter provides an outline of the research questions that will be addressed in the main part of the thesis.

The second chapter explains some major methodological issues that have to be addressed in the context of regional human capital formation in the long run. Thus, the different elements in the title of the thesis are highlighted. More specifically, the chapter refers to the three-dimensional approach of the thesis, some of the major theories in the areas of human capital, economic growth and regional analysis, as well as the employed regional classification concept.

The third chapter empirically introduces the long-run regional European perspective. It presents a collection of maps to provide a rough descriptive overview of the long-run evolution of the factors of production in the European regions. Some common factors of production in the standard economics literature are land, capital, labour, technological progress and human capital. The rationale of this chapter is to first consider all of the economic factors of production before the thesis turns to the most recent factor introduced in the literature: human capital. This procedure permits human capital to be regarded in a broader context, as one of several economic factors.

Based on these initial intuitions, the fourth chapter considers human capital in greater detail by highlighting the relationship among different human capital proxies. Because numeracy (as proxied by the age heaping method) is a fairly novel method, additional evidence is required to clarify its relationship with the other indicators. Literacy is an obvious candidate because it has similar characteristics to the numeracy proxy. Data from developing countries in Africa, Asia and Latin America from the second half of the 20th century have been employed in addition to the historical European data from the end of the 19th century. Moreover, the chapter takes accounts for factors other than literacy that might explain the age heaping phenomenon. The results show that numeracy correlates well with literacy both in historical Europe and in present-day developing countries.

These findings lay the foundations for the further use of the age heaping method. Therefore, the fifth chapter traces the evolution of numeracy in the European regions between 1790 and 1880. To this end, a new, large database on numeracy has been created. In addition, the chapter highlights the evolution of regional inequality in human capital using the coefficient of variation. The results indicate a significant gap in numeracy levels between advanced western and central European countries and the rest of Europe. Nevertheless, differences in basic numeracy between and within countries declined over the 19th century because the periphery was able to solve its basic numeracy problem.

Based on these findings, the sixth chapter aims to explain these regional differences in numeracy in Europe. Thus, it considers a possible hurdle to human capital formation: land inequality. As advanced by a corresponding economic theory, land inequality may have a negative impact on the formation of human capital due to the power of large landholders. Landowners may block changes because improved education for their workers may encourage the latter to initiate uprisings against the rule of the landowners. Moreover, workers may migrate to the industrial sector, which may offer higher salaries.

To analyse the effect of land inequality on human capital, the chapter approximates regional human capital using numeracy and land inequality using the share of large landholdings. It employs the earliest available regional land inequality data at the end of the 19th century and the first decades of the 20th century from several European countries. Using OLS and IV regression techniques, the results show a substantial negative effect of land inequality in less industrialised countries.

Subsequently, the seventh chapter investigates another factor in greater depth that may contribute to regional inequalities in human capital: space. Therefore, it employs rather recently developed techniques to enhance understandings of the geographical importance and the potential clustering of human capital in the European regions. To this end, ESDA is employed to validate the existence of positive and negative spatial autocorrelation and the formation of clusters in the distribution of regional human capital. Accordingly, the methods employed in this study are Moran's I, the Moran scatterplot and the Moran significance map. Human capital is proxied by numeracy in 1850 and literacy in 1930. The findings suggest that spatial clustering has characterised the spatial distribution of human capital in the European regions in the past.

The eighth chapter also considers the aspect of space but from a different perspective. It analyses the effect of market access on human capital by taking advantage of recent theoretical models and testing their pertinence in a long-term setting. Market access (or market potential) is an important factor in NEG models and has attracted the attention of numerous researchers. In a recent contribution, lower market access is assumed to generate disincentives to human capital investment, leading to a 'penalty of remoteness'. However, the models are primarily tested at either the regional level in Europe with fairly recent data or in cross-country comparisons for longer time periods. For this reason, the chapter considers the market access concept in Europe in 1850 and 1930, using numeracy and literacy, respectively. The results show that market access has a significant, positive effect on human capital in OLS, Tobit and IV regression models. Therefore, the chapter confirms the remoteness hypothesis for Europe in the long run.

Thus far, the thesis has considered human capital formation from the end of the 18th century until the first half of the 20th century. Therefore, the ninth chapter expands the temporal scope of the thesis and makes the connection to the present. Accordingly, it traces the long-term evolution of human capital at the regional level in Europe between 1850 and 2010. It focuses on three different variables that proxy for human capital: numeracy, literacy and educational attainment. Data have been collected for 1850, 1900, 1930, 1960, 2000 and 2010. The regional differences within European countries are demonstrated using boxplots. Moreover, the use of inequality measures (i.e., the coefficient of variation and the Gini coefficient) highlights the substantial variation in human capital. Regional inequalities appear to be characteristic for many countries and throughout time, again illustrating the importance of regional human capital analysis.

The tenth chapter uses the enlarged database on human capital to explore the relationship between historical human capital and current indicators of innovation and economic growth. In other words, it demonstrates the economic effects of human capital in the long run. The regression analysis employs proxies for human capital in conjunction with different control variables in 1850, 1900, 1930 and 1960. In addition, current innovation is measured using patent applications per capita to the European Patent Office and economic development using GDP per capita. The findings highlight that historical regional human capital is a key factor in explaining contemporary regional disparities in innovation and economic development.

Finally, the eleventh and final chapter summarises the results of the thesis and emphasises their implications for economic policy and future research.

2. Methodological background

2.1 Three-dimensional approach

The entirety of this thesis is characterised by an implicit three-dimensional approach. Conceptually, these three dimensions can be called space, time and attribute. The latter can be represented by economics (and statistics) in general and the analysis of human capital in particular. Moreover, one could include other factors that are important in such an analysis. One may present these three dimensions in the following manner:

- Space: regional (i.e., subnational) level in Europe
- Time: long-run evolution, ca. 1790-2010
- Attribute: human capital

Kant already recommended organising human knowledge in a similar three-dimensional manner (Martí-Henneberg 2011). Therefore, the thesis unites elements from the three research disciplines of economics, history and geography. Thus, it merges econometric methods and models from economic theory with historical and long-term applications, adopting a geographic and spatial perspective in a European context and employing corresponding geographic and spatial methods.

By definition, this is an interdisciplinary approach that has benefitted from advances in diverse areas. On the one hand, new and more elaborate economic modelling and theory have continued to be developed and advanced. In recent years, long-run theories such as Unified Growth Theory and geographic theories such as New Economic Geography have highlighted the importance of considering either time or space in economics. Previously, interest in historical and geographical aspects had been peripheral, and mainstream economics did not consider them explicitly (and this remains somewhat true even today). Moreover, econometric methods have been combined with and advanced by spatial econometric methods that account for the spatial dimension of economics. An improved historical understanding of institutions, processes and evolutions has provided a complementary direction. As Temple notes in his review on the new economic growth literature, "historical analysis must be a major way forward" (Temple 1999, p. 148). Accordingly, research continued to advance in (new) economic history in recent years (Costa *et al.* 2007). On the other hand, GIS technologies have exploited significant improvements in computer processing and are able to perform very complex spatial data analyses that would not have even been imaginable in the past. Projects on Historical GIS note that research is moving forward in this direction (e.g., Knowles 2008, Bailey and Schick 2009, Martí-Henneberg 2011).

For these reasons, it may be worthwhile not to restrict oneself to one of these dimensions but to take advantage of the substantial opportunities that particularly recent technological advances have allowed. Therefore, this thesis explicitly takes all three dimensions into account. It contributes to unify not only economics and history on the one hand and economics and geography on the other but economics with both history and geography.

Finally, this approach may also be considered a bridge to the increasingly evolving field of and literature on development economics. Development economics is in a sense closely related to its economic and historical aspects and also to its geographic dimension. The experiences of the early developing countries, i.e., those of the Western world in general and Europe in particular, may perhaps not directly provide guidance for contemporary developing countries in Asia, Latin America, Africa and Eastern Europe because the circumstances in the current globalising world differ from those in the past. Nonetheless, one may suggest that there are some important fundamentals that are common to the economic and social development of countries throughout time. This notion of commonalities is also implicitly considered by long-term economic growth theories such as Unified Growth Theory. In this way, long-term regional studies may help to understand the past and better relate the present to a much larger temporal context. Furthermore, they may allow new insights into the future evolution of developing and industrialised countries.

2.2 Definition and measurement of human capital

2.2.1 The notion of human capital

This thesis focuses on human capital formation. However, what is human capital? Nobel laureate Gary S. Becker defines human capital as "the knowledge, information, ideas, skills, and health of individuals" (Becker 2002, p. 3). Alternatively, the OECD takes a more economic approach, defining this form of capital as "the knowledge, skills, competencies and other attributes embodied in individuals that are relevant to economic activity" (OECD 1998, p. 25). Thus, the most common proxies for human capital are educational variables, which are also employed in this thesis.

According to Becker (2002), human capital is the most decisive type of capital in contemporary economies. He refers to studies showing that human capital accounts for over 70% of total capital accumulation in the US, representing more than a fifth of total GDP. Consequently, "[t]echnology may be the driver of a modern economy, especially of its high-tech sector, but human capital is certainly the fuel" (Becker 2002, p. 3).

Before characterising human capital in greater detail, the term itself and its history should be explained and briefly portrayed. According to one of the early proponents of the human capital concept, Nobel laureate Theodore W. Schultz, human capital "is a form of capital because it is the source of future earnings, or of future satisfactions, or of both of them. It is human because it is an integral part of man" (Schultz 1972, p. 5). What then is so special about human capital? "The most critical attribute of human capital arises from the fact that the person and his human capital are inseparable. The person must always be present wherever the services of his human capital are being rendered" (Schultz 1972, p. 8).

Nevertheless, it must be stressed that the concept of human capital has a much longer history. The notion of a specific capital invested in human beings appears in the works of Adam Smith. For example, he makes explicit reference to the acquired abilities of individuals: "[t]he acquisition of such talents, by the maintenance of the acquirer during his education, study, or apprenticeship, always costs a real expense, which is a capital fixed and realized, as it were, in his person. Those talents, as they make a part of his fortune, so do they likewise that of the society to which he belongs. The improved dexterity of a workman may be considered in the same light as a machine or instrument of trade which facilitates and abridges labor, and which, though it costs a certain expense, repays that expense with a profit" (Smith 1776/1976, p. 265-266; Folloni 2010).

Nevertheless, Smith was not the first author to reflect on the notion of capital that is invested in human beings (see Folloni 2010). Petty had already considered factors in addition to land and population to explain the wealth of nations in the 17th century. Therefore, the value of labour should be taken into account (Petty 1690/1899). Subsequently, Cantillon also referred to the value of labour but more in the sense of the cost entailed in the maintenance of slaves and their children (Cantillon 1755/1952, see also Hofflander 1966). Smith was less interested in measuring this value than in understanding the differences in remuneration across occupations (Folloni 2010).

Other prominent economists also discussed the notion of human capital in different ways. One may mention Mill (1848/1926) and Marshall (1890/1920) in this regard. For example, the latter defines human capital in the form of personal wealth "as to include all those energies, faculties and habits which directly contribute to making people industrially efficient" (Marshall 1890/1920, p. 58; Folloni 2010).

Human capital has always been a somewhat controversial topic because the notion of defining human beings as a form of capital has been considered repulsive by different authors. For instance, Mill believed that human beings cannot be defined as capital (Folloni 2010). Controversy persists in the current public discourse, despite that the scientific community has largely accepted the term.[1] Schultz's argument may be helpful here, stating that "our values and beliefs inhibit us from looking upon human beings as capital goods, except in slavery, and this we abhor [but…] there is nothing in the concept of human wealth contrary to [the] idea that it exists only for the advantage of people. By investing in themselves, people can enlarge the range of choice available to them. It is one way free men can enhance their welfare" (Schultz 1961, p. 2; Folloni 2010). Thus, it appears important to advance our knowledge regarding this crucial factor.

2.2.2 Attributes and effects of human capital

Human capital can be categorised according to different criteria. For example, one may refer to its attributes and effects, as emphasised by Kwon (2009). He shows that, on the one hand, Crawford (1991) distinguishes between different attributes or characteristics of human capital. These can be classified into two categories. The first category describes human capital as being self-generating and expandable. In fact, knowledge is boosted when it is used, differentiating it from other resources such as raw materials. The expansion of human capital may be due to exogenous reasons and endogenous causes. Thus, knowledge may be enlarged by the interplay of external factors such as external knowledge or information and endogenous skills or experiences.

[1] One may mention the fact that human capital was chosen as the 'non-word' of the year by German linguists in 2004 (FAZ 2005).

Second, it is also possible to diffuse this knowledge to other agents. This shows that human capital is sharable. Finally, the last characteristic of human capital is that it is transportable because it is an integral part of its holder.

In sum, one may say that the first two characteristics (expandable, self-generating) increase human capital's volume, while the latter two (sharable, transportable) increase its range (Crawford 1991).

On the other hand, Kwon (2009) states that human capital may have effects from the very micro level up to the macro level. In other words, it has repercussions on individuals but also on organisations and society as a whole.

He highlights this point by considering the labour market. In the internal labour market, human capital may have a positive effect on individuals' income by increasing their productivity (e.g., Schultz 1961, Becker 1964, Schultz 1971, Lucas 1988). Evidently, employers prefer highly productive employees because they maximise the profitability of a company. They are also more mobile in the labour market than other employees. These facts allow such employees to move upwards in the hierarchical structure (Sicherman and Galor 1990). In the external labour market, the employment possibilities of an unemployed person are also affected by his human capital (e.g., Vinokur *et al.* 2000). For example, it is easier for individuals with a high level of human capital to access information and obtain job opportunities.

Lastly, both former perspectives may be combined to underline the impact of human capital on the society (Kwon 2009). For example, human capital may affect political stability, democracy and human rights due to its impact on the individual's social consciousness of society's basic underlying framework (McMahon 1999, Beach 2009). The link between the individual's human capital and social consciousness may lead to socio-political developments (e.g., Sen 1999). Ultimately, human capital may affect economic growth (e.g., Romer 1990).

2.2.3 Proxies of human capital

This thesis focuses on the analysis of human capital proxies that measure rather basic skills, such as numeracy and literacy. In fact, the lack in basic education for the masses has characterised most of human history and, unfortunately, it is still a major issue in some regions of the world today. However, everybody needs to have a sufficient level of human capital in order to be productive in the respective economic areas. Accordingly, the Council of the European Union highlights that "[a]ll individuals need a core package of knowledge, skills and attitudes for employment, inclusion, subsequent learning as well as personal fulfilment and development" (Council of the European Union 2003, C 134/4). Therefore, some common human capital indicators are now briefly presented. These indicators will mostly be explained in greater detail in subsequent chapters.

2.2.3.1 A more recent indicator: numeracy and the age heaping method

The understanding of concepts of natural numbers is unique to humans (Hauser and Spelke 2004). However, research on numeracy has only been increasing during the last decades. This is both illustrated by the recent important number of publications on the subject and by the history of the term itself. In fact, the term 'numeracy' only appeared at the end of the 1950s in the so-called Crowther report. This report was presented to the Ministry of Education of the United Kingdom in 1959. Its aim was to give a "mirror image of literacy" (Central Advisory Council for Education 1959, p. 269) because the term 'literacy' had already been well-established at that time.

Until today, policy actions often focus more on literacy than on numeracy skills (UNESCO 2005). This may be surprising because we are surrounded by numbers in our everyday life and every form of organised society has been based on numbers throughout history (Cohen 2005). Accordingly, "[h]istory records the ever-increasing need for ordinary men and women to be able to count and to do simple arithmetic" (Cohen 2005, p. 18). Therefore, numeracy is an important aspect of human capital. In consequence, it is also on today's policy agenda. For example, it is seen to be "contributing to the empowerment, effective functioning, economic status, and well being of citizens and their communities" (Gal 2000, p. IX).

However, it is important to specify what is meant by numeracy, as the meaning of the term has been increasingly enlarged. Today, it describes the "ability to add, subtract, multiply and divide" (UNESCO 2005, p. 421). But it may also be defined as the "ability to process, interpret and communicate numerical, quantitative, spatial, statistical and even mathematical information in ways that are appropriate for a variety of contexts" (UNESCO 2005, p. 150). These different possibilities of defining numeracy reflect the non-existence of a standard and universal definition (O'Donoghue 2002).

Historical numeracy can be measured usng the age heaping method. Age heaping as such is nothing new for researchers in fields such as demography. Nevertheless, while it has been a major problem for these disciplines because it creates a bias to the true age distribution of a population, it enables researchers of human capital to approximate historical numeracy levels.

The basic idea is that there is a general phenomenon in historical censuses (and still in some censuses of current developing countries). In fact, there are more individuals who state that their age ends on the numbers 0 and 5 than on others. It is possible to show that this age heaping is primarily the result of lacking age awareness. In other words, there is an important part of individuals that did not know their own age and were not able to calculate it. This is why age heaping gives an indication of the numeracy of the population.

Given the very large availability of historical censuses, measuring numeracy by applying the age heaping method allows to trace back human capital over the long run and at small territorial units. This is a major advantage with regard to other human capital proxies which are more limited in these space and time dimensions.

Why is the heaping phenomenon on zero and five by far the most important one? What is so special about these numbers? Sheets-Johnstone (1990) gives a hint in indicating that our biology predetermines us to employ our bodies for the communication with other individuals. This circumstance can partly explain the fact that counting words are related to hands and fingers in an important number of worldwide languages. For instance, in the English language the number 'five' comes indirectly from the ancient term for 'fist' which was derived from the ancestor of the Proto-Indo-European languages, the so-called Nostratic (Winter 1992, Manaster Ramer *et al.* 1998, Harper 2008). The biological preconditions that humans have ten fingers and ten toes may thus explain this phenomenon.

Not least thanks to the large potential of sources, this method marks a new and innovative step towards a fuller comprehension of human capital formation.

2.2.3.2 A standard measure: literacy

Numeracy as proxied by the age heaping method is still a rather recent indicator of human capital. A standard way of measuring historical human capital is literacy. Literacy can be measured using different approaches. The most important ones are signature rates (e.g., of married couples or recruits) and the ability to read and write as stated in historical censuses. In this thesis, the most appropriate way is to employ the ability to read and write because its underlying population is similar to the one of numeracy. In contrast, marriage registers and recruitment lists are much more limited. Therefore, the use of censuses is a standard approach that has been recommended by international organisations such as UNESCO (UNESCO 1953).

Literacy has been a standard indicator of human capital and education for more than a century. Given the fact that in 2005 there remained about 771 million people above 15 years worldwide who did not have the basic competencies in literacy, it is evident that the study of literacy is an important contribution to improve many lives (UNESCO 2005). In this spirit, the United Nations launched the Literacy Decade (from 2003 to 2012) because "creating literate environments and societies is essential for achieving the goals of eradicating poverty, reducing child mortality, curbing population growth, achieving gender equality and ensuring sustainable development, peace and democracy" (United Nations 2002, p. 3).

2.2.3.3 Other human capital indicators and skill tests

Numeracy and literacy are only two options for approximating human capital. There are different other proxies that are used in analyses focusing on historical and current periods. Some of the most important ones are educational attainment and years of schooling (e.g., Benhabib and Spiegel 1994, Barro and Sala-i-Martin 1995, Barro 2001, Krueger and Lindahl 2001, Murtin and Viarengo 2010), school enrolment rates (for historical data: e.g., Becker and Woessmann 2009, Becker *et al.* 2012; for current data: e.g., Barro 1991, Levine and Renelt 1992, Mankiw *et al.* 1992), book production (Baten and van Zanden 2008), educational expenditure (e.g., Barro and Lee 1996) and education-augmented labour input (see Woessmann 2003 for a detailed discussion).

Moreover, various international studies on cognitive skills have been conducted over the last decades. Some of the most important ones are, among others,[2]

[2] See BICSE (1995) for a detailed overview.

- the International Adult Literacy Survey (IALS) and the Adult Literacy and Life skills (ALL) Survey (NCES 2005)
- the International Assessment of Educational Progress (IAEP) (1988, 1991) (BICSE 1995)
- the studies and surveys conducted by the International Association for the Evaluation of Educational Achievement (IEA) (e.g., FIMS 1964, FISS 1971, SIMS 1981, SISS 1984, Reading Literacy Study 1990, TIMSS 1995, PIRLS 2001, TIMSS 2003, SITES 2006, TIMSS 2007) (IEA 2011) and by the OECD (PISA in 2000, 2003, 2006, 2009, 2012; TALIS in 2008, 2013; PIAAC in 2011) (OECD 2010a, OECD 2010b, OECD 2012).

The abundance of these international studies highlights once more the importance that has been attributed by researchers and policy makers to measuring human capital over the last decades.

2.3 Theories of human capital, economic growth and regional development

2.3.1 Origins

The idea that human capital may be considered as a determinant of economic growth has a long history (see Demeulemeester and Diebolt 2011), and its key importance has increasingly been considered over the last decades. It has already been indicated that Smith and Marshall had incorporated the notion of something akin to human capital in their thinking. However, only after World War II a proper theory of human capital was developed. In particular, Becker was an important founder of this theory (e.g., Becker 1981). Initially, human capital theory was designed for microeconomics, relating incomes to human capital, but has subsequently been adapted to macroeconomics.

Later on, the relationship between human capital and growth was more directly examined. Human capital was deemed to explain whether the differences in human capital in the work force were able to give an explanation for the 'residual' total factor productivity (TFP) after having taken account of inputs from labour and capital (e.g., Denison 1967, Jorgenson and Griliches 1967; Schütt 2003).

Nevertheless, the real surge in contributions in this issue only emerged thanks to the 'New Growth Theory'. This new interest manifested itself by a bulk of convergence regressions in cross-country settings and the intention to reveal the fundamental causes of different growth patterns among countries. To this end, many variables have been introduced but one in particular stood out from the others: human capital (Schütt 2003).

Therefore, first, the following subsections give a very brief but not exhaustive introduction to some of the most important economic growth models that include human capital implicitly or explicitly: exogenous, endogenous and, in particular, unified growth theory.[3] Then, the spatial dimension of economic growth is presented by sketching some characteristics of New Economic Geography.

2.3.2 Exogenous growth models

2.3.2.1 Original neoclassical Solow model

The original Solow model, or alternatively, Solow-Swan model (Solow 1956, Swan 1956) defines the aggregate production function as

$$Y_t = F(K_t, L_t \cdot A_t), \tag{2.1}$$

with Y representing output, K representing capital, L representing labour and A representing a technology index. The characterising properties of this function are, first,

[3] The following presentation of the exogenous and endogenous growth models is based on Schütt (2003).

constant returns to scale, second, decreasing returns to inputs and, third, a constant elasticity of substitution.

The capital stock evolves at a constant savings rate and a constant depreciation rate. Moreover, L and A grow at exogenously given exponential rates. Without A growth would go to zero. Nevertheless, the diminishing returns to K are offset by increasing efficiency. For this reason, steady state convergence takes place in which both Y and K/L (the capital/labour ratio) are growing at the rate of A. In consequence, neither population growth nor the savings rate affect long-run economic growth. In other words, the slope of the growth path is unaffected in the long run, while changes in population growth or the savings rate modify its level (Schütt 2003). This model takes account of technological progress but not of human capital.

2.3.2.2 Human-capital augmented Solow model

In contrast, the human-capital augmented Solow model extends the original model by introducing human capital H as an input in the aggregate production function. This function is a Cobb-Douglas production function with labour-augmenting technological progress. According to the important contribution by Mankiw *et al.* (1992), it can be represented in the following form:

$$Y_t = K_t^{\alpha} H_t^{\beta} (A_t L_t)^{1-\alpha-\beta}, \tag{2.2}$$

with Y representing output, K representing capital, H representing the human capital stock, A representing the technological level and L representing labour. Output elasticity is measured by α to the input K, β to H and $1\text{-}\alpha\text{-}\beta$ to AL. Thus, the aggregate production function exhibits, first, constant returns to scale and, second, diminishing returns to the reproducible factors of production due to the assumption that $\alpha + \beta < 1$. This model has in common with the original Solow model that both population growth and technological progress are exogenously given and that capital depreciates (Schütt 2003).

One of the main disadvantages of these exogenous growth models is that human capital is exogenously given and not endogenously determined in the model. In contrast, this is the major advancement of the endogenous growth models.

2.3.3 Endogenous growth models

Endogenous growth models have been particularly influential over the last decades. Romer's work in 1986 (Romer 1986) can be named as the first important contribution in this field. These 'new growth models' aim at endogenising the different sources that lead to growth. In this way, the growth rate it is not exogenously given anymore but it is established within the endogenous growth model itself.

The overall category of endogenous growth models can be divided into two main approaches (Aghion and Howitt 1998; Schütt 2003). The first line of thought focuses on human capital accumulation as the main driver of growth. The second approach underlines the importance of technological change for the creation of economic growth. Both approaches will now be briefly presented.

2.3.3.1 Human capital accumulation

Lucas (1988) initiated this line of research which highlights the effects of human capital accumulation on economic growth. In his model, he assumes that the economy is constituted by identical individuals who maximise their life-time utility. These individuals, or agents, can control their degree of consumption and the time which is allocated between the acquisition of skills and work. Thus, physical capital is accumulated by the level of consumption and the future productivity is determined by the agent's allocation of time. The corresponding production function is

$$Y_t = AK_t^{\beta}(u_t h_t L_t)^{1-\beta} h_{a,t}^{\gamma}, \tag{2.3}$$

with Y representing output, K representing capital, L representing labour and A representing technology. Moreover, u is the share of time that an individual allocates to work (and thus $1\text{-}u$ is the fraction that he allocates to the accumulation of human capital) and h_a represents the average human capital that exists in this economy. Note that Lucas assumes that A is constant and population growth is exogenous. However, perhaps the most fundamental assumption concerns the relationship between $1\text{-}u$ and the rate at which human capital grows, i.e., $\dot{h}_t / h_t$. He assumes that this relationship is linear, that is

$$\dot{h}_t = h_t \delta(1 - u_t) \Leftrightarrow \frac{\dot{h}_t}{h_t} = \delta(1 - u_t), \tag{2.4}$$

where δ denotes the maximum level of growth that h can achieve. This is also called the "productivity of schooling" (Aghion and Howitt 1998, p. 330). In consequence, the level of human capital does not determine its growth rate. In simple terms, independent of the amount of already accumulated human capital, a certain effort always leads to an identical growth rate of human capital. More intuitively, one may explain this result by evoking the fact that already acquired skills make it easier to learn (Romer 2001). Acquiring skills is not subject to diminishing returns so that growth in human capital is unlimited in this model. This effect endogenously generates growth, growth being dependent on δ and $1\text{-}u$ (Schütt 2003).

2.3.3.2 Technological change

Technological change as the main driver of growth is put forward by a second line of thought (Schütt 2003). Its focus on technological change makes it similar to the original Solow-Swan models. Nevertheless, they differ in the way that technological change is no longer left unmodeled. Thus, they assume that many inventions are the consequence of the research and development (R&D) efforts made by firms. Therefore, human capital is conceived to be a catalyst that enables technological progress and that does not generate sustained growth by itself.

One important assumption is that human capital induces individuals to willingly accept change and adapt to the introduction of new technologies (e.g., Nelson and Phelps 1966). In consequence, the notions of knowledge spillovers and knowledge diffusion become central: a higher level of human capital incorporated in individuals speeds up the diffusion process of technology. Thus, human capital enables countries that lag behind to advance more rapidly and converge to the countries at the technological frontier.

In particular, the Romer model (1990) considers an economy with three sectors. These sectors are the final-goods sector, the intermediate-goods sector and the research sector. The latter produces designs for the creation of capital goods which are to be sold to the second, intermediate sector. Both human capital and the knowledge stock are used by the research sector. The intermediate sector produces new intermediate goods by the help of these designs and the savings of the economy. Lastly, the final-goods sector combines the intermediate goods with additional labour and human capital inputs to obtain the final output. In consequence, the final sector has the following production function:

$$Y = H_Y^{\alpha} L^{\beta} \sum_{i=1}^{A} x_i^{1-\alpha-\beta}, \tag{2.5}$$

where H_Y denotes the human capital that is employed in the production of the final good, A denotes the existing knowledge stock and x_i denotes the intermediate goods. One important feature of this model is that it disaggregates capital into various forms of intermediate goods which affect output in additive separable ways. Thus, A determines the number of intermediate goods that exist in the modelled economy. Romer takes the assumption that knowledge is a nonrival input and that L and H are constant.

The fact that knowledge is a nonrival good requires the introduction of imperfect competition into the model because perfect competition would lead to the non-existence of R&D. In this latter case, firms would produce at marginal costs and all revenues generated by the production of goods would be exhausted. This scenario does not allow firms to pay for their fixed costs and thus for R&D either. Therefore, Romer uses monopolistic competition to circumvent this issue.

Moreover, the knowledge stock is represented by the quantity of designs produced by the R&D sector. This knowledge stock evolves in the following way:

$$\dot{A} = \delta H_A A \Leftrightarrow \frac{\dot{A}}{A} = \delta H_A, \tag{2.6}$$

where H_A denotes the human capital that is employed in the research sector ($H = H_Y + H_A$) and δ is a measure of the productivity of the research sector. Thus, both the knowledge stock and the human capital allocated to research activities are needed to create new knowledge.

Finally, sustained growth is created by two reasons. Firstly, the variety of goods increases and thereby A expands. Secondly, the existence of knowledge spillovers is assumed. These spillovers are possible because all researchers have access to A. As in the human capital accumulation models, the assumption of a linear relationship as shown in (2.6) corresponds to the idea that human capital's productivity in research rises proportionally to A. Thus, this assumption allows to obtain endogenous growth by an unlimited growth of knowledge. In other words, the human capital stock is this time the factor that speeds up growth. This is the major difference between the two types of endogenous growth models. In Romer's model, it suffices to increase at one time the human capital stock in order to increase the growth rate, while in the Lucas model this rate requires a rise in the rate of human capital accumulation. These and other earlier mentioned differences between the different types of endogenous growth and the exogenous growth models are briefly summarised in Table 2.1 (Schütt 2003).

2.3.3.3 Critiques and advances

Since the original contributions by Lucas (1988) and Romer (1990), the literature has steadily been advancing further (see Ang and Madsen 2010). In fact, these initial models (together with models by e.g. Segerstrom *et al.* (1990), Grossman and Helpman (1991) or Aghion and Howitt (1992)) are now considered as the first generation of endogenous growth theory models. This first generation was criticised by Jones (1995). In particular, he refuted the scale effect property of the Romer model which postulates that an increase in the size of the population (and thus in the number of researchers) generates a simultaneous increase in the steady-state growth rate. As a consequence, a second generation of models takes semi-endogenous or Schumpeterian approaches. For example, Jones (2002) uses a semi-endogenous growth model which does not assume scale effects in knowledge production, so that human capital and R&D do not show level effects. In contrast, Schumpeterian growth models have been proposed by Aghion and Howitt (1998), Peretto (1998), Young (1998) and others and are characterised by the maintenance of scale effects. However, they take the assumption that economic growth leads to the propagation of products and in this way to a lower effectiveness of R&D. Still, sustainable growth is possible in such a world if the R&D effort is fixed at a constant ratio to the variety of products, itself being fixed to the relative population size. The appropriateness of these new types of models is still being debated and empirically tested (e.g., Madsen 2010, Madsen *et al.* 2010; Ang and Madsen 2010).

2.3.4 Unified Growth Theory

2.3.4.1 Definition and scope

Exogenous and endogenous growth models have been very important pillars in the evolution of growth models that include human capital. However, they are limited in the sense that they are not able to explain the process of economic growth since the beginning of the existence of humans (Galor 2005a). Therefore, a new ambitious theory has been developed aiming at understanding the growth patterns of human kind in the very long run. This Unified Growth Theory has particularly been advanced by Galor (e.g., Galor and Weil 2000, Galor and Moav 2002). The term “unified” refers to the explanation of economic development as a whole and the unification of the basic economic micro structure (Galor 2005a). Due to its long term and worldwide scope, it is clear that it has to be a very general theory. Nevertheless, it can be flexibly adapted to the characteristics of the countries under study (Snowdon 2008).

2.3.4.2 Explaining economic development in the very long run

To understand the concept of UGT, it is necessary to briefly review the fundamental development schemes to which it refers explicitly or implicitly in its models.

According to Galor (Snowdon 2008), economic development was more or less similar throughout the world during most of history when compared to today’s large differences among the major regions of the world. Figure 2.1 shows that GDP per capita in the different regions of the world were rather similar and stagnating during almost two millennia until the take off during the Industrial Revolution. From then onwards, substantial discrepancies have emerged. This evolution is commonly termed the Great Divergence. Thus, there are large differences between the most advances countries of the Western Offshoots (the United States, Canada, Australia, New Zealand), Western Europe and the other regions of the world in 2000. To refine this picture, economic growth in the very long run can be classified in the following way (see Figure 2.2).

First, a long phase of hunting and gathering was typical for human development between 1 million BC and 8000 BC. After the Neolithic Revolution, which changed the way humans lived by allowing agriculture and permanent settlement, began the Malthusian growth regime. It was characterised by a per capita income which always fluctuated around low levels of subsistence, i.e., the so-called 'Malthusian trap'. More specifically, technological progress and the expansion of land led to increases in the population size and a higher population density but income per capita was only temporarily positively affected. The economies that were more technologically advanced or which had more land at their disposal had population densities which were higher than others. Still, even those more advanced economies had comparable per income levels in the longer term.

It took about another 10,000 years to escape the Malthusian growth regime by means of the Industrial Revolution. One can see that the phases become smaller as we approach the current regime. Subsequently, the Post-Malthusian growth regime lasted until the demographic transition in Europe around 1870. The economies were able to escape this trap and grow sustainably during this epoch. Population growth still led to an increase of income per capita. However, and in contrast to the Malthusian growth regime, technology was advancing even further. For this reason, output was not completely offset by population growth as before. Thus, both population and per capita income increased at growing rates.

The final growth pattern is the modern growth regime which has characterised the world's advanced countries from the 19th century onwards. This regime shows an acceleration of technological progress. Moreover, human capital is increasingly demanded. Thus, UGT postulates that human capital played a crucial role in the transition from the

Post-Malthusian regime to modern growth. Human capital and technological progress brought about the demographic transition which has led to lower population growth. Because population growth does not counterbalance the gains achieved in total output anymore, the result is high and sustainable growth in per capita output.

What about the future? Galor believes that the gap between the advanced (Western) economies and the richer developing countries will be narrowed, while the gap of these two country groups to the LDCs will likely be further enlarged (Snowdon 2008). The future will prove if he is right.

2.3.4.3 Main building blocks

Galor (2005a) specifies the most important building blocks underlying UGT. The first important building block are the elements related to the Malthusian era. More specifically, individuals are initially limited to consumption at the subsistence level. In such a world, more income leads to a higher rate of population growth. Technological progress only temporarily increases per capita income. Therefore, an increase in technological progress results in a rise in population size, offsetting the per capita income gains because returns to labour are assumed to be diminishing. Nevertheless, technological progress ultimately outpaces population growth, leading to per capita income growth.

Second, one has to specify the forces that lead to technological progress in economic development. In this context, an important distinction has to be made between earlier and later stages. In early phases, technological progress is stimulated by population size. For example, population size had an effect on the provision of new ideas, technological diffusion and trade. In contrast, human capital is the prime factor for technological advancement in later phases.

Third, the bases of human capital formation are another building block. New technologies are mainly skill-biased, whereby they lead to a kind of disequilibrium which increases human capital demand. Individuals with a higher level of human capital are more adaptable to new technologies and thus have an advantage to individuals with less human capital in the new environment.

Finally, the last building block are the underlying paternal decisions that lead to changes in the quantity and quality of children. In other words, the decisions on the quantity and quality of offspring are related to a time constraint because parents can spend time either on child-raising or on working in the labour market. Ultimately, human capital becomes increasingly important in the economy, leading to an increase in its demand which gives parents incentives to substitute child quality for child quantity.

2.3.4.4 Challenges and innovations

In this way, Unified Growth Theory aims to develop a unified theory which is able to explain long-run growth. How does UGT practically endeavour to come up to its goals? First, UGT works in a dynamical system which allows to depart from a stable Malthusian steady-state. To achieve this, one could assume an important shock to the system which itself is subject to multiple equilibria. On the other hand, there could be a gradual change allowing to advance from the existing stable equilibrium. However, both explanations are inconsistent. First, because history shows that there has not been a major shock in economic development but rather a gradual advancement and, second, because it is not possible to get out of a stable equilibrium (Galor 2005a).

Therefore, other methodological ways have to be found. UGT chooses to use phase transition as a major tool. Phase transition works through evolving latent state variables which have an effect on the characteristic qualitative properties that are inherent to the dynamical system. These latent state variables are in particular technological progress and population. Both variables alter the system in a qualitative way although output per capita is assumed to be stagnating in the Malthusian regime. This alteration brings the Malthusian equilibrium to fall and enables upcoming forces to make the transition to the modern and sustainable growth regime.

In this way, it is possible to unify economic development in one framework. UGT shows that sustained economic growth is possible through the formation of human capital, highlighting the preponderant role that human capital plays in the explanation of economic growth. Moreover, UGT allows the identification of the responsible forces behind these major socioeconomic transitions in human history. Understanding these long-term growth paths shows the hurdles but also the way that less developed countries have to go to be able to advance economic growth and the future development of industrialised countries (Galor 2005a).

2.3.5 New Economic Geography

2.3.5.1 Space and (new) economic geography

After this insight into the long-term growth models, it is also important to consider the spatial character of the economy because this thesis focuses on regions as the unit of analysis. Economic geography models aim to explain the differences between regions and geographical units as well as the evolution of spatial disparities.

However, space was forgotten in mainstream economics for a long time (Baumont *et al.* 2000, Garretsen and Martin 2010). More specifically, regional or spatial analyses were a periphery phenomenon in economic theory and econometric studies. For the area of economic theory, Nobel laureate Samuelson affirms in 1952 that “[s]patial problems have been so neglected in economic theory that the field is of interest for its own sake” (Samuelson 1952, p. 284).

Still, the attention of economists to the study of regional phenomena has only increasingly emerged since the 1990s. One reason for this is that new models have been created which take explicitly regional evolutions into account. In particular, these models are part of New Economic Geography (NEG).

According to Samuelson (1983), the discipline of economic geography was primarily founded by von Thünen (see von Thünen 1826/2008; Fujita and Krugman 2004). Braudel also highlights the importance of von Thünen, stressing that he "ranks alongside Marx as the greatest German economist of the nineteenth century" (Braudel 1992, p. 38). Later important contributions in the field came most notably from Christaller (1933) and Lösch (1940).

So how can one define economic geography? In simple terms, economic geography is defined by Krugman as "the branch of economics that worries about where things happen in relation to one another" (Krugman 1991b, p. 1). The relationship of these 'things' to each other can be clarified by Tobler's first law of geography: "[e]verything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236).

New Economic Geography may be said to have its origins in 1991, when Krugman published the article "Increasing Returns and Economic Geography" (Krugman 1991a). One defining difference between the 'traditional' economic geography and this 'new' one resides in the fact that the NEG models are mathematically more sophisticated. Nevertheless, the idea that this theory of economic geography is 'new' has created many critics from 'traditional' economic geographers. The (initial) 'hostility' from different directions may be illustrated by the remark that referees claimed that "[i]t's obvious, it's wrong, and anyway they said it years ago" (Gans and Shepherd 1994, p. 178; Fujita and Krugman 2004). Moreover, NEG models have been considered to be too general and insufficiently realistic given their restrictive sets of hypotheses and their high level of complexity (Kleinewefers 2010).

2.3.5.2 Characteristics and modelling techniques

There are some particular characteristics of NEG that should be considered in the current context (see Fujita and Krugman 2004). NEG models are general equilibrium models with possible multiple equilibria which differentiates it from more traditional economic geography models. In addition, they take some important assumptions to model the geography of economics. First, increasing returns to scale (IRS) are fundamental to explain spatial concentration processes. IRS have the effect that concentration increases productivity and output. In other words, output is increased more than proportionally if the factors of production are multiplied by the same quantity. Second, and in consequence of the first condition, the assumption of imperfect competition is necessary because the inclusion of space generates a form of oligopolistic competition. That is, modelling space makes it necessary for firms to take strategic decisions (Combes *et al.* 2006). Third, the movement between locations incurs transport costs. In addition, the possibility for firms, workers and consumers to move is crucial because it is a precondition for the existence of agglomerations (Fujita and Krugman 2004). Finally, in which region the agglomeration is formed is arbitrary and is assumed to be dependent on initial conditions or historical accidents. This is why history may play an important role for the distribution of firms and consumers in space. In other words, 'history matters' (Garretsen and Martin 2010).

Thus, one important aim of NEG is to explain the concentration of economic activity (see e.g., Betran 2011, Combes *et al.* 2011). This means that economic actors are not distributed randomly in space. In fact, there are different forces at work which influence actors. These forces can be classified into two categories: centripetal and centrifugal forces (Table 2.2; Fujita and Krugman 2004). First, one may distinguish three main centripetal or agglomeration forces: linkages, thick markets and knowledge spillovers. Linkages refer to beneficial contacts between consumers, producers and suppliers of goods which are enhanced by spatial proximity. The idea of thick markets takes into account the advantages associated with the existence of a large and dense market. Finally, knowledge spillovers are also very important. The notion of knowledge spillovers was developed by Marshall (1890/1920), highlighting the positive externalities that may arise of locating near other actors of the same field and thus being in contact with the knowledge of others. Second, there are three main centrifugal or dispersion forces: immobile factors, land rent and commuting as well as congestion. For example, one immobile factor is land. These immobile factors are not mobile across space and their inherent characteristics are exogenous. Land rent and commuting refers to the fact that land rentals are lower in the countryside than in a city. That is, land and housing prices are lower in rural areas. Commuting enables individuals to live in another place as the location of the workplace. Finally, congestion describes the consistent phenomenon that traffic is quite dense in agglomerations. These generate traffic jams and other associated time consuming inconveniences.

Given these different forces, there are reasons for firms and workers to agglomerate or to disperse in space. The existence of cities and areas with highly concentrated economic activity shows that agglomeration forces can be very strong. In Krugman’s words, concentration is “the most striking feature of the geography of economic activity” (Krugman 1991b, p. 5).

NEG explains this observation by the model of circular causation (see Figure 2.3). Basic NEG models assume two identical regions, *A* and *B*, with the same real income in autarky (see Fujita and Thisse 1997, Eckey and Kosfeld 2004). If some individuals now move from region *B* to region *A* then the number of consumers in *A* rises at the expense of *B*, thereby increasing *A*’s market. In consequence, firms prefer to locate in *A* because the market in *A* is now larger than in *B*. This phenomenon is called the demand effect.

On the other hand, the fact that more firms locate in *A* also means that more goods and more varieties of goods are available for consumers in this region. In consequence, firms have to compete more fiercely with each other. This competition leads to a decrease in the price of products in *A*, increasing the real income of its consumers. This is the real income effect. Thus, the higher real incomes in *A* attract consumers in *B*. More and more consumers move from *B* to *A* until all consumers are located in *A*. Therefore, this effect has resulted in the creation of an agglomeration (Fujita and Thisse 1997). In sum, the described process has demonstrated the different agglomeration forces that come into play.

However, the centrifugal forces work against these centripetal forces. In fact, the concentration of firms and workers depends on the level of transport costs (Figure 2.4) (Combes *et al.* 2008). One may distinguish three stages of development. In the first one (1), both regions are still in autarky because transport costs have such a high level that trade and the transfer of capital or labour is impossible between the regions. A decrease in transport costs enables the regions to trade with each other. Trade benefits both regions by increasing their real incomes. Nevertheless, when the transport costs attain a critical low level the process of circular causation alters the distribution of capital and labour between the regions (2). The real income in *A* rises importantly while it falls in *B* because the creation of an agglomeration in *A* leads to a decrease of consumers in *B*. In this way, inequality increases between the regions.

Still, this evolution has not come to an end. If transport costs continue to fall to another critical point, the dispersion forces are so strong that it becomes increasingly unattractive for workers and firms to work and produce in *A*. For this reason, workers and firms now increasingly take the other direction and begin to return to *B*. In the end, given a relatively low level of transport costs, both regions have once again the same real income but this time on a significantly higher level (3). Therefore, policy makers would have to consider the level of transport costs to take the right policy choices.

One may summarise the main predictions of NEG in the following way. First, the existence of agglomeration forces leads to a rise in regional inequality when regions integrate. More specifically, capital and labour (and their incorporated human capital) flow from the poorer to the richer region, the latter being characterised by a larger market (the so-called home market effect). Second, richer regions export their manufacturing goods while poorer regions export agricultural products. Third, increasing concentration leads to rising congestion costs which can offset the forces of agglomeration. This effect can bring about a decrease in GDP growth in the central region and higher growth rates in the periphery region. In this way, a process of convergence appears.

In sum, NEG has brought new attention to regional growth and the analysis of the geography of the economy.

2.4 Definition of the regional level in Europe

This thesis employs regions as its basic unit of analysis. However, how can one define these regions? Clearly, there are different possibilities to define a region. In this thesis, the current classification for administrative boundaries of the European Union is used: the Nomenclature of Territorial Units for Statistics (abbreviated NUTS according to its French version). This classification distinguishes between several levels of regional aggregation (see Figure 2.5). Every higher level is totally constituted by regions that establish the lower level (for details see Table 2.3 and Table 2.4).

The highest level is NUTS 0 (Figure 2.6). This level is the national level. The next step is NUTS 1 (Figure 2.7), the first degree of disaggregation. In some countries, these NUTS 1 regions are identical to official national regions, such as the *Bundesländer* in Germany. In other countries, these are regional constructs that suffice the criteria of the EU but do not have any historical predecessors or political importance in the countries (e.g., in Hungary). A third category of countries comprises smaller countries. Given their relatively small size of population, the NUTS system classifies the entire country as NUTS 1 region (e.g., the Republic of Ireland, Luxembourg, Denmark, Finland, Estonia).

The following level, the NUTS 2 level, is a further step to smaller regional units (Figure 2.8). For example, these regions are the *départements* in France, the *Comunidades y ciudades Autonomas* in Spain and the *regioni* in Italy.

Finally, the NUTS 3 level is the smallest disaggregation level within the NUTS classification.[4] These NUTS 3 regions correspond, for instance, to the *oblasti* in Bulgaria, the *kraje* in the Czech Republic and the *Län* in Sweden.

How can one decide to what NUTS level a region or a group of regions belongs? The official delimitations are based on several criteria. In fact, there are three main underlying principles in the constitution of NUTS regions (see Eurostat 2011d).

The first principle is related to population size. The idea is that all regions of one NUTS level should have a similar population size. Therefore, the size of the population of a region has to fulfil certain thresholds to belong to a certain NUTS level (Table 2.5). Nevertheless, there remain important differences between the regions within one NUTS level because the minima and maxima are relatively large.

The second principle is to give preference to administrative divisions. The European classification is modelled on the national administrative classification to facilitate the correspondence between NUTS regions and national classification systems. Generally, two NUTS levels are constituted by these official regional units and a third one is added through aggregation.

[4] However, there are still smaller, disaggregated units, the so-called LAU (Local Area Units). However, we stay within the NUTS terminology and classification system within the framework of this thesis.

Finally, the third and last principle is to favour geographical units that are of general nature and not particular to any given area of activity. This facilitates the comparability and objectiveness of the NUTS regions.

However, the NUTS classification does not refer to all European countries. More specifically, it only includes members and Candidate Countries of the EU as well as EFTA countries. In particular, some countries in Eastern Europe do not fall into any of these categories. Thus, in these cases one has to take the current administrative borders of these countries (e.g., for Belarus, see Figure 2.9; for Russia, see Figure 2.10; for Ukraine, see Figure 2.11). In this way, all European countries have been disaggregated into their respective regions.

The time-invariant use of the NUTS framework allows to combine historical with current data. To this end, it was necessary to develop the correspondence between historical regions and the current classification. In some cases, this is not a matter of great effort because the historical regions have not been modified until today or there have only been minor changes. To mention two examples, the administrative structures of France and Spain have not considerably changed over the last two centuries.

In France, there have been modifications primarily due to gains and losses in wars or other military conflicts (in particular Alsace-Lorraine, Savoy and Alpes-Maritimes) and due to the restructuring of the greater Paris region (Ile-de-France). The existing regional structure of *départements* was created during the French revolution and replaced the historic provinces of the *Ancien régime* that had been characteristic for France's subnational administrative structure for several centuries.

Similarly, in Spain the current *provincias* have officially been established in 1833 and since then there have only been minor changes. For example, the province of the Canary Islands was split into two in 1927 (to Las Palmas and Santa Cruz de Tenerife), increasing the number of Spanish provinces to 50. Moreover, there have been some changes in the names of provinces over the last decades. For instance, the name of the province of Logroño was changed to La Rioja, of Santander to Cantabria and of Oviedo to Asturias. The domination in the regional language has also replaced the former Castilian version during the last years (e.g., Orense to Ourense, Alicante to Alacant, Vizcaya to Bizkaia). However, these changes have not affected the boundaries of the territories concerned.

In contrast, the case is not as easy for other European countries. In part, there are obvious reasons for the necessity of modifications. One can name territorial changes due to wars and here in particular the changes caused due to World War I and World War II. These have importantly modified the territorial structure of some countries, split up former empires and created new countries. In these cases, the correspondence is made as best as possible between the historical regions and the current NUTS regions. The employed data are in many cases more disaggregated than the constructed NUTS level which allows to avoid possible biases that may arise from this approach.

All in all, this methodology has the important advantage that it is possible to make long-term comparisons at the regional level which are a key characteristic of this thesis.

2.5 Appendix

2.5.1 Tables

Table 2.1 Summary of differences between exogenous and endogenous growth models

Category	Exogenous	Endogenous	
	Augmented Solow model	Lucas model	Romer model
Human capital is accumulated by…	investing a fraction of income	spending a fraction of time acquiring skills	not modeled
Technology for production of human capital	same production function for C, K and H	separate sector for production of H using only human capital	not modeled
Role of human capital	input in production	input in production of Y and H	input in production of Y and A
Growth rate determined…	outside of the model	within the model	within the model
Determinant of long-run growth	Exogenous technological change	rate of human capital accumulation	stock of human capital
Effect of a permanent change in the variable governing the accumulation of human capital	level effect (relevant variable: s_H)	rate effect (relevant variable: 1-u*)	rate effect (though not explicitly modeled)
Effect of a one-off increase in the stock of human capital	level effect	level effect	rate effect

Source: Schütt (2003).

Table 2.2 Centripetal and centrifugal forces

Centripetal forces	Centrifugal forces
Linkages	Immobile factors
Thick markets	Land rent / commuting
Knowledge spillovers and other pure external economies	Congestion and other pure diseconomies

Source: Fujita and Krugman (2004), © RSAI 2004.

Table 2.3 Overview NUTS levels 1, 2 and 3 in the EU

Country	NUTS 1	NUTS 2	NUTS 3
BE	Gewesten / Régions	Provincies / Provinces	Arrondissementen / Arrondissements
BG	Райони (Rajoni)	Райони за планиране (Rajoni za planirane)	Области (Oblasti)
CZ	Území	Oblasti	Kraje
DK	-	Regioner	Landsdeler
DE	Länder	Regierungs-bezirke	Kreise
EE	-	-	Groups of Maakond
IE	-	Regions	Regional Authority Regions
GR	Γεωγραφική Ομάδα (Groups of development regions)	Περιφέρειες (Periferies)	Νομοί (Nomoi)
ES	Agrupacion de comunidades Autonomas	Comunidades y ciudades Autonomas	Provincias + islas + Ceuta, Melilla
FR	Z.E.A.T + DOM	Régions + DOM	Départements + DOM
IT	Gruppi di regioni	Regioni	Provincie
CY	-	-	-
LV	-	-	Statistiskie reģioni
LT	-	-	Apskritys
LU	-	-	-
HU	Statisztikai nagyrégiók	Tervezési-statisztikai régiók	Megyék + Budapest
MT	-	-	Gzejjer
NL	Landsdelen	Provincies	COROP regio's
AT	Gruppen von Bundesländern	Bundesländer	Gruppen von politischen Bezirken
PL	Regiony	Województwa	Podregiony
PT	Continente + Regioes autonomas	Comissaoes de Coordenaçao regional + Regioes autonomas	Grupos de Concelhos
RO	Macroregiuni	Regiuni	Judet + Bucuresti
SI	-	Kohezijske regije	Statistične regije
SK	-	Oblasti	Kraje
FI	Manner-Suomi, Ahvenananmaa / Fasta Finland, Åland	Suuralueet / Storområden	Maakunnat / Landskap
SE	Grupper av riksområden	Riksområden	Län
UK	Government Regions; Country	Counties (some grouped); Inner and Outer London; Groups of unitary authorities	Upper tier authorities or groups of lower tier authorities (unitary authorities or districts)

Note: DOM stands for French overseas departments.
Source: Eurostat (2011b); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Table 2.4 Overview NUTS levels 1, 2 and 3 in Candidate Countries and EFTA

Country	NUTS 1	NUTS 2	NUTS 3
ME	-	-	-
HR	-	Regija	Županija
MK	-	-	Статистички региони (Statistički regioni)
TR	Bölgeler	Alt bölgeler	Iller
IS	-	-	Hagskýrslu-svæði
LI	-	-	-
NO	-	Landsdeler	Fylker
CH	-	Gross-regionen Grandes regions Grandi regioni	Kantone Cantons Cantoni

Note: Eurostat emphasises that the MK code is provisional.
Source: Eurostat (2011c); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

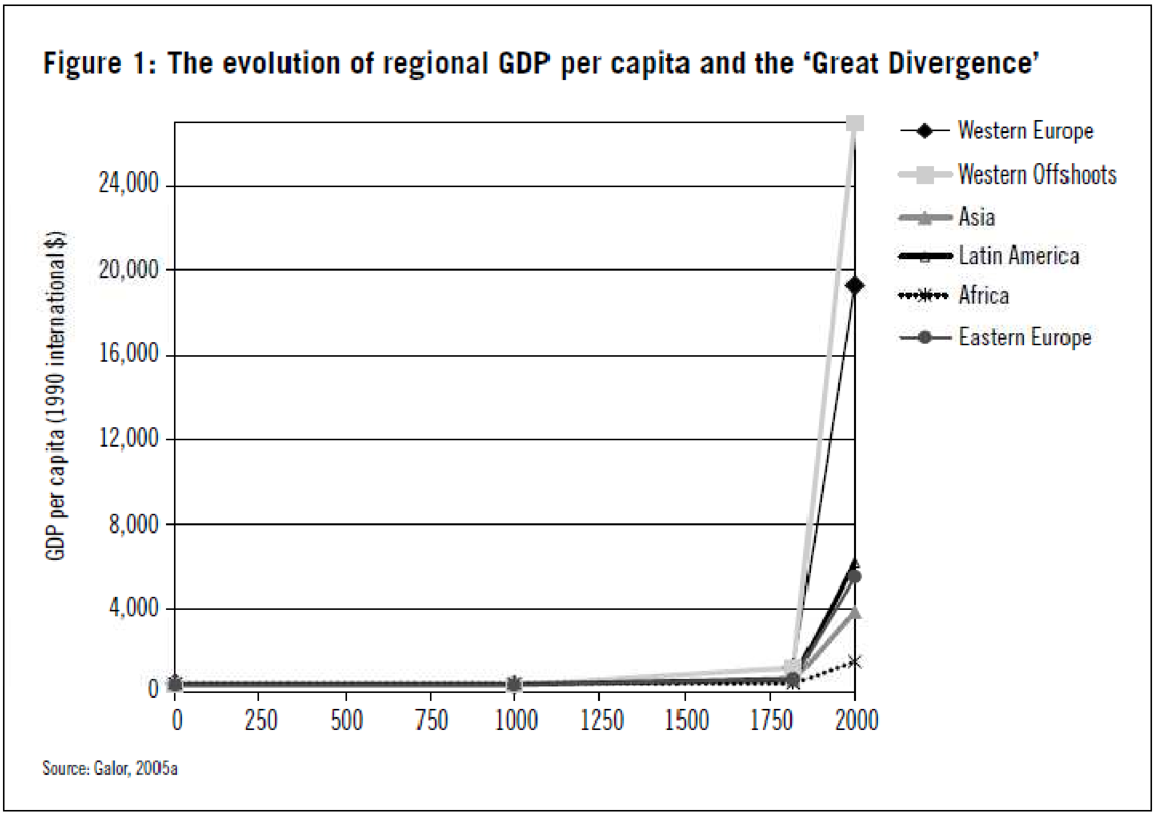
Table 2.5 Minima and maxima for population size

Level	Minimum	Maximum
NUTS 1	3 million	7 million
NUTS 2	800 000	3 million
NUTS 3	150 000	800 000

Source: Eurostat (2011d); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

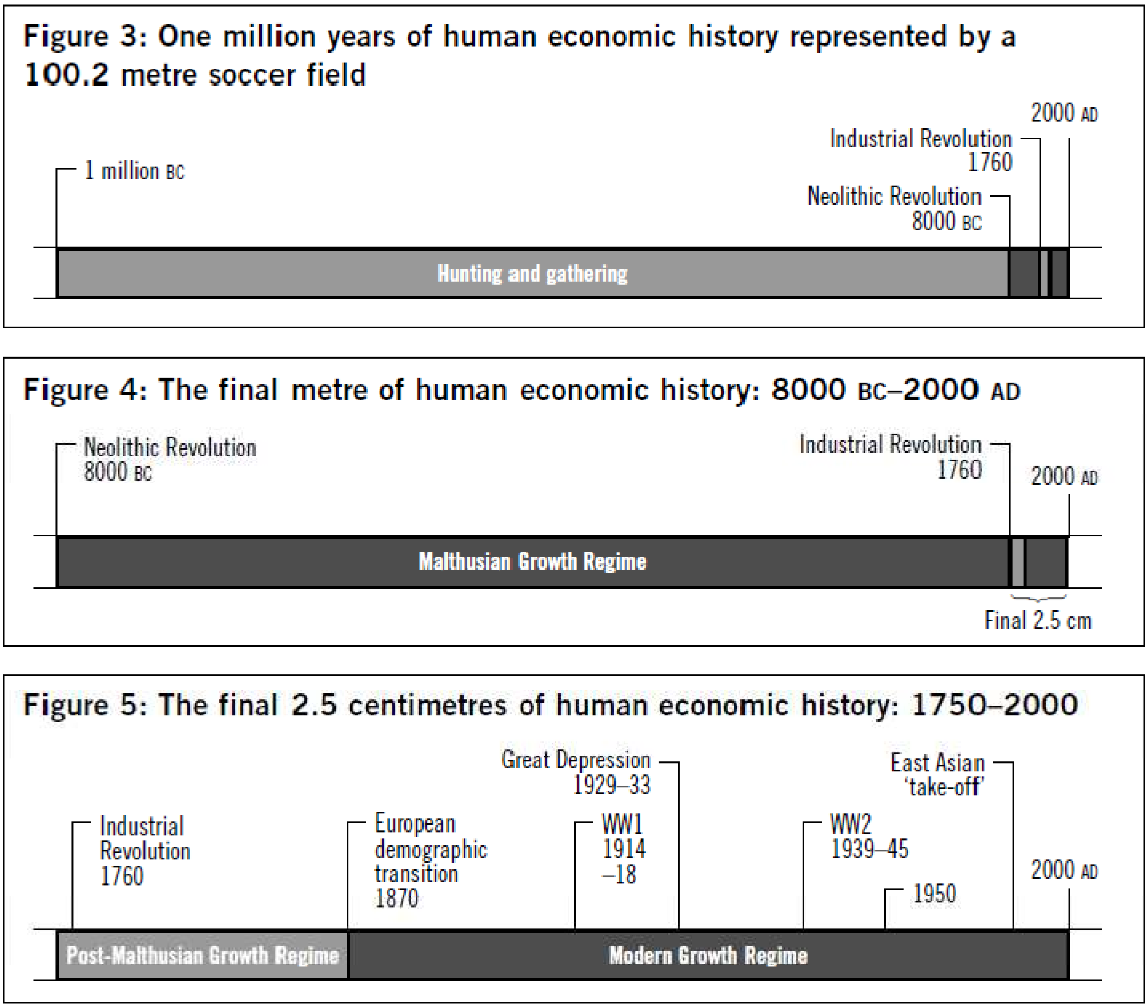
2.5.2 Figures

Figure 2.1 Long-run evolution of GDP per capita in the world



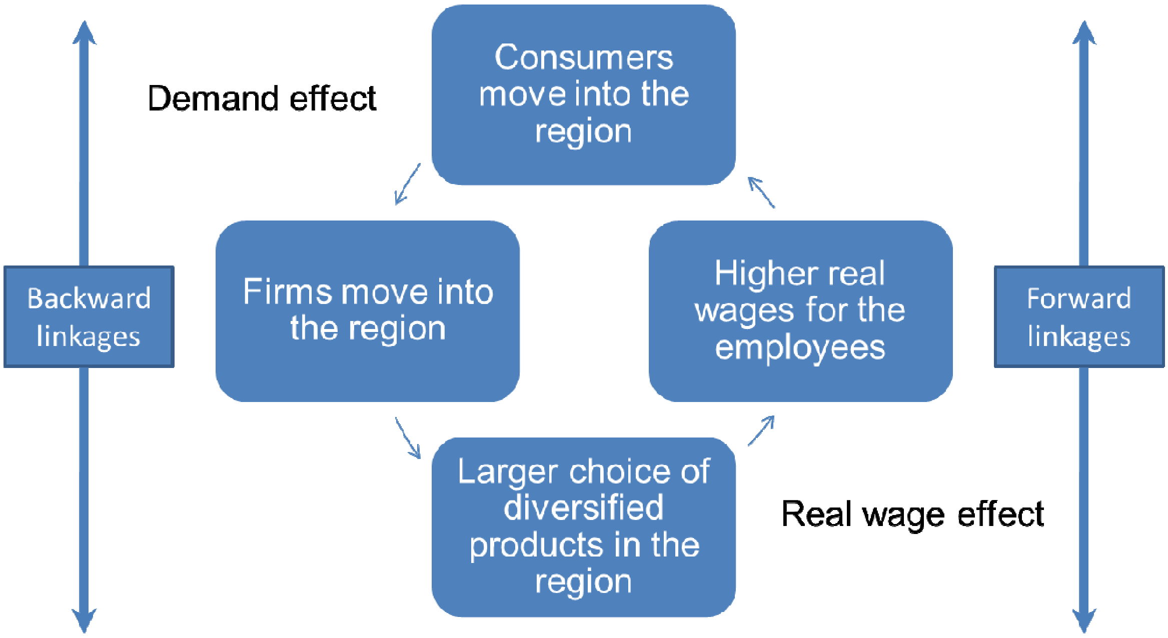
Source: Snowdon (2008).

Figure 2.2 Human history in the long run



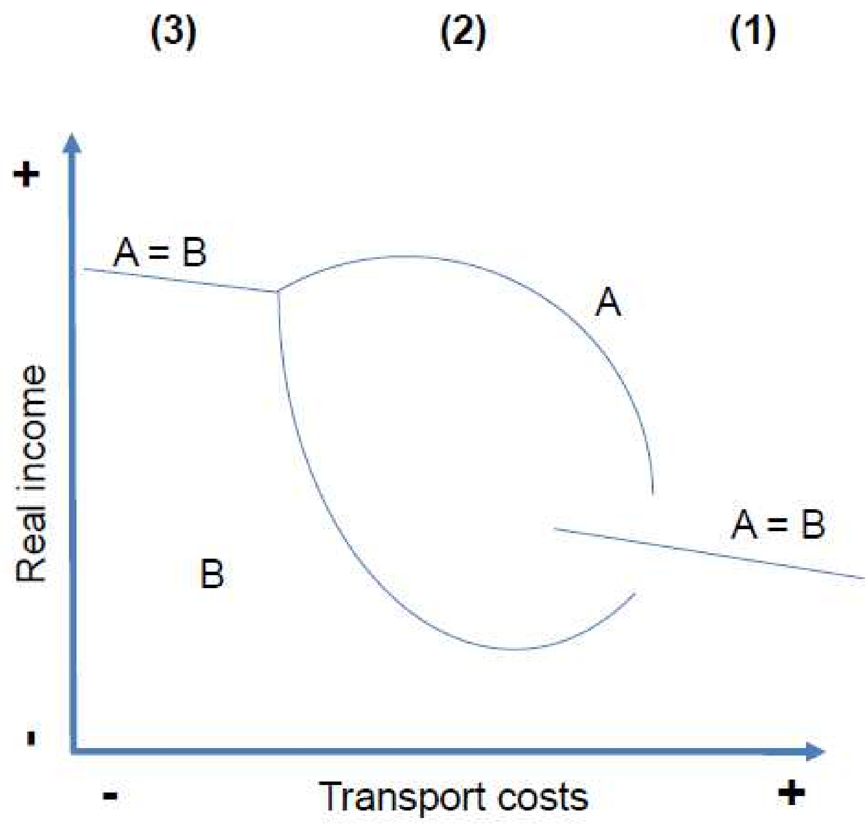
Source: Snowdon (2008).

Figure 2.3 Circular causation



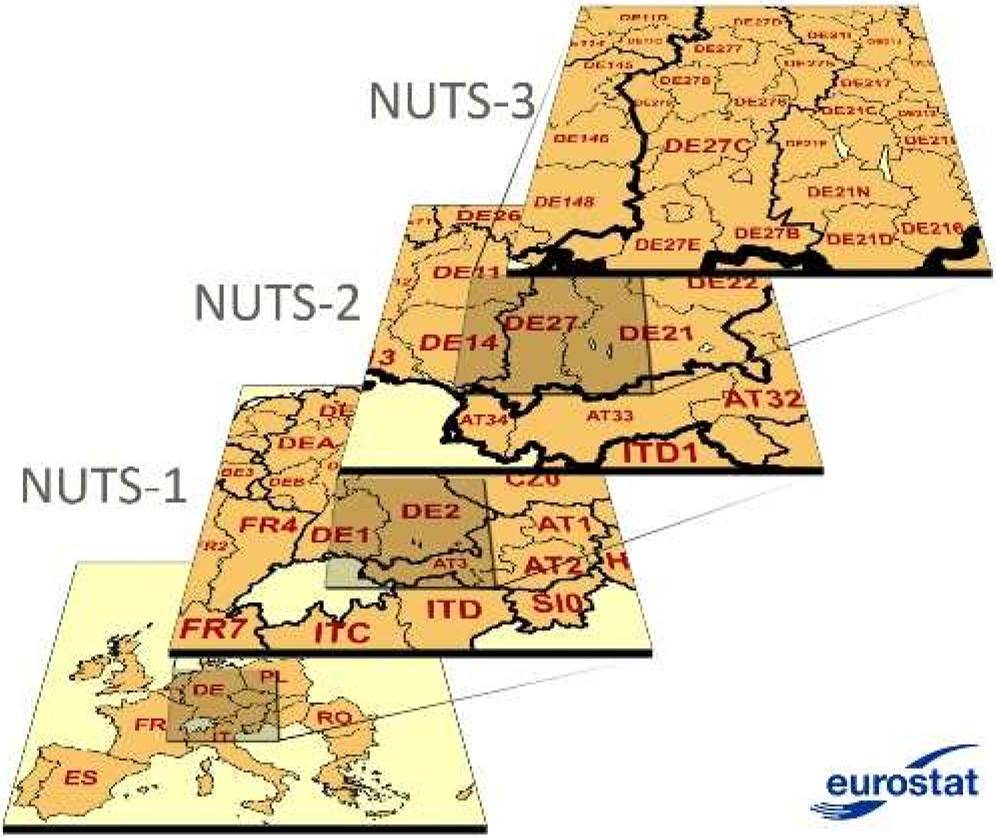
Source: Own presentation adapted from Eckey and Kosfeld (2004).

Figure 2.4 Transport costs and real income



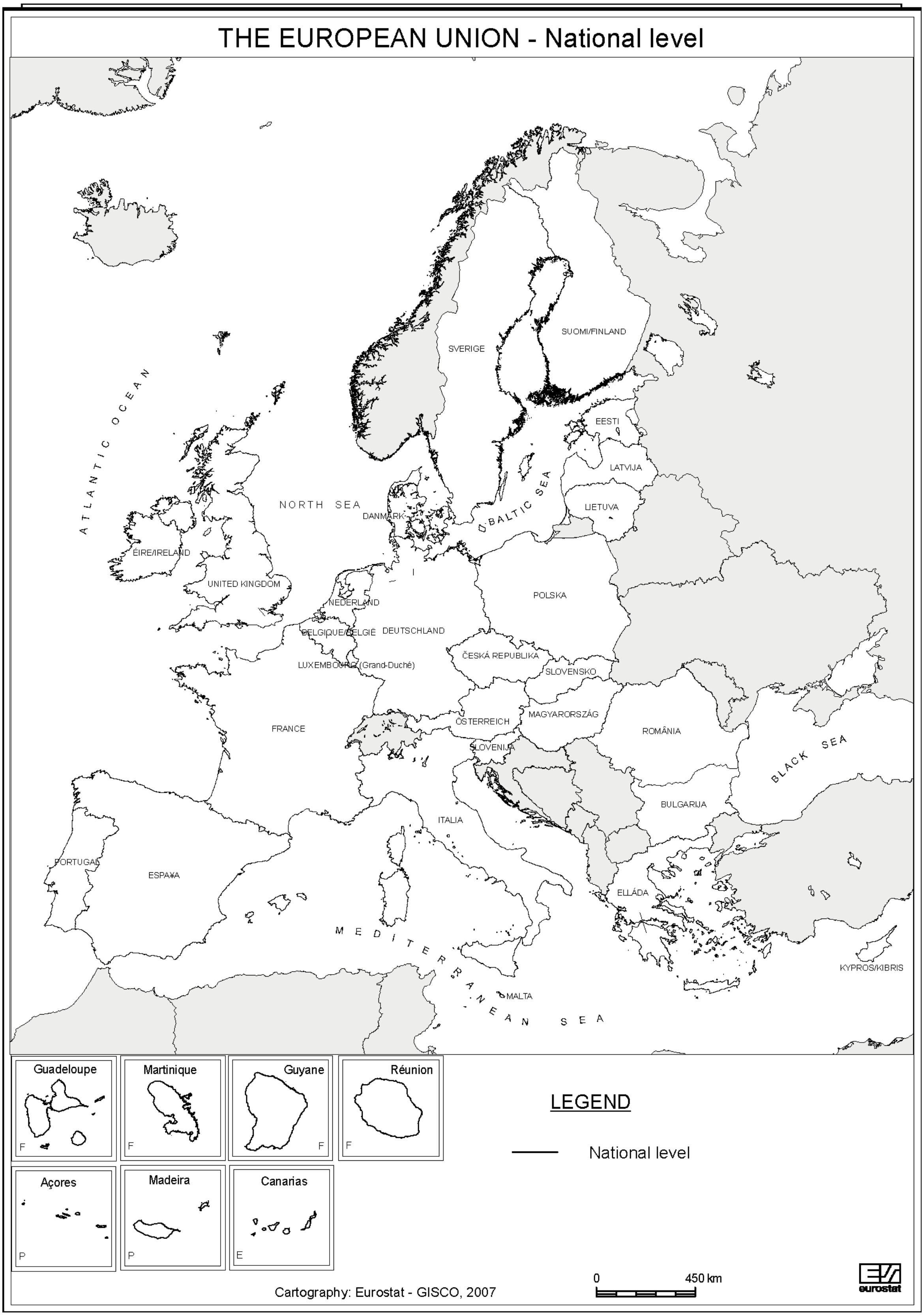
Source: Own presentation adapted from Combes *et al.* (2008), © OFCE 2008.

Figure 2.5 NUTS classification



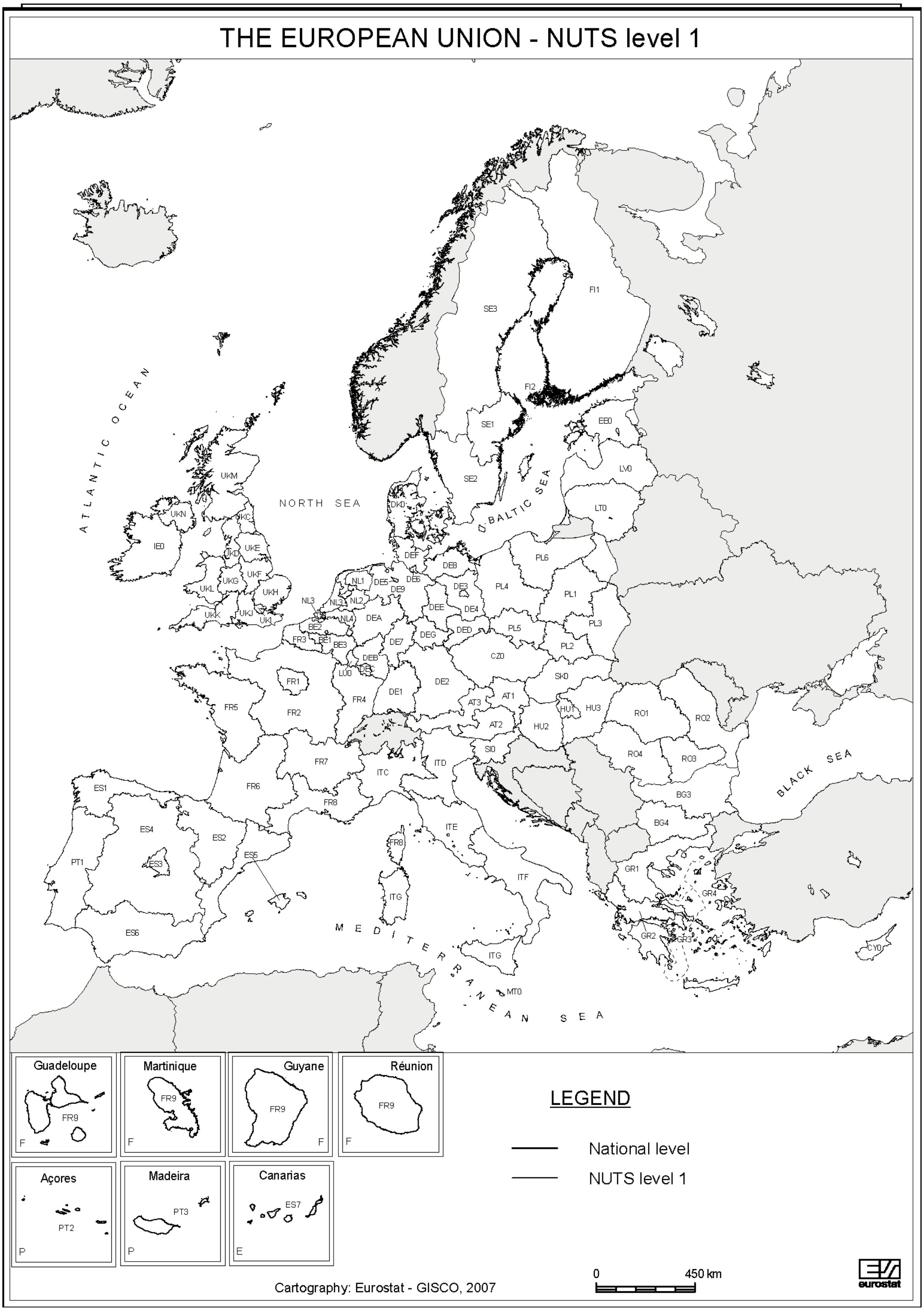
Source: Eurostat (2011a); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 2.6 NUTS 0 regions (EU)



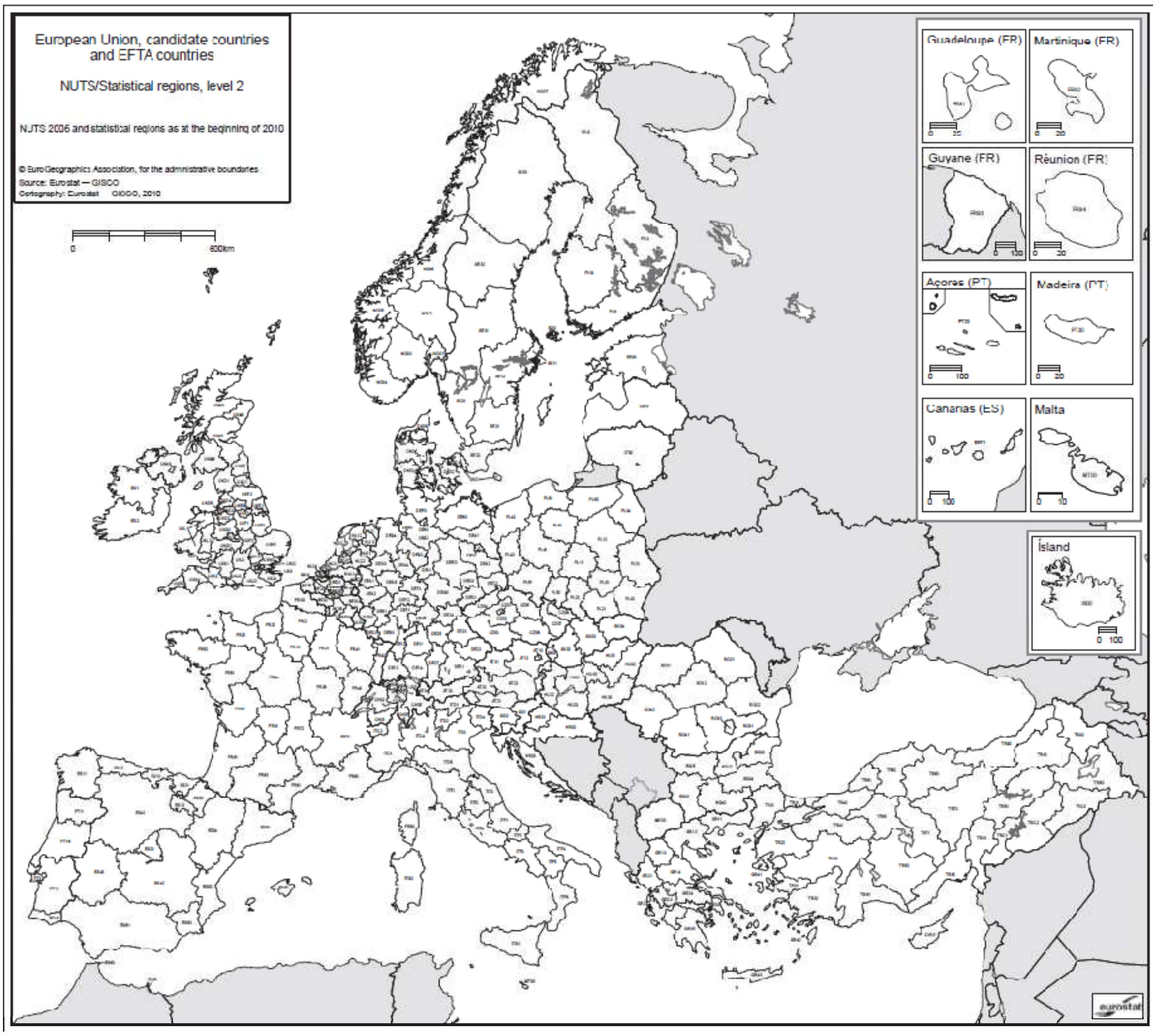
Source: Eurostat (2011e), © European Communities, 2007; http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 2.7 NUTS 1 regions (EU)



Source: Eurostat (2011e), © European Communities, 2007; http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 2.8 NUTS 2 regions (EU, Candidate Countries, EFTA)

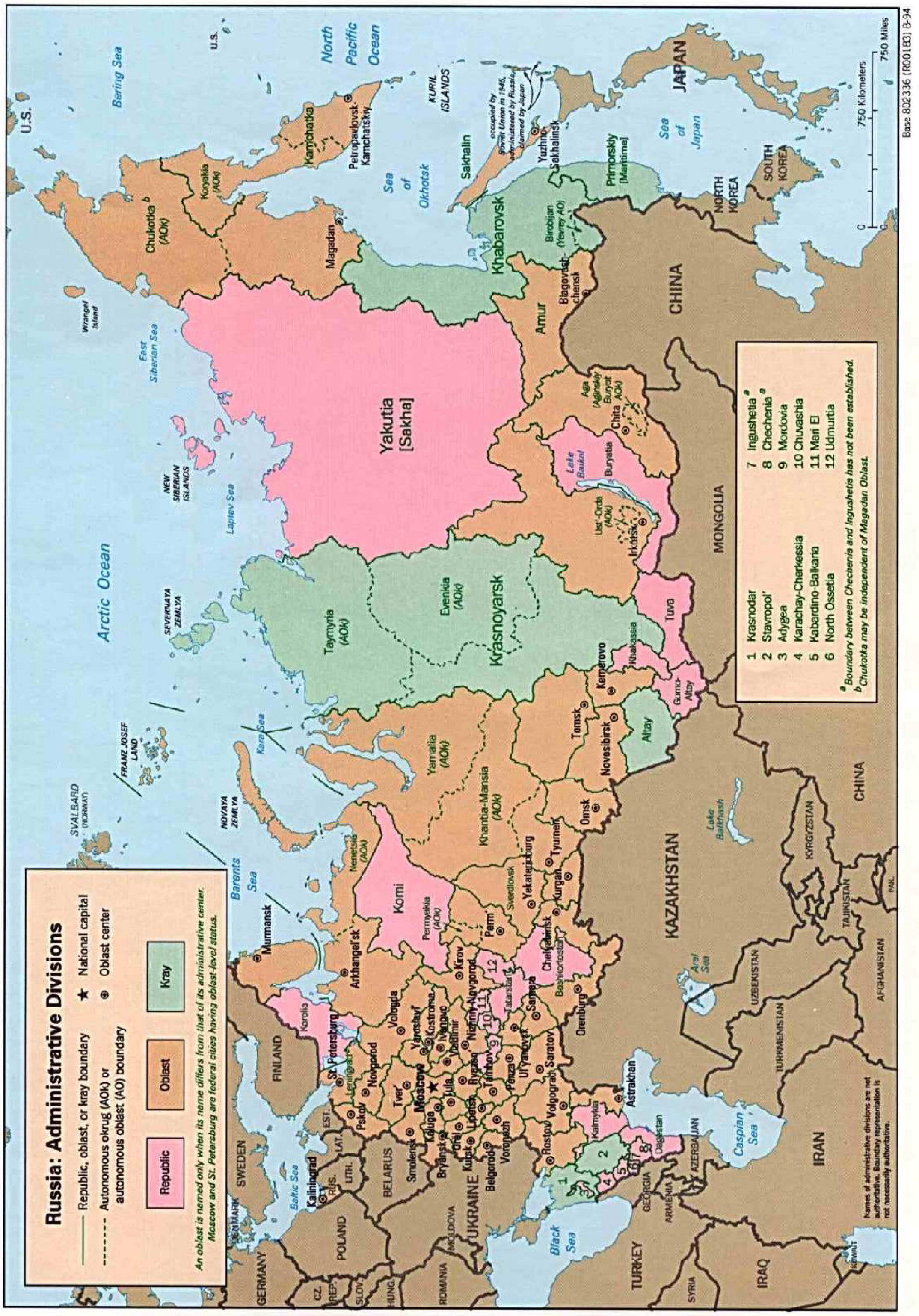


Source: Eurostat (2010); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 2.9 Administrative regions of Belarus

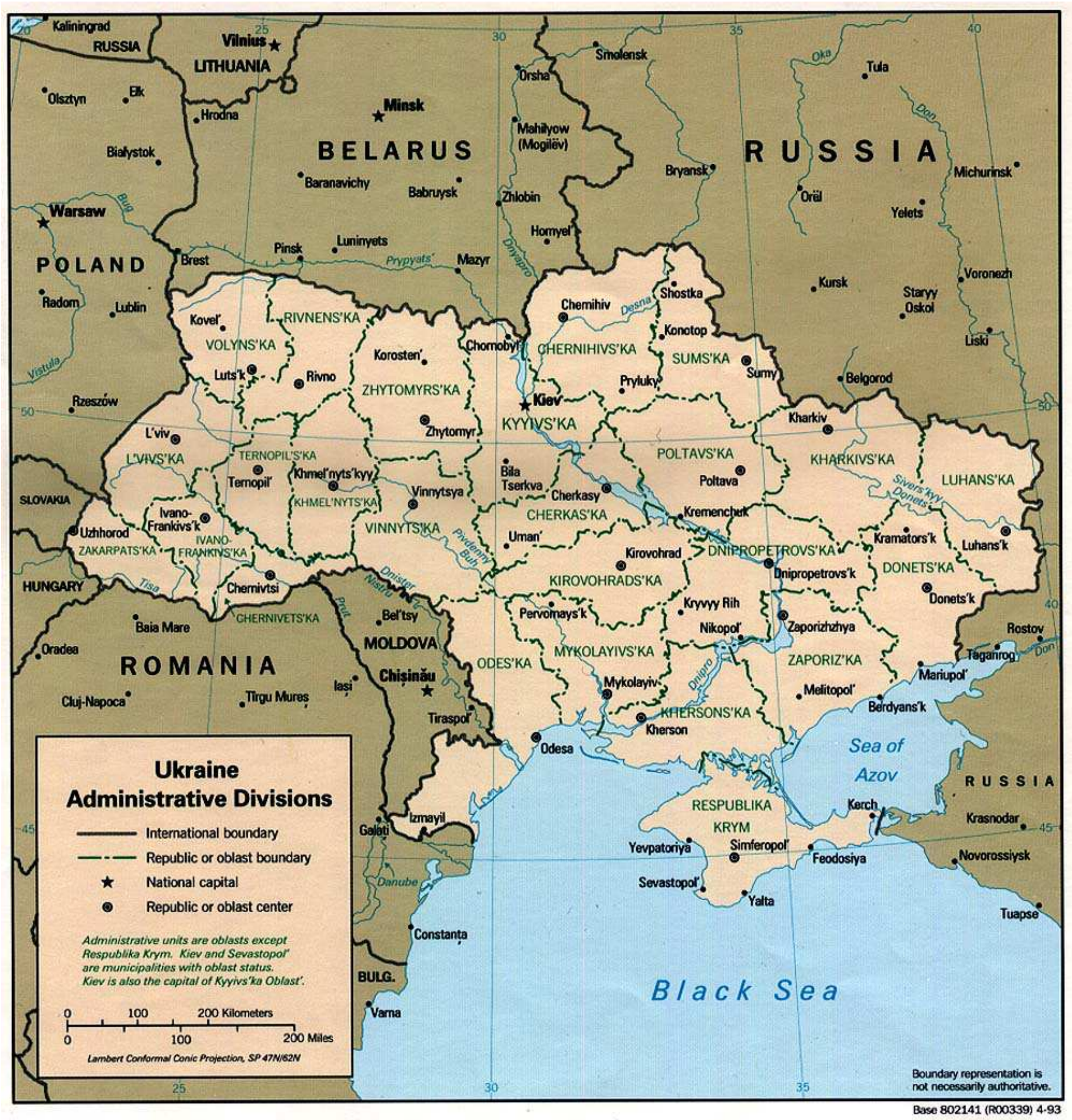
Source: University of Texas Libraries (2012a).

Figure 2.10 Administrative regions of Russia



Source: University of Texas Libraries (2012b).

Figure 2.11 Administrative regions of Ukraine



Source: University of Texas Libraries (2012c).

3. Are you NUTS? The factors of production and their long-run evolution in Europe from a regional perspective

Standard economic growth models generally consider different factors of production such as land, capital, labour, technology and human capital. These are common in theoretical models and empirical applications but more evidence is still needed for their long-term regional evolution. Therefore, this paper traces the evolution of specific aspects of these factors in the European regions and cities through the use of different proxies. The data have been collected and calculated from a wide range of diverse historical and spatial data bases. A particular feature is the definition of the European regions according to the NUTS classification of the European Union. Thus, the paper gives a rough outline of some of the most important long-term regional tendencies that should be taken into account in research directed to past and recent time periods.

An abridged version of this chapter has been published in *Historical Social Research*.

3.1 Introduction

Economic output is defined in the standard economics literature as a result of the interaction between different factors of production. The most common factors of production are land, capital, labour, technology and, more recently, human capital.[5] At least some of these factors are used in almost all theoretical economics models in this area and have been approximated using econometric applications in a variety of ways. Nevertheless, there is often a separating line between standard economics which commonly works with data from the last decades and economic history which is more concentrated in historical frameworks. In contrast, general and broad theories such as Unified Growth Theory (Galor 2005, Galor 2012) have enabled to get the full picture of economic development in the very long run.

On the other hand, economic geography theories such as New Economic Geography (e.g., Krugman 1991a, b, Fujita *et al.* 1999, Krugman 1999) explain the inequalities and convergence processes that characterise economic activity in space. They have also been applied to historical settings. Nevertheless, testing Unified Growth Theory models in particular needs more data for the long run, while New Economic Geography models rely on geographical and spatial components whose access has been progressing over the last years.[6] But there is still much more potential which has not yet been appropriately exploited.

For this reason, the aim of this paper is to give a rough descriptive outline of the various factors of production in Europe in a long-term regional perspective. Moreover, a

[5] Other factors of production are, for example, social capital (Felice 2012), entrepreneurship and natural resources (Baumol and Blinder 1991). See Xu *et al.* (2009) for a brief historical overview on the theories of the factors of production.

[6] In Europe, the EU has set up an initiative called Infrastructure for Spatial Information in the European Community (INSPIRE). It aims at providing more spatial data, even though primarily directed to the current time. The underlying directive was set up in 2007 and until now it is an ongoing project whose full implementation is obligatory by 2019 (European Commission 2012b).

particular feature of our approach is that we use common boundaries throughout time, enabling long-term regional comparisons. These boundaries are defined according to the Nomenclature for Territorial Units of Statistics (NUTS)[7] that has been developed by the European Union. It is the standard regional break down used for data in the current period, often provided by the official Statistical Office of the EU, Eurostat. In addition, we exploit city data to complement the picture on the spatial evolution of the factors of production. Although the employed data are in part standard in the literature, they have not been presented within such a common and combined framework. This feature may give new insights into the economic long-run evolution in Europe.

The paper is structured as follows: The first section gives a brief introduction to the factors of production and various economic growth models that rely on them. Afterwards, the methodology and the data are described in detail. Subsequently, we present the results of applying the methodology for the five most common factors of production: land, capital, labour, technological progress and human capital. A final conclusion sums up the paper.

3.2 Brief overview of economic growth models

Let us briefly take a look at the different economic growth models. Growth models may be categorised into exogenous and endogenous growth models (Schütt 2003). First, exogenous growth models do not explain growth in the model itself but growth is assumed to be given at an external rate. To name an example, neoclassical growth models *à la* Solow use the factors of production capital, labour and technological progress (Solow 1956). There are also varieties of these Solow models. For instance, the human capital

---

[7] The abbreviation comes from its French version, "*Nomenclature des Unités Territoriales Statistiques*".

augmented Solow model introduced explicitly human capital into the production function.[8] Second, endogenous growth models are different from exogenous growth models in the way that economic growth is not exogenously given but determined by the model. Because the rate of economic growth is very important and may be distinctive in different economies, this may allow an improved understanding of the dynamics behind economic growth.

There are two major different types of endogenous growth models, focussing either on human capital accumulation or technological change. Human capital accumulation is a strand that was initiated by the seminal work of Lucas (1988). In these models, growth is dependent on the maximum growth rate in the economy and the share of time that is spent by individuals on the acquisition of skills. Moreover, the second type of endogenous growth models was founded by Romer (1990). The importance of technological change for the creation of economic growth is underlined in these models. Economic growth depends, first, on the stock of ideas which expands as the variety of goods increases. Second, the existence of knowledge spillovers is assumed because all workers (here researchers) have equal access to this (increasing) stock (Schütt 2003).

The most recent contributions in the area of economic growth come from Unified Growth Theory and New Economic Geography. Unified Growth Theory aims at explaining economic growth in the very long run since the beginning of human kind. Therefore, it is a broad and flexible framework which can be adapted to special cases (Galor 2005a, Galor 2012). On the other hand, the goal of New Economic Geography is to understand why concentration of economic activity occurs in space (Krugman 1991a, b, Fujita *et al.* 1999). Concentration is a characteristic feature of economic activity and of human activity, which

[8] For more information on the link between economic growth and human capital, see Demeulemeester and Diebolt (2011).

is highlighted by economic hubs such as Silicon Valley and urban centres such as Tokyo. Both theories are helpful to consider both the very-long term and the spatial part of European development. Therefore, the information provided in this overview may contribute to bring together theory and facts.

3.3 Methodology and data

We use the NUTS classification which has progressively been developed by the European Union over the last decades.[9] This is a standard geographical break down of European space. More precisely, it includes the countries of the EU, EFTA and Candidate Countries of the EU.

It is defined by three basic principles which have been set up to improve comparability throughout Europe: population size, administrative divisions and geographical units (see Eurostat 2011d). What does this mean? The first principle is based on the criterion that the regions should have a similar population residing in them. For this reason, different thresholds have been defined which determine whether a region is classified as being a NUTS 1, 2 or 3 region (NUTS 3 from 150.000 to 800.000, NUTS 2 from 800.000 to 3 million and NUTS 1 from 3 to 7 million inhabitants). The second principle favours the use of already existing administrative regions. This rule makes data collection much easier because the national authorities already provide the data at the corresponding regional break down. Thus, there is no difference between the European classification and the national one, saving effort and avoiding biases and contradictions between different classification systems. Third, geographical units should not be defined according to a certain category but by general terms. This principle also improves the comparability of these regions.

[9] In this paper we always refer to the NUTS 2006 classification.

In this way, Europe is categorised into several layers. The lowest level of regional aggregation is the NUTS 0 level, i.e., the country level. The second level is NUTS 1, which in most cases corresponds to the first regional level. For small countries such as Luxembourg or Malta, the NUTS 1 level remains identical to the country level. The same principles apply to NUTS 2 and NUTS 3, each time dividing more aggregated regions into their constituting subregions. This means that e.g. a NUTS 2 region is always perfectly made up by (one or) several NUTS 3 regions, allowing a smooth change from one NUTS level to the next.

Because the NUTS classification is only available for the EU, EFTA and Candidate Countries to the EU, we use the current regional administrative units for the other European countries (in particular in eastern Europe).

To perform the analyses of this paper, a very broad range of sources have been used. First, we have used specialised publications that treat one (or more) particular aspect(s) in this paper. For example, we use city data by Bairoch *et al.* (1988) and data on universities by Rüegg (2004). These are data referring to spatial data points, that is cities or the locations of universities. Second, we use available data on historical regional development. For instance, Coale and Watkins (1986) provide data on fertility and marital status. The historical regions were brought into correspondence with current NUTS regions as best as is possible in such exercises. Third, we used spatial data (such as raster data) whose information has been averaged for each NUTS region. An example are altitude data provided by Hijmans *et al.* (2005). By combining these various methods and data sources, we get insights into the evolution of regions that have constituted Europe today and in the past. Note that the purpose of this paper is only to give a rough outline of the different factors and of their evolution. Therefore, it is neither intended nor attempted to explain regional differences or their evolution which would be beyond the limits of this paper. We have to limit ourselves to highlighting some particular striking aspects of the considered variables in time and space.

3.4 Results

3.4.1 Land

3.4.1.1 Altitude

Land is constituted by different aspects. One element that may be considered is altitude. Altitude affects economic actors in different ways. First, it increases the cost of transportation because it is more costly and time consuming to transport goods or people to locations that are on a higher altitude level. In this way, second, mountains may also constitute natural barriers. These barriers may define the limits of a territory (e.g., the Pyrenees for the Franco-Spanish border) and increase the likelihood of separate social, cultural and economic developments because they complicate exchange and communication between different peoples.

A map of average altitude in the European regions at the NUTS level is presented in Figure 3.1. The Alps are clearly visible. The fact that Spain is Europe's second country in altitude (behind Switzerland) can be easily guessed. Important parts of Norway, the Balkans and the Caucasus region are also located on rather mountainous areas. In contrast, low altitude characterises in particular England, northern Germany, Denmark, the Netherlands and many parts of European Russia.

3.4.1.2 Temperatures

Temperatures are also natural conditions that have been used in different publications explaining different growth patterns. Moreover, the soil of land is heavily affected by climatic conditions. Thus, temperatures have been used in the literature as instrumental variables (e.g., Galor *et al.* 2009) which underlines the appropriateness of considering them in the present context.

In international comparisons, temperature may be linked to the distance to the equator, the existence of deserts, tropical and subtropical weather conditions. The differences in Europe are not as high as in a worldwide perspective but still remarkable and important for shaping local culture and local economies.

Taking a look at Figure 3.2, one can see that the more southern a region is located, the higher are the average annual temperatures. However, the Alpine regions have lower values due to their high altitude. This observation can also be made e.g. for the Carpathian regions.

However, it may be interesting to ask how these temperatures have been evolving in the past. Given the context and importance of today's global warming discussions a clearer insight into past changes might be useful. For example, one may consider Figure 3.3, showing the change in the average annual temperature between 1930 and 1960. Note that the data are not calculated for individual regions this time. The average change in temperature in that period is between -0.5 and 0.0°C for most of central Europe. This means that there was no change or the temperature decreased on average. Moreover, at the outskirts of Europe, i.e., broadly speaking the Iberian Peninsula and Belarus, Ukraine, Russia and a part of the Baltic region, the changes were 0.0 to 0.5°C, that is there was no change or an increase in average annual temperature. The only regions were temperatures changed more than 0.5°C were Iceland and the very north of Russia (-1 to -0.5°C) and the very eastern parts of European Russia (0.5 to 1°C).

Moreover, Figure 3.4 shows the average annual temperatures between 1900 and 1910. Unfortunately, the available data do not take into account an important part of European Russia. No important changes are visible in such an overall map in comparison to current data but differences exist at the local level which cannot be explained in detail within the limits of this paper.

An additional insight can be obtained from Figure 3.5, representing the daily average temperature range in Europe between 1901 and 1910. There are some differences worth to be highlighted. The average daily temperature ranges are less important in Italy than in the Iberian Peninsula or the Balkans, although the average annual temperatures are similar. The Iberian Peninsula and important parts of Bulgaria and Romania have the maximum values in this category. Central European regions (e.g., in Germany and Denmark) have lower temperature ranges as their Nordic neighbours (e.g., Sweden).

3.4.1.3 Precipitation

Precipitation has also been used by recent publications as an instrumental variable (e.g., Galor *et al.* 2009). We proceed the same way as before. See Figure 3.6 for average annual precipitation between 1950 and 2000. In each case, highest precipitation values come from different geographic areas. The first group consists of regions with a direct contact with the Gulf Stream (e.g., Iceland, Ireland, Scotland, Norway, northern parts of Spain and Portugal). Second, a further area comprises the mountains in the Alpine region down to Greece. Low precipitation characterises many regions in Spain and in the south-east of European Russia.

In sum, the maps show that land conditions vary quite significantly in the European regions. These differences have an important impact on regional economic and social development.

3.4.2 Capital and labour

Unfortunately, there is a "lack of long-run series for capital across countries" (Prados de la Escosura and Rosés 2010, p. 141), in Europe and elsewhere. Recent efforts have been made for some European countries (e.g., Földvari and van Leeuwen 2010, Prados de la Escosura and Rosés 2010). Still, if this fact is true for series across countries, this conclusion applies even more so to regions. Therefore, we are not able to directly show capital accumulation in the European regions but have to be content with much more general notions. For this reason, we put both capital and labour in the same section. First, we present the overall economic development today and in the past before considering economic proxies and labour characteristics in closer terms.

3.4.2.1 General regional economic development, today and in the recent past

Important geographic disparities in economic activity have always existed in Europe and are still clearly observable. The explanation of these spatial regimes is a major challenge for academic researchers and policy makers alike. Therefore, EU policy attributes an important share of its budget to regional funds (i.e., the European Regional Development Fund (ERDF), the Cohesion Fund and the European Social Fund (ESF)) (European Commission 2008). In total, the overall Cohesion Policy for the period 2007-2013 has a budget of € 347.4 billion. This budget is spent on the 'Convergence objective' (81.5 %), the 'Regional Competitiveness and Employment' objective (16 %) and the 'European Territorial Co-operation' objective (2.5 %) (European Commission 2008). These funds are allocated to the regions most in need of economic progress (Figure 3.7). Globally, one can state that these funds are primarily focusing on countries that formerly constituted the eastern Communist countries (e.g., East Germany, Poland, Romania) or on countries in southern Europe (i.e., Greece, southern Italy, parts of Spain and most regions in Portugal).

This general distinction between two large groups of regions is helpful but not sufficient. There are also major differences within such groups. Therefore, one should consider alternative measures. GDP per capita is an obvious candidate. As shown in Figure 3.8, a cluster including the regions with the highest GDP is located in northern Italy, western Austria, southern Germany, the Benelux countries, the Greater London region and southern Ireland. Other more dispersed regions have also high GDP levels, such as southern Finland, the Stockholm region, the Madrid, Basque and Navarra regions, the Greater Paris region and eastern Scotland. It is apparent that within country differences are quite important in the European countries today, which is why regional funds will also be attributed in the future (see also Figure 3.9 and Figure 3.10).

Going a bit more back in time, the spatial concentration of GDP has evolved rather slowly but geographic concentration was still advancing in the period between 1984 and 1999 (López-Rodríguez *et al.* 2007, López-Rodríguez 2008). Nevertheless, there might be an overall convergence process underway, as the comparison of the year 2008 with 2000 suggests (Figure 3.11). Most of the regions that have a low GDP per inhabitant have higher growth rates than the richer regions. However, the current financial and debt crisis may threaten this convergence process as some countries particularly in southern Europe have to cope with major economic fallbacks.

3.4.2.2 General regional economic development in the long run

In addition, our aim is to consider the long-term evolution of economic activity in the European regions. To this end, one may refer to the concept of the Blue Banana (Brunet 2002, see also Hippe and Barmeyer 2009). This economic area has had a very import impact on economic development in Europe – and not only since a few decades. In fact, Braudel (1979) estimates that this region in central Europe was already crucial in the 13th century. Therborn (1995) even believes that it goes back to the 9th century. In sum, "[t]he core of European industrial societies proves to be almost identical with the city belt that dominated economic development in pre-industrial Europe" (Heidenreich 1998, p. 315).

Where is this area located and what are its origins? Figure 3.12 roughly highlights the basic concept (see Brunet 2002). Products from the eastern Mediterranean arrived in the commercial trade cities of northern Italy. There were two major possibilities to transport them to the consumers of north-western Europe: first, shipping the goods through the Mediterranean Sea all around the coasts of France and the Iberian Peninsula. The alternative was to transport them by means of rivers through central Europe. The Rhine has always constituted a natural trading route, being the second longest river in Europe and the longest in western Europe.[10] Subsequently, the Hanse transported these commodities to north-eastern Europe. For these reasons, the Rhine was an important means of transport and enabled the local communities to become rich trading cities. This, in turn, attracted individuals from other areas which led to the growth of the local population. Until today, the Rhine is a major transport axis and the most important waterway for trade in Europe. This is one reason why we find a banana in population density in Europe in the past (1930, Figure 3.13) and still in the present (Figure 3.14). These important populations in the Blue Banana area lead today to a prominent spatial distribution of the European population.

Therefore, the Blue Banana has been the driver of economic growth in Europe. In consequence, it has also spurred innovation and technological process over the last centuries, as Zündorf states: "[w]herever and whenever innovations occurred in Europe, imitators soon appeared who provoked a competitive struggle resulting in a rapid diffusion of these changes. Nowhere was this diffusion of changes faster and more lasting than in the blue banana that has always had the best developed lines of communication through a dense girdle of cities" (Zündorf 1997, p. 244; translated by Heidenreich 1998, p. 315).

These aspects make it particularly appropriate to make reference to the Blue Banana in a long term approach. Still, estimations have been made that this area has become and will become less important in the future. The decline of coal industries over the last decades which were located close to the Rhine has contributed to this tendency.

[10] The longest European river is the Danube.

Moreover, other geographic constructions such as the Mediterranean Arc or the Atlantic Arc have shown other geographical areas that have seen particular growth schemes due to improved European cooperation (Ministère des Affaires Etrangères 2006).

3.4.2.3 Measures of regional economic development in the long run

We can get an intuition of regional economic development in the past when considering the share of individuals dependent on agriculture in 1930 (Figure 3.15). This variable can be taken as a rough proxy for industrialisation. More specifically, a higher share stands for a lower degree of industrialisation. However, we are not able to distinguish between the industry and service sectors which may bias the results. Still, it is evident that most British regions are highly industrialised. In other western European countries we find several important regions with a low share of individuals dependent on agriculture as well, such as in Belgium, northern France and different parts of Germany and Scandinavia. But the Greater Athens region is also an important geographic outlier. In contrast, most regions in the Balkans (in particular Yugoslavia) and the USSR are highly based on agriculture. Only Moscow as the capital is more industrialised (and based on services).

Taking a closer glance at the agricultural sector as such, it is also possible to consider its per capita productivity in 1930 (Figure 3.16). In general, those regions that have a low share of individuals dependent on agriculture have also a high productivity. This result may come from the fact that less individuals are employed in the agricultural sector but those are very productive because they have better technologies at their disposal. Therefore, the most productive regions in agriculture are located in the UK, the Netherlands, Denmark, southern Belgium, northern France, northern Germany and Saxony. The USSR, Yugoslavia and Albania are the most important countries with low per capita agricultural productivity.

For much longer time spans, data are available for the economic prosperity of individuals living in European cities. One standard indicator are real wages of workers. Real wages are obtained by dividing nominal wages by an index for consumer prices. Alternatively, Allen (2001) also calculates welfare ratios which are based on real wages. These welfare ratios are defined as "average annual earnings divided by the cost of a poverty line consumption bundle for a family. A welfare ratio greater than one indicates an income above the poverty line, while a ratio less than one means the family is in poverty" (Allen 2001, p. 425).[11] Therefore, they give an indication of the relative poverty or richness of the individuals working in a city. Allen (2012) offers data for both real wages and welfare ratios from 1300 to 1900 on two different types of workers: building labourers and craftsmen (see Figure 3.17 to Figure 3.32). In general, craftsmen gain more than building labourers.

The data on building labourers indicate that the welfare ratios are lower in many European cities around 1500 than during the two centuries before. In most cities in the 16th century, the welfare ratios for the labourers are below the poverty line. Only in London, Anvers, Amsterdam, Warsaw and Vienna are the ratios above 1. The economic situation of building labourers does not improve over the following centuries; in fact, the contrary was often the case. The only cities above the poverty line are London, Anvers and Amsterdam during the 17th and the 18th century. Welfare ratios increase above the line in the 19th century, Paris and Warsaw joining the other three cities. However, building labourers in other cities in central Europe and southern Europe still live below the poverty line at the beginning of the 20th century.

[11] More specifically, Allen clarifies that this poverty line "is computed for a notional family of a man, a women, and two children, and the nonhousing component of their poverty line income is set equal to three times [a specific] basket of goods […] [which] provides only 1941 calories per day. […] It was possible to get by on less […] but this level of calories and variety of consumption mark a line between respectability and destitution" (Allen 2001, p. 425/426).

Similar results apply to craftsmen with the difference that craftsmen have higher real wages and their welfare ratios are mostly above the poverty line. Welfare ratios are often the highest in London and Anvers throughout time. Craftsmen in different German cities and later also in Italy see their welfare ratios decrease below the line. Overall, there are striking differences in the welfare ratios of the various European cities whose evolution diverged particularly between the 16th and the 18th century. At the end of the period, i.e., the beginning of the 19th century, welfare ratios increased almost everywhere and permitted higher real wages than over the past centuries.

3.4.2.4 Commodity prices (wheat)

Commodity prices play a fundamental role in our understanding of the evolution of the European economy and trade. The commodity which has (probably) obtained the highest degree of attention is grain in general and wheat in particular. Grain was an important consumed good before the 20th century and was a key product in European trade. For these reasons, data are much more available for this good than on others, allowing to trace its long-term evolution.

Thus, for the long run before 1800, the database by Allen and Unger (2012) provides a good starting point (Figure 3.33 to Figure 3.38). Note that only those cities are depicted where enough data are available for the considered time span.[12] In the 15th century, grain prices appear to be higher in Belgium and the Netherlands than in other European cities. The overall evolution from the 16th to the 17th century is one of increasing grain prices in most cities. The changes to the 18th century are much more ambiguous. In

[12] Data are available yearly and are averaged by larger time spans. We excluded those cities for each time span where the data are available for less than 7 years for the time span of 100 years between 1300 and 1800 and 25 years between 1800 and 1899. We took this threshold to avoid a possible bias resulting of the inclusion of only a part of a typical business cycle. Taking the assumption that a typical business cycle has a length of at least 7 years, we avoid this bias to an important part while retaining a maximum of the available data.

many French cities grain prices fall, while they remain constant in the UK and rise in cities such as Naples. Finally, the first 25 years of the 19th century see a sharp increase in grain prices almost everywhere in Europe. Moreover, wheat prices in eastern Europe are generally lower than in western Europe.

The evolution of wheat prices in a higher number of European cities between 1800 and 1913 is further highlighted in Figure 3.39 to Figure 3.43, using Jacks' database (Jacks 2005, 2006).[13] The relative price positions of the European cities to each other is generally similar to that shown using the Allen and Unger (2012) database, although the inclusion of cities is in many cases different (compare Figure 3.38 to Figure 3.39). Note also that another definition of wheat prices is used here.

In general, wheat prices fell throughout the first half of the 19th century and also over the first decades of the 20th century. However, there is a clear increase of wheat prices in the period 1850 to 1874 when comparing it to 1825 to 1849. Thus, although the overall trend was a decrease in wheat prices, important variations characterised the 19th century. These variations would be even more striking when using yearly or monthly data instead of the proposed 25 year averages (see Jacks 2005, 2006 for a detailed analysis).

3.4.2.5 Population

We consider different characteristics of regional and spatial population distribution. The first one is urbanisation. Urbanisation rates have been used in order to approximate economic development in the long run by authors such as Acemoglu *et al.* (Acemoglu *et al.* 2002, Acemoglu *et al.* 2005a). We present here the spatial distribution of these urbanised areas, that is the location and size of cities in the European space from 800 to

[13] Data are available monthly and are averaged by years and by larger time spans. We excluded those cities for each time span where the data are available for less than 7 years for the time span of 25 years between 1800 and 1899 and for the time span of 14 years between 1900 and 1913. This criterion follows the same principle as for the Allen and Unger (2012) database to improved comparability.

1850. The data stem from Bairoch *et al.* (1988) (Figure 3.44 to Figure 3.54). The threshold for inclusion in the database is a minimum of 5000 inhabitants at some point in time in the period.

Let us very briefly and roughly sketch some important changes during the millennium. At the beginning of our time period (800 AD), the largest European city is located in the southern part of Europe, i.e., in Spain under Muslim domination: Cordoba. By the year 1000, it is joined by another important city under Muslim rule, Palermo. This scheme remains for two centuries, until in 1200 the population of Paris is of similar size as some of its southern counterparts. In western Europe, the most important areas of agglomerations are located in northern Italy, northern France, Belgium, western Germany and in the very south of Spain and Italy. Thus, the focal point of the European economy shifts more and more to the north, "to an axis from the Low countries to Lombardy" (De Long and Shleifer 1993, p. 677). The entire eastern half of Europe (from central Germany to the east) is characterised by very few cities over the entire period which is why the urban population is mostly concentrated in western Europe.

This general distribution remains in existence for the next centuries with the exception that the southern agglomerations become less important and Paris grows to become the largest city of Europe. Moreover, Naples becomes an increasingly important city. London grows quickly and overtakes Paris in 1700. The dominance of London becomes more pronounced the more the Industrial Revolution advances until 1850. In eastern Europe, many more small cities are established. In addition, the spectacular growth of St. Petersburg is particularly impressive after its creation in 1703, overtaking Moscow by 1850.

To get an insight into the rather recent distribution of the population, we use Moriconi-Ebrard's (1994) dataset on European agglomerations and show their distribution in 1950 and 1990 (Figure 3.55 to Figure 3.56).[14] The general pattern from 1850 has remained intact but urbanisation is progressing even further, so that most agglomerations are becoming larger and larger. The most important agglomeration becomes Moscow in 1990, overtaking the former Russian capital of St. Petersburg and other major European capitals such as London (largest agglomeration in 1950) and Paris.

Another aspect has attracted the interest of researchers in the last decades and particularly in the last years: fertility. In fact, different fertility patterns have existed in Europe in the past. In most of Europe, a particular European Marriage Pattern (EMP) prevailed. The European Marriage Pattern refers to the regions west from an imagined line going from Russian St. Petersburg to Italian Trieste – the famous Hajnal line (Hajnal 1965, Hajnal 1982). In this area, women were married not after having become fertile but often many years later so that their average age at marriage was relatively high. After marriage no restrictions were set on fertility (Voigtländer and Voth 2009). This peculiar way of fertility limitation did not exist in the east of Europe.

Note that there are three conditions for fertility decline (in particular marital fertility) (Coale 1973). First, fertility has to be determined by individuals voluntarily, so that they can choose the number of their children. Second, it has to appear advantageous to have fewer children than the generations before. And third, it is important that appropriate and effective techniques are available to individuals to reduce fertility.

Recent growth models also take into account fertility. Unified Growth Theory, among others, explicitly refers to the demographic transition (e.g., Galor 2005a, Galor 2012). The well-known Quantity-Quality Trade-Off models a decision that parents have to take (e.g., Guinnane 2008, Bleakley and Lange 2009, Becker *et al.* 2010, Becker *et al.*

[14] The dataset includes agglomerations of at least 10000 inhabitants. See Moriconi-Ebrard (1994) for more details.

2012). Parents have to choose to invest either in the quantity (number) or the quality (education) of their children, being limited by time constraints. The increased importance of human capital for future earnings of children is seen to be an essential factor for the demographic transition and ultimately economic growth.

Related to fertility is also infant mortality. Falls in fertility are often accompanied by reductions in infant mortality, which is why there is still an ongoing discussion on the interrelatedness of these factors (see e.g., Eastwood and Lipton 2011, Galor 2012). For example, the data tell us that infant mortality was particularly high in eastern Europe during the interwar period (see Figure 3.57 for data in 1930-31). But also in the Iberian Peninsula, southern Italy and parts of Bavaria infant mortality was relatively important. The most advanced regions came from Scandinavia, England, the Netherlands, Switzerland and France.

A major project on the creation of a European fertility database was the European Fertility Project (EFP) (see Coale and Watkins 1986 for more information). It created an abundant collection of fertility data for Europe in the 19th and 20th century. We use three indicators to get an insight into the regional distribution of fertility and marriage behaviour in Europe between 1870 and 1960, the most important time span during which the European Demographic Transition took place. First, total fertility is defined as "a measure of the fertility of all women in the population" (Coale and Treadway 1986, p. 154). The same principle also applies to marital fertility, considering the fertility of married women. Second, the marital status is "the ratio of the number of births produced by married women in […] a population to to the number that would be produced if all women were married" (Coale and Treadway 1986, p. 154). Watkins clarifies that this indicator represents "the proportions married at each age" (Watkins 1986, p. 315) and can thus be used as a proxy for nuptiality.

We present a brief summary of the main tendencies of total fertility in Figure 3.58 to Figure 3.61. In 1870, the lowest fertility persists in France, which had already known low fertility throughout the 19th century, and in Ireland, the Scottish Highlands and parts of Sweden. On the other hand, the highest fertility has been calculated in southern Russia and Romania. The other countries lie between these two extremes. In 1900 most other regions of the UK join France with lower fertility, France still progressing in low fertility levels. But also in various other parts of Europe total fertility falls to similar levels (e.g., Norway, Latvia and Madrid).

The general tendency to lower fertility is clearly visible in most European countries in 1930. Nevertheless, while this trend continues to advance in most parts of Europe, in particular France, Germany, Austria, the UK and Norway show a pattern of increasing fertility in 1960. Total fertility is now highest in Albania.[15]

Let us further consider the fertility of married women. Marital fertility is shown in Figure 3.62 to Figure 3.65. The lowest values for marital fertility in 1870 are again found in France. France is followed by many regions of Transleithania (the Hungarian part of Austria-Hungary), German Mecklenburg, Italy's Lazio region and several Spanish regions (Madrid, Catalonia, Balearic Islands). Higher marital fertility dominates in particular in many regions of the Russian Empire and Scandinavia. However, the highest shares come from south and west Germany and Belgium.

In 1900, the picture remains globally the same. In most cases, marital fertility is reduced in comparison to 1870. In France, all regions have reached a similar low marital fertility index of about 0.3 to 0.45, only (historically Gaelic speaking) Bretagne stands out and, to a lesser extent, some regions in north-east and north France.

[15] Before 1960, there is no data available on Albania. Therefore, it is possible that Albania had already high fertility rates before that time.

A more radical evolution show the values for 1930. Marital fertility has further declined in most European regions. This trend is very strong in Germany, Austria and today's Czech Republic, where many regions are now on the forefront of low fertility. The shift is equally important in England and parts of Sweden. Regions of the USSR have now the highest marital fertility. Other regions in Europe characterised by relatively high marital fertility are located in particular in Ireland, the Iberian Peninsula, southern Italy and Yugoslavia.

Finally, the tendency to lower marital fertility is continued in many European regions in 1960. This is particularly true for the regions of the USSR. However, some countries are already experiencing a renewed increase in marital fertility. France is once again the leader in this trend, increasing its marital fertility in almost all its regions. Similar are the changes in parts of Germany and Austria. The highest marital fertility remains in Catholic Ireland, Muslim Albania and Buddhist Kalmykia.

In addition, Figure 3.66 to Figure 3.69 highlight the evolution of the marital status (nuptiality) during the demographic transition. Between 1870 and 1900, a reduction in the marital status, i.e., an increase in the average age at marriage, is apparent in the British Isles, northern Spain and Transleithania. In Germany and Scandinavia, one can observe the reverse tendency in several regions. The highest marital status index is generally found in eastern Europe.

In 1930, the higher marital status in eastern Europe is in most cases at the same level as in French regions and some German regions. An outlier are the Balkans where higher marital status is still common. A very low marital status is characteristic for Ireland, the Scottish Highlands, distinct regions of Norway, Sweden and Finland as well as in Alpine regions.

Contrary to what may have been expected from the earlier descriptions, the trend towards lower marital status is reversed between 1930 and 1960 and the European Marriage Pattern vanishes. Instead, in almost all Europe marital status increases, i.e., the average age at marriage falls. Particularly in Communist countries in eastern Europe but also in many parts of western Europe the increase is clearly visible.

Finally, note that the study of historical and long-term fertility trends in Europe is important to understand the current and future economic and social development in the world. As Engelen and Puschmann stress, the "marriage behavior in the present-day Arab world shows striking similarities to nuptiality patterns which have been described by Hajnal and adherents as typically Western European" (Engelen and Puschmann 2011, p. 387). One may also come to similar conclusions for other regions of the world.

3.4.3 Technological progress

Technological progress is a very important factor in models that intend to explain economic growth. Clearly, there are very different ways to highlight technological progress. For example, technological progress and innovation may be measured using patent grants (e.g., Acs *et al.* 2002, Diebolt and Pellier 2009a, b, c, Diebolt and Pellier 2011, Diebolt and Pellier 2012). Another key factor influencing the spatial distribution of economic activity is the transport of goods and labour. The Industrial Revolution brought about a revolution in transportation. An illustrative example of the advancement in transportation in Europe in the long run is given by Bretagnolle *et al.* (1997) (Figure 3.70). The authors highlight that the average time of travelling from Ireland to the south of Italy was two months in 1500. This time was reduced to only one day by 1990. The figure shows the space-time contraction that has characterised the further advancement of transportation in Europe (Janelle 1969, Juillard 1972, Bretagnolle *et al.* 1997).

Related to this is a key driver of transportation during the Industrial Revolution: the railway. The railway has been a quick and cost-effective means of transport. The railway network was very important in the further advancement of the Industrial Revolution. Its evolution is highlighted in Figure 3.71 (taken from Morillas-Torné 2012). The densest networks are located in the UK, Benelux, northern France and the German Empire in 1880. Scandinavia, the Iberian Peninsula, Italy and eastern Europe clearly lagged behind. Until 1940 the network was enlarged almost everywhere in Europe.[16] Given the progress of alternative means of transport (car, airplane), the lines in service have been reduced in some countries until 2000.

3.4.4 Human capital

3.4.4.1 Basic human capital

With the term basic human capital we refer to rather basic education, as highlighted by the ability to count or to read and write. A long-term perspective on basic human capital has been advanced by Hippe (2012a), referring also to earlier work, in particular by A'Hearn *et al.* (2009), Hippe (2012b) and Hippe and Baten (2012a). Given the detailed illustration in other work, this paper only highlights that, historically, the earliest and most advanced regions in education can be found in the Germanic countries (Scandinavia, Prussia, Netherlands). In contrast, the regions which caught up only very late (if at all) are located in the European periphery, i.e., in the Iberian Peninsula (in particular Portugal), eastern Europe (Belarus, Russia, Ukraine, etc.) and south-east Europe (the Balkans and the Caucasus region).

[16] Countries in the East of EU members are not taken into account in the figure (e.g., Ukraine, Russia).

3.4.4.2 Advanced human capital

More advanced human capital may refer to the study at a level of tertiary education. The number of students or similar variables are not available at a European regional scale. For this reason, we trace the evolution of the locations of enduring universities in the European space from 1300 to 1944. The original data have been categorised according to different centuries. Note that the data do not refer to all universities that existed at each particular point in time. They only include those universities that persisted at least until the first half of the 20th century, i.e., institutions of higher education which were not of temporary existence and did not close down. For this reason, the number of higher education institutions increases at each period in time. Still, it allows to take some general conclusions (see Figure 3.72 to Figure 3.80).

The first institutions of higher research (created before 1300) were located in Bologna, Cambridge, Lisbon, Montpellier, Naples, Orleans, Oxford, Padua, Paris, Salamanca, Siena and Toulouse. Thus, all these institutions were located in France, Italy, Portugal, Spain and the UK. These countries increased their number of higher education institutions over the following centuries and other countries followed their example. In particular, the Holy Roman Empire began to invest in higher education. First enduring Scandinavian universities were created between 1400 and 1500 and in today's Russia only between 1700 and 1800 (in Moscow and St. Petersburg). Only between 1900 and 1944 there was a massive increase in the number of universities in Russia, when most other parts of Europe had already a dense net of universities over their territories.

3.5 Conclusion

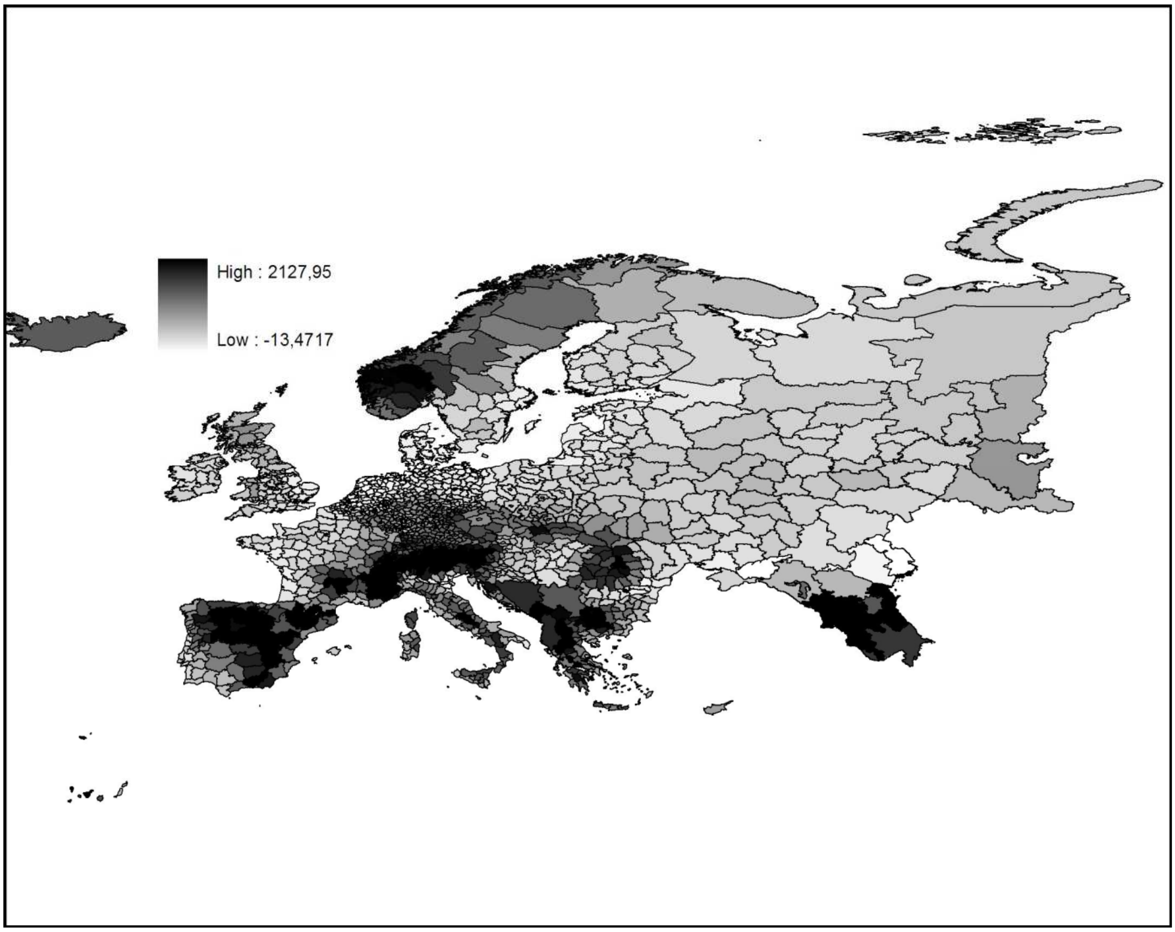
This paper has given an outline of the regional and long term evolution of the factors of production in Europe. The factors of production are key to standard economic growth models and their applications. We have used a wide range of different sources and proxies to trace the evolution of land, capital, labour, technological progress and human capital. Nevertheless, there are evidently many other alternative ways to measure these factors proposed by theory. Given the focus of this paper, we have advanced the following aspects of land: altitude, temperature and precipitation. Second, capital and labour have been highlighted by the general economic development, real wages (i.e., welfare ratios), commodity prices (wheat), (urban) population, infant mortality, fertility and marital status. Third, the evolution of technological progress was shown by the advance of transport technologies. Last, human capital evolution was particularly demonstrated by the location of institutions of higher education.

Although we have only outlined very roughly the regional distribution over time of these proxies, this sketch may allow a clearer overall picture of how economic activity has been evolving in space in Europe over the last centuries. In this way, it may underline the possibilities existing in this area of research and motivate more profound future analysis of the issues at stake. This will allow a much better and profound understanding of economic development in Europe at the regional level in the long run.

3.6 Appendix

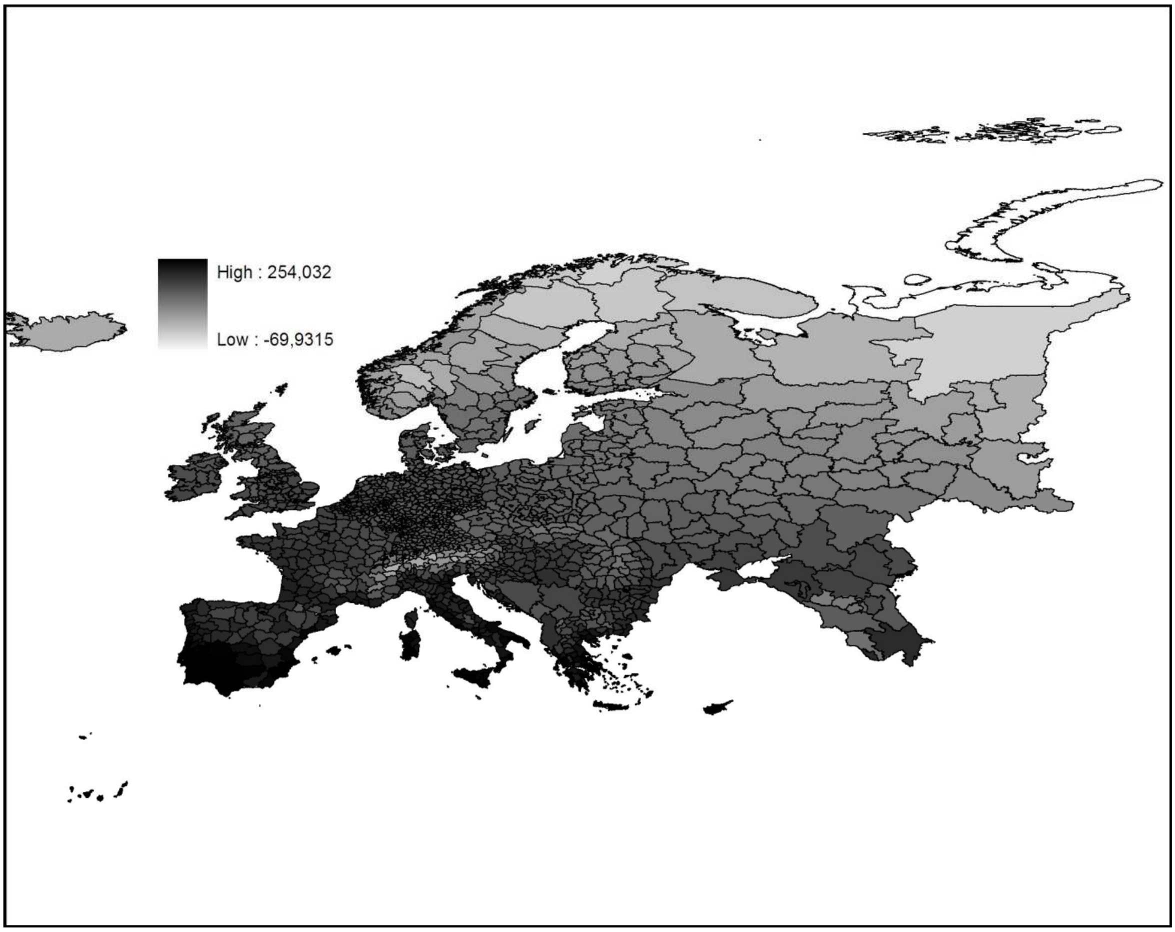
3.6.1 Figures

Figure 3.1 Average altitude by NUTS 3 regions, ca. 1950-2000



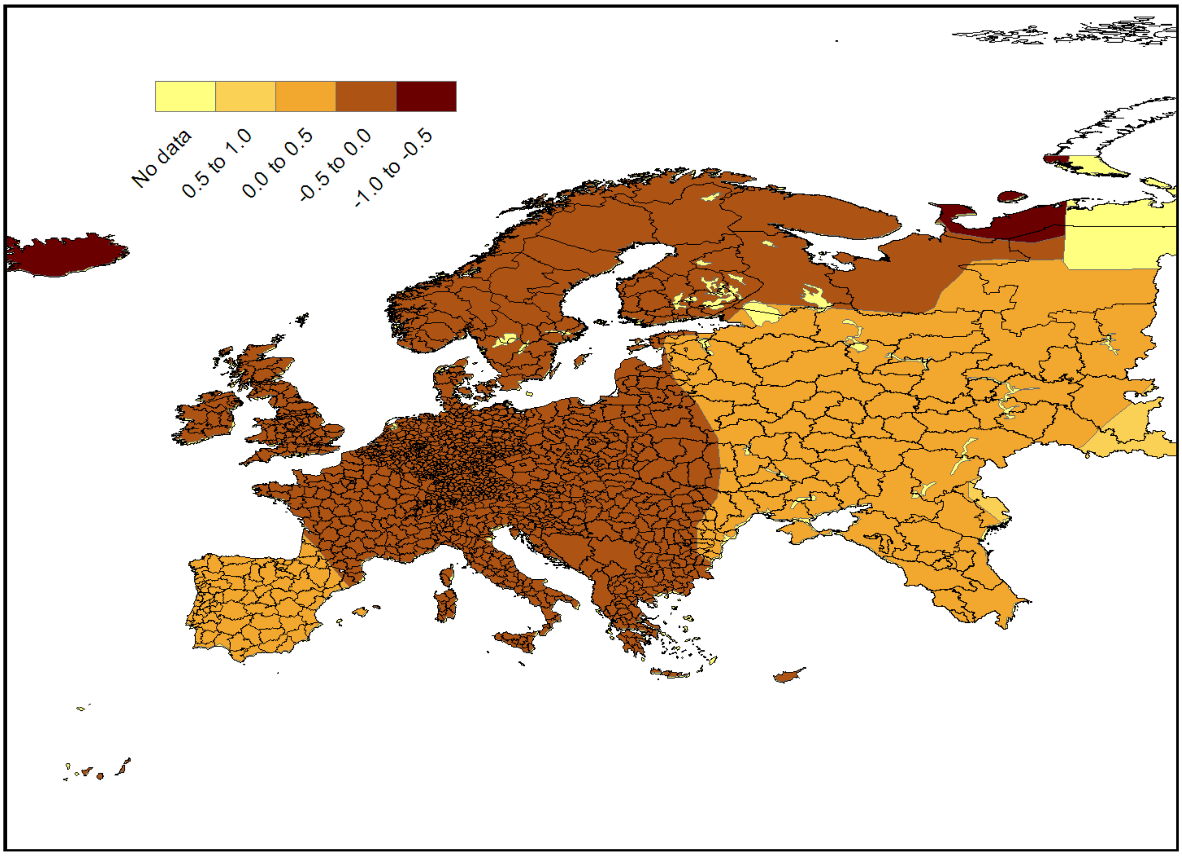
Source: Own calculations, based on data provided by Hijmans *et al.* (2005).

Figure 3.2 Average annual temperature by NUTS 3 regions, ca. 1950-2000



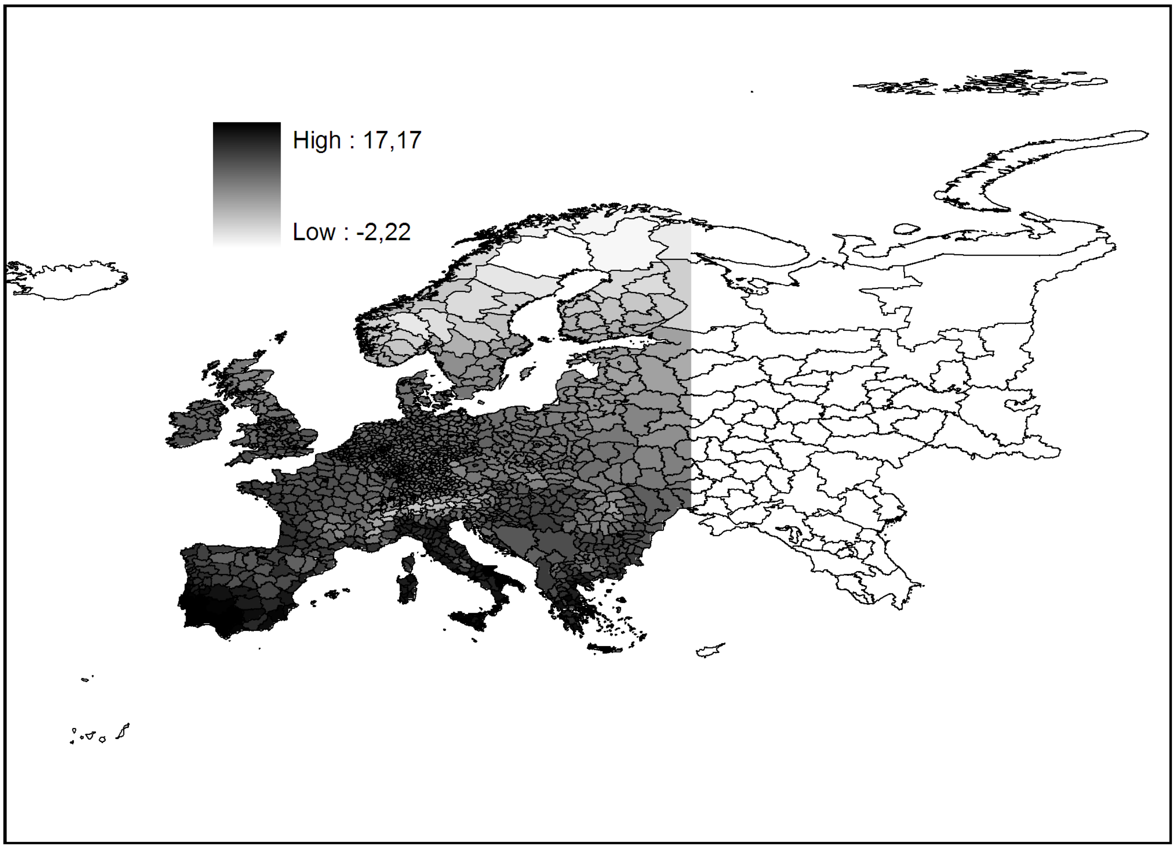
Note: Mean temperature is expressed in °C x 10.
Source: Own calculations, based on data provided by Hijmans *et al.* (2005).

Figure 3.3 Change in average annual temperature, 1930-60



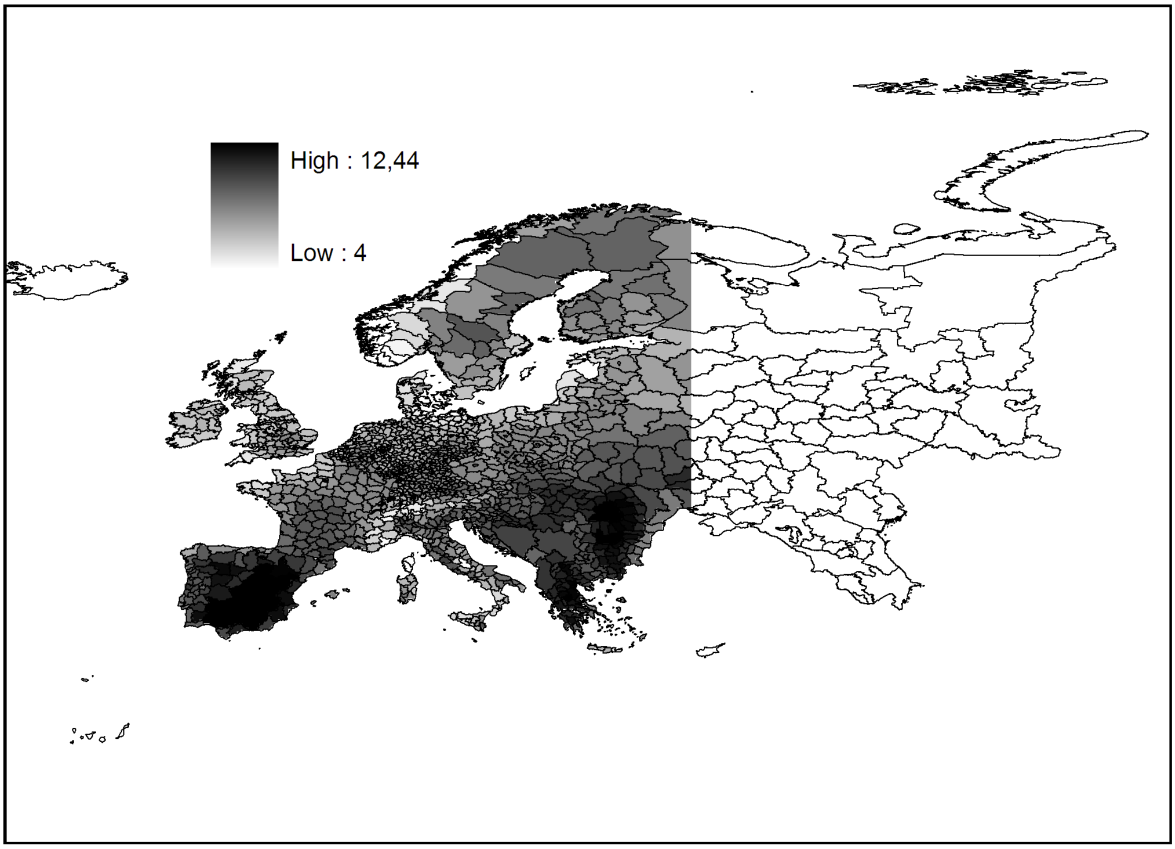
Note: Temperature is expressed in °C.
Source: Data provided by Deichmann and Eklundh (1991).

Figure 3.4 Average annual temperature by NUTS 3 regions, 1901-1910



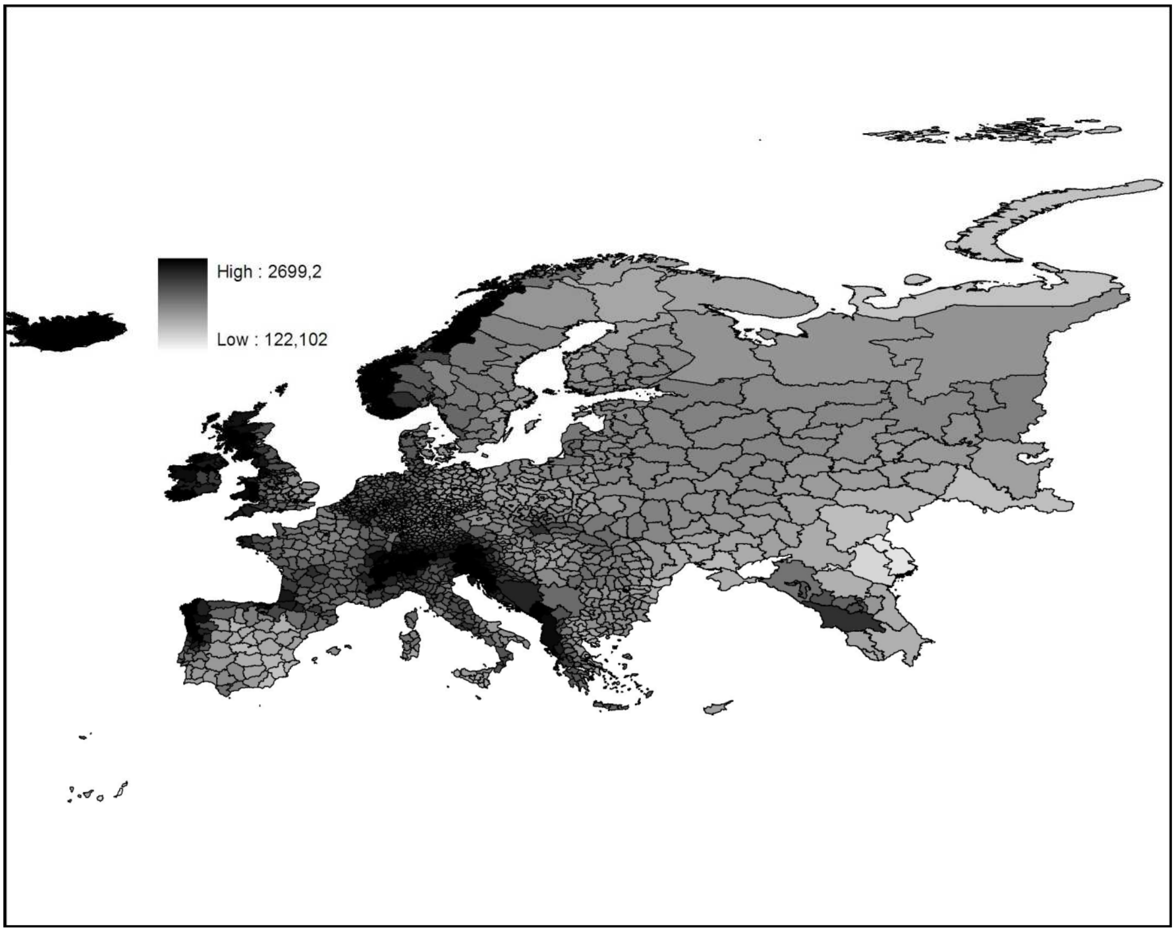
Note: Mean temperature is expressed in °C.
Source: Own calculations, based on data provided by Mitchell (2004) and Fronzek *et al.* (2012); see also ALARM (2012).

Figure 3.5 Average daily temperature range by NUTS 3 regions, 1901-1910



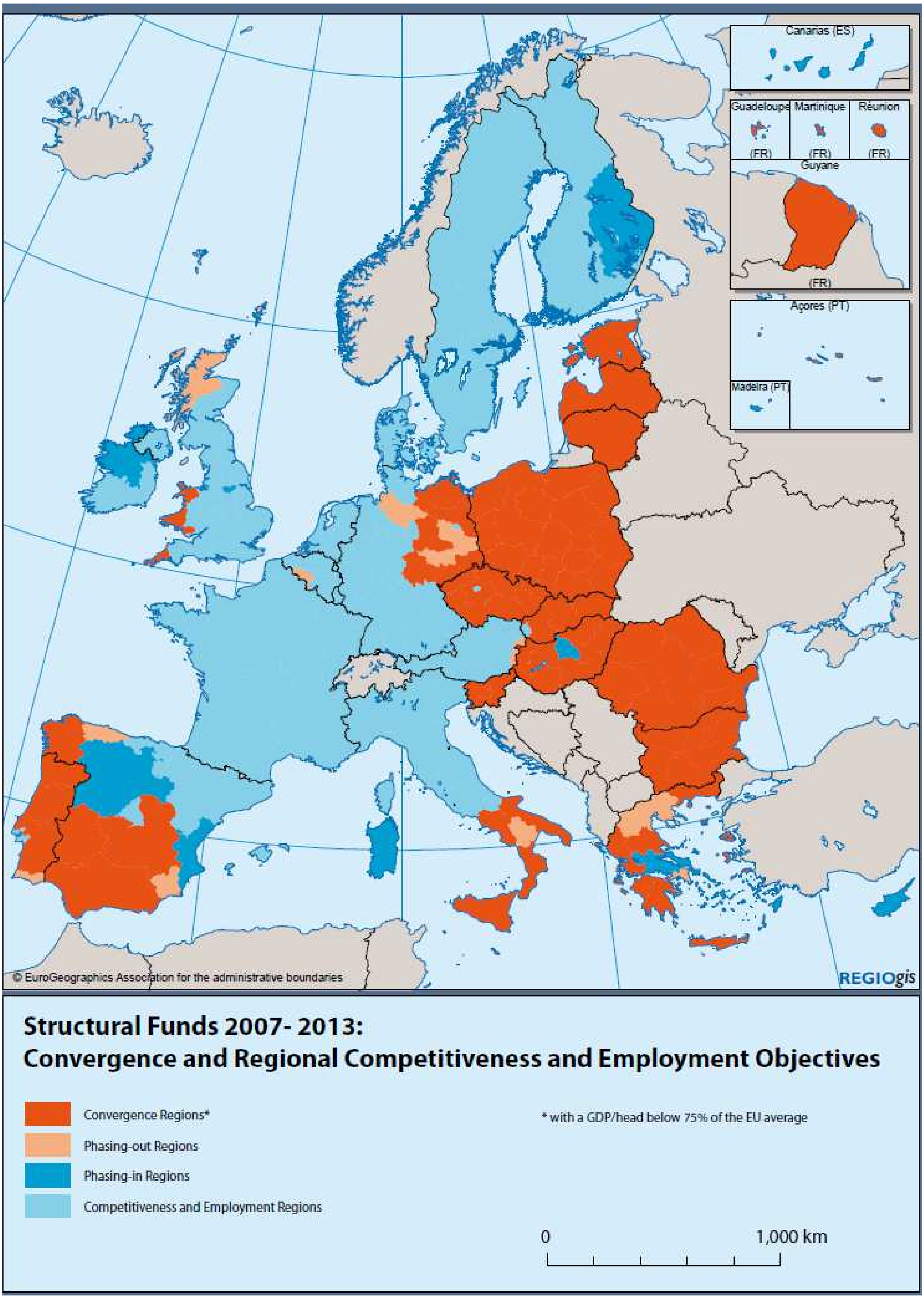
Note: Temperature range is expressed in °C.
Source: Own calculations, based on data provided by Mitchell (2004) and Fronzek *et al.* (2012); see also ALARM (2012).

Figure 3.6 Average annual precipitation by NUTS 3 regions, ca. 1950-2000



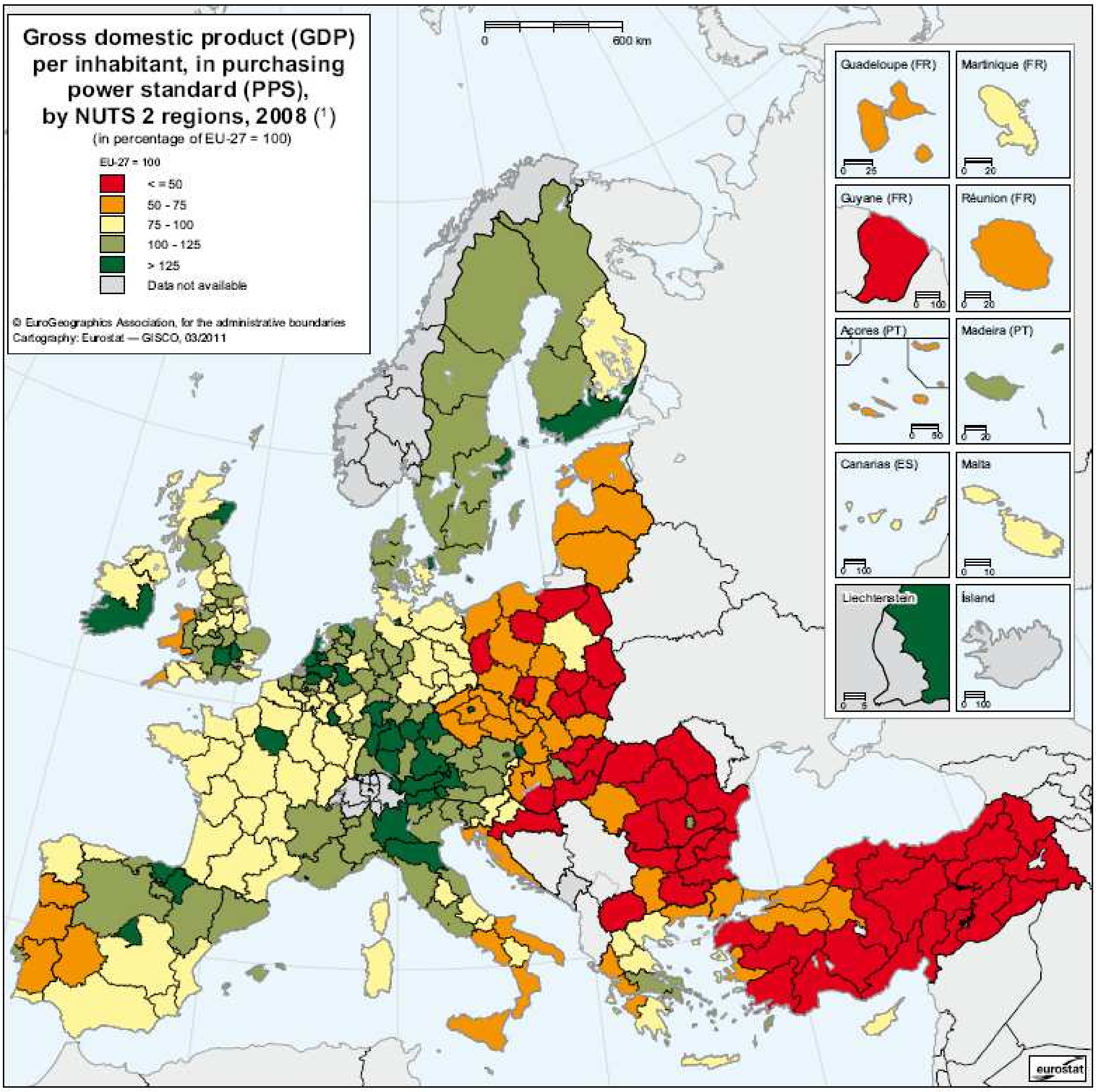
Note: Precipitation is expressed in mm (1 mm = 1 l / m²).
Source: Own calculations, based on data provided by Hijmans *et al.* (2005).

Figure 3.7 Structural Funds 2007-2013 in the EU



Source: Inforegio (2008); © European Union, 1995-2013.

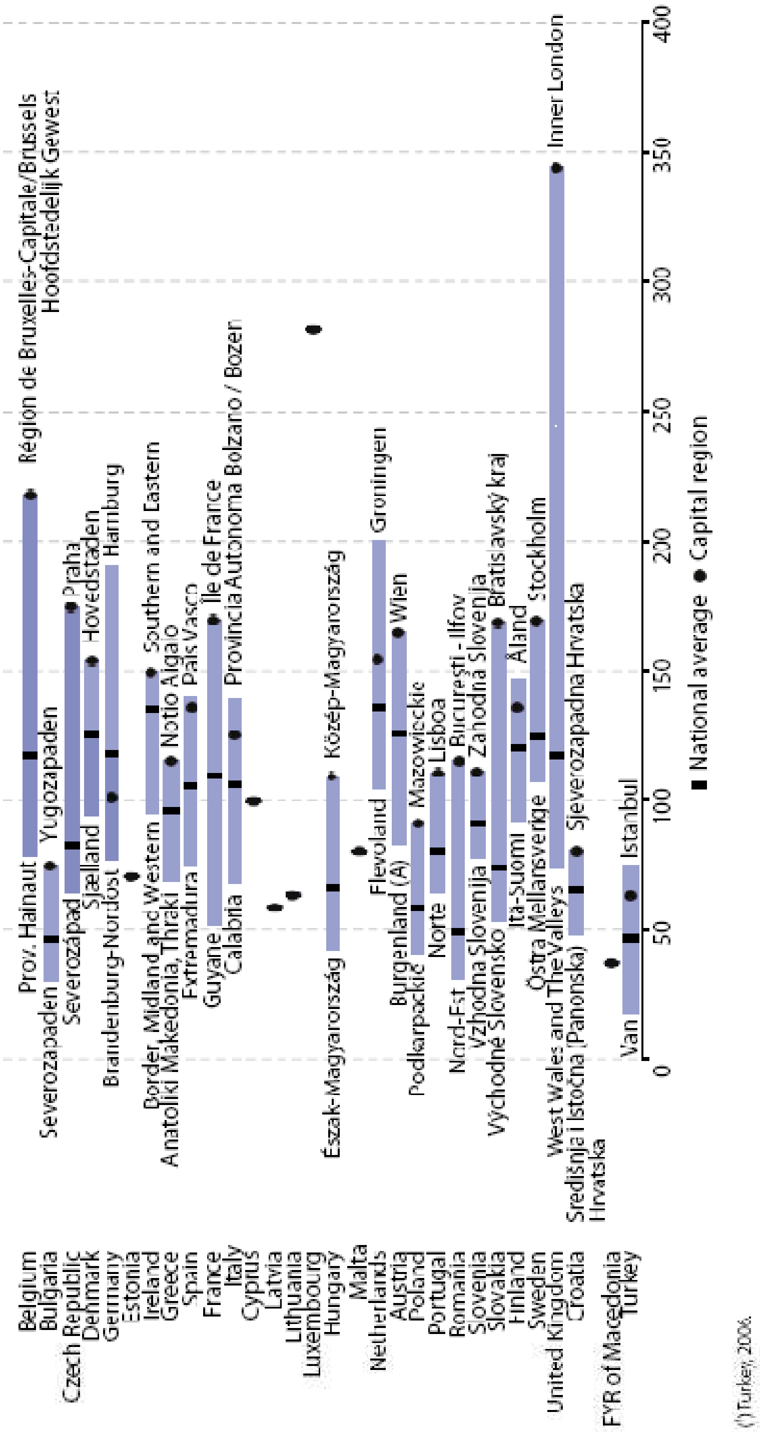
Figure 3.8 GDP per inhabitant (in PPS) in Europe, 2008



(1) Turkey, 2006.

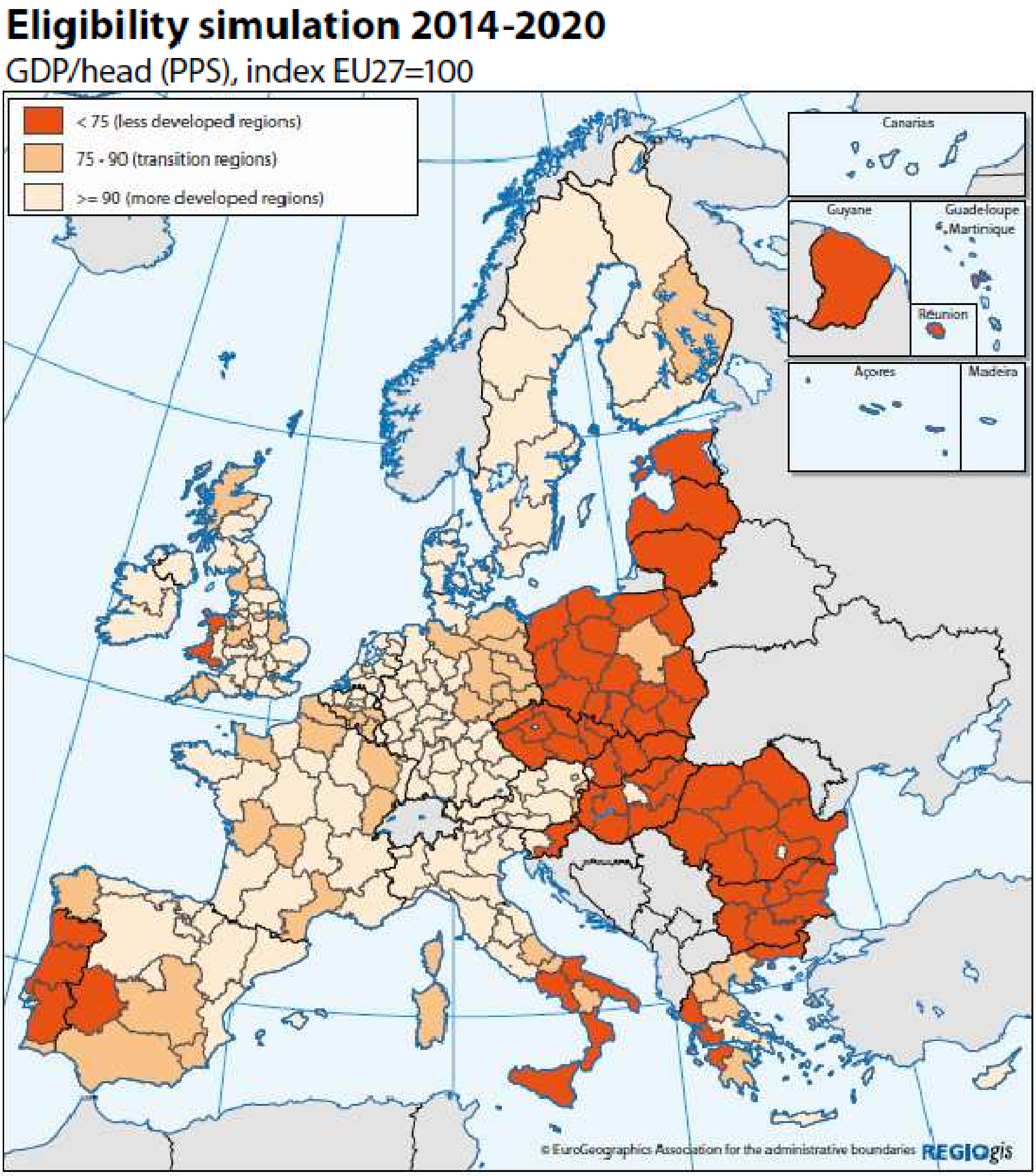
Source: Eurostat (2011f); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 3.9 Within country differences in GDP per inhabitant in Europe, 2008 (in PPS)



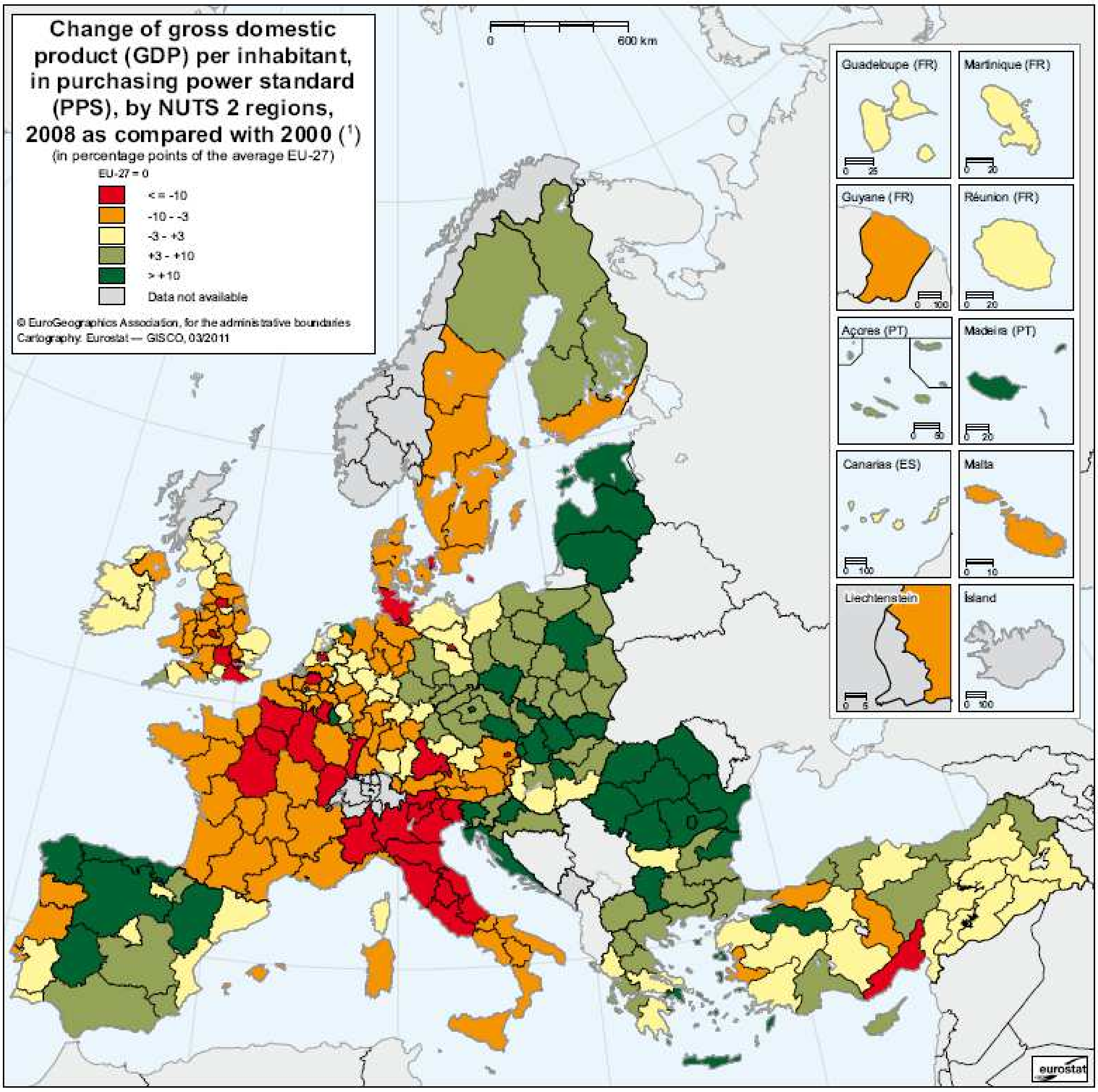
Note: In % of the EU-27 average, EU-27=100.
Source: Eurostat (2011g); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 3.10 Eligibility simulation 2014-2020



Source: European Commission (2011), © European Union, 2011.

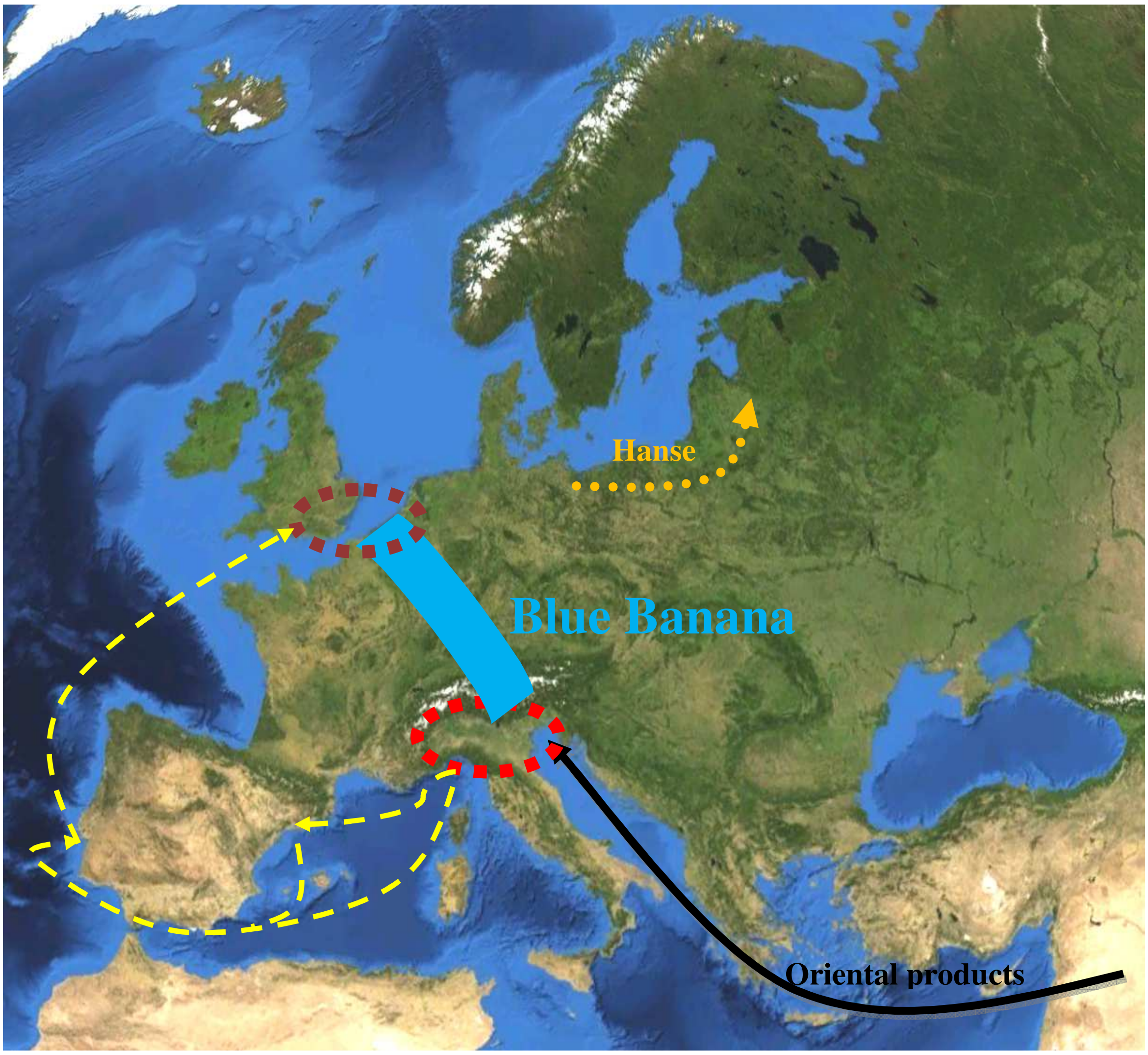
Figure 3.11 GDP per inhabitant (in PPS) in Europe, 2008 compared to 2000



(1) Denmark, Eurostat estimate; Turkey, 2006 as compared with 2000; Croatia, 2008 as compared with 2001.

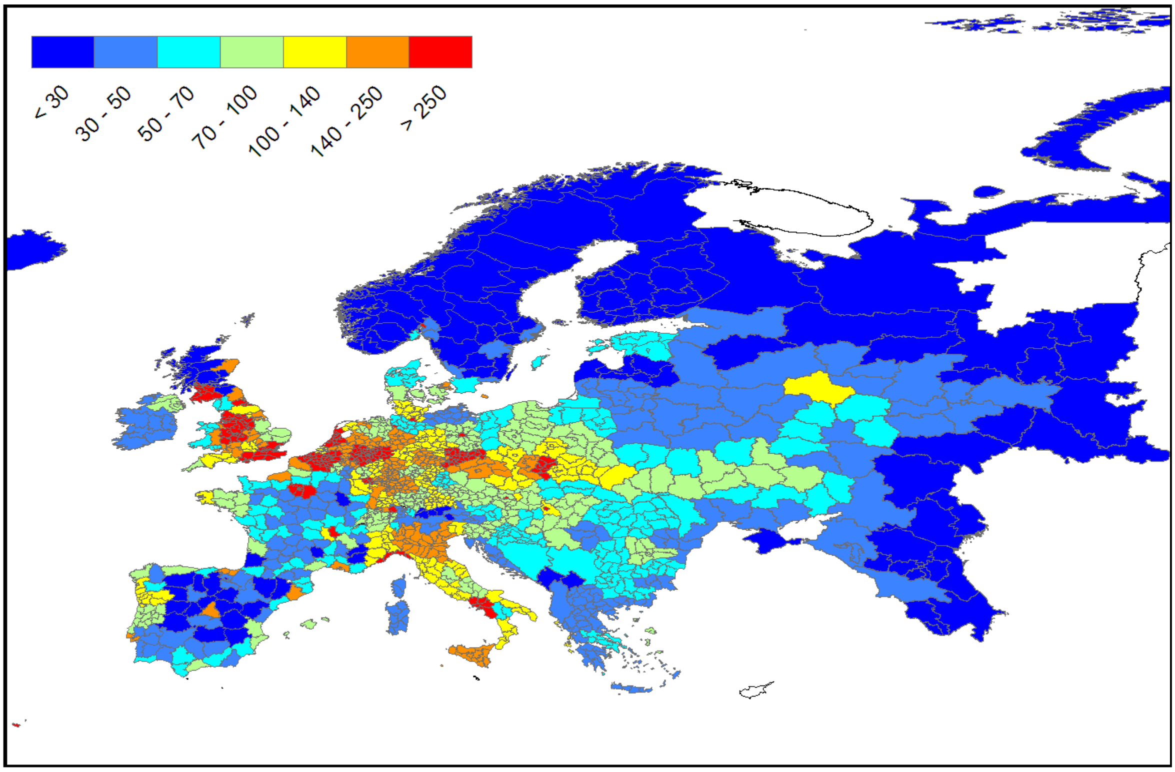
Source: Eurostat (2011h); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 3.12 The origins of the Blue Banana



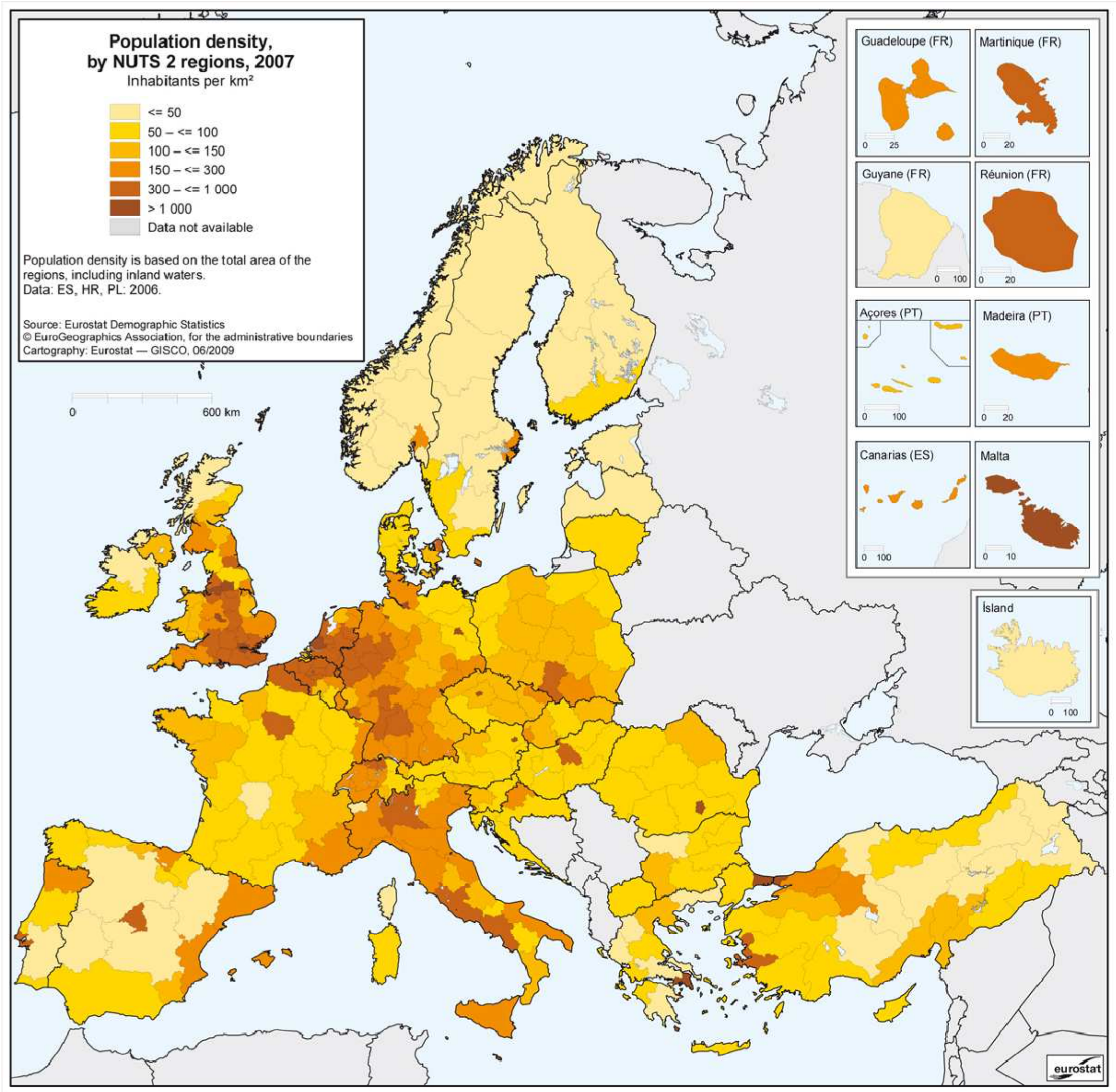
Source: Own presentation, based on Brunet (2002). Background map by Wikimedia Commons (2007).

Figure 3.13 Population density, ca. 1930



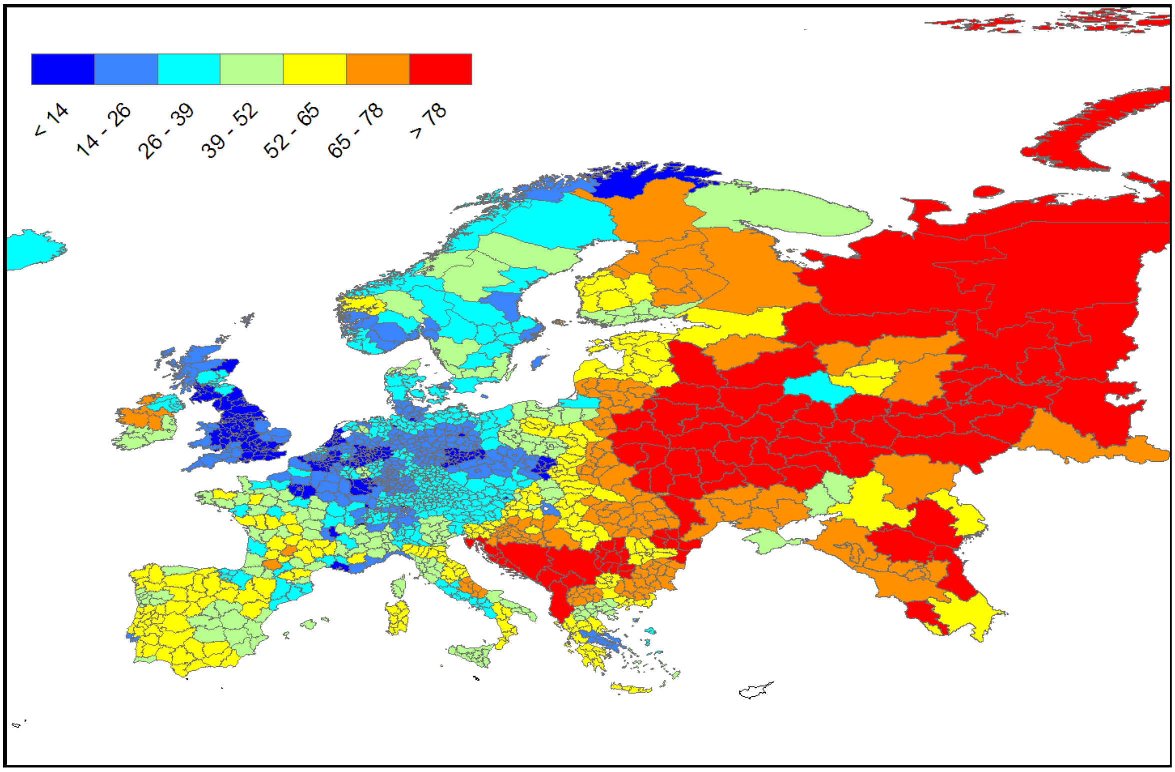
Note: Population density is defined as inhabitants per km². No data available for Komi Autonomous Region and Murmansk.
Source: Own calculations, data provided by Kirk (1946).

Figure 3.14 Population density in the European regions, 2007



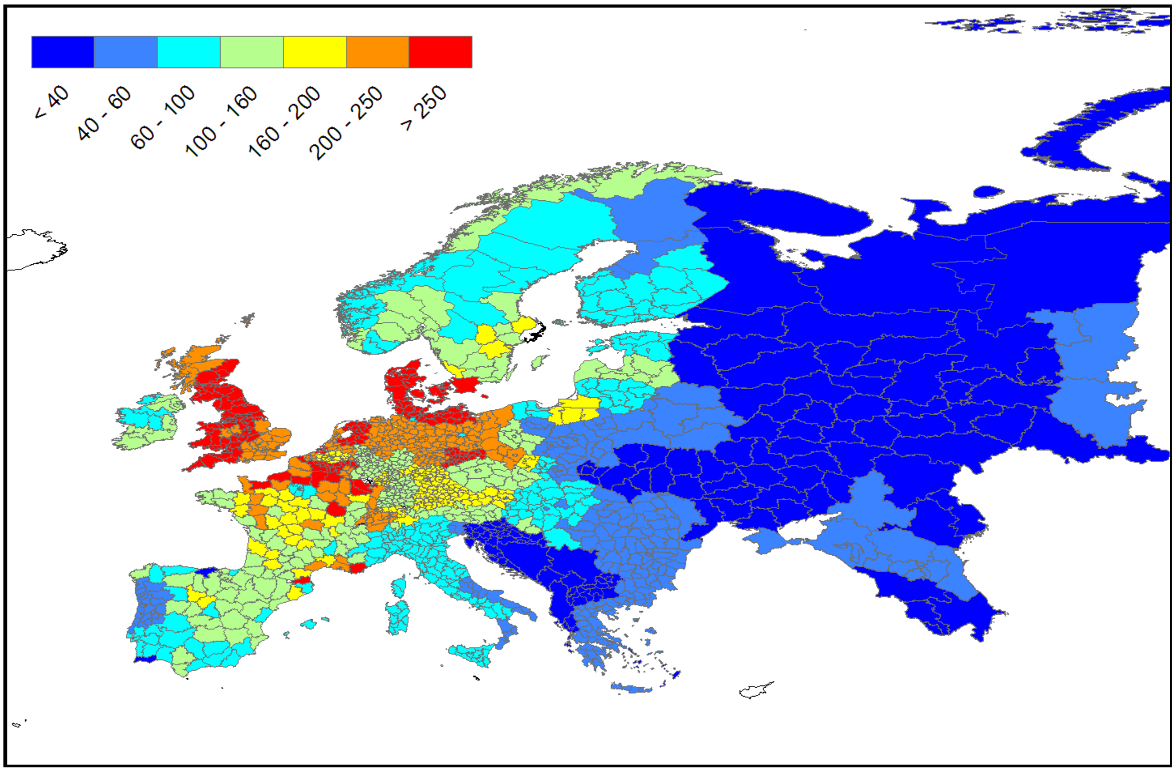
Source: Eurostat (2009), © European Communities, 2009; Eurostat, http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 3.15 Share of total population dependent on agriculture, ca. 1930



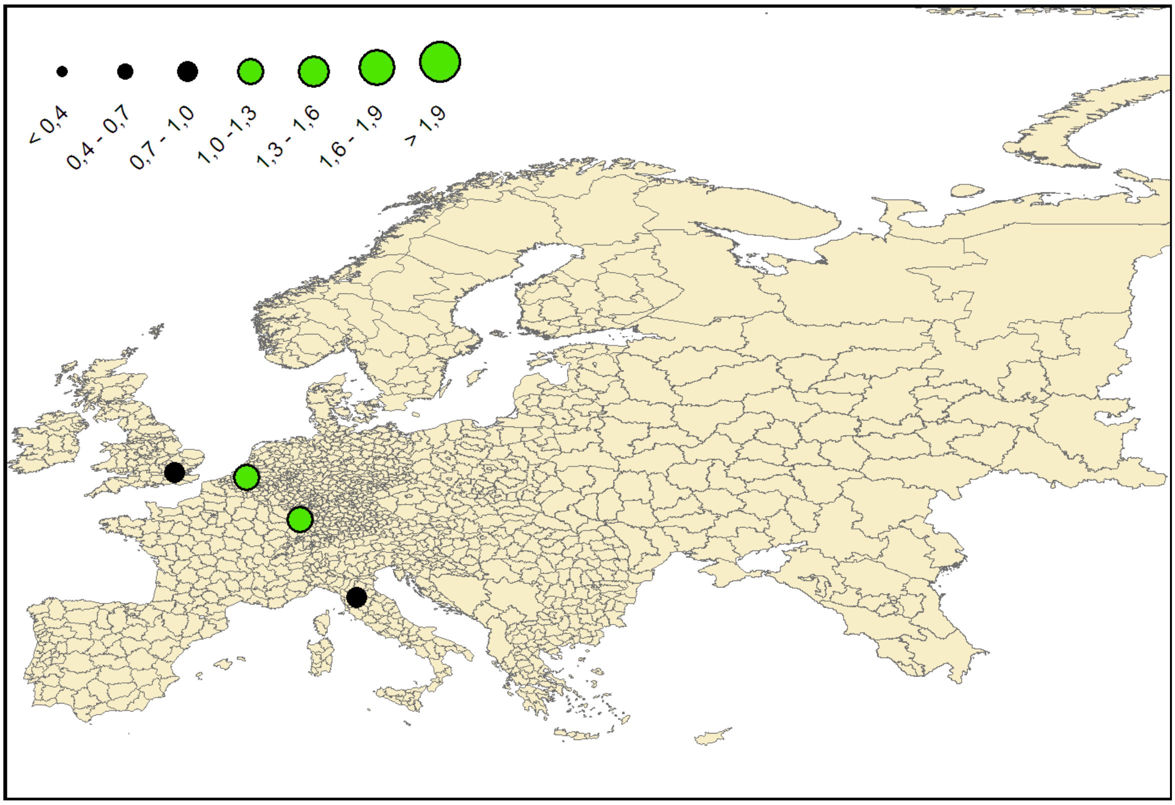
Source: Own calculations, based on data by Kirk (1946).

Figure 3.16 Agricultural productivity per capita, ca. 1930



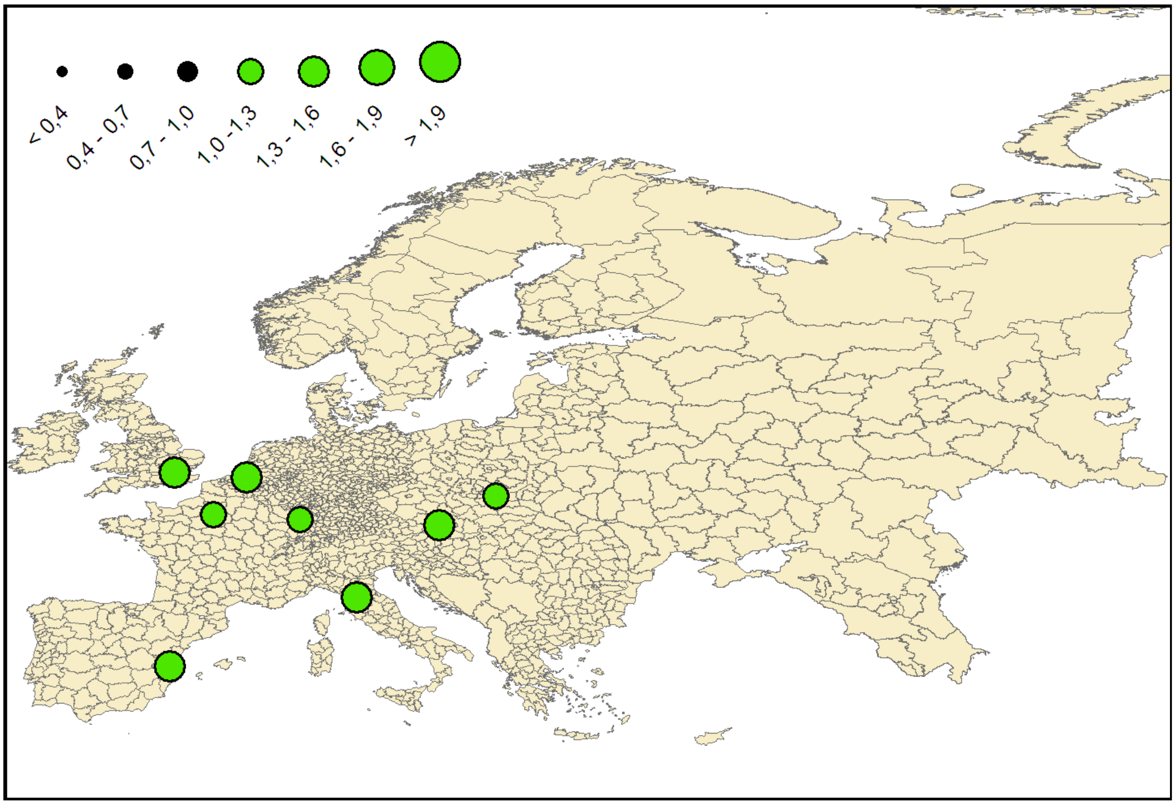
Note: The indicator is the "average per capita productivity of persons dependent on agriculture" (Kirk 1946, p. 262).
Source: Own calculations, based on data by Kirk (1946).

Figure 3.17 Welfare ratios for building labourers in European cities, 1300-1399



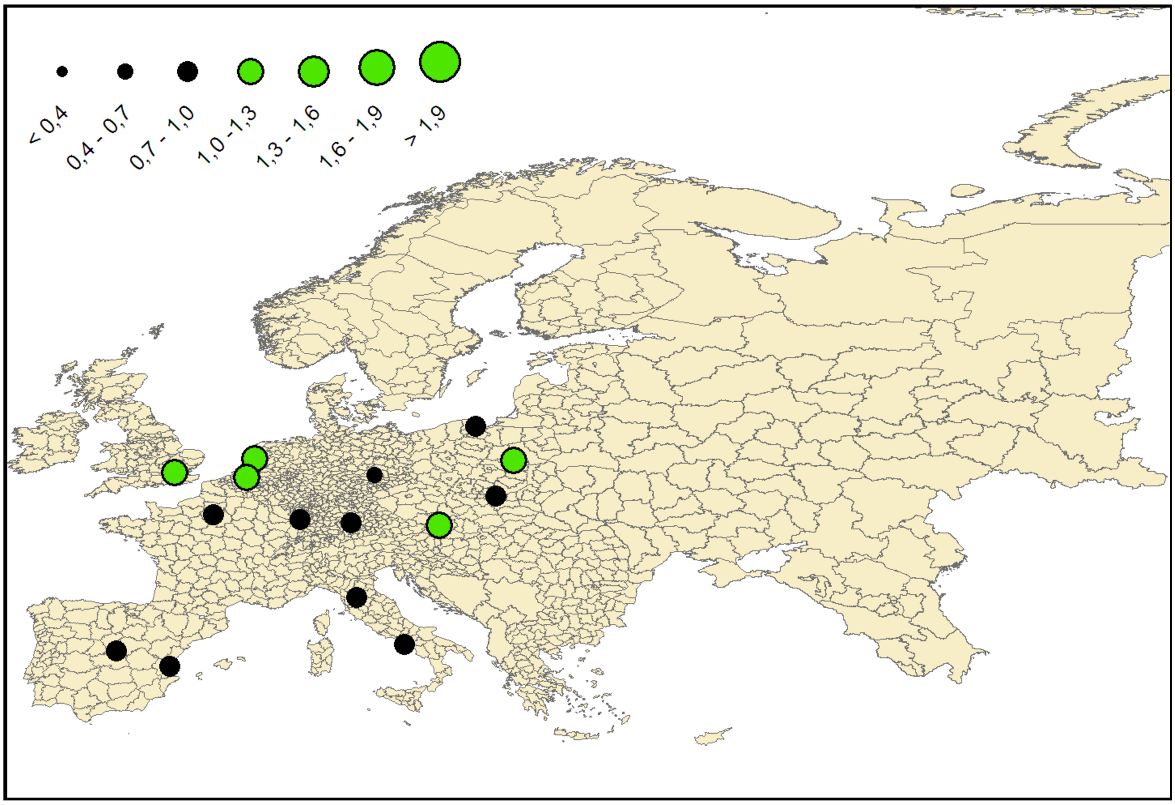
Source: Own calculations, based on data by Allen (2012).

Figure 3.18 Welfare ratios for building labourers in European cities, 1400-1499



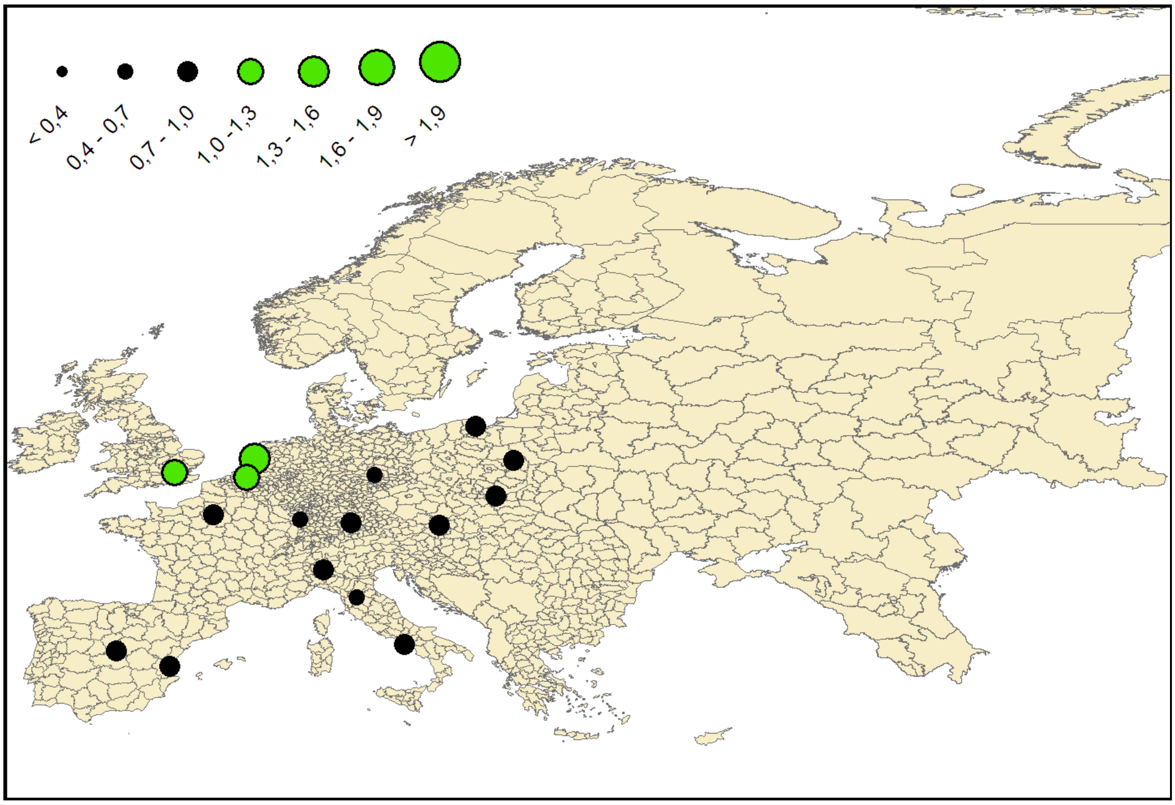
Source: Own calculations, based on data by Allen (2012).

Figure 3.19 Welfare ratios for building labourers in European cities, 1500-1599



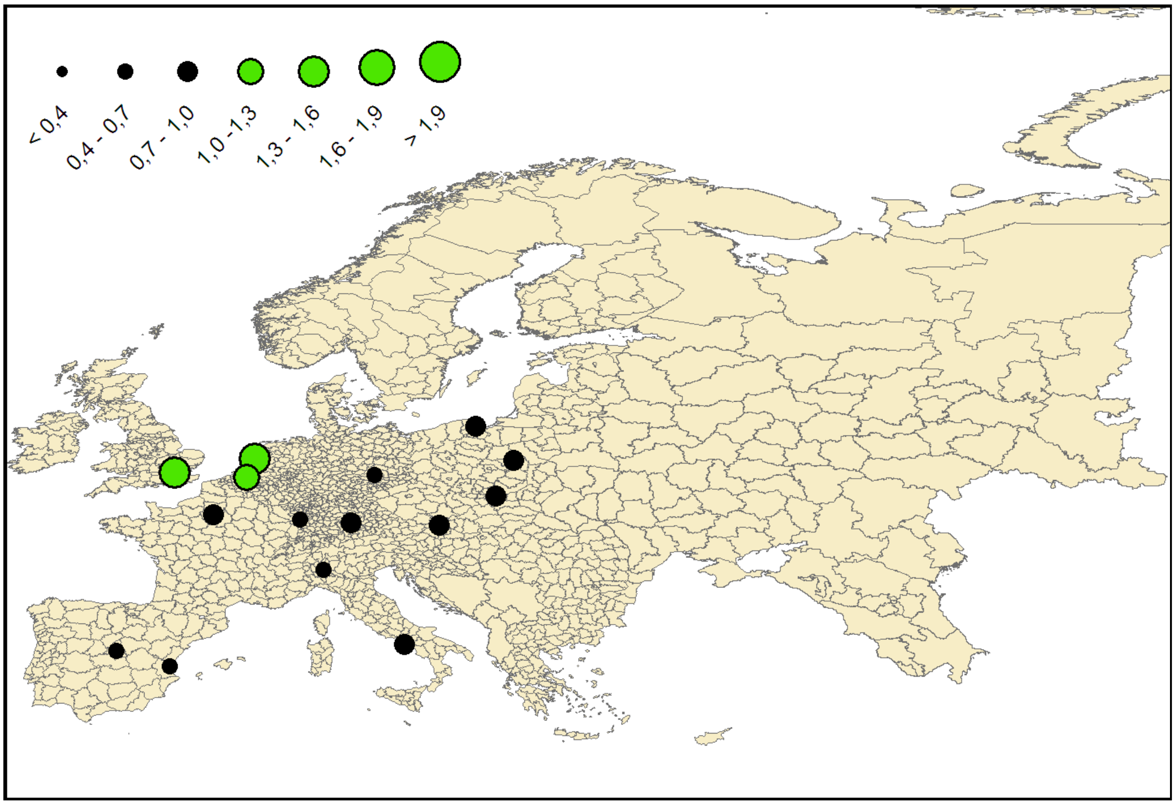
Source: Own calculations, based on data by Allen (2012).

Figure 3.20 Welfare ratios for building labourers in European cities, 1600-1699



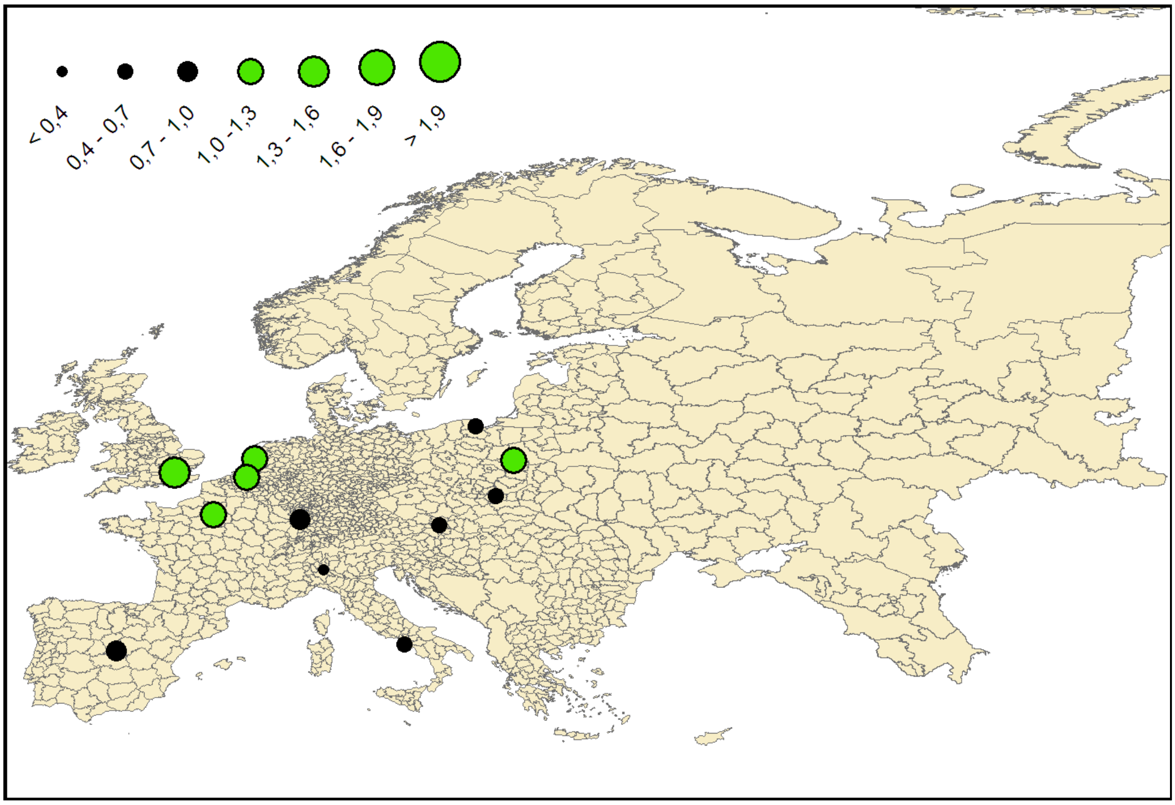
Source: Own calculations, based on data by Allen (2012).

Figure 3.21 Welfare ratios for building labourers in European cities, 1700-1799



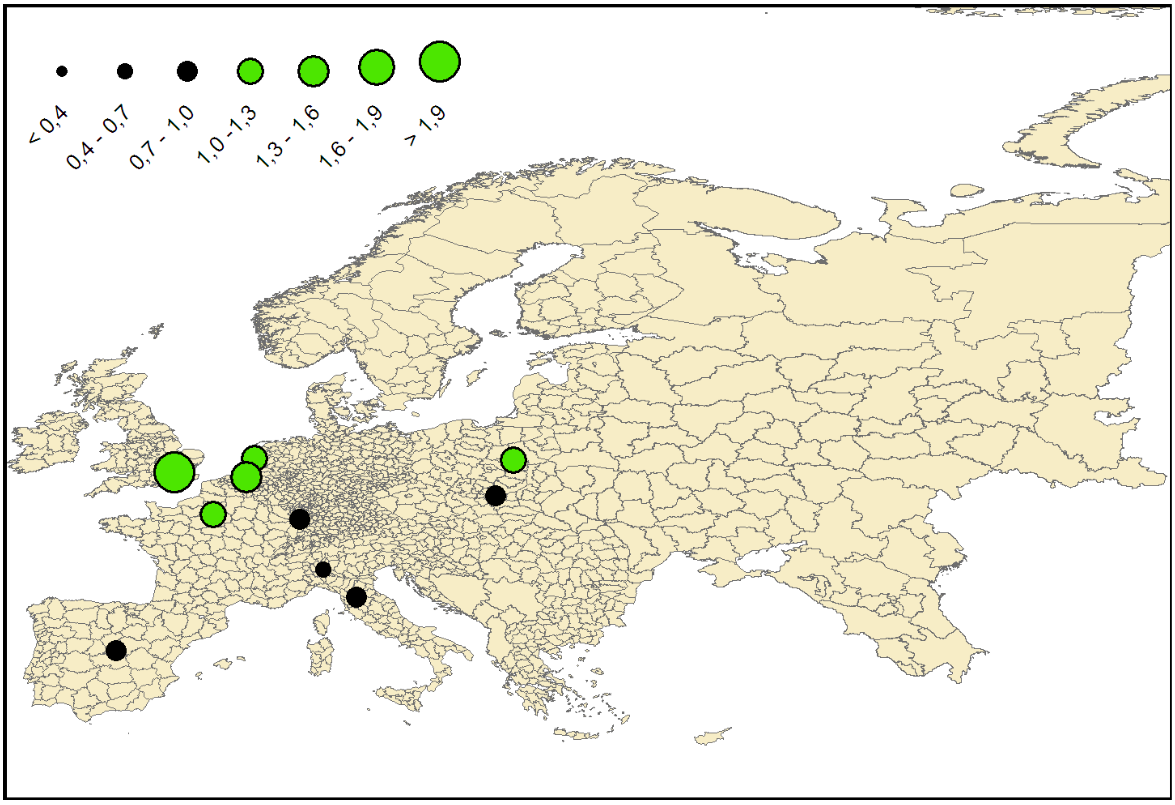
Source: Own calculations, based on data by Allen (2012).

Figure 3.22 Welfare ratios for building labourers in European cities, 1800-1849



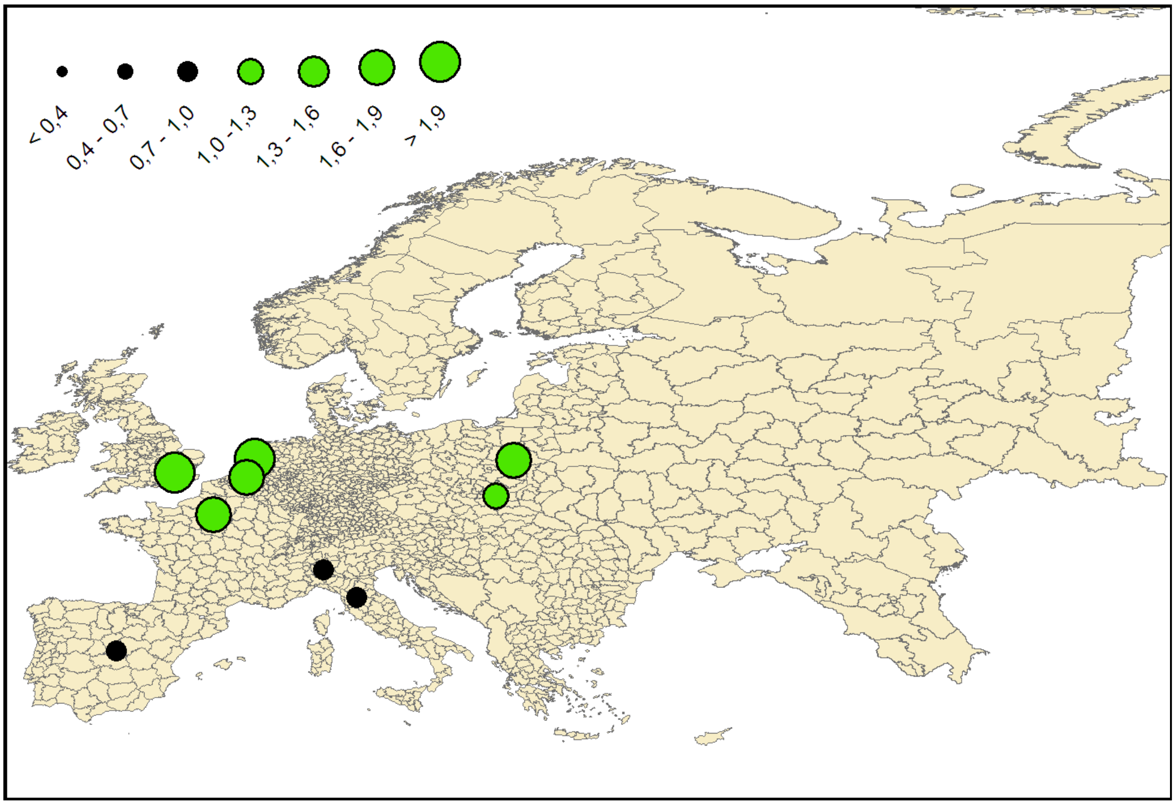
Source: Own calculations, based on data by Allen (2012).

Figure 3.23 Welfare ratios for building labourers in European cities, 1850-1899



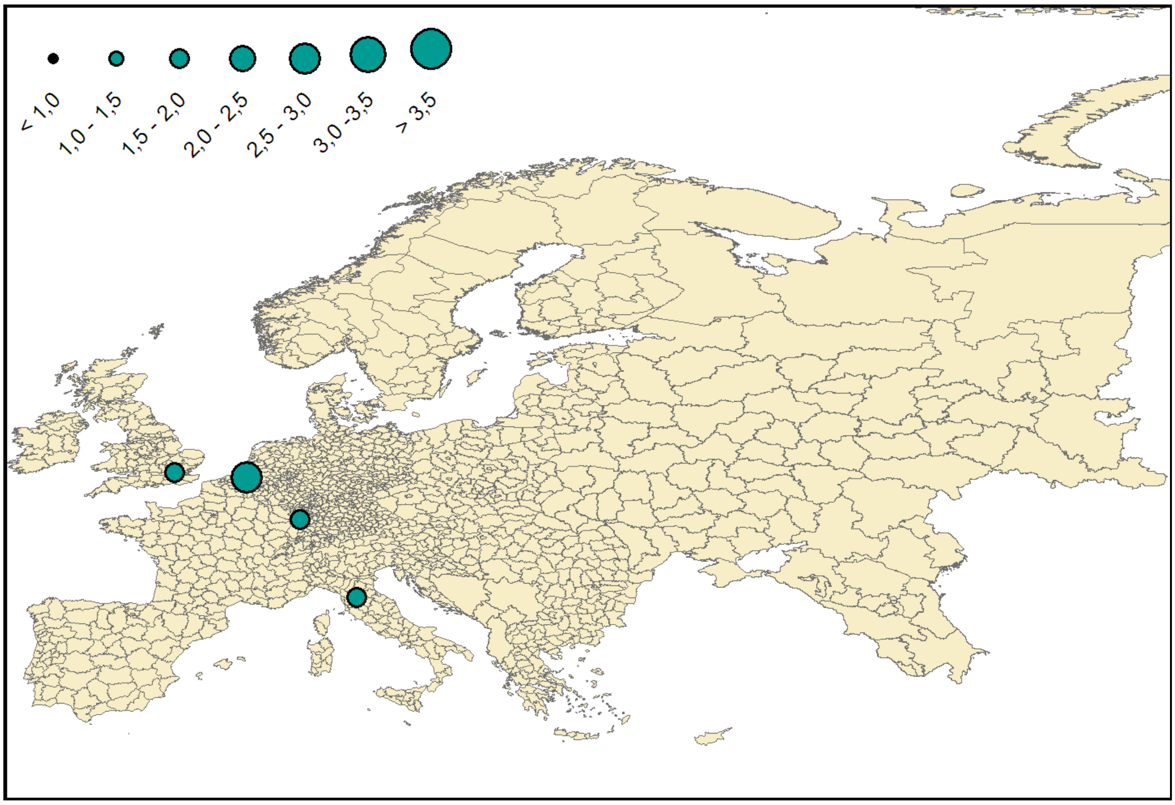
Source: Own calculations, based on data by Allen (2012).

Figure 3.24 Welfare ratios for building labourers in European cities, 1900-1914



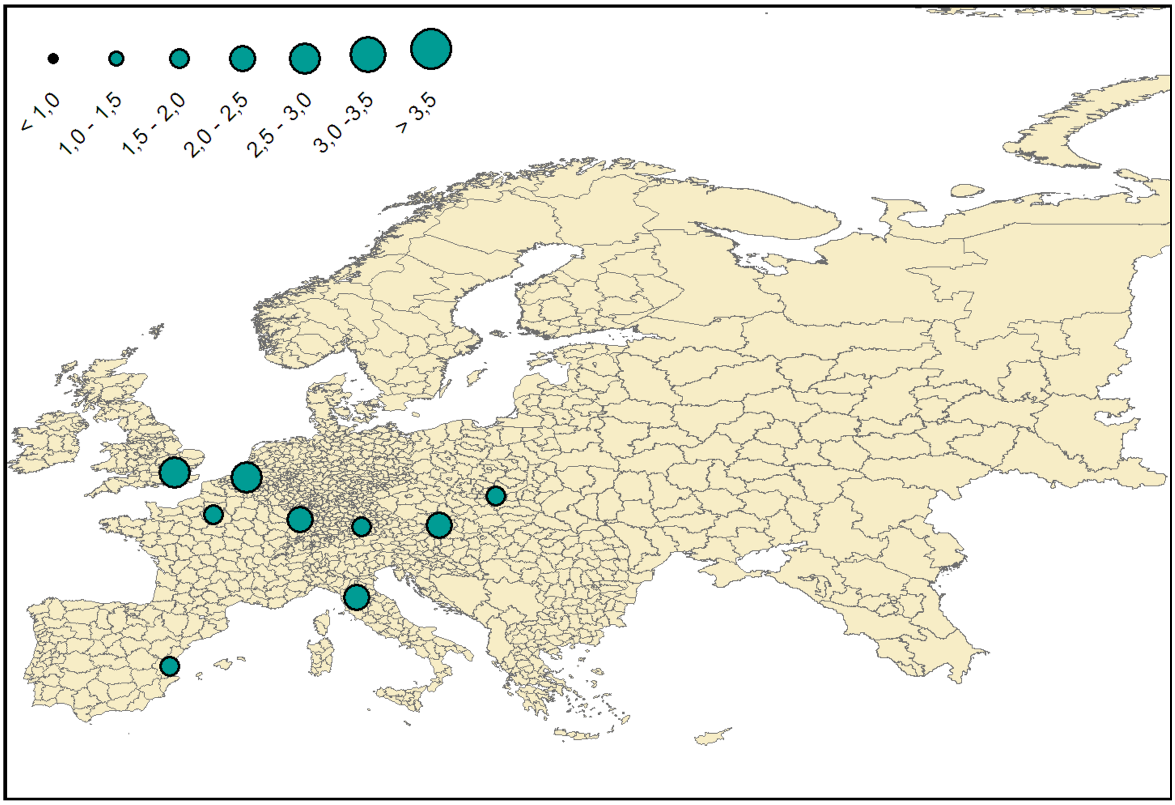
Source: Own calculations, based on data by Allen (2012).

Figure 3.25 Welfare ratios for craftsmen in European cities, 1300-1399



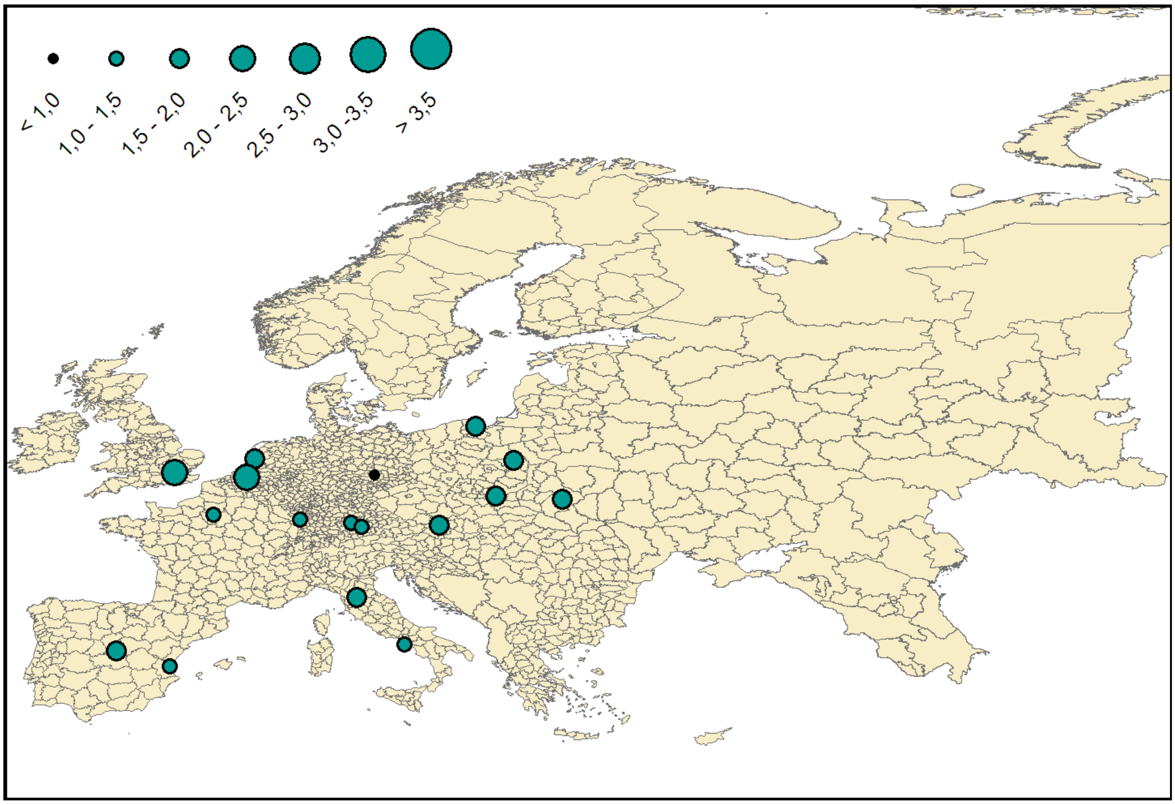
Source: Own calculations, based on data by Allen (2012).

Figure 3.26 Welfare ratios for craftsmen in European cities, 1400-1499



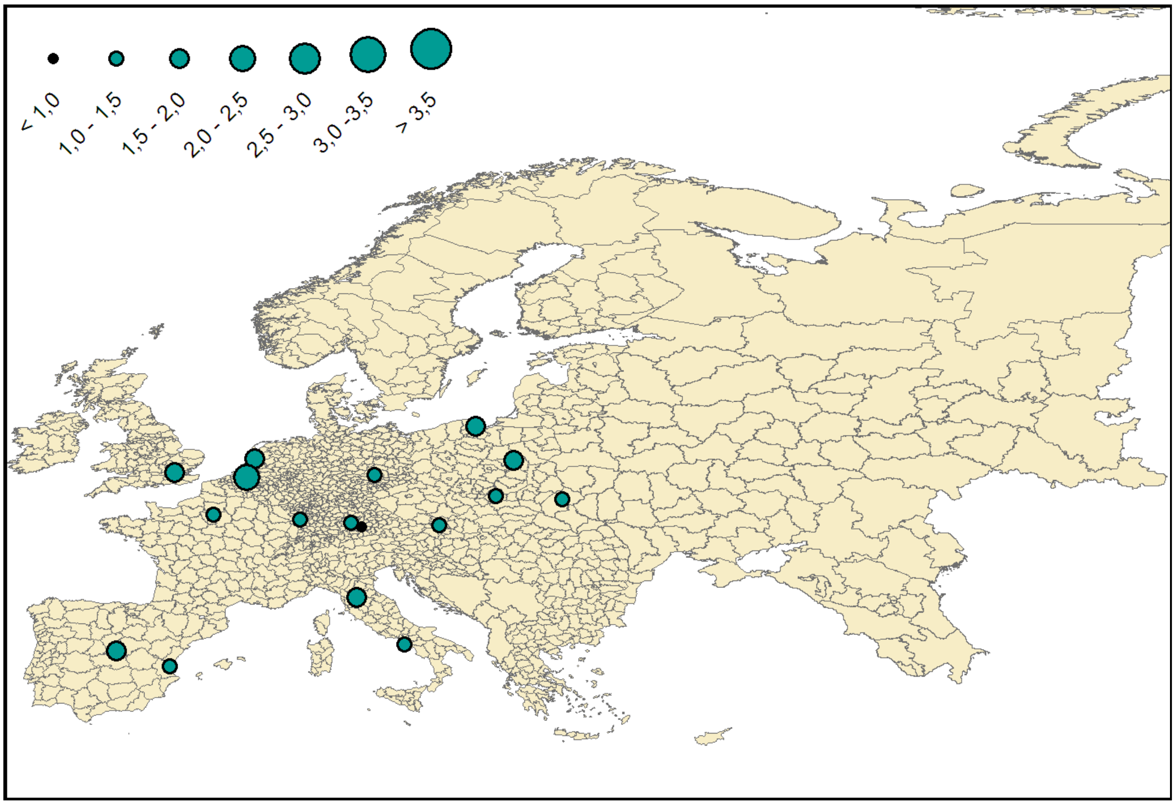
Source: Own calculations, based on data by Allen (2012).

Figure 3.27 Welfare ratios for craftsmen in European cities, 1500-1599



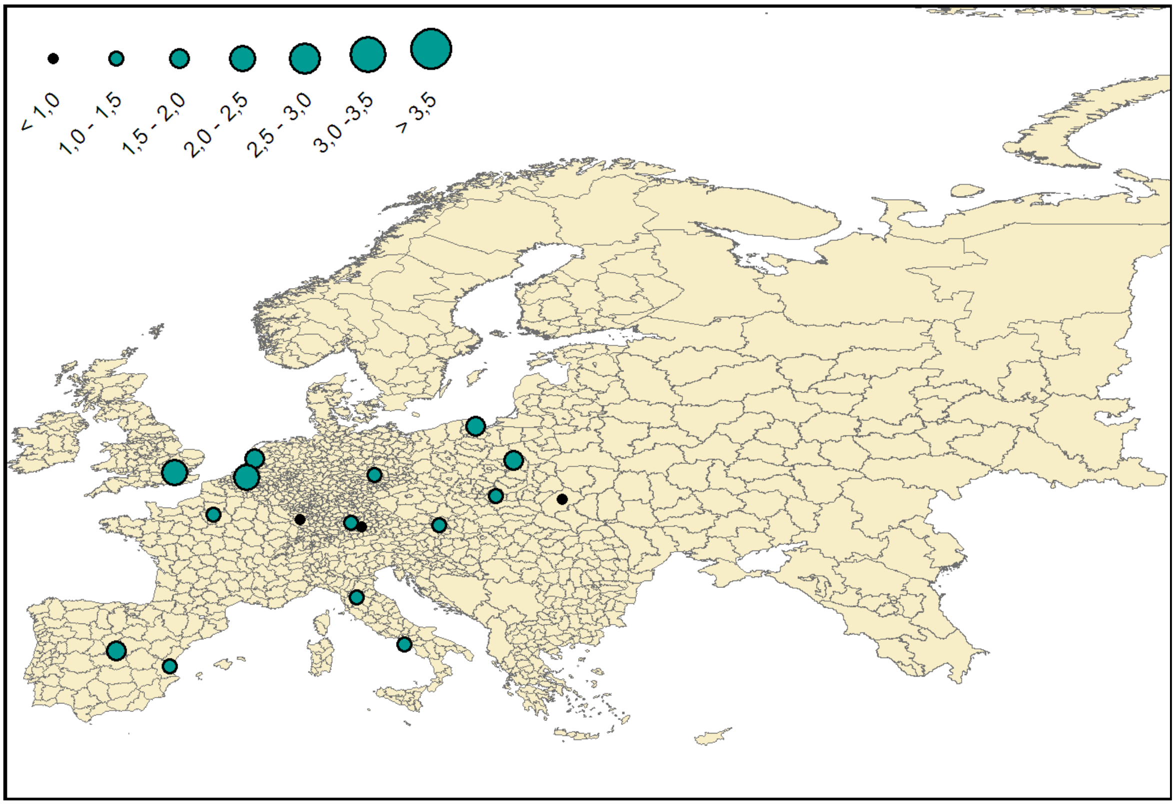
Source: Own calculations, based on data by Allen (2012).

Figure 3.28 Welfare ratios for craftsmen in European cities, 1600-1699



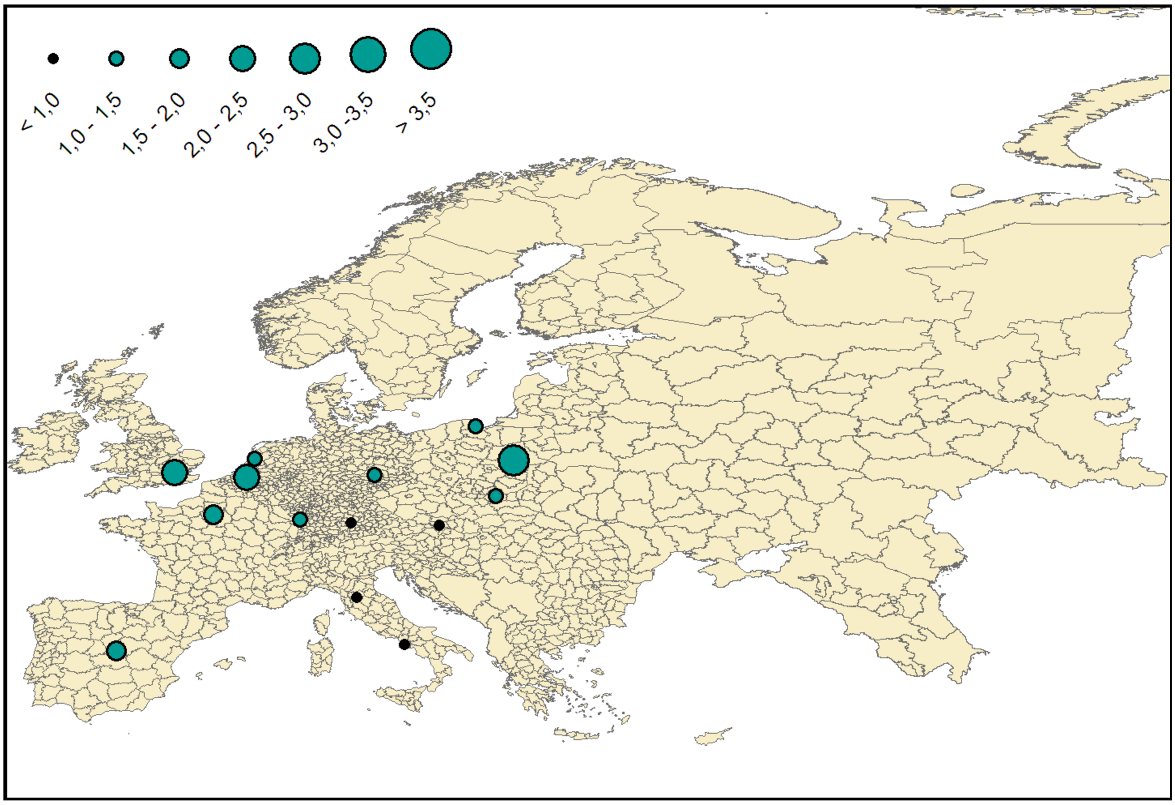
Source: Own calculations, based on data by Allen (2012).

Figure 3.29 Welfare ratios for craftsmen in European cities, 1700-1799



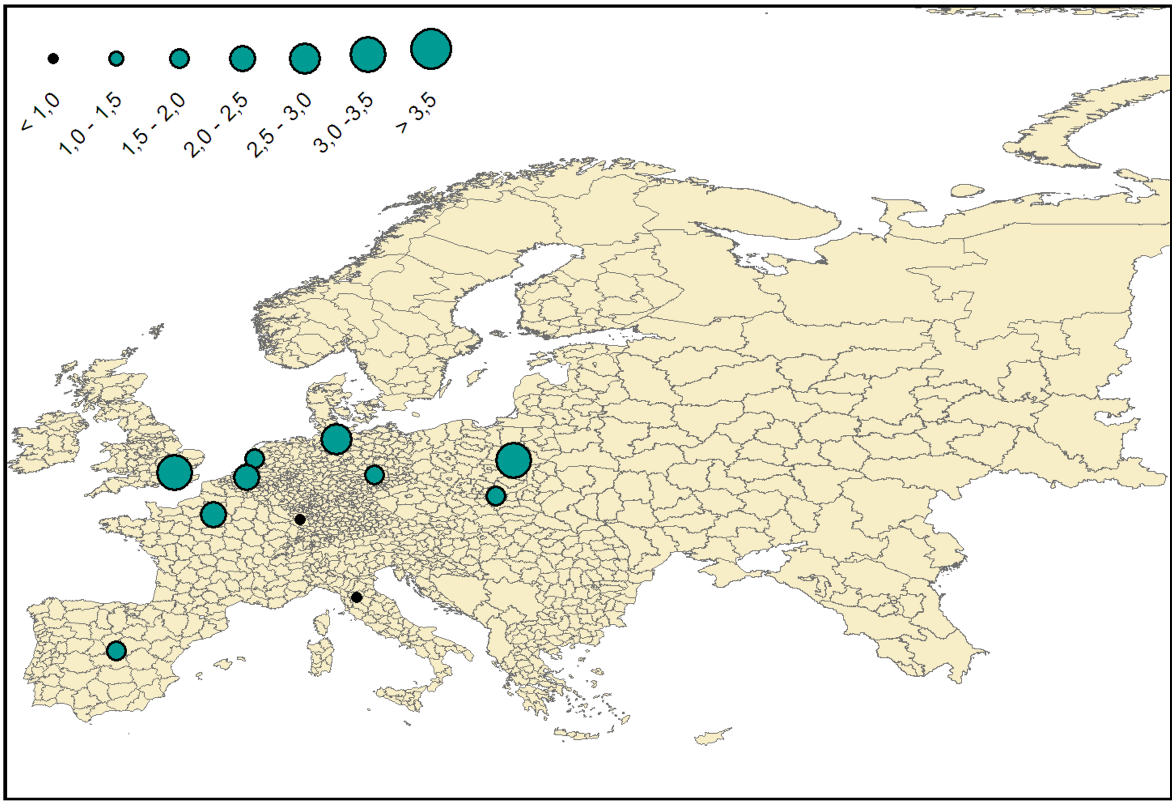
Source: Own calculations, based on data by Allen (2012).

Figure 3.30 Welfare ratios for craftsmen in European cities, 1800-1849



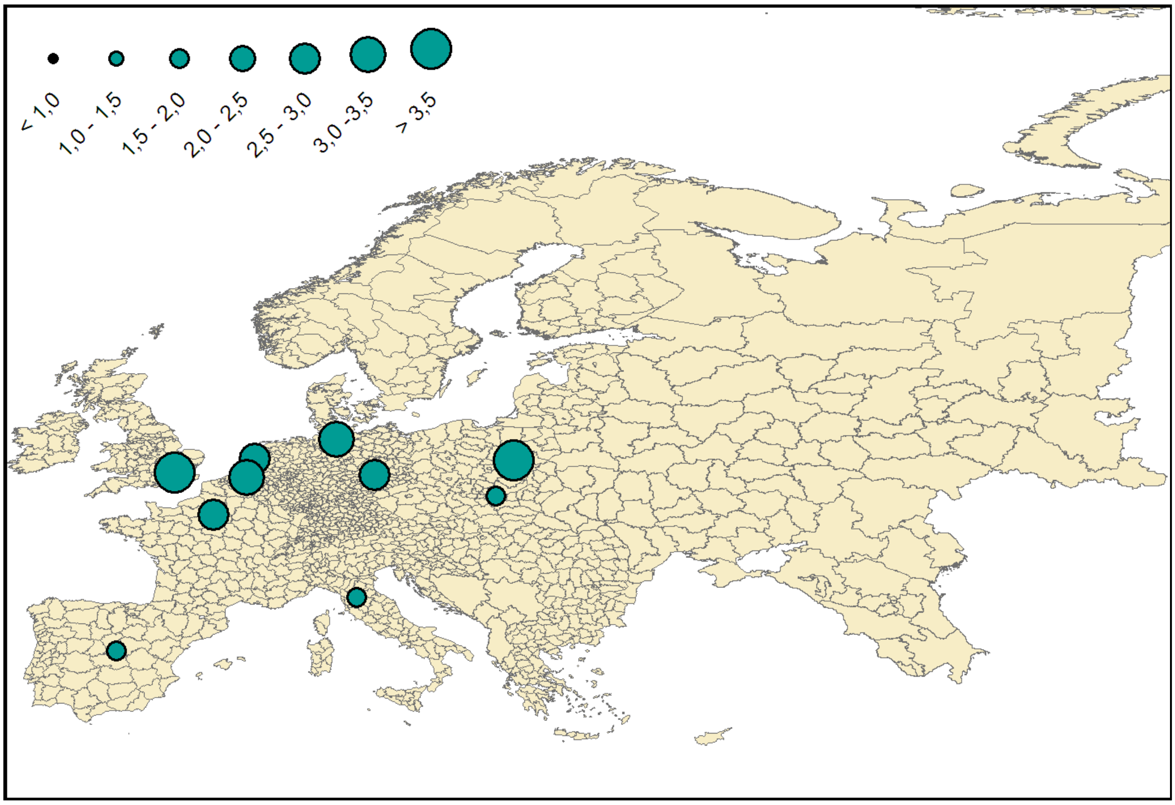
Source: Own calculations, based on data by Allen (2012).

Figure 3.31 Welfare ratios for craftsmen in European cities, 1850-1899



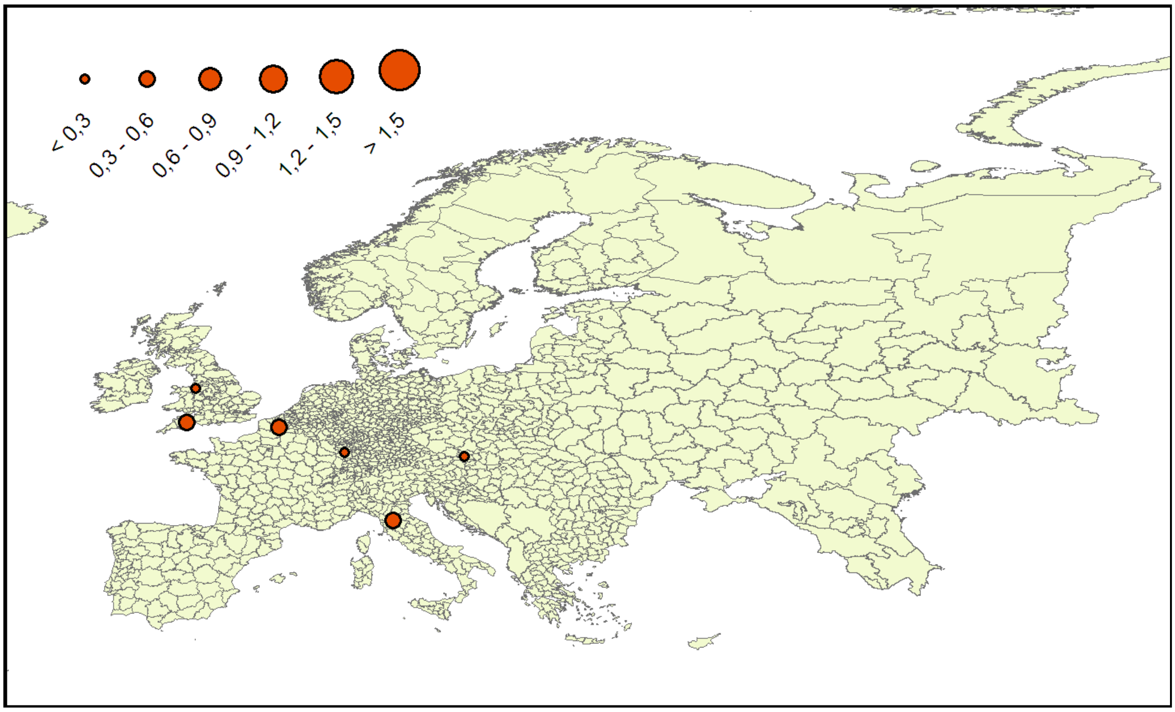
Source: Own calculations, based on data by Allen (2012).

Figure 3.32 Welfare ratios for craftsmen in European cities, 1900-1914



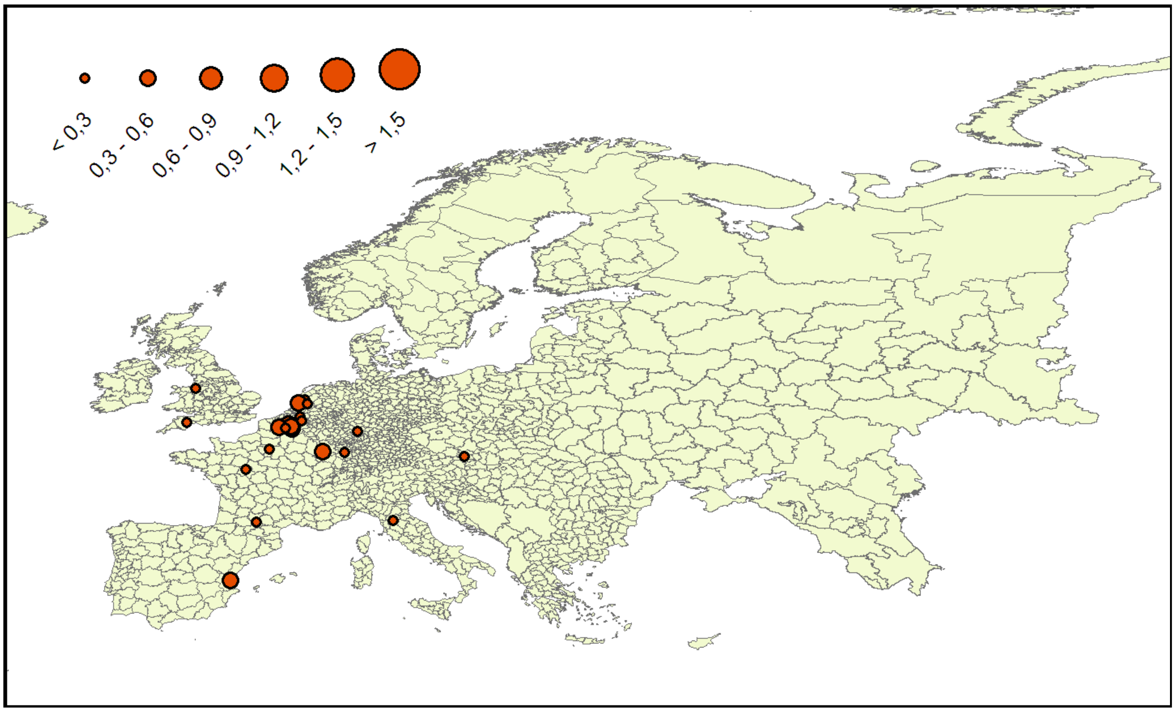
Source: Own calculations, based on data by Allen (2012).

Figure 3.33 Average wheat prices in European cities, 1300-1399



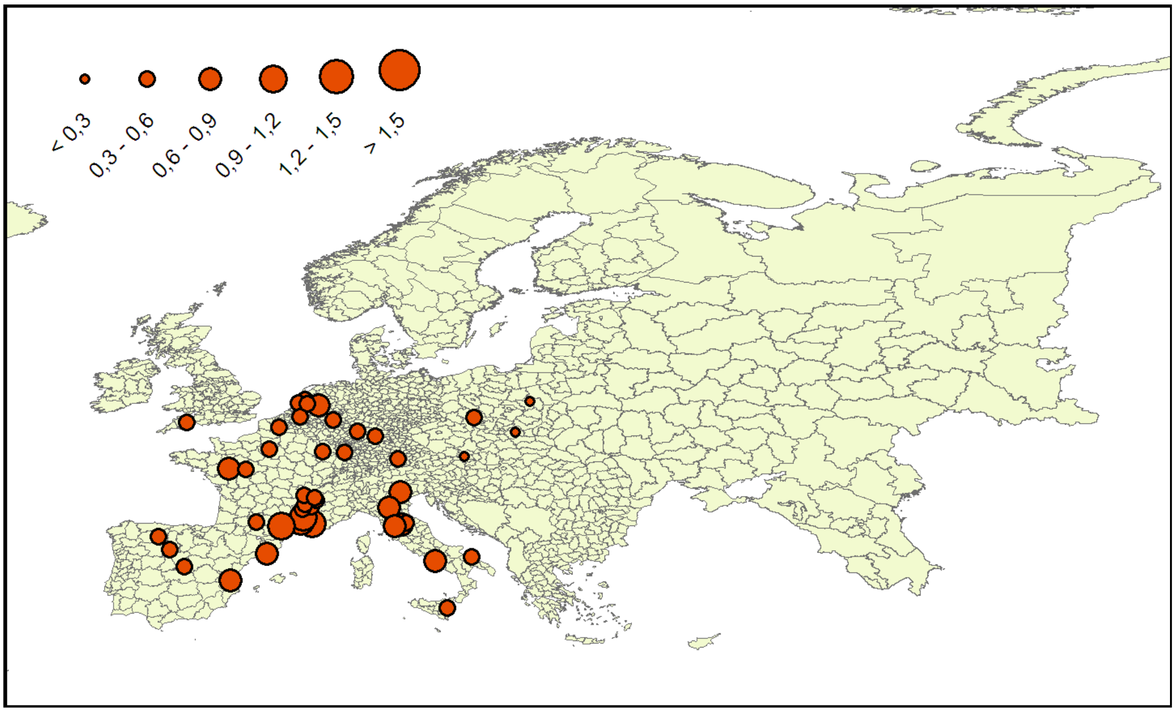
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.34 Average wheat prices in European cities, 1400-1499



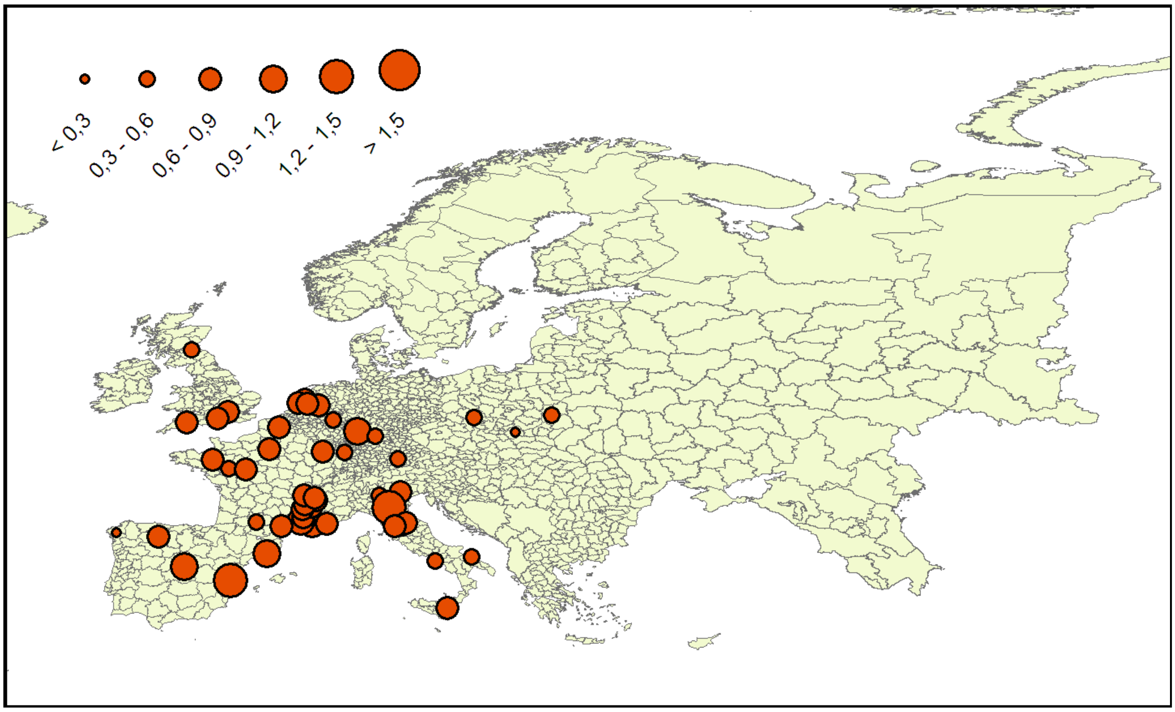
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.35 Average wheat prices in European cities, 1500-1599



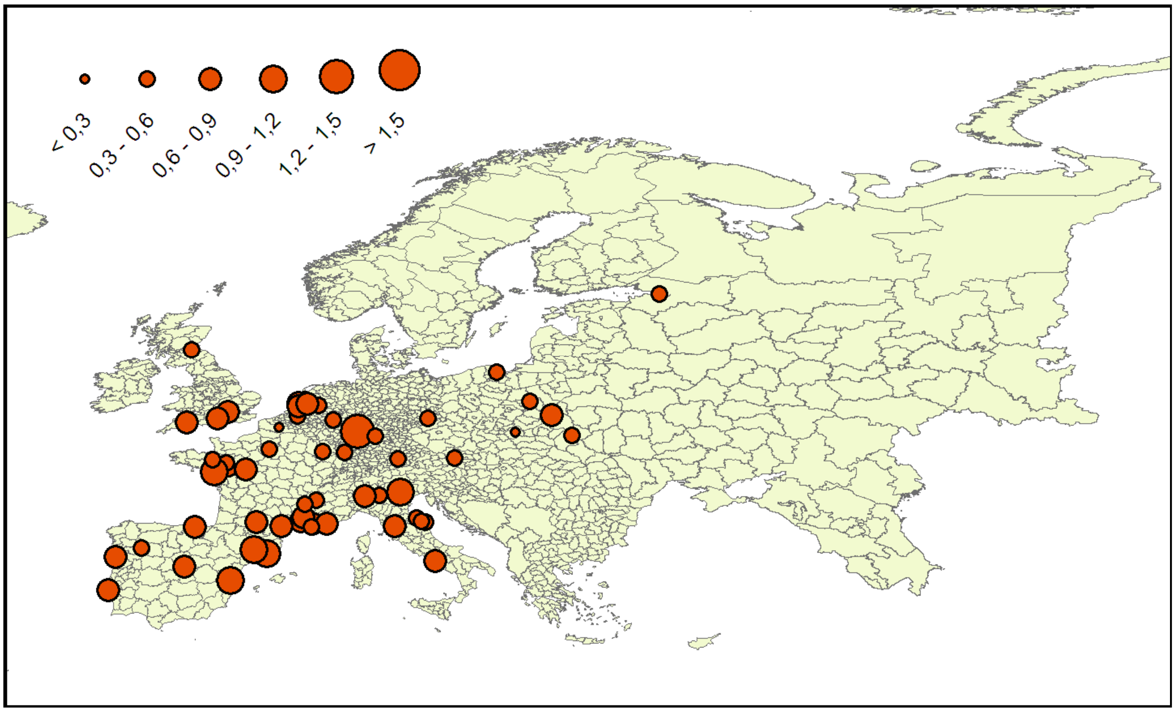
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.36 Average wheat prices in European cities, 1600-1699



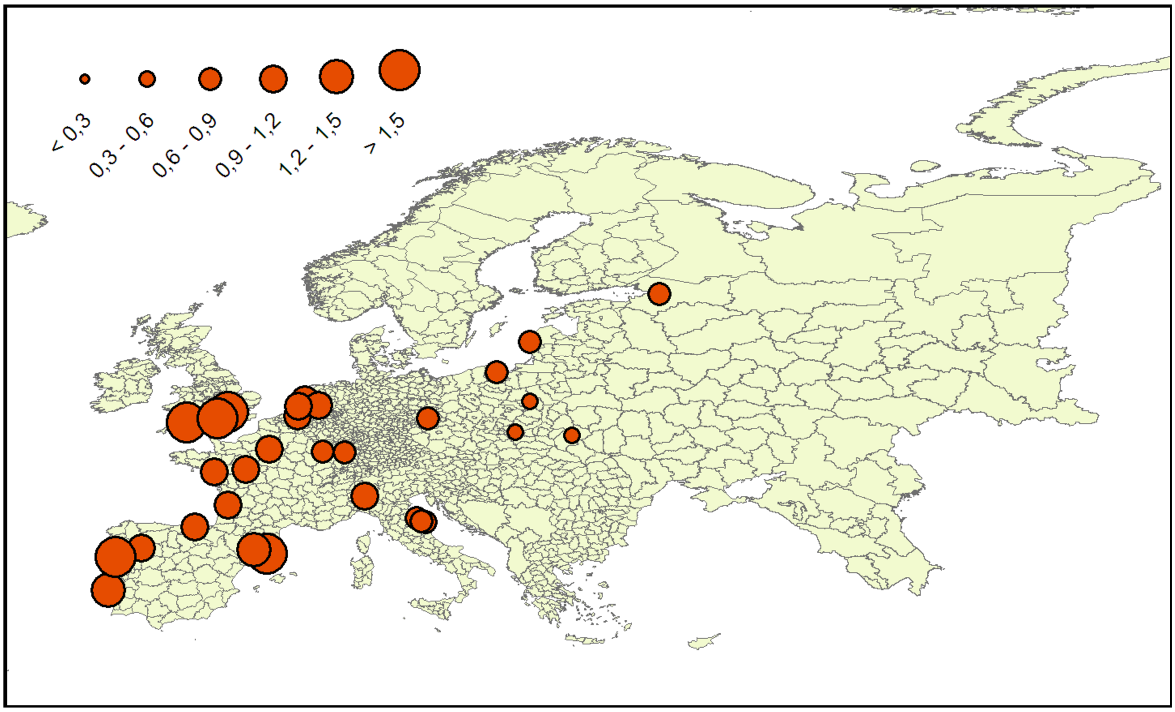
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.37 Average wheat prices in European cities, 1700-1799



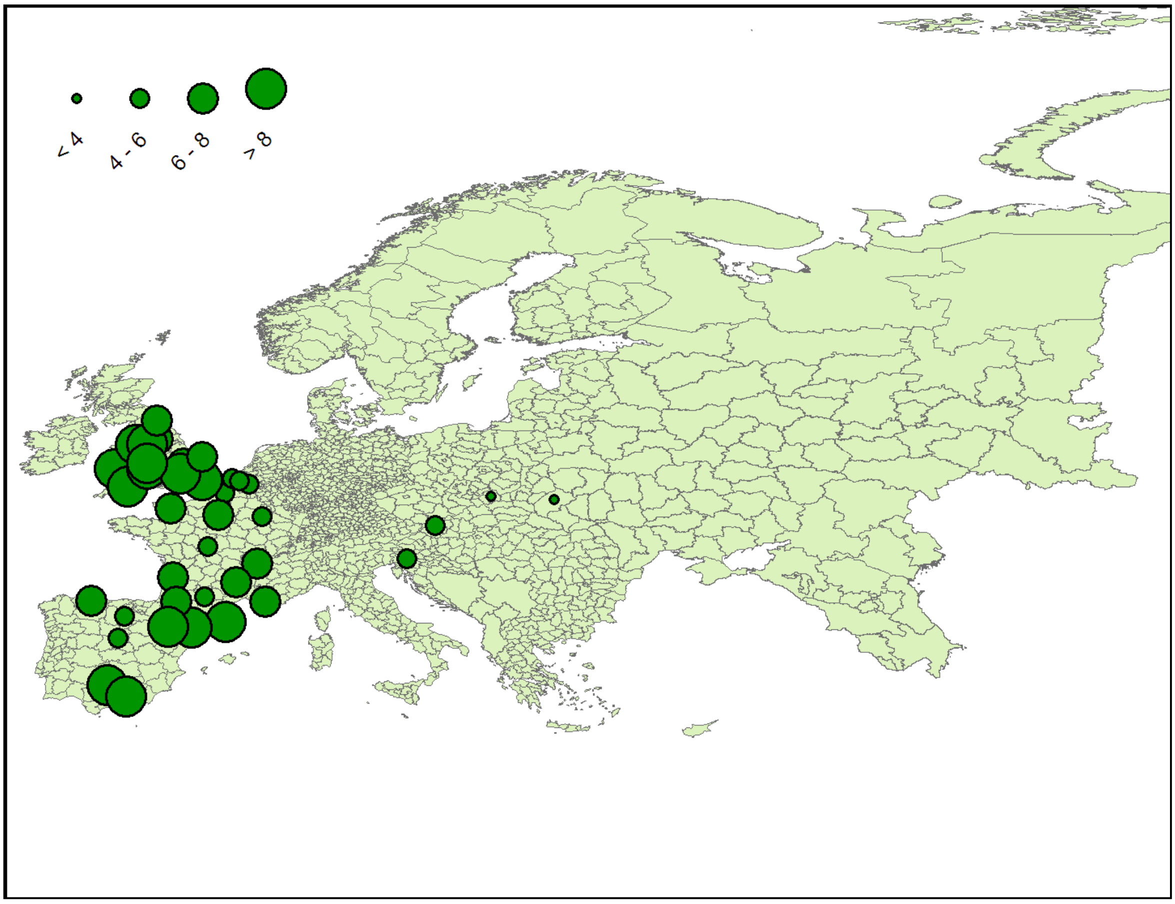
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.38 Average wheat prices in European cities, 1800-1824



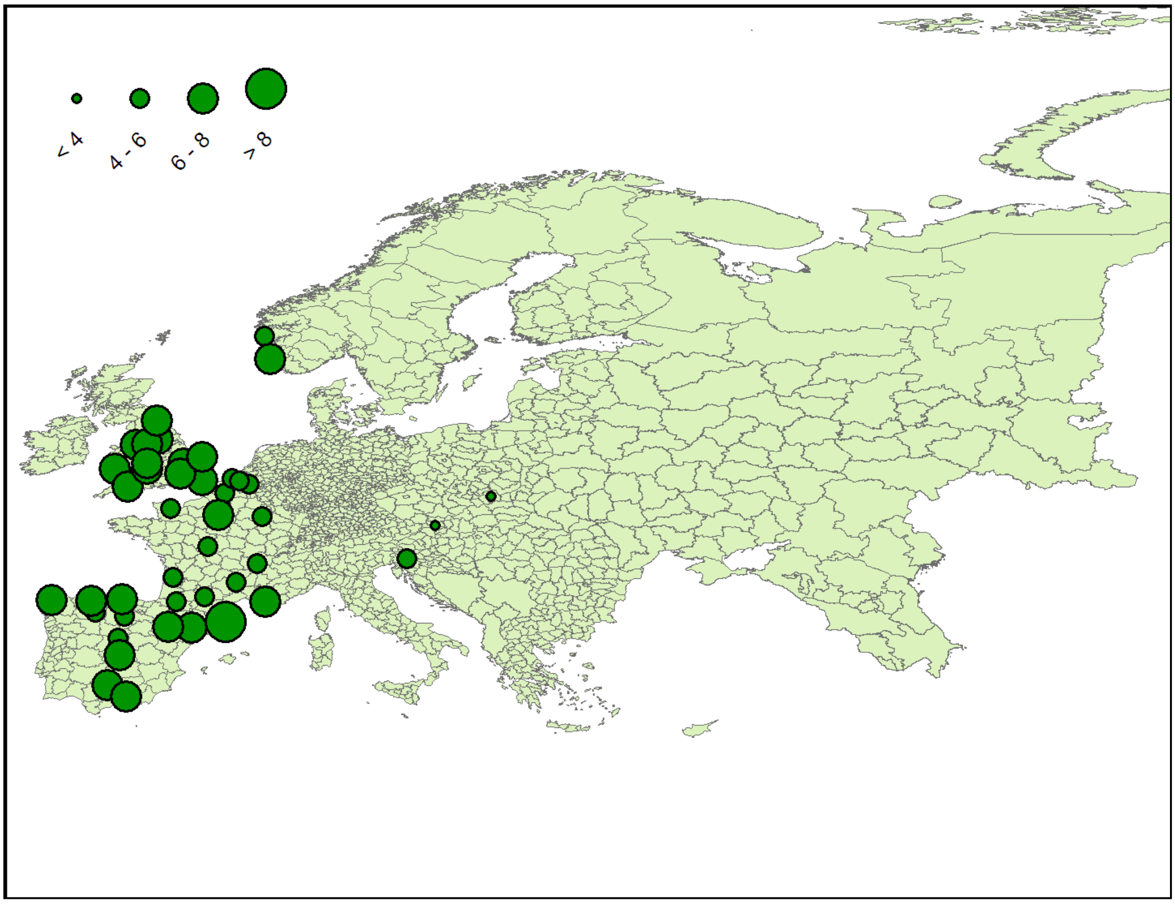
Note: Prices are in grams of silver for one litre.
Source: Own calculations, based on data by Allen and Unger (2012).

Figure 3.39 Average wheat prices in European cities, 1800-1824



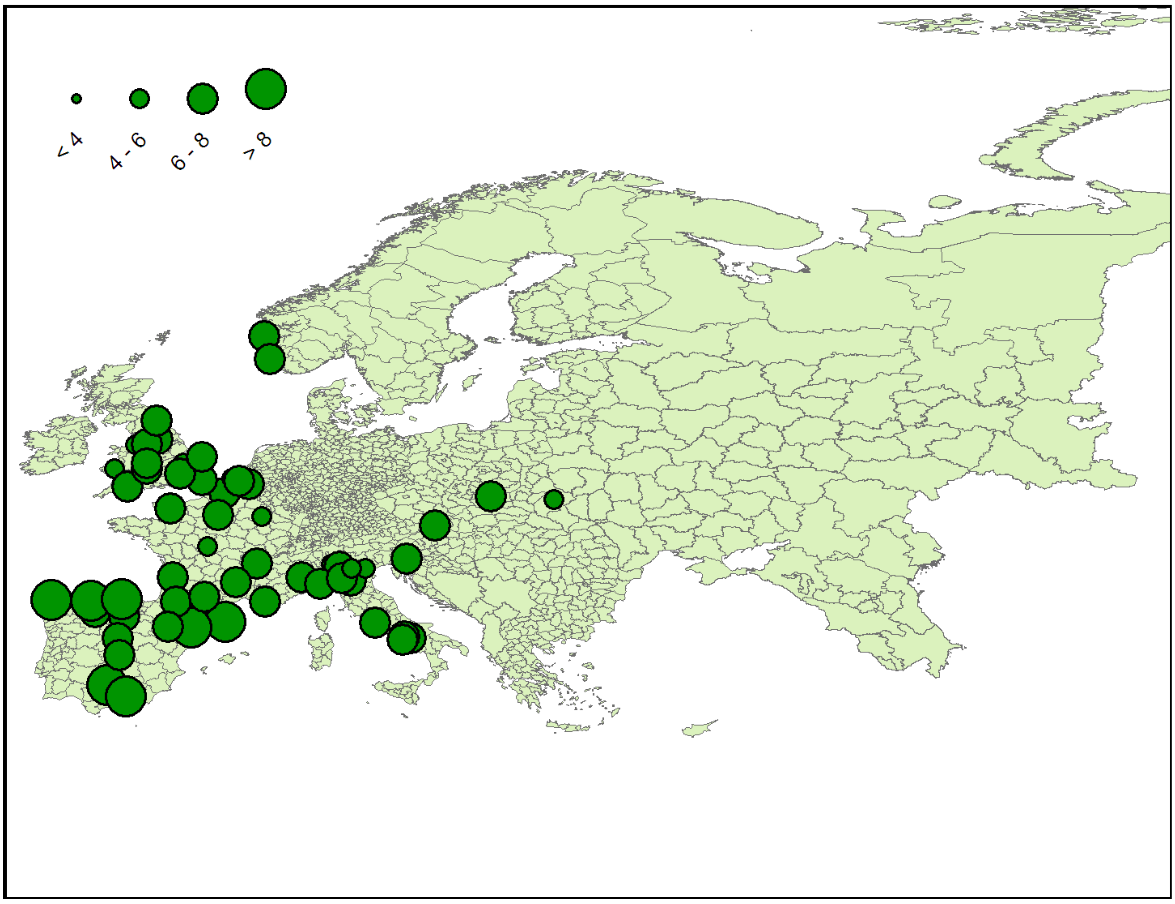
Note: Prices are given in dollars per 100 kg of wheat.
Source: Own calculations, data provided by Jacks (2005, 2006).

Figure 3.40 Average wheat prices in European cities, 1825-1849



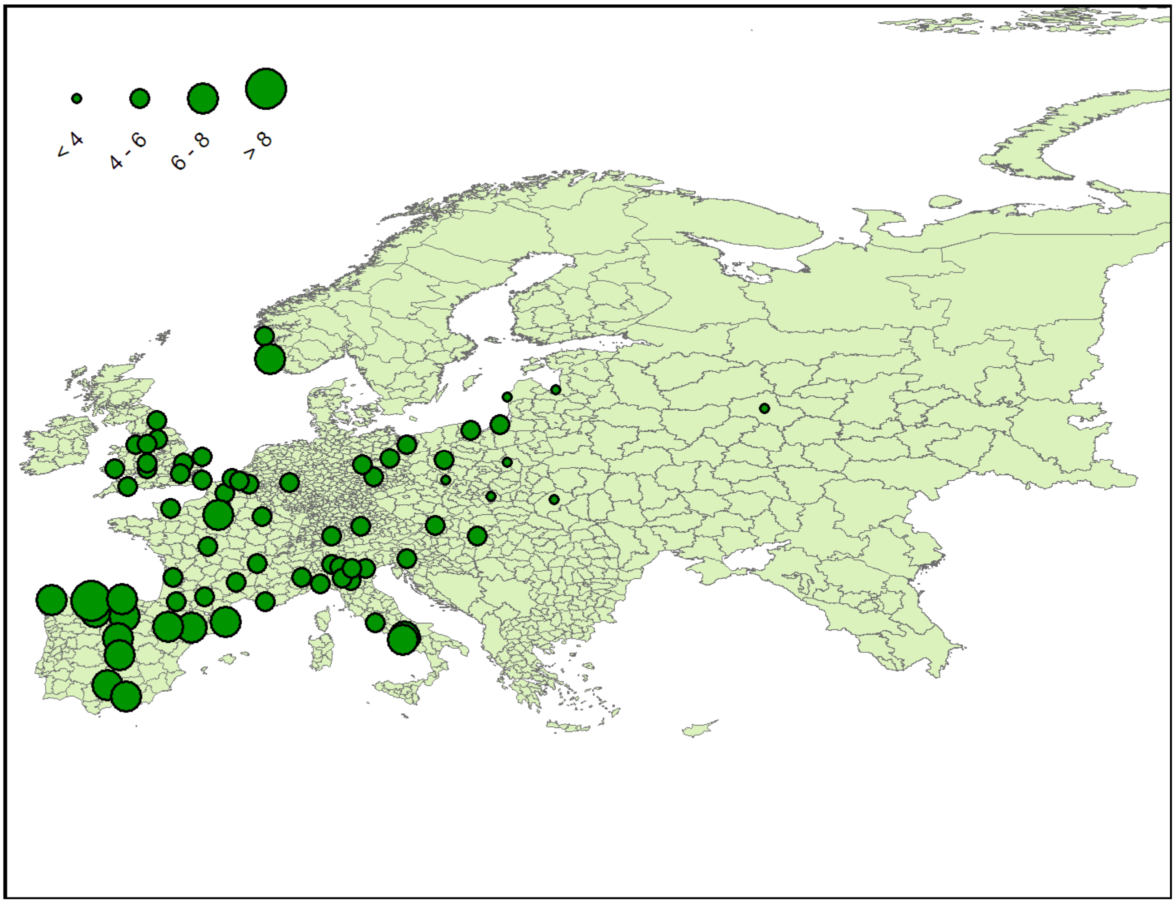
Note: Prices are given in dollars per 100 kg of wheat.
Source: Own calculations, data provided by Jacks (2005, 2006).

Figure 3.41 Average wheat prices in European cities, 1850-1874



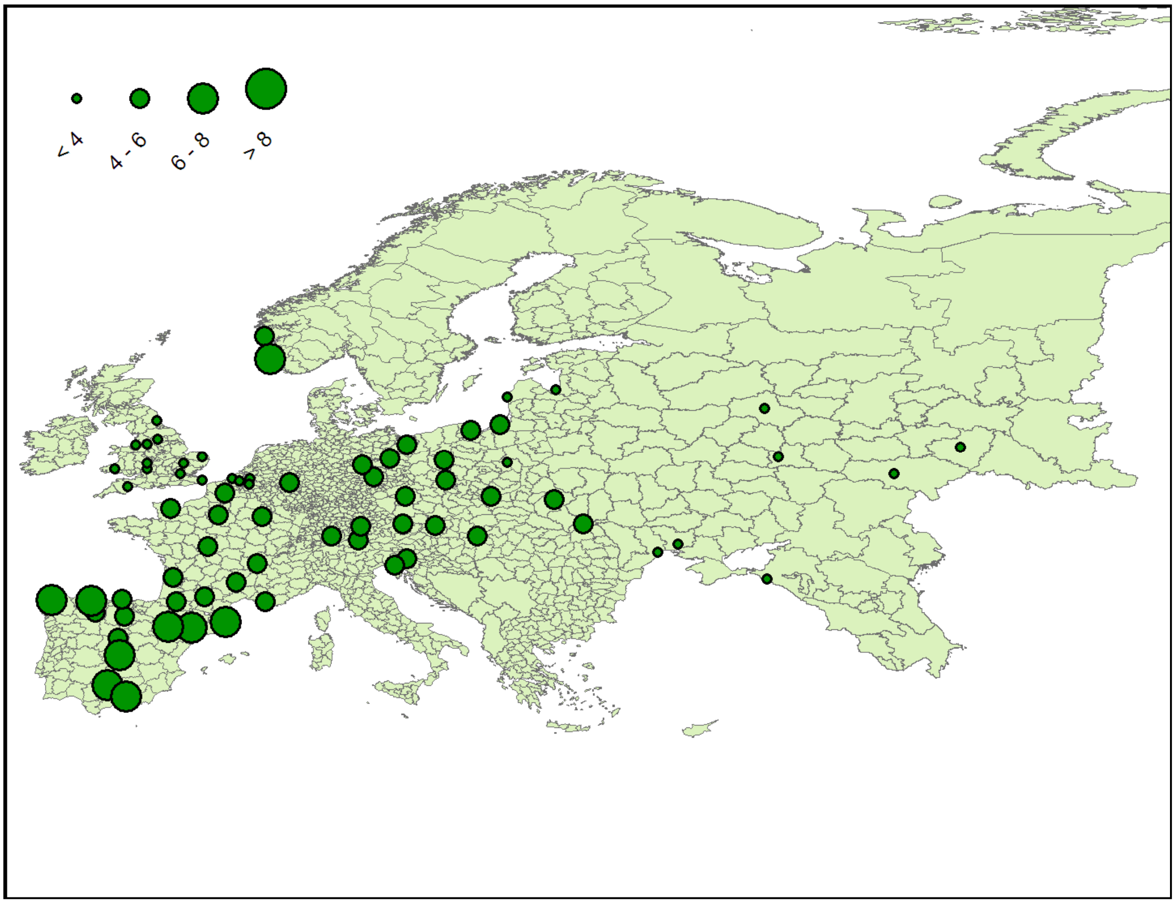
Note: Prices are given in dollars per 100 kg of wheat.
Source: Own calculations, data provided by Jacks (2005, 2006).

Figure 3.42 Average wheat prices in European cities, 1875-1899



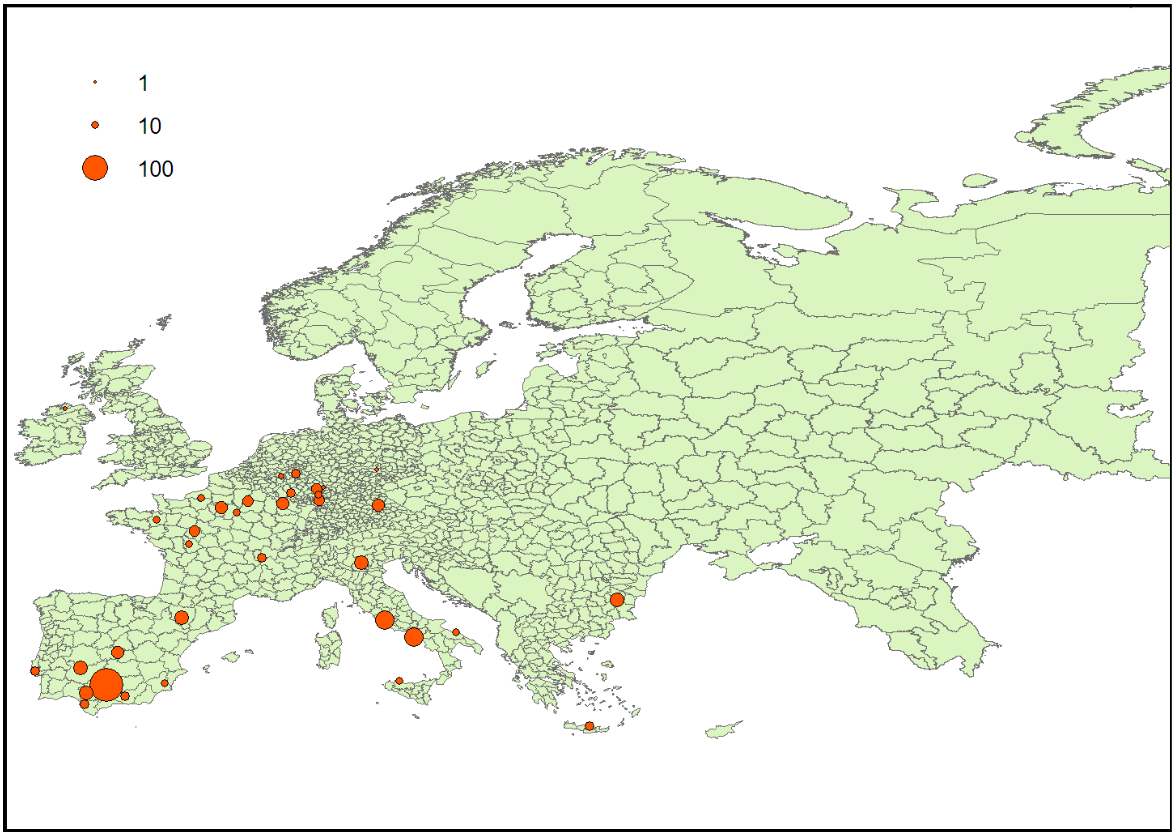
Note: Prices are given in dollars per 100 kg of wheat.
Source: Own calculations, data provided by Jacks (2005, 2006).

Figure 3.43 Average wheat prices in European cities, 1900-1913



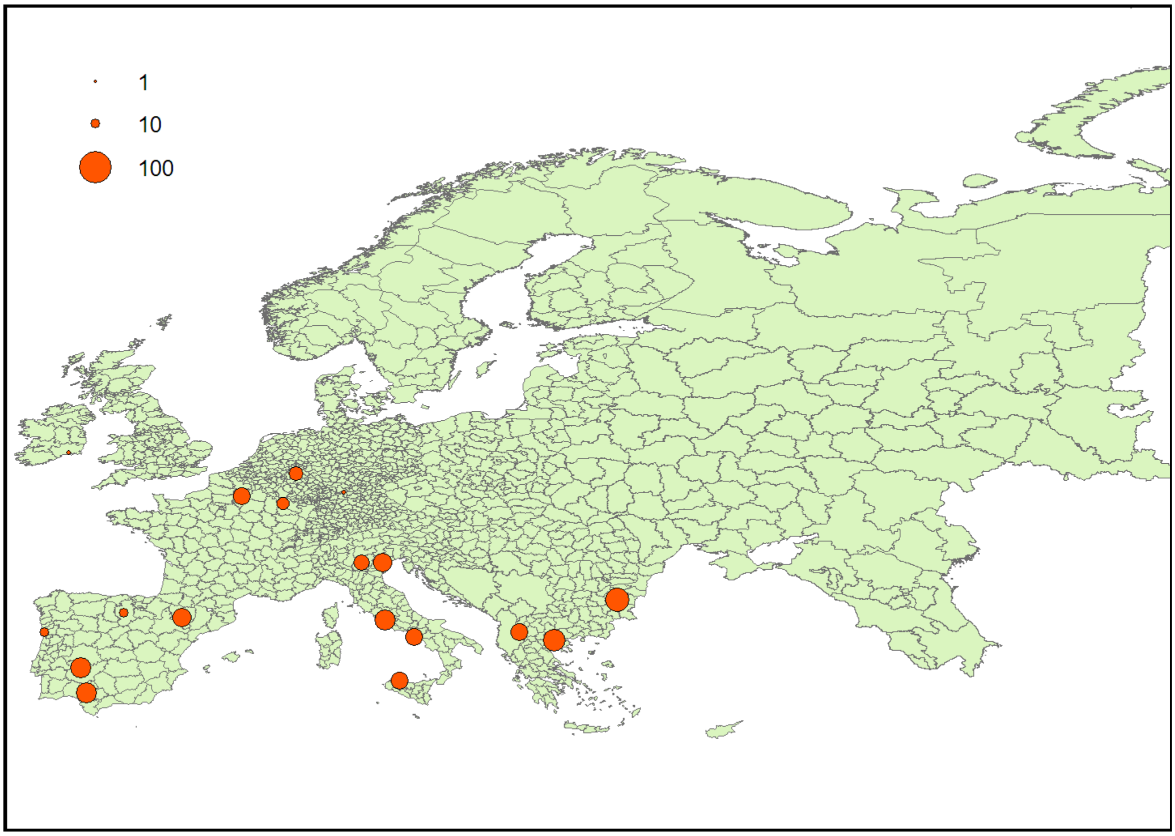
Note: Prices are given in dollars per 100 kg of wheat.
Source: Own calculations, data provided by Jacks (2005, 2006).

Figure 3.44 Location and size of European cities, 800



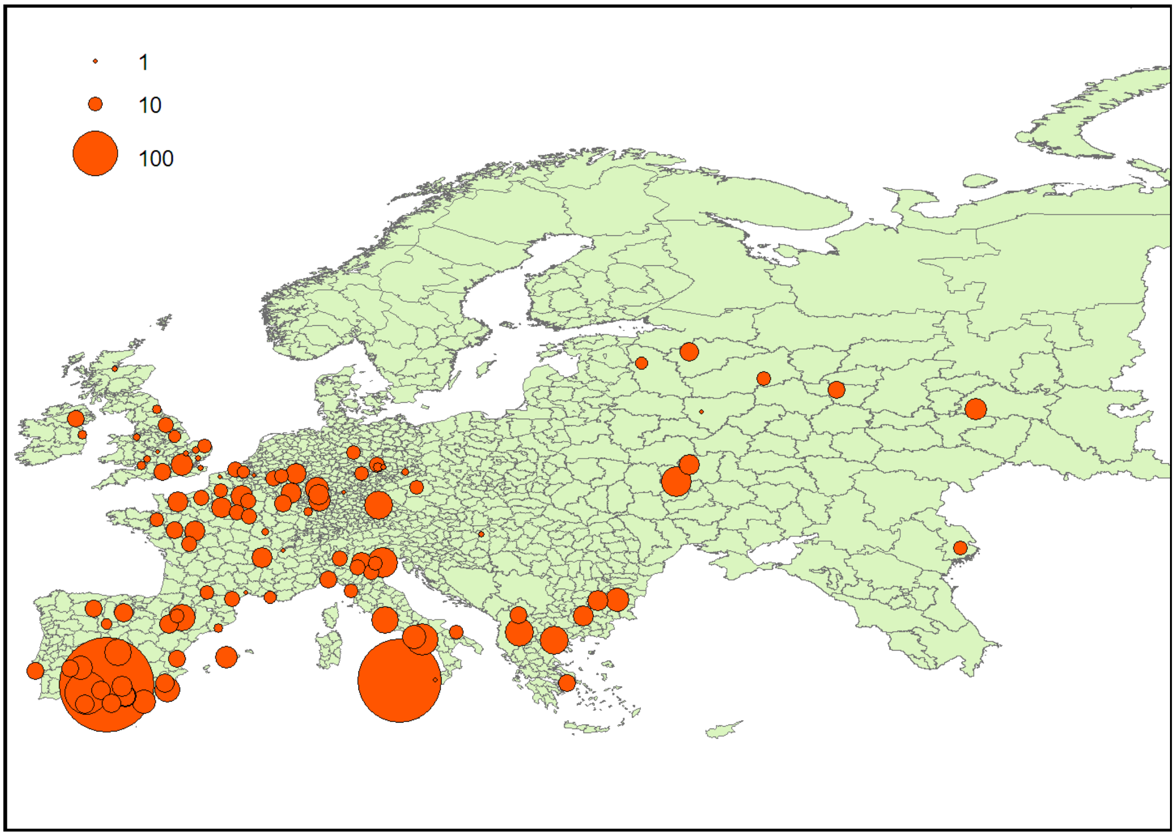
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.45 Location and size of European cities, 900



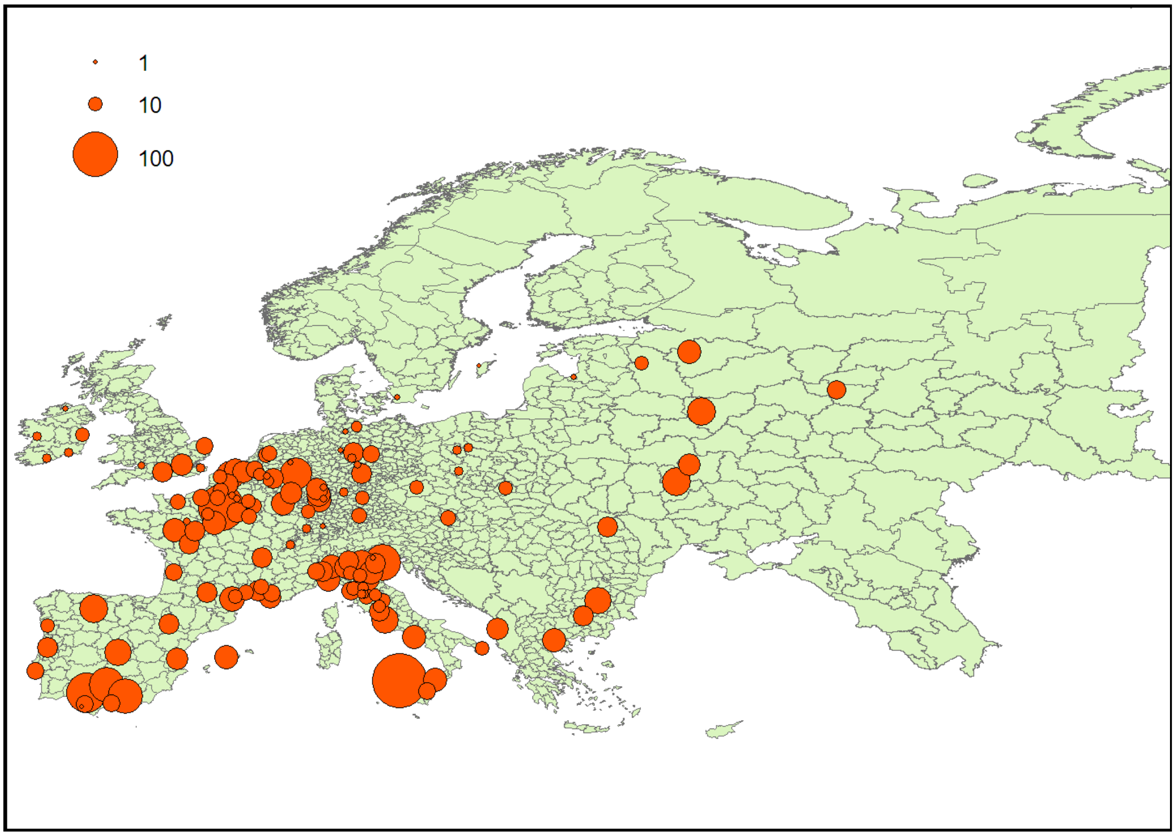
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.46 Location and size of European cities, 1000



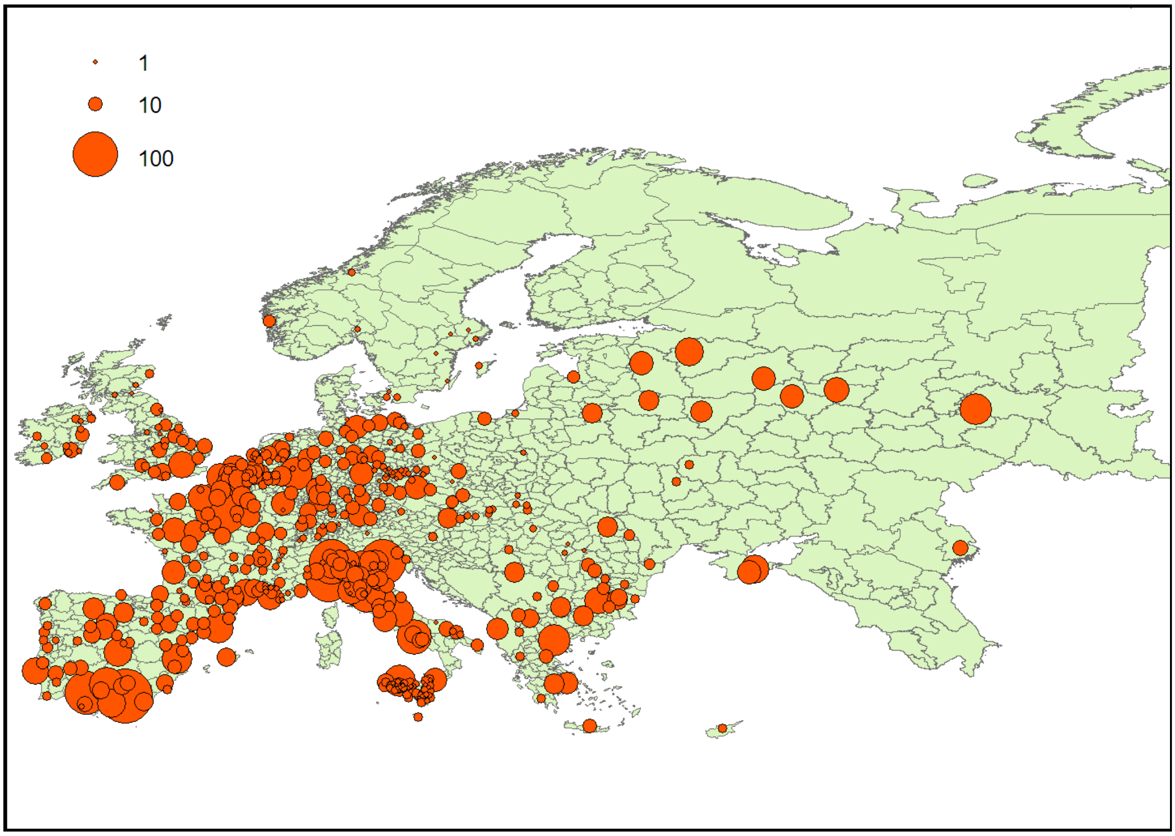
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.47 Location and size of European cities, 1200



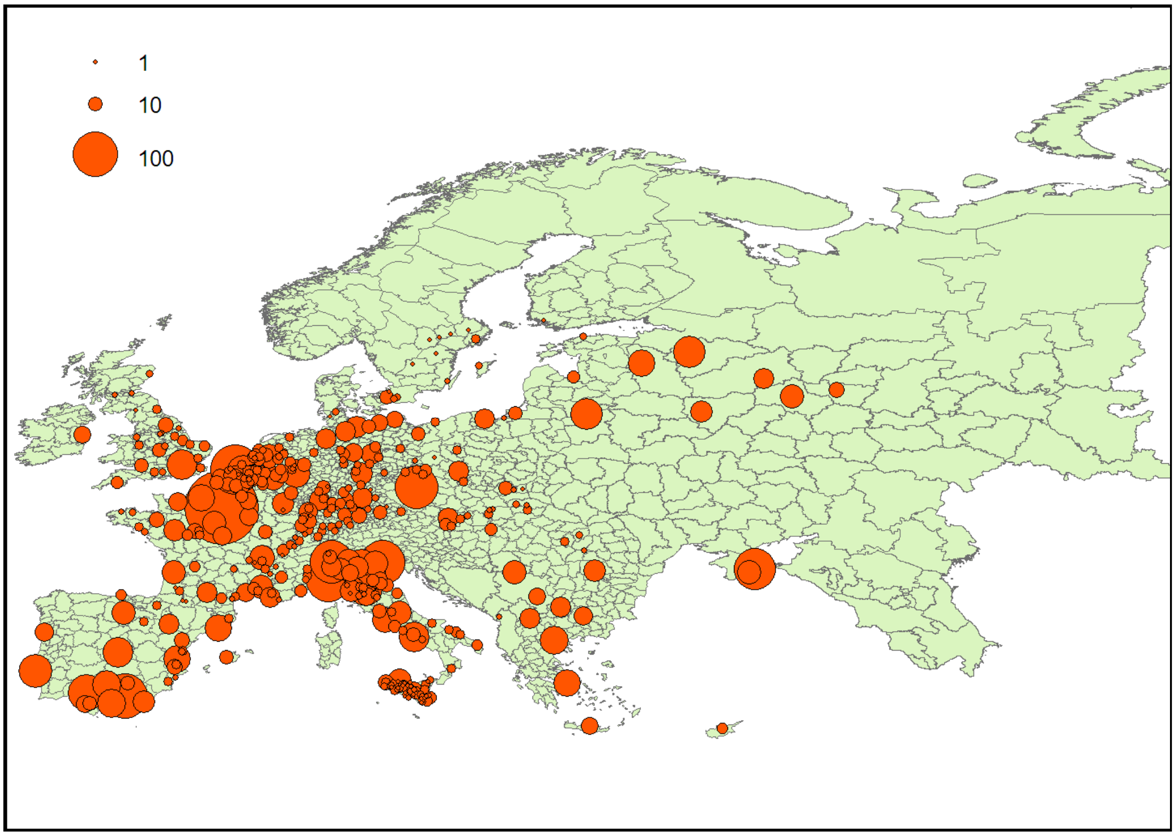
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.48 Location and size of European cities, 1300



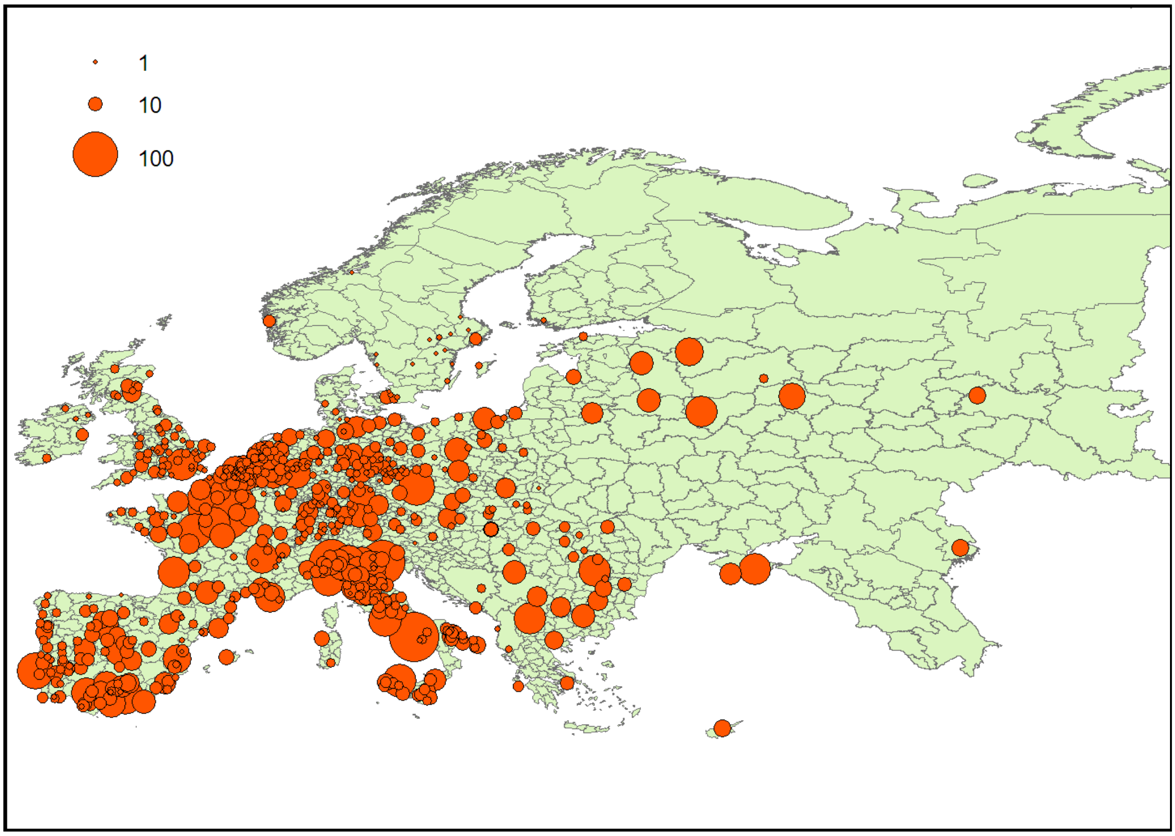
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.49 Location and size of European cities, 1400



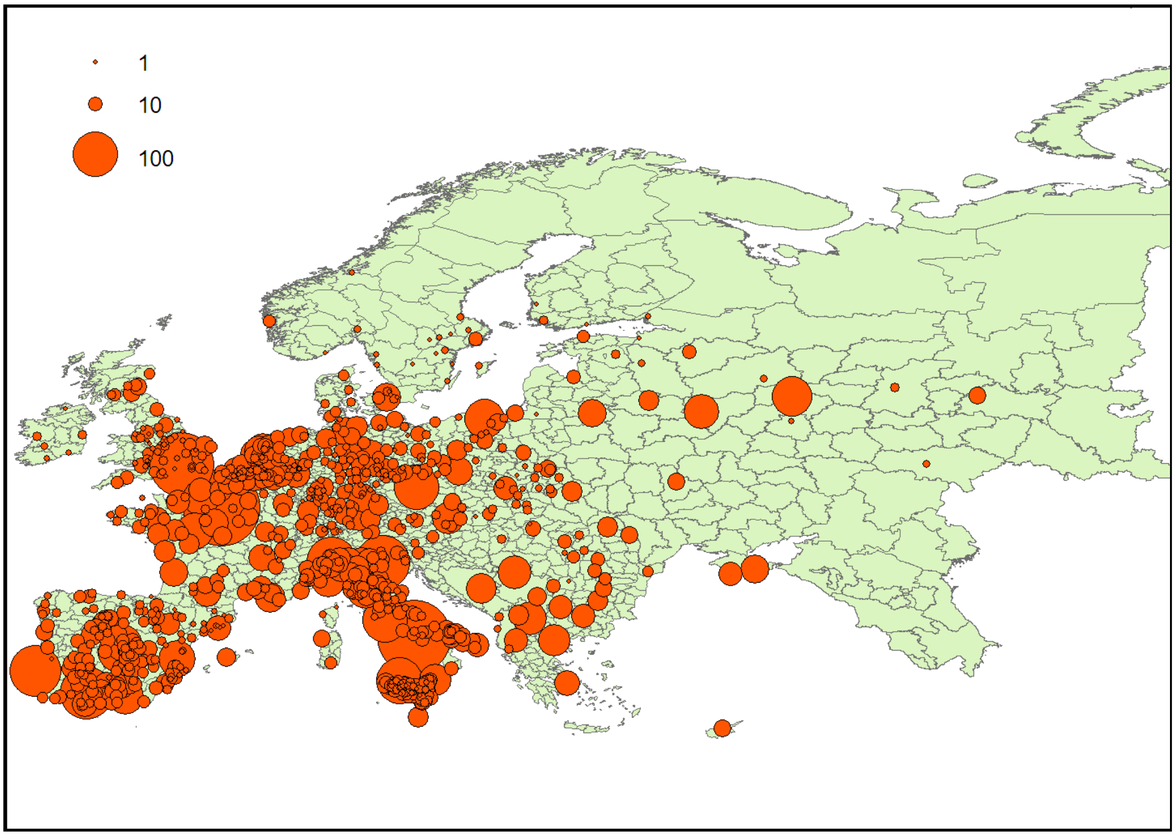
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.50 Location and size of European cities, 1500



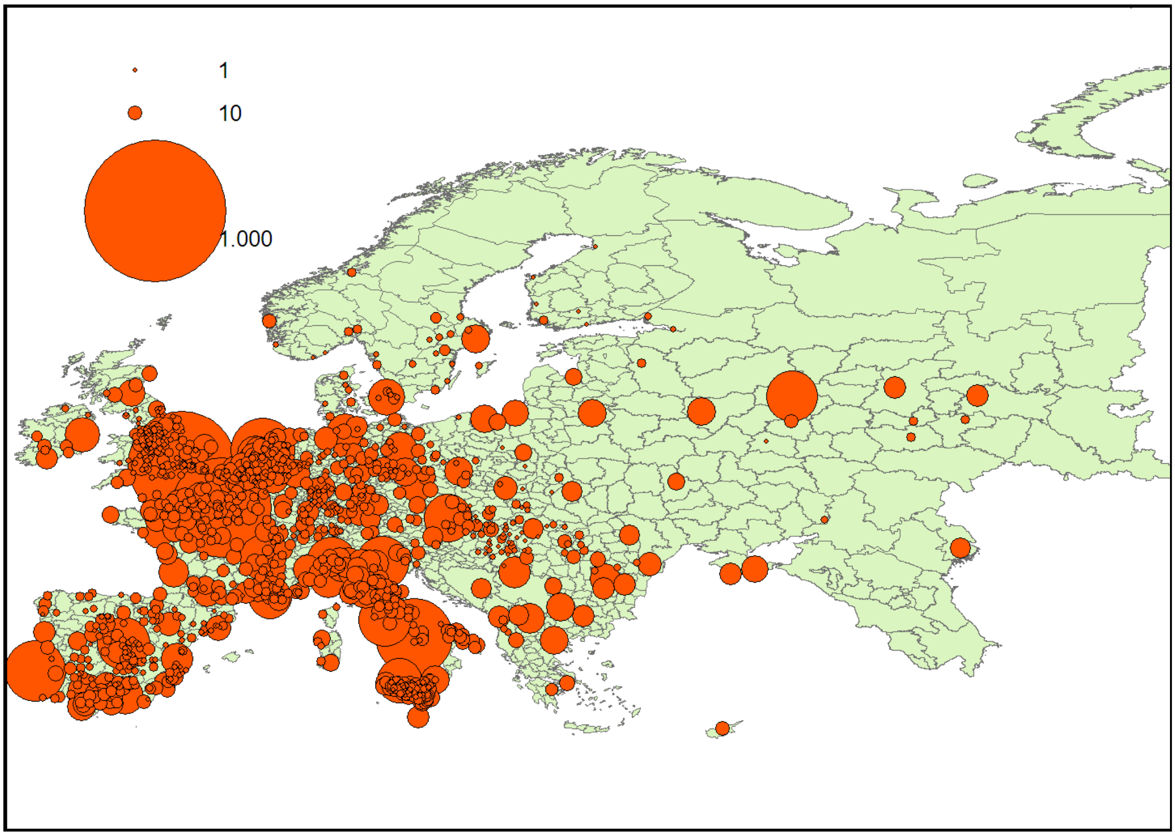
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.51 Location and size of European cities, 1600



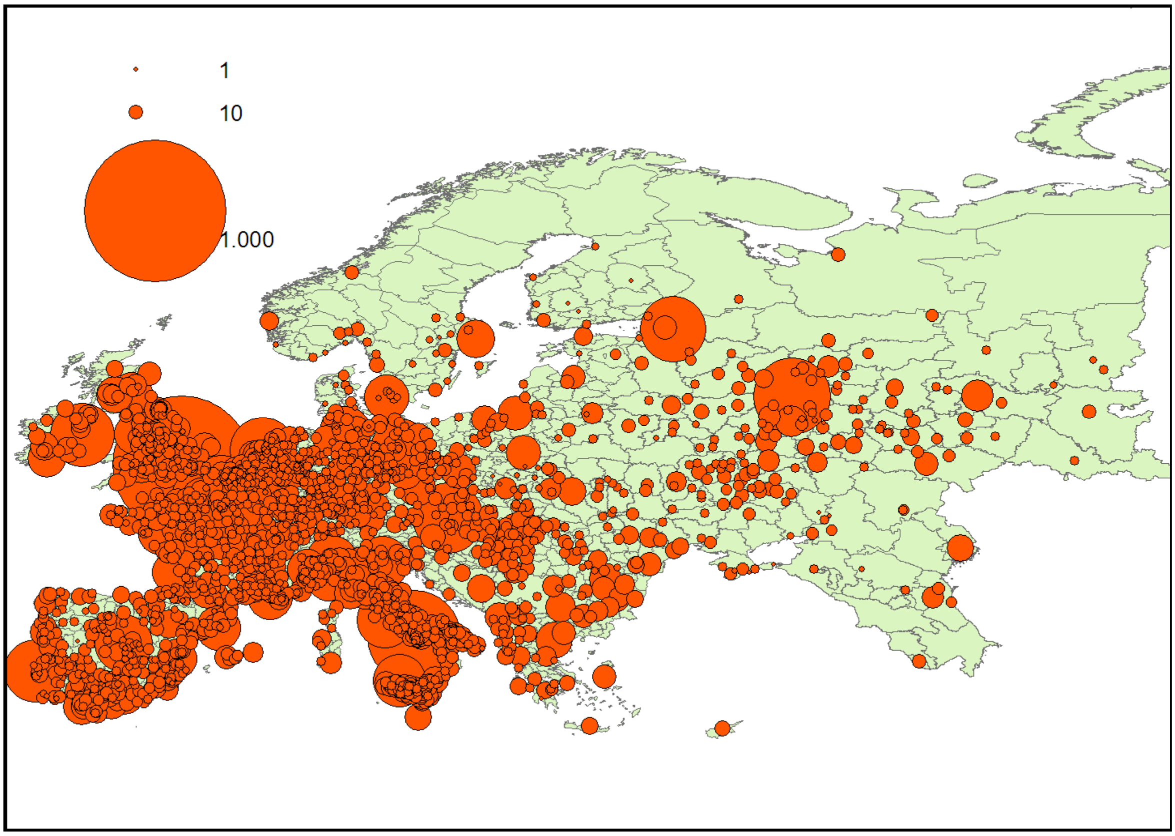
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.52 Location and size of European cities, 1700



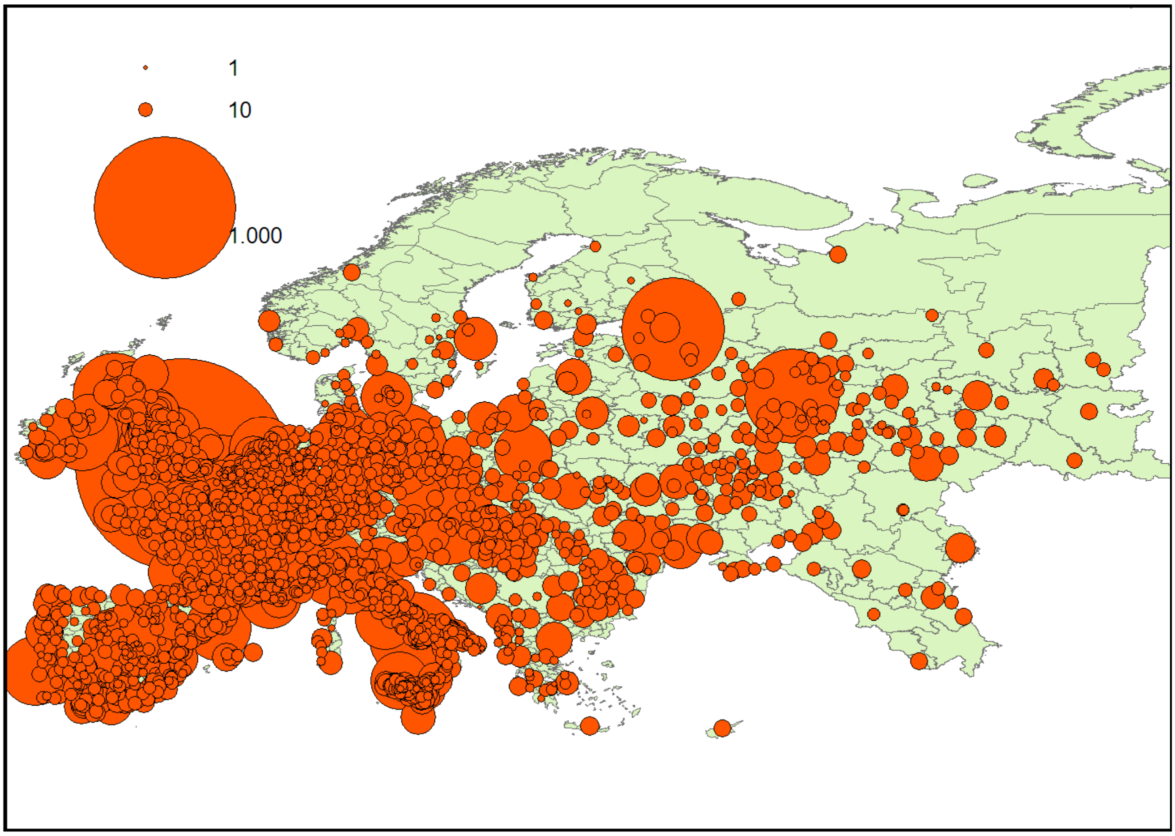
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.53 Location and size of European cities, 1800



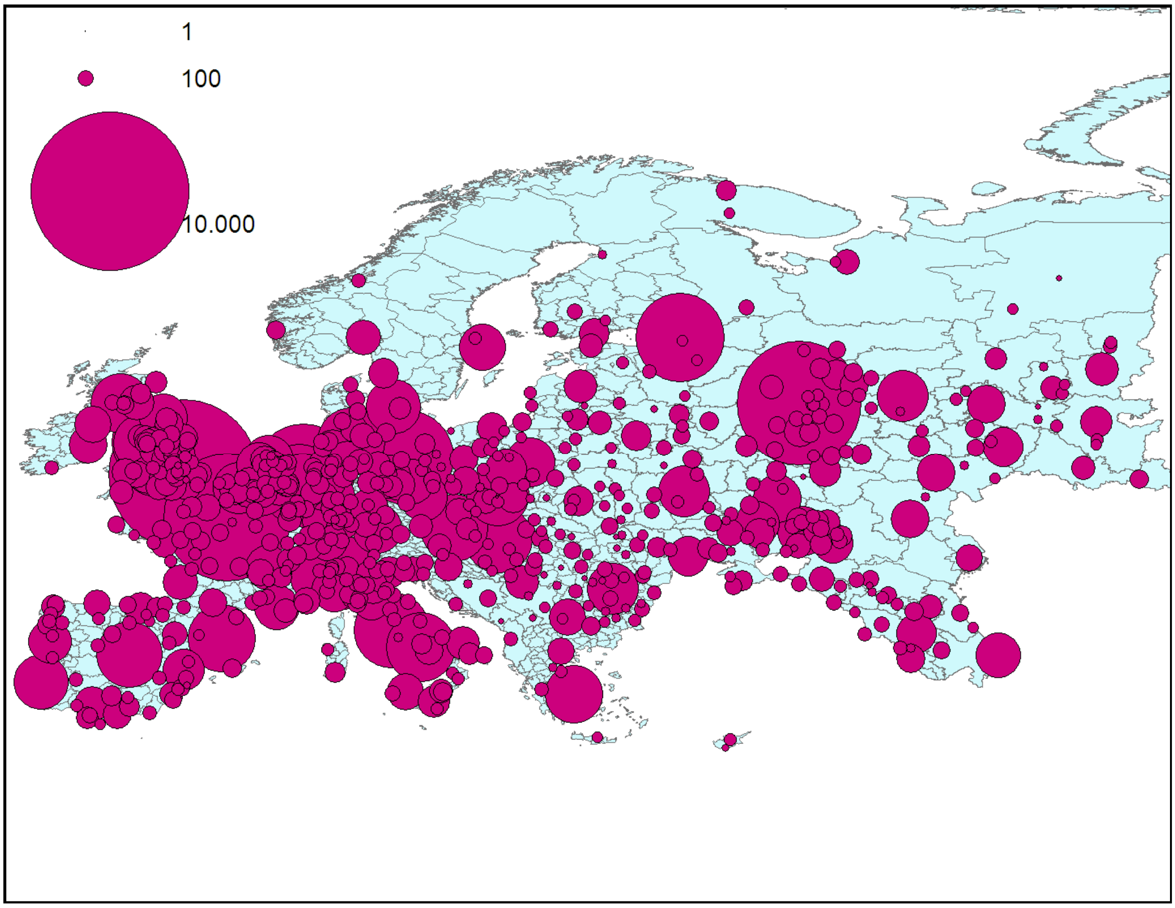
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.54 Location and size of European cities, 1850



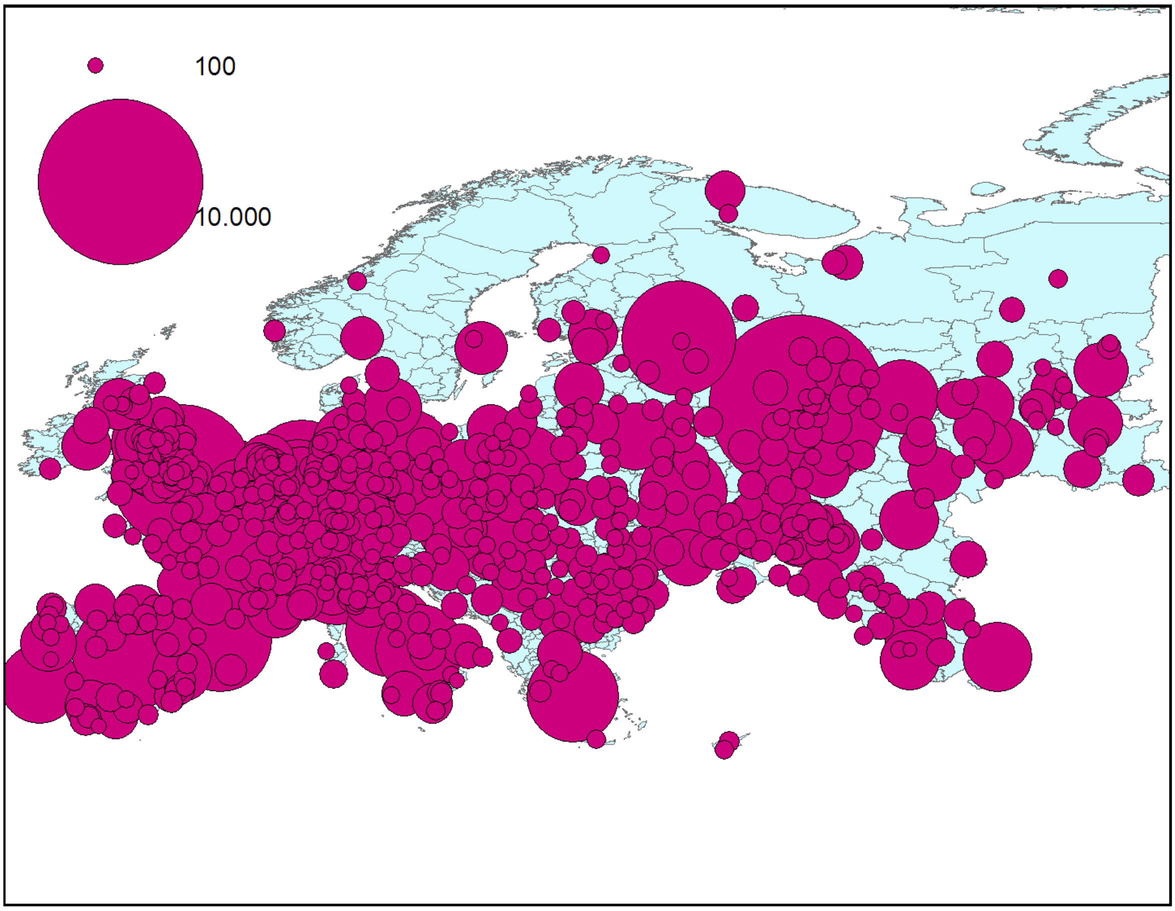
Note: City size in thousand inhabitants.
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988).

Figure 3.55 Location and size of European agglomerations, 1950



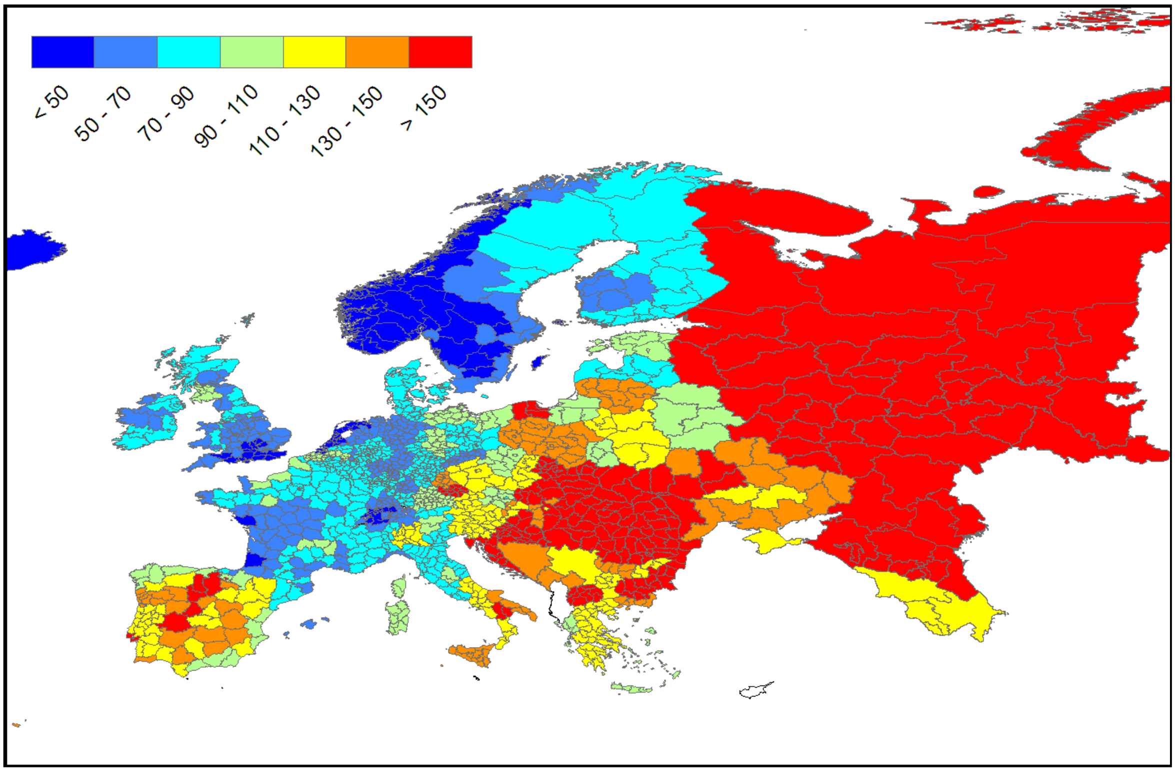
Note: Agglomeration size in thousand inhabitants.
Source: Own graphical presentation of data provided by Moriconi-Ebrard (1994).

Figure 3.56 Location and size of European agglomerations, 1990



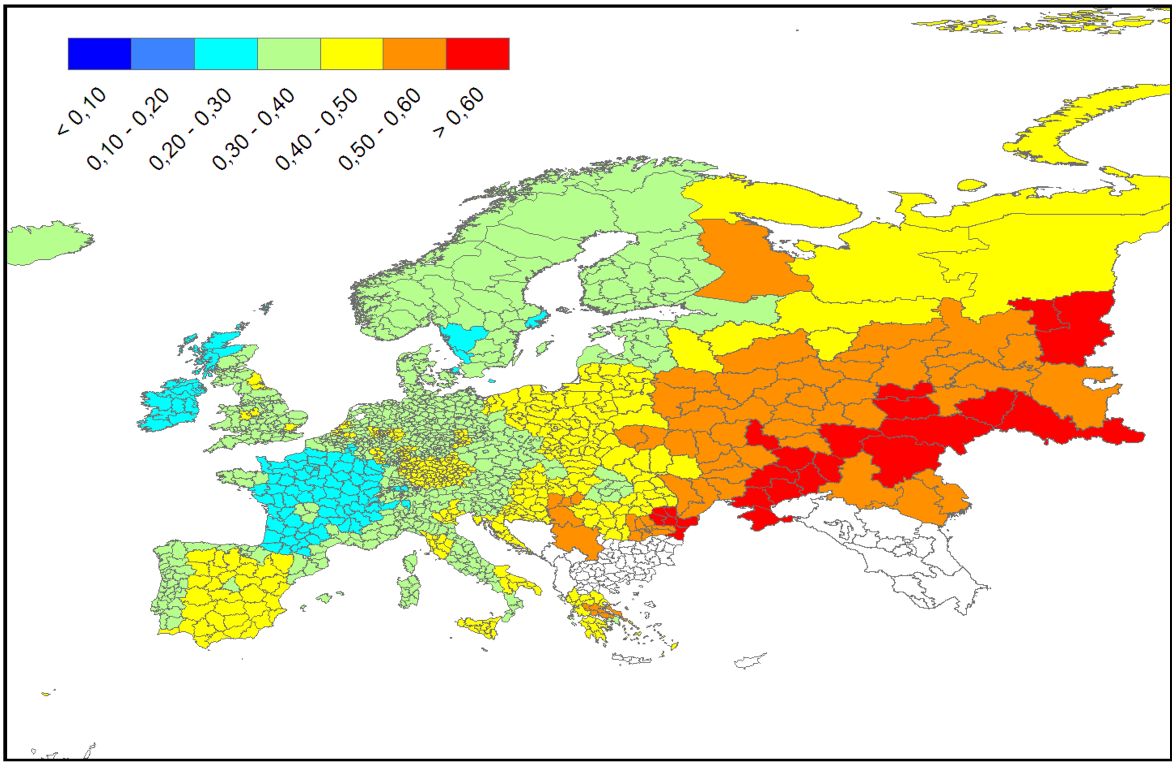
Note: Agglomeration size in thousand inhabitants.
Source: Own graphical presentation of data provided by Moriconi-Ebrard (1994).

Figure 3.57 Infant mortality, 1930-31



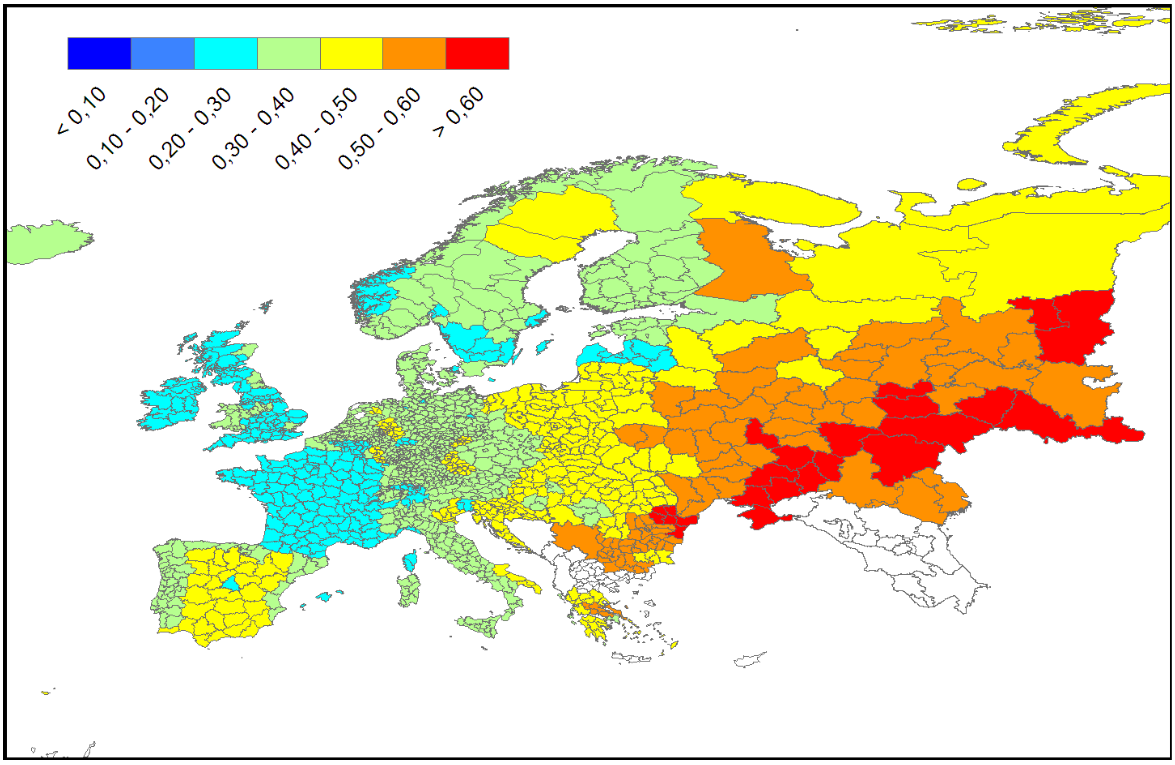
Note: Infant mortality is defined as "number of infant deaths per 1000 live births" (Kirk 1946, p. 261).
Source: Own calculations, data provided by Kirk (1946).

Figure 3.58 Total fertility, 1870



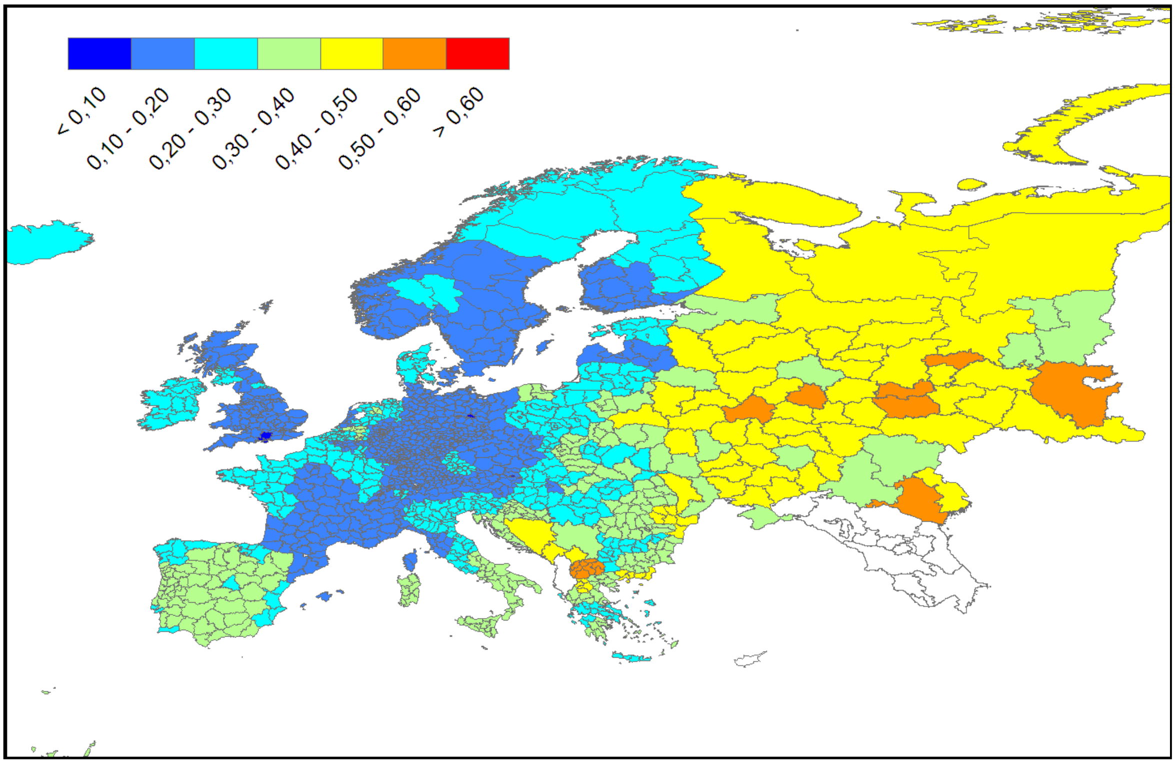
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.59 Total fertility, 1900



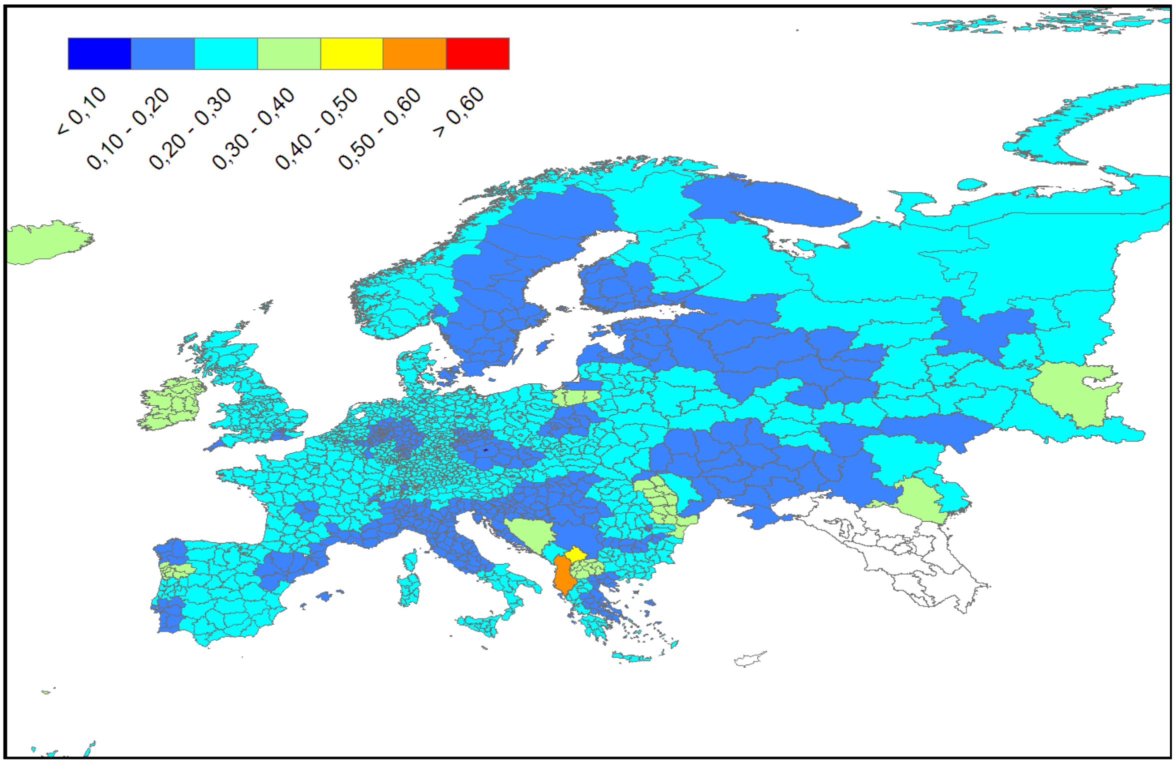
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.60 Total fertility, 1930



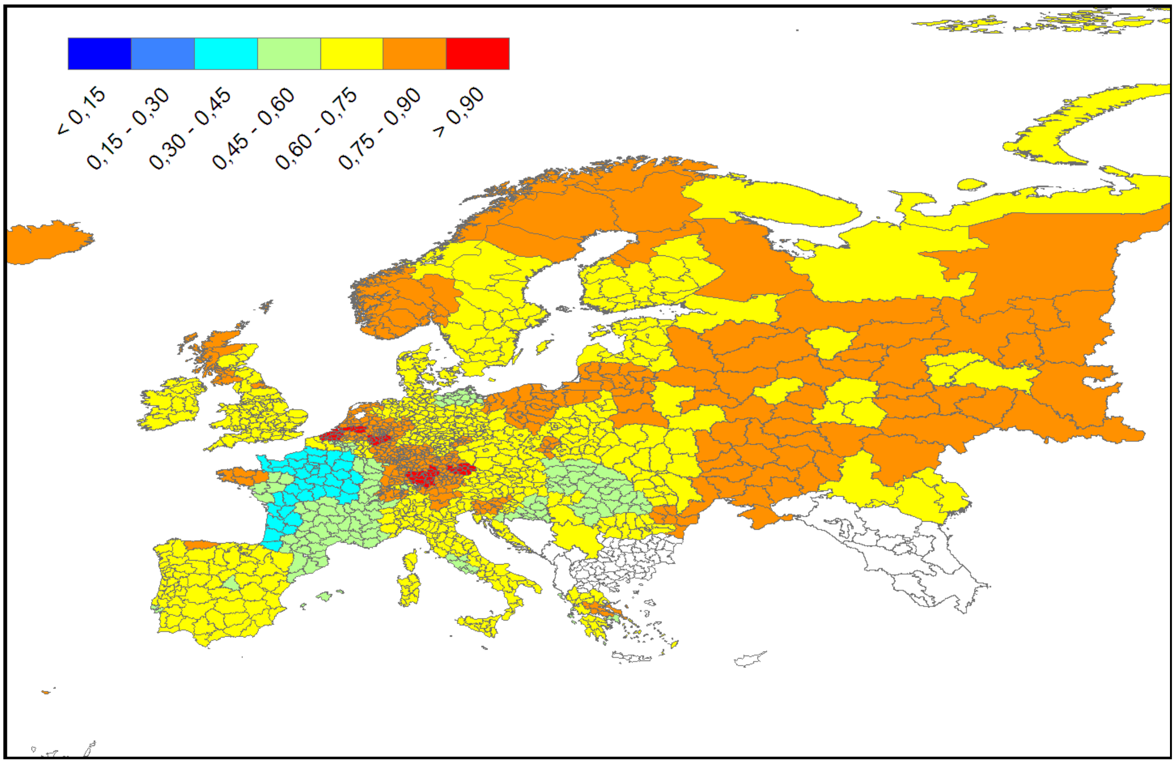
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.61 Total fertility, 1960



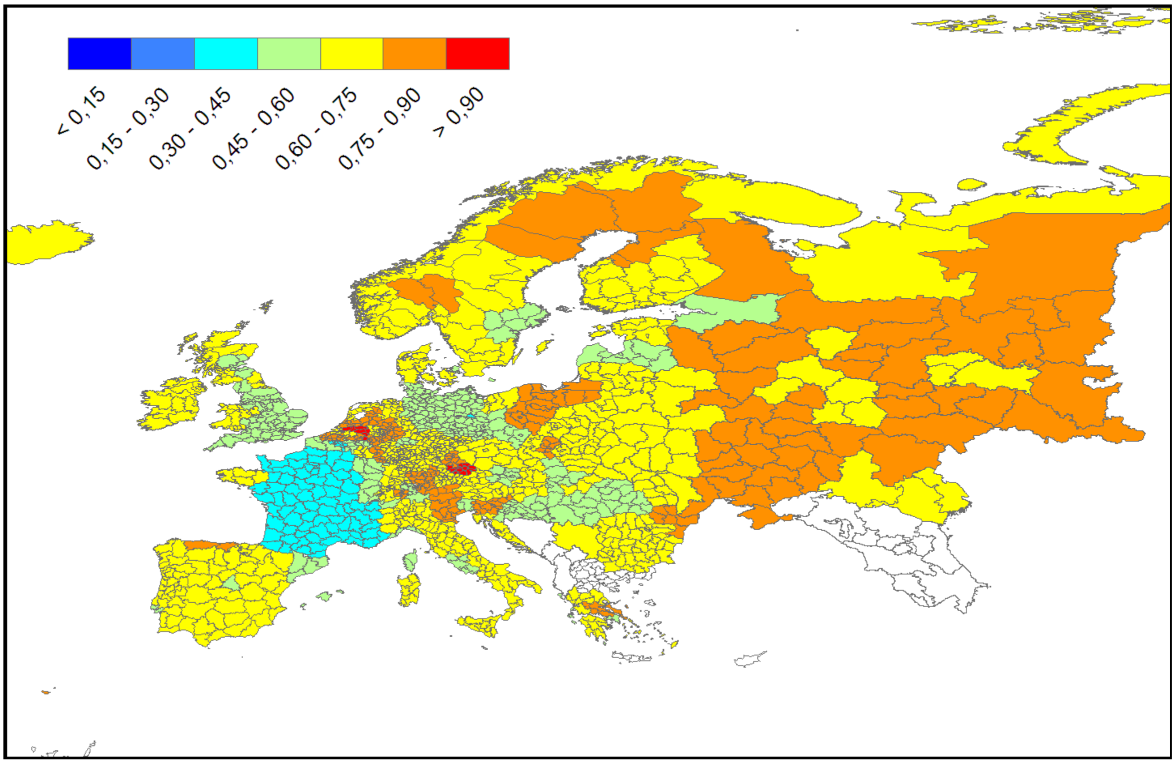
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.62 Marital fertility, 1870



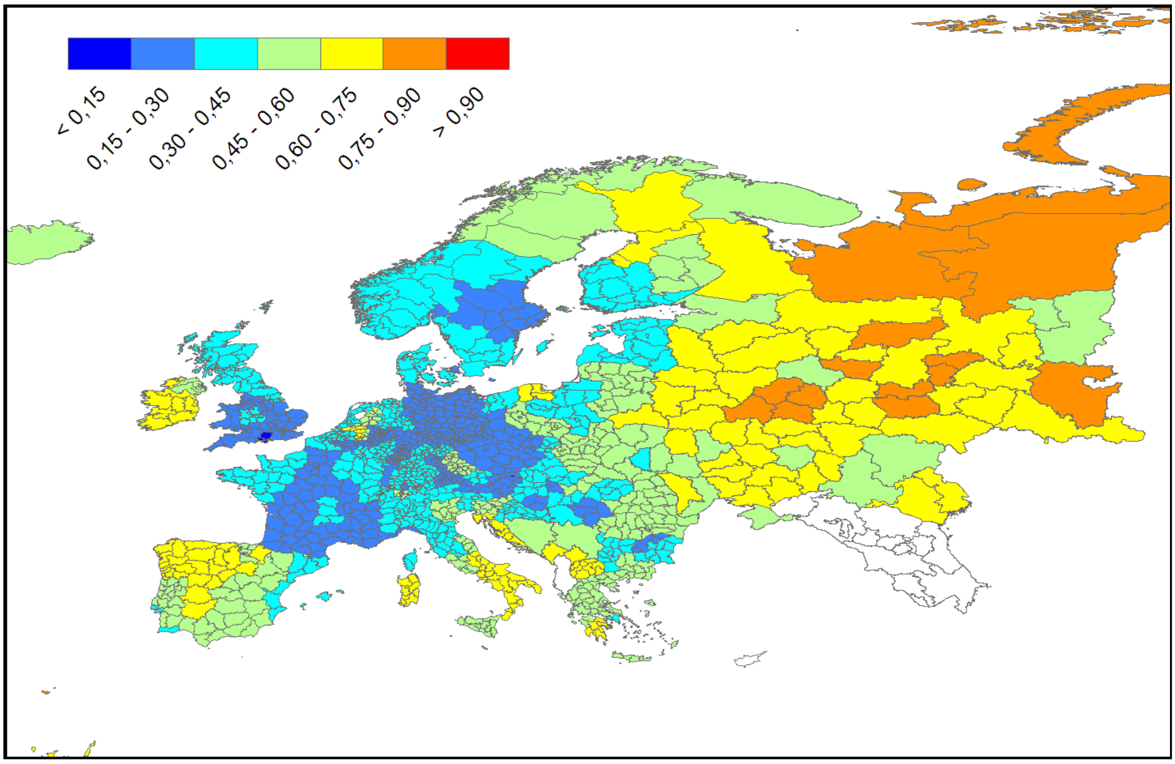
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.63 Marital fertility, 1900



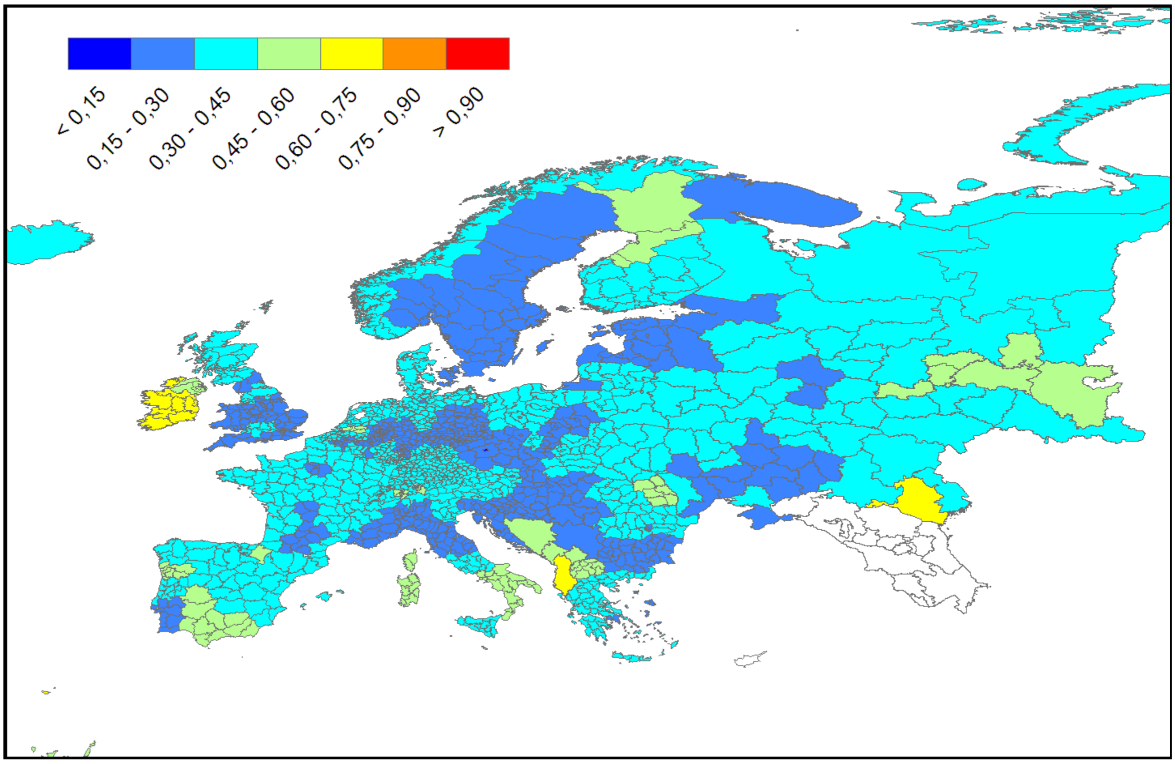
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.64 Marital fertility, 1930



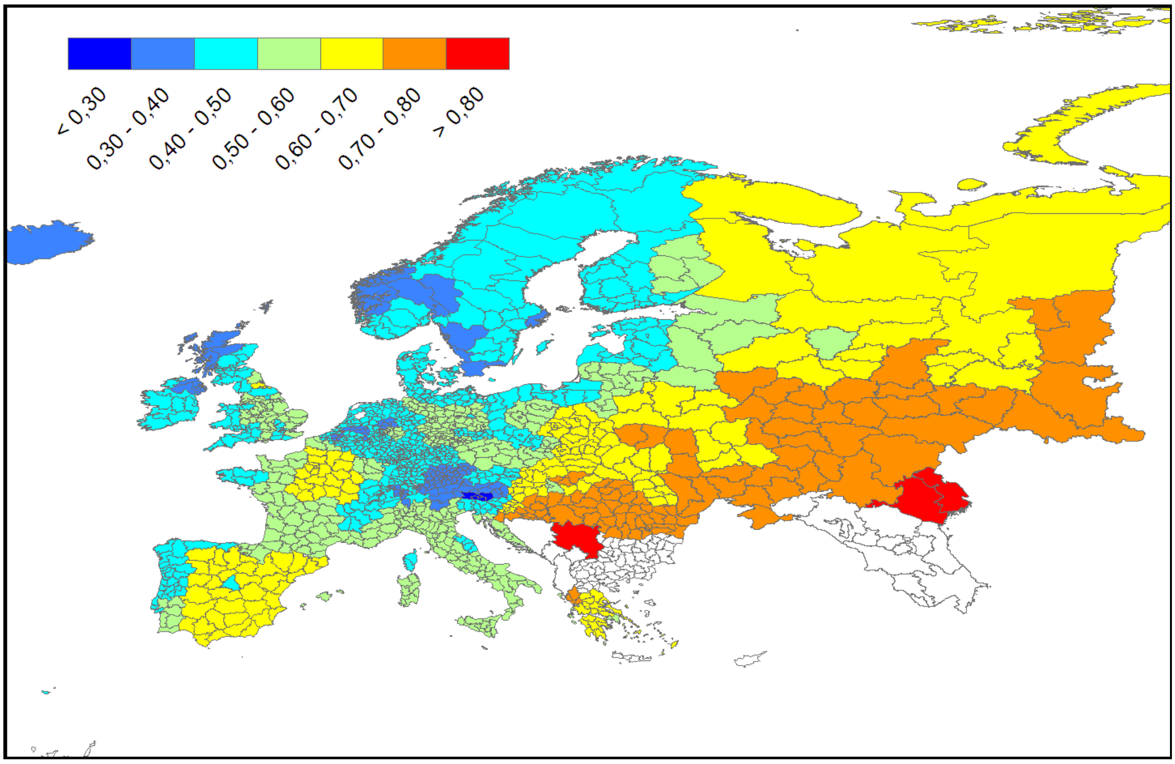
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.65 Marital fertility, 1960



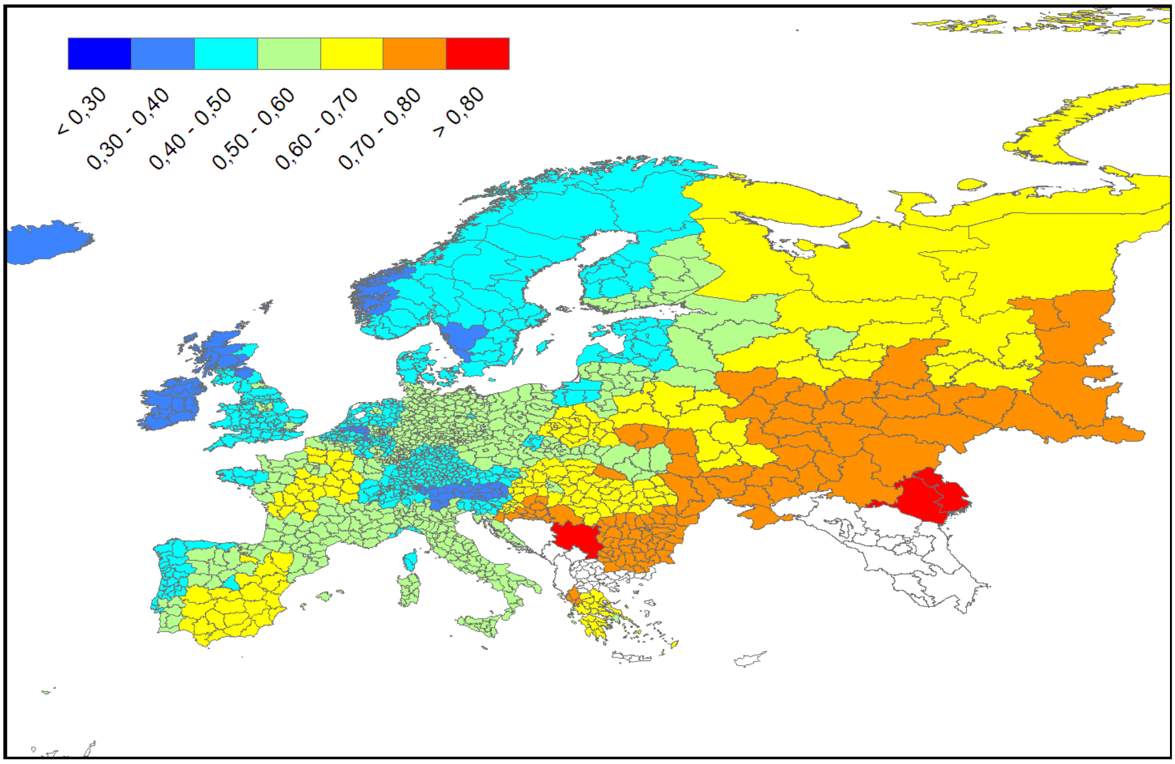
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.66 Marital status, 1870



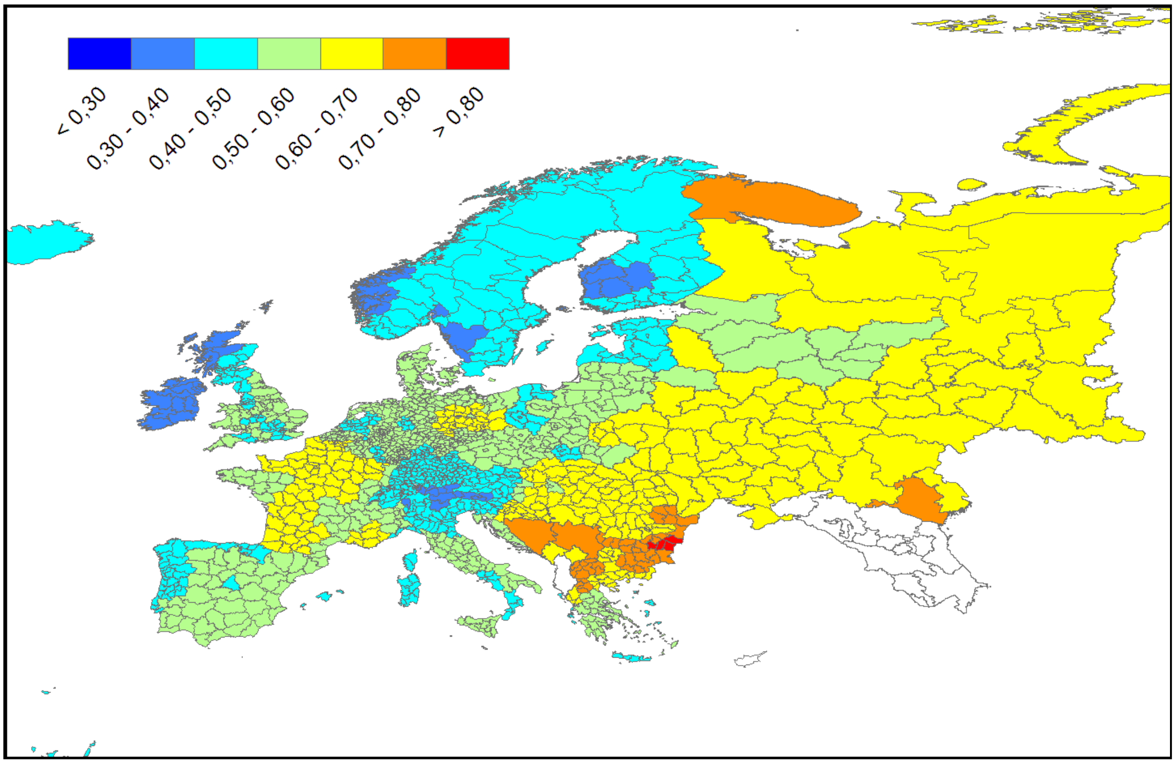
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.67 Marital status, 1900



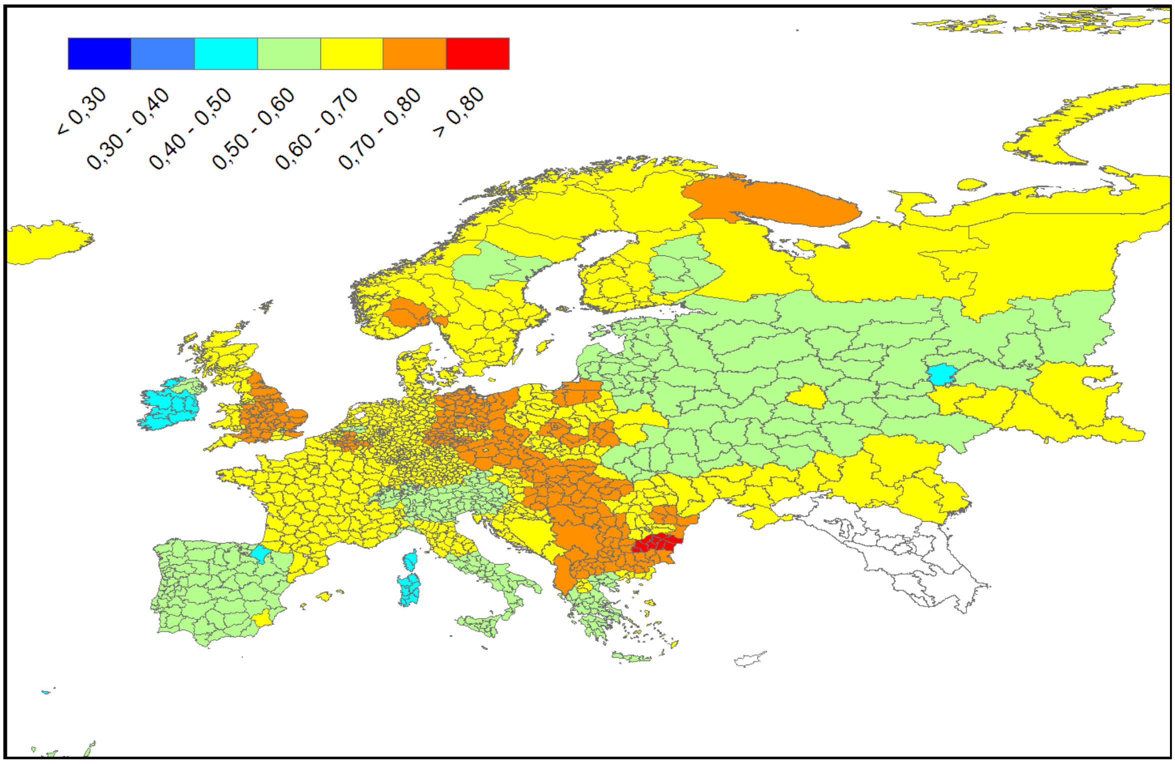
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.68 Marital status, 1930



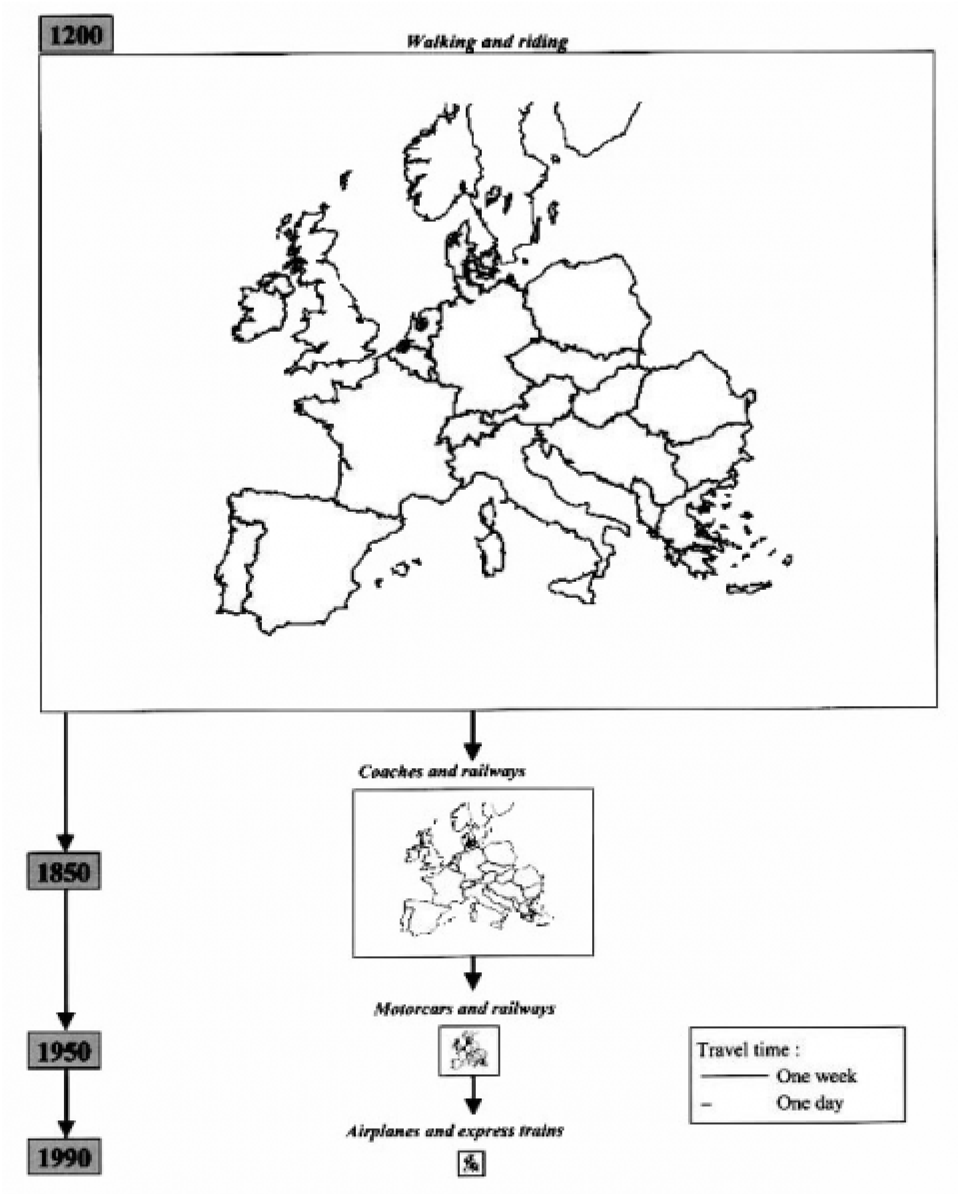
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.69 Marital status, 1960



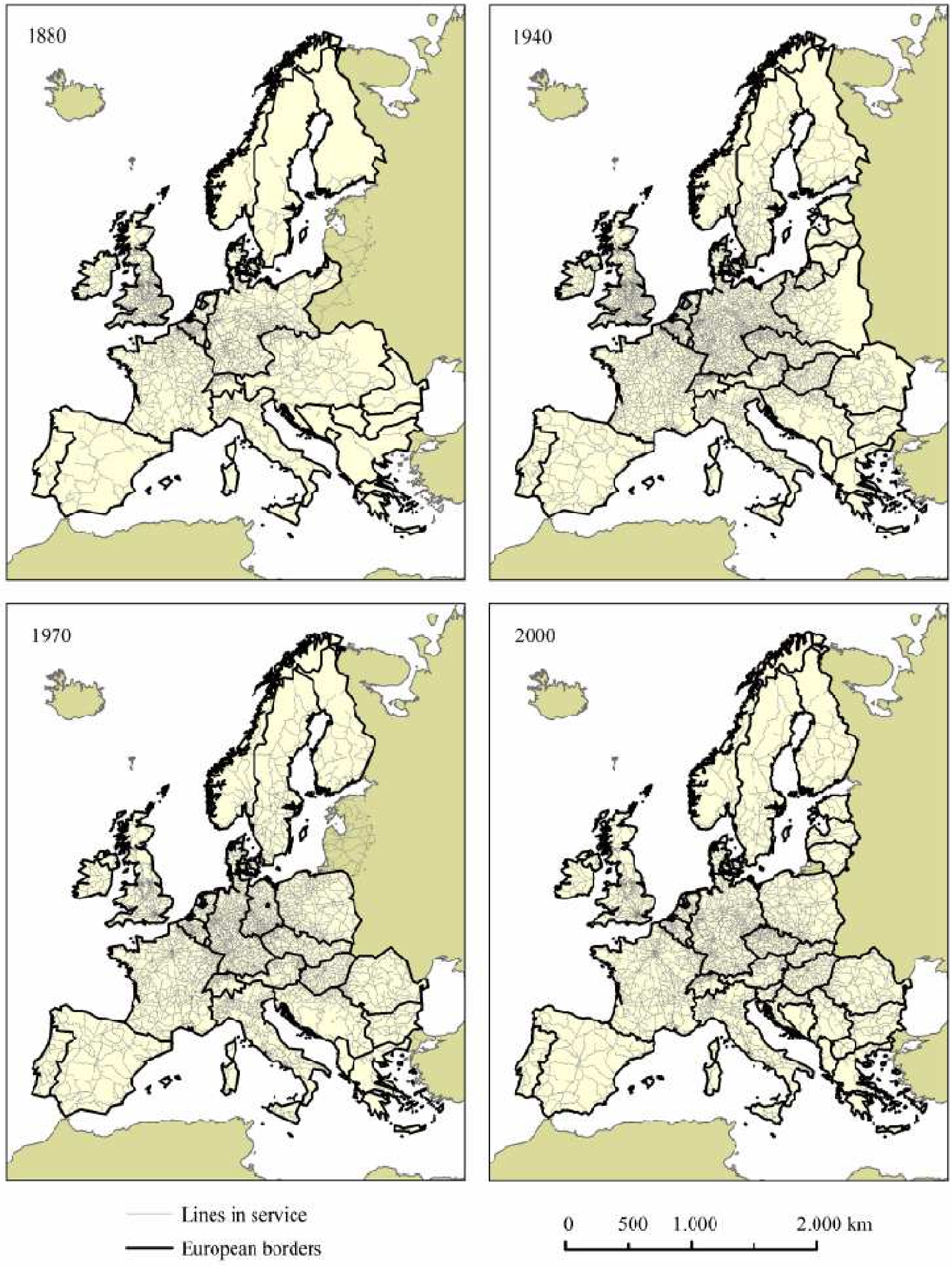
Source: Own calculations, data provided by Coale and Watkins (1986).

Figure 3.70 Transport revolution in Europe over time



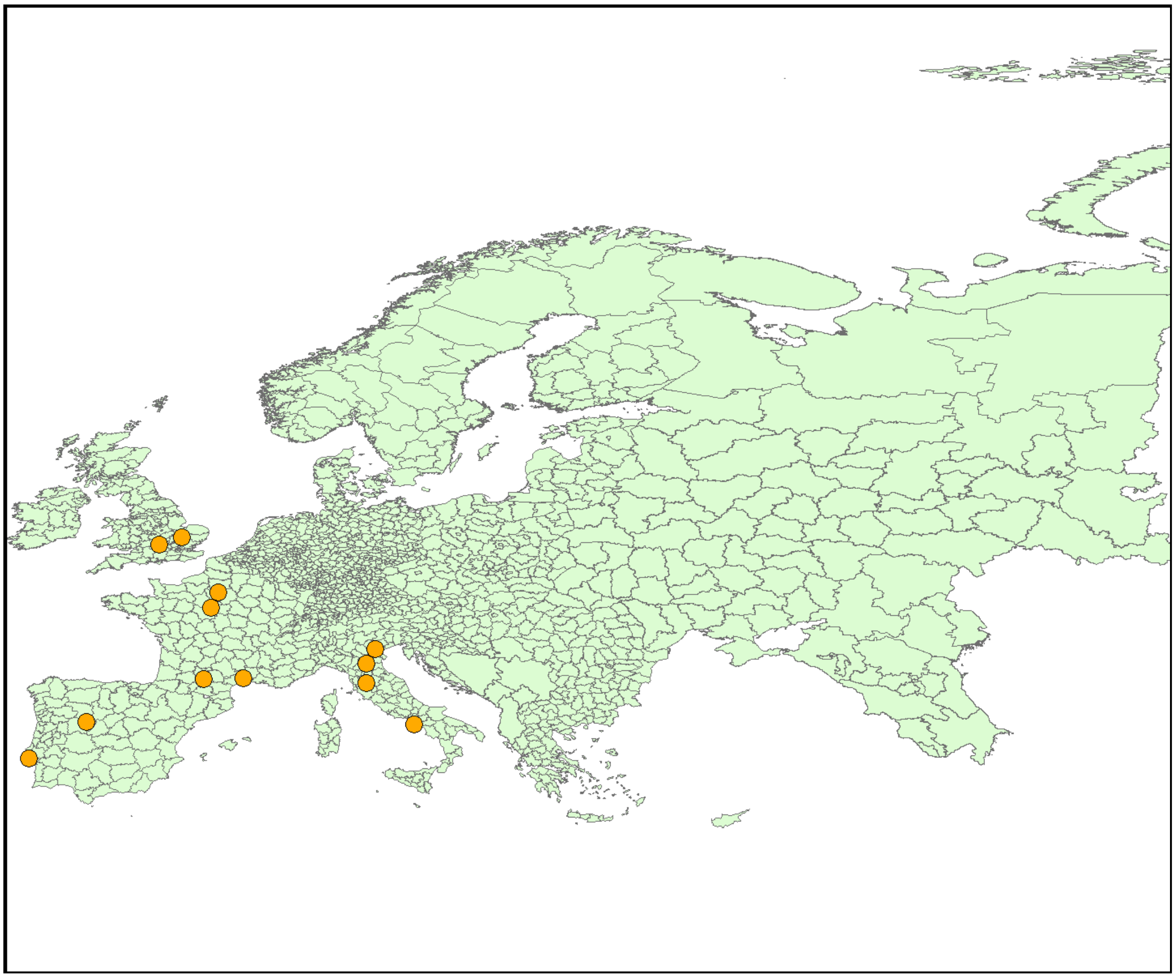
Source: Bretagnolle *et al.* (Cybergeo, 1997).

Figure 3.71 Evolution of railways in Europe, 1880-2000



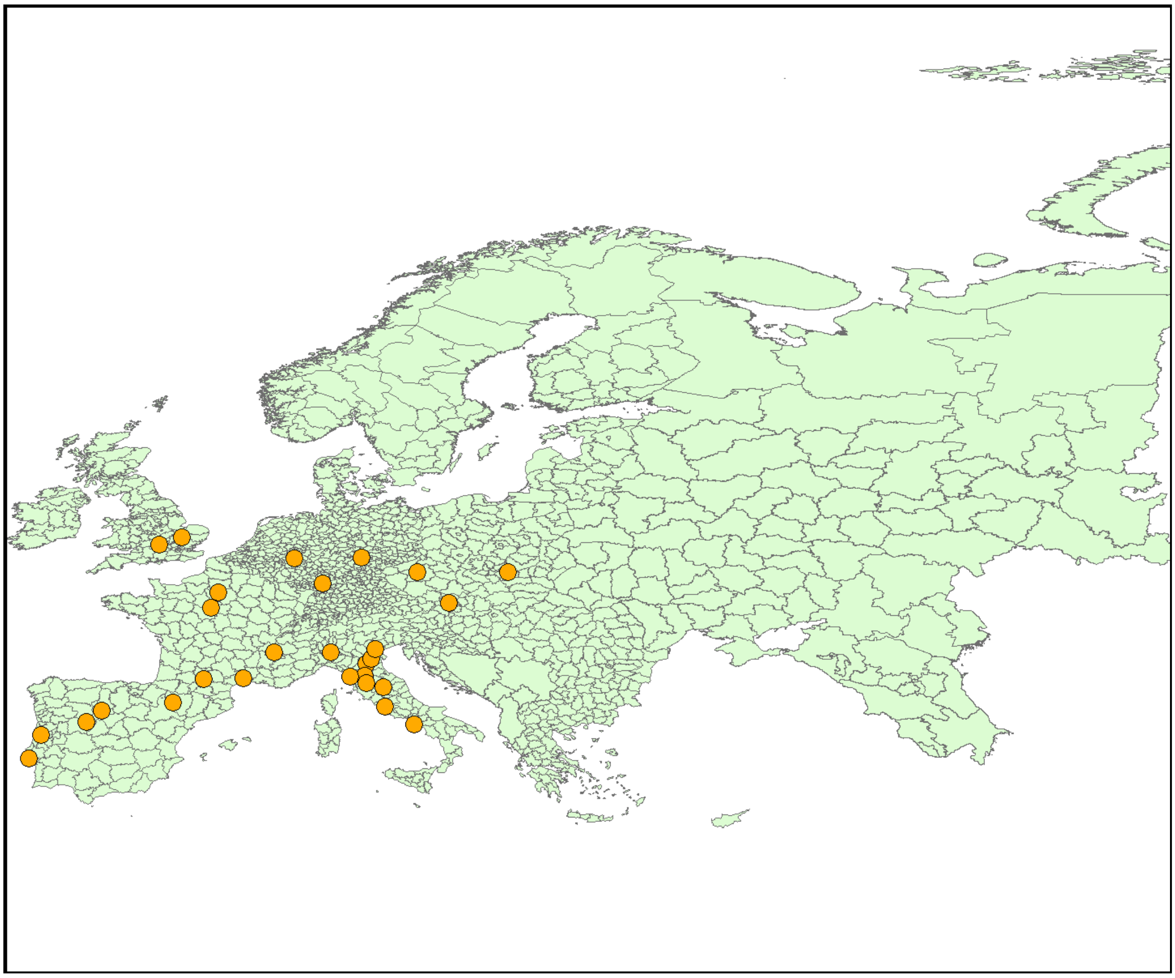
Source: Morillas-Torné (2012).

Figure 3.72 Location of enduring universities, 1300



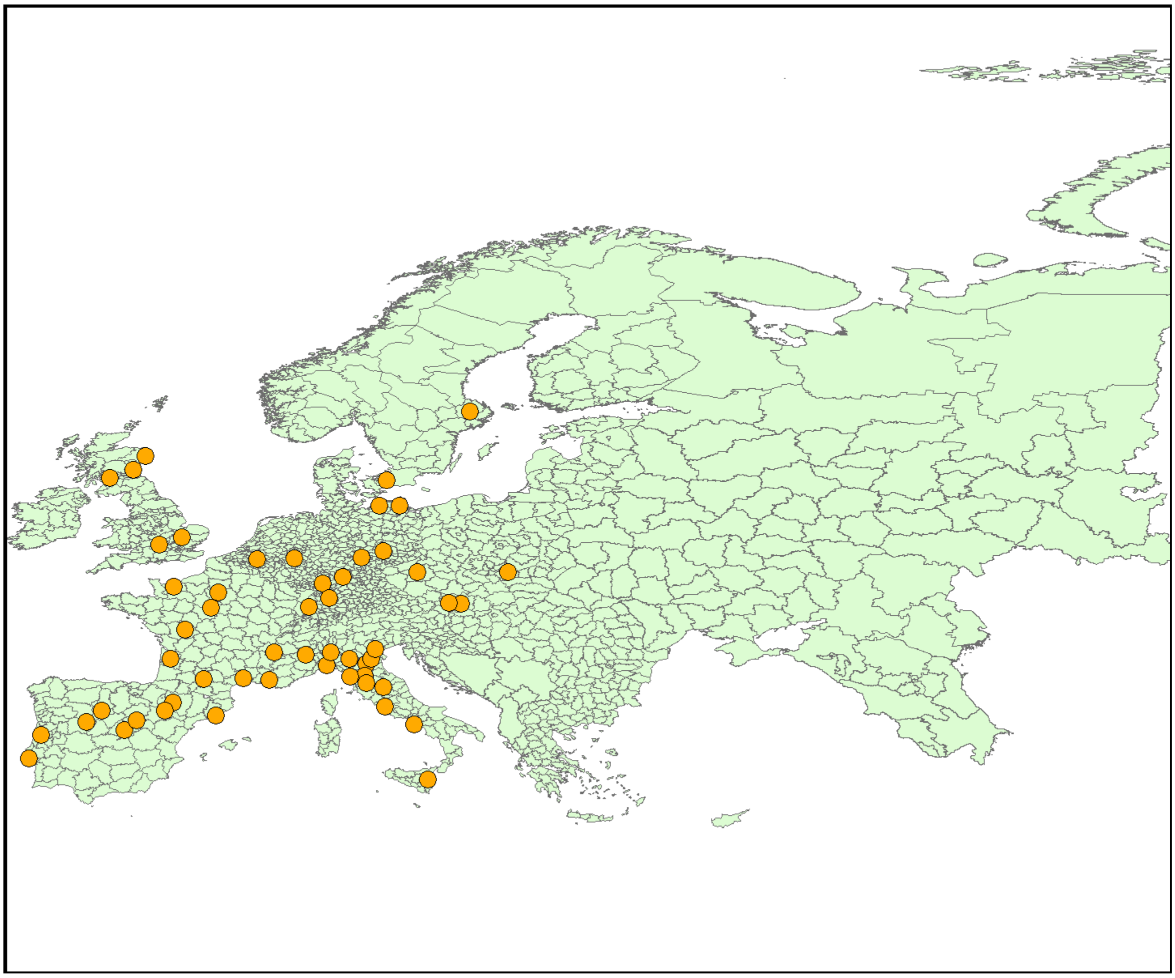
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.73 Location of enduring universities, 1400



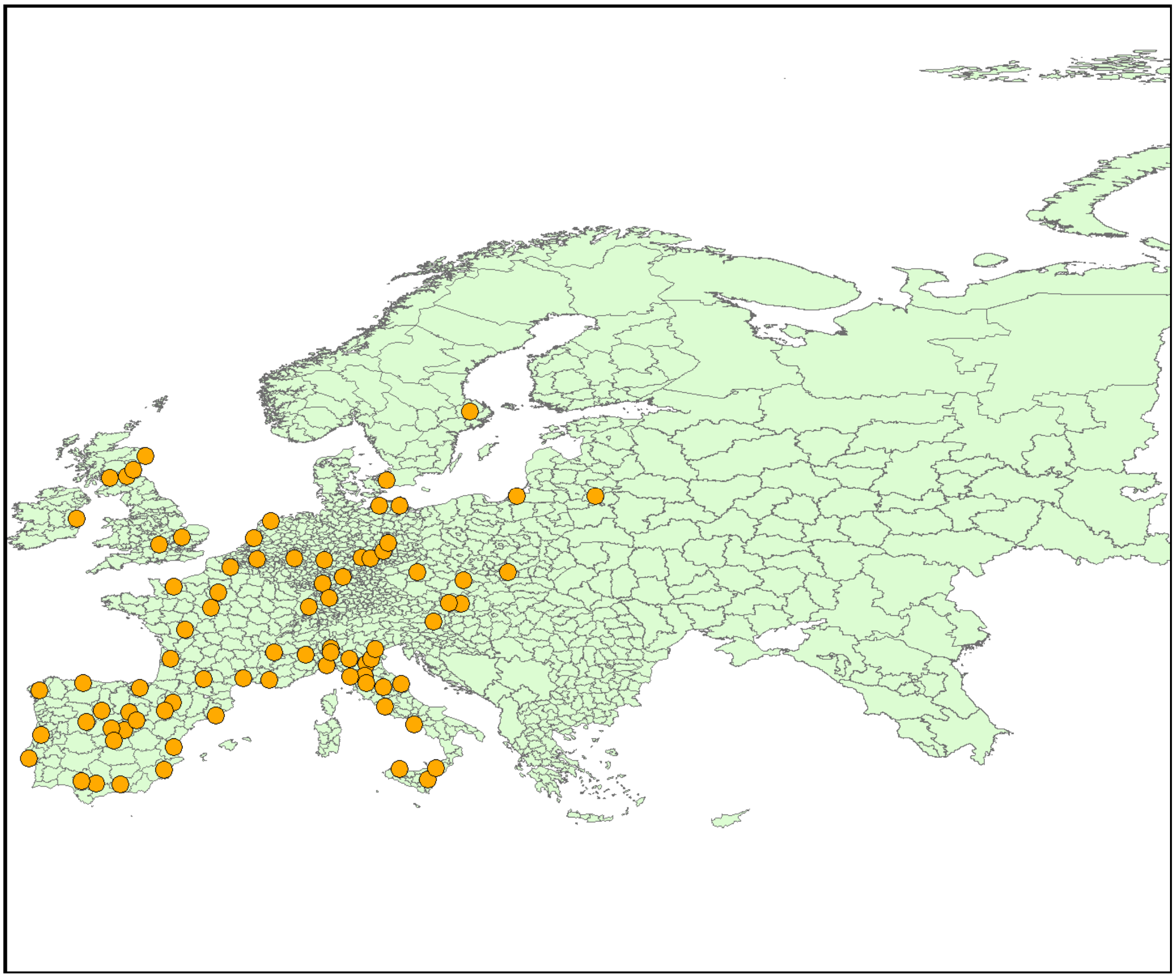
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.74 Location of enduring universities, 1500



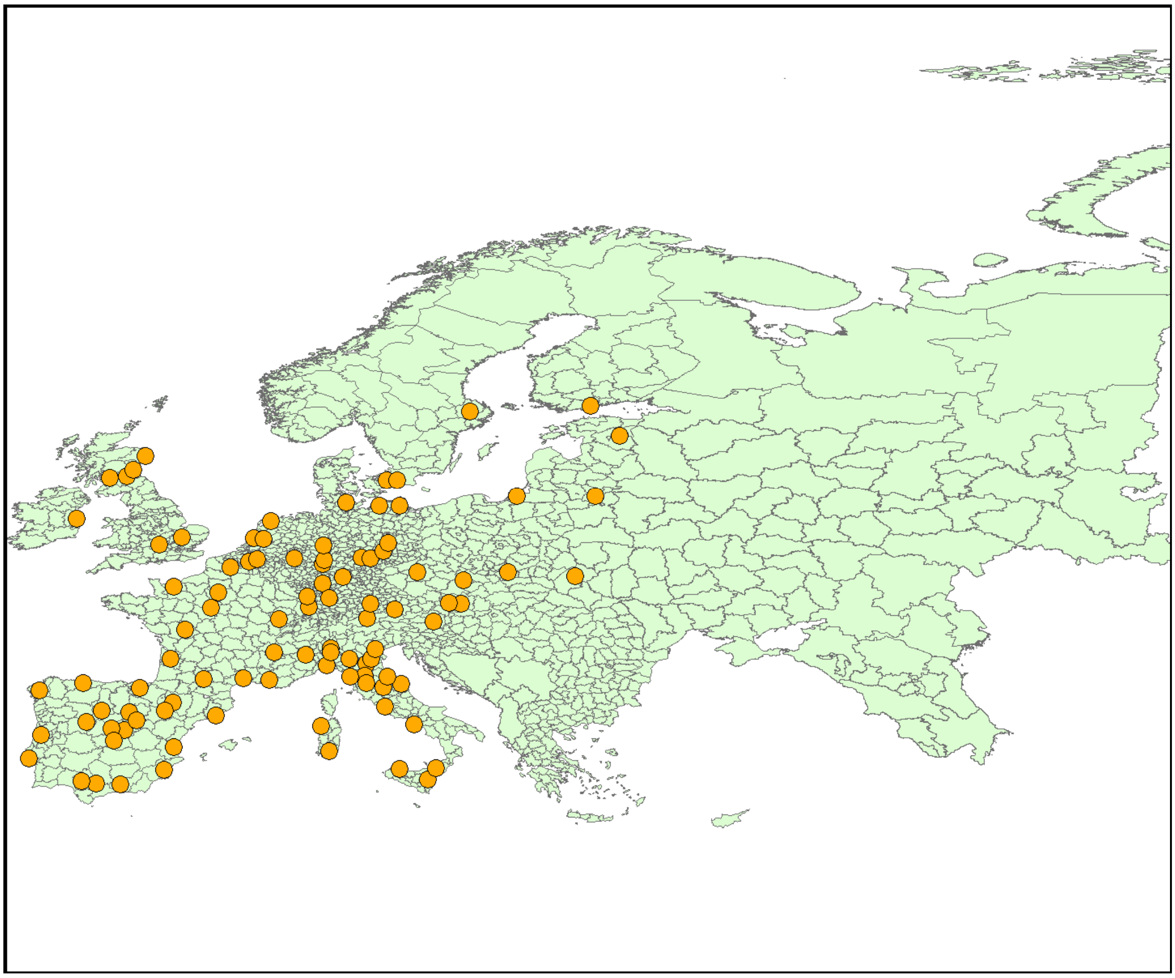
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.75 Location of enduring universities, 1600



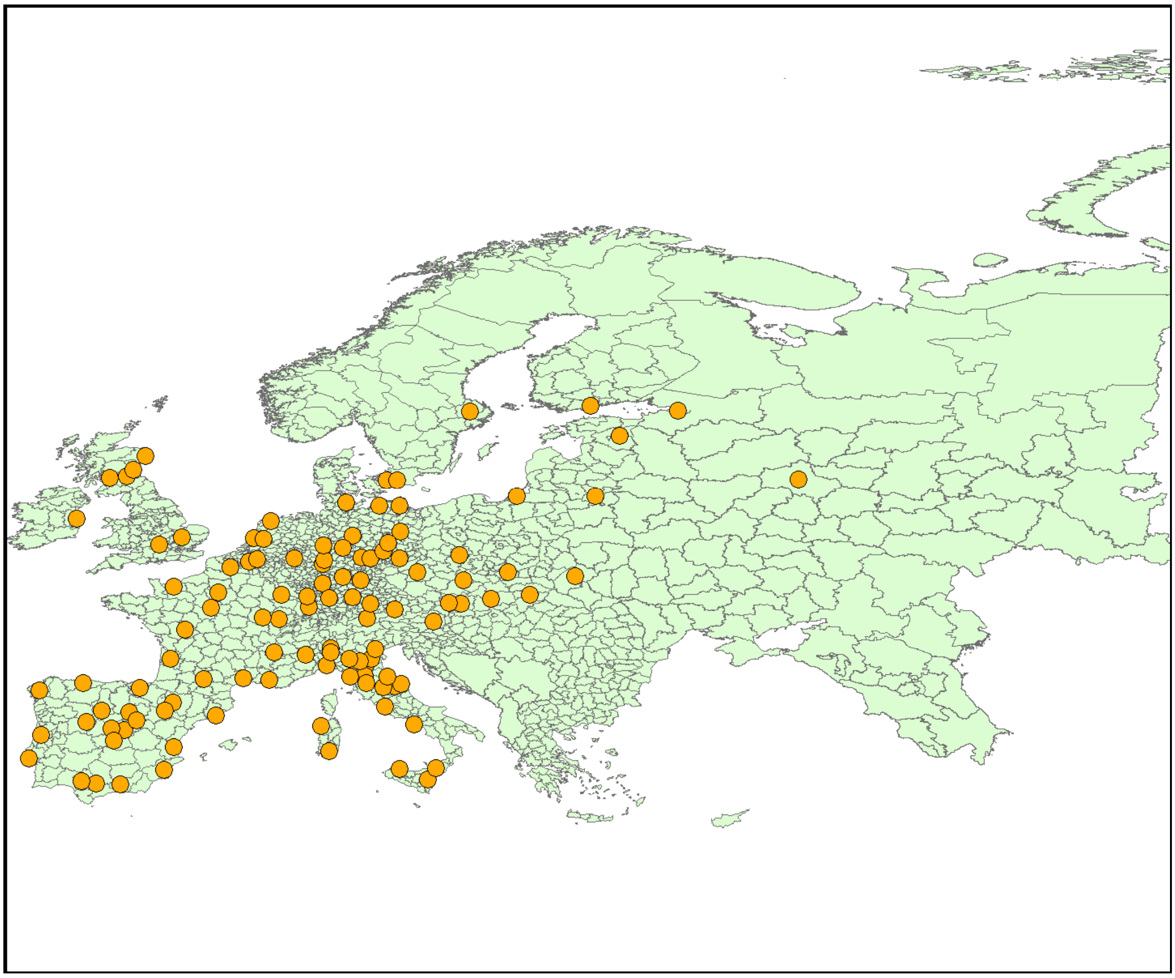
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.76 Location of enduring universities, 1700



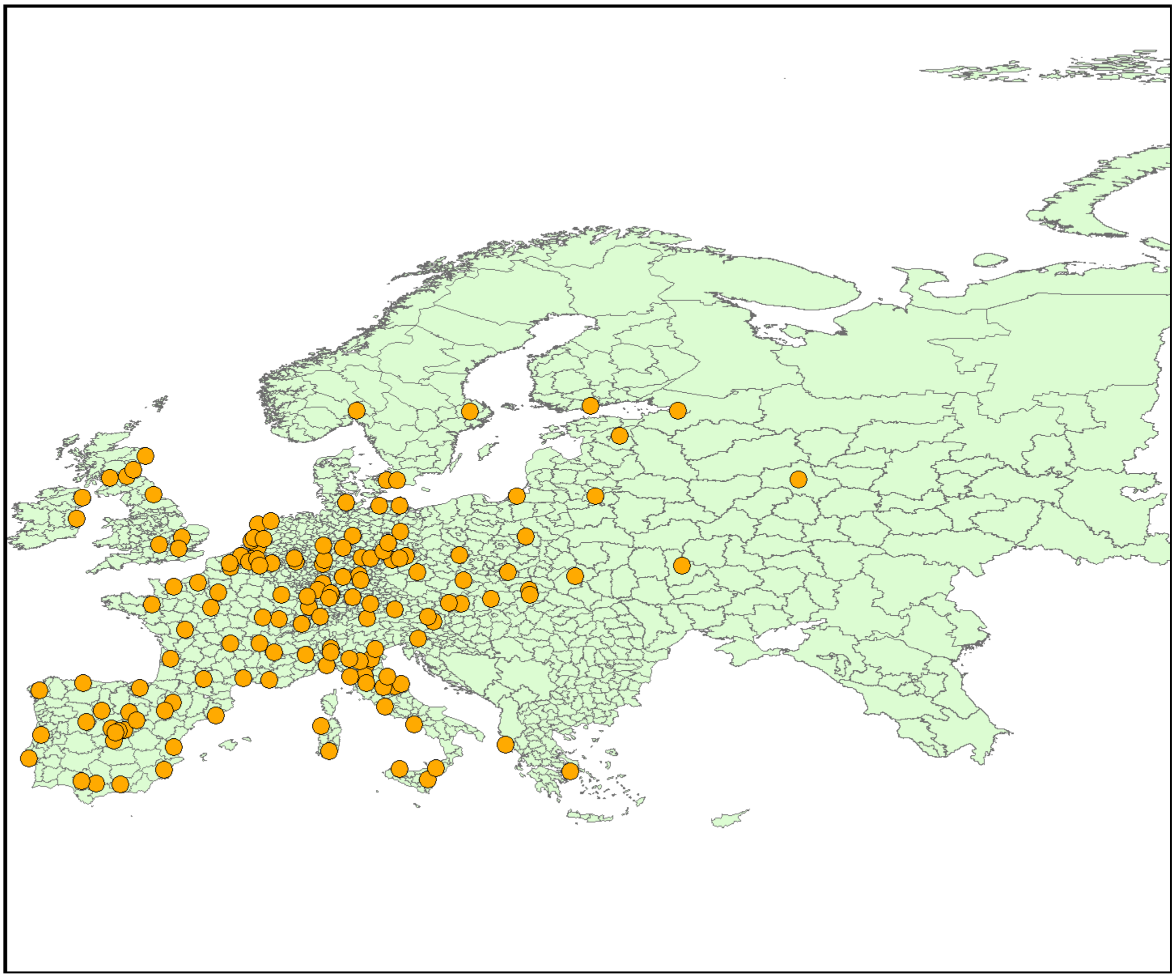
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.77 Location of enduring universities, 1800



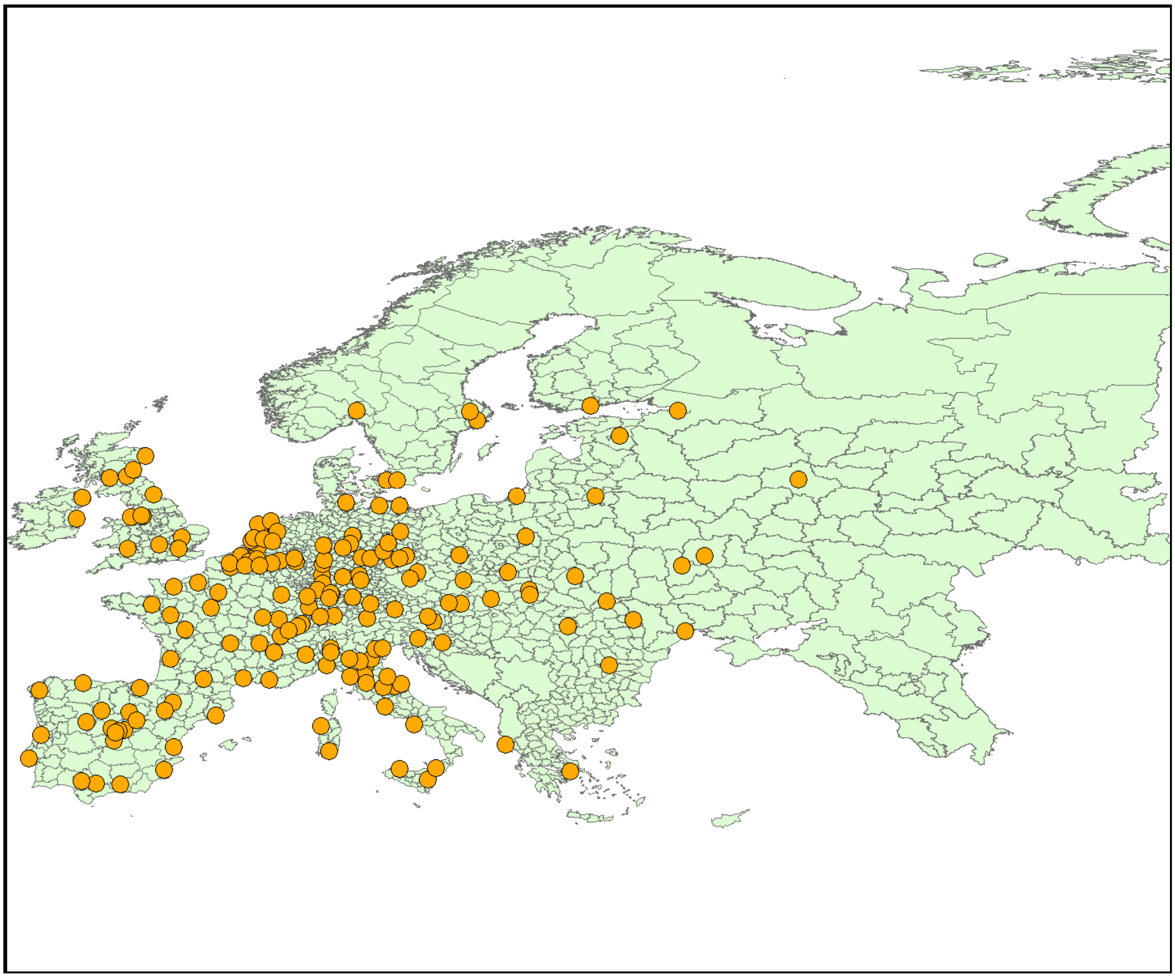
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.78 Location of enduring universities, 1850



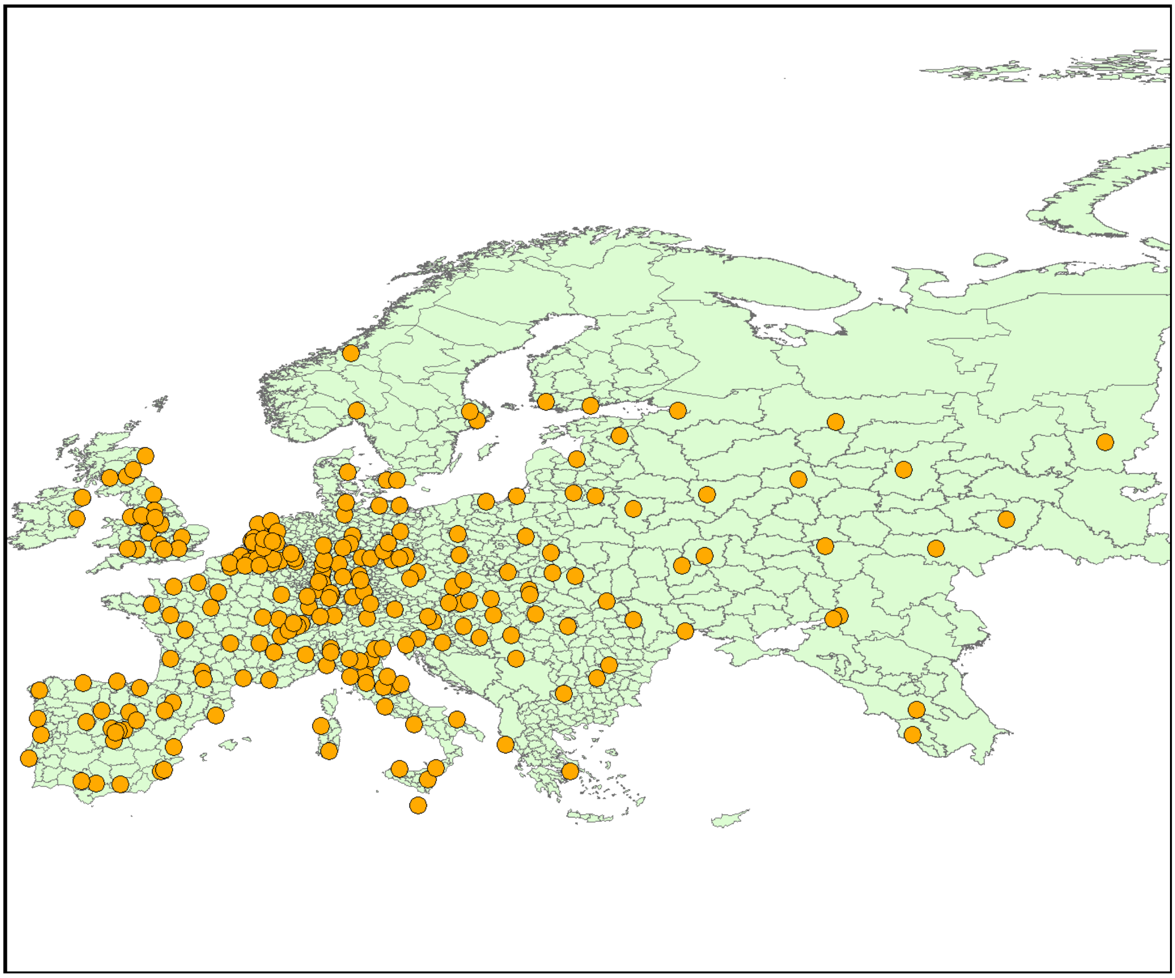
Source: Own calculations, data provided by Rüegg (2004).

Figure 3.79 Location of enduring universities, 1900



Source: Own calculations, data provided by Rüegg (2004).

Figure 3.80 Location of enduring universities, 1944



Source: Own calculations, data provided by Rüegg (2004).

4. How to measure human capital? The relationship between numeracy and literacy

Recent research has increasingly employed the age heaping method to approximate numeracy. But is this approach complementary to other more standard proxies of human capital? To answer this question, we compare numeracy with another major indicator of human capital: literacy. The results show that numeracy correlates well with literacy. This is underlined by descriptive historical evidence for regions in Europe and OLS regressions for a number of today's developing countries.

This chapter is based on a working paper which has been published in *Economies et Sociétés*.

4.1 Introduction

There has been a recent surge in research on human capital in general, as portrayed by Unified Growth Theory (e.g., Galor 2005a, Galor 2012), and on measuring human capital adequately today and in history. Concerning historical evidence, approaches focusing on literacy or signatures rates have been complemented by other proxies, such as book production or numeracy (e.g., Baten and van Zanden 2008, Crayen and Baten 2010a). In particular, numeracy has been approximated by the age heaping method in a range of recent studies. This method allows to obtain valuable information on basic levels of human capital which were characteristic for most historical societies. In this way, it is also possible to calculate earlier estimates on human capital than other methods (such as school enrolment rates or literacy). But to trace back the history of human capital in the long run, it appears crucial to make the link between human capital indicators to assess the long-term implications of human capital on the economy and the society. However, evidence on the relationship between numeracy and other human capital proxies has been restricted to a few studies mostly at the national level or at the regional level for one particular country (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010a).

For this reason, in this paper we use a recently constructed large data set on numeracy, covering most of the European regions in the 19th century, to advance the understanding of the relationship between numeracy and another major indicator of human capital: literacy. This dataset is complemented by additional regional data for developing countries outside Europe in the 20th century. By using these data we are able to compare numeracy and literacy data at the regional level in different parts of the world at different points in time. This enriches the existing literature in both space and time dimensions.

Our results indicate that there is a high correlation between numeracy and literacy indicators. First visual analyses by using scatter plots are verified by several OLS regressions of literacy and other possible determinants on numeracy. Literacy appears to be a robust determinant of numeracy in all specifications.

The paper is organised as follows: first, we review the literature on literacy and numeracy and the historical educational context. Then, we present the data and the methodology to compare the different proxies of human capital. In particular, the age heaping method is used to approximate numeracy and the ability to 'read and write' to measure literacy. The next section highlights the results. Finally, the last section concludes.

4.2 Literature

4.2.1 Literacy

There is a broad range of human capital proxies which are used today. Woessmann (2003) lists some of them. He names variables such as education-augmented labour input, adult literacy rates, level of educational attainment and average years of schooling. Still, not all of them can be used in an historical perspective because data are often not sufficiently available. Therefore, the most important historical proxies for human capital in Europe are literacy (taking the form of the ability to 'read and write' and signature rates), and, more recently, numeracy. Other indicators include book production, school enrolment ratios, the number of schools or the number of teachers.[17]

In the next sections, we review first some of the literature on literacy, before taking a closer look at the relationship between literacy and numeracy. Finally, numeracy itself and the broader context are highlighted, thus allowing a broader understanding of the issues at stake.

The capacity of writing has a very long tradition, although it was only open to the elites of society during most of history. Still in 1750, more than 90 % of the worldwide

[17] An example for the use of historical school enrolment ratios is the paper by Becker *et al.* (2012).

population was not able to write and did not have access to institutions teaching it (Cipolla 1969). Therefore, the broad majority of the population was excluded from literacy. However, literacy is very important because it is a “tool for enabling individuals and social groups to extend their understanding of themselves and their world” (Vincent 2000, p. 24). In this way, individuals are more receptive to new ideas and adapt themselves faster to the changing demands of their work and their environment. This is a crucial point because the advances in technology necessitated this flexibility in different areas of the job market throughout history.

In general, literacy is a very popular measurement method of human capital. However, interest in its historical development had languished for a long time. Pioneering works by Cipolla (1969) and Stone (1969) initiated this research field. Since then, research has been conducted much more widely on historical literacy (Graff 2009). As Graff (1991) points out, one can classify research on historical literacy over the last decades in three broad generations. The first generation was constituted by research during the end of the 1960s, in particular by Schofield (1968), Cipolla (1969) and Stone (1969). Nevertheless, already some works in the 1950s were predecessors of this research line (Webb 1955, Fleury and Valmary 1957). The first generation created a foundation for upcoming studies by stressing and demonstrating the importance of literacy. Furthermore, it indicated future research possibilities with respect to more extensive numerical sources and broader research themes. Based on these results, a second generation began its work in exploiting even more detailed quantitative data and distinguishing historical patterns of literacy. Finally, the third generation has particularly been opting for more interdisciplinary research between different fields to advance the knowledge on literacy. In this context, the study of literacy combined with the one of numeracy appears to be a logical and valuable addition to the existing literature.

Considering data availability, literacy rates in the form of reading (and writing) ability are available for most European countries only from the middle of the 19th century onwards. Accordingly, Cipolla states that "for the periods preceding the second half of the nineteenth century the information [on illiteracy] is very poor" (Cipolla 1969, p. 15). Therefore, many studies on early literacy developments use signature rates of conscripts or newly married couples (e.g., Schofield 1981, Mitch 1993, Reis 2005). In the case of the latter, marriage contracts or other official documents had to be signed by the eligible person. However, this person was not always able to sign the contract. For this reason, taking the share of people who were able to sign with respect to the entire population might be employed as an indicator of literacy.

The idea of taking signature rates of marriage contracts is nothing new. For instance, this proxy was already employed in *Statistique générale de la France* from 1854 onwards (Furet and Ozouf 1977). Interest in reconstructing educational levels is nothing new either. In 1877 Louis Maggiolo, ancient rector of Nancy Academy, began his work in recollecting marriage signatures from all over France for the years 1686 to 1690, 1786 to 1790, 1812 to 1816 and 1872 to 1876.

Still, there are potential biases and disadvantages of this method. For example, did the individual sign the corresponding document or was this done by another person, such as the bride or the priest? In France, newly married couples were obliged by law to sign their marriage contract since 1647, before it was often the priest who signed them. Yet this example may not be confounded with other countries less stringent on marriage signatures. For instance, it was not obligatory for a couple to sign the registers in Italy (Cipolla 1969). Moreover, even when laws were passed, their application was often a different matter, with biasing effects on the data. The degree of spatial and time coverage is thus very different from country to country in Europe.

More globally, a range of methodological and conceptual issues have to be addressed when measuring literacy. What is illiteracy? There are several possible answers which make it not always easy to define literacy in a unique way. One can define a literate person by his capacity to read and write and an illiterate one by the lack of this capacity. However, there is still a group of 'semi-illiterates' (Cipolla 1969) who can read but not write. In some censuses of the 19th century data were explicitly collected for those semi-illiterates, in others they were not. Another issue that arises is the question of the quality of reading and writing. More specifically, people may be able to read but not necessarily capable of understanding the content. On the other hand, a person who is able to sign may not be able to write anything else than the name he has been trained to write in some way in order to fulfil certain minimum requirements for contracts. Finally, literacy is measured using age categories which may be different from one country to another and may vary over time. For this reason, literacy data is available in some countries for the population e.g. above 7 years and in others the threshold is 15 years onwards, etc. This may possibly make it difficult to compare these data directly.

Furthermore, Woessmann (2003) explicitly addresses the lack of further skills in this proxy: "Any educational investment which occurs on top of the acquisition of basic literacy – e.g., the acquisition of numeracy, of logical and analytical reasoning, and of scientific and technical knowledge – is neglected in this measure" (Woessmann 2003, p. 243). Thus, it appears necessary to conduct further research on numeracy in general and on the link between literacy and numeracy in particular.

4.2.2 Literacy and numeracy

However, compared to the literature on literacy, research on numeracy is still in its infancy (see e.g., Mokyr 1983, Thomas 1987, Emigh 2002, Netz 2002). Why is there such a gap between these two research fields? A major problem for researchers has long been the quantification of numeracy (A'Hearn *et al.* 2009). A coherent measure was lacking because few statistics were collected on numeracy in the past (Thomas 1987, Vincent 2000). Some even believed that it was impossible to construct one (Cohen 1982). As we will see later on, this has been overcome by the age heaping strategy (A'Hearn *et al.* 2009).

Nevertheless, research on both indicators appears to be helpful to understand the development of human capital due to several reasons. First, their link is already identifiable in the terms themselves. In fact, an alternative term for numeracy is 'quantitative literacy' (Steen 1997, Chiswick *et al.* 2003, A'Hearn *et al.* 2009), suggesting that the concept of literacy is applied to quantitative capacities. Anecdotal evidence further suggests that there is a relationship between the two proxies. For example, young children who are good in literacy (here: reading) are often also high performing in numeracy (arithmetic) (Bulcock and Beebe 1981).

Second, literacy and numeracy have been closely intertwined throughout history. Accordingly, Netz (2002, p. 323) points out that "there is no difference between the history of numeracy and the history of literacy". In ancient cultures, the use of numerical symbols paved the way for verbal symbols. For instance, by analysing the emergence of writing in Mesopotamia, Schmandt-Besserat (1992) comes to the conclusion that "in early cultures, numeracy drives literacy rather than the other way around" (Netz 2002, p. 323).

An example for the linkage between literacy and numeracy are also the 'Arabic numerals' widely used around the world today. In fact, these Arabic numerals should rather be called 'Indian' numerals since they were invented in India. Their first recorded images on pillars date from around 250 BCE (Woods and Woods 2000). Subsequently, they were passed on to other neighbouring cultures which traded with India. In this way, people from the Middle East adopted and eventually adapted this numerical system. It took quite some time for its breakthrough in Europe, though. Only beginning from 976 CE onwards it became known in Europe because Europeans traded with Middle Easterners (Woods and Woods 2000). As Europeans did not know the origins of the numerals, they named them 'Arabic numerals'. Later on, the invention of the printing press led to a process of standardisation of the numerals which increased their acceptance among Europeans. Finally, Europeans brought this system to other parts of the world by means of their trades and conquests.

In consequence, Arabic numerals eventually replaced the old Roman system in Europe. They did so because they have several advantages. First, Roman numerals do not allow calculation as easy as Arabic ones and large sums cannot be expressed in a short manner. Second, there is no means to obtain fractions and a symbol representing zero does not exist. Third, it is possible to record transactions, in contrast to calculations which had to be done with counters such as an abacus (Thomas 1987). This, however, is a key feature of Arabic numerals. It combines numerical and verbal practices so that arithmetic can be put down by the use of paper and pen, and not by counters. In this way, both practices are linked to each other and their histories are intertwined.

4.2.3 Numeracy

Numeracy (and particular arithmetic) may be influenced by different additional factors, most importantly education, state bureaucracy and capitalism (Emigh 2002). Population statistics such as censuses and tax assessments carried out by the state may oblige individuals to keep records and correspond to the described requirements. Nevertheless, Emigh (2002) argues that the chain of causation runs the other way round because states may have also reacted to a rise in numeracy by collecting more thorough data. Otherwise the role and the power of the state on the population would be overemphasised (Tilly 1999, A'Hearn *et al.* 2009).

Moreover, market capitalism furthered numeracy due to the necessity to keep records and be able to calculate. Accordingly, arithmetic was also perceived as being at the core of trade (Hodder 1671, Thomas 1987). Tradesmen, state bureaucrats and plenty of other occupations needed arithmetic (Thomas 1987). Nevertheless, numeracy was also important in local economies without market capitalism (Emigh 2002). Therefore, numeracy already played an important role in everyday life before the Industrial Revolution. As the Catasto of 1427 in Florence highlights, the ability to work with numbers was needed for transactions such as property sales, payment for whatever service, testaments or dowries (A'Hearn *et al.* 2009).

By contrast, calculating was not accessible to and not perceived to be needed by everyone. Taking the example of early modern England, most grammar schools did not teach arithmetic before 1660 and afterwards only in form of an extra. As a result, still during the last decades of the 17th century, "probably fewer than four hundred men could be said to be mathematically minded" (Cohen 1982, p. 39). The focus lay on literary classes at school, also because mathematics was still disregarded in society. The attitude by the public was rather negative, mathematics was perceived as an anti-social object (Thomas 1987). Moreover, arithmetic was considered to be only important for certain occupations and not for the general training and education of everyone. These occupations were to be found in commerce and trade. In contrast, higher society despised such occupations to be beneath its status. On the other hand, the bottom of the society had no access to educational facilities. Only at the end of the 17th century did this slowly change and arithmetic was included in the curricula of more grammar schools. But this was not always the case. For example, arithmetic became obligatory at Eton only in 1851 (Houston 2002).

The importance for improved navigation and rising overseas trade helped to spur this formation of arithmetic skills alongside commercial developments. In addition, reformers favoured the inclusion of arithmetic because "it disciplined the mind, encouraged inductive thought, and developed habits of precision, attention to detail, and a love of factual knowledge" (Houston 2002, p. 164). Finally, as Thomas (1987) resumes the research on this point, this development led to the acknowledgement of mathematics being a fundamental part of the education of a gentleman in England in the 18th century. However, the change in attitude towards a gentleman's education was not equivalent to the one of a lady. Mathematics was not deemed to be appropriate for women. Accordingly, it was mostly not taught to girls. Not surprisingly, women are mostly less numerate than men in many historical numeracy studies in Europe and elsewhere (e.g., Manzel and Baten 2009).

4.2.4 Some economic and social implications

The evolution and relationship of literacy and numeracy has also to be seen in a broader social and economic environment. Because this is a too large domain to be treated in detail within the framework of this paper, it appears more appropriate to focus on a few important issues in this context.[18]

First, Baten and van Zanden (2008) show that the accumulation of human capital had an important positive influence on economic growth before the Industrial Revolution. In contrast, Mitch (1993) and Mokyr (2002) argue that during the first phase of the Industrial Revolution in England, the basic school system was not substantially contributing to economic growth. Economic growth is rather accelerated despite the education system to that time. Only during the further advancement of the economy

[18] For a detailed discussion of the literature on the link between education and economic growth, see Demeulemeester and Diebolt (2011).

became education increasingly important because some basic knowledge in different disciplines was required for more and more occupations. This need for skilled personnel after mid-19th century finally evoked a surge for educational facilities and education in general. Education became much more widely available than before.

Social developments encouraged this spread of education. For example, illiteracy became to be deemed to be a national disgrace (Cipolla 1969). Moreover, education was perceived as a requirement and a manifestation of the state's authority (Green 1990, Vincent 2000). More schools were set up and thus school attendance rose. These facts illustrate the new importance given by the state to the education of the people. The state became the principal promoter of schooling and replaced in some ways the church in enforcing education, although the latter still played an important role. In fact, no institution other than the church had the necessary facilities, the means and the experience of managing the bureaucratic requirements in order to provide large-scale teaching. Therefore, as Vincent (2000, p. 7) puts it, "the capacity of […] structures of power [other than the state], the family, private philanthropy, the market place and the church, to realize this objective was called into question". In consequence, the church joined more vigorously the efforts of the state later on.

Nevertheless, even if the church and the state endeavoured to improve the quantity and the quality of education, parents still had a preeminent influence on the education of their children. Until the final compulsion by the state (and afterwards), there were at the beginning very convincing reasons not to let one's children go to school. In addition to fees, clothing and stationary, parents had to do without the contributions that the work of the children made to the income of the family. Moreover, the training given by schools was not always regarded as important for the future work of their children either. For this reason, authorities aimed at meeting the preferences and demands of parents in order to increase the effective enforcement of schooling laws. For example, they adapted the school calendar more to the needs of the parents. Still, school attendance was not granted itself by a law (Vincent 2000).

Therefore, "across Europe as a whole, formal schooling cannot be taken as a proxy for the distribution of reading and writing skills" (Vincent 2000, p. 57; Houston 1985). For this reason, direct literacy values as stated by individuals during censuses appear to be better suited for this purpose. This underlines once more the meaningfulness to use literacy data (to be able to 'read and write') in the present study.

4.3 Data

We analyse numeracy and literacy by using historical census data from Europe in the 19th century and more recent data from other continents during the middle to the end of the 20th century. Almost all data stem from official census publications.[19] First, historical numeracy estimates are derived from the database collected by Hippe and Baten (2012a). This database covers almost all European countries at the regional level in the 19th century. Literacy data have been added to this source where available, mostly taken from the same official publications. Thus, the European countries included in our dataset are Greece, Hungary, Ireland, Italy, Russia, Serbia and Spain.

Second, we use the Integrated Public Use Microdata Series (IPUMS) database which includes population microdata for many countries in the world from the 1960s onwards (Minnesota Population Center 2011). Clearly, age heaping and illiteracy have still to be sufficiently present in these censuses to compare both human capital indicators. Crayen and Baten (2010a) and Hippe and Baten (2012a) have already demonstrated that age heaping was already low in many of today's industrialised countries at the national and

[19] All data except the data from India. These are derived from an employment survey (see Table 2).

at the regional level at the end of the 19th century. For this reason, we use available IPUMS data from developing countries in the Americas, Africa and Asia for the second half of the 20th century. As before, because literacy and numeracy are both upper bounded, we always use the earliest census data for each country available in IPUMS to avoid biases as best as possible.[20] By using these criteria, we selected data from Bolivia, Brazil, Chile, Colombia, Ecuador, India, Kenya, Mexico, Panama and Tanzania where both age heaping and illiteracy were still observable. Table 4.1 and Table 4.2 give details on both datasets.

4.4 Methodology

What are the advantages of using regional data instead of national ones? Cipolla (1969) explicitly points out that *intra*national differences in human capital can be equally important as *inter*national ones. The importance of interregional variation is confirmed for many countries in the 19th century by Hippe and Baten (2012a). The data of that paper suggest that regional variation appears to be quite significant in numeracy, allowing a closer comparison of these regional numeracy data with literacy data.

Taking regions as the standard unit of analysis can considerably improve and complement the existing results obtained at the national level. This reasoning has also been recently promoted by other economic theories as, for example, by New Economic Geography and its Nobel laureate, Paul Krugman (e.g., Krugman 1991b). The regional perspective has the advantage that inherent cross-country differences do not bias the results as might be the case when performing pure country comparisons. This is particularly helpful for literacy comparisons because literacy was not always measured in the same way in each country and at each point in time. In contrast, numeracy values are directly derived from total census outcomes and do not rely on changing definitions throughout

[20] They are also lower bounded but this is not an issue concerning the data used here.

history. This appears to be an important advantage of numeracy for means of comparison. Moreover, early signature rates typically rely on rather small samples and do not cover the whole population but only a specified fraction such as married people at different ages, military recruits, etc. The majority of literacy data in this study are derived from the ability to 'read and write' as stated by all individuals in censuses and not only specific categories of the population. Therefore, it is a more modern and more complex interpretation of literacy which is still used in developing countries today.

Because the majority of age heaping studies are still rather recent (e. g., A'Hearn *et al.* 2009, Manzel and Baten 2009, Manzel *et al.* 2012, Stolz and Baten 2012), it appears necessary to explain the underlying methodology. In general, the age heaping method takes advantage of the fact that in many historical official documents the ages of the concerned group of people are listed. This is notably the case in population censuses but also other material can be used.

More specifically, individuals were asked their ages by a census taker. These statements are available either in individual or aggregate form. However, individuals did often not know their exact age. This is why a heaping is discernible on certain ages, i.e., the so-called age heaping. For example, an individual was 33 years old but told the census taker that he was 35. This means that there are clear rounding effects because the individual was not able to count correctly. In consequence, a heaping on '0' and '5' is visible in many historical cases.

Evidently, reasons other than human capital might be attributable to this age heaping effect. These include the role of administrations in public affairs when conducting the census and false age declarations on purpose. However, Crayen and Baten (2010a) have shown that the influence of human capital is the most important factor. Moreover, earlier studies on this numeracy proxy have shown a high correlation between numeracy and literacy on a national and in some cases at the regional level (Crayen and Baten 2010a, Hippe and Baten 2012a). These findings underline the significance of the human capital effect. The inherent characteristics of this method are also quite advantageous. Age statements are available for (almost) all time periods over the last couple of centuries and beyond. Thus, the long-term measurement of human capital becomes possible. Moreover, these data are often quite more spatially available, meaning that the analysis can be brought to smaller spatial units than before.

Nonetheless, it is clear that the age heaping method only captures very basic numerical abilities. In this sense, it is a proxy for very basic human capital values. These, however, persisted in today's industrialised countries until the 19th century and in many developing countries until today.

In comparison, literacy as measured by the ability to 'read and write' concerns already higher human capital levels. A more basic indicator would be the ability to 'read only'. This indicator is given only in few historical censuses, the majority preferring both capacities, i.e., to read *and* write. Exemptions prove the rule, as one of the early leaders in literacy, Sweden, was actually a leader in *reading* but not necessarily in *writing*. The aim was to enable believers to read the bible. Writing, in contrast, was not strictly relevant for this purpose. Still, the higher competencies required for reading and/or writing lead generally to the observation that numeracy values as measured by age heaping are lower than literacy values. This is important to know in the further analysis of the data.

In addition, both literacy and numeracy proxies used in this study rely on output values, i.e., they measure the performance of individuals in given tasks such as reading. On the other hand, other human capital measures such as enrolment rates, the number of schools or teachers are input indicators. They describe, for example, how many children go to school but it is apparent that the attendance of a school is more beneficial for some students than for others. The learning success is not equal among all individuals. In contrast, the human capital proxies analysed here measure the actual acquisition of some basic component of human capital. This makes them better comparable than using input proxies.

Moreover, proxies such as enrolment rates do not capture the labour force since children are still going to school. For this reason, their educational level does not represent the one of the labour force in production (Woessmann 2003). Using the standard age definition of numeracy, however, this is broadly the case.

The calculation of the ABCC Index, which is able to capture this age heaping behaviour of individuals, is as follows. The most intuitive way is to begin with the Whipple Index (which is also used by the United Nations). It is defined as

$$WI = \left(\sum_{i=5}^{14} n_{5i} \Big/ \frac{1}{5} \sum_{i=23}^{72} n_i \right) \times 100, \tag{4.1}$$

where *i* stands for years of age and *n* for the number of observations. Values range from 100 to 500. A value of 100 means that age heaping is not present and one of 500 that all age observations end on '0' and '5'. Because this range is not very intuitive, A'Hearn *et al.* (2009) propose a new index, the ABCC Index, which is a linear transformation of the Whipple Index:

$$ABCC = \left(1 - \frac{WI - 100}{400} \right) \times 100. \tag{4.2}$$

The ABCC Index has the advantage to be handier than the Whipple Index. Here, values range from 0 to 100, where 100 is the maximum numeracy level and 0 the lowest. This makes the analysis much easier because literacy rates are commonly defined as percentage shares. Thus, the ABCC Index achieves a higher comparability between literacy and numeracy values than the Whipple Index. Therefore, the following analyses are performed by using the ABCC Index.

Ideally, literacy should be defined in this study as

$$Literacy = \sum_{i=23}^{72} rw_i \Big/ \sum_{i=23}^{72} n_i, \tag{4.3}$$

where *rw* stands for the number of individuals able to 'read and write'.[21] This would give the same age range as for ABCCs and would arguably allow a maximum of comparability. Unfortunately, this perfect standardisation is not always possible because literacy is often defined as a share of individuals able to read and write above a certain age, often the age of 7 years. Therefore, the ABCCs are compared with the available literacy definition in each case. Details on the literacy definition of each country can be seen in Table 4.1 and Table 4.2. Moreover, Table 4.3 and Table 4.4 present some descriptive statistics.

4.5 Results

To get a first insight into the quantitative relationship between numeracy and literacy, Figure 4.1 to Figure 4.6 present scatter plots for both indicators at the regional level for the different countries under study in 19th century Europe.

Literacy and numeracy appear to be well correlated at the regional level in these countries. However, there are some apparent outliers. These outliers are mostly highly urbanised areas which are characterised by (far) higher literacy rates than other regions. This is notably the case in Greece (Attica, i.e., the greater Athens region; Figure 4.2), in Russia (e.g., the largest cities Saint Petersburg and Moscow; Figure 4.5) and Serbia (the largest cities Belgrade and Niš; Figure 4.6). But also in Italy (Figure 4.7) the regions of two of the three largest cities, i.e., Naples and Rome, are more advanced in literacy than their rather average numeracy level would suggest (between an ABCC of 80 to 85). A possible reason might be that bureaucracy was important in the capital and that literacy

[21] Including if a person is only able to 'read'.

was thus very important. Moreover, access to schools was much higher and easier in these urbanised regions and the focus was possibly still more on reading and writing ability than on numerical capacities.

In addition, the relationship between the ABCCs and literacy is not always the same. When the ABCC values are already quite high, literacy rates vary although the range of ABCCs is quite limited (e.g., in Spain). This changes when the ABCCs are on a lower level. A similar observation can be made when literacy rates are fairly low. In this case, literacy rates do not vary very much but ABCCs do, as for example in Serbia.[22] What might cause these differences in the slope? Clearly, ABCCs and literacy rates are bounded by 0 and 100 (%). This means that as ABCCs get closer to the upper limit, literacy rates are wider ranging and the relationship becomes less important. Therefore, when the mean values of the ABCC are higher, the slope is less high. This is a common issue for bounded variables and so for the two variables under analysis too.

Furthermore, Figure 4.8 displays the relationship in today's developing countries where the ABCC did not yet attain its upper bound at the time when the censuses were taken. Similar tendencies appear as in the former dataset. However, the literacy values are generally not as low as in some European countries in the 19th century. More precisely, Chile has on average the highest ABCC and literacy. Bolivia and Kenya are characterised by the highest ranges in literacy. In contrast, the highest deviation from the mean ABCC value appears in India. The fact that the data come from an employment survey might be related to this. Still, the relationship also appears to hold in these cases.

To test these first visual impressions, we use OLS regressions to evaluate the relationship between literacy and the level of ABCCs in the latter IPUMS sample for

[22] Not taking account here of the apparent urban outliers.

developing countries.[23] An advantage of the IPUMS database is that it allows us to include several explanatory variables to perform the regressions. A caveat, however, is that all variables are not always available for all countries. For this reason, we include only variables that are available for most countries in the sample to achieve a maximum of representativeness.

We test the impact of the different determinants on the ABCC by several OLS regressions. The regression model including all explanatory variables is

$$ABCC_j = \beta_0 + \beta_1\, literacy_j + \beta_2\, urbanisation_j + \beta_3\, electricity_j + \beta_4\, young_j + \beta_5\, capital_j + \varepsilon_j, \tag{4.4}$$

where j is a region, *literacy* is the share of literate individuals in the population (23-72 years), *urbanisation* is the share of individuals living in urban areas (as defined in the IPUMS database), *electricity* is the share of individuals with electricity in their households, *young* is the share of individuals aged 18 or less, *capital* is a capital region dummy and ε are the non-observed influences on the ABCC.

The first explanatory variable is literacy. Evidently, this is the crucial explanatory variable because we want to test the relationship between numeracy and literacy. Second, urbanisation may also have an impact on the ABCC. That urbanisation might play a role has been indicated by the earlier scatter plot analysis, where outlier regions where often more urbanised ones. Third, the level of infrastructure and administration in a region might have an impact on the ABCC. We proxy for this level through the variable electricity. The level of infrastructure and administration within a region might have a positive impact on the ABCC because basic education may benefit from infrastructure and infrastructural improvements (schools, etc.). Fourth, we include the share of individuals aged 18 or less

[23] In an alternative model, we run Tobit regressions because the given upper limit for the ABCC (100) might bias the results (the lower limit of 0 is not relevant given our dataset). We find that the use of the Tobit model does not affect the results obtained by running OLS.

(young) to check for the demographic composition of the population. Younger people can generally be expected to be better educated on average than their parents. Therefore, the higher the share of young people in a population, the higher their later contribution to the ABCC, increasing its overall level (see also A'Hearn *et al.* 2009). Finally, we control for the influence of capital regions by including a capital dummy.[24] Capital regions are often quite distinct from other regions because of the importance of the capital for the administration and governance of a country but also in economic and cultural terms. This reasoning was also highlighted by the fact that some capitals appeared to be outliers in the aforementioned scatter plots.

We proceed in the following manner: first, we regress the different explanatory variables individually on the ABCC to check the individual significance of each variable. In the second round, we run the regressions in the same way but keep always literacy in the model. In this way, we evaluate the importance of literacy with regard to other variables on the ABCC. Finally, the last regression includes all explanatory variables.

Note that we always control for individual country fixed effects by including country dummies.[25] Because countries have e.g. different geographical and institutional characteristics, the country dummies allow us to take these inherent differences between countries explicitly into account so that they do not bias our results. Moreover, as mentioned before, because all variables are not always available for all countries, the number of regions varies depending on the variables included.

The results are shown in Table 4.5. In the first five regressions of individual explanatory variables, all explanatory variables except young are significant and have the

[24] Capital regions are the following: in Bolivia 'Oropeza', in Brazil 'Distrito Federal', in Colombia 'Bogotá', in Chile 'Santiago', in Ecuador 'Pichincha', in India 'Delhi (Delhi)', in Kenya 'Nairobi', in Panama 'Panamá', in Mexico 'Distrito Federal' and in Tanzania 'Dodoma urban'.

[25] In a further specification (not shown), we also include time dummies in the form of decade dummies to control for the possibility of time effects. The results are unaffected.

expected sign. Literacy is highly positively correlated with the ABCC at the 1 % level. A positive correlation exists also with the level of urbanisation and electricity. Furthermore, capital regions have a higher ABCC than other regions.

The next set of regressions includes literacy and each of the other explanatory variables. Interestingly, literacy remains highly positively correlated (significant at the 1 % level) and its coefficient is very stable in all regressions between 30 and 35. In contrast, the other explanatory variables are not stable but change even in part their sign. This is the case for urbanisation and electricity. On the other hand, young turns positive and significant, whereas the capital dummy is now insignificant.

Finally, the last regression includes all explanatory variables. Again, literacy's sign, level of significance and its coefficient is stable with regard to the other specifications. However, this is not the case for the other variables. In particular, electricity changes its sign once again and the coefficient of urbanisation more than doubles. The respective sign and the significance of the young and capital variables remain, while both coefficients increase.

Therefore, the importance of literacy in explaining the ABCC is nearly unaffected by the individual or joint inclusion of the other variables. This is not the case for the other variables in our regressions. The robustness of these results underlines once more the positive and significant relationship in our sample between literacy and numeracy, as measured by the ABCC.

4.6 Conclusion

This paper has analysed the relationship between two proxies of human capital, numeracy and literacy. The importance of human capital is underlined both by recent theoretical work such as Unified Growth Theory (e.g., Galor 2005a) and numerous empirical studies. When comparing the literature on both indicators, the literature on literacy appears to be large with regard to the one on numeracy. Furthermore, the literature and the empirical evidence employing both human capital indicators at the same time are limited. For this reason, the present study addresses this gap in the literature. We show that numeracy and literacy have been historically intertwined so that, at least theoretically, a connection between the two variables should be at hand.

This link has been further investigated by using historical census data from Europe in the 19th century and more recent data from developing countries outside Europe in the 20th century. Numeracy was proxied by the ABCC Index and literacy by the ability to ‘read and write’. To allow a better comparability between the two human capital proxies, similar age ranges of the individuals were used, i.e., in most cases individuals between the ages of 23 and 72 years.

First descriptive evidence shows that numeracy and literacy indicators are well correlated both historically and in more recent data. The slopes become less high when ABCCs come closer to its maximum value. On the other hand, very low values of the ability to ‘read and write’ lead to a much more important range in ABCC values. This observation is common for bounded variables. Moreover, the historical data of several European countries show some outlier regions. These are (mostly) attributable to comparatively highly urbanised areas where literacy values are often much higher than numeracy ones. The importance of literacy for administrative tasks might play a role here. Still, both indicators of literacy and numeracy appear to be well linked to each other.

This outcome is underlined by further evidence from developing countries using the IPUMS database. By regressing several possible determinants on numeracy, literacy is shown to be its main determinant. This is indicated by the robustness of its high significance and by the stability of its coefficient. These results confirm once more the positive correlation between literacy and numeracy.

Still, neither numeracy nor literacy may appropriately capture human capital in today's more advanced countries. However, these indicators may deliver very valuable information on human capital formation in the past and for some regions of the world even today.

4.7 Appendix

4.7.1 Data

Data on European countries: see Hippe and Baten (2012a)

Data on Developing countries: Minnesota Population Center (2011). Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]. Minneapolis: University of Minnesota

4.7.2 Tables

Table 4.1 Details on historical European data

Country	Code	Census	Sexes	ABCC definition	Literacy definition
Greece	GR	1907	Both	23-32 yrs.	23-32 yrs.
Hungary	HU	1869	Both	23-72 yrs.	7+ yrs.
Ireland	IE	1841	Both	23-72 yrs.	26-75 yrs.
Italy	IT	1871	Both	23-72 yrs.	23-72 yrs.
Russia	RU	1897	Both	23-72 yrs.	23-72 yrs.
Serbia	SR	1895	Both	23-72 yrs.	7+ yrs.
Spain	ES	1887	Males	23-72 yrs.	21-70 yrs.

Table 4.2 Details on data from developing countries

Country	Code	Census years	Density	Sexes	ABCC definition	Literacy definition
Bolivia[a]	BO	1976	10 %	Both	23-72 yrs.	23-72 yrs.
Brazil	BR	1960	5 %	Both	23-72 yrs.	23-72 yrs.
Chile	CL	1960	1 %	Both	23-72 yrs.	23-72 yrs.
Colombia	CO	1964	2 %	Both	23-72 yrs.	23-72 yrs.
Ecuador	EC	1962	3 %	Both	23-72 yrs.	23-72 yrs.
India[b]	IN	1983	0.091 %	Both	23-72 yrs.	23-72 yrs.
Kenya	KE	1989	5 %	Both	23-72 yrs.	23-72 yrs.
Mexico	MX	1960	1.5 %	Both	23-72 yrs.	23-72 yrs.
Panama	PA	1960	5 %	Both	23-72 yrs.	23-72 yrs.
Tanzania	TZ	1988	10 %	Both	23-72 yrs.	23-72 yrs.

Note: a) Excludes 11 states in the north. b) Employment survey.

Table 4.3 Descriptive statistics for ABCC and literacy in Europe

Code	Obs.	ABCC				Literacy			
		mean	*sd*	*min*	*max*	*mean*	*sd*	*min*	*max*
ES[c]	50	93.34	4.70	83.90	100.00	0.53	0.17	0.27	0.84
GR	26	61.43	8.17	50.64	79.83	0.37	0.08	0.27	0.64
HU	18	89.97	3.29	82.23	94.73	0.49	0.09	0.31	0.63
IE	32	73.24	4.13	65.39	81.88	0.50	0.16	0.21	0.82
IT	69	88.62	6.93	72.62	99.01	0.29	0.15	0.11	0.65
RU	35	80.06	6.39	66.85	91.78	0.24	0.09	0.16	0.60
SR	18	60.08	9.30	41.77	82.75	0.17	0.14	0.06	0.61

Note: c) Spain includes as 50th observation its African possessions (Alhucemas, Ceuta, Chafarinas, Melilla, Penon de Velez de la Gomera and Rio de Oro) which are still a part of Spain.

Table 4.4 Descriptive statistics for ABCC and literacy in developing countries

Code	Obs.	ABCC				Literacy			
		mean	*sd*	*min*	*max*	*mean*	*sd*	*min*	*max*
BO	80	84.99	7.64	65.47	98.09	0.48	0.19	0.13	0.85
BR	16	86.87	5.95	74.53	95.57	0.52	0.15	0.31	0.75
CL	26	90.59	2.87	83.12	95.88	0.78	0.09	0.63	0.94
CO	30	86.37	6.44	65.27	98.27	0.64	0.15	0.31	0.95
EC	19	76.90	6.94	67.53	88.39	0.60	0.12	0.39	0.82
IN	75	62.40	8.54	44.34	85.86	0.45	0.15	0.21	0.94
KE	41	84.39	9.02	53.70	97.22	0.55	0.20	0.14	0.92
MX	32	80.15	7.06	65.25	91.78	0.65	0.15	0.35	0.88
PA	28	92.15	5.12	77.39	100.00	0.64	0.18	0.20	0.93
TZ	113	75.34	8.99	47.36	89.58	0.54	0.13	0.22	0.90

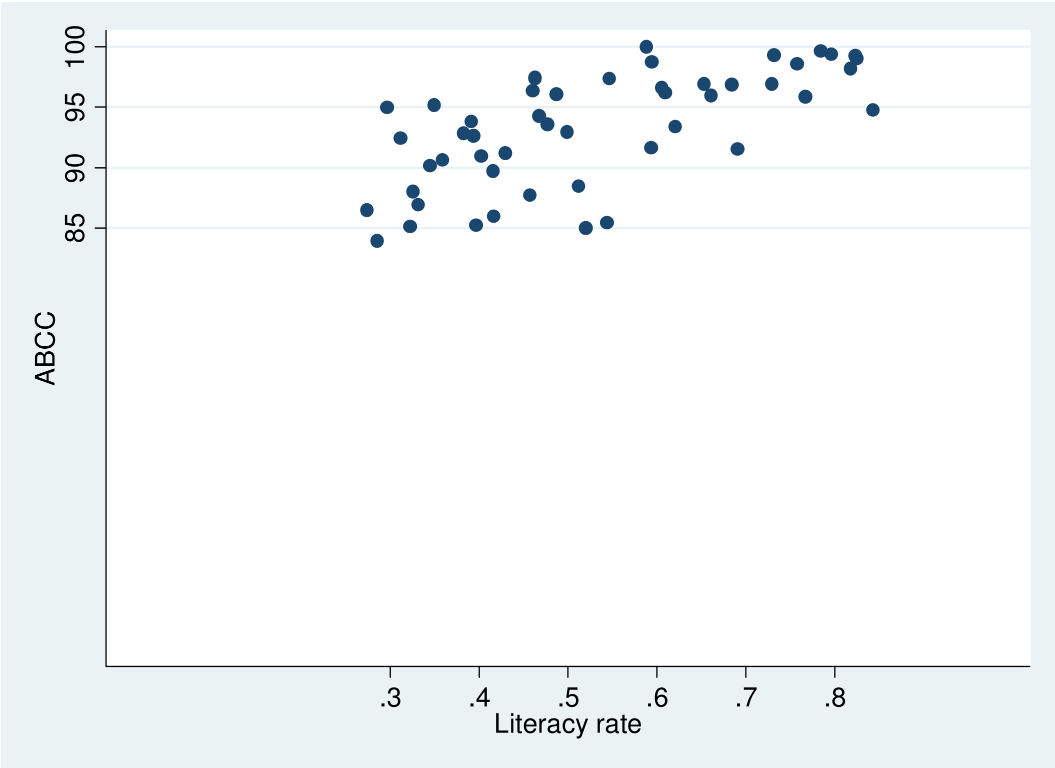
Table 4.5 Regressions of literacy and other determinants on ABCC

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable						ABCC				
Literacy	30.88***					33.37***	34.56***	31.51***	30.80***	30.80***
	(0.000)					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urbanisation		12.25***				-3.16*				-7.17**
		(0.000)				(0.083)				(0.040)
Electricity			15.94***				-5.67**			8.60***
			(0.000)				(0.042)			(0.009)
Young				-5.69				18.01**		24.66**
				(0.603)				(0.013)		(0.012)
Capital					6.62***				0.44	2.22
					(0.000)				(0.689)	(0.102)
Constant	70.17***	82.29***	82.04***	87.85***	84.91***	69.67***	69.45***	60.82***	70.20***	57.78***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	460	347	295	460	460	347	295	460	460	182
R-squared	0.74	0.70	0.38	0.58	0.58	0.83	0.58	0.74	0.74	0.67

Note : ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

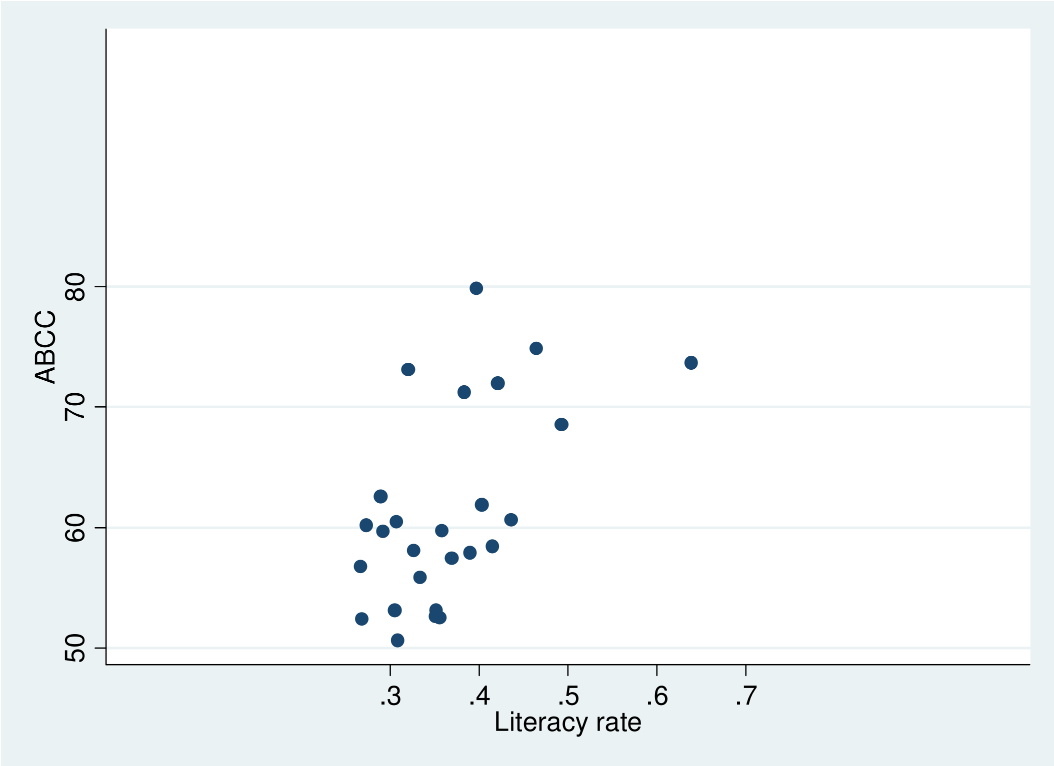
4.7.3 Figures

Figure 4.1 ABCC and literacy in Spain



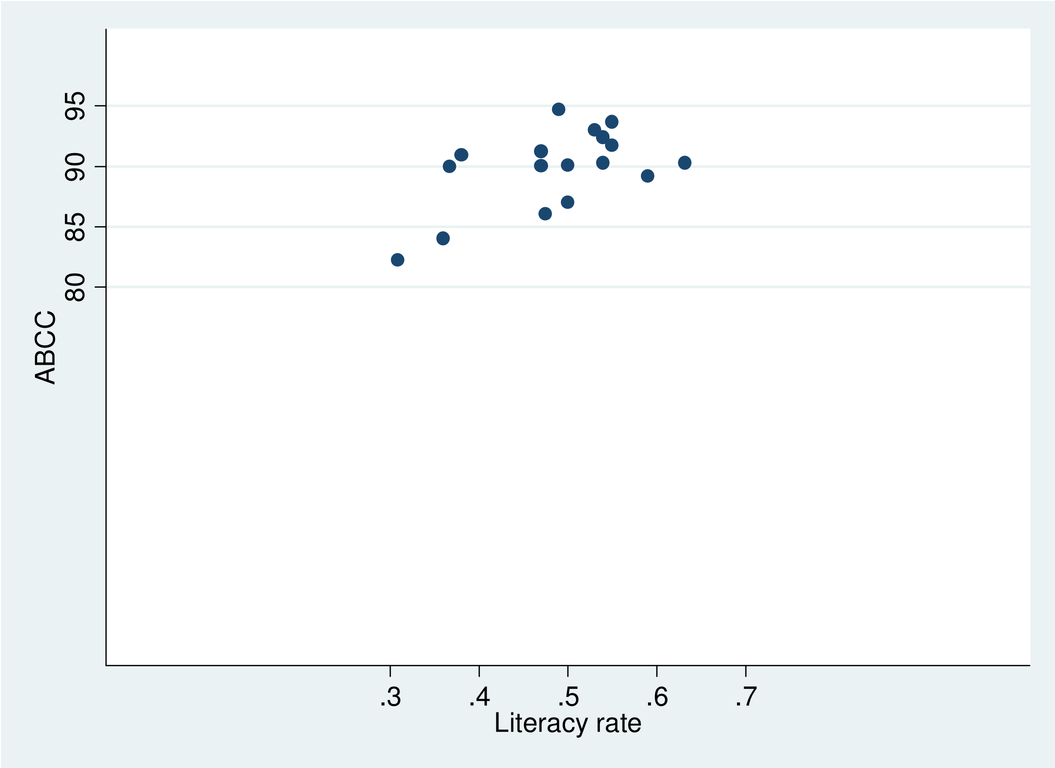
Note: Only data for the male population are available.

Figure 4.2 ABCC and literacy in Greece



Note: Only data for the ages 23 to 32 years are available.

Figure 4.3 ABCC and literacy in Hungary



Note: Hungary within today's borders.

Figure 4.4 ABCC and literacy in Ireland

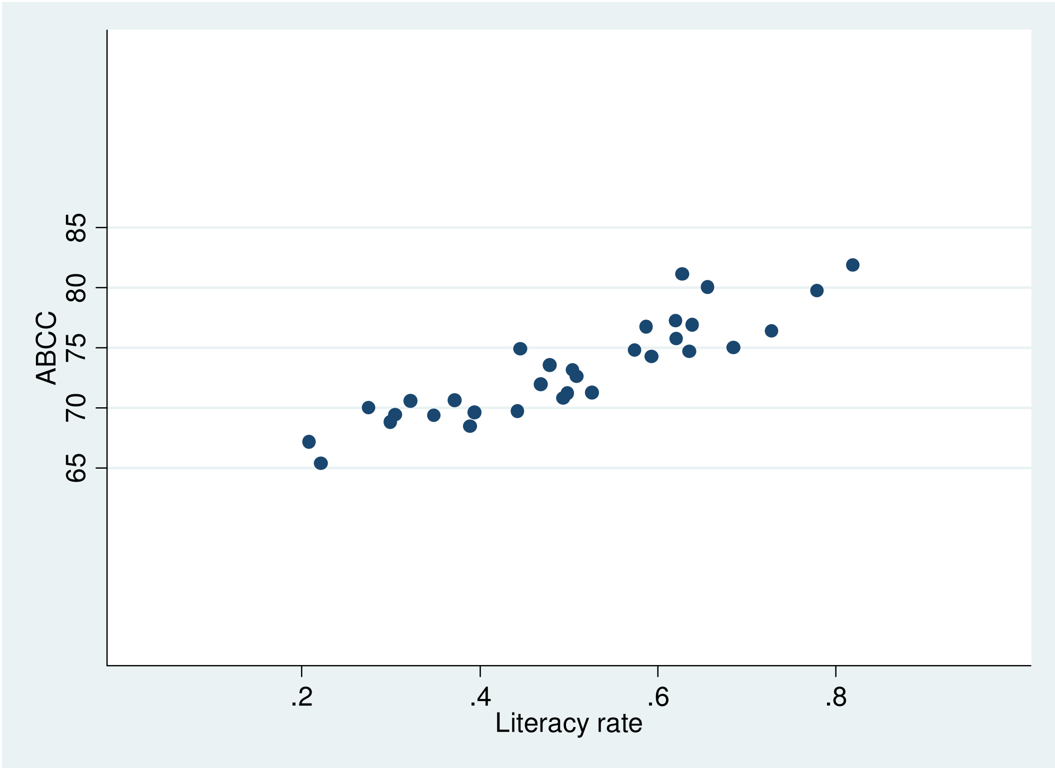
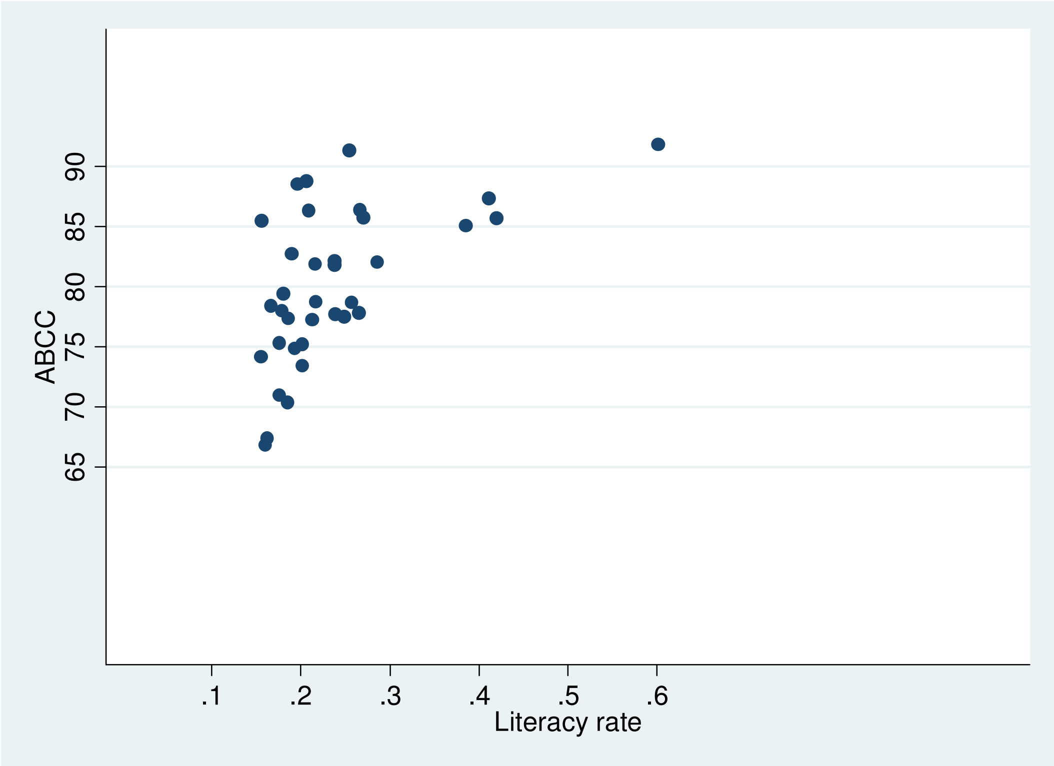


Figure 4.5 ABCC and literacy in Russia



Note: Russia comprises Russia's European part in today's borders.

Figure 4.6 ABCC and literacy in Serbia

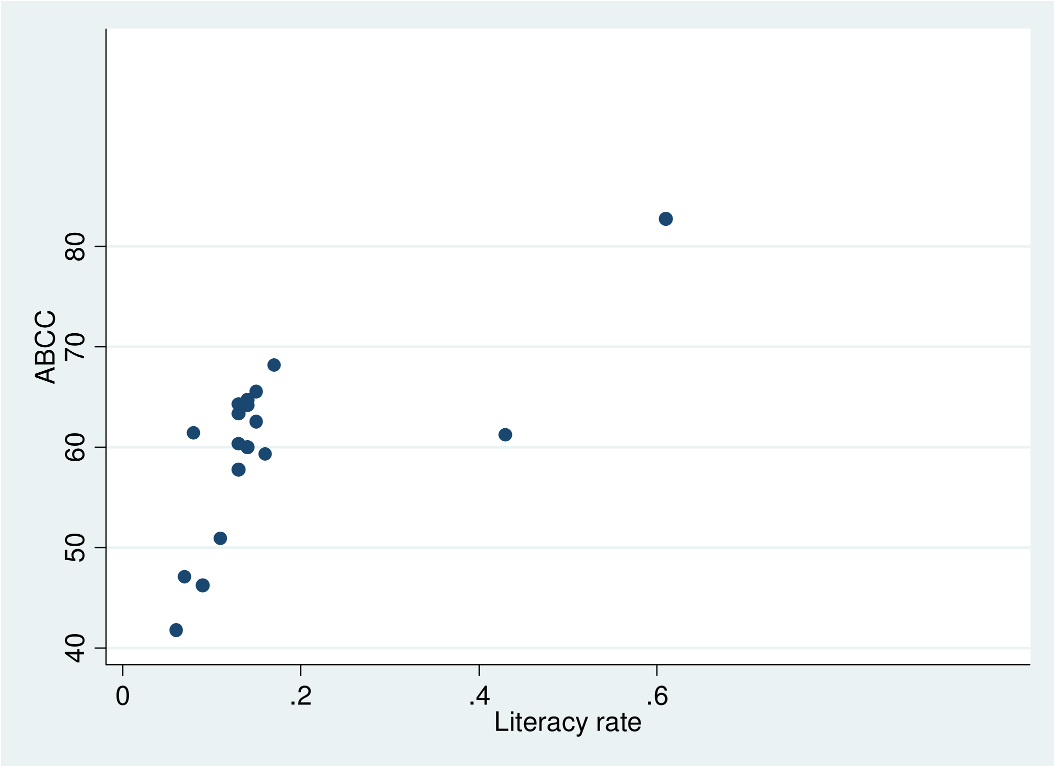


Figure 4.7 ABCC and literacy in Italy

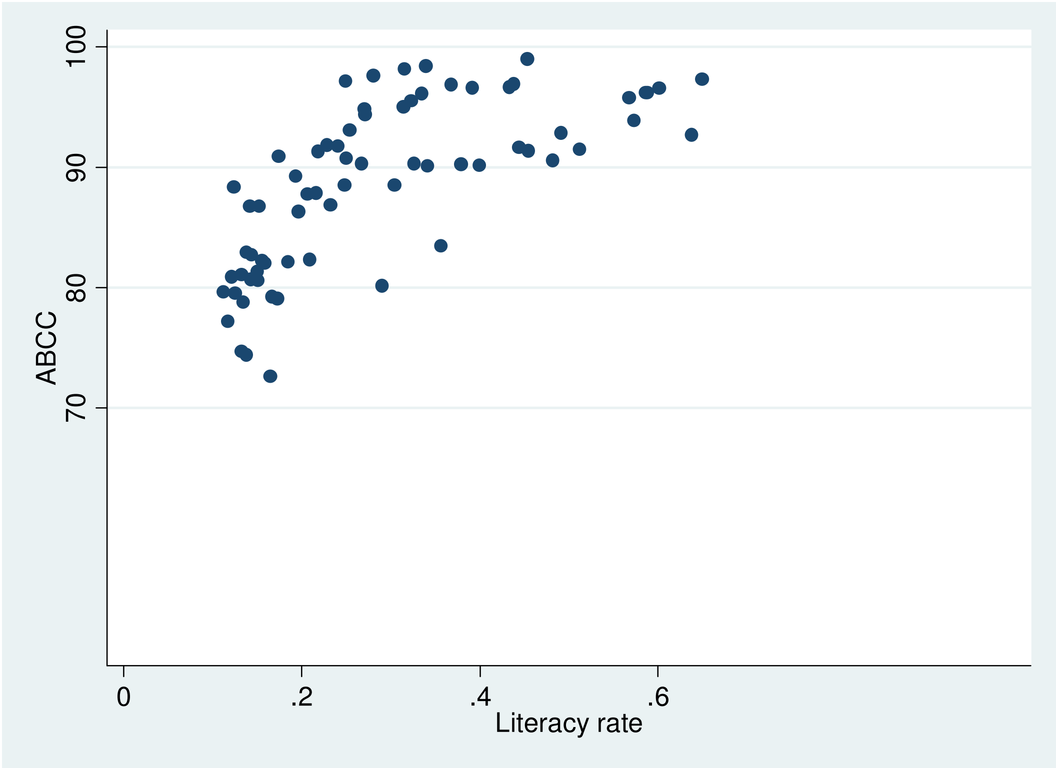


Figure 4.8 ABCC and literacy in developing countries

Note: bo = Bolivia, br = Brazil, cl = Chile, co = Colombia, ec = Ecuador, in = India, ke = Kenya, mx = Mexico, pa = Panama, tz = Tanzania.

5. Regional inequality in human capital formation in Europe, 1790 – 1880

Recent theoretical advances reveal the importance of human capital for long-run economic growth. However, the absence of data makes it difficult to measure human capital before 1870 at the national level, let alone at the regional level within countries. By using the age heaping method and a large, new dataset, we approximate the numeracy values in more than 570 regions in Europe between 1790 and 1880. The results indicate a significant gap in numeracy levels between advanced west and central European countries and the rest of Europe. Nevertheless, differences in basic numeracy between and within countries became smaller over the 19th century, as the periphery solved its basic numeracy problem.

This chapter is based on a working paper co-authored with Joerg Baten (University of Tuebingen). The concept for the paper was developed jointly; the analyses and writing were equally shared. It has been published in the *Scandinavian Economic History Review*.

5.1 Introduction

Human capital is one of the most important determinants of economic growth, particularly during the transition from Malthusian stagnation to modern growth, as highlighted by Unified Growth Theory (Galor and Weil 2000, Galor and Moav 2002, Galor 2005a, Galor 2012). Already, the first endogenous growth models in the 1980s revealed its importance (Lucas 1988, Romer 1990). However, were regional inequalities in education aggravated or smoothened within countries over the 19th century? Answering such a question using quantitative techniques has turned out to be difficult, although doing so is important in order to understand the importance of human capital in the long run. Because human capital cannot be measured directly, proxies such as literacy, schooling or numeracy are commonly employed. Thus, estimations of human capital levels in Europe have been possible for some countries at the national level

One of the key characteristics of this study is that we are able to take a look at the regional level. In fact, most previous studies on human capital have either considered the national level, thereby comparing different European countries, or have shown the differences within one country (e.g., Allen 2009, Becker and Woessmann 2009). We go beyond these limits and explore the evolution of and the inequality in numeracy of most European countries at the regional level. Considering regions in lieu of countries allows to improve our understanding of underlying determinants and consequences. Renowned authors such as Cipolla highlight that these regional differences in human capital within a country can be important, and can in part be even higher than between countries, and therefore they should be taken into account (Cipolla 1969). More generally, recent economic theories such as New Economic Geography also underline that the regional level is the most pertinent level of analysis (Krugman 1991b). For these reasons, it appears crucial to obtain more evidence on regional inequalities that provide much more profound insights than mere national averages. Therefore, this paper is a contribution toward filling this gap in the literature. It explores human capital levels for the first time in the regions of Europe during the period between 1790 and 1880, by employing the age heaping method to calculate numeracy values. This method has increasingly been used in the recent literature on long-run human capital formation (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010a). As a result of our large, new data set, we are able to cover nearly all of the countries of Europe, totalling more than 570 regions ranging from Portugal to Russia and from Norway to Italy.

The study is structured as follows: first, we outline some important aspects of the existing evidence on the development of economic performance and of human capital in Europe in the 19th century. Then, the age heaping method is presented. Subsequently, the characteristics of the data and the spatial methodology are discussed. Our data stem from population censuses that were taken mainly in the 19th century. The historical regions are coded into today's current regions. For this reason, we employ the NUTS classification developed by the European Union. Descriptive, cartographic and statistical results on mean numeracy values follow in the next section. We analyse numeracy first on the national and then on the regional level. Based on these results, we compute regional inequalities in numeracy. Regional inequality within a country is measured by the coefficient of variation, thus allowing comparison across all European countries. A final conclusion summarises the results of the paper.

5.2 Economic differences between European countries in the 19th century

How did the economies of the various European countries in the 19th century develop in comparison to each other? O'Rourke and Williamson state that, during the second part of the 19th century, economic performance was converging between the 'core' and the 'periphery' countries in Europe (O'Rourke and Williamson 1997). The term 'core' describes the geographic location of countries in Europe and their industrial output at the same time. In this context, the core countries include Belgium, France, Germany, Great Britain, the Netherlands and Switzerland, whereas the periphery countries include Finland, Ireland, Italy and Spain. Taking GDP per capita as an example, the authors conclude that periphery countries had values of only about a third of the core countries' values in 1870. Although the periphery advanced until 1914, the following decades revealed important growth differences within this group. In particular, while the Scandinavian countries caught up to the core, countries in southern Europe fell behind it. Slow growth for the southern European countries was found by Good and Ma (Good and Ma 1999). In contrast, countries in central and eastern Europe grew at much higher rates, resulting in a pattern of overall convergence. Ivanov and Tooze are less optimistic for the case of the Balkans because their data suggest that GNP per capita stagnated in Bulgaria throughout the period (Ivanov and Tooze 2007). Similarly, Foreman-Peck and Lains argue that countries such as Serbia and Greece were not able to generate high growth rates (Foreman-Peck and Lains 2000).

How can we explain these differences in economic growth? For example, Tortella describes the case of the 'Latin' countries in detail, i.e., Italy, Spain and Portugal (Tortella 1994). These countries share common cultural and geographic attributes, which could partially explain their similar economic retardation up to that time. He argues that agricultural conditions were worse than in the northern countries of Europe due to the adverse characteristics of the climate and the soil, which decreased agricultural productivity and the possibilities of technological progress. Long-lasting budget deficits of the governments and similar, detrimental land tenure systems were other drawbacks to these countries' efforts to catch up economically. Moreover, the education of the population, as approximated by literacy rates, was on a much lower level than in Great

Britain, France or Belgium. Basic education, as measured by literacy (and, we may add, numeracy), could have an impact on the individual in several ways: on the one hand, workers may become more productive, and on the other hand, they may be able to adapt faster to changing environments and new challenges. Finally, they acquire new skills and learn more quickly. Thus, the lack of education may be a factor that limited the economic performance of these countries.

5.3 Human capital and education in 19th century Europe

In line with this reasoning, Sandberg showed that education was an important factor for long-run economic development in Europe (Sandberg 1979). Support for the argument that education mattered for economic growth has been established for a range of countries, including Prussia and Italy (e.g., Zamagni 1978, Becker *et al.* 2011). Nevertheless, the exact relationship between human capital and growth in the 19th century is still not very clear.[26]

More generally, the question of the contribution of human capital to economic growth has a long history. These discussions have been going on during many decades if not centuries.[27] An important part of those discussions with a historical dimension have focused on the contribution of human capital to the Industrial Revolution (IR). David Mitch found that literacy and educational development did not increase during the period of the IR in England (Mitch 1993). Joel Mokyr supported this point in his knowledge based interpretation. However, England was already on a very high level of advanced human capital before its Industrial Revolution really started (Mokyr 2002). Baten and van Zanden

[26]As an example, O'Rourke and Williamson argue that, for the period between 1870 and 1913, forces of globalisation were more important for economic development than education. See O'Rourke and Williamson (1997).
[27] For an excellent overview see Demeulemeester and Diebolt 2011).

highlight that book production was unusually high in the UK during the early 18th century (Baten and van Zanden 2008).

A higher level of knowledge of individuals may generally be argued to be beneficial for society as a whole. For this reason, in an ideal world all social groups should have the maximum level of any kind of knowledge. To be more restrictive, the notion of useful knowledge appears to be important in our context. Useful knowledge is defined by Mokyr as knowledge dealing "with natural phenomena that potentially lend themselves to manipulation, such as artifacts, materials, energy, and living beings" (Mokyr 2002, p. 3). However, to be even more specific, one can say that the three most important aspects of useful investment in knowledge or education in the 19th century were numeracy, literacy and skills that were specific to trade (Allen 2009). Acquiring knowledge in these areas was helpful or even necessary for different parts of the population. For example, trade by its very essence deals with quantities and thus with numbers. Being engaged in trading needed a certain basic knowledge of numeracy. But also literacy skills were quite valuable for economic development, as Becker and Woessmann argue (Becker and Woessmann 2009). This was not only true for the upper classes. Tollnek and Baten show that even farmers had to deal with numbers (Tollnek and Baten 2011). For example, they had to be numerate in climate-sensitive decisions about harvest timing and they were also engaged in small trade operations to sell their products in the market. This made it necessary for them to acquire basic numeracy skills.

One central limiting factor in this analysis is the availability of data on human capital. Unfortunately, evidence on regional development and regional inequality of human capital in Europe before the 20th century is still rather scarce, particularly for the time before 1870. Human capital cannot be measured directly. Instead, it has to be approximated by related, quantifiable variables. Examples of such proxies employed for modern times include literacy, numeracy, enrolment rates, years of schooling or books per capita (e.g., Benavot and Riddle 1988, Baten and van Zanden 2008, A'Hearn *et al.* 2009, Becker and Woessmann 2009).

In addition to these indicators, the beginning of compulsory schooling had a strong effect on subsequent educational levels. A minimum level of the education of soldiers and ordinary citizens was supposed to ameliorate the military and economic position of the country (Brint 2006). This motivation was one of the reasons why more and more countries constructed mass education systems during the 19th century. These systems replaced, integrated or partly expanded the more informal private or religious institutions that had existed until then (Soysal and Strang 1989).

Religious institutions had played a major role in providing and thus enforcing specialised instruction for centuries, particularly in Protestant countries (Vincent 2000). Consequently, alliances between the state and the national church facilitated nationwide education in some countries. In others, however, schooling laws even led to conflict, as both the state and religious organisations struggled to control schooling. For instance, the Catholic Church fought in France against the state in order to teach conservative values, whereas the state preferred to promote Republican ideas. Hence, compulsory schooling was enacted relatively recently and in a successful way (by the Ferry law in 1882). As a consequence, many different religious and societal groups tried to inhibit the development of a unified educational system in Great Britain (Soysal and Strang 1989). Furthermore, formal education stressed the individual's educational achievement and socialisation. Weber stated that these attributes are similar to the reasoning given by Protestants of having a direct link to God (Weber 1958). As an alternative to his famous Protestant working ethic, Becker and Woessmann argue that Protestantism encouraged reading the Bible, so Protestants became more literate than other religious groups (Becker and

Woessmann 2009). This generated a more educated population which could also use these skills for economic activities. Finally, this practise spurred economic development in these countries. In any case, national Protestant churches generally supported mass schooling, as they did in Prussia, while the Catholic Church was often opposed to it, e.g., in France.

Consequently, the introduction of compulsory schooling varied enormously in the European countries. Prussia, Sweden and Scotland were the pioneers in introducing compulsory schooling. Denmark (1814), Greece (1834) and Spain (1838) were the first to follow in the 19th century, whereas other western European countries, such as Belgium (1914) and the Netherlands (1900), were very late in this pursuit.[28]

However, merely passing schooling laws and the reality of schooling have often been two different matters (Flora 1975). For instance, states such as Greece, Portugal and Spain were eager to achieve mass education, but their influence at the local and the societal level was often too weak to achieve this goal. Other, additional methodological concerns arise as well, because compulsory schooling laws required, in part, merely erecting school buildings. This, however, is not always equivalent to teaching pupils on a compulsory basis. For these reasons, compulsory schooling laws are not an appropriate means to measure time differences in human capital development, as early schooling laws did not ensure higher enrolment rates in later decades and centuries (Vincent 2000, Adick 2003). For instance, Prussia's initial lead in schooling laws in the 18th century did not result in higher enrolment rates at the end of the 19th century than in countries that had not had such laws passed until then (Schneider 1982).

In addition, most other methods are not able to estimate human capital levels before the second part of the 19th century. For instance, school enrolment data are

[28] These dates refer to the years when the first important general schooling laws were passed. These laws aimed at enforcing the attendance of all children until they reached some specific age.

generally rather scarce.[29] To obtain literacy rates, signature rates are used in a range of studies (e.g., Schofield 1981, Mitch 1993, Reis 2005). The potential disadvantages of this method are openly acknowledged by their applicants. For instance, it is not always possible to discern if the person himself signed a marriage contract, nor is it always clear what importance has to be attributed to the responsible priest in this context. Additionally, this indicator is not always available in order to compare the regions of Europe on a larger scale.

Still, we may take a glance at some results of these studies. An important study on literacy was published by Cipolla already in 1969 (Cipolla 1969). This study, although not the first,[30] opened a major discussion on literacy in the following years. Graff has added a number of studies on this issue (Graff 1987, Graff 1991, Graff *et al.* 2009).

Recently, Allen published his influential book on the British Industrial Revolution in which he also provides estimates for the proportion of adults who could sign their names in different parts of Europe (Allen 2009, see also Cressy 1980, Graff 1987). His estimates show that literacy in 1800 was by far the highest in the Netherlands (68 %), followed by England (53 %) and Belgium (49 %). France (37 %) and Germany (35 %) lay in the middle, while the lowest shares were found in Italy (22 %), Austria-Hungary (21 %), Poland (21 %) and Spain (20 %). Globally, one may summarise that there is a pattern of north-western countries which had signature rates well above those in eastern Europe and in southern Europe. Unfortunately, there is no evidence included on other important European regions and countries such as Scandinavia, Ireland, Russia and south-east Europe as a whole.

[29] A notable exception is Prussia, see e.g., Becker and Woessmann (2009).

[30] Earlier studies on literacy include, e.g., the one by Fleury and Valmary (1957).

By contrast, the 'age heaping' strategy allows one to go beyond most of the limitations already mentioned. Thus, A'Hearn *et al.* trace numeracy levels for 16 European countries between 1350 and 1840 (A'Hearn *et al.* 2009). They observe a striking discrepancy between numeracy in western Europe (e.g., France, UK, the Netherlands) and eastern Europe. High numeracy levels were already found for the western European and Scandinavian countries in the 17th and 18th centuries. Central Europe generally comes close to western European numeracy levels, whereas eastern Europe stays far behind. The differences are greatest at the beginning of the data for eastern Europe in the 17th century. Similar results were established in an even more recent paper on the development of global numeracy (Crayen and Baten 2010a).

5.4 Deriving age heaping from historical censuses

The age heaping method investigates the numeric skills of a population and has been used in a multitude of recent studies (e.g., de Moor and van Zanden 2008, A'Hearn *et al.* 2009, Manzel and Baten 2009). It uses the declarations made in census records or other documents, as is also the case in this study. Census taking itself has a very long history and, therefore, the age heaping technique can be used for long-run estimations of human capital. The ancient Egyptians, Persians, Hebrews, Greeks, Romans and Japanese had already conducted censuses. The motivation for these historical censuses was "limited to matters of wealth, military campaigns, or defense" (Goyer and Draaijer 1992, p. 6). During the Middle Ages, censuses and the counting of people were rather rare. A well-known and, for its time, very detailed example of a census was the Domesday Book of 1085, which counted the English population for tax purposes. Moreover, censuses on the city level were reported in Europe during the 15th and 16th century. As trade between cities and countries increased during the Renaissance, not only were goods counted increasingly more often, but the rulers also became interested in knowing the number of their subordinates. This went hand in hand with a change in the perspective of the 'value' of the people. Individuals were no longer regarded only as taxpayers or soldiers but also as economic assets by the state authorities. The importance of this manpower further increased during the Industrial Revolution.

The first prototypes of what is commonly called the modern census include those of New France (Québec) in 1666 and of Iceland in 1703. On a much larger scale, Sweden undertook its first census in 1749 (Goyer and Draaijer 1992). Modern times witnessed a great proliferation of census taking, beginning on a decennial basis in the United States (1790) and England (1801) and quinquennially in countries such as France (1801). On the international level, states exchanged their views on census taking increasingly more often, culminating in the International Statistical Congresses being held. The first of these congresses took place in Brussels (1853). They made recommendations and the first principal requirements for census taking (Sahai 1988).

The wealth of this census material generated in the 19th century allows us to construct estimates of human capital for the whole of Europe. By using the age heaping method, we can take advantage of the fact that one of the basic questions posed to individuals by a census taker was their age. However, parts of a population did not know their exact age in Europe at that time.[31] Consequently, individuals rounded their ages by '0' and '5'.[32] For example, a 47 year old man would erroneously have told the census taker that he was 50 years old. Census records therefore depict a typical heaping of these ages. The census of Bulgaria serves as an example (Figure 5.1).

[31] Yet age heaping is still observable in current censuses in parts of Asia and Africa.

[32] The question may arise whether all individuals stated their ages by themselves. Friesen *et al.* find by distinguishing between males and females in their data that a large gender gap in literacy existed also in numeracy. This is a strong indication that females were asked themselves because the differences between both gender gaps would be much more pronounced (Friesen *et al.* 2012).

One can discern also other heaping patterns. For example, there is a modest preference for ages ending on '8' compared to '7' and '9'. However, people rounding on '8' are often a middle group who can specify their age more precisely than people rounding on '0' or '5', but they cannot report an exact age. In previous studies, the best way to model age heaping has turned out to focus on the extreme cases (A'Hearn *et al.* 2009). These extreme cases are very clearly the rounding effects on '0' and '5'. Empirically, the correlation with other human capital indicators is maximised.[33]

Factors other than human capital could also be attributed to the age heaping phenomenon. For instance, intentionally false age declarations, which were made in order to avoid the negative consequences of being part of a particular age group, and bad state administrations played some role. Also, the awareness of one's age in early adulthood is often increased because of events such as marriage. However, other studies have already demonstrated that educational investments play the most important role in age heaping, when they are included in multiple regressions, and that other institutional factors have no systematic impact (Crayen and Baten 2010a).

Moreover, age heaping is highly correlated with literacy indicators, which has already been demonstrated by previous research (e.g., Crayen and Baten 2010a). In order to make an analysis with regional data, we plot age heaping against literacy data. The literacy rate is defined here as the share of individuals in the population that is able to read and write. We use the smallest available administrative unit. With respect to our data, this unit is most often the county level. For example, we find a high correlation between literacy and numeracy in Ireland in the census of 1841 (Figure 5.2). We are also able to plot the data separately, age group by age group. The differences between one age group

[33] Moreover, because this has been the standard way of calculating numeracy values derived from age heaping, it allows to be in line with the other contributions. This standardisation enables us and other researchers later on to compare the numeracy values from different studies directly.

and another are thus shared by both indicators. Another case is Serbia in 1895 (Figure 5.3). However, literacy data are not distinguishable by age in the case of Serbia. For this reason, we look at the relationship between age heaping for all ages between 23 and 72 years and the overall literacy data. We obtain a similar picture with corresponding results.[34]

One advantage of age heaping is that, in contrast to literacy rates, the indicator measuring age heaping is calculated using the data on the age distribution in the population statistics. Therefore, it is not explicitly given in the statistics. As a consequence, it is less prone to voluntary manipulation by state authorities. These authorities may have had, in some cases, an interest in governing a population characterised by high literacy levels in order to hush up the backwardness of their education and their economic system. Nevertheless, statisticians may be tempted to smooth the peaks in the age distributions to arrive at the real age structure.[35] In addition to factors such as costs and time consumption being used in establishing detailed statistics for individual age years, this may be another reason for why a range of census publications does not contain tables on individual ages but on age groups.

All in all, age heaping allows for the measurement of the basic numeric skills of a population in general and an analysis of the development of human capital in most European regions in a long-term perspective in particular. In this paper, it is calculated by using a transformed Whipple Index, the ABCC Index. The original Whipple Index (WI) relates the number of age observations on '0' and '5' to the total of observations. It is defined as follows:

[34] Nevertheless, we want to indicate that the possibility remains that numeracy could improve in a population independently of improved literacy or schooling. Baten *et al.* consider China and find evidence for such a case in this East Asian country (Baten *et al.* 2010). Basic numeracy is in fact an ability which can also be trained in households by using calendars, numerically demanding games for children and similar devices.

[35] Yet if they did so, then age heaping would normally be equal to zero, which can mostly be identified.

$$WI = \left(\sum_{i=5}^{14} n_{5i} \middle/ \frac{1}{5} \sum_{i=23}^{72} n_i \right) \times 100, \tag{5.1}$$

where *i* stands for the years of age and *n* stands for the number of observations. Values range between 100 and 500, where a value of 100 means that age heaping is not present and 500 means that all age observations are in multiples of five.[36] Because this range is not very intuitive, A'Hearn *et al.* proposed a new index, the ABCC Index (A'Hearn *et al.* 2009). It is a linear transformation of the Whipple Index, as can be seen by the following formula:

$$ABCC = \left(1 - \frac{WI - 100}{400}\right) \times 100. \tag{5.2}$$

The ABCC Index has the advantage that it is more accessible and more comprehensive than the Whipple Index. Values are from 0 to 100, where 100 is the maximum numeracy level and 0 the lowest. Consequently, the range of this index is identical to the ranges of other conventional indices, which makes comparisons easier. Therefore, the following analyses are performed using the ABCC Index.

Additionally, as in previous studies, we limit our age data to the years above 23 and below 72. Under the threshold of 23, it is possible that individuals did not themselves declare their age and their parents did it for them. Moreover, ages above 72 may be prone to a selection bias because only those who are still alive can be counted. Because the individuals benefiting from a long life represent only a part of the total, it is advisable to exclude them from the analysis. Consequently, the numeracy values can be calculated for five subsequent birth decades. Moreover, as proposed by Crayen and Baten, we have adjusted the age group ranging from 23 to 32 years because this age group heaps systematically more on multiples of 2 and less on multiples of 5 (Crayen and Baten

[36] Values below 100 are also possible but are normally found in samples with a low number of observations.

2010a).[37] Because the ABCC is derived from the Whipple Index, the use of this formula automatically gives us the age corrected ABCC value.

Moreover, it is conceivable that the individuals at the higher end of the age range (e.g., between 53 and 72 years) heap considerably more than their younger counterparts. However, older people do not significantly misreport more often their age relative to what can be expected, as has also been put forward by Crayen and Baten (Crayen and Baten 2010a). In other words, it is true that older people, for example those aged 63-72, seem to have lower numeracy. However, during the 19th and 20th century this can be entirely explained by the fact that they were born earlier in time, when educational inputs were still lower.

5.5 Data

To compare the development of numeracy in most European regions, a large, new dataset has been assembled from many individual sources, typically census records. This has the advantage that official documents are used. These are often well documented with respect to the methods employed during census taking. Possible measurement problems can thus be avoided or corrected. Additionally, the data employed are total population censuses, meaning that we have included all individuals of each country between the ages of 23 and 72 years.[38] This makes the data much more representative than alternative measures which only use selected groups, such as married people or military recruits. Altogether, the database includes samples covering over 570 regions in 39 European countries (in today's borders) for the time period between 1790 and 1880.

---

37 In practice, the precise adjustment of the corresponding age group is an addition of "0.2 Whipple units for every Whipple unit above 100" (Crayen and Baten 2010a, p. 95). This implies a reduction of ABCC estimates by roughly one quarter.

38 Except for the census of Greece in 1907, where it is only possible to include age data from people between the ages of 23 and 32 because data for individual years is only available up to the age of 34.

An overview of the countries covered in this study and the corresponding time frame is given in Table 5.1. Clearly, not all the data refer to the same time period. The reason for this is that the censuses included were taken in different years and decades. Whereas the earliest data are from Ireland, France, the Netherlands, Belgium and Denmark, the most recent data come from Portugal and Cyprus.[39] In general, there is a tendency to have the earliest censuses from countries in north-western Europe and the last in south-eastern Europe. We will see later on in the paper that this globally corresponds to the core-periphery in numeracy. For this reason, one could come to the conclusion that censuses are only held after some basic level of numeracy or literacy is established. Yet evidence from other world regions does not support this scheme, because there were already censuses in e.g. Roman Palestine in biblical times where literacy or numeracy levels were very low. Still, the costs of taking and publishing a census inhibited these statistical analyses in less advanced regions for a longer time than in the more advanced western Europe.

Still, the detailed geographical coverage allows analyses of the regional distribution of human capital, which has not been possible until now. Moreover, future research will enable us to fill in some of the missing data.

An issue that has to be taken into account when studying data from different census years is border changes. Because national borders changed during the 19th century in some cases, due to wars or other events (e.g., the French region Alsace was annexed by the German Empire in 1871), it is possible that a region is not listed in any available census or is included in censuses of two different countries. In the former case, we had to limit ourselves to these data restrictions, and in the latter case, we interpolated the corresponding regional values or opted for the most appropriate one for intra-country comparisons.

[39] The exact sources are documented in the appendix.

For the age heaping method to be employed, data for individual ages are required. Unfortunately, some countries preferred not to ask the individuals of the population their exact age. In its place, these had to declare their age in certain age groups (20 – 25 years or 20 – 30 years, etc.), for example, as in some German states prior to unification. In these cases, it is not possible to use the age heaping method. In contrast, due to space requirements or other reasons, countries sometimes indicated aggregated age groups in their official census publications instead of individual years. This is true for some available publications, e.g., for Greece or Malta in the period of our study. Unfortunately, original census lists have not always survived to allow the compilation of individual, disaggregated data, which could otherwise be used to obtain the required individual ages.

Finally, Sweden is a special case, due to its tradition of using birth registers and similar documents to construct population data, instead of questioning the population itself in a real census. This means that age data were not derived by the responses of individuals. For this reason, we cannot use these data. The same case can be made for Finland, which inherited the same counting strategy from its time under Swedish rule. Population registers were not always used but it was clearly the preferred method to estimate data on the whole population.[40] Consequently, the evidence does not yet cover Sweden and Finland.[41] However, it is still possible that further research might allow us to collect data for these countries and from sources not yet available to us. Still, we were able to include all other countries of Europe, except small area states, such as Malta or Vatican City, in this study.[42] This yields a geographical coverage previously unachieved in the literature.

---

[40] In contrast, data on the population of the major cities were collected by census takers, as in other European countries. Thus, these data are also characterised by age heaping. Nevertheless, because all our other data are homogeneous in the sense that they always cover the whole population, and because the exact relationship between the urban and rural population is not yet sufficiently clear, we decided not to include these data in the framework of this study.

[41] Nevertheless, the Finnish data already indicate that numeracy should be at a high level in this country.

[42] For Romania, we only have data for the parts of the Hungarian Kingdom.

However, it is important to explicitly define what is meant by a 'region'. Our territorial definition of a region corresponds to the current NUTS classification employed by the European Union. The 'Nomenclature des Unités Territoriales Statistiques' (NUTS, Nomenclature of Territorial Units for Statistics) was adopted in 1988 to produce regional statistics within the European Union. NUTS covers all the member states of the EU[43], the EFTA countries[44] and Candidate Countries[45] whose aim is to join the EU. It is obvious that territorial units in the European countries in the 19th century were, in many cases, quite different from today.[46] In particular, two World Wars changed the territories of many existing countries, and new countries were formed by dividing old Empires (in particular, Austria-Hungary). In addition, World War II had major impacts on the (ethnic and linguistic) composition of the population of many regions and administrative reforms carried out by the states changed the territorial characteristics of a multitude of regions.

By contrast, European population density patterns have stayed more or less the same in Europe for more than a hundred years (Martí-Henneberg 2005). Martí-Henneberg measured a high correlation of 0.83 by comparing the population density patterns in Europe in the years 1870 and 2000. Thus, many highly populated areas have attracted individuals for decades and centuries. Based on this result, current population density patterns may correspond roughly to those in the second half of the 19th century and, in many cases, even before then. With this in mind, it is clear that the use of NUTS territorial categories does allow for a rough estimation of actual regional human capital values.

---

[43] These are Belgium, Bulgaria, Czech Republic, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden and the United Kingdom.

[44] Lichtenstein, Norway and Switzerland.

[45] Montenegro, Croatia, FYROM, Turkey and Iceland.

[46] Notable exceptions include Spain and France.

In addition, using the administrative units of the 19th century would not be currently possible, as there is no data set available to us with boundaries at such a detailed subnational level for the whole of Europe. However, employing current NUTS regions has some additional advantages: first, it allows us to make the data more comparable over time because we look at constant and fixed boundaries. Second, it gives us the ability to compare the data from the 19th century with more recent data, in future research.

Furthermore, we take into account the changing national boundaries by attributing today's NUTS regions to national (or, in part, ethnic) boundaries of the 19th century, where necessary. For instance, Poland is defined by its borders within the Russian Empire during the 19th century. For this reason, some current Polish NUTS regions are treated as German regions, as they formed part of the German Empire when the census was taken. We proceed similarly with other regions of this kind.[47]

Consequently, we choose to adapt the territorial administrative divisions of the 19th century as closely as possible to those of the current NUTS classification, although this inevitably leads to some geographical inaccuracies. Thus, data between 1790 and 1880 are available for some countries at the NUTS 3 level and others are available at the NUTS 2 level. To harmonise territorial sizes, we opted for the smaller NUTS 2 level for all regions in our further data analysis.[48] Unfortunately, the NUTS classification is only available for the member states of the European Union and Candidate Countries, as well as EFTA members. For this reason, a somewhat corresponding classification had to be found for other European countries. This concerned, in particular, countries located in east and

---

[47] As a caveat, we should note that we cannot directly compare the historical numeracy and today's numeracy level due to population movements, but this comparison is the purpose of this study.

[48] For mapping purposes, we decided to use the smallest available territorial unit to highlight the regional differences as clearly as possible.

south-east Europe, such as Russia, Ukraine, Belarus and Serbia. In these cases, we take the current territorial administrative division of these countries.

Table 5.2 gives more details on the countries and the regions contained in our data set. For many countries, we have data for the lowest regional classification units, i.e., NUTS 3.[49] Still, this does not necessarily mean that data would not be available for even smaller units. In the case of Austria, the 33 NUTS 3 regions have been formed from more than 70 original counties (‘*Bezirks-Hauptmannschaften*’) in the Austrian Empire located in today’s Austria. The collected data are also similarly detailed for other countries, such as the other parts of the Austrian-Hungarian Empire (NUTS 3) and the United Kingdom and Italy (NUTS 2). Clearly, this wealth of regional data reduces possible biases arising from the use of current administrative borders.

5.6 Results

5.6.1 The development of human capital in the European countries

To obtain a general idea of the distribution of ABCC values for Europe, we have calculated the ABCCs for all European countries between 1790 and 1880 in our data set. The results are depicted in Figure 5.4. Its sole purpose is to show general tendencies and, hence, country labels have been omitted. As our first conclusion, we state that ABCC values differ importantly throughout Europe. Numeracy ranges between 25 and 100 ABCC points.

In our second step, we have divided the European countries into several macroregions, due to the large number of countries under study. These are the core industrial European countries[50], Austria-Hungary, the western European periphery

[49] Or similarly small units for countries not included in the NUTS classification.

[50] Our classification of the core countries corresponds to the one employed by O’Rourke and Williamson, except that it has been extended by also including Luxembourg. See O’Rourke and Williamson (1997).

countries, the east European countries and the south-east European countries.[51] The attribution to one macroregion was effected mainly by reason of geographical location and economic output. Core Europe is constituted by countries such as France, Germany and the United Kingdom. The case for the macroregion Austria-Hungary is self-evident, as it is made up of the regions of the dual monarchy. The western periphery consists of countries from Scandinavia (Denmark, Iceland, Norway) plus Ireland, Italy, Spain and Portugal. The regions of the Russian Empire have been attributed to eastern Europe and, in part, south-east Europe (Armenia, Azerbaijan, Georgia), which also comprises countries from the Balkans. By employing this classification, all macroregions also comprise a similar number of countries.

The details on the development of ABCCs in the various geographical macroregions can be found in Figure 5.5 to Figure 5.9. Because the focus here is on the differences within the macroregions and on a clear visualisation of the ABCC trends, the scale of the ABCC mean values is *very* different in each figure. This always has to be taken into account when interpreting the development of ABCC values. But, it makes the interpretation of trends within the macroregions easier.

Several results can be highlighted: countries from the European 'core' are already characterised by high ABCC values at the beginning of the 19th century (Figure 5.5). The differences between these countries are minor, although France catches up to Belgium and the Netherlands between 1790 and 1820.

The regions of the Austro-Hungarian Empire are more diverse (Figure 5.6). For Cisleithania (today mainly constituted by Austria, the Czech Republic and Slovenia), the

[51] Core Europe is comprised of Belgium, Switzerland, Germany, France, Luxembourg, Netherlands and the United Kingdom, Austria-Hungary of Austria, Czech Republic, Croatia, Hungary, Romania, Slovenia and Slovakia, the Western Periphery of Denmark, Spain, Ireland, Iceland, Italy, Norway and Portugal, Eastern Europe of Belarus, Lithuania, Latvia, Estonia, Moldova, Poland, Russia and Ukraine and South-East Europe of Albania, Armenia, Azerbaijan, Bosnia-Hercegovina, Bulgaria, Cyprus, Georgia, Greece, Macedonia and Serbia.

ABCC values are high and similar to those in the core. The picture is different for Transleithania (today mainly constituted by Hungary, Slovakia, parts of Romania and Croatia[52]); Croatia enjoyed the highest ABCC values initially, followed by Hungary, Slovakia and today's Romanian provinces. Notably, this order stays rather constant over time, except that Romania overtook Slovakia later in the period. The Croatian and Romanian provinces follow similar patterns; this is also true for Hungary and Slovakia. All parts of Transleithania are converging over the time span covered.

Taking a closer look at the western periphery (Figure 5.7), the Scandinavian countries are characterised by very high ABCC values. Italy, Spain and Portugal are evolving rather slowly with values at around an ABCC of 90. Ireland is on a much lower level than other western European countries, but it is progressing.

Most regions of the Russian Empire form the east European countries (Figure 5.8). The range of ABCC values is very striking. It is astonishing that the Baltic States have such varied ABCC levels: Estonia is characterised by ABCC levels on the same level as in the core countries, whereas Latvia and Lithuania follow only after large intervals. Lithuania finds itself with Belarus at the end of the numeracy ladder. Moreover, Poland initially has an important lead over Russia and Ukraine.

Finally, south-east Europe (Figure 5.9) features the lowest numeracy values in our data set. The Caucasus regions (forming part of the Russian Empire) and Albania are the least numerate of all European regions. Numeracy in Bulgaria is somewhat higher and is increasing. Serbia, Bosnia-Herzegovina, Macedonia and Cyprus do better, but the earliest data are already relatively late with respect to the other countries.

[52] Today's Croatia was split between Cisleithania (mostly Dalmatia) and Transleithania (Croatia-Slavonia). Here we refer only to the part belonging to Transleithania.

All in all, we find noticeable differences in numeracy between and within the observed macroregions.

These national results may be compared to other indicators which have been available before. In particular, we have already mentioned Allen's data on signature rates in 1800 (Allen 2009). The overall picture that we get from our numeracy data may be described as being similar to his data, but not always in all respects. Germany is one of the top numeracy countries, whereas it is only in the middle for signature rates.[53] Austria-Hungary also performs better in numeracy than in literacy. In contrast, Poland, Italy and Spain are (or can be estimated from our data to be) on a similar numeracy level as in literacy in the first decades of the 19th century. Moreover, the Netherlands is the clear leader in literacy in Allen's data and our data come to the same conclusion. Belgium and England had similar literacy levels while our data for Belgium and the United Kingdom would suggest a slight advantage for Belgium. This result may, however, be very well driven by our inclusion of Scotland and Wales, which is why these data in the above presentation are not exactly comparable to Allen's. The point of view that France was on a lower level than its advanced neighbours from the Netherlands, Belgium and England is also confirmed by our data. We may add that France was probably a late arriving nation in the top group of numerate countries but progressing very quickly. In sum, our numeracy estimates globally support Allen's evidence, but some important differences remain. Moreover, we are able to provide evidence for most European countries, whereas the data on signature rates are much more limited. This enlarges considerably the geographical scope of previous studies.

[53] However, there might be a bias in signature rates because we are considering all of Germany, whereas signature rates typically come from few locations. Depending on the location this may explain the relatively low literacy values for Germany at that time.

5.6.2 Taking a closer look at the educational differences in Europe

How can we explain these differences in human capital? A first look might be attributed to the possibilities of generating human capital: educational policies and schooling institutions. Because the western European countries and parts of the western periphery have been shown to feature very high numeracy values in general, it appears more important to consider the less numerate countries, particularly in eastern and south-eastern Europe.

Why were the differences so large in the Russian Empire? In particular, the Baltic states appear to be anything but homogeneous. Let us consider these regions in greater detail. Lithuanian education faced important restrictions under Russian rule, e.g., Lithuanian schools and publications in Lithuanian were not allowed (Giedraitiene *et al.* 2007). Similarly, in Belarus, education in the form of schools was mostly limited to populations in cities (Sroka 2007). Because the ABCC differences in the Baltic States are similar to the literacy rates, one may refer to Mironov, who attributes these differences to religion (above all, to Protestantism in the region) (Mironov 1991). The Lutheran Church played an important role in promoting literacy in Estonia (Raun 1986). In this sense, the numeracy results also reflect the heritance of Swedish rule in the country (1561-1721). From the 17th century onwards[54], the ability to read became obligatory for everyone in the Swedish Empire, i.e., mainly today's Sweden, Finland and Estonia (Johansson 2009). This tradition of providing education may also have contributed to the high numeracy values in Estonia.

Moreover, countries formerly belonging to the Ottoman Empire do worse in terms of numeracy than most other regions.[55] This particularly concerns the Balkans.[56] The

[54] Particularly important was the Church Law of 1686.
[55] The most important exception to this trend is the Caucasus region, which has equally little numeracy.

Ottoman education system was not well developed during this period. After the educational lead of the Islamic world in medieval times, the human capital revolution that took place in the west did not happen in the Ottoman Empire.

As an example, we can take a look at Bulgaria which, as we have seen, has one of the lowest ABCC rates in our sample. Bulgaria lost its independence in 1396 and from that time was ruled by the Ottoman Empire until its final *de jure* independence in 1908.[57] Very few educational facilities were available until the 19th century, and these were mostly small cell schools. This also highlights the importance of the Orthodox Church in Bulgaria and in other countries in the Balkans. It "promoted education, which it considered an important means for upholding the Christian faith under the Ottomans" (Daskalova 1996, p. 6). Nevertheless, nothing comparable to a modern education system was established. This is also illustrated by the fact that only 142 cell schools existed in 1762 and this number grew to only 235 by 1835 (Crampton 2007). A widespread, 'nationwide' drive to foster education was created much later than in other countries. Moreover, the teaching that was available was mostly in Greek. However, the intellectual elite of Bulgaria disapproved of the increasing cultural hegemonialism by the Greeks more and more with the passage of time. They wanted to establish schools in which the teachers used their own language. The wealthy mostly sent their children abroad, preferably to Russia, Constantinople or western Europe. Until the liberation, this amounted to about 700 Bulgarians in total (Crampton 2007). These expatriates also had a major influence on the development of the Bulgarian educational system, as they urged the advancement of the system. Thus, moving towards

[56] In particular, this is Serbia (independence fully recognised internationally in 1878), Bosnia-Herzegovina (occupied by the Austro-Hungarian Empire in 1878 and later annexed in 1908), Montenegro (independent in 1878), Albania, Bulgaria (self-government re-established in 1878 (Principality of Bulgaria), unification with Eastern Rumelia in 1885, *de jure* independence in 1908), Romania (independence in 1878), Macedonia (part of Serbia in 1913) and to some extent Greece (already independent in 1830).

[57] The principality of Bulgaria was already *de facto* independent in 1878 and was unified with Eastern Rumelia in 1885.

the use of the Bulgarian language in educational facilities and opening up new lay schools marked a new, second step in Bulgarian development. Finally, Bulgarians advanced from mutual to new class schools, the first being established in Gabrovo in 1840.

Clearly, the struggle for the use of their own language played an important part in the development of Bulgarian education. The first textbook that was written in Bulgarian was published in 1824, but there was hitherto no standardised form of Bulgarian; only in the 1870s did a standard, written language emerge. Another factor promoting educational development came from the increasingly prosperous towns. With rising economic progress and growing financial possibilities, the town councils were the driving forces behind the improved education of the people because they felt the need for better educated workers. Furthermore, they were able to bear some of the costs connected with setting up schools and training teachers (Daskalova 1996). Still, the overall education level of the Bulgarian people remained low during the 19th century although it progressed continuously.

Albania provides a similar picture, as the Albanians had to fight for Albanian-language schools and education, which were banned by the Ottomans (Kostovicova 2005). One issue was the establishment of a written language, and another one was the determination of the type of alphabet to be used (i.e., Latin, Greek or Arabic). As a consequence, the struggle for education in Albanian marked the move towards Albanian nationalism. The way of ruling practised by the Ottomans in south-east Europe might therefore be an important factor in explaining the low educational levels obtained.

Finally, the examples above can also highlight the difficulties and potential biases that may arise when using or approximating literacy rates as long as a written language is not clearly defined by a state, people or region. In contrast, numeracy uses the statements of numbers, which are less dependent on language formation, particularly in south-east Europe.

5.6.3 Regional differences at the European level

To get insight into the regional disparities between European regions, Figure 5.10 to Figure 5.17 depict the regional ABCC values between 1800 and 1870.[58] Some supplementary information is revealed by the maps. For example, there is a general north-south difference in Italy: northern regions feature higher ABCC values than southern regions. This is not unexpected, as differences between the south and the north of Italy have prevailed until today. A similar tendency can be seen in France, even though there is not such a clear divide as in Italy. Nevertheless, the north of France has a lead on the south (albeit with some exceptions), a characteristic also apparent in the literacy data (Furet and Ozouf 1977).

In contrast, the reverse geographic case can be found in Norway. Here, the distant and less populated northern provinces, particularly Finnmark, are not at the same level as the southern ones. Moreover, these regions were not only far away but also sparsely populated. Schools could only be provided at great cost because pupils had to walk long distances and urban teachers were typically not willing to work there. For these reasons, schooling had already been delayed with respect to the other provinces at the end of the 18th century (Guttormsson 1990). An analogous observation can be made for the first birth decades observed in Great Britain, where in particular, the Highlands of Scotland have lower ABCC values than the Lowlands and some parts of England. This is interesting because previous historians of education saw Scotland as a pioneer in Europe (e.g., Lockridge 1974). Still, this generally refers to the success of educational campaigns in the Scottish Lowlands. By contrast, the Highlands possessed “a primitive economic and social structure, few schools, and poor communications” (Houston 1987, p. 60), augmented by a

[58] In order to discern as many regional differences as possible, the lowest available administrative classification (NUTS 3 or NUTS 2) has been selected in the cartographic representation.

language barrier due to the wide diffusion of Gaelic and adverse geographical conditions. These factors also led to lower literacy attainment rates than in the Lowlands in the first half of the 19th century. The ABCC values suggest a similar regional disparity between these regions.[59]

In Spain, a core-periphery pattern is discernible. The regions in southern Andalusia and north-western Galicia are characterised by low numeracy. The area in central-northern Spain is the most advanced in the country. One might initially expect north-eastern Catalonia, and particularly the industrial region of Barcelona, to be among the leading regions within Spain. However, this was not the case. Neither in numeracy nor in literacy was Barcelona leading, as can be derived from figures published by Núñez (Núñez 1992).

In the Russian Empire, we highlight the case of the city of Odessa.[60] Its high level of numeracy is in line with literacy figures, with Odessa on a similar level as the two major cities of the Russian Empire, namely, St. Petersburg and Moscow.[61] Odessa itself is located on the north-west shore of the Black Sea and featured (until today) a major seaport. Either the economic success engendered a need for trained (and also literate and numerate) workers or the remarkable numeracy (especially of the Jewish minority) enabled the harbour city to grow to a commercial centre. Also, the peasants were surprisingly apt in calculating (Goodwin and Grennes 1998). Moreover, in the Polish provinces, the south-western and eventually the central provinces around Warsaw have a lead on the eastern ones.

[59] Nevertheless, the reasons why the Lowlands of Scotland do not generally show higher numeracy rates than many regions in England still have to be explored further.

[60] The numeracy of the surrounding region is shown in the maps; the numeracy of the city itself is still higher throughout time. We refer here to the numeracy of the city.

[61] Odessa had higher literacy rates than St. Petersburg but lower rates than Moscow. See Herlihy (1986).

Finally, numeracy is very high in the regions of the German Empire and the German speaking parts of Austria-Hungary. With regard to the German Empire, the regions with the lowest ABCC values are found in West Prussia, Upper Silesia and partly East Prussia.[62]

In general, throughout Europe, it appears that the most backward regions within countries, as judged by numeracy, are often found in peripheral geographic locations, which are characterised by a lack of infrastructure and less economic power. However, more research is still needed to clarify the determinants of regional human capital formation on a European scale.

5.6.4 Inequalities of regional human capital distribution

Based on these first cartographic impressions, can we statistically observe regional differences in the distribution of human capital within countries? To answer this question, we measure regional inequality by using the coefficient of variation (CV). Using the CV has the advantages that it is a dimensionless number and that it allows comparisons between the different countries, even though they have different means. It is defined as the standard deviation of regional ABCC values of a country (σ) divided by the average ABCC value of a country (μ), multiplied by 100:

$$CV = \frac{\sigma}{\mu} \times 100. \tag{5.3}$$

We proceed similarly to the description of ABCC means.[63] However, we exclude those countries that had nearly solved their basic numeracy problem because the proximity to an ABCC value of 100 would bias the CVs. This applies to core and Scandinavian

[62] Still, one has to take into consideration the already high level of attained numeracy in Prussia. For this reason, the differences are not very large.

[63] Not all countries are included for the calculation of the ABCC means since only data on the national level is available for the smaller countries (e.g., Luxembourg, Iceland, Estonia, Cyprus) and thus regional inequalities cannot be measured.

countries. Hence, we avoid potential 'bounded variable' problems. Thus, Figure 5.18 to Figure 5.21 highlight the outcomes for the macroregions. As in the case of the ABCC means, the different scales have to be taken into consideration.

In general, regional variation is shrinking throughout time in most countries. Considerable differences in the CV can initially be found in Transleithania (eastern Austria-Hungary), Ireland, Italy and Spain. However, they decrease over the decades. The paths of the different countries are different and far from a linear, straight improvement in numeracy over time. Still, nationwide homogeneity is increasingly being attained within these countries. Numeracy in east European regions within their modern frontiers is still widely dispersed, but it is also diminishing. European Russia is characterised by the highest regional variation. Given the vast territory covered by this country, even when only considering its European part, this is not a very surprising result. A particular case is Serbia. Serbia already includes the region of Vojvodina[64], which belonged to Austria-Hungary but was united with Serbia after World War I. Thus, it benefited from the educational infrastructure of the Kingdom of Hungary. Accordingly, this northern region is characterised by considerably higher ABCC values than the rest of Serbia. This has led to a very high coefficient of variation and it may also highlight the persistence of regional human capital patterns over time. The only country with widening regional disparities in our sample is Bulgaria. In Bulgaria, the southern regions advance much more with respect to ABCC values than do their northern counterparts, resulting in an increasing south-north gap. This result may also be the outcome of the different timing of independence in the north and the south. Moreover, the northern regions could have benefited from the proximity to the Danube, a major trade route even today. Still, Bulgaria was geographically and ethnically diverse. These differences may have contributed to the overall divergence.

[64] The underlying census was the first undertaken in Yugoslavia in 1921.

In summary, one can generally state that a reduction of basic numeracy differences took place within and among most European countries in the 19th century.

5.7 Conclusion

This paper has examined the long-term development of human capital in more than 570 regions in Europe between 1790 and 1880. We have used the age heaping method to approximate human capital values and the NUTS classification to categorise regions according to current national borders. Even though both methods have their limitations and are prone to some possible biases, this has enabled us to estimate, for the first time, the levels of human capital for most European regions in the 19th century and to calculate regional inequalities over time.

Due to the large number of regions and countries under study, we have divided the European countries into five macroregions: core industrial European countries, Austria-Hungary, western European periphery countries, east European countries and south-east European countries. Core western and central European countries enjoyed high numeracy levels throughout the period. Low and medium levels of numeracy were dominant in periphery countries, particularly in eastern and south-eastern Europe. However, many of these countries were solving their basic numeracy problem by the middle of the 19th century.

Regional inequalities, as measured by the coefficient of variation, were also important in many countries. As the population of a country improves its numeric abilities, regional differences become less striking. Still, the persisting inequalities in many countries highlight the importance of our regional approach. In fact, working with data on the national level does, in part, conceal major differences within countries. For this reason, future research should focus on the human capital distribution at the regional level and on the factors explaining why some regions perform better than others.

5.8 Appendix

5.8.1 Data

Country	Census year	Source
Albania	1918	Preliminary dataset "Albanische Volkszählung von 1918", entstanden an der Karl-Franzens-Universität Graz unter Mitarbeit von Helmut Eberhart, Karl Kaser, Siegfried Gruber, Gentiana Kera, Enriketa Papa-Pandelejmoni und finanziert durch Mittel des Österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Special thanks to Siegfried Gruber for providing the data.
Austria (Cisleithania)	1880	K. K. Statistische Central-Commission (1882). Oesterreichische Statistik. Ergebnisse der Volkszählung und der mit derselben verbundenen Zählung der häuslichen Nutzthiere vom 31. December 1880, 2. Band, 1. Heft [Austrian Statistics. Results of the census and of the count of domestic working animals on December, 31st 1880, Vol. 2, number 1], Wien: Kaiserlich-königliche Hof- und Staatsdruckerei.
Belgium	1846	Statistique de la Belgique (1849). Population. Recensement General. 15 octobre 1846 [Population. General census on October, 15th 1846], Bruxelles: Imprimerie de Th. Lesigne.
Bulgaria	1893	Statistique de la principauté de Bulgarie (1893). Résultats du recensement pour la Principauté de Bulgarie au 1er janvier 1893 [Results of the census of the Principality of Bulgaria on January, 1st 1893], Sofia: Imprimerie de l'Etat.
Cyprus	1946	Government of Cyprus (1946). Census of Population and Agriculture 1946, Nikosia: Government Printing Office.
Denmark	1845, 1801/1803, 1787	Dansk Demografisk Database, Kildeindtastningsprojekte og Dansk Data Arkiv, Folketaellingen 1845, CD-ROM; Dansk Demografisk Database, Kildeindtastningsprojektet og Dansk Data Arkiv, Folketaellingen 1801 og 1803, CD-ROM; Dansk Demografisk Database, Kildeindtastningsprojektet og Dansk Data Arkiv, Folketaellingen 1787, CD-ROM.
France	1851	Statistique Générale de la France (1855). Statistique de la France, Tome II, 2e série, Territoire et Population, 1851 [Statistics of France, Vol. II, 2nd series, Territory and Population, 1851], Paris: Imprimerie impériale.
German Empire (incl. Luxembourg)	1880	Statistik des Deutschen Reichs (1883). Volkszählung im Deutschen Reich am 1. Dezember 1880, Band LVII [Census in the German Empire on December, 1st 1880, Vol. LVII], Berlin : Verlag von Puttkammer &

		Mühlbrecht.
Greece	1907	Royaume de Grèce (1909). Résultats statistiques du recensement général de la population effectué le 27 octobre 1907, Tome I [Statistical results of the general census of population on October, 27th 1907, Vol. I], Athènes: Imprimerie nationale.
Hungary (Trans-leithania)	1869	Königliches Ungarisches Statistisches Bureau (1871). Ergebnisse der in den Ländern der ungarischen Krone am Anfange des Jahres 1870 vollzogenen Volkszählung sammt Nachweisung der nutzbaren Hausthiere [Results of the census in the countries of the Hungarian crown at the beginning of 1870, including the results for the domestic working animals], Pest: Druck des Athenaeum.
Iceland	1901	Rothenbacher, F. (2002). The European population 1850 –1945, Basingstoke: Palgrave Macmillan.
Ireland	1841, 1851, 1861	Census of Ireland (1843). Report of the commissioners appointed to take the census of Ireland, for the year 1841, Dublin: Alexander Thom; (1855). The Census Of Ireland for the year 1851, Part IV, Report on Ages and Education, Dublin: Alexander Thom and Sons; (1863). The census of Ireland for the year 1861. Part II. Report and tables on ages and education. Vol. I., Dublin: Alexander Thom and Sons; (1863). The census of Ireland for the year 1861. Part II. Report and tables on ages and education. Vol. II., Dublin: Alexander Thom and Sons; Database of Irish Historical Statistics : Literacy, 1841-1911, UK Data Archive, no. 3582; Database of Irish Historical Statistics : Age, 1821-1911, UK Data Archive, no. 3574.
Italy	1871	Statistica del Regno d'Italia (1874). Censimento 31 Dicembre 1871. Popolazione per età , sesso, stato civile ed istruzione, Vol. II [Census on December, 31st 1871. Population according to age, sex, civil status and instruction, Vol. II], Roma: Tipografia Cenniniana.
Netherlands	1846	Volkstellingen 1795-1971, online, last accessed 12 June 2012, http://www.volkstellingen.nl/nl/volkstelling/jaarview/1849/.
Norway	1865	Minnesota Population Center (2008). North Atlantic Population Project: Complete Count Microdata. Version 2.0 [Machine-readable database]. Minneapolis: Minnesota Population Center. Dataset: The Digital Archive (The National Archive), Norwegian Historical Data Centre (University of Tromsø) and the Minnesota Population Center (2008). National Sample of the 1865 Census of Norway, Version 2.0. Tromsø, Norway: University of Tromsø.
Portugal	1940	Instituto Nacional de Estatística (1945). VIII

		Recensamento Geral Da Populacao no Continente e Ilhas Adjacentes em 12 de Dezembro de 1940, Vol. I [8th General census of population of the Continent and the adjacent islands on December 12th, 1940, Vol. I], Lisboa: Imprensa Nacional de Lisboa.
Russian Empire	1897	издание центрального статистического комитета министерства внутренних (1899), первая всеобщая. перепись населения, российской империи, 1897 г. [The first universal population census of the Russian Empire, 1897], с.-петербург.
Serbia and Yugoslavia	1895, 1921	Servia (1899). Dénombrement de la Population dans le Royaume de Serbie le 31. Décembre 1895. Deuxième Partie [Census of the population in the Kingdom of Serbia on December, 31st 1895. 2nd part], Belgrade: Imprimerie de l'Etat du Royaume de Serbie; Kraljevina jugoslavija (1932). Opsta drzavna statistika. Definitivni rezultati popisa stanovnistva od 31 januara 1921 god [General government statistics. Final results of the national census on January, 31st 1921], Sarajevo: Drzavna stamparija.
Spain	1900	Dirección general del Instituto geográfico y estadístico (1903). Censo de la Poblacion de Espana, según el Empadronamiento hecho en la Península é Islas adyacentes en 31 de diciembre de 1900 (1907). Tomo III. Clasificación de los habitantes por su edad, combinada con sexo, estado civil e instrucción elemental [Census of the Population of Spain, according to the census lists for the Peninsula and the adjacent Islands on December, 31st 1900 (1907). Tomb III. Classification of the inhabitants according to age, combined with sex, civil status and elemental education], Madrid: Imprenta de la Dirección general del Instituto geográfico y estadístico.
Switzerland	1870	Statistisches Bureau des eidgenössischen Departement des Innern (1874). Schweizerische Statistik. Eidgenössische Volkszählung vom 1. December 1870. Zweiter Band. Die Bevölkerung nach Alter, Geschlecht und Familienstand [Swiss Statistics. Census on December, 1st 1870. Vol. II. The populaction according to age, sex and civil status], Zürich: Orell, Füssli & Cie.
United Kingdom	1881	Minnesota Population Center (2008). North Atlantic Population Project: Complete Count Microdata. Version 2.0 [Machine-readable database]. Minneapolis: Minnesota Population Center. Dataset: K. Schürer and M. Woollard (2003). National Sample from the 1881 Census of Great Britain [computer file], Colchester, Essex: History Data Service, UK Data Archive [distributor].

5.8.2 Tables

Table 5.1 Time span of covered countries

	Birth decade											
	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890
Cyprus										x	x	x
Portugal										x	x	x
Greece										x		
Albania								x	x	x	x	x
Spain						x	x	x	x	x		
Iceland						x	x	x	x	x		
Bulgaria					x	x	x	x	x			
Russian Empire					x	x	x	x	x			
Austria (Cisleithania)					x	x	x	x				
German Empire				x	x	x	x	x				
Luxembourg				x	x	x	x	x				
UK				x	x	x	x	x				
Serbia (and Yugoslavia)			x	x	x	x	x	x	x	x	x	
Hungary (Transleithania)			x	x	x	x						
Italy			x	x	x	x						
Switzerland			x	x	x	x						
Norway		x	x	x	x	x						
Ireland	x	x	x	x	x	x						
France	x	x	x	x	x							
Netherlands	x	x	x	x	x							
Belgium	x	x	x	x								
Denmark	x	x	x	x								

Note: Grey-shaded are birth-decades covered with census evidence.

Table 5.2 Regional classification units

Code	Country	NUTS 3	NUTS 2	Non-NUTS
AL	Albania			1
AM	Armenia			1
AT	Austria	33		
AZ	Azerbaijan			1
BA	Bosnia-Herzegovina			1
BE	Belgium		11	
BG	Bulgaria		5	
BY	Belarus			4
CH	Switzerland	25		
CY	Cyprus	1		
CZ	Czech Republic	14		
DE	Germany		46	
DK	Denmark	10		
EE	Estonia		1	
ES	Spain	49		
FR	France	85		
GE	Georgia			1
GR	Greece	24		
HR	Croatia	11		
HU	Hungary	19		
IE	Ireland		2	
IS	Iceland		1	
IT	Italy		22	
LT	Lithuania		1	
LU	Luxembourg	1		
LV	Latvia		1	
MD	Moldova			1
MK	FYROM			1
NL	Netherlands		11	
NO	Norway	19		
PL	Poland		7	
PT	Portugal		7	
RO	Romania	16		
RS	Serbia			3
RU	Russia			34
SI	Slovenia	11		
SK	Slovakia	7		
UA	Ukraine			15
UK	United Kingdom		32	

Note: Always the lowest available administrative division is listed. "Non-NUTS" refers to countries which are not in the NUTS classification. Serbia includes Vojvodina, Montenegro and Kosovo. Russia includes only European Russia.

5.8.3 Figures

Figure 5.1 Age heaping in the census of Bulgaria (1893)

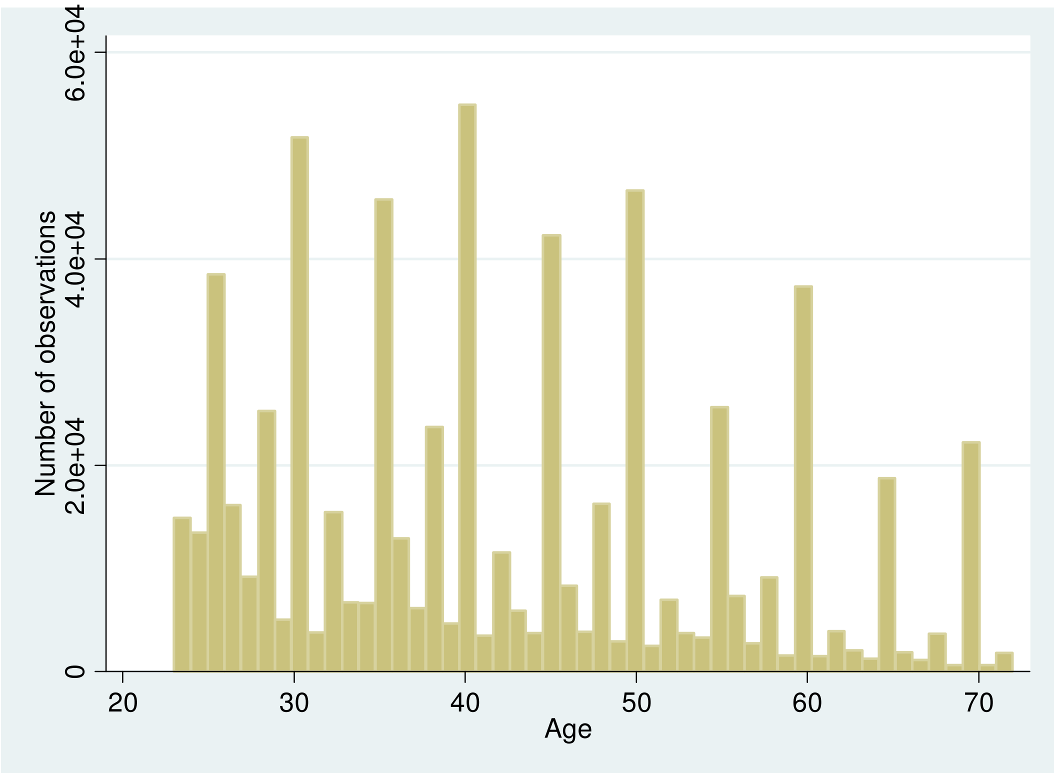
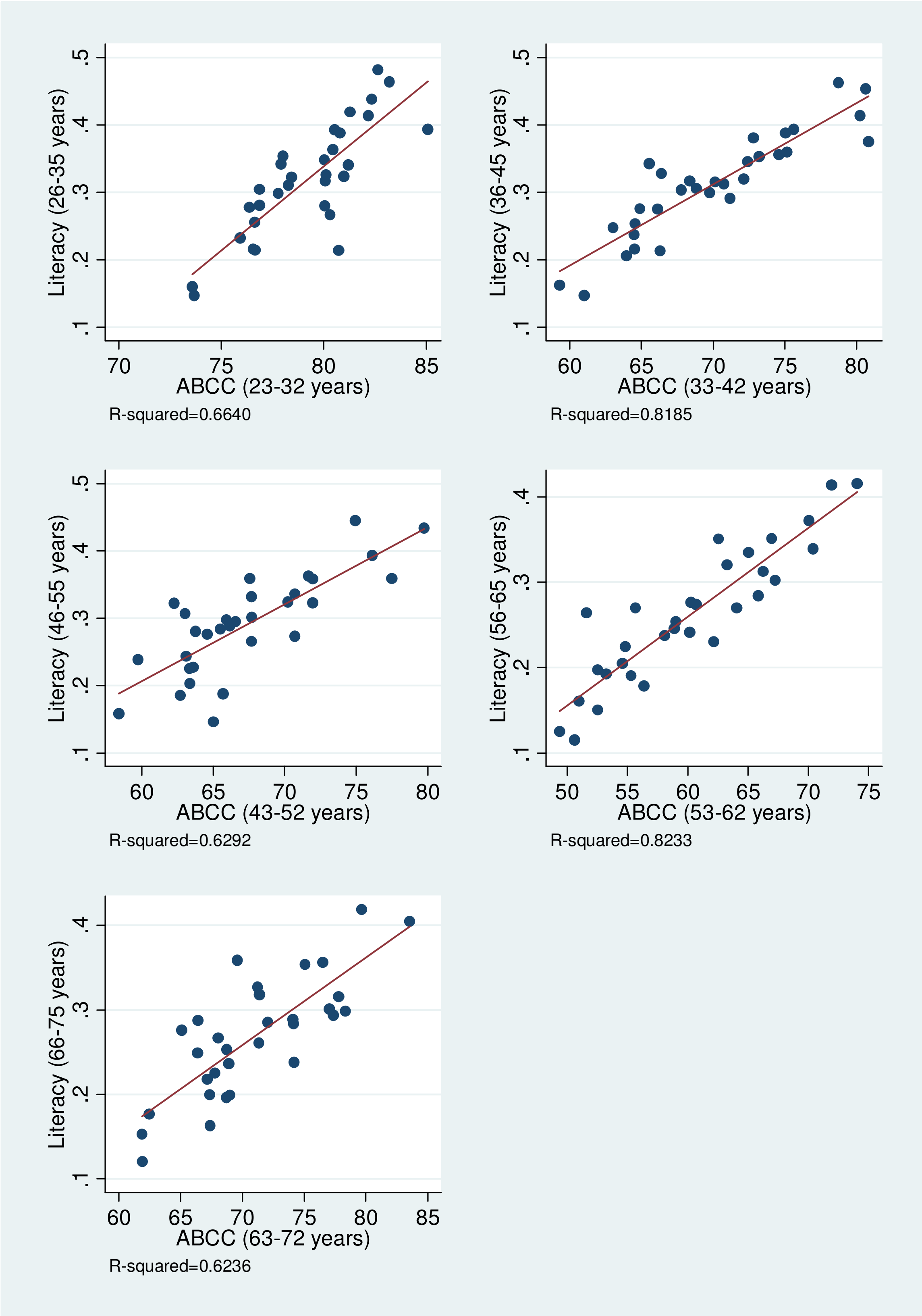
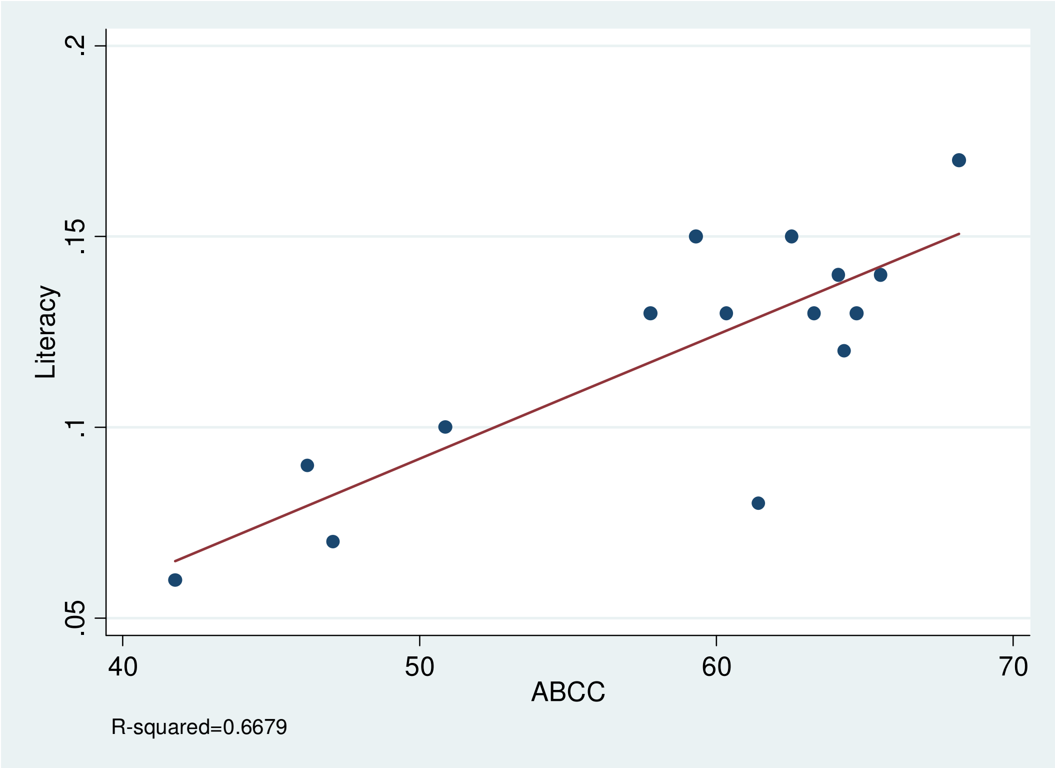


Figure 5.2 Regional relationship of age heaping and literacy in Ireland 1841



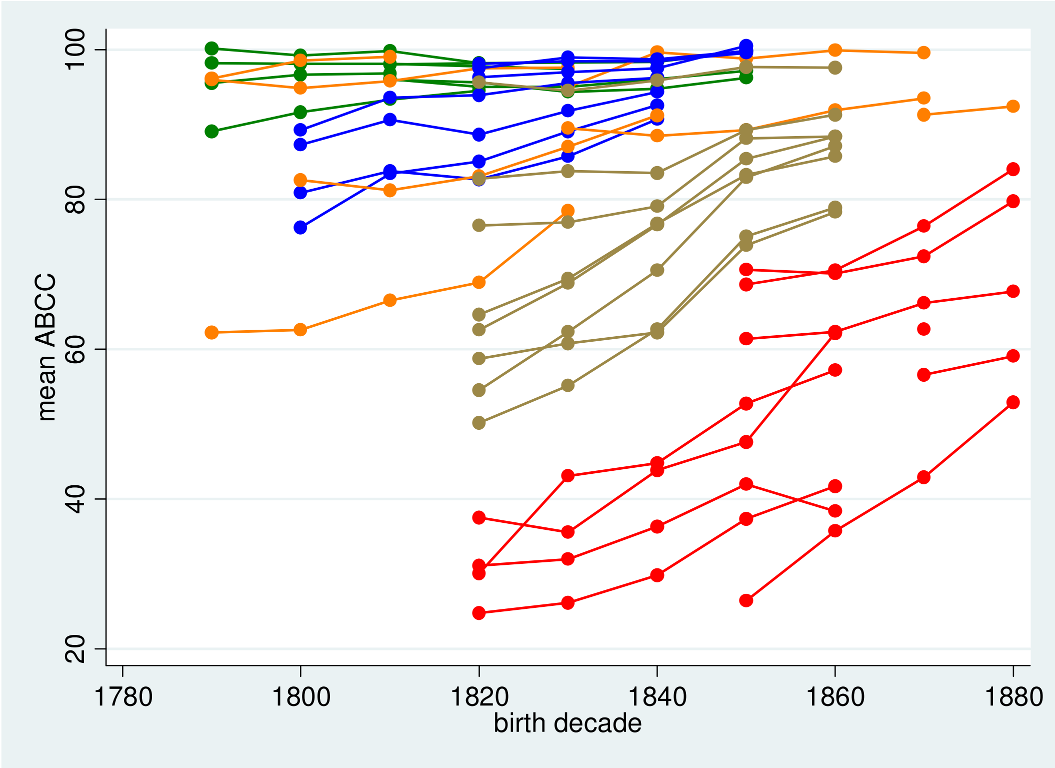
Note: Literacy is defined here as the ability to read and write. County data have been used. The region around the capital, i.e., county Dublin, is an outlier and thus has been excluded.

Figure 5.3 Relationship of age heaping and literacy in Serbia 1895



Note: Literacy refers to the ability to read and write. The ABCC is calculated by merging all individual years between 23 and 72 years old since literacy data are not available for individual age groups.

Figure 5.4 Mean ABCC values for all European countries



Note: Colours correspond to macroregions. Green = Core Industrial Europe, blue = Austria-Hungary, orange = western Periphery, brown = eastern Europe, red = south-east Europe.

Figure 5.5 ABCC mean of Core European countries

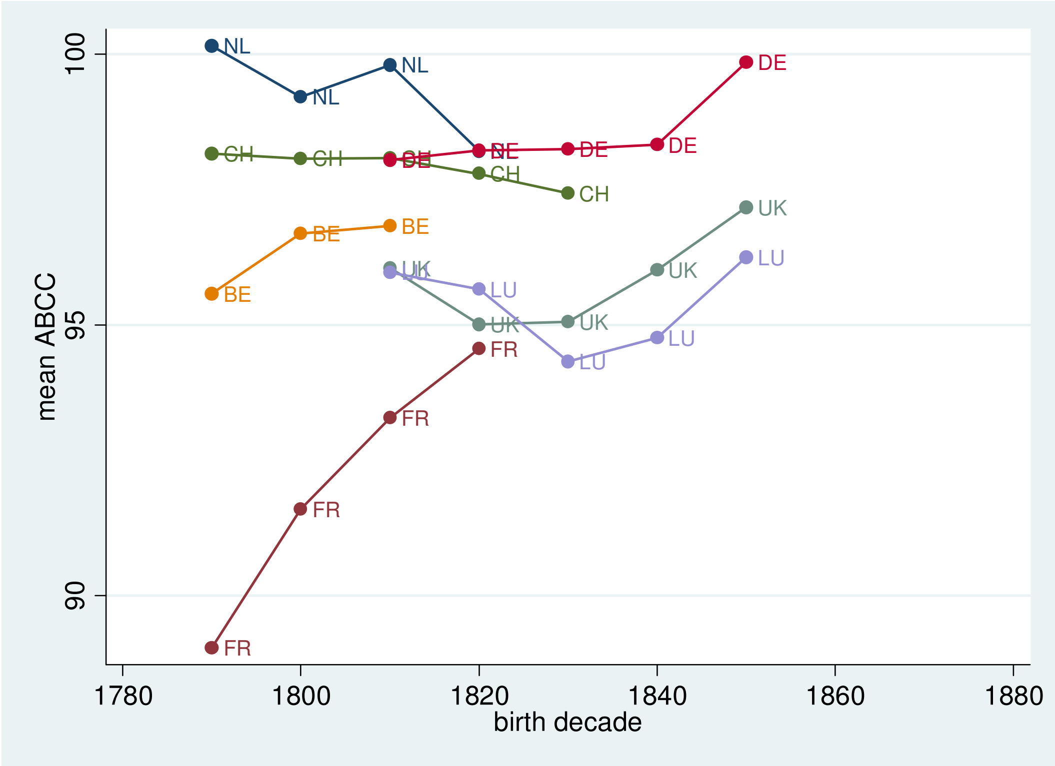


Figure 5.6 ABCC mean of the regions of Austria-Hungary

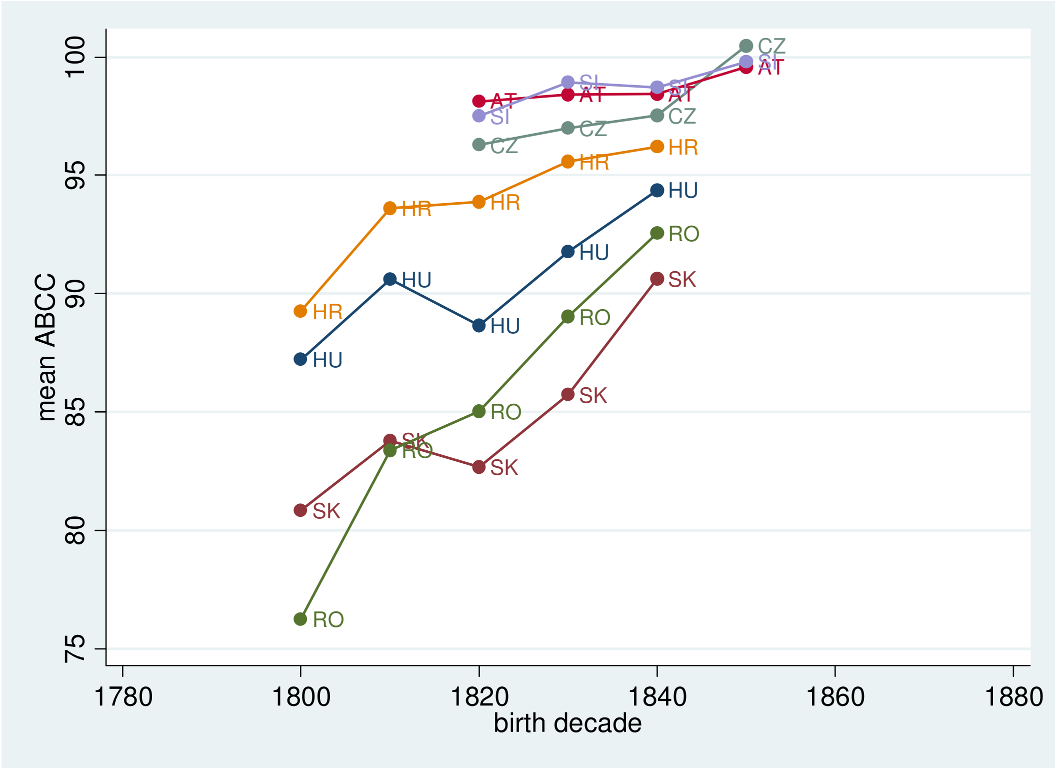


Figure 5.7 ABCC mean of the western periphery

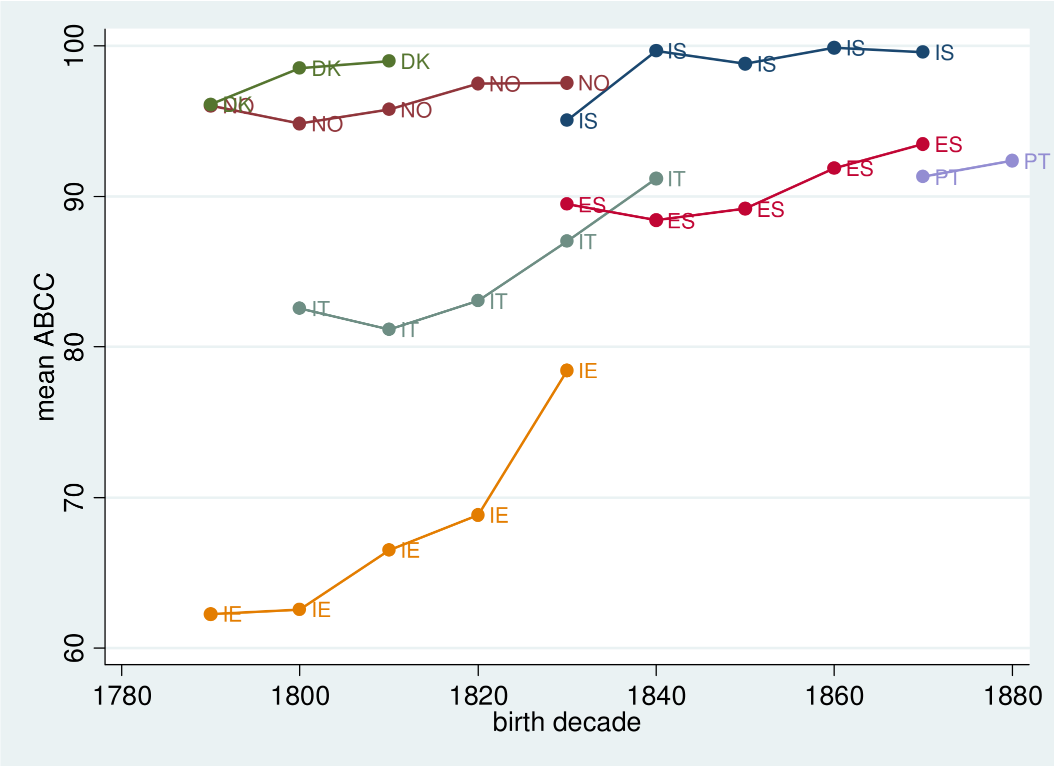


Figure 5.8 ABCC mean of east European countries

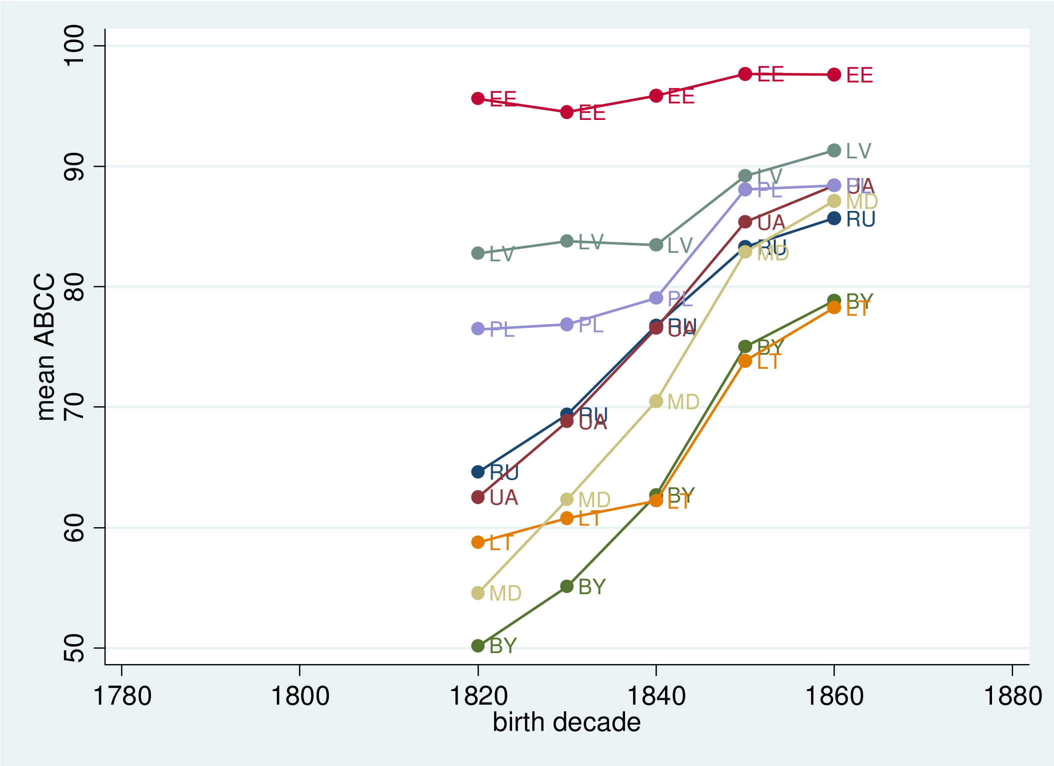


Figure 5.9 ABCC mean of south-east European countries

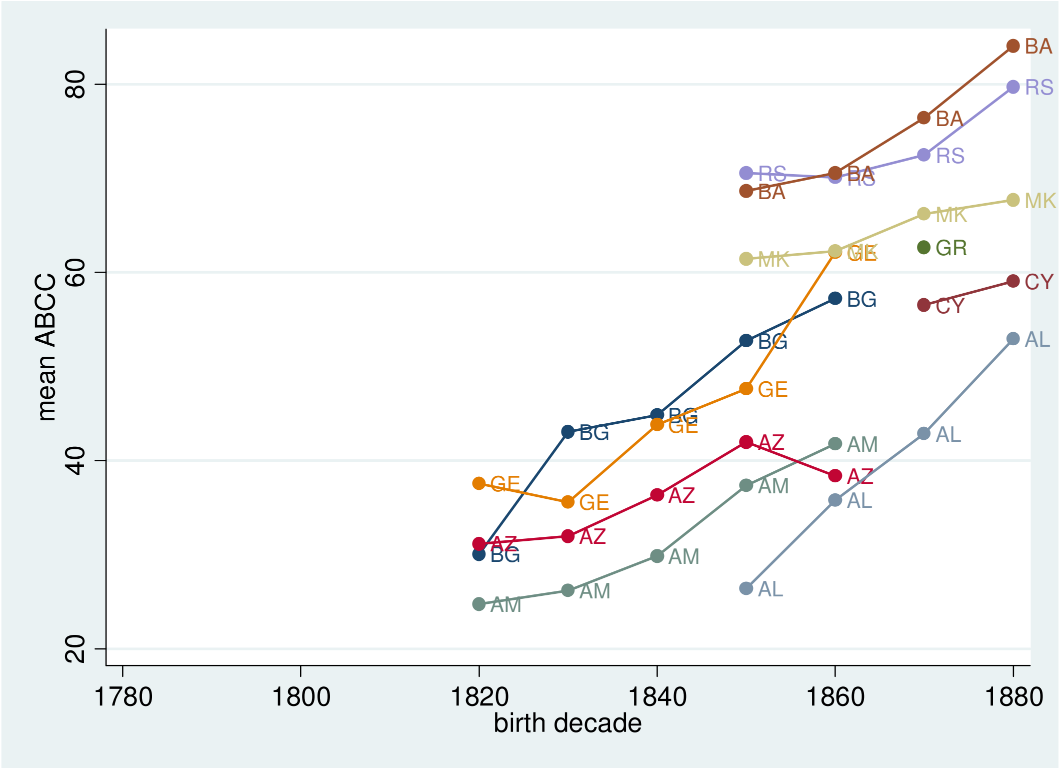


Figure 5.10 Regional ABCC differences in 1800

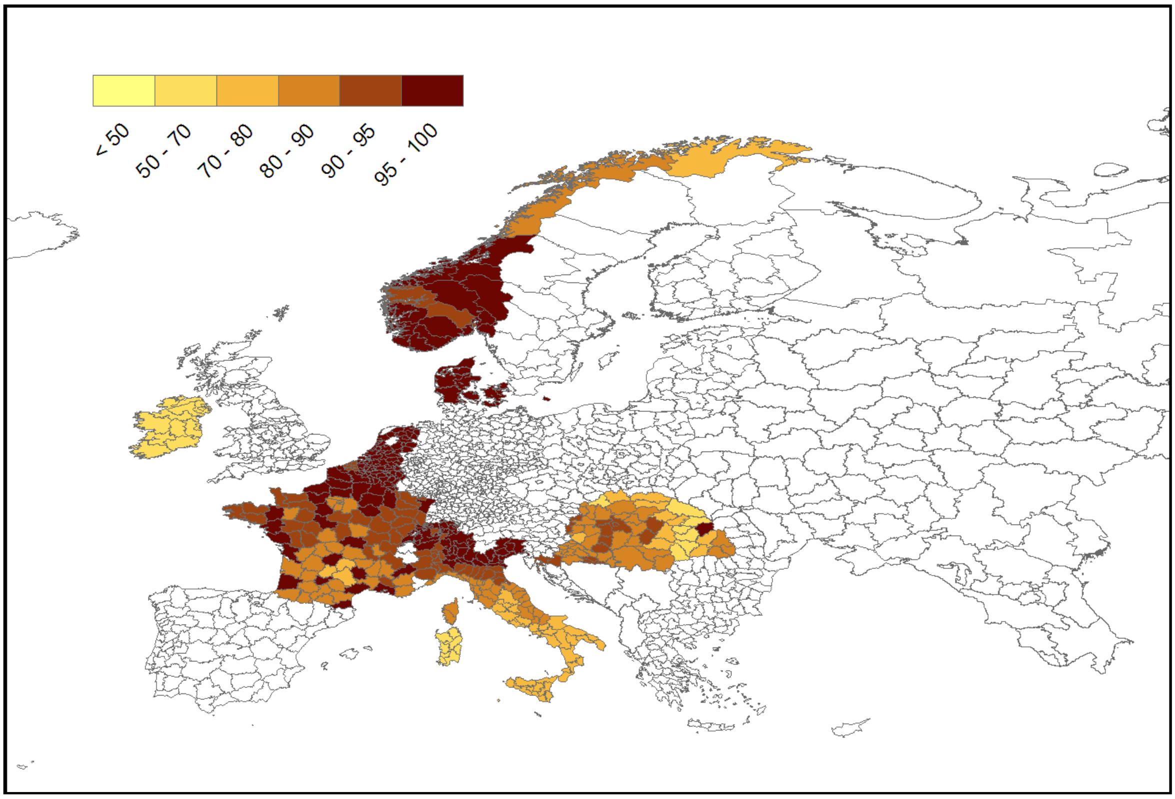


Figure 5.11 Regional ABCC differences in 1810

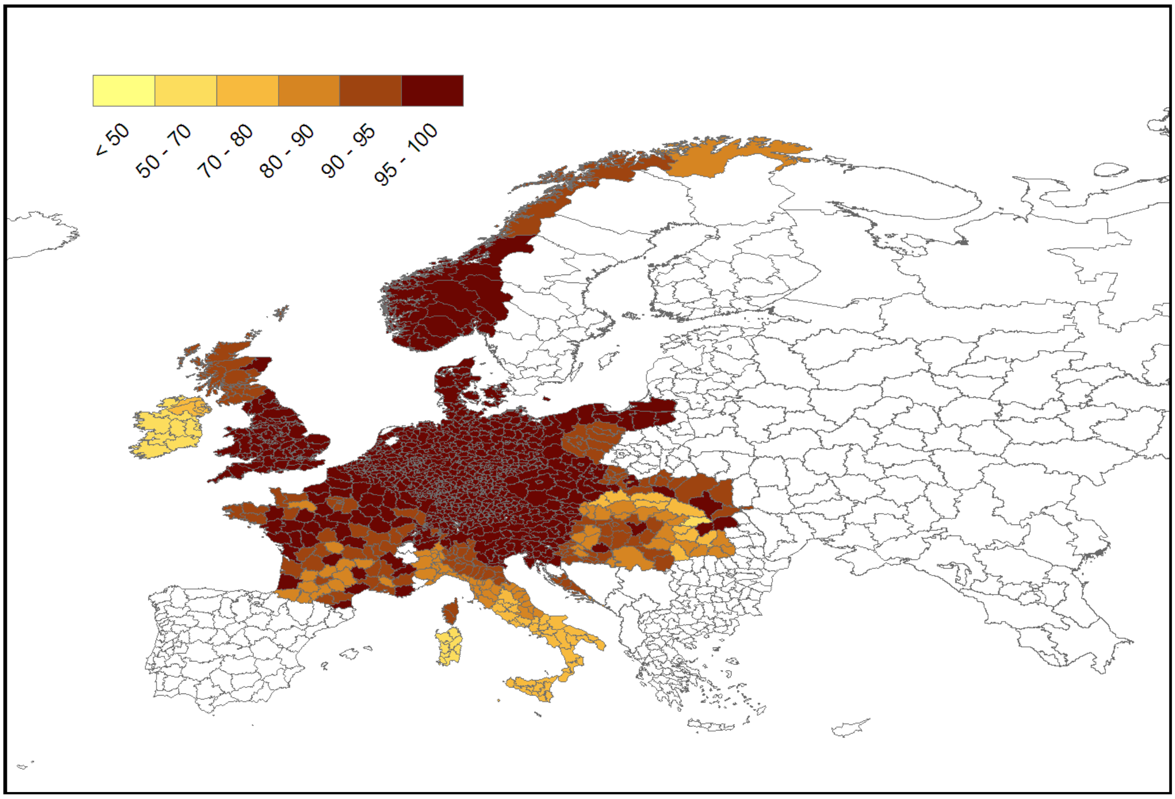


Figure 5.12 Regional ABCC differences in 1820

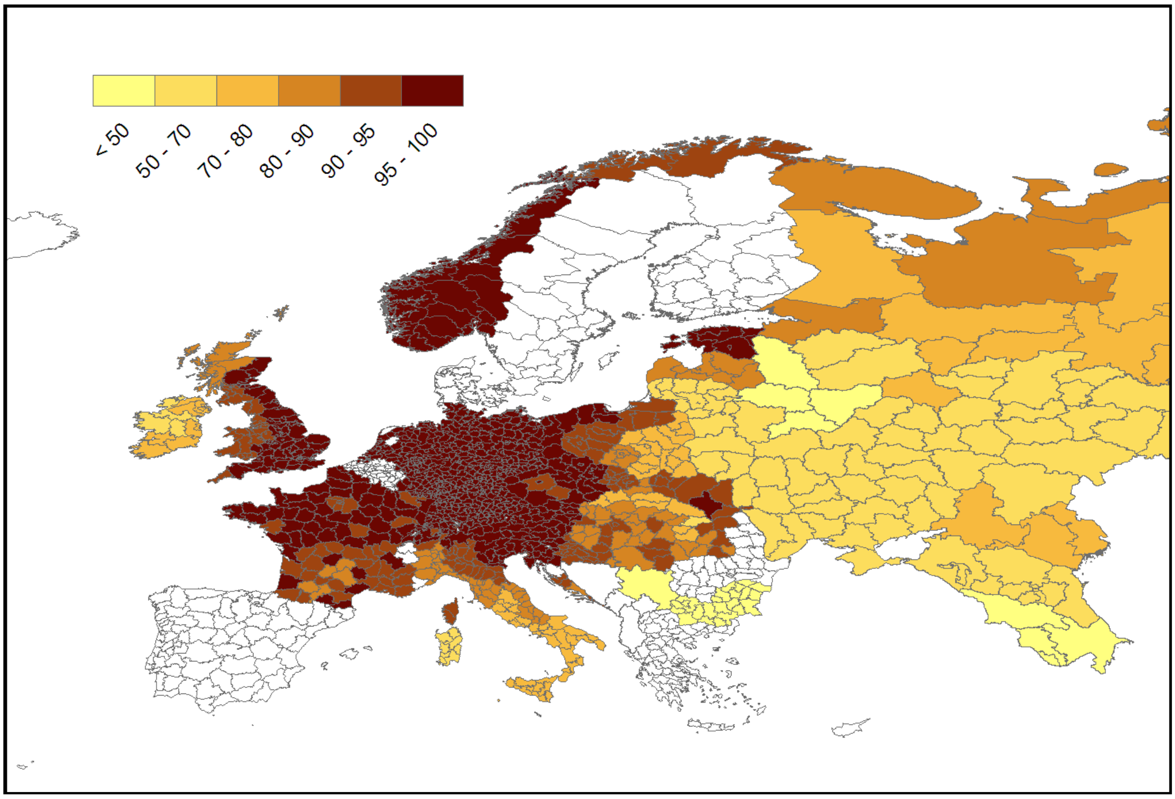


Figure 5.13 Regional ABCC differences in 1830

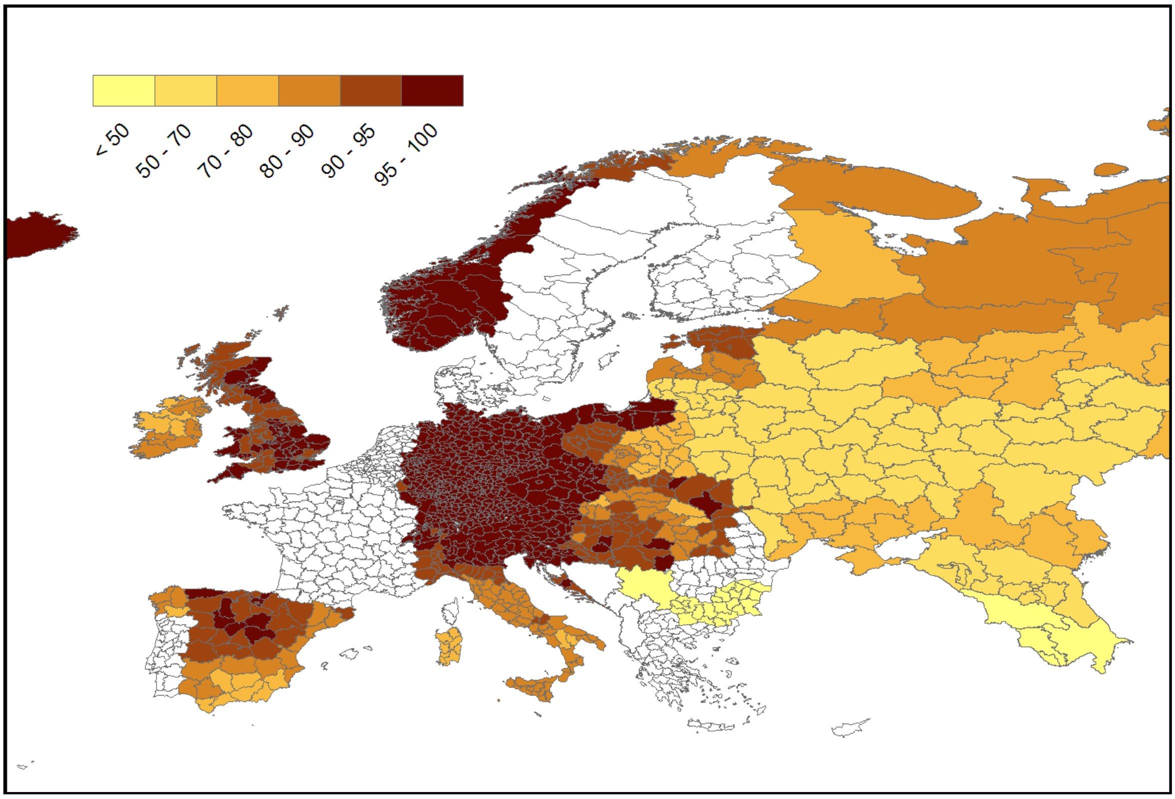


Figure 5.14 Regional ABCC differences in 1840

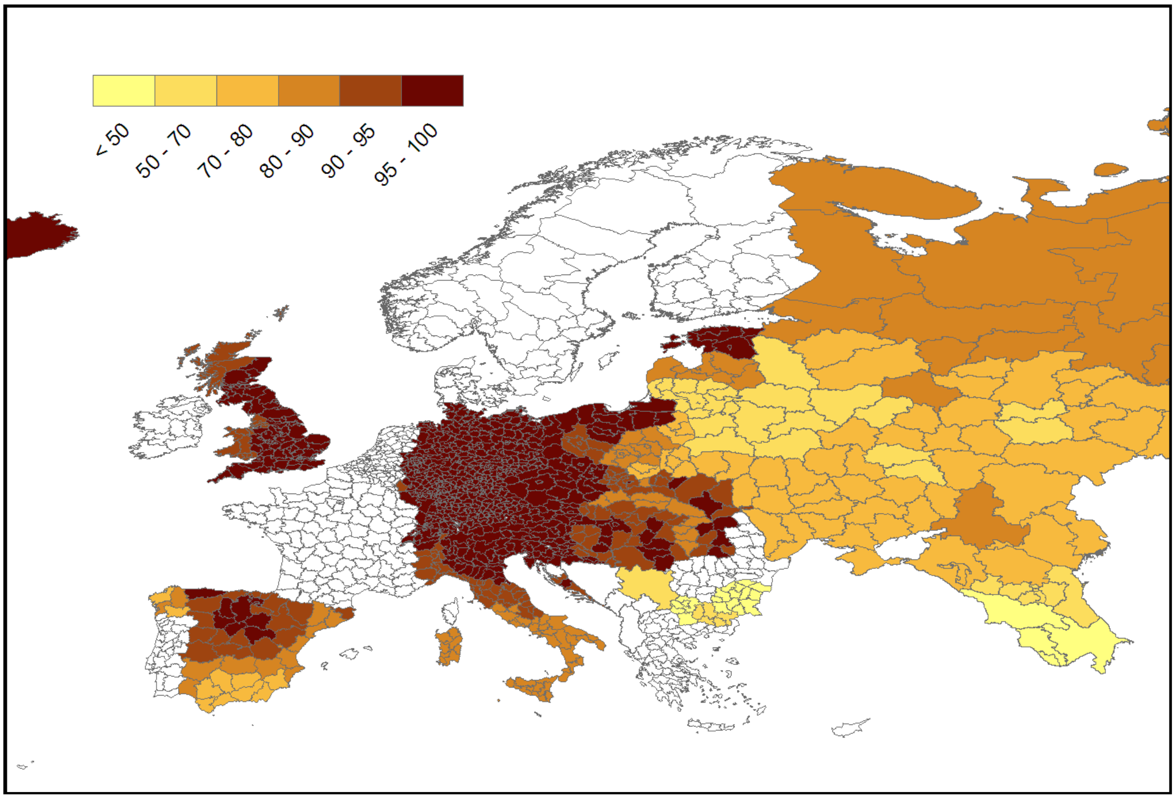


Figure 5.15 Regional ABCC differences in 1850

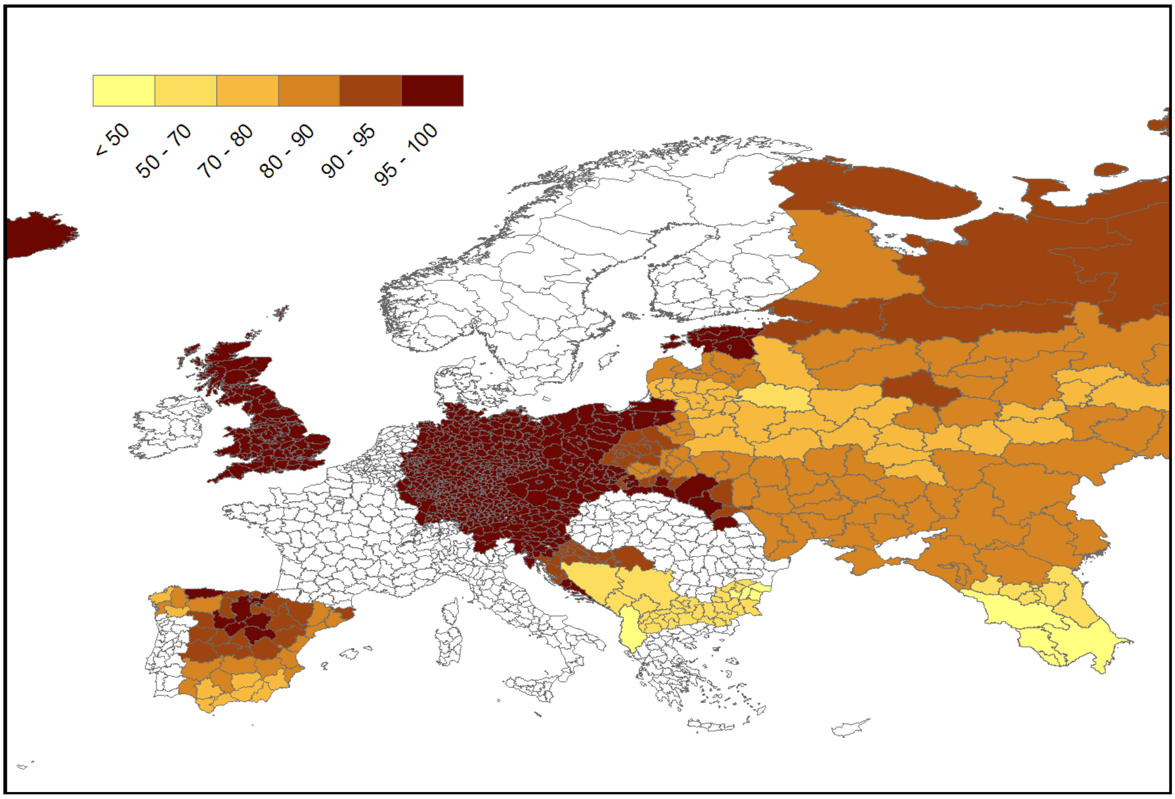


Figure 5.16 Regional ABCC differences in 1860

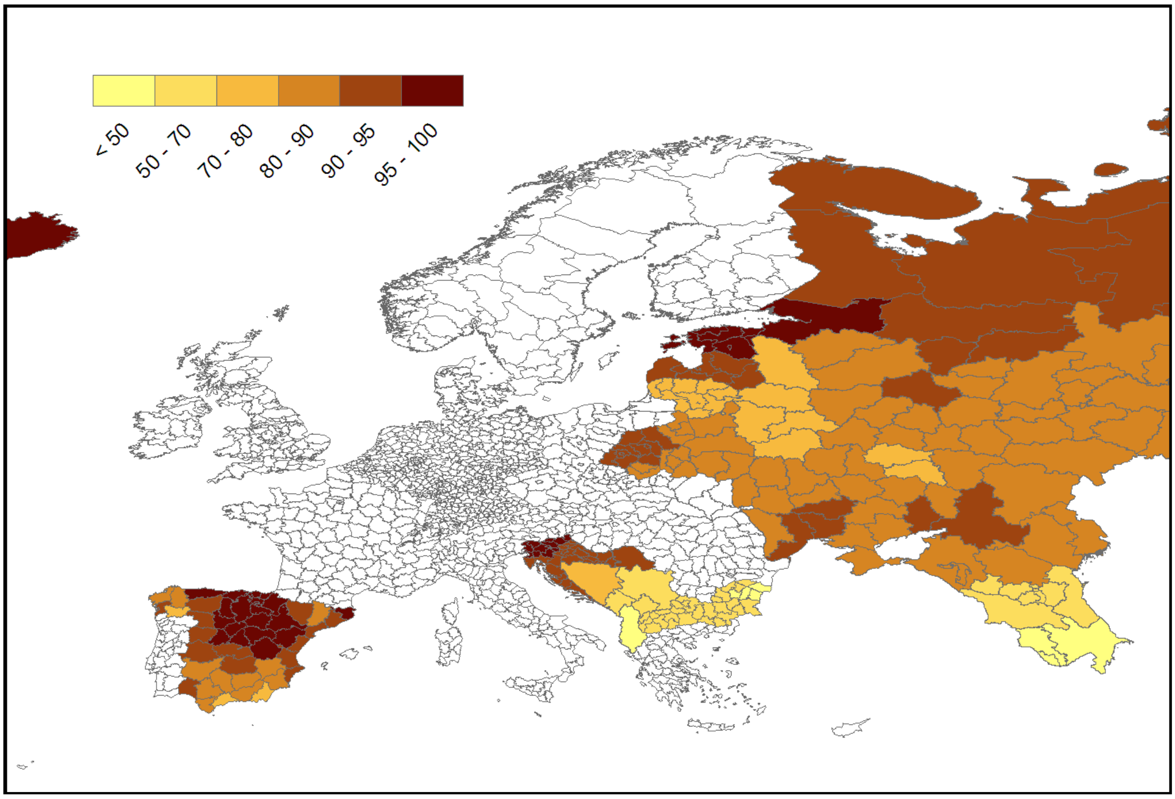


Figure 5.17 Regional ABCC differences in 1870

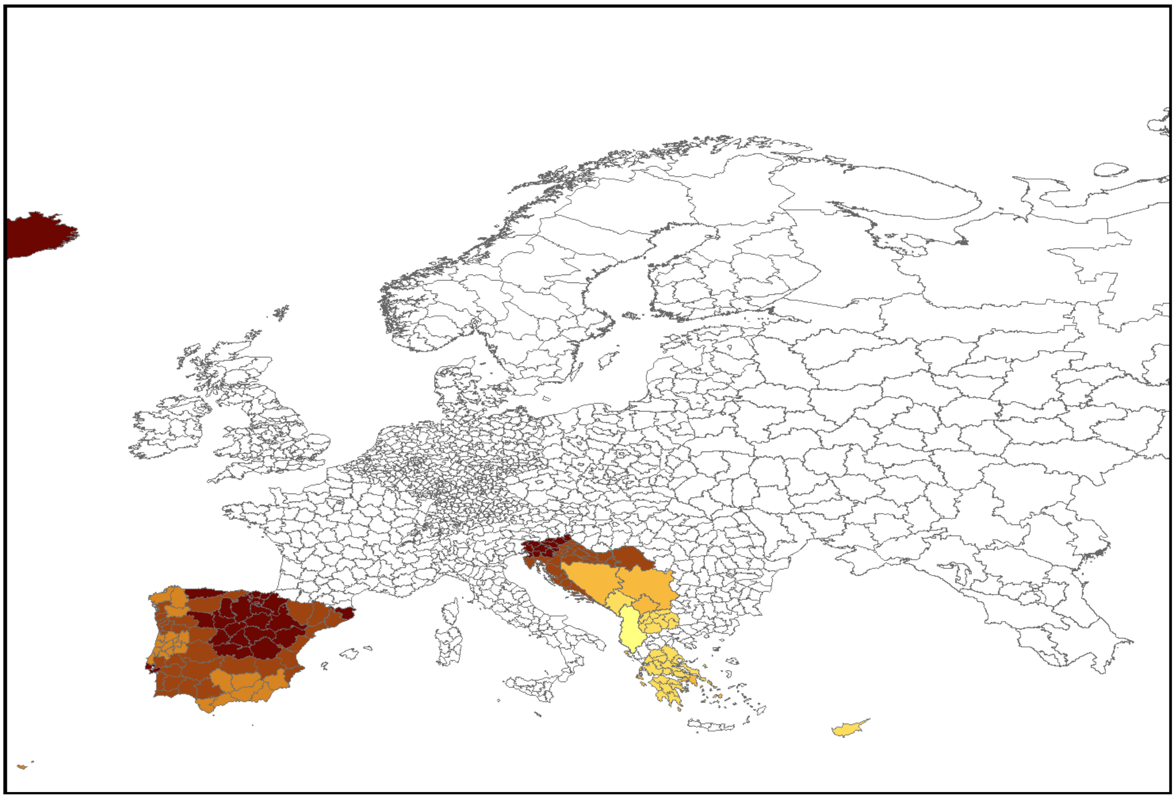


Figure 5.18 ABCC CV of the regions of Austria-Hungary

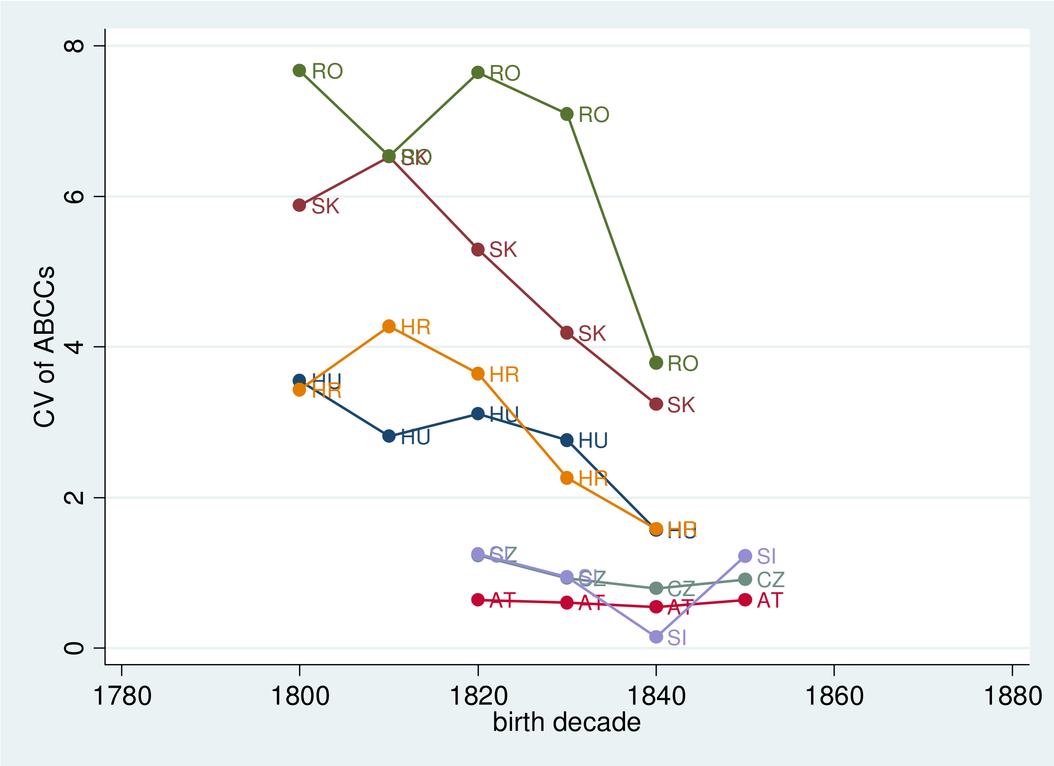


Figure 5.19 ABCC CV of the western periphery

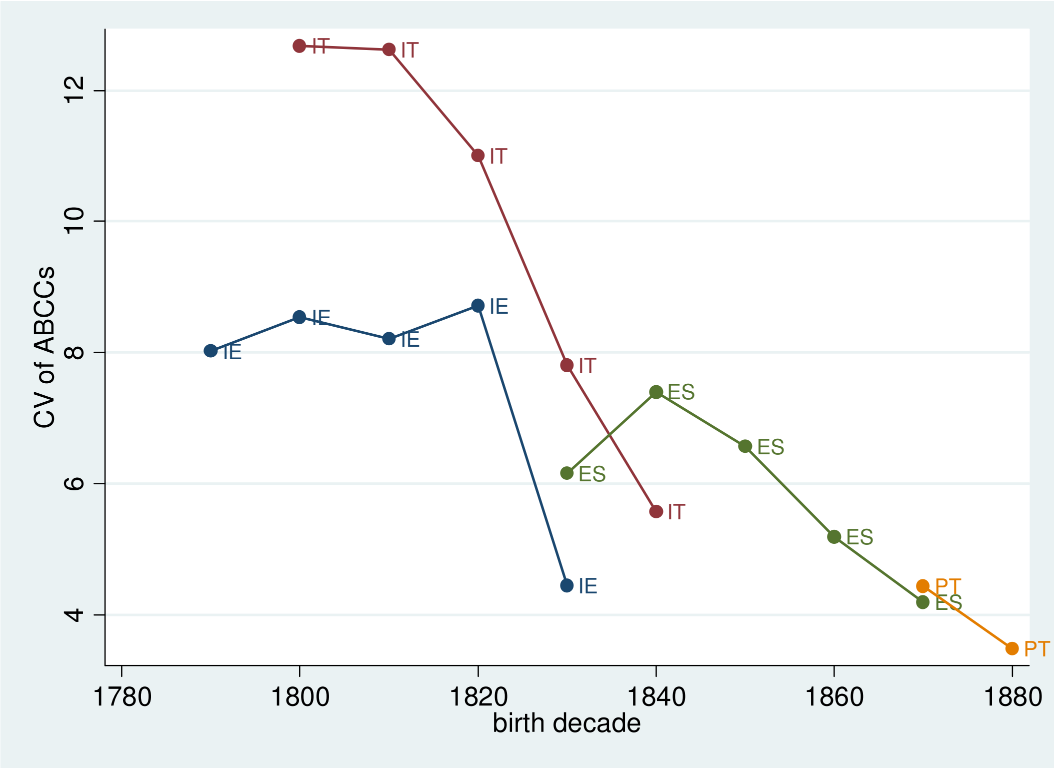


Figure 5.20 ABCC CV of east European countries

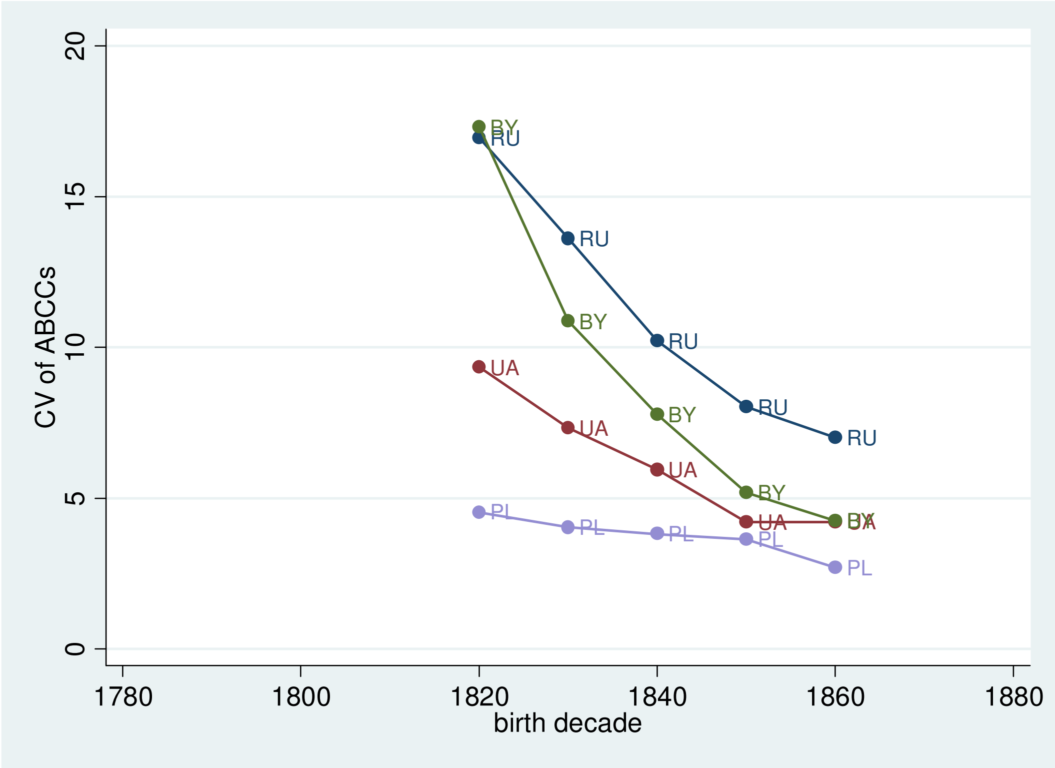
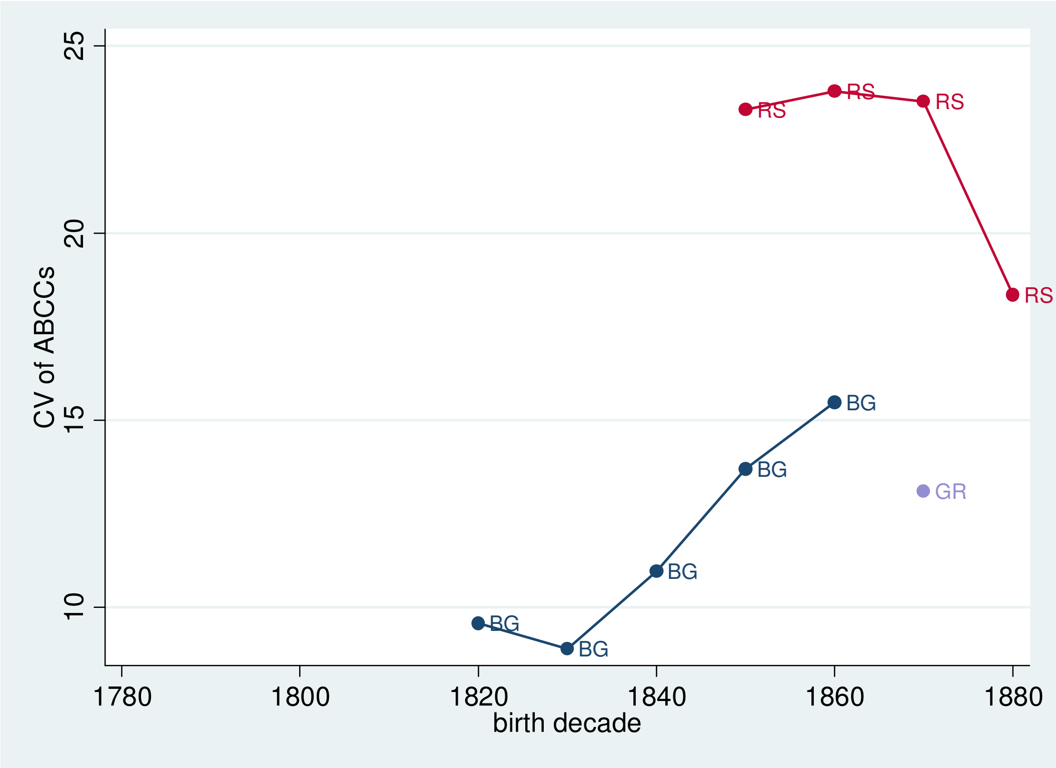


Figure 5.21 ABCC CV of south-east European countries



6. 'Keep them ignorant.' Did inequality in land distribution delay regional numeracy development?

A number of theoretical studies in the Unified Growth Theory framework argue that inequality in land distribution has a negative effect on human capital formation because landowners do not have incentives to promote educational institutions or are not willing to pay the necessary taxes. Using a large data set on regional numeracy in 19th century Europe, we analyse the relationship between land inequality and human capital. We find a substantial negative effect from land inequality in less industrialised countries, especially when instrumenting land inequality with soil and climatic suitability indicators.

This chapter is based on a working paper co-authored with Joerg Baten (University of Tuebingen). The concept for the paper was developed jointly; the analyses and writing were equally shared. It has been submitted to an academic journal.

6.1 Introduction

In their Unified Growth Theory, Galor and Weil (2000) and Galor and Moav (2002) consider human capital to be a crucial factor for explaining the explosion in economic growth. In a recent paper, Galor, Moav and Vollrath (2009; GMV) argue that inequality in land distribution has a negative effect on the development of institutions that promote human capital. Although land availability allowed the creation of economic prosperity in some land-abundant countries during the early phases of development, the former beneficial effects of land were reversed after a period of population growth because of the subsequent inequality in land distribution.

During the transition from an agricultural to an industrial economy, a major conflict arose between agricultural landholders, on the one hand, and capitalists, on the other hand. Landholders did not benefit as much from an increase in the human capital of their workers as did capitalists. According to GMV, the reason for this result is that human capital raises the productivity of workers much more in industry than in agriculture because land and human capital are less complementary. As a consequence, the return on land declines as the wages of workers rise because of the higher level of education. Moreover, educated workers have more incentives to migrate to urban, industrial areas than do their less educated counterparts. This departure of workers from their fields is obviously contrary to the interests of landholders. As a consequence, they tried to hinder the developing 'rural exodus'.

For this reason, as GMV describe the process, landowners inhibited educational policies and reforms that aimed at augmenting the general education of the people. In contrast, capitalists could take advantage of the higher standards in education by increasing their industrial output. Therefore, as long as landholders had sufficient political power to hinder education reforms, they did so. As a consequence, inequality in land distribution may be seen as an obstacle to human capital formation and, thus, as a factor in slowing down the process of industrialisation and the creation of economic growth. We would argue that this effect might have been even more pronounced the case in countries where landed elites were quite powerful, such as Russia and Spain. In this respect, land inequality is a driving force in the emergence of the great divergence in income per capita, which has long-run implications that we can still see in the world today.

GMV assess their theoretical model by empirical evidence from the US in the first decades of the 20th century, even if, in this case, the industrial elites might have dominated national politics. In contrast, in this study we assess the detrimental effects of land inequality on the development of basic numerical skills in Europe in the 19th century. We employ the age heaping method which measures the share of the population which could at least report its own age exactly, rather than reporting a rounded age in censuses. By using a new and large data set on numeracy, adding relevant data on land distribution and including several control variables, we trace the effect of land inequality on human capital in Europe at the regional level.

We also take care to differentiate between the situation in the less industrialised east and south and the more industrial western part of Europe. Based on GMV's theory, we would expect that the land inequality effect depends on bargaining between large landowners and industrialists. It is very likely that in the less industrialised east and south, the political power of landowners was still considerably stronger than in the western European countries, in which the industrialists had taken over a large share of the decision power.

We employ different econometric estimation strategies to check the robustness of our results and to address endogeneity. First, we use OLS models constructed as panels and as country-specific regressions. Second, we check the causality of land inequality on numeracy by employing an Instrumental Variable (IV) approach. Similar to GMV and Easterly (2007), we use geological and geographical variables (cereal suitability, altitude, ruggedness) as instruments which have the advantage to be intrinsically exogenous. To ensure that our instruments are uncorrelated, we use Principal Component Analysis (PCA). Our results suggest that land inequality has a substantial negative effect on numeracy in less industrialised countries.

The paper is structured as follows. First, we revisit the literature on human capital, land distribution and land reforms. Second, we describe the data used in this study and underline the regional land distribution in Europe in the past. We highlight the modelling strategy and the empirical results in the next section. The last section concludes.

6.2 Literature review

6.2.1 Economic growth and inequality

Inequality is one important mechanism that explains different growth patterns around the world that led to the great divergence. Initially, the inequality literature focused on redistributive policies. In particular, proponents of the political economy channel first argued that the majority of voters would vote for policies that favour redistribution in contrast to growth. For this reason, higher tax rates might be expected in more equal societies (Alesina and Rodrick 1994, Persson and Tabellini 1994). However, Perotti (1996) did not find evidence for this mechanism.

Later, research concentrated more on the lobbying ability of different societal groups to maintain the status quo or change the existing distribution of property. By influencing politics, powerful landowners were able to delay policies aimed at redistributing property and at enabling poor workers to obtain a higher level of education. Unlike other sources that concentrate on *political* changes, GMV focus on the *economic* incentives for landowners to hinder institutional changes that would promote education and, thus, economic growth. However, the dynamics of economic development may bring about changes in the behaviour of landowners, even when political structures stay the same.

Research has long focused on *income* inequality in contrast to *land* inequality (e.g., Kuznets 1955). Deininger and Squire (1998) investigate the relationship of both variants of inequality with growth, using worldwide country data from the second half of the 20th century. One of their results is that initial land inequality has a negative and significant influence on later growth, in contrast to initial income inequality. On a theoretical level, this relationship may be explained in two ways. First, it is possible that "credit rationing in the presence of indivisible investments – in schooling, for example – may prevent the asset-poor from making economically profitable investments" (Deininger and Squire 1998, p. 260). In this way, land inequality influences the individual ability to invest. Second, the possession of assets may be an important factor in determining one's individual capacity to take part in the political bargaining process and in promoting one's interests, which are affected by political decisions. Moreover, the significant relationship between initial land inequality and growth is only valid for lower-income countries. It actually does not hold for high-income national entities. Nevertheless, initial land inequality still has an influence on school attainment. Therefore, the link between land inequality and human capital formation and accumulation is confirmed.

Mechanisms influencing growth other than inequality have been discussed in the literature. For example, the quality of institutions may affect growth because good institutions may protect property rights and be conducive to fostering policies that promote technological and economic progress. Acemoglu *et al.* (2005a) argue that different types of colonies, depending on whether settlers or non-settlers were in the majority, led to different types of institutions, which may or may not have promoted growth.

In addition, geography and natural geographical conditions affect the distribution of wealth and, thus, the economic growth process. Detrimental climate and disease could delay the transition from an agricultural to an industrialised economy. Similarly, Engerman and Sokoloff (2000) argue that initial factor endowments in the New World had a major influence on the development of institutions and, thereby, on economic growth. These endowments may explain the differences in institutions. Because factor endowments were quite unique in the colonies, they may have had important and enduring effects on the economy. For example, the quality of soil and climate determined which commodities could be grown in these colonies. Moreover, some commodities are much more suitable for large-scale production because considerable economies of scale can be achieved. A good example of this economic advantage are sugar plantations. The advantage of their large-scale production led to larger farms and favoured the import of cheap slaves to work on these large plantations. In this situation, inequality rose in terms of landownership, wealth and human capital. As a consequence, inequality was cemented between a small elite class owning large tracts of land and the majority of individuals who owned only a small portion. As a result, the elite class had the power to maintain its economic influence by shaping political institutions. For this reason, initial factor endowments led to inequality promoted by institutions and, therefore, to differences in economic growth during later stages of development. Although Williamson (2010) and Dobado and García (2010) recently challenged the long-run view of Engerman and Sokoloff, the basic mechanism for the period since the 19th century is still plausible.

Easterly (2007) also addresses the inequality-growth relationship. He comes to the conclusion that a high level of structural inequality hinders the development of mechanisms, such as schooling, that enable economic development. He further argues that this structural inequality was brought about by agrarian conditions, as presented by Engerman and Sokoloff (2000).

Nevertheless, Glaeser *et al.* (2004) see human capital accumulation as a more important factor for economic growth than institutions. They argue that human capital had a positive influence on the development of institutions, and not *vice versa*. They show that the previous standard measures of institutions were not able to establish a causal relationship between institutions and growth. Instead, their results suggest that human capital plays a more important causal role. The initial level of education is strongly associated with later economic growth. This connection supports the human capital approach taken by GMV.

In view of these considerations, does the argument of GMV hold, that landowners used their economic power to influence the political process against educational and other reforms? Literature on this phenomenon remains relatively scarce (Wegenast 2009). One of the first theoretical considerations was made by Bowles (1978). He argued that there was an important difference between capitalists (promoting education) and landowners (hindering education), as put forward by GMV.

In general, anecdotal historical evidence shows that landowners have used their local power to influence policy, particularly with regard to education. In Spain, a land reform was initiated at the beginning of the 1930s under the Second Spanish Republic also because leftist politicians thought that large landowners would influence local voters in order to obtain seats in the Spanish parliament (Pidal and Rosés 2011) and pursue their own interests.

In Italy, the government was not eager to promote land redistribution measures that would have improved productivity in the agricultural sector. It did not want to decrease landowners' income because they had an important share of the votes until 1911 (Federico 2009). These clientele politics had an impact on the policies of the state. This impact was marked by an absence of action in the south and an activist help for commercial landowners in the north (Elazar 1996).

Lobbying was also an important factor in preserving the large estates held by the landed aristocracy in countries like Hungary. The aristocracy controlled the state and could therefore protect its farms from competition with smaller and more efficient farms (Kopsidis 2009).

In England and Wales, large landowners had a 60 % share of parliamentary seats before the Voting Reform Act in 1885, which gave farm workers the right to vote (Swinnen 2002). As a consequence, the landowner share dropped considerably, to 30 % in 1885 and to 10 % in 1919. Up until this time, landowners dominated the political scene and could influence the government actions that suited their needs.

In addition, Lindert (2004a, b) emphasises the importance of the public's right to vote in order to give more children the chance to attend primary schooling. In an historical comparison, Lindert analyses the factors accounting for the differences in primary school enrollment for a range of countries in the period from 1881 to 1937. Taking a closer look at voting rights, he assesses whether the opposition by landowners significantly impeded the speed of schooling progress. He relates the share of power held by landowners inversely to the overall share of men who were allowed to vote. He distinguishes between non-democracies (such as the Austria-Hungarian Empire), elite democracies (e.g., the UK) and more widespread democracies. He concludes that an elite democracy did not have, on average, a higher schooling rate than a non-democracy. Higher schooling rates were, however, the case for more widespread democracies. As a consequence, Lindert stresses that landowners were an important obstacle to education reforms. They feared paying higher taxes for financing the education of the masses. Moreover, they were afraid that workers would leave agriculture and that they would be increasingly opposed to domination by landowners. These three 'threats' or arguments against education are also in line with the predictions by GMV.

Utilising more recent data, Erickson and Vollrath (2004) also show that a lower degree of land inequality goes hand in hand with a larger public provision for education. Other studies, for India (Banerjee and Iyer 2005) and Latin America (Wegenast 2009), come to similar conclusions. In addition to using the indirect effect of education, landowners tried to directly reduce the mobility of workers by legal means, as highlighted by Huber and Safford (1995).

Clearly, a number of other factors might have had an impact on numeracy development, both across regions and over time. One of the important forces could be integration into markets. As there was a broad trend towards basic numeracy between the Middle Ages and the early 20th century in most countries, one could hypothesise that one of the driving forces was the integration into market economies, which required the ability to process numbers, at least in a very basic way. However, some societies were more successful than others in developing those basic skills. The Istanbul region of the Ottoman Empire, for example, was a market economy since the Ottomans 'inherited' a highly urbanized and commercialized economy from the Byzantine Empire. However, Crayen and Baten (2010a) found that this region had a much lower numeracy during the 19th century than very isolated subsistence regions in northern Scandinavia or north-eastern Russia, as we present here in this study. Hence while there is clearly some contribution of the market economy driving basic numeracy over time, it is not so clear that cross-sectional patterns can be easily explained with this factor alone.

6.2.2 The economic and social effects of large farms: the example of England

The question of the right size of a farm was hotly debated in many European countries for a long time. For example, Levy (1911) discusses the particular case of England, which was the leader of industrialisation. He states that the large farm system was perceived to be the most appropriate method of agriculture in England up to 1880. For this reason, there was also a tendency towards the concentration of land holdings. However, this tendency radically changed afterwards, and the merits of small land holdings were emphasised.

Before 1880, English economists argued that large land holdings would increase agricultural productivity, particularly in grain growing, as large farmers had the necessary capital resources to employ the most up-to-date agricultural methods and tools to cultivate their land. Moreover, the large farmer had the human capital to "appreciate and to apply the advances made in agricultural science. He travelled in order to enlarge his mind, and he read the scientific treatises on agriculture, which the small holder regarded as the height of folly" (Levy 1911, p. 2, from Middleton 1807).[65] As grain prices rose between 1765 and 1815, capital was increasingly invested in this sector, and even people from other classes, such as doctors and lawyers, wanted to participate in these attractive profits by becoming farmers (Levy 1911). Land was also let in larger holdings than before. To the landowner, this 'engrossing' of farm land had several advantages. First, he could increase his rent because large farmers could afford to pay higher rents. Second, he was able to save costs on repairs. Third, it was easier for him to collect the rent from a small number of tenants who possessed much land.

Nevertheless, large farmers preferred to employ labourers who did not possess any land. The large farmer wanted to manage his farm as efficiently as possible. The ownership of land and the consequent duties involved, in his view, would distract the

[65] See Weiss-Bartenstein (1917) for a similar argument in favour of large farms in Bulgaria.

attention of the labourer from his work at the large farm. "[H]e wanted his men to be as dependent as possible upon their employer, and, consequently, to depend for their livelihood on their wages" (Levy 1911, p. 37).

This dependence not only concerned the ownership of land but also the acquisition of knowledge and education (e.g., Lindert 2004a). The opposition of landowners in England is well illustrated by a comment made by Davies Giddy (later president of the Royal Society) in the House of Commons in 1807: "giving education to the laboring classes of the poor … would … be prejudicial to their morals and happiness; it would teach them to despise their lot in life, instead of making them good servants in agriculture […]; it would render them insolent to their superiors…" (quoted in Lindert 2004a, p. 100).[66]

Low wages in agriculture and the expropriation of small landholders contributed to the exodus from rural to urban areas and, thus, from agriculture to industry. Workers hoped to ameliorate their position by working in a town or city. For this reason, the detrimental social consequences of large land holdings eventually changed the opinion of some advocates of large farming in England.[67]

The description of the English case illustrates and confirms the general hypothesis of GMV that large landowners had an interest in keeping their labourers and in hindering educational policies. As a consequence, England had a modest level of schooling before 1891 (Lindert 2004b).

---

[66] Similarly, taking another example from Germany, a landed conservative from Silesia in 1800 wrote the following: "[i]s it not true that the lords experience far more difficulty in maintaining authority over their serfs than they did when the latter were still illiterate?" (quoted in Lindert 2004a, p. 101).

[67] However, the system of large farms still progressed in the period after the introduction of the Corn Laws from 1815 to 1846. Neither during this time nor after the abolition of the Corn Laws did corn prices decrease (contrary to general expectations). Therefore, the profitability of arable farming still attracted more capital. Large farms also advanced in this period. Thus, the number of large holdings (here defined as 100 or more acres) increased in the period between 1870 and 1885. Nevertheless, private and state efforts aimed at lowering the social costs of large farms took hold from 1880 onwards, and, as a consequence, the number of large holdings decreased in favour of smaller holdings.

6.3 Data

6.3.1 Overview

To assess the relationship of land inequality and numeracy, we primarily use data from population and agricultural censuses from European countries in the 19th and 20th centuries.

First, how can we specify inequality in land distribution? We define a large agricultural land holding as extending to more than 100 hectares, as is also indicated by contemporary classification.[68] Note that we take into consideration the *area* of these holdings, in contrast to their *number*.[69] One possibility would be to use the distribution of land holdings as an indicator of land inequality. But the importance of differences in the number of large holdings over total holdings may not be regarded as the crucial determinant of the impact of large land holdings because, typically, the shares involved in considering the *number* of large holdings are very small.[70] By contrast, the share of *area* is much higher and differs much more between countries. A further advantage of using area-based rather than owner-based measures is that, in cross-country comparisons, the inclusion of very small holdings in the statistics varies considerably. Because the very small holdings constitute an important share of the number of holdings, the share in the number of holdings would be very much biased if inclusion definitions are different. The use of the area shares of large holdings avoids this problem because the area share of the

[68] For England, Levy (1911) defines large farms as above 250 to 300 acres, which is approximately 100 to 120 hectares. A similar designation can be applied to Germany (1882), Cisleithania (1902) and Italy (1930). As already shown, the underlying theory highlights the importance of large landowners. Thus, the inequality measure should take into account their share with respect to the total land available. In contrast, other inequality measures that measure *overall* land inequality, such as the Gini coefficient, do not accurately reflect the basic argument of the theory.

[69] For example, Becker *et al.* (2010) use this as an instrument.

[70] In Prussia, for example, 0.2 % of farmers were large landowners (Becker *et al.* 2010). The difference between 0.1 and 0.3 might be less important, compared to the share of land owned by large landowners being 40 to 60 %.

smallest holdings is very small and does hardly affect the area share of the largest land holdings.

Data on the area of agricultural holdings are normally available by bin size (e.g., 0-5 hectares (ha), 5-20 ha, 20-50 ha, 50-100 ha, and over 100 ha), which allows us to directly calculate the share of large holdings by dividing the total area of holdings larger than 100 hectares by the total area of all holdings.[71] However, the measurement unit 'hectare' was not used in all European countries. For instance, holds were used in Hungary and acres were employed in the United Kingdom. As a consequence, we converted the data into hectares and weighted the size categories closest to 100 hectares by their corresponding share.[72] By using this methodology, we were able to compare the share of large land holdings throughout Europe. Summary statistics on this variable are given in Table 6.1.

One potential caveat is that detailed agricultural censuses, which included information on the size of farms, were taken in most European countries only in the second

[71] This calculation is used for all countries except Cisleithania, where only data on the number of agricultural land holdings by size categories are available at the county level. Here, we assume that the area of all farms within a given size category is equal to the value of the average size of that particular category, as proposed by GMV. For instance, given a category from 10 to 20 hectares, we consider all farms to be 15 ha large. As the bin sizes for smaller and medium holdings are quite detailed (below 0.5, 0.5-1, 1-2, 2-5, 5-10, 10-20, 20-50, 50-100 ha), we estimate large landholdings by calculating the share of smaller holdings and taking the difference of the smaller holdings below 100 ha to 1.

In contrast to data on the counties, data on the area of agricultural holdings is also available for the crown lands of Cisleithania. Thus, we are able to compare the obtained values with the official estimations. Excluding the Austrian Littoral and Dalmatia, the estimated values are highly and significantly correlated, by 96 %, with the given values in the statistics. However, the official calculation, particularly for Dalmatia, raises important doubts because, even if we assume that the area of all farms in each size category corresponds to the upper bound of the size category, the share of the land holdings below 100 ha is still far too low. It may be possible that the missing knowledge of the local people of the hectare system led to this discrepancy (which is indicated in the introduction of the agricultural census) or that not all land holdings were included.

Furthermore, the overlapping regions from Cisleithania in 1902 and Italy in 1930 (Alto Adige, Trentino, Gorizia and Trieste) are highly and significantly correlated (about 99 %), indicating that the Austrian Littoral may well have had a much higher share of large holdings than indicated in the official statistics. This result would confirm our estimations for these regions.

Moreover, in Russia, the shares were calculated by using two separate publications (see appendix).

[72] For example, 1 acre is about 0.405 hectares. In England and Scotland, size categories are 100-300 acres and 300-500 acres. Therefore, we calculated the share of the holdings larger than 100 acres (about 40 ha) and larger than 300 acres (about 120 ha). We then added them by weighting the former by 0.264 and the latter by 0.736 because the latter is closer to the measure 'above 100 hectares'.

half of the 19th and the first decades of the 20th centuries. Our numeracy evidence refers to the same period, but we would have preferred inequality data from an earlier period to avoid potential contemporaneous correlation problems. For example, the first agricultural census was taken in the Austrian part of Austria-Hungary (Cisleithania) in 1902. The earliest evidence (see Table 6.2) comes from the United Kingdom and Ireland (1875 and 1881, respectively) and the latest from Italy and Spain (1930). Clearly, the unavailability of earlier data may appear to be a major obstacle in obtaining meaningful results. However, land inequality can generally be seen as quite stable over time. Agrarian reforms were enacted in western Europe as a reaction to the French revolution at the beginning of the 19th century, but, afterwards, no radical changes were enforced. This agrarian situation is confirmed by taking a closer look at the evidence from some of the countries. For example, Eddie (1967) shows that the share of large land holdings, with a definition similar to that used in this paper (above 200 holds or 115 hectares), remained almost constant in Transleithania (the Hungarian part of the Habsburg Empire) in the period between 1867 and 1914, although changes occurred within the different size groups. For example, at the regional level, the share (in area) of large land holdings in England, Wales and Scotland, changed, on average, by less than 2 % between 1875 and 1895 (see details in Table 6.3). According to the statistics, the largest differences between 1875 and 1895 occurred in England in the small counties of Rutland (- 6 %) and Stafford (+ 5 %), in Wales in Carnarvon (+ 4 %) and in Scotland in Clackmannanshire (+ 7 %) and Ayr (+ 6 %). In all other counties, the differences were even smaller; in many counties, almost no changes occurred. Thus, these results highlight, once more, the stability of the share of large land holdings over time.

In eastern Europe, similar questions about land distribution and land reform were more intensively discussed than in western Europe about one century later. The First

World War altered the political and economic positions of many countries. In addition, the creation of new countries as a consequence of the war encouraged major agricultural reforms in some European countries. Nevertheless, our data stem from all countries from before this period of major transition, except in the case of Spain and Italy, where earlier data is not available. The agricultural censuses were the first ones in Spain and Italy. How, then, was land distribution affected in the 19th century and the first decades of the 20th century in Italy and Spain?

The south of Spain had an important share of large estates since the Reconquista. A Liberal Reform was enacted at the turn of the 19th century but did not have a very strong effect on land distribution – the large estates of Andalusia remained intact. Church lands were expropriated and privatised, but much of the land went into the hands of already large landowners. The next important change was a reform that was only approved in 1936. However, it was not entirely implemented because of the Civil War (Pidal and Rosés 2011). For this reason, we argue that the share of large land holdings in the 1930 data is a reasonable indicator of late 19th century land inequality.[73]

In Italy, discussions on the redistribution of land began after the unification of the country. Nevertheless, few policy activities were introduced and fewer still had any strong effect (Dovring 1965). Although minor transfers between large landowners and smaller peasant owners took place until later in the century, no major redistribution of farm sizes took place. Our argument is therefore similar to that of Spain.

What about land reforms in other countries in the 19th century? In France, the government did not make great efforts to change the farm structure existing since the French Revolution. Dovring (1965) notes that until 1940, France was one of the least

[73] The regions in the north of Spain are not included in this census. The first time that they were included was in the 1960s. Because this agricultural census is very late, we have not included these data.

performing actors of political land redistribution. Although the intention of improving the situation of small farmers was expressed, action was mostly limited to words alone.

In Austria-Hungary, no active land reform was undertaken in the considered time period. Hungary was characterised by large estates that were the result of the long, ongoing threat of the Ottoman Empire. These estates belonged to a nobility class that led the military and could therefore justify its ownership and protection of large lands. However, land changes occurred because of the liberation of the Hungarian serfs in the middle of the 19th century, which created new smaller farms but also combined existing larger ones. No major changes to the overall pattern occurred until the First World War (Dovring 1965).

Liberating the serfs in 1861 also had repercussions in Russia. In contrast to the generally held idea that consolidation of the farms benefited the serfs, the reality was that peasants only obtained their previous share of the land, or less. Nevertheless, the incompetence of many landowners with regard to agrarian organisation and production and the competition arising from peasant labour led to a decrease in large estates after 1861 (Dovring 1965). However, the farming structure at the end of the 19th century underestimates the importance of large farms in the first half of the century.

In Serbia, farms were mostly small after independence from the Ottoman Empire in 1882. Although no official agrarian reform took place, parts of the land were redistributed because peasants took over the land from the former Ottoman landowners who fled from the country. Given the departure of the foreign landowners, large farms disappeared, and Serbia became a country of small- and medium-sized peasant farms (Von Franges 1937). Thus, a new land reform that aimed at breaking up larger farms was not necessary. Furthermore, although a major redistribution of land was enacted in other regions of the newly created Kingdoms of the Serbs, Croats and Slovenes after World War I, Old Serbia was almost completely unaffected. A similar statement can be made about

Bulgaria. After independence in 1878, the Bulgarian government redistributed some of the land from the departing Turks in small pieces (Strong 2004). Therefore, agricultural census data from the end of the 19th century might underestimate the share of large farms in these two countries. Nevertheless, both countries had previously been characterised by a rather small share of large farms. Thus, the role of the Ottoman agricultural system and the retreat of the large Turkish farmers after independence in Serbia and Bulgaria are important in analysing farm sizes in the Balkans.

In addition to regional distribution of land, the measurement and development of regional numeracy in Europe in the 19th century is a focal point of this study. In contrast to literacy data, evidence on numeracy has the advantage of being available earlier and more broadly at the regional level. Therefore, by making use of a new and large dataset, Hippe and Baten (2012a) highlight the regional inequalities in numeracy development in more than 570 European regions between 1790 and 1880. Whereas numeracy was already quite high in many central European and Scandinavian countries, southern and eastern Europe lagged behind; even in western Europe, some regions had relatively low numeracy levels (for an example, see Figure 6.1). The lowest numeracy values were found in south-eastern Europe. Nevertheless, many of these 'periphery' countries were able to catch up to the other countries over time. Moreover, the fact that many countries were characterised by considerable regional differences in numeracy underlined the importance of taking a more disaggregated view. Rather than simply focusing on cross-country differences, we see the need to examine within-country variation more thoroughly.

Therefore, in this paper, we use the database from Hippe and Baten (2012a) to model the impact of regional differences in land distribution on numeracy. However, as some central and north European countries already attained the maximum numeracy level by the middle or the end of the 19th century, this upper-bound problem would bias the results. For this reason, we excluded these countries from the analysis. Consequently, we primarily concentrate on periphery countries in southern and eastern Europe, but also include two advanced countries with sufficient regional variation in numeracy (France and the United Kingdom). More specifically, the included countries are Austria (Cisleithania), Hungary (Transleithania), the UK, France, Italy, Russia[74] and Spain (see Table 6.1). In total, these countries include more than 300 regions.

Our indicator of numeracy is the ABCC Index. The ABCC has been used in a multitude of recent publications (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010a, Hippe and Baten 2012a). The ABCC Index takes advantage of the fact that in historical censuses (and in contemporary ones in some developing countries), many people did not state their exact age correctly. Many individuals rounded their ages to values ending in '0' and '5'. This preference for certain values leads to a characteristic 'age heaping' pattern in demographic statistics and can be used to infer the numeracy of the population. As has been shown by Crayen and Baten (2010a), the most important factor for explaining age heaping is education. Not all individuals knew their exact age and thus had to estimate it. Therefore, this age heaping allows us to measure the very basic numerical ability of a population. The ABCC Index is calculated as follows:

$$ABCC = 125 - 125 \times \left(\sum_{i=5}^{14} n_{5i} \Big/ \sum_{i=23}^{72} n_i \right), \qquad (6.1)$$

where i stands for the years of age and n stands for the number of observations.[75] As in the literature, we take only ages between 23 and 72.[76] Because we are only interested in the level of numeracy in this study, we take all ages (23 to 72 years) together.

[74] Not all regions of European Russia (excluding Poland and the Caucasus regions) are included because data on small land holdings are not available for these regions.

[75] This formula is derived from transforming the original Whipple Index into the ABCC Index and simplifying the formula.

This indicator is clearly not a perfect measure for human capital, but rather a proxy for *basic* numerical skills. However, basic numerical skills are preconditions for more advanced skills. Persons who are not able to perform those basic skills will also not be able to work as accountants or engineers, but on the other hand only a fraction of them will develop the skills to become accountants or engineers. A perfect human capital measure in our view would be a composite index of basic and advanced text-related skills, of basic and advanced numerical skills, of technological, social and organizational creativity, and perhaps other components. However, given that such perfect composite indexes are impossible to construct for most real world situations, scholars often use proxy indicators for more broad concepts and this is what we do as well.

We use current regional boundaries as defined by the NUTS classification of the European Union.[77] This classification begins at the national level (NUTS 0) and goes down to the county level (NUTS 3).[78] Clearly, there were considerable border changes in Europe in the 20th century, particularly in eastern Europe. However, in some countries, such as France and Spain, internal and external borders hardly changed. Nevertheless, border changes are taken into account as well as possible by attributing the historical regions to the current NUTS regions.[79] Because our data are quite detailed for many countries, such as for Austria-Hungary, and sometimes even much more detailed than NUTS 3, we avoid border problems in many cases because smaller regions are accumulated into larger NUTS regions. In general, the data are mostly available for NUTS 2 and NUTS 3.[80]

---

[76] We also adjust for the first birth decade (23 to 32 years), as proposed by Crayen and Baten (2010a).

[77] NUTS (Nomenclature of Territorial Units for Statistics) is used by EU and EFTA members as well as official Candidate Countries for the EU. For countries outside the NUTS classification, in particular in Eastern Europe, we use the current regional administrative classification used by these countries.

[78] Moreover, the LAU (Local Area Units) are the next smaller level, including districts.

[79] That is, historical regions were matched as closely as possible to current NUTS regions.

[80] More precisely, data are available on NUTS 3 for the former regions of Austria-Hungary (i.e. Austria, Croatia, Czech Republic, Hungary, Slovakia, Slovenia and parts of Italy, Poland, Romania, Ukraine and Serbia), France and Spain as well as on NUTS 2 for Bulgaria, Ireland, Italy and the UK. Exceptions are the

6.3.2 Regional land distribution in Europe

Having looked at the national level, let us now use a much more detailed dataset to see how land was distributed in Europe at the regional level. In Figure 6.2, we show the share of large land holdings in the European regions. The south of Spain (particularly, Andalusia) and the south of Italy are generally viewed as a stronghold of latifundists and large farms. This perception is partly confirmed by our data. In Italy, the southern region of Calabria and the region around Rome (today's Lazio), in particular, are characterised by a high share of large farms.

Along with the Austro-Hungarian parts of Cisleithania and Transleithania, the Czech lands have a lower share of large farms, while Galicia is much more unequal. Hungary is characterised by a higher land inequality that, however, differs regionally. Interestingly, the Transylvanian region appears to be the least unequal.

In the United Kingdom, large farms prevail in the eastern and southern parts, as well as in the border regions up to Scotland. In Scotland, the Lowlands have a higher share than the Highlands. In contrast to the situation in the United Kingdom, land appears to be rather equally distributed.

With regard to the European part of the Russian Empire, a high share of large farms can be found in the western parts (and, in particular, in the Belorussian regions) and also in the east.

In contrast to more northern countries, the Balkan countries of Serbia and Bulgaria are characterised by a very low share of large land holdings. As has already been highlighted, both countries had mostly peasant farms of small size, which made it very difficult for farmers to generate sufficient output for a living (Berend 1985), even though

greater regions of London (on NUTS 1) and Paris (on NUTS 2). For countries where the NUTS classification is not applicable (e.g., Russia), we use data for the current administrative provinces.

Bulgaria was considered the "peasant land *par excellence*" (Eddie 1967, p. 304, accentuation in original). Large land holdings were almost non-existent in these countries when the agricultural censuses were taken.

6.4 Results

6.4.1 OLS models

Did the share of area owned by large landowners have a systematic impact on the numeracy of the regional population?

We plot the relationship between numeracy and land inequality in the European countries that are available for study in Figure 6.3. As can be seen, there is a negative slope for most countries. Exceptions are France and, less visibly, the United Kingdom. In these countries, a positive relationship prevails.

To better understand these differences, we use the following OLS regression model:

$$ABCC_i = \beta_0 + \beta_1\, share\_100ha_i + \beta_2\, popdens_i + \beta_3\, metropolis_i + \beta_4\, share\ prot_i + \varepsilon_i, \quad (6.2)$$

where *i* denotes the specific country, *ABCC* denotes the level of numeracy, *share_100ha* is the share of the land above 100 ha (= share of large farmers), *popdens* is the population density[81], *metropolis* is a dummy variable for metropole regions[82], *share prot* is the percentage share of protestants[83], and ε comprises the non-observed influences on the ABCC.

---

[81] That is, the number of individuals in a region per square kilometre.

[82] This variable includes the major cities in the countries under study. These are always the capital cities except for Russia, where historically both Moscow and St. Petersburg have been capital cities and have obtained similar importance regarding population size and political influence. Therefore, metropole regions are in Russia St. Petersburg and Moscow, in Austria Vienna, in France Paris in Italy Rome, in Hungary Budapest, in Spain Madrid and in the UK London.

[83] More concretely, it is the number of Protestants in a region divided by the total population in a region.

First, we have included population density because a more dense population allows schooling with much lower commuting costs (Bouccekine *et al.* 2007). Second, we added a dummy variable for metropoles because, in those large cities, more public goods in education were created by the government. Finally, we constructed a Protestant share variable. Protestantism has a reputation for more substantial educational inputs for large parts of the population (Becker and Woessmann 2009). Jewish minorities accounted for only small shares of the population.

We take a two step approach, first constructing a panel and then considering each country separately. We decided to report robust p-values in order to avoid heteroskedasticity problems, although we treat each region with equal weight here, cumulatively, as one historical experience.

For the panel analysis, we group our data into four different categories (see Table 6.4). The panel approach has the advantage to allow the inclusion of more observations than otherwise possible. Note that we always include country dummies to control for country fixed effects.

In column 1, we include those countries that we are most interested in, i.e., today's Russia, Hungary, Spain and Italy. These regions were less industrialised around 1900; hence, their focus on agricultural production might have given power to regional landowners. In fact, our results indicate that land inequality has a highly significant negative impact on numeracy at the 1 % level. In addition, the metropolis variable is positively significant, showing the possible impact of lower commuting costs. All other control variables are insignificant.

If we enlarge our sample to all regions except for the most industrialised countries (column 2), the negative sign for land inequality is once more confirmed, even though it has turned insignificant (p-value = 0.121).

Next, in column 3, we only take into account the most industrialised European countries in our sample, i.e., Austria, France and the United Kingdom. In contrast to our previous results, the coefficient of land inequality is now positive and significant. The metropolis and the share of Protestants variables also change their sign. The negative sign for the share of Protestants can be explained by the fact that both France and the UK were quite homogeneous in religious terms, France being mostly Catholic and the UK mostly Protestant. We have kept this variable in this column for better comparability with our previous results but it is clearly better suited for the less religious homogeneous countries of Hungary (Transleithania) and the Russian Empire. Moreover, population density is positively significant this time. Overall, the more advanced development status of these countries has altered the relationship between numeracy and our explanatory variables. Therefore, the subsequent analysis will clarify this finding.

Finally, we include all countries in our panel regression in column 4. Unsurprisingly, land inequality is neither negatively nor positively significant given our previous ambiguous results. All other controls are also insignificant.

After these panel results, we consider each historical country separately. This strategy allows us to specify our panel results for each country and check their robustness. For example, there remains the possibility that the obtained findings might not be entirely comparable across countries because the definition of land holding might vary slightly between historical countries (even if the country dummies should capture this effect in the panel regressions). Regressing each country individually avoids possibly arising biases due to the use of different regional units.

We included only countries for which sufficient observations were available. However, given the modest number of regions in some countries, an outlying region can readily make a coefficient statistically insignificant. This is the case in Hungary. For this reason, we have decided to report the coefficient for the whole country, as well as for the country minus one region of Hungary. In addition, we added the Protestant share variable only for Hungary and Russia because, as noted above, most of the other countries were relatively homogenous in religious terms.

As a result, all of the coefficients for the share of area owned by large landowners are negative and substantial in southern and eastern Europe (Table 6.5). In southern and central Spain, as well as in Italy, the effect is evident. In the 50 Hungarian regions, the coefficient is not statistically significant because the central Hungarian region of Fejer had a special development. Its proximity to the Hungarian capital might have played a role here. After reducing N to 49, the Hungarian landowner coefficient is significant. In Russia, the coefficient is substantial and significant. The investment in metropolitan schooling was significant and large in the cases of Spain and Russia, the opposite being true for Italy and Hungary. Finally, we obtain a negative sign for population densities in southern and eastern Europe, after controlling for land inequality and metropolitan effects. In Spain and Russia, this sign is statistically significant, in spite of the relatively small number of cases.

The share of Protestants had a positive influence in both Hungary and Russia. The very high percentage of Protestants in the Baltic region has probably led to the significant coefficient in Russia. In Hungary, although there are positive educational and economic results, the coefficient is insignificant. These results confirm the widespread literature on higher educational achievement in Protestant regions, even though, in Russia, there also might be a hidden cultural effect. In general, the explanatory shares are quite substantial in southern and eastern Europe, maxima being around 0.5 in Spain, Italy and Russia.

In central and western Europe, we find again opposite signs for most of the variables in France and the UK (Table 6.6). Large landownership had a negative sign in Austria, which, however, was a borderline case in our regional classification because it included some eastern European regions in today's Ukraine and some Balkan regions. Land inequality was statistically insignificant, with a p-value of 0.148. It seems that at the higher level of industrial and educational development, as in France and the UK, the effect of large landownership loses its importance and might even be a sign of human capital-intensive agriculture (note the positive significance in the UK). Numeracy is dependent on other factors. Interestingly, large urban agglomerations appear not yet to be a driving force, either – London and Paris have negative coefficients. A potential explanation might be the presence of large numbers of factory workers, who had less financial means to invest in their children's education.

In sum, the OLS models confirm the large landowner effect for our southern and eastern European samples. At this point, it seems to have been a core effect in European development.

6.4.2 Instrumental Variable Models

Next, we need to consider endogeneity. The results of the ordinary least squares regressions could be affected by reverse causality. For example, apart from the direction of causation running from the inequality of land via the political economy of landlords opposing primary schooling (and the tax burden that comes with it), one can also imagine that in the long run, regions with relatively good education even for small-holders could reach a lower inequality level of land distribution, as those peasants could buy more and more land. They might also influence political activity in favour of land reforms, as Cinnirella and Hornung (2011) have pointed out for the case of the historical German kingdom of Prussia. On the other hand, educated smallholders might decide to sell their plots in order to obtain the return to their human capital investment in other industries – in nearby cities, for example. Instrumental variable estimation allows to circumvent these issues of endogeneity. Easterly (2007) and GMV have recently advocated the use of climatic, geological and similar variables which allow types of agriculture that are correlated either with higher or lower efficient sizes of scale. Cereal production, for example, is already highly productive on quite small farm units as has been amply demonstrated in the agricultural economics literature, whereas many other crops are more efficient to use in large units. Cereal production requires relatively clear-cut climatic and soil characteristics; hence we can use those to proxy for the suitability for cereals.

The advantage of climatic and geological suitability variables are their intrinsically exogenous nature, whereas the actual crop use would be influenced by educational levels. Similarly, altitude and its variability (ruggedness) are quite exogenous variables. In order to cope with potential non-linearities, we also include square root terms next to the linear terms of cereal suitability, altitude and ruggedness.

It has become a standard in the literature to use a principal component analysis (PCA) if a number of soil and climatic variables are used as instruments which are highly correlated. For example, the correlation coefficient of cereal suitability and altitude is 0.68, the one with ruggedness is 0.70, and between the latter two 0.78 (all are significant at levels below 1 %). This high degree of correlation would influence the first stage of the regression, weakening the performance of the instruments. The PCA components in contrast are uncorrelated by construction. Our first component explains 83 % of the total variation of the six suitability, altitude and ruggedness variables (including their square root terms), the second component an additional 10 % and the third component 7 %. All three components together explain a cumulative 99 %of the variation.[84]

[84] The first component loads positively on altitude (+0.41 both for the linear and root term) and ruggedness (+0.41 both for the linear and root term), and negatively on cereal suitability (-0.40 and –0.41 for linear and root term). The second component particularly positive on cereal suitability (+0.62 and +0.55, root), and less so on the other four (between +0.24 and +0.31). Principal component 3 loads positively on altitude (+0.55 and +0.44) and negatively on ruggedness (-0.48 and -0.51), almost no loading for suitability (+0.09, -0.08).

We base our first stage of the two stage-least-square estimate on the following equation:

$$Landineq_i = \beta_1 + \beta_2\, PCA_1 + \beta_3\, PCA_2 + \beta_4\, PCA_3 + \beta_5\, X_i + \varepsilon_i, \qquad (6.3)$$

where PCA_1, PCA_2 and PCA_3 are the principal components of altitude, ruggedness and suitability and *X* is a vector of other exogenous variables.

We have also to take into account the exclusion restriction. The exclusion restriction implies that the instrumental variables do not have a direct influence on the ultimate dependent variable. The research on land inequality has found broad support that the exclusion restriction holds, as illustrated e.g. by Easterly (2007), Ramcharan (2010) and Cinnirella and Hornung (2011). In fact, Easterly (2007) has carefully studied the applicability of the exclusion restriction of soil and climatic suitability. He addresses the issue by referring to theoretical considerations and corresponding econometric testing methods. One possibility for such a direct causal channel is the possibility that wheat/rice and sugar have different effects on the wealth of the local population. This could be a potential direct causal influence on cognitive abilities, as those might depend on different investment possibilities. On the other hand, Easterly argues convincingly that the difference in wealth effects of those agricultural goods are quite limited, compared to all the other goods which countries are producing. In addition, he uses tests of overidentification for the econometric testing of identification. His arguments provide clear theoretical justifications of the fact that climatic and soil quality variables are instruments that do not violate the exclusion restriction. Of course, we will use also overidentification tests below, bearing in mind the limitations of this strategy.

In Table 6.7, we show the results of the two-stage-least-squares (2SLS) regressions. Column and 1 and 2 refer to the eastern and southern European countries which interest us most, and column 3 reproduces the results for all European regions.

Firstly the results confirm that the three principal component variables are modestly good instruments of land inequality. They jointly correlate with land inequality, as is documented by the 'first stage' section of Table 6.7. The F-statistic of about 7 is still below the usual threshold of 10, hence we employ a Limited Information Maximum Likelihood (LIML) estimator in column 2. The results barely change. Secondly, the instruments influence the dependent variable only through the potentially endogenous variable, land inequality, as we discussed theoretically before. As we have three instruments, there is the possibility of overidentification. However, the Sargan test for over-identification cannot reject the null hypothesis that all three instruments are uncorrelated with the error term. This might be taken as additional tentative evidence that the exclusion restriction holds.

As a main result, the significant impact of land inequality on early numeracy remains a very consistent determinant of human capital in eastern and southern Europe (column 1 and 2). Most of the other controls are insignificant, except for the metropolis variable, which has a positive and significant impact. In all of Europe, the land inequality variable has a negative and marginally significant impact on human capital. The change of signs compared to the panel analysis we used above demonstrates that the instrumental variables probably also reduce measurement error bias in this case.

We may also check whether land inequality is economically significant. We use the method of calculating the effect of one standard deviation of the explanatory variable on the dependent variable.[85] We find that the effect of one standard deviation of land inequality is quite remarkable: in eastern and southern Europe, it explains about 40 % of the standard deviation of the dependent variable numeracy. Considering all of Europe, it

[85] The calculation is done as follows: multiplication of the coefficient of the explanatory variable in the IV regression (see Table 7) with its standard deviation and then division of the result by the standard deviation of the dependent variable (see Table 8).

still accounts for 20 % of the standard deviation of numeracy. In consequence, these results confirm once more the important effect of land inequality on numeracy.

6.4.3 Comparison of our results with other data

Other data available at the national level allow us to compare and validate our results for the relationship between numeracy and land inequality in Europe in the 19th and 20th centuries. As an alternative to our main database, we take data from Vanhanen (2003) on the share of literates as a proxy for human capital in a population and data on the share of family farms as a rough proxy for small and medium-sized farms (in contrast to large farms). Clearly, family farms can lead to quite different extensions. Thus far, this indicator does not correspond to the reverse of large farms. Nevertheless, this analysis allows us to obtain an impression of the education-inequality relationship in European countries over time.

Figure 6.4 to Figure 6.6 show this relationship in 1858, 1888 and 1918 for a range of European countries. A clear, positive relationship is discernible in the second half of the 19th century. This result confirms the prediction that a rather equal distribution of farm sizes favours human capital formation. However, this relationship becomes less significant over time. The figures show that the share of literates mostly rises in many countries, whereas the share of family farms is more or less constant. Countries such as the United Kingdom, France, Belgium, Switzerland, Denmark, Sweden and Norway had very high literacy values in 1918. These high rates contributed to the dissolving relationship in later decades. Moreover, as countries became more industrialised, the power of landowners decreased because the newly rich capitalists exercised an increasingly higher degree of influence in politics and in the economy, promoting education as argued by GMV. Therefore, the inequality-education relationship should have become less significant as countries were more and more industrialised and landowners were less important. These factors may explain the changing relationship between 1858 and 1918. Nevertheless, the proposed negative relationship between land inequality and human capital is confirmed by the data.

6.5 Conclusion

This paper has analysed the relationship between inequality in land distribution and numeracy, as proposed by Unified Growth Theory, using a large dataset for European regions in the 19th and the first decades of the 20th century.

Galor *et al.* (2009) argue that land inequality plays a major role in delaying educational improvements. They support their theoretical model by analysing the experience of the US at the beginning of the 20th century. However, is this situation an exceptional case or is it also true for Europe?

We employed a large and new dataset on numeracy (in the form of age-heaping-based ABCC indices) and land inequality in more than 320 European regions in the second part of the 19th and the first part of the 20th century. In addition, we controlled for several other explanatory factors. More specifically, we included population density, a dummy variable for metropole regions and the share of Protestants in the population. Moreover, we used panel and country-specific OLS models to check the robustness of our results. Finally, we employed Instrumental Variable estimation models to establish causality, taking advantage of intrinsically exogenous geological and geographical variables. As is standard in the literature, we used Principal Component Analysis to ensure that our instruments are uncorrelated.

Our results suggest that, in earlier phases of industrialisation, the unequal distribution of land did indeed have a negative effect on numeracy development. The finding is consistent with the theory of GMV, because they argued that the effect depends on bargaining between large landowners and industrialists. It is very likely that in the less industrialised east and south, the political power of landowners was still considerably stronger than in the western European countries, in which the industrialists had taken over a large share of the decision power. This impact implies that a more equal distribution of land, for example, by implementing a land reform, may help foster educational attainment and economic growth at this stage of development.

6.6 Appendix

6.6.1 Data

Data on land inequality

Country	Census year	Source
Austria (Cisleithania)	1902	K. K. Statistische Zentral-Kommission (1909). Österreichische Statistik. Band LXXXIII. Ergebnisse der Landwirtschaftlichen Betriebszählung vom 3. Juni 1902, Kaiserlich-Königliche Hof- und Staatsdruckerei, Wien. K.K. Statistische Zentralkommission (1903). Gemeindelexikon der im Reichsrate vertretenen Königreiche und Länder, I-XIV, K.K Hof- und Staatsdruckerei, Wien.
Bulgaria	1908	Direction Générale de la Statistique (1914). Statistique de la Propriété Foncière pendant l'année 1908, Imprimerie de l'Etat, Sofia
France	1882	Ministère de l'Agriculture (1887). Statistique Agricole de la France (Algérie et Colonies), Résultats généraux de l'enquête décennale de 1882, Imprimerie administrative Berger-Levrault et Cie, Nancy.
Hungary (Transleithania)	1895	Ungarische Statistische Mitteilungen (1900). Neue Folge. XXIV. Band. Landwirtschaftliche Statistik der Länder der Ungarischen Krone, 3. Teil, Pester Buchdruckerei-Actien-Gesellschaft, Budapest
Ireland	1881	Census of Ireland (1882). Part II. General report, with illustrative maps and diagrams, tables, and appendix., Alexander Thom and Sons, Dublin.
Italy	1930	Istituto Centrale di Statistica del Regno d'Italia (1936). Censimento Generale dell'Agricoltura 19 Marzo 1930, VIII, Vol. II, Tipografia I. Failli, Roma.
Russian Empire	1905/06	ЦЕНТРАЛЬНЫЙ СТАТИСТИЧЕСК1Й КОМИТЕТ (1907). Statistika zemlevladeniia 1905 g, Svoddannykhpo 50-ti guberbiiam Evropeiskoi Rossii, St. Petersburg; (1907). Kizdaniiu 'Statisticheskaia svedeniiapo zemel'nomvoprosu v Rossii, St. Petersburg.
Serbia	1905	Palairet, M. (1979). Fiscal Pressure and Peasant Impoverishment in Serbia before World War I, *Journal of Economic History*, 3: 719-740.
Spain	1930	Carrión, P. (1932). Los latifundios en Espana, Gráficas Reunidas, Madrid.
United Kingdom	1875, 1895	(1875). Agricultural returns of Great Britain, with abstract returns for the United Kingdom, British possessions, and foreign countries, 1875, Eyre and

	Spottiswoode, London; (1896). Agricultural returns of Great Britain, with abstract returns for the United Kingdom, British possessions, and foreign countries, 1895, Eyre and Spottiswoode, London.

Data on numeracy

See Hippe and Baten (2012a).

Data on population density and protestant share

Same sources as for the ABCC in Hippe and Baten (2012a), except supplementary area data for Hungary from Ungarisches Statistisches Jahrbuch (1911). Neue Folge. XIX Buchdruckerei der Aktiengesellschaft Athenaeum, Budapest.

Data on cereal suitability

Cereal suitability data are raster data with a resolution of 5 arc-minutes provided by FAO and IIASA (2007). *Suitability of global land area for rainfed production of cereals (intermediate level of inputs) (FGGD)*, online, last accessed 5 December 2012, dataset downloadable at http://www.fao.org:80/geonetwork/srv/en/resources.get?id=14077&fname =cereal_int.zip&access=private, see also documentation at http://www.fao.org/ geonetwork/srv/en/metadata.show?id=14077). For more details, see Van Velthuizen, V., Huddelston, B., Fischer, G., Salvatore, M., Ataman, E., Nachtergaele, F., *et al.* (2007). *Mapping biophysical factors that influence agricultural production and rural vulnerability,* Rome: FAO. Regional data have been derived from this dataset.

The suitability of land for a crop (in this case cereals) is estimated by "comparing likely attainable yields with the maximum biological yield for that crop in ideal environmental conditions. Land where attainable yields are very close to the maximum potential yield is classified as very suitable for that crop, whereas land where attainable yields are far below the potential maximum is classified as only marginally suitable or not suitable" (Van Velthuizen *et al.* (2007, p. 2). More specifically, the employed suitability index (SI) "reflects the suitability make-up of a particular gridcell. In this index VS represents the portion of the gridcell with attainable yields that are 80 percent or more of the maximum potential yield for the specified input scenario. Similarly, S, MS and mS represent portions of the gridcell with attainable yields 60–80 percent, 40–60 percent, and 20–40 percent of the maximum potential yield, respectively. SI is calculated using the following equation: *SI = VS*0.9 + S*0.7 + MS*0.5 + mS*0.3, where: VS = very suitable; S = suitable; MS = moderately suitable; mS = marginally suitable*" (Van Velthuizen *et al.* (2007, p. 26).

The intermediate level of input assumption is as follows: "Under the intermediate level of input, improved management assumption, the farming system is partly market oriented[.] Production for subsistence plus commercial sale is a management objective. Production is based on improved varieties, on manual labour with hand tools and/or animal traction and some mechanization, is medium labour intensive, uses some fertilizer application and chemical pest disease and weed control, adequate fallows and some conservation measures." (Van Velthuizen *et al.* (2007, p. 23).

Thus, similarly to Haber (2012), we argue that the intermediate level of inputs adequately represents the available level of inputs and the management methods of our historical European context.

Data on altitude and ruggedness

Data on altitude (median) and ruggedness (standard deviation of altitude) are ESRI grid raster data with a resolution of 30 arc-seconds provided by Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and A. Jarvis (2005). Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology*, 25: 1965-1978. Regional data have been derived from this dataset.

Altitude is defined as elevation above sea level (in meters).

6.6.2 Tables

Table 6.1 Summary statistics for landshare > 100ha

Census	Obs.	Share >100ha				Definition
		mean	*sd*	*min*	*max*	
AT 1902 (Austria)	75	0.41	0.20	0.03	0.83	Total area
BG 1908 (Bulgaria)	6	0.06	0.05	0.02	0.15	Total area
ES 1930 (Spain)	27	0.37	0.19	0.02	0.70	Total area
FR 1882 (France)	85	0.31	0.15	0.00	0.65	Total area
HU 1895 (Hungary)	51	0.40	0.16	0.00	0.71	Total area
IE 1881 (Ireland)	3	0.21	0.08	0.12	0.27	Total area
IT 1930 (Italy)	21	0.36	0.17	0.10	0.74	Total area
RU 1906 (Russia)	39	0.39	0.12	0.09	0.69	Total area
UK 1875 (United Kingdom)	31	0.37	0.14	0.13	0.62	Crops and grass
SR 1905 (Serbia)	1	0.04	---	0.04	0.04	Arable area

Table 6.2 Census years of data for land inequality and human capital

Country	Census years for data on	
	Share > 100ha	ABCC
Austria (Cisleithania)	1902	1880
Bulgaria	1908	1893
France	1882	1851
Hungary (Transleithania)	1895	1869
Ireland	1881	1841, 1851, 1861
Italy	1930	1871
Russian Empire	1905/06	1897
Serbia	1905	1895
Spain	1930	1900
United Kingdom	1875, 1895	1881

Table 6.3 Changes in shares of large holdings in the United Kingdom

County	Difference*	County	Difference*
England		**Wales**	
Bedford	0.00	Anglesey	0.03
Berks	0.03	Brecknock	0.03
Buckingham	0.00	Cardigan	0.03
Cambridge	-0.01	Carmarthen	0.01
Chester	-0.01	Carnavon	0.04
Cornwall	0.00	Denbigh	0.00
Cumberland	-0.01	Flint	0.01
Derby	0.01	Glamorgan	0.02
Devon	-0.02	Merioneth	0.02
Dorset	0.02	Montgomery	0.01
Durham	-0.02	Pembroke	0.00
Essex	-0.01	Radnor	0.02
Gloucester	0.00		
Hereford	0.01	**Scotland**	
Hertford	-0.03	Aberdeenshire	0.00
Huntigdon	0.00	Angus	0.01
Kent	0.01	Argyll	0.00
Lancaster	-0.01	Ayr	0.06
Leicester	-0.01	Banff	0.00
Lincoln	-0.01	Berwickshire	0.00
Middlesex	0.00	Bute	0.00
Monmouth	0.01	Caithness	0.05
Norfolk	-0.01	Clackmannanshire	0.07
Northhampton	-0.01	Dumfriesshire	-0.05
Northumberland	0.02	Dunbartonshire	-0.04
Nottingham	0.03	East Lothian	0.00
Oxford	0.00	Fife	0.00
Rutland	-0.06	Inverness-shire	-0.01
Salop	0.02	Kincardinshire	0.03
Somerset	-0.01	Kinross-shire	0.02
Southhamptom	0.03	Kirkudbrightshire	-0.01
Stafford	0.05	Lanarkshire	0.01
Suffolk	-0.01	MidLothian	0.00
Surrey	-0.01	Morayshire	-0.01
Sussex	-0.01	Nairnshire	-0.01
Warwick	0.00	Orkney	0.04
Westmoreland	0.00	Peeblesshire	-0.03
Wilts	0.01	Perthshire	0.01
Worcester	0.01	Renfrewshire	0.00
York, East Riding	0.03	Ross andCromarty	0.00
York, North Riding	0.00	Roxburghshire	0.00
York, West Riding	0.00	Selkirk	-0.03
		Shetland	-0.03
		Stirlingshire	-0.01
		Sutherland	0.02

County	Difference*	County	Difference*
		Wigtown	0.00
		West Lothian	0.00

Note: London has been omitted since it was not yet listed in 1875. * Difference is the difference in the share (in area) of large land holdings (1875-1895).

Table 6.4 Panel OLS regressions of numeracy on land owned by large landowners and other determinants

	(1)	(2)	(3)	(4)
Regions	East/South†	East/South enlarged§	Centre/West‡	All Europe
Share >100ha	-18.11***	-5.72	4.73**	0.19
	(0.000)	(0.117)	(0.023)	(0.922)
Pop density	2.76	27.65	2.48*	0.06
	(0.965)	(0.398)	(0.099)	(0.968)
Metropolis	6.10**	1.88	-2.29**	1.64
	(0.044)	(0.377)	(0.041)	(0.319)
Share Prot.	0.05	0.05	-0.04***	0.05
	(0.216)	(0.172)	(0.000)	(0.165)
Constant	97.54***	84.53***	96.68***	98.67***
	(0.000)	(0.000)	(0.000)	(0.000)
Country FE	Yes	Yes	Yes	Yes
Observations	92	183	145	328
R-squared	0.59	0.68	0.27	0.70

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. † Russia, Hungary, Spain, Italy. § All countries except Austria, UK, France. ‡ Austria, UK, France. In Russia, metropolis includes St. Petersburg and Moscow, in Spain Madrid, in Italy Rome, in Hungary Budapest, in Austria Vienna, in UK London, in France Paris.

Table 6.5 OLS regressions of numeracy in southern and eastern Europe on land owned by large landowners and other determinants

	(1)	(2)	(3)	(4)	(5)
Country	Spain	Italy	Hungary (Tr.)	Hungary (Tr.)	Russia
Regions excl.	0	0	0	1	0
Share >100ha	-16.61***	-24.44*	-6.78	-9.76*	-14.08*
	(0.000)	(0.094)	(0.193)	(0.052)	(0.067)
Pop density	-0.33***	0.15	0.16	0.18	-0.35*
	(0.008)	(0.237)	(0.144)	(0.103)	(0.053)
Metropolis	17.19***	2.27	2.96	3.13	14.92***
	(0.001)	(0.528)	(0.336)	(0.299)	(0.002)
Share Prot.			0.05	0.05	19.27***
			(0.200)	(0.214)	(0.000)
Constant	100.60***	85.35***	85.23***	85.74***	83.23***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	27	18	50	49	43
R-squared	0.54	0.40	0.07	0.09	0.33

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Excluded region in Hungary is Fejer. Hungary includes Transleithania, Spain includes only central and southern Spain, Russia includes European Russia (modern Russia, Ukraine, Belarus and Baltic States). In Russia, metropolis includes St. Petersburg and Moscow, in Spain Madrid, in Italy Rome, in Hungary Budapest, in Austria Vienna, in UK London, in France Paris.

Table 6.6 OLS regressions of numeracy in central and western Europe on land owned by large landowners and other determinants

	(1)	(2)	(3)
Country	Austria (Cisl.)	France	UK
Regions excl.	0	0	0
Share >100ha	-1.53	9.93***	14.00
	(0.143)	(0.007)	(0.180)
Pop density	-0.01	0.06*	0.02
	(0.270)	(0.055)	(0.207)
Metropolis	4.19	-7.94***	-30.06
	(0.160)	(0.007)	(0.204)
Constant	98.61***	88.00***	88.71***
	(0.000)	(0.000)	(0.000)
Observations	75	81	32
R-squared	0.03	0.10	0.19

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Excluded region in Hungary is Fejer. Hungary includes Transleithania, Spain includes only central and southern Spain, Russia includes European Russia (modern Russia, Ukraine, Belarus and Baltic States). In Russia, metropolis includes St. Petersburg and Moscow, in Spain Madrid, in Italy Rome, in Hungary Budapest, in Austria Vienna, in UK London, in France Paris.

Table 6.7 IV regressions

	(1)	(2)	(3)
Countries	East/South*	East/South*	All
Estimation	2SLS	LIML	2SLS
Panel A. First Stage. Dependent variable: land inequality			
PCA component 1	-.021	-.021	.029
	(0.231)	(0.231)	(0.000)
PCA component 2	-.035	-.035	.004
	(0.096)	(0.096)	(0.754)
PCA component 3	-.092	-.092	-.007
	(0.001)	(0.001)	(0.627)
Panel B. 2SLS estimation. Dependent variable: numeracy			
Share > 100ha	-24.13**	-24.36**	-11.18*
	(0.019)	(0.020)	(0.098)
Pop density	-7.37	-7.78	-1.47
	(0.898)	(0.892)	(0.638)
Metropolis	6.63**	6.65**	2.51
	(0.023)	(0.023)	(0.167)
Protestant	0.05	0.05	0.05
	(0.498)	(0.498)	(0.315)
Constant	100.78***	100.91***	103.77***
	(0.000)	(0.000)	(0.000)
Adj. R-squared	0.547	0.546	0.643
First-stage F-stat	6.819	6.819	9.671
Over-ID test (p-val.)§	0.707	0.706	0.098

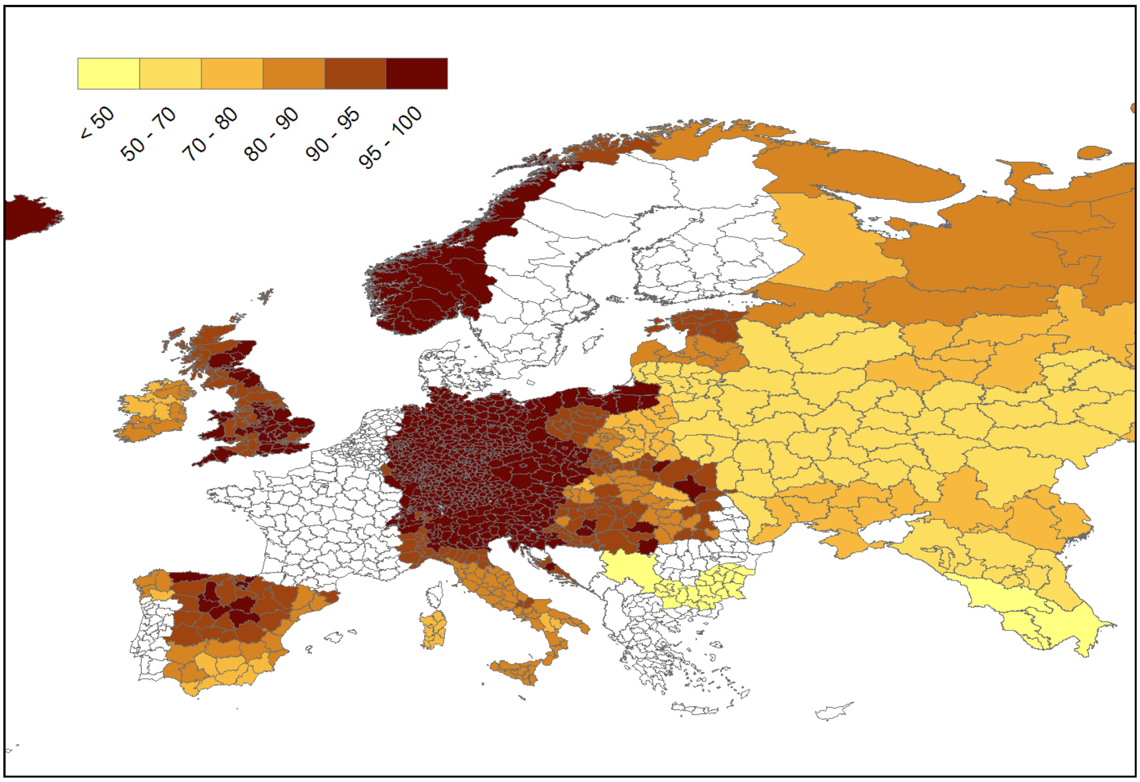
Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. † Russia, Hungary, Spain, Italy. § for the LIML regression, we report the p-value of the Anderson-Rubin chi-square test. In Russia, metropolis includes St. Petersburg and Moscow, in Spain Madrid, in Italy Rome, in Hungary Budapest, in Austria Vienna, in UK London, in France Paris.

Table 6.8 Descriptives of the IV regression

	East/South					All				
Variable	obs	mean	sd	min	max	obs	mean	sd	min	max
ABCC	91	84.39	9.13	60.48	98.12	327	89.93	9.46	51.70	100.00
Share > 100ha	91	0.39	0.15	0.02	0.71	327	0.37	0.17	0.00	0.83
Pop density	91	0.02	0.02	0.00	0.08	327	0.04	0.11	0.00	1.98
Metropolis	91	0.05	0.23	0.00	1.00	327	0.04	0.19	0.00	1.00
Protestant	91	7.25	15.48	0.00	80.00	327	13.61	30.90	0.00	100.00

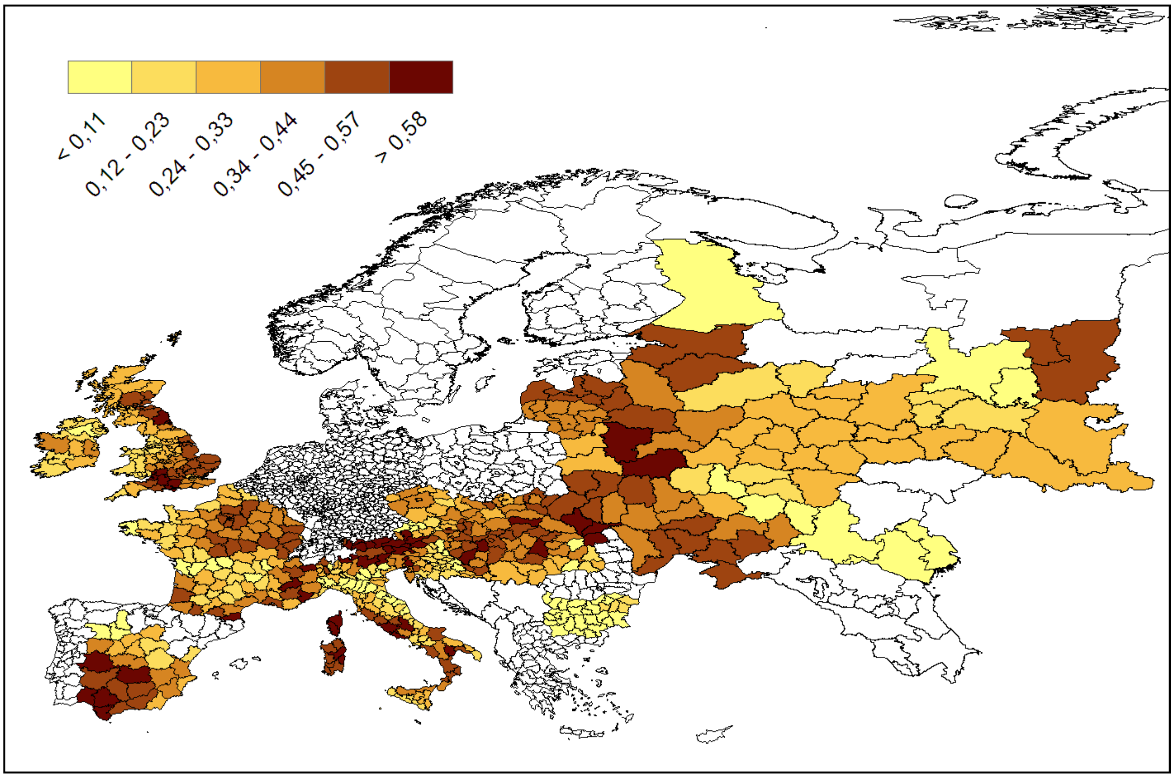
6.6.3 Figures

Figure 6.1 Regional ABCC differences in 1830



Source: Hippe and Baten (2012a).

Figure 6.2 Share of large land holdings > 100ha in European regions



Note: For Poland, Portugal, Greece and north-western/north-eastern Spain, no early land inequality data of comparable quality are available; central and northern Europe were not included because ABCC values did not vary sufficiently.

Figure 6.3 ABCC and landshare > 100ha in selected European countries

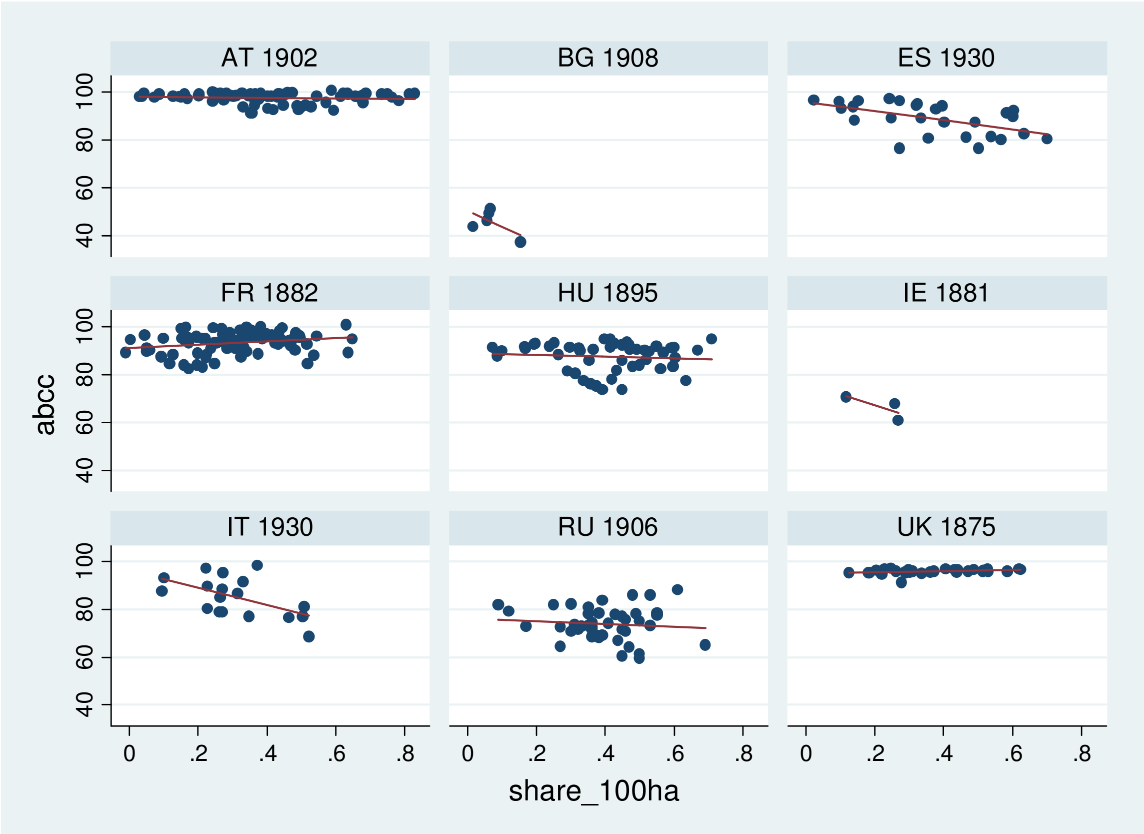
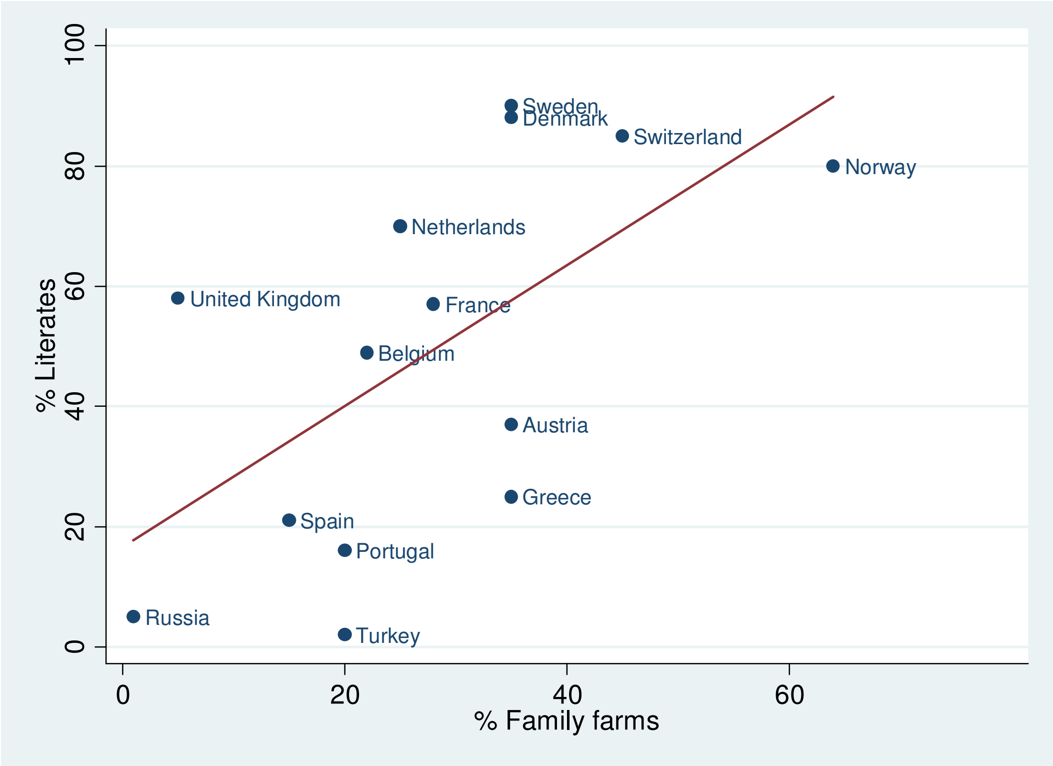
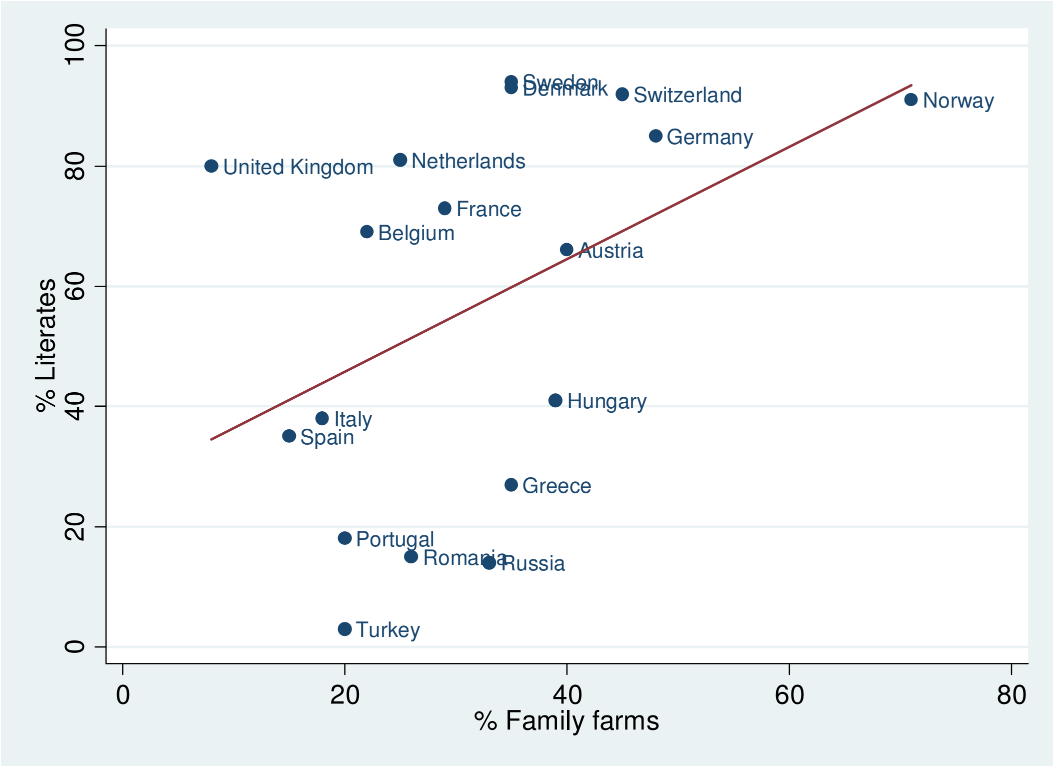


Figure 6.4 Relationship of literacy and family farms in Europe, 1858



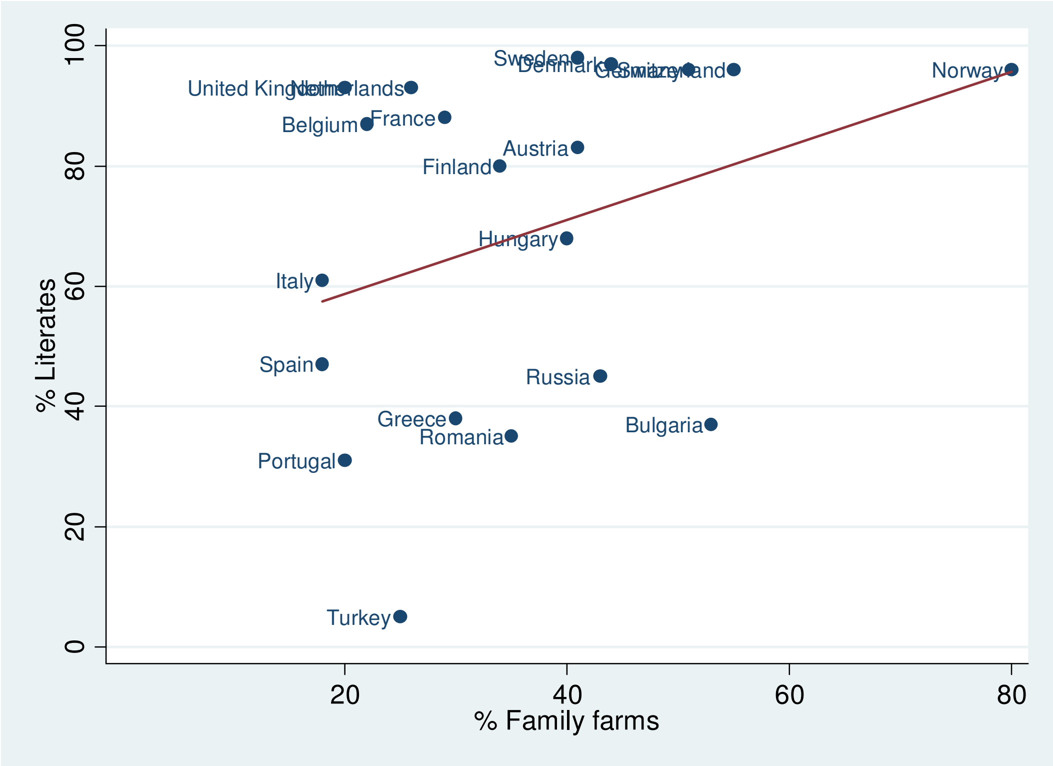
Source: Data by Vanhanen (2003).

Figure 6.5 Relationship of literacy and family farms in Europe, 1888



Source: Data by Vanhanen (2003).

Figure 6.6 Relationship of literacy and family farms in Europe, 1918



Source: Data by Vanhanen (2003).

7. Spatial clustering of human capital in the European regions

Human capital is an important factor for economic growth. However, studies comparing (almost) the entire European continent at the regional level are rare. Therefore, we investigate in this study the spatial distribution of basic human capital in Europe in the long run. To this end, we employ the methods of Explanatory Spatial Data Analysis to analyse the spatial heterogeneity of human capital in the European regions by using cross-sectional evidence for 1850 and 1930. The results show a pattern with one core and several periphery clusters. Moreover, spatial clustering persists throughout the time period. Spatial outliers are rather rare. Geographic location and spatial interactions appear to be important for the human capital of a region.

This chapter is based on a working paper which has been published in *Economies et Sociétés*.

7.1 Introduction

Human capital is an important factor for economic growth. Particularly Unified Growth Theory underlines that human capital plays a major role in the long-run economic development from stagnation to growth (e.g., Galor and Weil 2000, Galor and Moav 2002, Galor *et al.* 2009, Galor 2012).

Still, many studies compare human capital levels only at the national level across countries. However, considerable differences may reside within nations (e.g., Canals *et al.* 2003, Diebolt *et al.* 2005). Hippe and Baten (2012a) show that regional inequality in human capital was considerably high (but decreasing) for many European countries in the 19th century. Thus, taking a closer comparative look at the differences within countries at the European level may considerably advance our knowledge on the importance of human capital.

Economic theories such as New Economic Geography (e.g., Fujita *et al.* 1999) have further highlighted the appropriateness of the regional unit as the basic unit of analysis. Regional data also allow to evaluate the diffusion process of knowledge from one region to another. Accordingly, knowledge spillovers at the local level have attracted the interest of many researchers (e.g., Jaffe *et al.* 1993, Anselin *et al.* 2000, Del Barrio-Castro and García-Quevedo 2005).

For these reasons, the aim of this paper is to analyse explicitly the spatial distribution of human capital in Europe in the long run. What role did geographical proximity play for human capital formation? To this end, we take the new and large database used by Hippe and Baten (2012a) for regional numeracy in the 19th century and add supplementary data for human capital in the 20th century. In particular, we employ regional literacy data from censuses for the 1930s (by Kirk 1946).

The spatial distribution of these data are investigated by the use of the methods of Exploratory Spatial Data Analysis (ESDA). These rather recent methods have increasingly been employed to investigate the role of space and spatial effects on economic and social variables. For example, one application of these methods has been the analysis of the spatial distribution of GDP and patents as well as convergence processes at the regional level (López-Bazo *et al.* 1999, Baumont *et al.* 2003, Le Gallo and Ertur 2003a, Diebolt and Pellier 2009a). In fact, it is not sufficient to detect clusters by only mapping data because human perception tends to find patterns even in random data (Messner *et al.* 1999). In this way, we would run the risk of taking spurious conclusions. Furthermore, it is not possible to evaluate the significance of these clusters. In contrast, ESDA allows a clear identification of spatial clusters and to check their significance. Therefore, we can evaluate the evolving spatial heterogeneity of basic human capital in the European regions by employing these methods.

The specific methods used in this study are Moran's I, the Moran scatter plot and the Moran significance map. We also check the robustness of the applied methodology. The results highlight the spatial clustering of regional numeracy and literacy throughout our period. The overall pattern in 1850 consisting in a clustering of high and low values can still be observed in the 1930s. Hence, geographical proximity to central European countries appears to be an important explanatory factor for regional human capital distribution in Europe. However, we also identify several distinct spatial regimes that do not follow this pattern.

The paper is structured as follows. First, we describe the general evolution of education in the 19th and the first decades of the 20th century. Second, we present the data and the methodology. We employ the age heaping method to proxy for basic numeracy in 1850 and use literacy data in 1930. Our data are disaggregated at the regional level. For this reason, we have attributed the historical boundaries to a constant set of current administrative units. These are the NUTS categories as used by the European Union.[86] This strategy allows us to improve the comparability of our data throughout time. Third, we introduce ESDA as a tool for evaluating the spatial heterogeneity of human capital inequality in Europe. The fourth section presents the results of the application of these methods. Finally, the last section concludes.

7.2 Evolution of basic education in the European regions

Prior to the 19th century, education was not perceived to be important for broad segments of the population. This is also highlighted by the fact that more than 90 % of the population of all continents was not able to write in 1750 (Cipolla 1969). Therefore, education had been a privilege for the upper classes for a very long time. Still in the 19th century, this fact can be illustrated for the case of Serbia: "[e]ducation in Servia is strong at the top and weak at the bottom" (Report of the International Commission 1914, p. 270). Nevertheless, this attitude changed radically as education became to be seen crucial not only for achieving the authority of the state (Green 1990, Vincent 2000) but also for nation building, national integration and for encouraging economic development (Mishkova 1994).

In the decades between 1870 and the First World War, education played an increasing role in public discussion. The general education of the people was, however, not only perceived by western European countries as an important stake for development. In eastern Europe important efforts were also made to close the gap to the more advanced Western societies. At the beginning of the period, this region had very low enrolment rates

[86] NUTS stands for Nomenclature of Territorial Units for Statistics. Eurostat distinguishes between different NUTS levels, beginning with NUTS 0 (the country level) and going down to NUTS 3 (approximately the county level).

which improved during the following years (Benavot and Riddle 1988). Bulgaria is a particular appropriate example. After the creation of the Bulgarian Principality in the year 1878, primary education was made compulsory by the Bulgarian constitution in 1879 (Mishkova 1994). Subsequently, several laws reiterated the endeavour to significantly improve the basic education of the people. Moreover, the state actively supported the construction of new schools and more pupils were allowed to acquire basic education. Finally, the effects of this policy bore fruits: international observers evaluated that the educational situation in Bulgaria was better than in other Balkan states in the 1910s (Report of the Commission 1914).[87]

What about the regional differences in basic human capital in the European countries to that time? A general improvement of basic education throughout the 19th century can be seen in the development of regional numeracy (Hippe and Baten 2012a). Scandinavia, the UK and central Europe had already attained very high numeracy levels in the first decades of the 19th century, while southern and eastern countries such as Spain, the Balkans and the Russian Empire lagged behind. Moreover, regional patterns are observable. For example, Spain is characterised by a core-periphery structure where periphery regions, in particular in the west (Galicia) and in the south (Andalusia), have lower levels of basic numeracy than the more central regions. In contrast, Italy is divided into a northern part with high numeracy and a southern one with rather low values.

However, the authors not only considered the geographical aspect of the regional distribution of human capital but also its evolution throughout time and the inequalities that existed in the different European countries. They show that, in general, the inequalities (as measured by the coefficient of variation) diminished in these countries during the 19th century.

[87] The same was still true 30 years later (Kirk 1946).

The main tendencies of this study are corroborated by research using other human capital indicators. For example, the data used by Cinnirella and Hornung (2011) indicate that enrolment rates of 6 to 14 year olds were already high for contemporary standards in Prussia in 1816 (on average 60.3 %). These rates increased during the century. In 1849, they already reached 80.2 %. The lowest rates were mostly found in the Poznan provinces, which were also those provinces with the lowest ABCCs in Hippe and Baten's (2012a) study. Regional variation was even less important at the end of the 19th century, as demonstrated by average enrolment rates of 93.5 % in 1886 and 94.4 % in 1896.

Moreover, Felice (2012) assembled human capital data on larger Italian regions in the 19th and 20th centuries. His human capital index is a combination of literacy and enrolment rates. He finds that there were large differences in 1871 which decreased during the following decades. North-western regions had the highest values, followed by those from the centre/north-east. Finally, the south and the Italian islands were characterised by lower human capital.

For Spain, Núñez (1992) shows that literacy rates follow a similar core-periphery pattern as indicated by numeracy data. Literacy rates increased so that almost the whole population in the northern regions was literate in the 1930s. However, other regions, particularly in the south, lagged behind. Thus, the correspondence between the overall regional variation in numeracy and literacy has also been demonstrated by Hippe (2012b) for a set of historical European regions and some of today's developing countries in Africa, Asia and Latin America.

Moreover, Kirk (1946) considers literacy in the whole of Europe during the first decades of the 20th century. He shows that there remains important regional variation in literacy around 1930. In fact, the overall patterns are similar to those detected by Hippe and Baten (2012a) for the 19th century. The most advanced countries are located in central, northern Europe and western Europe (see Figure 7.1).

There are some elements that may explain the differences between the countries and the regions within the countries. In particular, Kirk names language, the dominant religion and political components as explanatory factors. Language might play a role when the language used by parts of the population differs from the official language of the state. This may lead to a disadvantage for children that go to school and have to learn a new language.

Second, religion might be considered as a factor due to the importance attributed to reading the scriptures. Protestant regions are generally more advanced in literacy than others, a result that has once more been highlighted by Becker and Woessmann (2009).

Third, history counts because the former political boundaries that had been modified by World War I are still visible in the 1930s. Kirk takes the examples of Alsace-Lorraine being the most literate region in France due to its heritage of the public school system in Germany. Moreover, the borders of the former Austro-Hungarian Empire are still apparent.

Fourth, government policies contributed to the degree illiteracy was fought in the first decades of the 19th century. In particular, Latin countries did not sufficiently succeed in putting educational policies into place. Portugal is just one obvious case. In contrast, countries in eastern Europe were better able to introduce more thoroughly general education as a means to boost their development in the years before 1930.

Nevertheless, apart from these factors, Kirk states quite clearly that space plays the predominant role in the explication of the spatial distribution of literacy: "the most important element in the degree of literacy was geographical proximity to, and cultural intercourse with, the more literate regions of Northwestern Europe" (Kirk 1946, p. 187).

These examples demonstrate once more the importance of not only analysing national human capital formation but of looking at what happens inside countries. However, Kirk makes this hypothetical statement without any (spatial) econometrical analysis. In contrast, we are able to test this hypothesis about the crucial impact of space on regional human capital distribution in Europe.

7.3 Data

The data used in this paper are taken from several sources. First, the database created by Hippe and Baten (2012a) has been used to estimate the regional distribution of numeracy in Europe around 1850. Regional numeracy values are calculated through the use of the age heaping method (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010b, Stolz *et al.* 2013). Age heaping describes the fact that in most historical censuses (and other sources) the actual age distribution deviates from the expected one.[88] More precisely, much more individuals reported ages on 0 and 5 as could be actually the case. Hence, these people were not aware of their own age. For this reason, they rounded (or heaped) their ages. This age heaping phenomenon can be taken advantage of to calculate an index which proxies numeracy.[89] We use the ABCC Index in this study because it is characterised by the same value range as literacy data, i.e., it varies from 0 to 100. The ABCC is calculated in the following way:

$$ABCC = 125 - 125 \times \left(\sum_{i=5}^{14} n_{5i} \Bigg/ \sum_{i=23}^{72} n_i \right), \qquad (7.1)$$

[88] This pattern is still observable in today's censuses of a variety of developing countries during the second part of the 20th century (see Hippe 2012b).

[89] Other reasons leading to age heaping can also be imagined. However, Crayen and Baten (2010a) and Hippe (2012b) show that education is clearly the most important factor.

where *i* stands for the years of age and *n* stands for the number of observations.[90] As it is the standard definition in the literature, we take the ages between 23 and 72 to proxy for basic numerical skills around 1850.[91]

Second, Kirk (1946) collected an impressive data base on Europe in the interwar period. We use the data on regional literacy from his data set. The definition of literacy used by Kirk (1946) is

$$Literacy = \left(\sum_{i=10}^{N} rw_i \middle/ \sum_{i=10}^{N} n_i \right) \times 100, \tag{7.2}$$

where *rw* corresponds to the number of individuals in a region able to read and write, *n* to the total number of individuals in that region and *i* to the years of age. As can be derived from the formula, the age threshold of individuals to be included in the data is 10 years.

Two questions might arise: first, why do we use both numeracy and literacy data? Second, can we use both proxies in combination for the sake of evaluating the evolution of human capital? To answer the first question, we have to deal with the fact that literacy data are not available for many countries as early as 1850. In consequence, we have to rely on another proxy for this point in time. This proxy is the ABCC. On the other hand, the ABCC has already reached or is close to its maximum level in 1930 for an important range of countries, even more than it is the case for literacy. For this reason, the ABCC cannot be used throughout our time period either. Second, we can use numeracy and literacy also because a significant positive correlation between these two indicators has been found in several studies (A'Hearn *et al.* 2009, Crayen and Baten 2010a, Hippe and Baten 2012a, Hippe 2012b). Moreover, both numeracy and literacy are output measures, measuring the

[90] This formula is the result of a linear transformation of the Whipple index into the ABCC Index.

[91] Note that we adjust for the first birth decade (23 to 32 years), as proposed by Crayen and Baten (2010a). In addition, we take earlier or later ABCC values and estimated these values by linear extrapolation because not all the data are available for this point in time.

actual performance of an individual in numeracy and literacy. This makes them better comparable than using an output measure in conjunction with an input proxy such as school enrolment. School enrolment only measures the proportion of pupils enrolled at school without giving us information on the actual knowledge obtained in school. This knowledge might evidently vary very importantly (see also Vincent 2000).[92]

Nevertheless, we are aware of the potential measurement bias using different proxies. Still, we believe that this strategy allows us to pursuit the goal of this paper to highlight the spatial distribution of basic human capital in our period.

After discussing the human capital proxies employed in this study, we have to define the notion of a 'region'. To make the data better comparable, we use a standard administrative definition. This means that we have converted historical administrative borders into today's NUTS regions (see also Hippe and Baten 2012a). Obviously, internal and external borders have changed throughout our period. This is very much the case for eastern European countries such as Poland, Hungary and Russia due to wars and revolutions. In contrast, countries such as France and Spain have almost completely remained with identical administrative divisions during the last 150 years. The latter countries clearly allow a much easier comparison between historical data and current data. Where administrative units have changed, we adapted our methodology as best as possible to this. However, caution has to be taken with the interpretation of individual cases when borders have changed importantly. Although this strategy has the aforementioned inconveniences, we proceed in this way because it has the very important advantage to

[92] Apart from the fact that there is (to our knowledge) no database available yet for school enrolment or other indicators, such as years of schooling, at the regional level to that time.

allow a look at more or less the same regional structure throughout our period. Nevertheless, we had to adapt some regions to the necessities in the different censuses.[93]

How aggregated are our data? We have data available on all NUTS levels. For example, we could use the lowest NUTS level, NUTS 3, in Spain, France, Austria or Slovakia. Clearly, we would prefer to analyse the data always at the highest level of regional disaggregation to use all the information available, i.e., to maximise the number of regions and avoid possible biases involved in using higher aggregated regional units. However, we prefer to use NUTS 2 in general and, in some particular cases, NUTS 1[94], to standardise the regional classification. It is important for the following ESDA methods to operate on similar levels because our estimations would otherwise be potentially prone to biases related to different sizes of our unit of analysis. For this reason, we merged NUTS 3 regions to NUTS 2 regions and weighted the human capital values according to the total population in the region.

This strategy allows us to construct a data set of human capital proxies in the European regions. However, according to Kirk (1946) some countries do not report any literacy rates in the 1930s anymore. This has already been the case since the beginning of the century (UNESCO 1953). These are generally countries with very high literacy rates such as the Scandinavian countries, Germany or the UK.[95] In consequence, Kirk estimated that these countries had illiteracy rates between 0 and 5 %. Because there is no regional variation in these estimates, we would automatically (and intentionally) create positive spatial autocorrelation if we left these countries in the data. This would clearly violate the

[93] This is particularly true for Russia. In fact, the provinces of the USSR, particularly of the Ukrainian SSR, as listed in Kirk (1946), are bigger in 1930 than the ones in 1897. Moreover, the current administrative structure does not have an equivalent to NUTS 2 (only NUTS 0 or NUTS 1 and NUTS 3). In consequence, we had to merge several smaller regions into bigger regions which would constitute NUTS 2 regions. These bigger regions are formed in accordance to historical predecessors.

[94] These are the greater Paris and London regions.

[95] More specifically, UNESCO (1953) mentions Denmark, Germany, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom.

fundamental concepts of the analysis strategy employed. In consequence, we excluded the countries concerned from northern and central Europe. Note that our results are, therefore, limited to this specific setup and could be altered if all European regions could be included. This has always to be taken into account in the later analysis. Nevertheless, we can still work with a large data base for western, southern and eastern Europe and are able to evaluate spatial clustering for these large parts of Europe.

Descriptive statistics of the data can be found in Table 7.1. There are around 190 regions in our data set for each point in time. Values range between 26 and 100 for the ABCC in 1850 and between 17 and 100 for literacy in 1930. The regions belong to the following countries (within current borders): Albania, Armenia, Austria, Azerbaijan, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, France, Georgia, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Macedonia, Moldova, Montenegro, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain and Ukraine.[96]

7.4 Exploratory Spatial Data Analysis

Geographic location has an important influence on the growth of the economy (e.g., Fujita *et al.* 1999, Azomahou *et al.* 2009) but also on knowledge flows (Paci and Usai 2009). Maps and other means of visualisation may help to identify specific spatial patterns in the data. However, they are not sufficient because their interpretation is subjective. This is also why it is possible to perceive spatial patterns when the data are clearly random (Messner *et al.* 1999). Moreover, the evaluation of the significance of potential spatial clusters is out of reach without further tools of analysis. Therefore, the use of appropriate methods is important to identify significant spatial autocorrelation.

[96] We do not have ABCC data neither for some regions that were not Greek at the date of the underlying census nor for the principalities of Romania in 1850.

What is spatial autocorrelation? Spatial autocorrelation is at hand when similar values coincide with similar location (see Anselin 2001). There are two types of spatial autocorrelation: positive and negative spatial autocorrelation. First, positive spatial autocorrelation means that there is a clustering of high (or low) values of a given random variable in space, whereas negative spatial autocorrelation refers to a clustering of dissimilar values. In the latter case, a region is surrounded by neighbours with significantly higher or lower values. Particular spatial patterns may arise due to spatial heterogeneity, resulting in clusters of regions with high educational levels (the core regions) and clusters of regions with low levels (the periphery regions).

In this paper, we investigate the spatial heterogeneity of human capital among the European regions. By comparing several points in time we are able to get an insight into the persistence of spatial inequality in human capital during our period. For this reason, we identify global and local spatial autocorrelation by using Exploratory Spatial Data Analysis.[97]

ESDA is a "set of techniques aimed at describing and visualizing spatial distributions, at identifying atypical localizations or spatial outliers, at detecting patterns of spatial association, clusters or hot spots, and at suggesting spatial regimes or other forms of spatial heterogeneity" (Le Gallo and Ertur 2003a, p. 177; see also Haining 1990, Bailey and Gatrell 1995, Anselin 1998a, b). These techniques allow the computation of, firstly, global and, secondly, local spatial autocorrelation.

On the one hand, Moran's I statistic is used in this paper to detect global spatial autocorrelation. This statistic is a standard approach to calculate global spatial

[97] The following presentation of ESDA (Moran's I, Moran scatter plot, LISA) is based on Le Gallo and Ertur (2003a, b) and Dall'erba (2005).

autocorrelation. It is defined in the following way (Cliff and Ord 1981, Le Gallo and Ertur 2003b):

$$I_t = \frac{n}{S_0} \cdot \frac{\sum_i \sum_j w_{ij}(x_{i,t} - \mu_t)(x_{j,t} - \mu_t)}{\sum_t (x_{i,t} - \mu_t)^2}, \tag{7.3}$$

where t is the year of observation (here: t = 1850, 1930), n is the number of NUTS regions, x is an observation, μ is the mean value of the observations and S_0 represents a standardisation factor which is equal to the total sum of all the elements w of the spatial weight matrix.[98] The elements of the spatial weight matrix which lie on the diagonal, w_{ii}, are equal to zero and the other elements, w_{ij}, denote in each case the spatial connection of a region i to a region j.

It is possible to rewrite Moran's I statistic in a matrix form by defining a vector z_t of the human capital observations for a given year which is in deviation from μ. Denoting W as the spatial weight matrix, this gives us:

$$I_t = \frac{n}{S_0} \cdot \frac{z'_t W z_t}{z'_t z_t}. \tag{7.4}$$

Note that the vector Wz_t is also called spatially lagged vector, representing the averages of human capital values of the neighbours by using spatial weights. Therefore, Moran's I indicates the level of linear dependence between z_t and Wz_t.

The outside influence which possibly affects each region can be normalised by row-standardising the spatial weight matrix, so that each individual row sums to 1. Consequently, the scaling factor S_0 is now equal to n, simplifying Moran's I statistic to

$$I_t = \frac{z'_t W z_t}{z'_t z_t}. \tag{7.5}$$

Moreover, the expected value $E(I)$ of Moran's I is equal to

98 That is $S_0 = \sum_i \sum_j w_{ij}$.

$$E(I) = \frac{-1}{n-1}. \tag{7.6}$$

Thus, when Moran's I is larger than *E(I)*, there is positive spatial autocorrelation in the data, when it is smaller, there is negative spatial autocorrelation.

Clearly, there are different ways of constructing the spatial weight matrix. However, the inherent characteristics of our regional European data make it particularly appropriate to use a spatial weight matrix which is derived from the k-nearest neighbours within the great circle distance d_{ij} between the centroids of regions. This matrix has been used by earlier publications referring explicitly to the NUTS 2 classification at the European level (e.g., Le Gallo and Ertur 2003a, Dall'erba 2005) and in other domains (e.g., Pace and Barry 1997, Pinkse and Slade 1998). In accordance with Le Gallo and Ertur (2003a), the spatial weight matrix takes the following form:

$$\begin{cases} w_{ij}(k) = 0 \text{ if } i = j \\ w_{ij}(k) = 1 \text{ if } d_{ij} \leq d_i(k) \text{ and } w_{ij}^*(k) = \dfrac{w_{ij}(k)}{\sum_j w_{ij}(k)} \\ w_{ij}(k) = 0 \text{ if } d_{ij} \geq d_i(k), \end{cases} \tag{7.7}$$

$d_i(k)$ being the critical cut-off distance for *i*, i.e., it "is the k^th^ order smallest distance between regions *i* and *j* such that each region *i* has exactly *k* neighbors" (Dall'erba 2005). Row standardisation allows to account for relative in lieu of absolute distance. The resulting matrix is w^*. As Le Gallo and Ertur (2003a) and Dall'erba (2005), we use the great circle distance with a minimum of *k* equal to 10 in order to allow the connection of islands such as Sicily, Corsica or the Greek Islands to the mainland. Otherwise we would have zero values for some rows and columns. Because our later empirical results are based on the selected criterion of the spatial weight matrix, we check the robustness of our results by increasing the number of *k* to 15 and 20. The increase in the number of *k* results in an increase in the part of international connections (Dall'erba 2005).

On the other hand, the global Moran's I statistic does not allow a closer examination of outliers and regional spatial clustering. For example, it is possible that high values of human capital are concentrated in some particular clusters and low values in others. Moreover, outlying regions that deviate significantly from their surrounding neighbours should be taken into focus. Therefore, we use the Moran scatter plot (Anselin 1996) and the Local Indicators of Spatial Association (LISA) (Anselin 1995) to analyse the contributions of individual regions and clusters of regions to the overall pattern of global spatial autocorrelation.

First, the Moran scatter plot allows to study spatial instability at the local level. It is divided into four quadrants by the zero values on each axis. The horizontal axis depicts the human capital values in units of standard deviations (vector z_t). The vertical axis reflects the standardised spatially weighted average for the human capital values (Wz_t).[99] The quadrants represent the four different types of spatial association between a region and the surrounding neighbours. In the case of this study,

- Quadrant I (high-high (HH); upper right) shows the regions in the data set which have human capital values above the mean, the average of their neighbours' human capital also being above the mean,
- Quadrant II (low-high (LH); upper left) shows the regions which have human capital values below the mean, the average of their neighbours' human capital being above the mean,
- Quadrant III (low-low (LL); lower left) shows the regions which have human capital values below the mean, the average of their neighbours' human capital being below the mean, and

[99] That is the average in human capital of the regions surrounding a region.

- Quadrant IV (high-low (HL); lower right) shows the regions which have human capital values above the mean, the average of their neighbours' human capital being below the mean.

These quadrants can be classified into two categories: first, HH and LL indicate positive spatial autocorrelation, i.e., a region is surrounded by regions with similar values. The reverse case is given in the second category of negative spatial correlation, i.e., the quadrants HL and LH. Spatial outliers can be easily identified by the use of these scatter plots.

Nevertheless, Moran scatter plots do not allow to judge whether the detected local spatial clusters are significant or not. For this reason, we use a Local Indicator of Spatial Association (LISA). According to Anselin, a LISA has to fulfil two criteria: first, "the LISA for each observation gives an indication of the extent of significant spatial clustering of similar values around that observation" and second, "the sum of LISAs for all observations is proportional to a global indicator of spatial association" (Anselin 1995, p. 94). Thus, we employ a local type of Moran's I statistic (Anselin 1995):

$$I_{i,t} = \frac{(x_{i,t} - \mu_t)}{m_0} \sum_j w_{ij}(x_{j,t} - \mu_t) \text{ with } m_0 = \sum_i (x_{j,t} - \mu_t)^2 /n. \tag{7.8}$$

The interpretation of this LISA is similar to the one of the global Moran's I statistic: there is positive local spatial autocorrelation when $I_{i,t}$ is positive and negative spatial autocorrelation when it is negative. Putting the information obtained by the Moran scatter plot and the LISA together gives us the Moran significance map (Anselin and Bao 1997). This map highlights, through the use of different colours, the regions that are characterised by significant (positive or negative) spatial autocorrelation.

7.5 Results

We highlight first the results for global spatial autocorrelation before taking a closer look at local spatial autocorrelation (Moran scatter plot, Moran significance map).

7.5.1 Global spatial autocorrelation

Table 7.2 shows Moran's I statistic for the proxies of human capital at the different points in time.[100] We employ a permutation approach using 10000 permutations as proposed by the literature (Anselin 1995). It appears from the table that there is positive spatial autocorrelation in our data because the Moran's I statistics are always significant at the level $p = 0.0001$. Moran's I is generally lower for the ABCC in 1850 than for literacy in 1930. More specifically, Moran's I of the ABCC has a value for $k = 10$ of 0.6031. In contrast, literacy's Moran's I is 0.7140 in 1930. Therefore, the statistics imply that regional human capital is clustered in space in 1850 and 1930. This means that we do not find a random distribution of human capital but that higher levels of human capital cluster with higher ones and *vice versa*. Thus, we find a significant clustering of European regions by using this global indicator.

When we use more nearest neighbours (15, 20) and calculate the Moran's I statistics, the sign and significance of the global spatial autocorrelation do not change. The expected value in both cases (ABCC in 1850 and literacy in 1930) is -0.005. Since the number of k-neighbours is higher, Moran's I is lower. Globally, these results confirm the appropriateness of our spatial weight matrix.

7.5.2 Moran scatter plots

In this section, we want to investigate whether there are regions that do not fit into the overall spatial pattern as indicated by the Moran's I statistics. In other words, we want to

[100] All calculations in the following sections were performed by using GeoDa.

detect spatial outliers or atypical locations. For this purpose, we use first Moran scatter plots and then Moran significance maps in the next section.

The results for the Moran scatter plot can be seen in Figure 7.2 and Figure 7.3. Note that we have used k = 10 in this case. Most of the observations are located in quadrants I (HH) and III (LL), i.e., these regions are positively spatially autocorrelated. As can be seen in Table 7.3, about 85 % (HH = 55.0 %, LL = 30.2 %) of all regions are associated with similar values in 1850 and 88 % (HH = 39.7 %, LL = 47.9 %) in 1930. The share of positive spatial autocorrelation is therefore quite similar but the distribution between HH and LL clusters is different.

Moreover, there are a number of regions which show an association with dissimilar values, i.e., LH or HL. It appears that there are more regions that are able to have higher basic human capital values than their neighbours than the other way round.

In fact, in 1850 there are 9.5 % of all regions in the HL quadrant and only 5.3 % in the LH quadrant. HL regions are predominantly northern Russian regions (including Moscow), Estonia, Latvia and a range of regions located on similar latitude from Serbia (Vojvodina) over Romania (Sud-Muntenia), some Ukrainian regions to Russia (Rostov). A somewhat third scheme is constituted by Puglia and Calabria in southern Italy. On the other hand, there are three clusters discerning for the LH quadrant. One is located in the western and southern peripheral regions of Iberia (Portugal's Centro, Spain's Galicia, Andalucía, Murcia and Comunidad Valenciana). A second cluster comprises Eastern Polish and western Ukrainian regions. Finally, another LH region is the Border, Midland and Western region in west/north-west Ireland. The area was the harshest hit by the Irish famine of the 1840s. This might explain its underperformance in human capital.

In 1930 we have a similar scheme with 7.2 % in the HL quadrant and almost identically 5.2 % in the LH quadrant. The HL regions are all located in eastern and south-east Europe, except for Asturias in Spain. In particular, these are Moscow and some regions in its north-east, the St. Petersburg region, Estonia, Latvia, the Ukrainian Krim island, Polish Podkarpackie, Romanian Vest and Bulgary's Severen tsentralen and Yugozapaden (including the capital, Sofia). The case of Bulgaria might confirm our introductory comments on this country, stating that it was particularly successful in elevating the overall human capital of its population during the decades following its independence. The picture is much more diverse for the LH regions, coming from different corners of Europe. More specifically, LH regions are located in Spain (Aragon, Castilla-La Mancha, Comunidad Valenciana), Italy (Sardegna), Croatia (Jadranska Hrvatska), Poland (Swietokrzyskie) and neighbouring Ukraine (Zakarpattia), Romania (Nord-Vest) and Russia (Kareliya).

Some regions are negatively spatially autocorrelated and are located far from the mean value of basic human capital. In 1850, these regions are from the Balkan and Caucasus regions and have very low values. The same observation can be made for the 1930s, when southern Balkan countries (Albania and Macedonia) and eastern Caucasus regions have still not caught up to other regions.

Because the Moran scatter plot does not give any indication of the significance of the obtained results, we use in the next step the Moran significance maps.

7.5.3 Moran significance maps

Local Indicators of Spatial Association are helpful to extend our understanding of the results previously obtained. Therefore, we use Moran significance maps. The statistical results of these maps are summarised in Table 7.4. First, most of the regions which are significant (also called hot spots) are found in either the HH or LL quadrants. This shows a tendency of positive spatial autocorrelation in space. In each year, most regions are in the HH cluster. Second, there are not many observations with significant values which fall into the other two (HL and LH) quadrants. In fact, outliers in the LH and LH clusters are almost not existent (0.1 to 1.5 %).

After these first impressions, we can have a direct look at the Moran significance maps. Note that in the maps (Figure 7.4 and Figure 7.5) the regions with significant values are coloured according to the underlying spatial regime.[101]

In both 1850 and 1930, there is a clear pattern of regions in central Europe to be in the HH quadrant. The regions concerned are primarily from France, northern Italy, Austria, (to that time) western Poland, the Czech Republic, Slovenia and Hungary. Cataluña in Spain is also associated to this cluster in both maps. On the other hand, a first LL cluster is located in the southern Balkans (primarily Serbia, Macedonia, Albania and Greece). In 1850, two other ones are visible in Belarus/western Russia and the greater Caucasus regions. The same can be said for 1930, except that much more Russian (and Ukrainian) regions are now also significantly in this LL cluster. In this case, the southern regions of Portugal, Spain and Italy are also concerned. Whereas other clusters appear to be in line with the notion of geographical proximity to north-western European countries as an explanatory factor, the Belarusian case shows that there have to be other additional factors. In fact, one would expect only very eastern regions in Russia to be characterised by LL clusters because they are the most distant regions. However, this is not true for today's Belarus. An alternative explanatory factor is land inequality which is shown to significantly affect numeracy formation in the Russian Empire (Hippe and Baten 2012b).

In addition, some regions do not fit the general pattern of positive spatial autocorrelation. In 1850, for example, this is the case for today's Ukrainian region Zakarpattia Oblast, an historical multiethnic region. This region had lower ABCC values than its surrounding regions. Cultural and language barriers might have contributed to this

[101] Dark red = HH, light red = HL, dark blue = LL, light blue = LH.

relatively low result. In general, the atypical regions are mostly at the outskirts of the neighbouring cluster. In 1930, among others, this is true for Crimea in Ukraine and the greater region around Bulgaria's capital Sofia (today Yugozapaden). In the latter case, Bulgaria's success in the fight against illiteracy might explain this positive spatial outlier. Still, negative spatial autocorrelation is rare in the data set.

7.5.4 Robustness checks

We checked the robustness of our results by increasing the number of k-nearest neighbours to 15 and 20. If the choice of k does not have a significant impact on the results, most observations would have to lie on the diagonal. That is, they would not change from one category to another. The results in our case are summarised in Table 7.5 and Table 7.6. Evidently, the use of alternatives for k does not significantly change our results. In fact, most of the observations lie on the diagonal. Due to the larger inclusion of neighbouring regions, the first line for k = 15 and k = 20 indicates that there are some regions that formerly were not significant which now become significant. As can be seen in the case of k = 15, these regions become to be positively autocorrelated in the majority of cases (HH = 4.2 %, LL = 5.5 %), whereas the move to negative autocorrelation almost only happens to LH regions (LH = 2.1 %, HL = 0.5 %). Similar are the conclusions when increasing k to 20. Therefore, these results confirm the appropriateness of the choice of the spatial weight matrix. For this reason, we can infer from these checks that our results are robust to different specifications.

7.6 Conclusion

This paper has analysed the spatial clustering of the regional distribution of human capital in Europe in 1850 and 1930. We have used the databases by Hippe and Baten (2012a) and Kirk (1946) to construct evidence on the basis of two cross-sections. To this means, we have employed several proxies for human capital. First, we have used the ABCC method to proxy for basic numeracy levels around 1850. Then, we have added literacy levels for the 1930s. Regions were defined according to the NUTS classification of the European Union. This fixed classification allowed a maximum of comparability of the evolution of regional disparities in basic human capital throughout time.

By using Exploratory Spatial Data Analysis (ESDA) methods (Moran's I, Moran scatter plot, Moran significance map), we investigated the spatial heterogeneity among European regions in human capital. We had to exclude northern and central European regions due to imposed data constraints, so that the focus is on western, southern and eastern European regions. Therefore, our findings apply to this specific regional setup.

The results show that Europe was characterised by an important core-periphery pattern, as evidenced by important HH and LL clusters. The HH cluster is located in central Europe, while several LL regimes exist in eastern Europe and in 1930 also in south-western Europe. Nevertheless, some regions were able to perform better than their neighbours or had significantly lower basic human capital. These outliers showing negative spatial correlation are rare in the sample.

These results imply that there were clear spatial clusters in Europe in the past. These spatial clusters persisted over longer time periods. In consequence, the influence of the surrounding regions has been an important determinant of a region's own human capital. In particular, geographic proximity to the most advanced countries in north-western Europe appears to be an important explanatory factor for the regional distribution of human capital in Europe. Nevertheless, the existence of clusters and outliers which do not fit to this overall pattern show that there are other determinants of human capital. The effect of land inequality, for example, explains regional numeracy distribution which cannot be explained by geographical proximity (Hippe and Baten 2012b). These results are complementary to better understand and explain regional human capital distribution and indicate the road to take in future research policy on this topic. Space is important for regional human capital formation but the right educational, social and economic policies may allow a region or a country to escape from its spatially given destiny.

7.7 Appendix

7.7.1 Data

ABCC data: Hippe and Baten (2012a)

Literacy data: Kirk (1946)

7.7.2 Tables

Table 7.1 Descriptive statistics for human capital proxies

Year	Proxy	obs.	mean	sd	min	max
1850	ABCC	189	87.23	13.70	26.38	100.00
1930	Literacy (> 10 yrs.)	194	69.96	20.75	17.00	99.51

Table 7.2 Moran's I statistic for regional human capital proxies, 1850 and 1930

		k=10		k=15		k=20	
Year	Proxy	Moran's I	sd	Moran's I	sd	Moran's I	sd
1850	ABCC	0.6031	0.0291	0.5471	0.0234	0.4930	0.0202
1930	Literacy (> 10 yrs.)	0.7140	0.0295	0.6634	0.0238	0.6214	0.0205

Note: The expected value for Moran's I statistics in 1850 and in 1930 is -0.005. All statistics are significant at p = 0.0001. The number of random permutations is 10000.

Table 7.3 Percentage of observations in each quadrant of Moran's scatter plot (k=10)

Year	Proxy	HH	LL	HL	LH
1850	ABCC	55.0%	30.2%	9.5%	5.3%
1930	Literacy (> 10 yrs.)	39.7%	47.9%	7.2%	5.2%

Table 7.4 Percentage of observations in Moran's significance map (k = 10)

Year	Proxy	Not sig.	HH	LL	HL	LH
1850	ABCC	45.5%	37.0%	15.3%	1.1%	1.1%
1930	Literacy (> 10 yrs.)	32.9%	34.0%	30.4%	1.5%	0.1%

Table 7.5 Robustness analysis for 1850 and 1930 (k = 10 to k = 15)

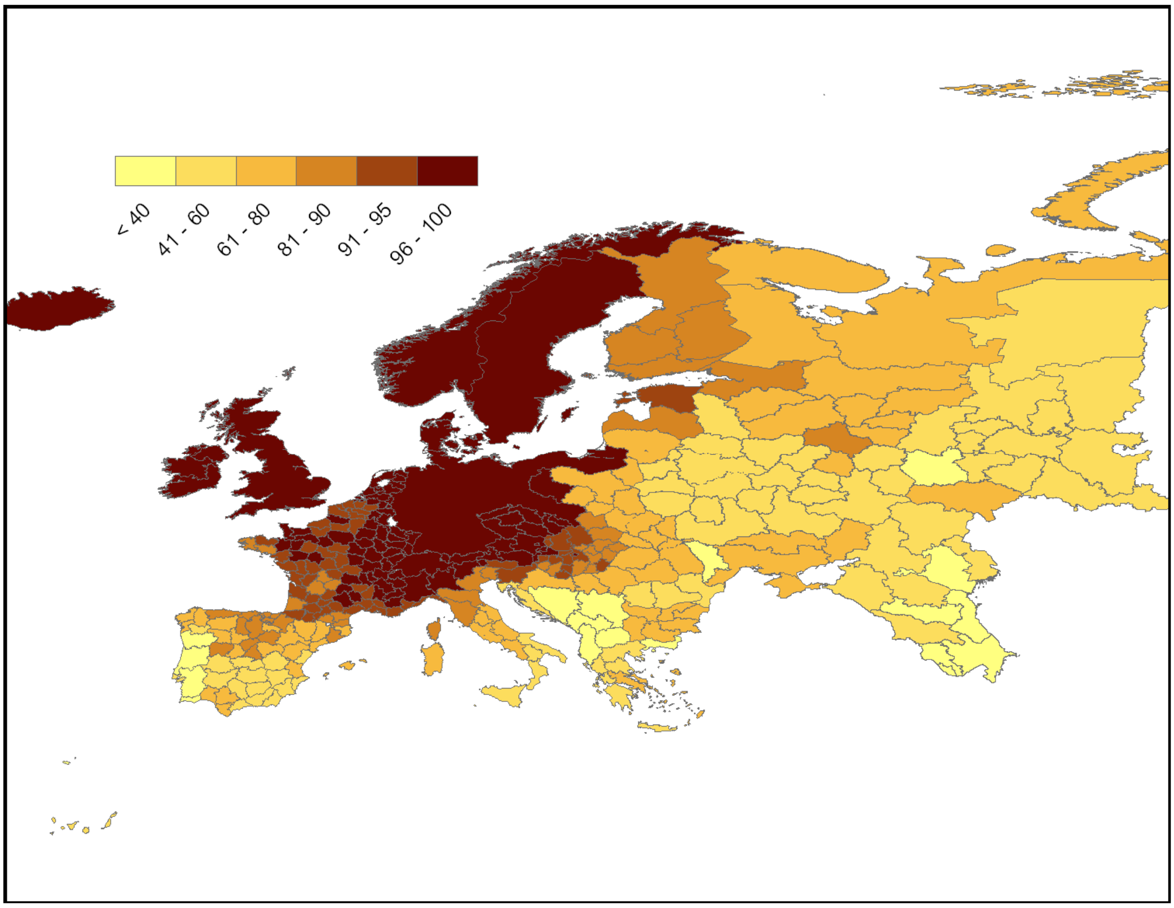
k = 10 / k = 15	Not sig.	HH	LL	HL	LH
Not sig.	87.7%	4.2%	5.5%	0.5%	2.1%
HH	0.3%	99.7%	0.0%	0.0%	0.0%
LL	2.1%	0.0%	97.9%	0.0%	0.0%
HL	0.0%	0.0%	0.0%	100.0%	0.0%
LH	0.0%	0.0%	0.0%	0.0%	100.0%

Table 7.6 Robustness analysis for 1850 and 1930 (k = 10 to k = 20)

k = 10 / k = 20	Not sig.	HH	LL	HL	LH
Not sig.	84.6%	4.7%	7.0%	0.8%	2.9%
HH	0.5%	99.5%	0.0%	0.0%	0.0%
LL	2.6%	0.0%	97.4%	0.0%	0.0%
HL	0.0%	0.0%	0.0%	100.0%	0.0%
LH	0.0%	0.0%	0.0%	0.0%	100.0%

7.7.3 Figures

Figure 7.1 Literacy (in %) in the European regions, ca. 1930



Note: There was no estimation on literacy of Luxembourg available but it should be between 0 and 5 % as in the neighbouring regions.
Source: Own calculations based on data by Kirk (1946).

Figure 7.2 Moran scatter plot for ABCC in Europe, ca. 1850 (k=10)

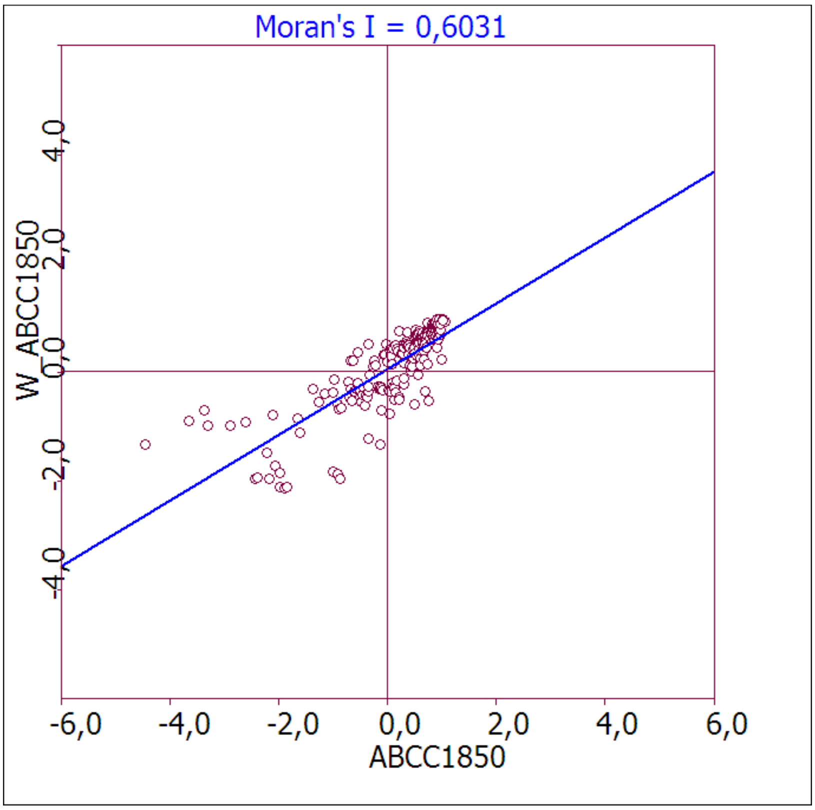


Figure 7.3 Moran scatter plot for literacy in Europe, ca. 1930 (k=10)

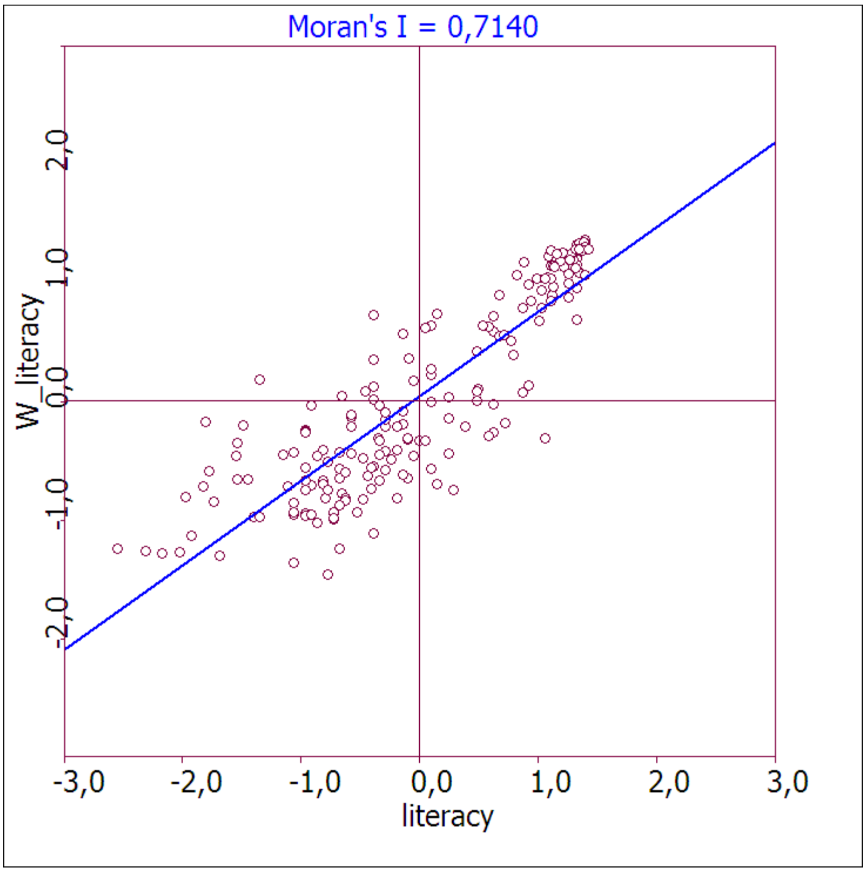


Figure 7.4 Moran significance map for ABCC in Europe, ca. 1850 (5 % pseudo-significance level, k=10)

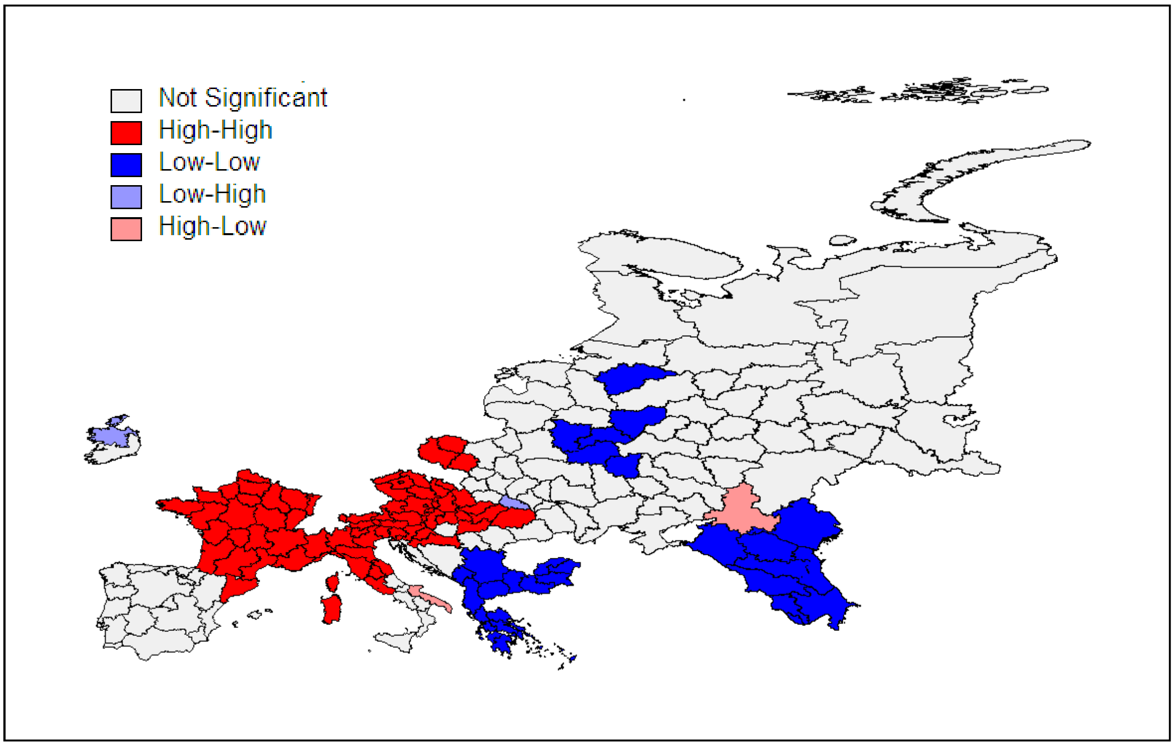
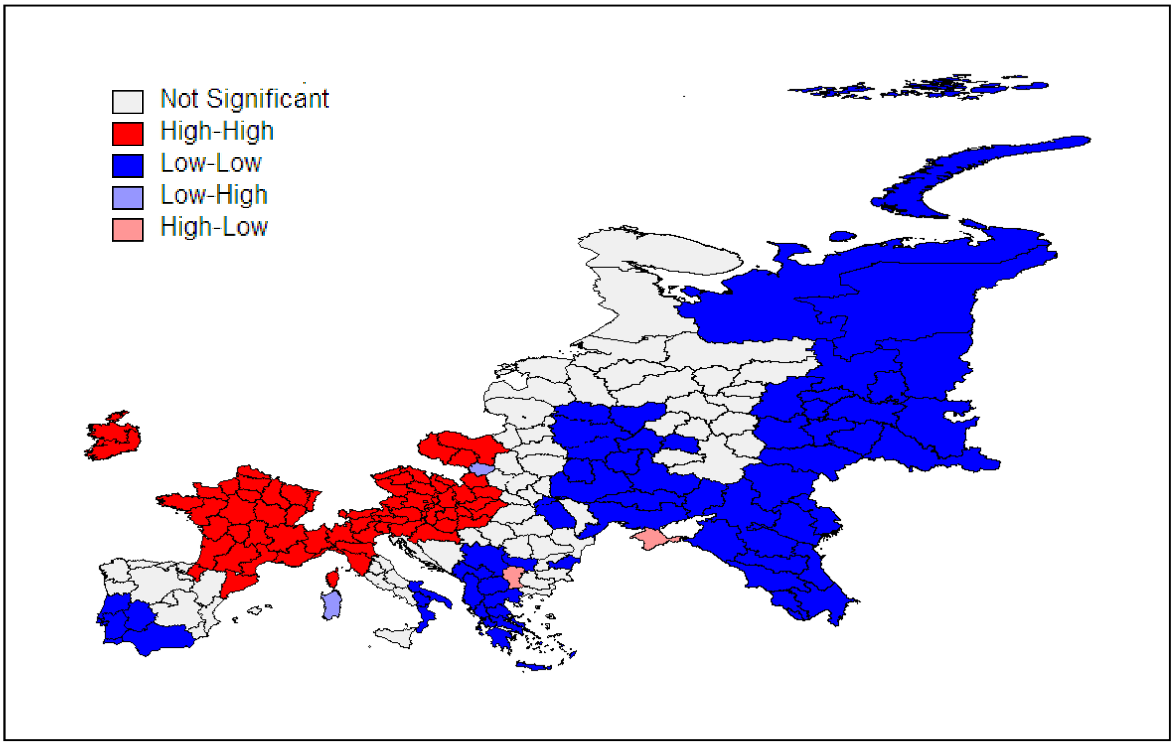


Figure 7.5 Moran significance map for literacy in Europe, ca. 1930 (5 % pseudo-significance level, k=10)



8. Remoteness equals backwardness? Human capital and market access in the European regions: insights from the long run

Human capital is generally considered to be an important factor to generate economic growth. Moreover, New Economic Geography (NEG) has become a major tool to address questions of regional inequalities and convergence processes. In an important recent contribution, Redding and Schott (2003) add human capital to a two sector NEG model, highlighting that remoteness represents a penalty that gives disincentives to invest in human capital. But is this hypothesis consistent with long-term evidence? We test the persistence of this effect at the regional level in an historical setting. In particular, we use numeracy and population potential to proxy for human capital and market access in the European regions in 1850. Alternatively, we also use literacy to measure human capital in 1930. The results show that market access has a significant positive influence on human capital in OLS, Tobit and IV regression models. Thus, the paper confirms the 'penalty of remoteness' hypothesis for Europe in the long run.

This chapter is based on a working paper which has been submitted to an academic journal.

8.1 Introduction

Human capital is generally perceived to be a key factor for today's knowledge-driven economies. This is particularly true for Europe and the European Union. For this reason, the Council of the European Union highlights that "[e]ducation and training have made a substantial contribution towards achieving the long-term goals of the Lisbon strategy for growth and jobs" (Council of the European Union 2009, C 119/2).

Still, the EU is facing important challenges in its regional policy. Although the EU has aimed to decrease economic and social regional inequalities over the last decades, there still remain important differences between and within countries. This is also true for education which is not equally distributed in space. Thus, the question arises how one can explain these differences.

One possible explanation advanced by theory and in particular by models from New Economic Geography (NEG) is that consumer markets play an important role in the distribution of economic development. These models have already been tested empirically for the last decades (e.g., Redding and Venables 2004, López-Rodríguez *et al.* 2005, Breinlich 2006, López-Rodríguez and Faíña 2006, Head and Mayer 2011) and have confirmed the predictions provided by these rather recently developed theoretical models. A particular case including human capital formation is presented by Redding and Schott (2003). The authors develop a theoretical NEG model showing that remoteness from large consumer markets gives disincentives to individuals to increase their human capital. For this reason, this 'penalty of remoteness' explains worldwide inequalities in human capital accumulation. Subsequent empirical studies for the European regions for the last couple of years have also confirmed the predictions of the model (e.g., López-Rodríguez *et al.* 2005).

Nevertheless, (to our knowledge) there has not yet been any evidence for the long-term evolution of market access and human capital in such a setting. This, however, appears particularly important to understand the changes that have shaped today's European regions in the long run. This may considerably enlarge the recent analyses for the short term which may be only part of a much larger long-term process.

For these reasons, this paper explores the importance of market access for the spatial distribution of human capital in the past. More specifically, we use regional numeracy values that have been calculated by the age heaping method for 1850 as our human capital proxy. As has been proposed by the recent literature, we exploit data on the distribution and size of cities in Europe to model historical market access. Moreover, we test the robustness of our results by employing literacy as an alternative indicator of human capital in 1930.

The results show that market access has a significant negative influence in OLS, Tobit and IV regressions. In the latter case we use distance to Luxembourg and area size of European countries as instrumental variables. The picture that we get from these regressions is that the 'penalty of remoteness' hypothesis theoretically advanced by Redding and Schott (2003) is confirmed by our historical data. This implies that the 'penalty of remoteness' is not a current trend but has existed for (very) long time periods, the present being only a very special case of a larger phenomenon.

The paper is structured as followed: first, we consider the literature on human capital formation in the European regions in the past and the main contributions of New Economic Geography. Then, we present the underlying theoretical NEG framework which has been originally proposed by Redding and Schott (2003). Subsequently, the data and the econometric specifications are discussed. In the fourth section we show the results. The final section concludes.

8.2 Regional human capital formation in Europe, today and in the past

Human capital formation in the European regions has attracted the attention from many researchers (e.g., López-Rodríguez *et al.* 2005, Breinlich 2006, López-Rodríguez and Faíña 2006). For example, one can highlight the contribution by Rodríguez-Pose and Tselios (2011) for the present. They use Exploratory Spatial Data Analysis to test the spatial distribution of educational attainment in western Europe between 1995 and 2010. They find that educational attainment is strongly correlated with inequality and that regions tend to cluster in space. Proximity plays an important role for educational attainment even today. Moreover, there are noticeable differences between the north and the south of western Europe and the urban and rural communities.

However, as the authors state, "[t]he geography of education, especially at [the] subnational level, is a huge black box" (Rodríguez-Pose and Tselios 2011, p. 358). If this is still true for today, one can imagine how the situation is for the past. New evidence on the regional distribution of human capital in the past has recently been provided by Hippe and Baten (2012a) and Hippe (2012c). For example, clusters were already clearly identifiable in 1850 and 1930, as evidenced by numeracy and literacy data (Hippe 2012c). Using ESDA, a core-periphery pattern is discernible with regions close to the (potentially) most developed countries in human capital having the highest human capital values (see Figure 8.1 for 1850). In general, when going further away from the centre, human capital levels decrease. The regions in the Belarusian region are particular outliers because they have consistently lower values than their neighbours which is why they fall into the LL (low-low) cluster. Moreover, the fact that the north-western part of Ireland was severely hit by the Great Famine between 1845 and 1852 may have contributed to the fact that it has fallen into the LH (low-high) cluster, meaning that it is a region with low human capital values which is surrounded by regions with high human capital values. Finally, the

Zakarpattia region in today's Ukraine has historically been quite ethnically mixed which can explain the fact that it is also characterised by being in the LH quadrant.

8.3 NEG and the economic geography of Europe

New Economic Geography has become an important field in economics over the last years. This is also highlighted by the recent attribution of the Nobel Prize for Economics to its founder, Paul R. Krugman, in 2008. New Economic Geography models enable to understand why economic activity and individuals cluster in space. In other words, it is possible to clarify the reasons for the existence of urban agglomerations, e.g., Tokyo and Mexico City, and areas with concentrated activity, such as the Manufacturing Belt in the United States and the Blue Banana in Europe. In fact, concentration is the most evident characteristic of economic geography (Krugman 1991b). This statement is confirmed by taking a look at the unequal regional distribution of GDP per capita in Europe (Figure 8.2). That economic geography still plays an important role for the distribution of economic development can also be illustrated by the use of Figure 8.3. There is a clear correlation between GDP per capita and the approximate economic centre of the EU in 1999, Luxembourg.

The increasing and in part spectacular growth of urban agglomerations particular in developing countries further shows that economic geography is an important factor for the distribution of the population in the past, today and probably in the future. The distribution of the European population may be illustrated by using population density data (Figure 8.4). The most densely concentrated populations are predominately located in England, Belgium, France, Germany and Italy.

Given these facts, it is not astonishing that policy makers are faced with the question how to deal with these inequalities. Economic geography in general and NEG in particular have gained attention due to the process of European integration and its consequences for regional inequalities (Fujita *et al.* 1999). In this context, human capital is clearly one important factor that has to be taken into account. In fact, when workers move from one location to another they do not only take with them the unit of labour that is available in the economy but also their incorporated human capital. The accumulation of human capital has important socioeconomic implications. It may possibly enable higher growth rates and lead to convergence or divergence processes. However, the incentives for individuals to invest in their human capital and their geographic location are not independent, as can be shown by the following NEG model.

8.4 Theoretical model

The proposed NEG model has originally been developed by Fujita *et al.* (1999). This model has two sectors, i.e., agriculture and manufacturing. However, the model does not take into account human capital accumulation. This factor has only been added by Redding and Schott (2003). Their model focuses on the interaction between human capital and input-output linkages, taking account of transport costs and assuming increasing returns to scale. One of their main results is that countries that are remotely located from main markets have to face higher trade costs and a decrease in the skill premium than other countries if one assumes that manufactures are relatively more skill intensive than agricultural goods. In this way, the effect of a remote location has the same consequences as a reduction in the relative price level of manufactures. Due to the assumption that the required skills in the manufacturing sector are higher than in agriculture, skilled workers face a fall in their relative wages. Thus, the incentive for an unskilled worker to invest in human capital and become skilled is decreased.

López-Rodríguez *et al.* (2005) have tested the model with recent data. Therefore, in line with López-Rodríguez *et al.* (2005), we will investigate econometrically the importance of market access in the long run later on in the paper. In the following, we first present the model by Redding and Schott (2003) in a similar way as put forward by the authors and López-Rodríguez *et al.* (2005). However, we adapt the model to the context of this paper by explicitly considering regions (instead of countries as in the original model).

First, we consider the preferences and the endowments that have to be modelled. Accordingly, Europe is constituted by $i \in \{1, \ldots, R\}$ regions. Every region is characterised by an endowment of L_i consumers. Every consumer has one single unit of labour. The supply of this unit of labour is inelastic, i.e., there is no disutility. The units of labour are initially unskilled but can endogenously decide whether they invest in order to become skilled. Consumer preferences are identical for all L_i. Consumption is restricted to two types of goods: first, the production of the agricultural sector is limited to one homogenous good. Second, the manufacturing sector produces a range of differentiated manufactures. The preferences follow a utility function in Cobb-Douglas form,

$$U_j = A_j^{1-\mu} M_j^{\mu}, \qquad 0 < \mu < 1 \tag{8.1}$$

where *j* denotes a region which demands (or imports) a good, *U* represents the utility function, *A* is the quantity that is consumed of the agricultural good, *M* denotes a consumption index which consists of varieties of differentiated manufactures and μ defines the share that is expended on manufactures.

The consumption index *M* is defined by

$$\begin{aligned} M_j &= \left[\sum_{i=1}^{R} \int_0^{n_i} m_{i,j}^c(z)^{(\sigma-1)/\sigma} dz \right]^{\sigma/(\sigma-1)} \\ &= \left[\sum_{i=1}^{R} n_i (m_{i,j}^c)^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \end{aligned} \tag{8.2}$$

where *i* denotes a region which produces (or exports) a good, σ denotes the elasticities of substitution between the pairs of varieties of manufactured goods. Note that consumers prefer more variety: an increase in σ leads to a rise in the substitutability among the varieties, engendering a decline in the preference to extend consumption over the differentiated manufactures. Moreover, n_i is the quantity of varieties which are produced in i and $m^c_{i,j}$ is the quantity of every variety that is produced in i for consumption in j. The second part of (8.2) takes advantage of the later result (see below) that in equilibrium all products and their corresponding quantities that region i produces are at the same time demanded by j. For this reason, we can take out the index z and in consequence the integral can be rewritten in the form of a product.

In addition to the aforementioned consumption index M_j we introduce G_j as a manufactures price index which takes into account the prices of the distinct varieties ($p^M_{i,j}$) that are produced in region i and that are sold in region j:

$$G_j = \left[\sum_{i=1}^{R} \int_0^{n_i} p^M_{i,j}(z)^{(1-\sigma)} dz\right]^{1/(1-\sigma)} = \left[\sum_{i=1}^{R} n_i (p^M_{i,j})^{(1-\sigma)}\right]^{1/(1-\sigma)}, \qquad (8.3)$$

where the second part takes advantage of the fact that we have symmetrical equilibrium prices.

Second, we define the production technologies involved in the two sectors. In the first sector, the produced agricultural good is homogeneous. Production is set within the framework of perfect competition and is characterised by constant returns to scale,

$$Y_i = \theta^Y_i (S^Y_i)^{\phi} (L^Y_i)^{1-\phi}, \qquad 0 \prec \phi \prec 1 \qquad (8.4)$$

where Y_i is the agricultural sector's output, L^Y_i is the quantity of unskilled workers in the sector, S^Y_i is the quantity of skilled workers and θ^Y_i is an index of agricultural productivity. For simplicity, Redding and Schott (2003) take the standard assumption that trade does not involve any costs. However, they relax this assumption later on to allow for shipping costs.

In the second sector, the production of the differentiated manufactured goods is characterised by increasing returns to scale and uses a combination of the two types of labour (skilled, unskilled) and of the intermediate inputs of manufactured goods. According to this setup, the cost function which a representative firm of region i has to face is

$$\Gamma_i = (w_i^S)^\alpha (w_i^U)^\beta G_i^{(1-\alpha-\beta)} c_i [F + x_i], \tag{8.5}$$

where w_i^S denotes the wage of workers that are skilled (and α their input share), w_i^U denotes the wage of workers that are unskilled (and β their input share), G_i denotes the price index that applies for manufactured goods (with $(1 - \alpha - \beta)$ its input share). Moreover, c_i stands for an inverse indicator of technological efficiency, $c_i F$ denotes the requirement of fixed inputs and $x_i = \sum_{j=1}^{R} x_{ij}$ denotes the overall output that the firm produces for all European markets. Note that the above mentioned combination takes again the form of a Cobb-Douglas function.

Furthermore, trade in manufactures involve iceberg costs, that is a part of the carried load is lost during transportation from i to j so that $T_{ij}^M > 1$ units of a good have to be shipped to ensure that one unit arrives in j. In consequence, $T_{ij}^M = 1$ means that trade does not incur any trade costs. T_{ij}^M is assumed to include all possible trade costs (e.g., transportation and communication costs).

In the next step, we introduce now endogenous investment in human capital into the model. Redding and Schott (2003) take the assumption that a conversion is possible from an unskilled to a skilled worker. Denoting an individual as z, this conversion incurs a fixed cost of education $\Omega_i(z)$ units in terms of unskilled labour. The underlying idea is that real resources are consumed to become skilled which results in the fact that the education cost is a proportion of the wage of unskilled labour. Moreover, the quantity of unskilled labour that is needed to become skilled is dependent on two factors. In particular, $\Omega_i(z) = \frac{h_i}{a(z)}$, where h_i denotes the overall environment provided by institutions and government policies that have repercussions on the education cost and $a(z)$ denotes the individual's personal ability. This ability ranges between the limits of a lower $(\underline{a})$ and an upper bound $(\overline{a})$ which are subject to human biology. Its cumulative distribution function is $\Lambda(a) = \int_{\underline{a}}^{a} \lambda(a)da$, where $\lambda(a)$ denotes its probability density function which is also assumed to be subject to human biology and equal for all regions.

Given the above mentioned assumptions, an individual z will only take the decision to invest in human capital if

$$w_i^S - w_i^U \geq \frac{h_i}{a(z)} w_i^U, \tag{8.6}$$

i.e., if education costs are lower than (or equal to) the difference between the wages of a skilled and an unskilled worker. The equation defines an implicit critical value for a above which all individuals choose to invest in human capital. This value a_i^* giving the supply of skills in equilibrium is

$$a_i^* = \frac{h_i}{({w_i^S}/{w_i^U} - 1)}. \tag{8.7}$$

An individual having the ability a_i^* does neither prefer to become skilled nor to remain unskilled but is indifferent to both options. Therefore, this equation is termed the 'skill indifference condition'. A graphical representation of the relationship between a_i^* and ${w_i^S}/{w_i^U}$ is provided by Redding and Schott (2003), see Figure 8.5. Only if an individual has an ability above a_i^* he will choose to get further education. For more intuition we additionally plot the corresponding relationship for ability and cost in education (Figure 8.6).

We consider next the consumer equilibrium. The consumer equilibrium is derived by the maximisation of consumers' utility. This utility is restricted by a budget constraint.

Denoting E_j^C the total expenditure on manufactures by consumers in j, the demand of consumers for varieties that have been produced in region i is given by (by application of Shepherd's lemma to G_j)

$$m_{ij}^C = (p_{ij}^M)^{-\sigma} E_j^C G_j^{\sigma-1}, \tag{8.8}$$

$E_j^C G_j^{\sigma-1}$ representing the market capacity ($m_j^C = E_j^C G_j^{\sigma-1}$) which measures the demand in j, i.e., the importing region. This market capacity comprises the total expenditure for manufactures in j (E_j^C, being constituted by intermediates and final consumption) and the price index for manufactures (G_j), the latter corresponding to the prices charged by the competing firms.

Finally, let us turn to the producer equilibrium. In the agricultural sector, profit maximisation combined with constant returns to scale give the condition that the unit cost of production is equal to the price, i.e.,

$$p_i^Y = 1 = \frac{1}{\theta_i^Y} (w_i^S)^{\phi} (w_i^U)^{1-\phi}, \tag{8.9}$$

where the numeraire is the homogeneous agricultural good, giving $p_i^Y = 1$ for all regions i.

Considering the manufacturing sector, profit maximisation by the representative region i firm is equal to the function

$$\Pi_i = \sum_{j=1}^{R} \frac{p_{ij}^M x_{ij}}{T_{ij}^M} - (w_i^S)^{\alpha} (w_i^U)^{\beta} G_i^{(1-\alpha-\beta)} c_i [F + x_i]. \tag{8.10}$$

As the varieties have equal weights in the above mentioned utility function, all firms in region i have an identical equilibrium price. To obtain this price one has to solve the first order conditions, resulting in

$$p_i^M = \left(\frac{\sigma}{\sigma - 1}\right) (w_i^S)^{\alpha} (w_i^U)^{\beta} G_i^{(1-\alpha-\beta)} c_i, \tag{8.11}$$

meaning that a firm takes advantage of a constant mark-up which is above the marginal cost. By substitution of (8.11) into (8.10), the profit function in equilibrium takes the following form:

$$\Pi_i = \left(\frac{p_i^M}{\sigma}\right)[x_i - (\sigma - 1)F]. \tag{8.12}$$

Thus, a firm breaks even (zero profits) given free entry into the market if its output is equal to the constant $\overline{x}_i = (\sigma - 1)F$. To sell this quantity of units, the firm has to set a price established by its demand function (8.4). Thus, we use (8.4) to obtain the break-even price for the quantity $\overline{x}$ sold by a firm in *i*:

$$(p_i^M)^\sigma = \frac{1}{\overline{x}}\sum_{j=1}^{R} E_j G_j^{\sigma-1}(T_{ij}^M)^{1-\sigma}. \tag{8.13}$$

We combine (8.13) with (8.11) to obtain the zero-profit condition

$$\left[\left(\frac{\sigma}{\sigma-1}\right)(w_i^S)^\alpha (w_i^U)^\beta G_i^{(1-\alpha-\beta)} c_i\right]^\sigma = \frac{1}{\overline{x}}\sum_{j=1}^{R} E_j G_j^{\sigma-1}(T_{ij}^M)^{1-\sigma}, \tag{8.14}$$

to which the literature refers as the ‘wage equation’. It defines the maximum level of wages that receive both categories of workers that firms in *i* are able to pay, taking account of the demand for its produced goods (in the equation on the right-hand side) and of the costs involved for the production of the intermediate inputs (left-hand side). In other words, the wage level in *i* is dependent on the weighted total purchasing power of all regions *j*. The weighting function declines the more the distances between *i* and *j* increase. This equation is also the beginning for the analysis of spatial wage structures in many contributions in the literature. We will call it the ‘market access’ of region *i*.

This equation can be taken as a function of the demand for labour in space under the condition of perfectly mobile labour. In fact, when the income of the neighbouring regions rises both the demand from labour and wages will increase. On the other hand, the demand and the wages decrease when transport costs rise to the considered locations. This leads to the fundamental result from New Economic Geography models that advantages in market access (or market potential) increase factor prices (Head and Mayer 2004). Thus, a good and easy access to important markets leads to the result that the concerned production sites recompense production factors (higher wages, higher land rentals) because they benefit from lower trade costs.

We can rewrite the wage equation as

$$(w_i^S)^\alpha (w_i^U)^\beta = \xi \frac{1}{c_i} (MA_i)^{\frac{1}{\sigma}} G_i^{(\alpha+\beta-1)}, \quad (8.15)$$

where ξ replaces the different former constants and MA_i is the market access of i. By using (8.3) we can represent G_i as $G_i = [SA_i]^{1/(1-\sigma)}$, so that the wages for the two types of labour can be expressed as

$$(w_i^S)^\alpha (w_i^U)^\beta = \xi \frac{1}{c_i} (MA_i)^{\frac{1}{\sigma}} (SA_i)^{\frac{(1-\alpha-\beta)}{(\sigma-1)}} \quad (8.16)$$

where SA_i is the supplier access of region i. In other terms, wages are dependent on both market and supplier access of i. Those regions that have a good access to the supply of manufactures (i.e., SA_i is high) have a lower index of manufacturing prices (G_i). For this reason, they are able to pay higher maximum wages because they face lower production costs than other regions. If the access to the markets for manufactures is also good (i.e., MA_i is high), they can increase the price for their goods while at the same time they are still able to sell sufficient output in order to face the given production costs. In this latter case, maximum wages which firms are able to pay are increased once again.

Combining equations (8.9) and (8.16) (i.e., the zero-profit conditions for agriculture and manufacturing, respectively) gives the wages that are paid to workers with and without skills in equilibrium. If we join equation (8.7), i.e., the skill indifference condition, it is possible to obtain the equilibrium relationship that exists between the geographical location of a region and endogenous investments in human capital. Thus, we have,

$$0 = \phi \frac{dw_i^S}{w_i^S} + (1 - \phi) \frac{dw_i^U}{w_i^U} \tag{8.17}$$

$$\alpha \frac{dw_i^S}{w_i^S} + \beta \frac{dw_i^U}{w_i^U} = \frac{1}{\sigma} \frac{dMA_i}{MA_i} + \frac{(1 - \alpha - \beta)}{(\sigma - 1)} \frac{dSA_i}{SA_i} \tag{8.18}$$

By means of (8.17) and (8.18) Redding and Schott (2003) show that if both the equilibrium MA_i and SA_i decrease, if the manufacturing sector is assumed to be skill intensive with regard to the agricultural sector and if the region is incompletely specialised, then the equilibrium moves to a new equilibrium with lower skilled wages but higher unskilled wages (see Figure 8.7), i.e., the isoprice line of manufactures is shifted inwardly. This implies that the critical ability level a_i^* increases. This change induces a lower supply of skilled and a higher supply of unskilled labour.

More specifically, the decrease of MA_i and SA_i has led to a smaller size of the skill intensive manufacturing sector. The reduction in size means that there are now more skilled workers in the market than there is demand for them in agriculture. Therefore, the wages of skilled workers decrease whereby their relative wages in comparison to the ones of unskilled workers fall.

In this way, remoteness leads to smaller incentives to invest in human capital. This means that the model predicts a positive relationship between market access and human capital investment.

8.5 Data and methodology

We test the theoretical model by the use of different datasets. We employ numeracy as a proxy for regional human capital formation in Europe in 1850. Numeracy is derived from the age heaping method. Age heaping as a method for calculating basic human capital values has been established by the rather recent literature (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010a, Hippe and Baten 2012a, b, Hippe 2012b). In particular, we use the ABCC Index to measure numerical abilities. This index has the advantage that it is comparable to other standard proxies because its range goes from 0 to 100, as is also the case for other share proxies. In fact, it measures the share of individuals that are able to calculate. More specifically, historical census data (and other sources), and in part even data for today's LDCs, show a clear pattern of rounding. Many people were not able to calculate their age. Therefore, they guessed their age to fulfil the census requirements set up by the state. Given that human biology serves as a first aid for calculations (e.g., five fingers on one hand, ten fingers in total), they rounded their ages on 0 and 5. It has been shown that this rough proxy is well correlated with other standard human capital proxies such as literacy (A'Hearn *et al.* 2009, Hippe 2012b) and primary school enrolment (Crayen and Baten 2010a).

The underlying formula of the ABCC Index is

$$ABCC_{it} = 125 - 125 \times \left(\sum_{k=5}^{14} n_{5k,it} \Big/ \sum_{k=23}^{72} n_{k,it} \right), \qquad (8.19)$$

where i denotes a region, k the number of years, n the number of individuals and t the time period.

The human capital data have been taken from the new and large database provided by Hippe and Baten (2012a). These data are based on original historical census data. The advantage of this measurement method is that it always takes into account the entire population and not, as other historical proxies of human capital (e.g., signature rates) only parts of it. For this reason, it is representative for the whole population and is not prone to biases that naturally reside in more partial indicators. In this way, we are able to measure the regional distribution of basic numeracy from Portugal to Russia. In total, there are 299 regions in our dataset (see Table 8.1 for descriptive statistics).[102]

Moreover, the data on urbanisation are provided by Bairoch *et al.* (1988). It is, alongside with a similar database by De Vries (1984), the standard database on urbanisation in the long run. In fact, the data trace back the cities of Europe until the year 800, starting from 1850. For a general geographical illustration of the data for 1850 see Figure 8.8. Cities are included if they fulfil a minimum threshold of population size between 800 and 1800. This threshold is 5000 inhabitants. In total, there are 2201 cities in our database. We excluded two observations because they were geographical outliers so that we have used the remaining 2199 cities for our calculations.[103]

Market access has been proxied by population potential in the recent literature (e.g., López-Rodríguez *et al.* 2005). Population potential is a standard way in geography of representing changes in the pattern in which cities are distributed in space. This method is much more precise than the mere comparison of maps. It enables to identify the relative location of a city within a greater network of other cities. Two factors are essential in the evaluation process: first, the size of the population of cities. Second, the distance of a city to the other regions in the network. In practice, one adds to the population size of a city the population sizes of the other cities, each time divided by their distances to the original city. This is done for every city in the data. In this way, a potential value is assigned to each city. To be more precise, the mathematical formula in correspondence to López-Rodríguez *et al.* (2005) is

[102] Given the variable "Distance to Luxembourg", we have excluded Luxembourg in all our regressions and do not list it here.

[103] These outliers are Ponte Delgada which is on the Azores Islands and far off the European continent. Moreover, we excluded Oral which is not located in the limits of today's definition of Europe.

$$MA_i = Po_i + \frac{Po_i}{D_{i,1}} + \cdots + \frac{Po_n}{D_{i,n}} = Po_i + \sum_{j \neq i, j=1}^{N} \frac{Po_j}{D_{i,j}}, \tag{8.20}$$

where MA_i stands for the market potential at i, Po_i is the population of i and $D_{i,j}$ is the distance that exists between i and j, each i and j representing individual nodes.

For the econometric specification of the relationship between investment in human capital and market access, we first test a standard OLS regression model as used by the literature. The basic framework is the following:

$$\ln(ABCC_i) = \beta_0 + \beta_1 \ln(MA_i) + \varepsilon_i, \tag{8.21}$$

where *ABCC* is the numeracy index (in logarithmic terms), *MA* is the market access (in logarithmic terms), i is a region and ε are the unexplained residuals. The basic OLS framework is later complemented by Tobit and Instrumental Variable regressions.

In addition, note that 'region' stands for a NUTS region in our case. NUTS is the official Nomenclature for Territorial Units of Statistics which has been developed by the European Union. It comprises all countries of the EU, EFTA and Candidate Countries of the EU. For countries outside this area, e.g., Russia, we used the current administrative division. This allows us to make our data comparable to current data and other research. Given the fact that market access and distance involves point data (cities and the central point of each region, respectively), the NUTS level can just be attributed without any further difficulties. The case is different for human capital data which were available only for the historical regions. In this case, we developed the correspondence of these historical regions to current regions as best as possible. Because we have often more detailed data than needed for this study (e.g., the *départements* in France, the *provincias* in Spain or the *Bezirks-Hauptmannschaften* in Cisleithania (the Austrian part of Austria-Hungary), the possible biases are importantly reduced because we can easily aggregate our data from province or county level. As a standard, we use NUTS 2 as the basic unit of analysis which is also the standard unit in most other contributions in our area.[104] In this way, we were able to create a unique dataset for the European regions in 1850.

8.6 Results

The calculated population potential values are illustrated in Figure 8.9 and for western Europe in greater detail in Figure 8.10. In the following we also refer to countries and regions in their current boundaries. It is apparent that the highest population potentials are found in the areas of Paris, London and Manchester and the wider locus until the western parts of Belgium. Given the size of the aforementioned cities, in particular of Paris and London, this is not surprising because these were the two most populated cities of Europe in 1850. Still, the figure highlights that they were not isolated from other population hubs but were the centre of a greater accumulation of population in western Europe. This can be explained by the long-term geographic change of economic importance from northern Italy to this area, as has also been postulated by Braudel (1979). This is also in line with the concept of the existence of a "blue banana" which has been put forward by Brunet (2002), a concentration of population and economic active from northern Italy over the course of the Rhine River until the UK and even Ireland.

In general, the more one distances oneself from the centre in western Europe, the lower are the potential population values. Going farther away from the centre, the highest estimated values are located in the regions of the UK, France, Germany, Switzerland, Italy, parts of Austria and a part of Spain (Cataluña). Polish regions are already in the next level. Nevertheless, there are some outliers to the overall rule. Large cities create their own high local population potential which explains the different shading in the areas of e.g., Madrid, Hamburg, Berlin, Prague, Vienna, St. Petersburg and Moscow.

[104] An exception is Greater London, where we had to use the NUTS 1 level due to unavailability of more disaggregated data.

In the next step, we investigate the relationship between market access and human capital. To this end, we plot market access against the ABCC (Figure 8.11). Unfortunately, the ABCC has already achieved its maximum level of 100 in several countries. This is why there are a number of regions that are limited by the upper bound. Nevertheless, there is a clear relationship between market access and the ABCC. Outlying regions are in particular Greater London (UKI) at the very right and Albania at the bottom of the figure.

To test this relationship econometrically, we performed different regression models. First, we began with standard OLS regressions. The results are shown in Table 8.2. Market access has a highly significant positive influence on the ABCC (column 1).[105] To compare our results for market access with distance to Luxembourg as proposed in the literature, we also computed this distance (in natural logarithm) and show the results in column 2 (and also in the subsequent steps). Distance to Luxembourg is negatively significantly correlated to the ABCC. To avoid biases, we also include dummies for the most important outlying regions in our data as López-Rodríguez *et al.* (2007). These regions are London (with by far the highest market potential) and Albania (with by far the lowest ABCC) (column 3). Both London and Albania are negatively significant. The case for Albania is clear due to its very low ABCC. London is at the top of the market potential scale. Still, it has not achieved the maximum attainable level of numeracy, which one may expect from its substantial market potential. Therefore, its coefficient is negative but not very large. The inclusion of these outliers improves considerably the explanation of the ABCC, as the increase in the (adjusted) R^2 underlines.

However, we have seen in the scatter plot that there are a number of regions that have already achieved the upper bound of 100 ABCC points in 1850. This given upper

[105] Note that we have opted for the presentation of the results with the logarithmic form of the ABCC. We have also done all regressions without this transformation and obtained the same results (only the value of the coefficients changed which is a logical consequence of the transformation).

limit may bias our results because some of these regions would have had higher numeracy values. For this reason, we take this fact explicitly into account by running the same regressions with the Tobit model. The Tobit model incorporates the problem of upper or lower bounds in its estimations. The lower bound is not important in our case but the upper limit is. Thus, in total, there are 41 regions which are right-censored by the model.[106] The results when using the alternative Tobit model are shown in columns 5 to 8. Globally, there are not any big differences to our OLS estimates and, therefore, the Tobit model confirms the robustness of our former results.

Nevertheless, it is still possible that our results are biased by endogeneity. In fact, one can imagine that market access is correlated with alternative variables that may have a significant influence on numeracy. Thus, to be able to identify whether there is causality between market access and numeracy, we also perform instrumental variable regressions. In the given case, an instrumental variable has to be a determinant of market access but also has to be exogenous to numeracy. Moreover, the variable should not be prone to influences of another underlying variable which may drive its values and affects both market access and numeracy.

As López-Rodríguez *et al.* highlight, “geographic variables seem to be the most adequate candidates for such an instrumental variable estimation” (López-Rodríguez *et al.* 2007, p. 223). Thus, in line with Redding and Venables (2004), Breinlich (2006) and López-Rodríguez *et al.* (2007), we take the distance from Luxembourg as our first instrumental variable. This variable captures the advantages conferred by being close to the centre of Europe. Second, as proposed by the same authors, we use the (area) “size of a region’s home country” (López-Rodríguez *et al.* 2007, p. 223), capturing the advantages

[106] Because we use the logarithmic form of the ABCC here, the upper limit (corresponding to 100) is approximately 4.6052.

that are created by big national markets for the market access of a region.[107] The results are shown in column 9. The IV estimates are once again highly significant and show the expected signs. This confirms once more the importance of market access for numeracy.

However, one may wonder if our results are robust to the use of other human capital variables and other time periods in the past. One standard indicator for human capital in the past (and still in today's developing countries) is literacy. For this reason, our results would need to be confirmed by the use of this alternative indicator. However, European wide data for literacy is only available around 1900 onwards but the earliest data on cities or, in this case, agglomerations (taken from Moriconi-Ebrard 1994) after 1850 are only available for 1950. Therefore, we use literacy data from 1930 (by Kirk 1946) and take as the best approximation of market access in 1930 data on European agglomerations in 1950.[108]

We take the same approach as for numeracy in 1850 and find that the agglomeration data appear to be quite similar around 1950 (see Figure 8.12 to Figure 8.14). The relationship between market access and literacy is also evident (see Figure 8.15). Note that there are no literacy data for several developed countries in 1930 such as the Scandinavian countries, Germany or the UK. Kirk (1946) estimates that these countries had literacy rates between 95 and 100. In the following, we exclude the regions from these countries (as has been done in Figure 8.15). Alternatively, we can also take the hypothesis that these regions had a literacy rate of 100. In any case, there are no apparent outliers. Therefore, we do not include additional dummies as in the previous ABCC analysis. The

[107] Borders and countries in ca. 1850 are considered. Because we are interested in the domestic market and trade advantages, we consider Germany as being constituted by those countries that had joined the *Zollverein* (German Customs Union). Data on country sizes (in geographical square miles) come from *Annuaire Statistique et Historique Belge* (1857).

[108] We are very well aware of the fact that World War II affected important portions of regional populations which may have a biasing effect on our estimates. However, authors such as Martí-Henneberg (2005) show that population concentrations are highly correlated at the regional level between 1870 and 2000 which suggests that data from 1950 are still a good approximation for 1930.

corresponding regression results are shown in Table 8.3. The coefficients are higher than in 1850 which can be explained by the fact that literacy rates are more dispersed than numeracy rates. The exclusion of the developed countries without any official literacy data (column 1 to 3) or their inclusion (column 4 to 8) does not affect the significance of the results. Market access appears to explain significantly human capital in every model.

All in all, we find a core-periphery pattern in Europe similarly to the literature that analyses the EU today. Market access has a significant influence on human capital, confirming the 'penalty of remoteness' hypothesis. Moreover, because we are referring to the rather distant past with our data, the current regional distribution of human capital and economic development appears to be rather stable in the longer run.

This gives important implications for regional policy. Remote regions need to obtain better access to the main European markets. For this reason, it appears to be essential to advance improvements in infrastructure and focus even more on investments in human capital.

8.7 Conclusions

This paper has analysed the importance of market access to explain the spatial distribution of human capital levels in the European regions in the past. The central question is if remoteness was connected to backwardness in the past, as has been postulated by Redding and Schott (2003) and tested by e.g., López-Rodríguez *et al.* (2005) for the present.

In particular, we used the age heaping method in order to approximate numeracy values for 1850. Moreover, data on European cities have been used to proxy for market access. In this direction, the standard concept of population potential has been employed to generate average market access estimations for the European regions.

The results show that market access is highest in the regions constituting England, northern France, Belgium and the Netherlands. In general, the farther away one goes from this centre, the lower is the market access. Therefore, we find a core-periphery structure also in the past.

Moreover, OLS, Tobit and IV regressions of market access on numeracy highlight that numeracy is significantly higher in regions with higher market access. We also control for outlying regions which improve the explanatory power of the model. Further robustness checks have been performed by using literacy data for 1930. Thus, after the literature has in particular used educational attainment for the current period, our numeracy estimates show that the 'penalty of remoteness' hypothesis is not only valid for today but that its importance can be traced back even to the middle of the 19th century. This underlines once more that this penalty has existed for a long time in Europe. Thus, it may continue to exist also in the future if not the right policy decisions are taken. Improved infrastructure and greater incentives for investment in human capital appear to be very important in this context.

8.8 Appendix

8.8.1 Tables

Table 8.1 Descriptive statistics for ABCC and market access, ca. 1850

Variable	obs.	mean	sd	min	max
ABCC	299	90.81	12.27	26.38	100.00
Market access	299	4465.12	2200.10	995.13	17165.47
Distance	299	14.19	11.36	0.61	51.50

Table 8.2 Market access and ABCC, ca. 1850

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable	ln(ABCC)									
ln(MA)	0.17***		0.17***		0.18***		0.18***		0.19***	0.19***
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	(0.000)
Dummy London			-0.18***	0.03***			-0.21***	0.00		-0.21***
			(0.000)	(0.000)			(0.000)	(0.679)		(0.000)
Dummy Albania			-1.20***	-1.18***			-1.21***	-1.19***		-1.19***
			(0.000)	(0.000)			(0.000)	(0.000)		(0.000)
ln(Dist. to Lux.)		-0.10***		-0.09***		-0.12***		-0.11***		
		(0.000)		(0.000)		(0.000)		(0.000)		
Constant	3.12***	4.72***	3.12***	4.71***	3.01***	4.78***	3.02***	4.78***	2.93***	2.94***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estimation	OLS	OLS	OLS	OLS	Tobit	Tobit	Tobit	Tobit	IV	IV
Observations	299	299	299	299	299	299	299	299	299	299
Prob. (F-Statistic)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.27	0.25	0.44	0.42					0.27	0.45

Note : ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

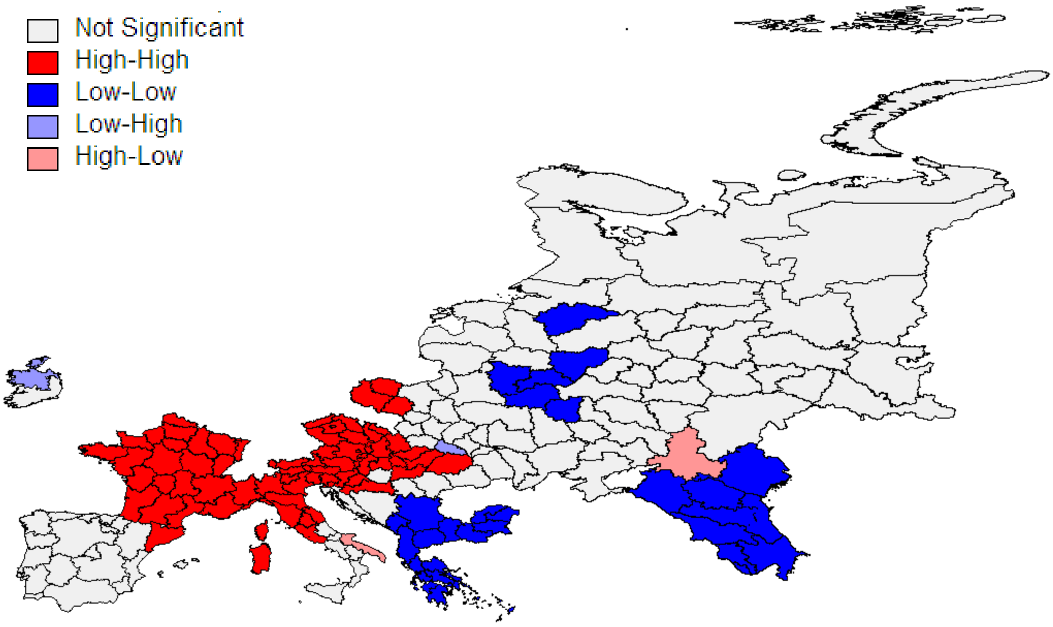
Table 8.3 Market access and literacy, ca. 1930

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable				ln(Literacy)				
ln(MA)	0.56***		0.57***	0.51***		0.77***		0.57***
	(0.000)		(0.000)	(0.000)		(0.000)		(0.000)
ln(Dist. to Lux.)		-0.24***			-0.24***		-0.36***	
		(0.000)			(0.000)		(0.000)	
Constant	-5.72***	0.25***	-5.84***	-5.21***	0.32***	-7.62***	0.72***	-5.78***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estimation	OLS	OLS	IV	OLS	OLS	Tobit	Tobit	IV
Obs. excluded	Estimated	Estimated	Estimated	None	None	None	None	None
Observations	201	187	186	327	306	327	306	301
Prob. (F-Statistic)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.44	0.38	0.49	0.50	0.44			0.56

Note : ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

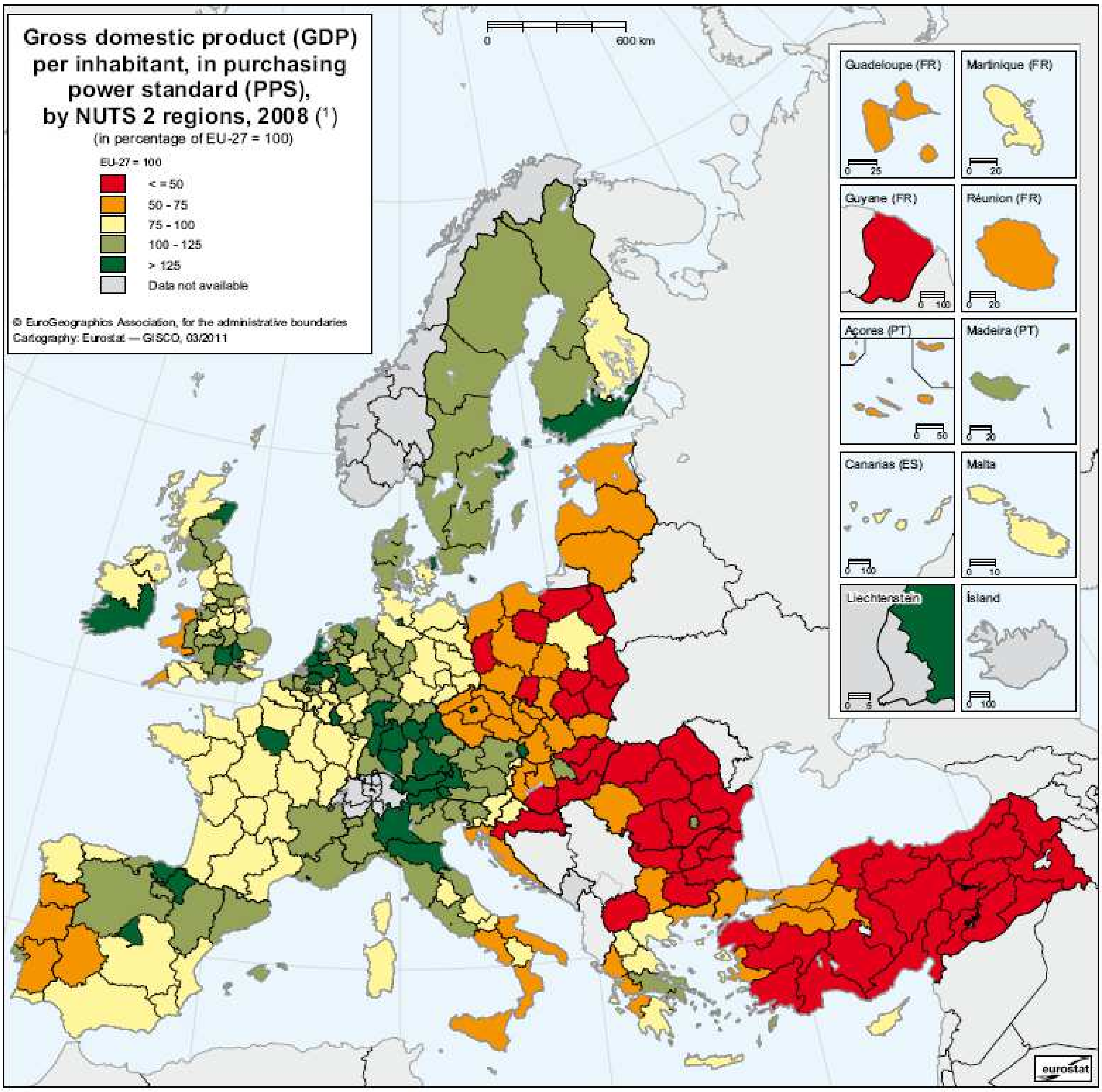
8.8.2 Figures

Figure 8.1 Moran significance map for ABCC in Europe, ca. 1850



Source: Hippe (2012c).

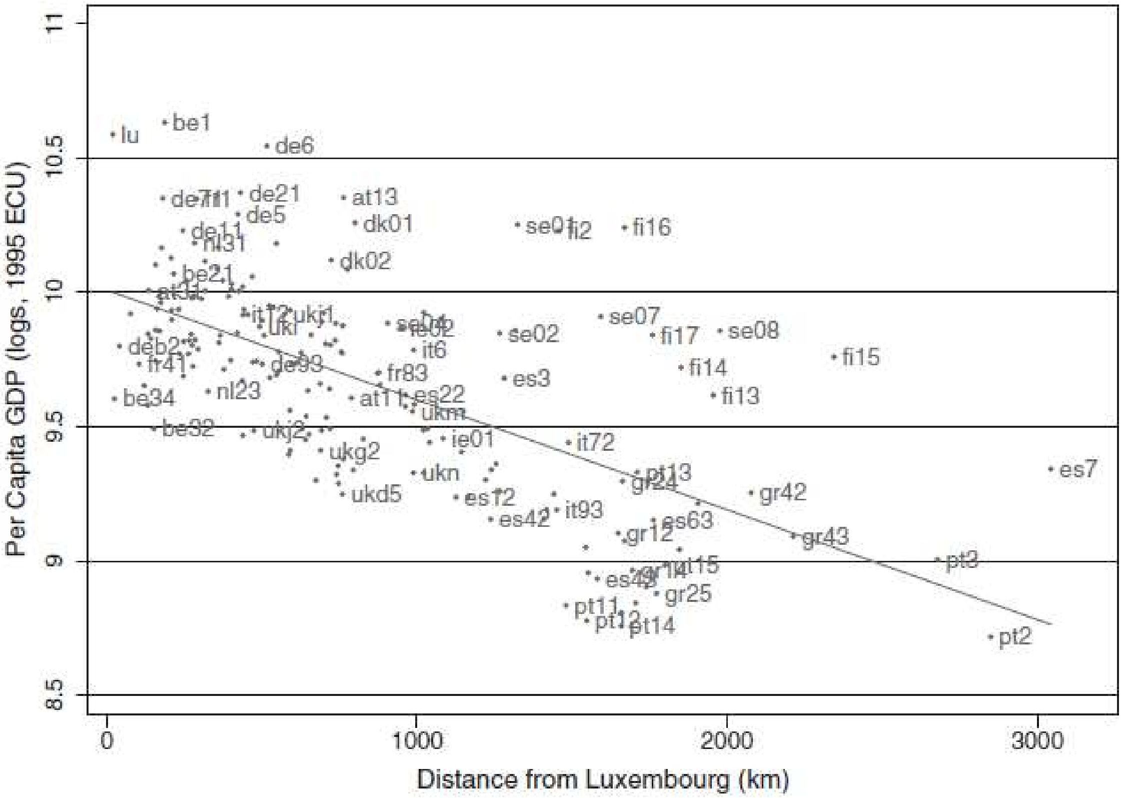
Figure 8.2 Regional GDP per capita (in PPS), 2008



(1) Turkey, 2006.

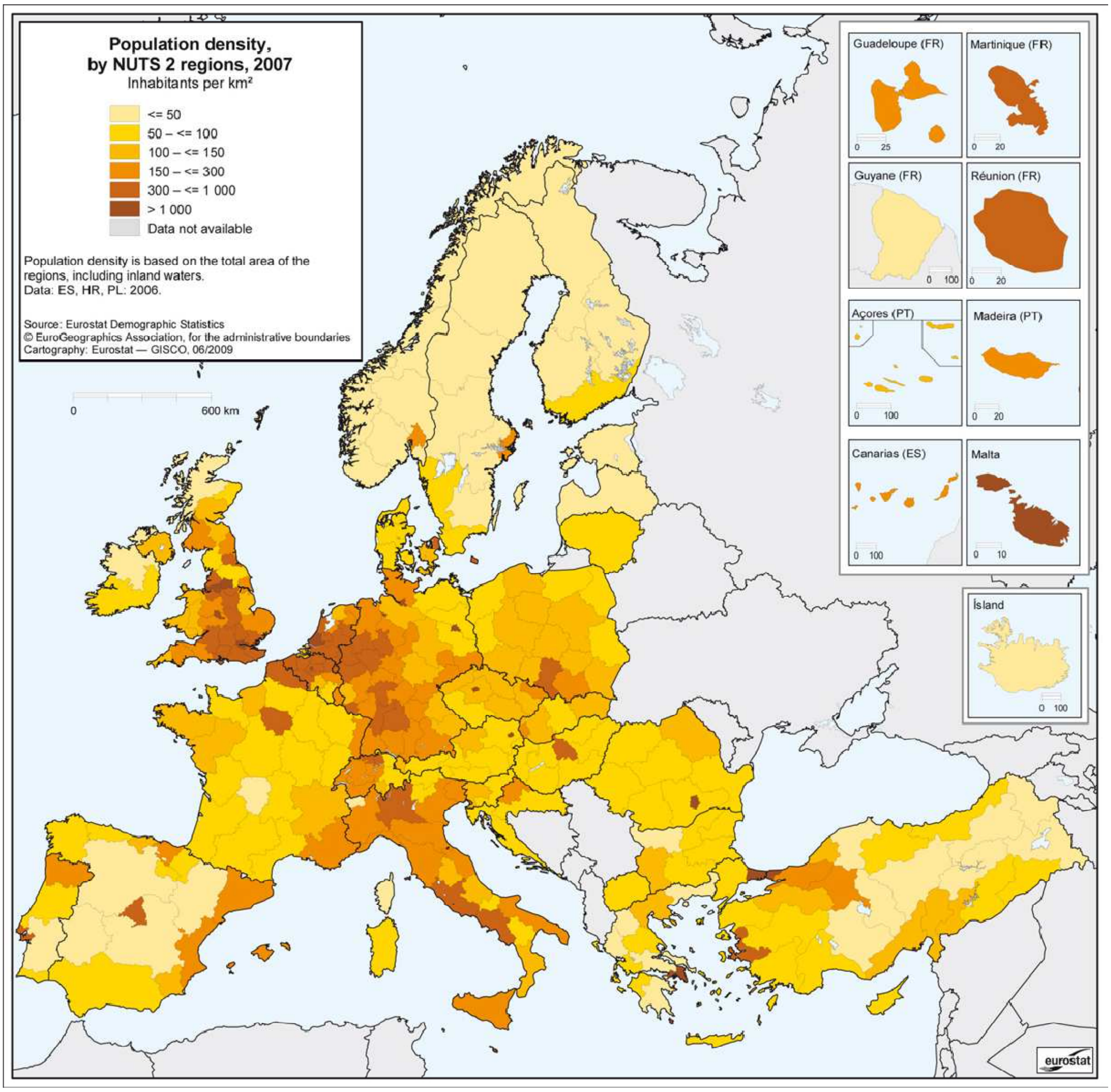
Source: Eurostat (2011f); http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 8.3 GDP per capita and distance from Luxembourg in EU, 1999



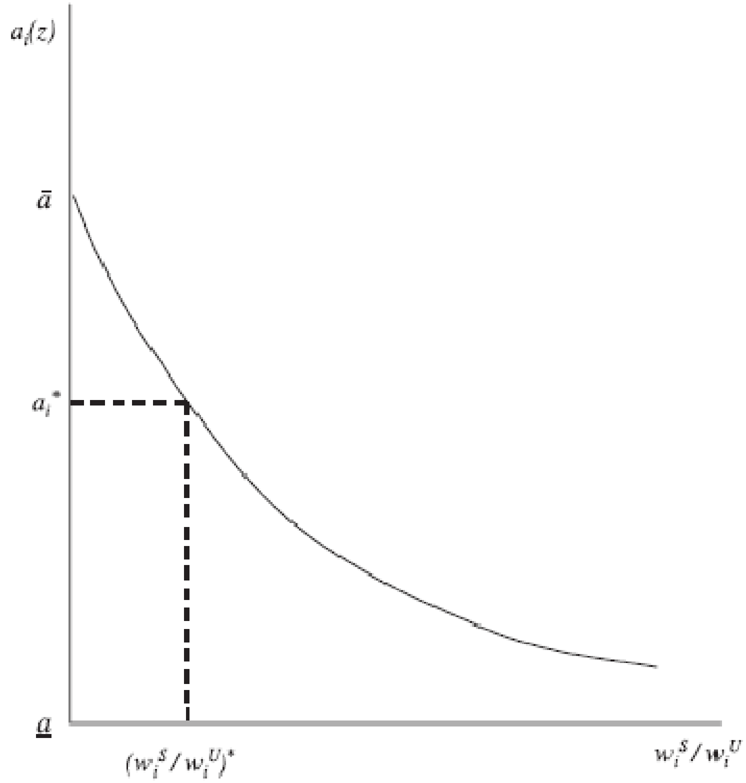
Source: Breinlich (2006), by permission of Oxford University Press.

Figure 8.4 Regional population density, 2007



Source: Eurostat (2009), © European Communities, 2009; Eurostat, http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Figure 8.5 Relationship between ability and relative wage



Source: Reprinted from Redding and Schott (2003), Copyright (2003), with permission from Elsevier.

Figure 8.6 Relationship between ability and cost in education

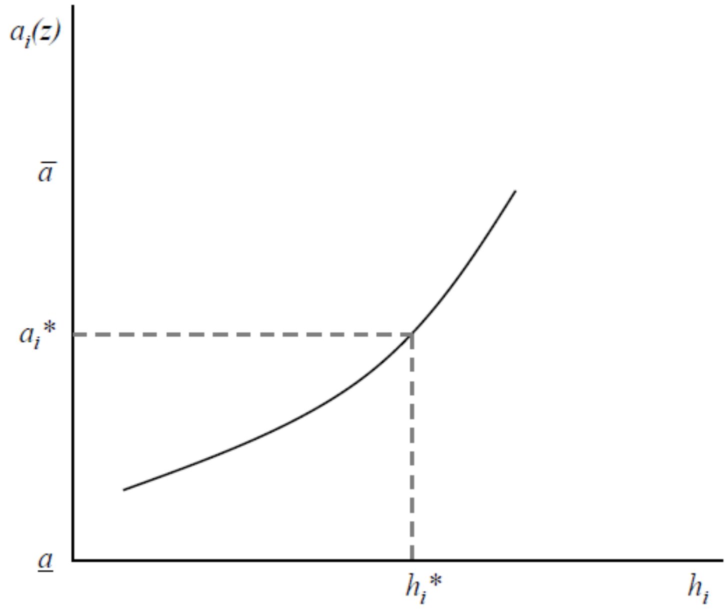
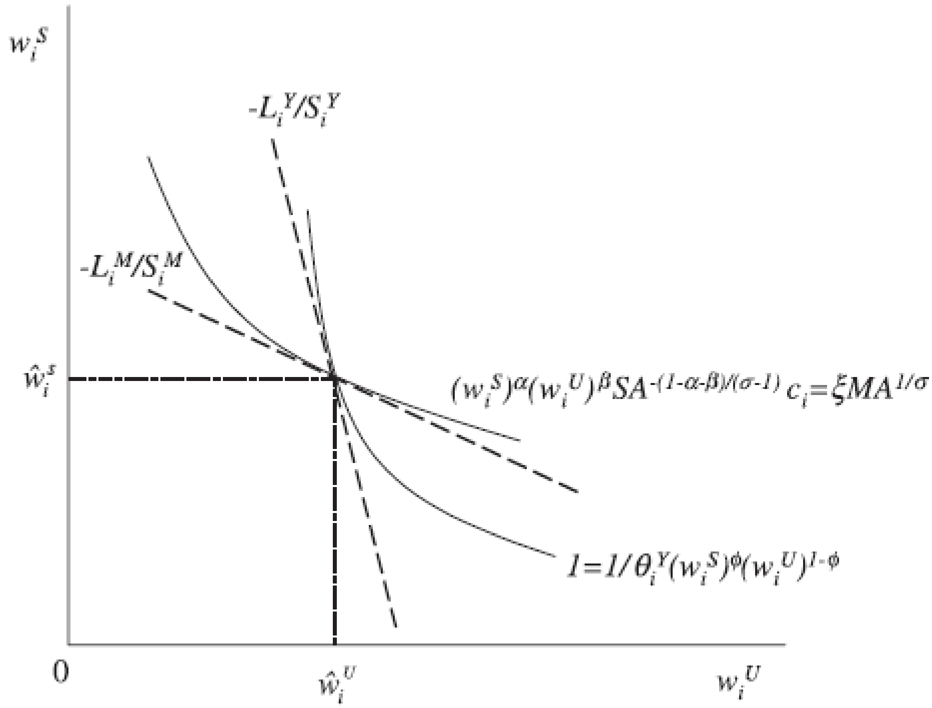
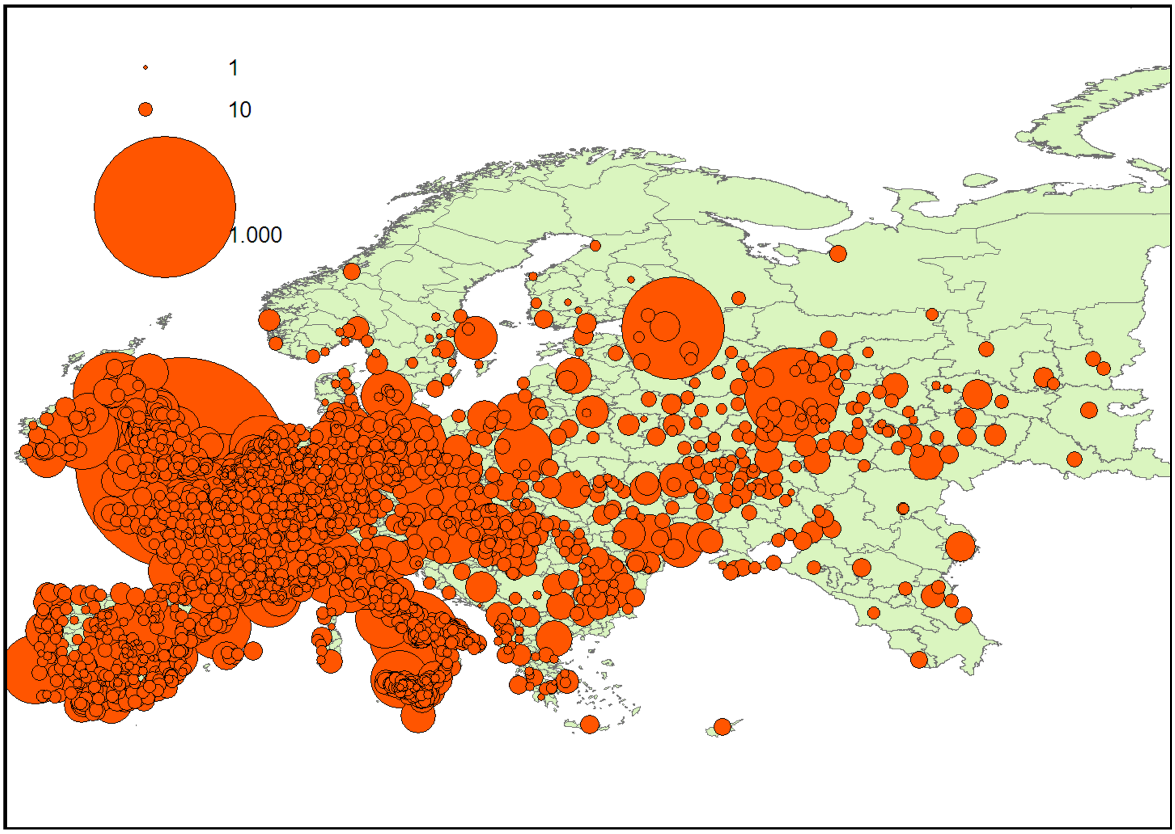


Figure 8.7 Equilibrium wages (skilled, unskilled) and relative unit factor requirements



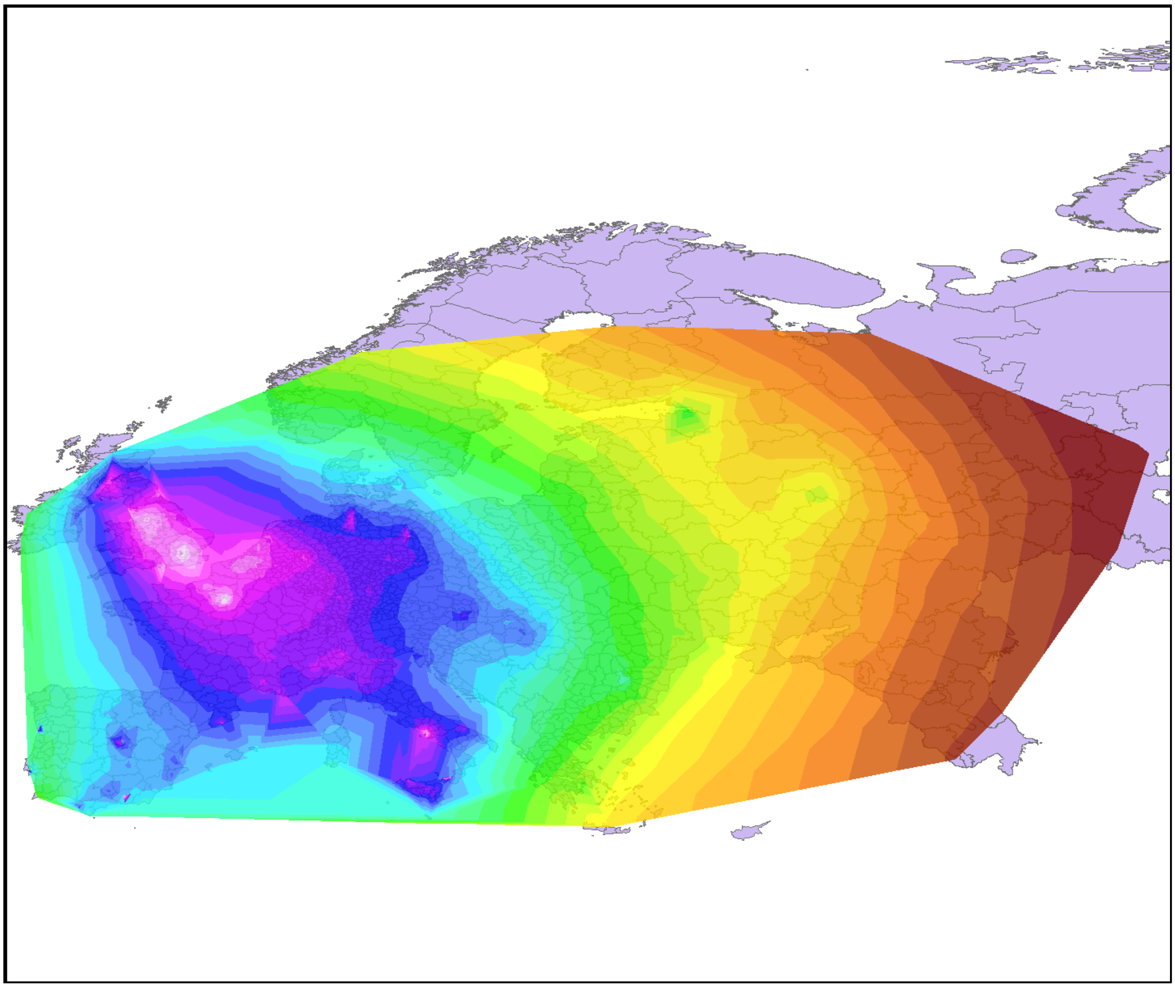
Source: Reprinted from Redding and Schott (2003), Copyright (2003), with permission from Elsevier.

Figure 8.8 Location and size of European cities, 1850



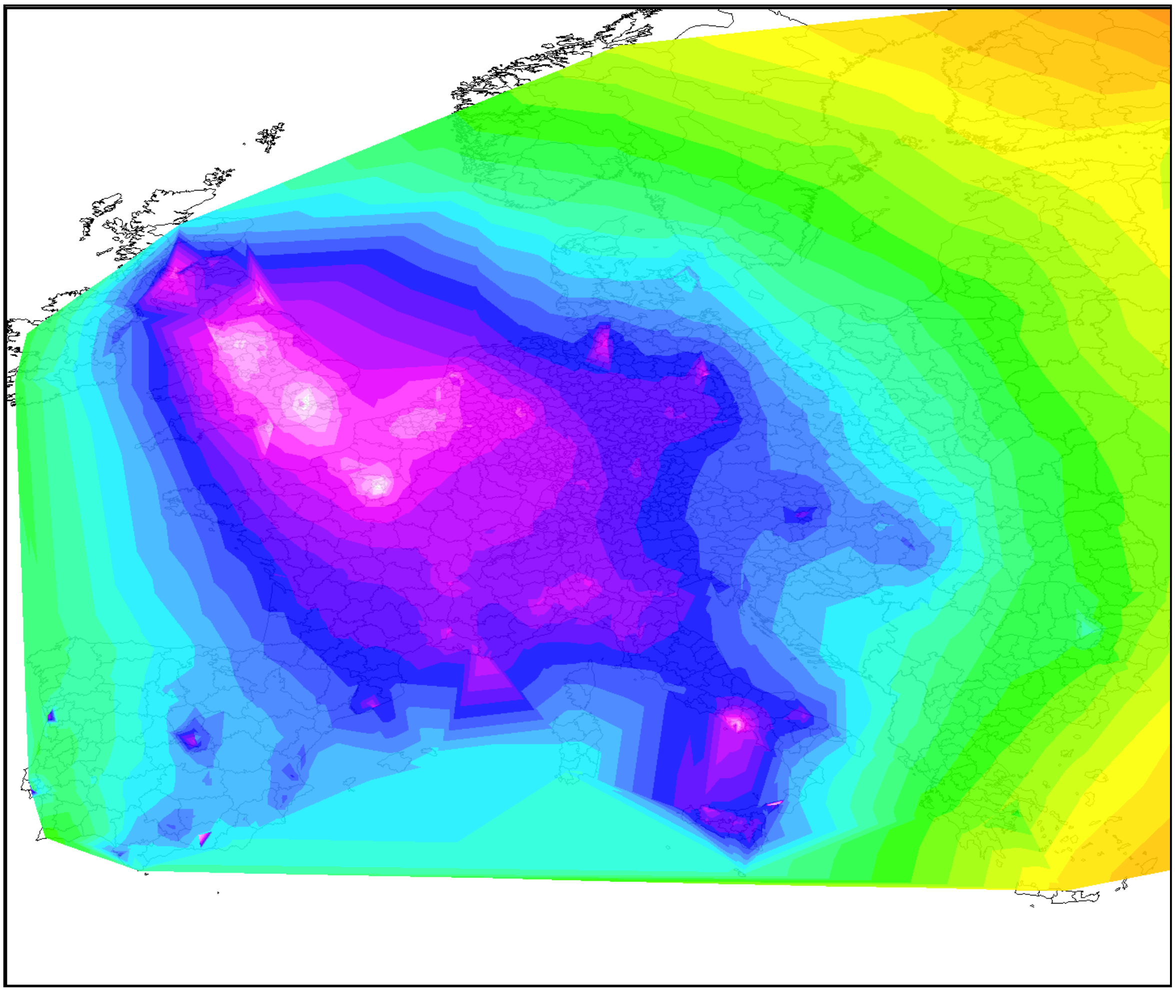
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988). Size of cities is shown in thousand inhabitants.

Figure 8.9 Population potential in Europe in 1850



Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, city data provided by Bairoch *et al.* (1988).

Figure 8.10 Population potential in Europe in 1850, zoom to western Europe



Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, city data provided by Bairoch *et al.* (1988).

Figure 8.11 ABCC and market access, 1850

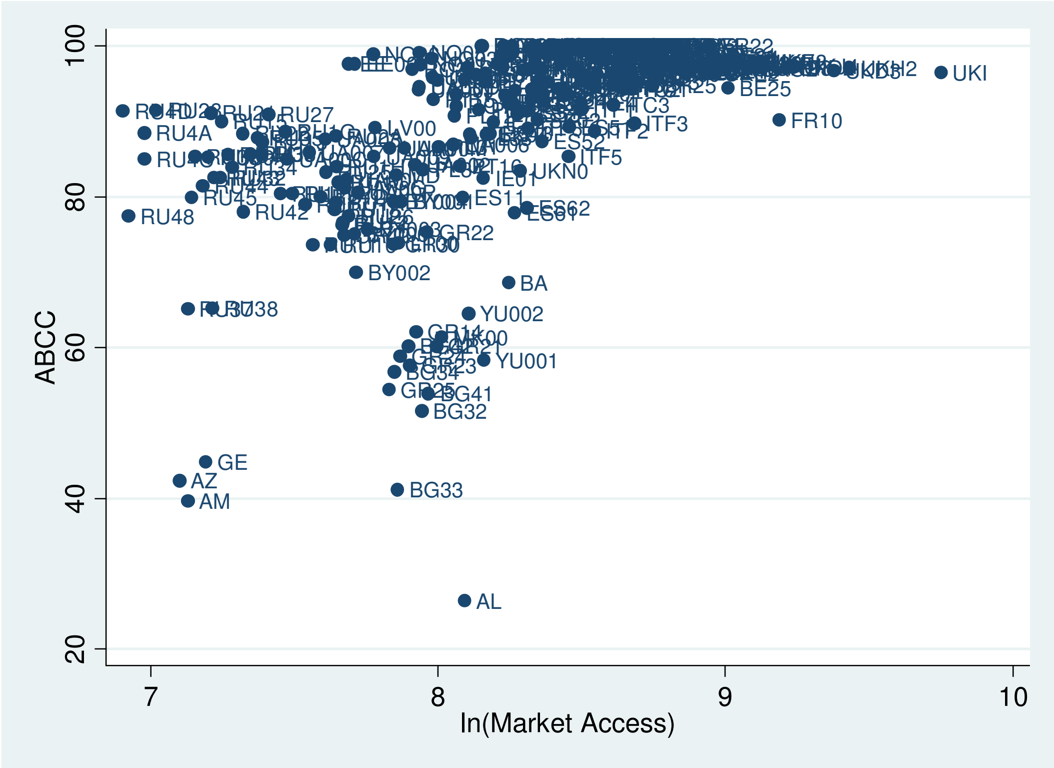
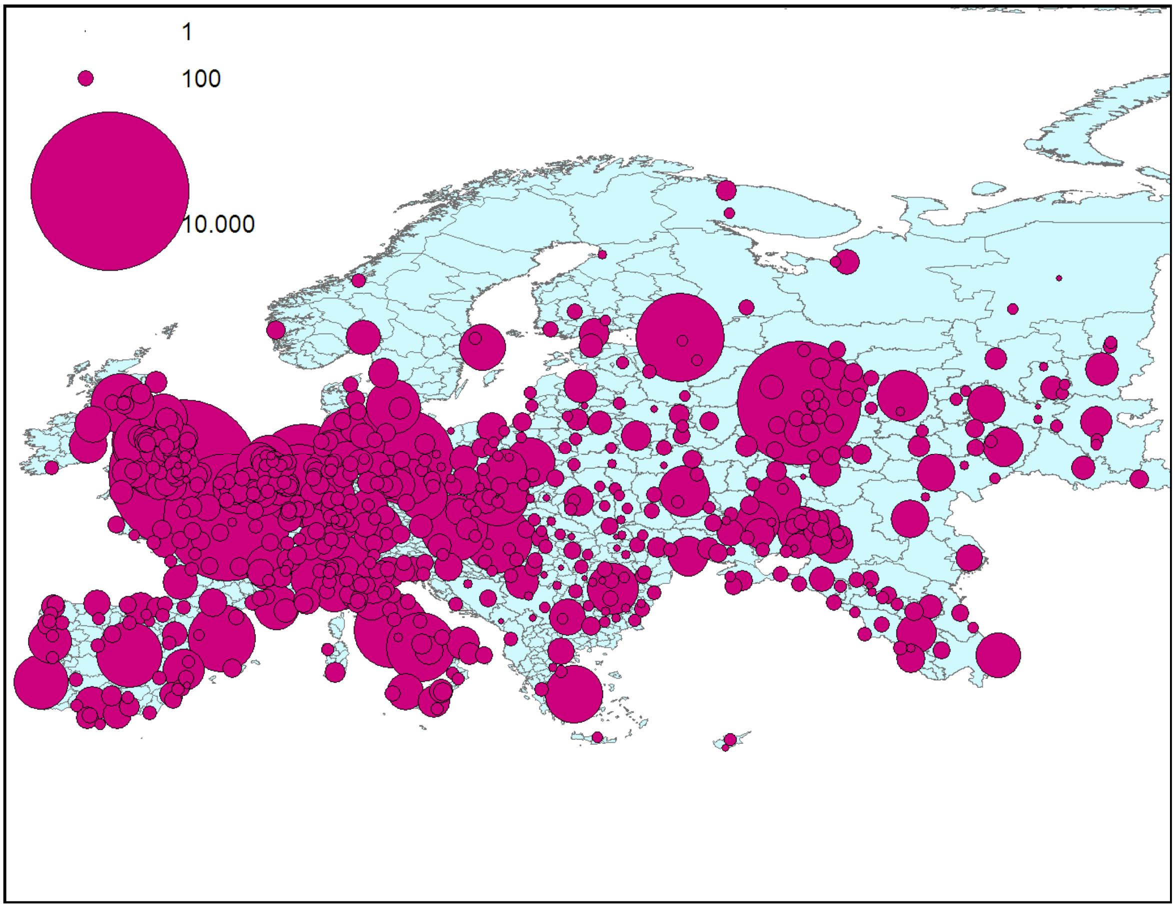
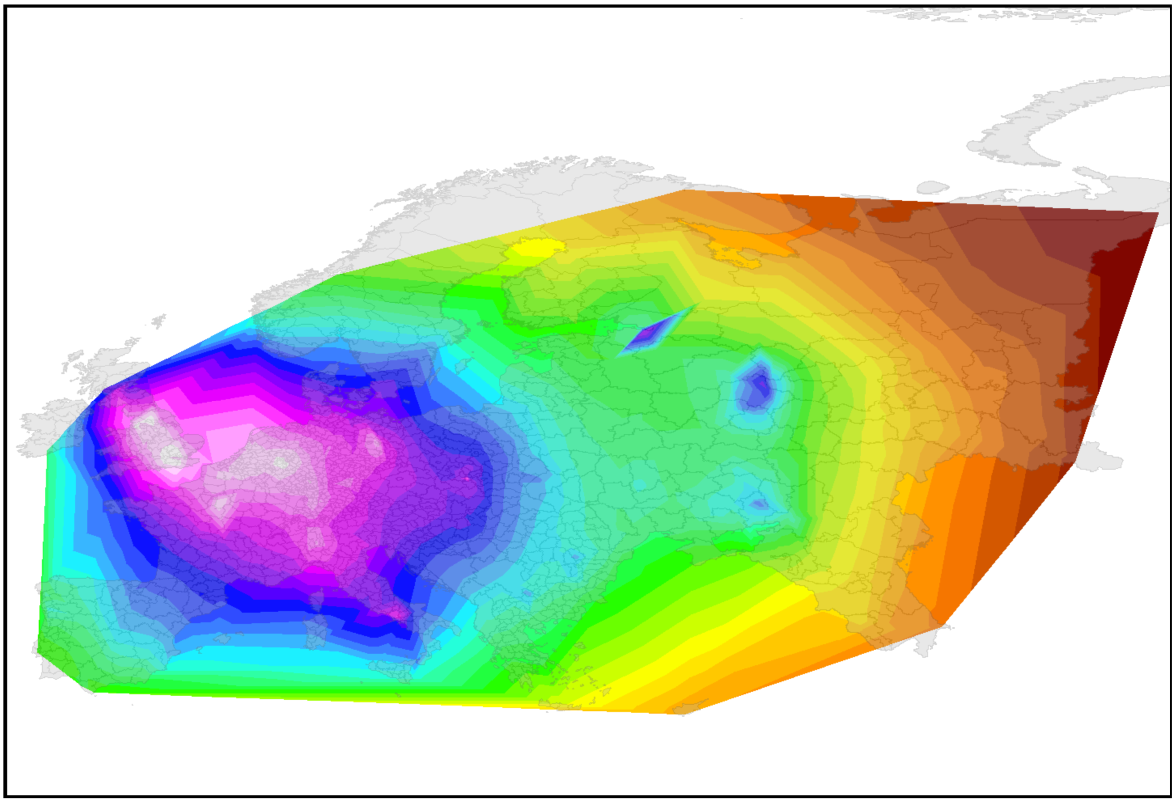


Figure 8.12 Location and size of European agglomerations, 1950



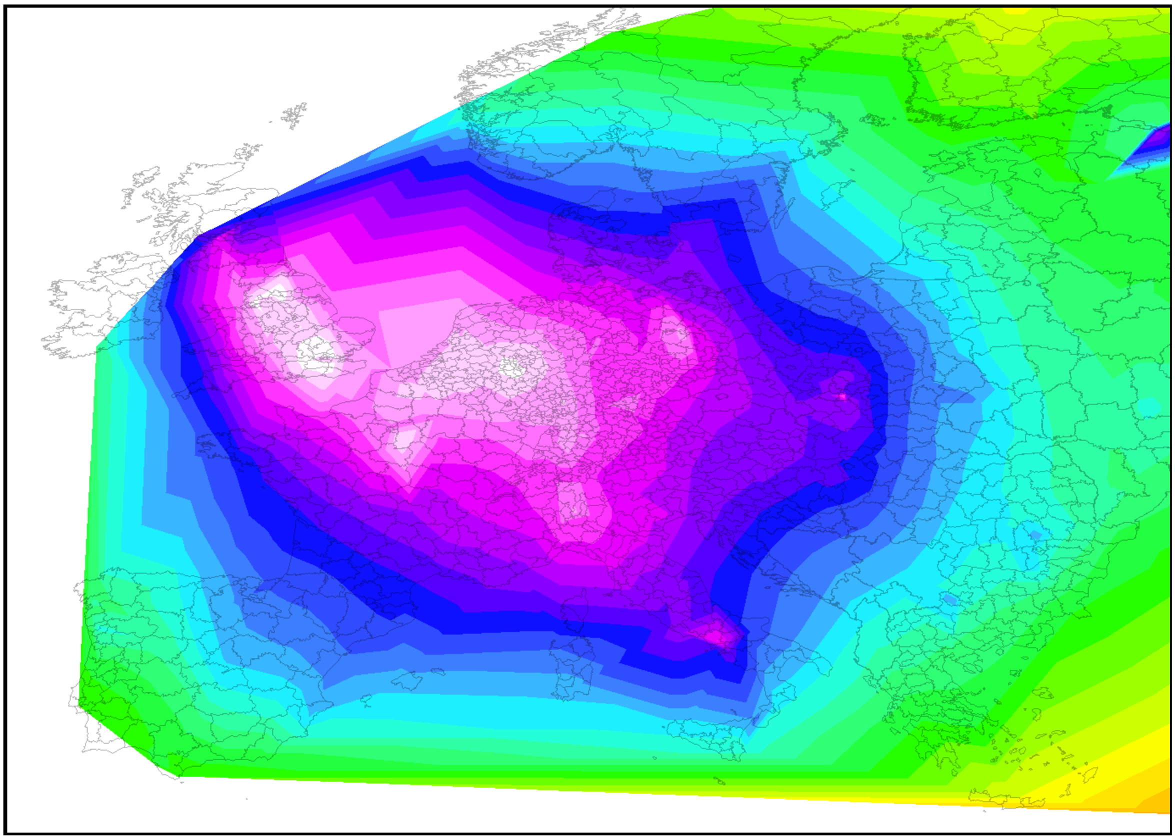
Source: Own graphical presentation of data provided by Moriconi-Ebrard (1994). Size of agglomerations is shown in thousand inhabitants.

Figure 8.13 Population potential in Europe in 1950



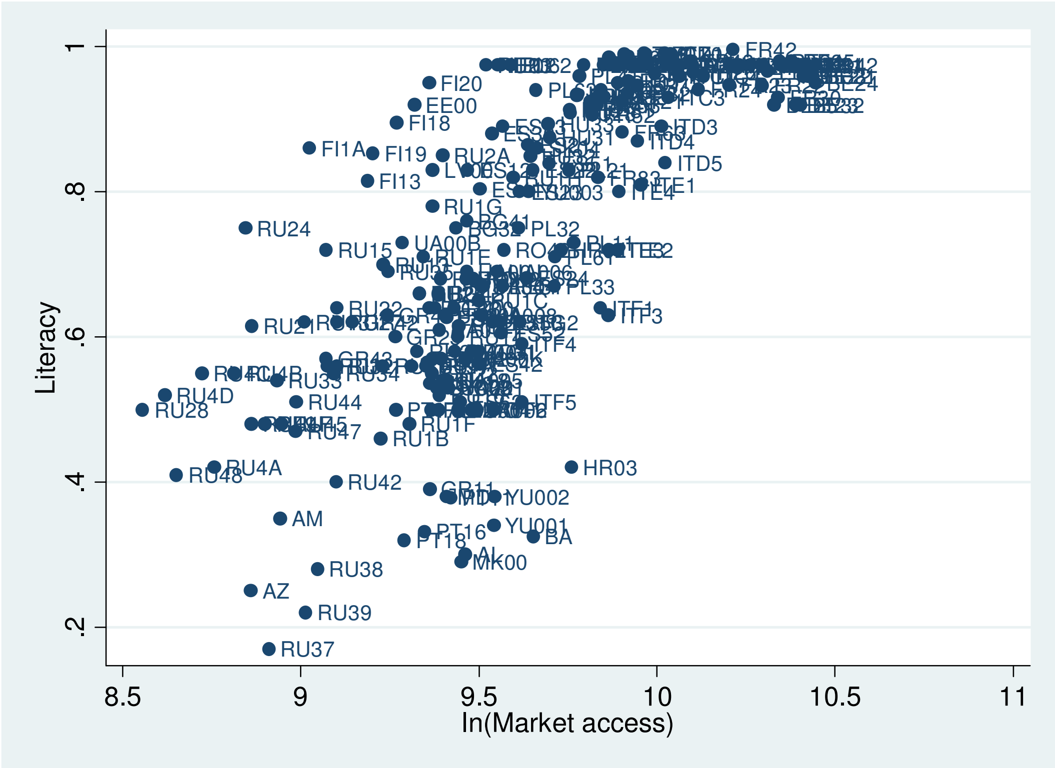
Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, data on agglomerations provided by Moriconi-Ebrard (1994).

Figure 8.14 Population potential in Europe in 1950, zoom to western Europe



Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, data on agglomerations provided by Moriconi-Ebrard (1994).

Figure 8.15 Literacy and market access, ca. 1930



9. Regional human capital formation in Europe in the long run, 1850 – 2010

Human capital is an important factor for economic and social development, as has been underlined by recent theoretical models. A range of contributions has focused on the international evolution of human capital over the last decades and beyond. However, the regional dimension of human capital in Europe remains insufficiently explored, particularly in a long-run perspective. For this reason, this paper addresses this gap in the literature and highlights the regional evolution of human capital in Europe between 1850 and 2010 by using numeracy, literacy and educational attainment proxies. The results show that intranational inequalities in human capital have always been important and are often more important than international differences.

This chapter is based on a working paper which has been submitted to an academic journal.

9.1 Introduction

Human capital has obtained considerable attention from both researchers and public policy makers recently and the in more distant past. Human capital is often assumed to positively affect a variety of socio-economic factors such as economic development (Lucas 1988, Romer 1990, Galor 2005a, b, Galor 2012), democracy and human rights (McMahon 1999, Sen 1999, Beach 2009).

Nevertheless, many authors in the human capital literature concentrate either on recent developments or on the evolution in history. Still, there are some studies that establish a link between historical and more recent data and that try to follow the development of human capital in the longer run. However, these studies typically take as the basic unit of analysis either the country level (e.g., Barro and Lee 2001, Morrisson and Murtin 2009) in order to make international comparisons or they take the regional level and in one country (e.g., Felice 2012).

In fact, studies that focus on the regional level in Europe *and* take a long-term approach, bringing together current data with distant historical ones, are rare. Yet this appears particularly important given the fact that human capital plays an important role for society and for economic development in the long run.

For this reason, this paper explores the long-term evolution of human capital in Europe at the regional level. We employ different indicators for subsequent periods of time. In particular, we use three proxies: numeracy, literacy and educational attainment. These proxies have some important characteristics in common which make them particularly appropriate for the study of their time period, thus allowing an evaluation of human capital throughout time. In this way, we can underline the evolution of human capital by taking into account the years 1850, 1900, 1930, 1960, 2000 and 2010. In addition, we address the question on the evolving inequality in human capital, employing two standard measures, the coefficient of variation and the Gini coefficient. For a better comparison, we also adapt historical regional boundaries to current NUTS 2 regions. In total, we have between 160 and 340 NUTS 2 regions in our database for the different years.

The results show that intranational inequalities in human capital have been important at the different points in time. Regional differences are in many cases quite persistent and are often higher than international ones. Moreover, the coefficient of variation and the Gini coefficient indicate important variations in inequality between countries throughout time. These findings underline the limitations of cross-country analyses and the need for further human capital research at the regional level in Europe.

The paper is structured as follows: first, we highlight some of the most important contributions in the human capital literature that make long-term comparisons or trace the long-term evolution of human capital in Europe. The second part explains the basic underlying methodology. The third section portrays the different data sources that have been used in this study. Subsequently, the results on the regional evolution of human capital and on the differences in regional inequalities are highlighted. Finally, a conclusion sums up the paper.

9.2 Human capital formation in Europe in the (very) long run

Human capital has been emphasised to be a crucial factor to improve the lives of individuals (e.g., Vincent 2000). For example, endogenous growth models have stressed the important role of human capital (Lucas 1988, Romer 1990). Furthermore, a long-run view on economic growth has been proposed by Unified Growth Theory (e.g., Galor 2005a, b). Unified Growth Theory highlights that human capital is essential for the creation of long-run growth. However, human capital is a theoretical concept that cannot be measured easily empirically. This is especially true in the long run. Nevertheless, the literature has put forward different proxies for the long-run formation of human capital in Europe. In particular, it is possible to trace its evolution in Europe by the use of numeracy, literacy and book production.

First, A'Hearn *et al.* (2006) give some first indications on numeracy formation in Europe in the very long run (see Figure 9.1). They calculate numeracy values for some European countries or larger regions (i.e., the Alps) and the United States up to Roman times by employing the age heaping method. More specifically, they find that the Alpine region was characterised by the lowest numeracy values (i.e., the highest Whipple Index (WI)). It had a WI of about 400 in 300 AD. In contrast, Italy had the highest numeracy level, followed by France and Germany. Clearly, the fact that Italy had the highest numeracy values is not a very surprising but rather a confirming finding due to the high level of general living standards, infrastructure and technology that characterised the Roman civilisation. The right hand part of the figure nicely illustrates the gradual decrease of the WI and thus the evolving progress in numeracy from 1300 to 1900 in the different countries.

More detailed information on a higher number of European countries allow Figure 9.2 for western and northern and Figure 9.3 for central and eastern Europe (A'Hearn *et al.* 2009). Note that this time the ABCC Index is used so that numeracy increases with higher ABCC values. The general tendency of increasing numeracy values is also confirmed when considering more European countries. Around 1450, the Netherlands were already more advanced in numeracy than (the more developed) northern Italy. The split between the north and the south of Italy is already clearly visible because the southern part of Italy had very low numeracy levels both around 1450 and 1500. Data for southern Italy is lacking for the centuries afterwards but we know from Hippe and Baten (2012a) and Felice (2012) that important differences were still visible at the beginning of the 19th century which became consecutively less pronounced until around the middle of the 20th century. According to Felice (2012), there is a renewed (but small) tendency of divergence in the decades after 1960.

The differences in the other western and northern European countries between 1600 and 1850 are less striking. The UK was a numeracy leader in 1700 but other countries such as Denmark, the Netherlands, Belgium, France and Norway reached soon higher numeracy levels. Belgium appears to have been quite rapidly advancing as its catch-up phase was relatively short between 1700 and 1800.

On the other hand, if one takes a closer look at Figure 9.3, it becomes clear that many central and eastern European countries had lower numeracy values than their western and northern European counterparts. In Germany, there is a similar divide as in Italy in 1700 but this time it is not a clear geographical criterion but a religious one: the Protestant region was more advanced in numeracy than the Catholic one, giving further evidence to theories underlying the positive influence of Protestantism on human capital such as Weber (1958) and Becker and Woessmann (2009). Central European German speaking countries (Germany, Austria, Switzerland) had higher numeracy than countries to their east in 1700. Nevertheless, the latter progressed throughout the time period whereas Protestant Germany fell back between 1750 and 1800, Switzerland becoming the numeracy leader before Austria and Poland. This description of numeracy has until now been limited to the country level (with some exceptions) but different projects are underway that will also highlight more regional differences within these countries in numeracy in the near future (e.g., Juif and Baten 2013). These new projects will considerably improve our understanding on the formation of human capital in the European regions in the very long run.

Second, research on literacy has allowed further insights into European human capital formation. For example, Houston (2001, Figure 9.4) portrays the evolution of regional male literacy in western Europe from before 1700 until 1970. He defines a threshold (i.e., at least 50 % of males between 20 and 50 years have to be literate) and divides the European regions and countries into different categories. These categories show when a region has surpassed this threshold level. Similar to the evolution in numeracy, it can be seen that Germanic countries (Germany, Switzerland, the Netherlands, Sweden) were the leaders in literacy, having surpassed the threshold already before 1700. This might also be due to the fact that those countries were at least in part Protestant countries. The south-east of England and the larger Edinburgh areas in Scotland were similarly quick as the aforementioned countries. Moreover, literacy spread gradually to neighbouring regions in Belgium, France and northern Italy as well as to the other regions in Great Britain (except Wales) and Iceland until 1790. Geographical proximity appears to have been a decisive factor in the diffusion of knowledge in general and of literacy in particular.

The pattern is still visible but less striking for the next category, the regions surpassing the threshold until 1850. French men from almost all regions became by majority literate during this time period. Exceptions are the Celtic region of Bretagne, some central regions and Corsica. It is possible that the language barrier which separated some of these regions from their French speaking neighbours played a role here. The same suggestion can be made for the late progress in Wales and Ireland. The strongholds of Gaelic speaking regions in Ireland's western part only surpassed the threshold until 1900. However, one has to keep in mind that literacy is defined by reading and writing a particular language (often the official and not the regional language) and this might have biased the results here. In Spain, the northern regions of Asturias and Castilla La Vieja were the literacy leaders. Most other Spanish regions surpassed the limit only until 1900 (except in the south where it took more time). In the same class fall the north-western regions in Portugal, the remaining regions in Ireland and France as well as the north to central Italian regions. Finally, the last European regions became by a majority (male) literate, except for some southern and northern Portuguese regions which were the last ones to achieve the threshold until 1970.

Third, literacy is linked to the production of books. After Gutenberg's invention of the printing press, an increasing number of books were printed. We may also briefly consider its evolution in different European countries between the 15th and the 18th century (Figure 9.5, Baten and van Zanden 2008). The leaders in book production were Italy and the Netherlands in 1450. In contrast, Sweden had the lowest number of produced books. This might be astonishing as Sweden was one of the early high achievers in male literacy. In fact, we see that the number of new editions per million inhabitants (log scale) caught up in an impressive manner to the other countries between 1550 and 1600. Later on, a church law was passed in 1686 to advance the reading abilities with the aim that every Swede should be able to read the bible. This policy action may have contributed to the fact that it became the second most advanced country for book production in 1750, just ahead of the UK but still largely behind the Netherlands. A different case is the evolution of book production in Spain. Spain had a number of produced books similar to the UK in 1450. Whereas the UK increased it importantly afterwards, Spanish book production advanced only slowly; too slowly in comparison to the other more dynamic countries.

The evolution presented above mostly stops in the 18th and 19th century. For the second half of the 19th century until today, there are different databases available. These are in part international databases, including European data mostly at the national level or sometimes referring to large regions constituting those countries. Some of the most well-known are provided by Banks (1971), Flora (1983), Benavot and Riddle (1988), Barro and Lee (2001), Mitchell (2003), De La Fuente and Doménech (2006), Cohen and Soto (2007) and Morrisson and Murtin (2009) (see an overview in Table 9.1). One has to add that these databases are not always independent from another. In particular, the most recent ones are in part constructed from earlier databases and take different measurement and correction methods. As can be easily seen, the most popular proxies for the last decades have been educational attainment and years of schooling. The discussions and the intention to improve these databases underline once more the need for more and better data (e.g., Krueger and Lindahl 2001, De La Fuente and Doménech 2006).

Nevertheless, even given these ongoing improvements at the international country level, the regional level is still not adequately represented in Europe. Only for the last 10 to 15 years are comparable regional data available, especially from the European Statistical Office Eurostat. If one wants to go back further in time, data collection becomes much more difficult. This is true in particular if one is interested not only in western but also in eastern Europe.

Moreover, the differences between regions have not attracted sufficient attention. Even for current periods, Rodríguez-Pose and Tselios (2001, p. 358) clearly conclude that "despite [the] flourishing of education research, the geographical dimensions of education remain largely underexplored" (see also Hippe 2012c). Thus, it is apparent that more needs to be done in this direction. Therefore, this paper contributes to fill this gap in the literature.

9.3 Methodology and data

For the comparison of regional human capital in the long run it is important that the employed variables follow some basic common principles. Clearly, this is a difficult task as, theoretically, there are many different possibilities to measure human capital. However, in practice their number is substantially reduced due to lacking data availability in the different European countries. For these reasons, we use three variables that may each be considered the representative variable of their respective time period.

The first variable is numeracy, calculated by the use of the age heaping method. It is still a rather recent method but has become a very dynamic research field as evidenced by the number of publications over the last years (e.g., A'Hearn *et al.* 2009, Manzel and Baten 2009, Crayen and Baten 2010a, Hippe and Baten 2012a). Numeracy is a particular appropriate proxy if the aim is to measure human capital in early periods of human development. This proxy can be employed until the 20th century in many European countries and in some of them even later on. Because regional data for other human capital variables is much more restricted and less broad in geographical coverage, it may be expected that this method will still have more success and more contributions in the future. The details of this method have already been discussed in earlier publications (see e.g., Hippe and Baten 2012a, Hippe 2012b). Thus, we do not go into more detail in this paper. Let us only mention that there are characteristic heaping patterns on ages in historical and even in some current LDCs' censuses. In particular, a part of individuals did not report their exact age but rounded it on 0 and 5. The most important reason for this was that they did not know and were not able to calculate their age. It can be shown that one can measure numerical capacities by taking advantage of this rounding pattern. In practice, the ABCC Index is defined as

$$ABCC_{jt} = 125 - 125 \times \left(\sum_{i=5}^{14} n_{5ijt} \Big/ \sum_{i=23}^{72} n_{ijt} \right), \qquad (9.1)$$

where *n* is the number of all observations and *i* is the number of years in region *j* at point in time *t*. Numeracy is used to measure human capital around 1850.

Second, the next variable is literacy. Literacy is measured by the ability to 'read and write'. This proxy has been used for a long time and is still used in many international publications today (see e.g., UNESCO 2005). In short, it measures the reading and writing abilities of individuals as stated to census takers or as filled out in census forms. We take one common definition for literacy that was used in the earlier decades of the 20th century and that is similarly used until today:[109]

$$Literacy_{jt} = \sum_{i=10}^{N} rw_{ijt} \Big/ \sum_{i=10}^{N} n_{ijt}, \qquad (9.2)$$

where *rw* is the number of individuals who are able to 'read and write' and *N* is the total number of years of age. Unfortunately, some countries do not collect information on the literacy of the total population, as is the case e.g. for the Scandinavian countries at the turn and the first decades of the 20th century. Hence, they cannot be included in this study. Finally, literacy as the measure of the education of the population is progressively replaced by the level of educational attainment during the 20th century in most countries.

Therefore, the third and last variable is educational attainment. It is one of the standard measures used in today's official publications and has been widely used in the literature on human capital today and in the recent past (and in part beyond) (e.g., Redding and Schott 2003, López-Rodríguez *et al.* 2005, Breinlich 2006, Rodríguez-Pose and Tselios 2011). It measures the share of individuals that have surpassed a certain educational threshold level, in particular primary, secondary or tertiary education. Clearly, education systems vary importantly throughout Europe and have been subject to changes throughout history. This makes it more difficult to compare educational attainment. Still, it is possible to obtain a level of sufficient standardisation which allows to compute human capital values. This is common practice in international publications in general, comparing different countries in the world, and in publications on Europe and the European Union in particular. Eurostat provides standardised measures for all of its members. A further advantage of taking regions as the unit of analysis as compared to countries is that the regions are bound to the same educational system and have to adhere to the same ruling

[109] For more details see appendix.

principles. Therefore, within country comparisons are generally not biased by differences in the education systems. We measure educational attainment in the following way:

$$Notlowedu_{jt} = 1 - \left[\sum_{i=15}^{N} le_{ijt} \Big/ \sum_{i=15}^{N} n_{ijt} \right], \tag{9.3}$$

where *le* is the number of individuals who have achieved pre-primary, primary and lower secondary education as highest level of education and *n* is the number of all individuals.[110] We have opted for this definition because both numeracy and literacy indicators are proxies of rather basic human capital. In contrast, taking e.g. the share of individuals with tertiary education would clearly be an indicator of more advanced human capital. This measure would be less revealing on the abilities of the overall population. For this reason, it appears more appropriate to choose a proxy for the attainment of rather low education. This proxy captures the basic attainment of the entire population.

For long-term comparisons, although it is clear that each variable does not measure the same attributes of human capital, it is important that the variables have some common characteristics to improve their comparability. There are several features that are common or at least similar to all three variables which underline the meaningful use of these variables in the present study.

First, all variables are in some sense representative of the contemporaneous period at which there were collected.

The reason for this is related to the second characteristic that all educational variables have mostly been calculated (or are in part directly taken) from official census

[110] This definition is based on the availability of data by Eurostat, referring to ISCED levels 0 to 2. Note that the Eurostat data are derived from the EU Labour Force Survey and refer to the “economically active population”, including both employed and unemployed individuals and following the principles set up by the ILO (see e.g., ILO 1982). In contrast, the census data for Russia in 2012 refer to the overall population. Alternatively, it is also possible to take the age range of 25 to 64 years old for the countries provided by Eurostat. Because both age ranges are correlated to 99.5 %, the results do not importantly change when using the alternative age range. We prefer the definition of ages above 15 years because it allows us to include the data on Russia in 2010 and enables easier comparison with literacy.

publications or official surveys. These official documents are generally intended to provide public policy makers and the general public with indications on the state of the population in different contexts. They may be better suited than other rather unofficial documents. For this reason, the common underlying document types underline the methodology common to all variables.

Third, some of the variables may be argued to be more or less direct output measures. This is clear for the case of literacy where the ability to 'read and write' is measured. This is also true for numeracy although the output has to be computed from an age distribution. In view of educational attainment, this is also the case for tertiary education. Even though one may consider secondary education not to be a direct output measure, it is still a relevant variable characterised by distinctive regional inequalities. In particular, early school leavers play a significant role here. They are those individuals "who leave education and training with only lower secondary education or less, and who are no longer in education and training" (Council of the European Union 2011, C 191/01). The Council of the European Union confirms that this issue is important. It stresses that the reduction in the number of these early school leavers is essential to achieve some of the key objectives of the Europe 2020 strategy (Council of the European Union 2011). Still in 2009, the early school leavers constituted a share of 14.4 %.[111] This arguably still (too) high share illustrates that the successful accomplishment of primary education and parts of secondary education cannot be taken for granted for the whole population until today. In consequence, this has even more so been the case in the past.

Fourth, all numeracy, literacy and educational attainment measures are considering an important part of the total population. In other words, they are not restricted

[111] Therefore, the goal of the Europe 2020 strategy is to reduce this share to 10 % until 2020 (Council of the European Union 2011).

to some particular social groups in society such as military recruits or married couples which are commonly used to approximate literacy by using signature rates. Furthermore, they normally consider both sexes (not only males) in contrast to military recruitment data. Thus, all variables are a measure of the basic human capital of the overall population and are not subject to biases that may arise when only a part of the population is considered.

Fifth, and directly connected to this, one should consider a similar part of the population with regard to age. All variables are commonly defined by the use of a certain age threshold, i.e., they only take into account the individuals above a certain age threshold. Numeracy takes account of the great majority of the individuals that constitute the population of the other two variables. In consequence, all proxies measure the share of individuals that have similar years of ages.

Finally, the sixth and maybe the most important common characteristic for the actual measurement of human capital is the definition of the variable at stake. All three variables are defined by an identical value range which goes from 0 (or 0 %) to 100 (or 100 %). The reason for this is that every variable considers the share of individuals that have some form of education. The rest of the individuals does not have this attribute. In other words, every variable is derived from a binary indicator. In consequence, all variables are subject to the same advantages and disadvantages inherent to share measures (for consequences see also Hippe 2012b). This common measurement framework makes it particularly appropriate to consider numeracy, literacy and educational attainment for estimating regional human capital in Europe from 1850 until today.

Given this common measurement framework, human capital data have been collected from a variety of sources. First, the data on numeracy have been taken from the large database by Hippe and Baten (2012a). Second, literacy data have been added for 1900 and 1960 from the census publications of the different European countries under study (see appendix for more details). Moreover, the data referring to 1930 have been taken from Kirk (1946). Finally, educational attainment data for 2000 and 2010 have been collected from Eurostat (Eurostat 2011i). They have been supplemented by census data from Russia in 2010.[112]

9.4 Results

9.4.1 Evolution of human capital in the European regions, 1850-2010

A first intuition on the overall evolution may be derived from descriptive statistics for all human capital variables. They are given in Table 9.2. Before 2010, we have the highest number of regions in 1850, taking advantage of the large database by Hippe and Baten (2012a). The number of observations is constant or decreasing during the 20th century until the most recent data for 2000 and 2010. Interestingly, although the indicators and the time periods are very different, the descriptive statistics show that not only the number of regions but also the standard deviation, minimum and maximum values are very similar in 1850 and 2010.

Overall, we may summarise that there is sufficient variation for all variables at all points in time to obtain relevant and pertinent results. The definition of literacy between 1900 and 1960 is (almost) identical so that we may make some general comments on its evolution. Note, however, that not always the same countries are included (in particular the data for 1960 refer only to peripheral European countries) which is why this first very general impression of the data can be biased and needs to be confirmed by more detailed analyses in later parts of the paper. Nevertheless, as one would have hypothesised, literacy is progressing over the time period 1900 to 1960 and regional inequalities decrease as

[112] The Russian census data have been classified according to similar classes as specified in the Eurostat data, i.e. ISCED levels 0 to 2.

literacy approaches its maximum level. Moreover, one may indicate that the number of observations and the distribution of our measure of educational attainment in 2000 are relatively similar to those of literacy in 1930. Data on more regions become available in 2010, but the minimum and mean values show that educational attainment is progressing until the present.

Let us now turn to the more detailed analysis of the data. Due to the different points in time covered in this paper, we limit our analysis of each point in time to the most important aspects and discuss the most noticeable changes.

One possibility is to use boxplots to see directly intranational variation and possible regional outliers. Boxplots allow an easy evaluation of regional variation by providing different pieces of information at the same time. First, boxplots highlight the minimum and maximum regional values within a country. In between, they also depict the lower quartile, median and upper quartile values. In this way, it is possible to get a clearer understanding of the regional distribution of human capital values. In addition, outliers can be distinguished from the other observations so that those regions with exceptional high or low values can be identified. For this reason, this approach enables to get a closer picture of regional variation.

The results are shown in Figure 9.6 to Figure 9.15. Note that there are always two figures for one point in time (except for the most recent points): first, the historical countries are considered and second, the country within their current administrative borders. We have opted to highlight the results for both possibilities because there are different implications stemming from the use of different administrative borders.

By considering historical countries, one can take into account the historical aspects of human capital formation. For example, why did an empire break up? One reason could lie in regional inequalities. Still, one has to be aware that these historical borders changed and that e.g. the Ottoman Empire does not refer to the same territory at each point in time.

On the other hand, the use of current borders allows to trace back the evolution of regional inequalities throughout time. This methodology enables us to infer conclusions for the long-term evolution of human capital and human capital inequality in a particular country. Therefore, these results are important for the present and may allow policy implications for today's world. Nevertheless, we do not always explicitly refer to each figure in detail to avoid unnecessary repetition.

To begin with the data on the ABCC around 1850, the most important regional differences exist within the (European part of the) Ottoman Empire, (the European part of) Russia, Spain, Portugal, Transleithania and Italy (Figure 9.6). It is clear that the Ottoman Empire had the highest regional variation and in general the lowest ABCCs. Moreover, the Irish regions that were part of the United Kingdom were clear outliers. As Hippe and Baten (2012a) note, there is also an important north-south difference in Italy and France, to a (much) lesser degree the reverse can be said of Norway. Spain appears to be characterised by a core-periphery pattern. The case in Russia is more complicated: the Caucasus region and the Belarusian regions are the least numerate while in particular Estonia has the highest numeracy values. The latter may be attributed to their historical and cultural ties to the most advanced countries in literacy (and numeracy) in Scandinavia.

In the next step, we consider literacy in 1900. Because literacy did not yet attain its maximum value in many countries, one can observe more regional variation than in the ABCC (Figure 9.8). Cisleithania (AT) has a very high range in literacy to that time (from about 0.2 to almost 1). Regional inequalities are also persistent in a number of other countries between 1850 and 1900. This is the case e.g. for Spain, Hungary and Italy. Less but still important variation is apparent in the Ottoman Empire, Portugal and the Russian

Empire. In Greece, the Greater Athens region leaves all other regions far behind. Comparing our results to those using current administrative borders, one may add that in today's Russia Moscow and Yaroslavl were important outliers. Kärnten is estimated to have a lower literacy rate than the other Austrian regions.

Advancing to 1930 (Figure 9.10), many results of the previous points in time are confirmed, underlining the persistence of human capital inequalities. Regional variation is impressive in the new state of Poland which combines former regions from the Russian, German and Austro-Hungarian Empire before World War I. This circumstance may probably well explain the stark contrasts still in existence in the 1930s. In addition, most outliers are regions that are in some way or other atypical for their country. For example, the Zakarpatska region was part of Czechoslovakia until World War II (and is now part of Ukraine) but has historically been a rather ethnically mixed region. In France, the island of Corsica has a lower level of literacy. In general, islands are often different from the mainland and are often less developed. In Greece, the Thrace region was annexed in 1923 so that the fruits of the government efforts cannot yet be observed. In Russia, the two outliers Kalmykia and Dagestan are also different from many other regions. Kalmykia is the sole European Buddhist region, while Dagestan is a mixed region with numerous ethnic (and non-Russian) groups, the majority being Muslim. In view of the current country borders, one should take into account the fact that Russian Kaliningrad was still part of Germany (Königsberg) which explains the very high literacy level for Russian standards.

In 1960, the literacy scale already indicates that regional variation has considerably decreased. For this reason, note that the number of considered countries is limited to less developed regions in western, eastern and south-eastern Europe. The highest regional variation can be observed in Spain, Italy and Yugoslavia (Figure 9.12). Portugal has the lowest average literacy rates, even lower than the European part of the Soviet

Union. The northern, Baltic and larger urban regions of Estonia, St. Petersburg, Moscow and Murmansk are outliers in the latter (taking current borders, the northern region of Komi is added to this group of regions).

Taking educational attainment in 2000 (Figure 9.14), we can include much more countries. Still, Portugal is at the lower bound of educational attainment. This highlights the historical continuity of Portuguese low human capital performance in a European comparison. Its capital, Lisbon, has significant higher levels of educational attainment. Nevertheless, Lisbon is still at the very bottom of the educational attainment of other countries (except Malta which has still lower educational attainment). The highest average values come from the Czech Republic and Slovakia which constituted one country until 1992. Other former Communist countries have also high educational attainment levels, such as Eastern Germany (e.g., Dresden, Leipzig, Chemnitz), Poland or Hungary. Given our indicator, within-country regional inequalities are relatively small when compared to literacy data in 1900 or 1930. For example, Italy's regional educational disparities are relatively small. Still the highest inequalities are found in Spain.

Finally, the current situation in 2010 is highlighted by Figure 9.15. Clearly, the overall picture is similar to that in 2000 even though some countries have been added. Thus, Portugal takes once again the last position on educational development. Malta follows. Russian disparities are relatively small but historical literacy leaders are once more significant positive outliers (Moscow, St. Petersburg, Murmansk). In general, one can state that educational attainment has been rising during the first decade of the 21st century. Nevertheless, regional inequalities are striking until the present, underlining the importance of the regional level for academic research and policy makers.

9.4.2 Evolution of intranational inequality

The evolution of inequality within a country may also be an interesting aspect to take account of. Hippe and Baten (2012a) have applied this idea to the evolution of numeracy in the European regions between 1790 and 1880. We employ their methodology because it is also particularly suitable for the present context. More specifically, we use the coefficient of variation (CV) as our regional inequality measure. The CV is defined as

$$CV = \frac{\sigma}{\mu} \times 100, \qquad (9.4)$$

where σ is the standard deviation of regional human capital values and μ represents the average population-weighted human capital value. This measure is especially appropriate for our study because it is a number without dimensions and thus enables an easy comparison between the countries of our dataset which are characterised by very different mean values.

The complete results for all points in time are shown in Figure 9.16 to Figure 9.19. Note that we show the results for 1900 to 1960 with and without Serbia which is an apparent outlier (Figure 9.17 and Figure 9.18). Conclusions can only be cautious because the number of countries and regions vary in the dataset. Nevertheless, it is clear that Serbia was marked by very high regional differences. These differences may result from the fact that Serbia was created from regions that formerly belonged to countries with very different levels of development, e.g., Austria-Hungary and the Ottoman Empire, and that we consider Kosovo to be still part of Serbia. Although regional inequalities decrease until 1960, Serbia was not able to relinquish these differences as all other countries.

Furthermore, Portugal appears to have been characterised by important regional differences throughout history even though this country is rather small and homogeneous in cultural terms. This finding underlines once more that regional variation can be striking even for small countries. Other countries with high CVs are in particular Italy, Spain, Greece and Russia. Still, these CVs have different causes in each case. For example, Italy shows striking north-south differences over the centuries. More specifically, the north has relatively high levels of human capital while the south lags behind. The core-periphery pattern in Spain has led to similar important regional disparities. The Greater Athens region in Greece is not only the capital regions of the country but in this way also the administrative, economic and educational centre, giving it an advance to other Greek regions. Finally, although we are only considering the European part of Russia, this is still a huge area, comprising different ethnical and cultural groups, so that regions do not develop their human capital at an identical pace throughout time.

As an alternative standard measure of inequality, it is also possible to construct population-weighted Gini coefficients.[113] The results (Figure 9.20 to Figure 9.23) are quite similar to those obtained for the CV. For example, Serbia had historically the highest inequality in human capital. In general, the outliers that have very high Ginis are not as pronounced as for the CV. Still, the ranking among the different countries at each point in time is almost unchanged. Therefore, the use of the Gini coefficient validates the results obtained by using the CV.

9.5 Conclusion

This paper has traced the long-run evolution of human capital in the European regions between 1850 and 2010. Human capital is an important factor that has to be considered in a variety of setups because it affects economic and social developments. The role of human capital has particularly been stressed by parts of the economic growth literature in the last decades, from Endogenous Growth Theories (Lucas 1988, Romer 1990) until the recent contributions by Unified Growth Theory (e.g., Galor 2005a).

[113] See Jenkins (1999/2010) for details of calculation.

We have constructed a new and large database from a variety of sources. The data from different points in time have allowed to construct a dataset that covers at least an important share of all European countries in 1850, 1900, 1930, 1960, 2000 and 2010. To our knowledge this is the first time that it has been possible to show and analyse the regional evolution of human capital during such a long time period for such an important number of countries located in the European continent. To this end, three different proxies have been used: numeracy, literacy and educational attainment. More specifically, numeracy is measured by the ABCC Index, literacy by the share of individuals able to 'read and write' and educational attainment by the share of employed individuals having attained an education level above lower secondary education. We show that this choice is not arbitrary but that the inherent characteristics of these variables make them appropriate for the purposes of this study.

For a general overview of the evolution of human capital in the long run, we have presented some of the literature that has contributed evidence on human capital in a longer historical perspective. More precisely, we have put forward the evolution of numeracy, literacy and book production.

After this point of departure, we have traced the evolution of human capital in the European regions between 1850 and 2010. The most striking result is that regional differences have been characteristic for many European countries in the past and the present. Regional persistence in human capital is striking in many cases. In addition, we measure regional inequality by the coefficient of variation and the Gini coefficient. Both show that some countries have been much more unequal than others. The high regional differences within countries underline that cross-country analyses miss an important part of the human capital story. For this reason, more research needs to be done at the regional level in Europe. This paper has done a further step into this direction. The new evidence is important both for academics and policy makers because it contributes to the understanding of long-term developments which are important for the present and the future.

9.6 Appendix

9.6.1 Data

Data for 1850

Hippe, R. and J. Baten (2012a). Regional Inequality in Human Capital Formation in Europe, 1790 – 1880, *Scandinavian Economic History Review*, 60 (3): 254-289

Data for 1900

Country	Census year	Source
Albania	1918	Preliminary dataset "Albanische Volkszählung von 1918", entstanden an der Karl-Franzens-Universität Graz unter Mitarbeit von Helmut Eberhart, Karl Kaser, Siegfried Gruber, Gentiana Kera, Enriketa Papa-Pandelejmoni und finanziert durch Mittel des Österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Special thanks to Siegfried Gruber for providing the data.
Austria (Cisleithania)	1900	K. K. Statistische Central-Commission (1903). Oesterreichische Statistik. Ergebnisse der Volkszählung vom 31. December 1900, 2. Band, 2. Heft, Wien, Kaiserlich-Königliche Hof- und Staatsdruckerei.
Bosnia-Herzegovina	1910	Mayer, M. (1995). Elementarbildung in Jugoslawien (1918-1941), München, R. Oldenburg Verlag.
Belgium	1900	Statistique de la Belgique (1903). Population. Recensement général. 31 décembre 1900, Bruxelles, Typographie-Lithographie A. Lesigne.
Bulgaria	1900	Principauté de la Bulgarie (1906). Résultats généraux du recensement de la population dans la principauté de Bulgarie au 31 décembre 1900, 1-ère livraison, Sophia: Imprimerie "Gabrovo".
Cyprus	1911	Mavrogordato, A. (1912). Cyprus. Report and general abstracts of the census of 1911, London: Waterlow & Sons Ltd.
France	1906	Statistique Générale de la France (1908). Résultats statistiques du recensement général de la population effectué le 6 mars 1906, Paris.
Greece	1907	Royaume de Grèce (1909). Résultats statistiques du recensement général de la population effectué le 27 octobre 1907, Tome I, Athènes: Imprimerie nationale.
Hungary	1900	Magyar Statisztikai Közlemények (1907). A magyar szent

Country	Census year	Source
(Trans-leithania)		korona országainak 1900. Evi. Népszámlálása. Harmadik rész. A népesség részletes leirása. Budapest: Pesti Könyvnyomda-Részvénytársaság.
Ireland	1901	Census of Ireland, 1901 (1902). Part II. General Report, Dublin: Brown & Nolan Ltd.
Italy	1901	Ministero di agricultura, industria e commercio (1907). Anuario statistico italiano 1905-1907, Fascicolo Primo, Roma: G. Bertero e C.
Montenegro	1900	MacKenzie Wallace, D. (2006). A short History of Russia and the Balkan States, Elibron Classics, Adamant Media Corporation.
Portugal	1900	Ministerio dos negocios da fazenda (1906). Censo Da Populacao Do Reino De Portugal No 1. De Dezembro De 1900, Vol. II, Lisboa: A Editora.
Romania	1899	Royaume de Roumanie (1905). Résultats définitifs du dénombrement de la population (décembre 1899), Bucarest, Eminesco.
Russian Empire	1897	издание центрального статистического комитета министерства внутренних (1899-1905). первая всеобщая. перепись населения, российской империи, 1897 г., с.-петербург. Various tombs. Note that data are unavailable for some gubernias.
Serbia	1906	Direction de la Statistique d'Etat du Royaume de Serbie (1908). Annuaire statistique du Royaume de Serbie, Onzième Tome, Belgrade, Imprimerie de l'Etat du Royaume de Serbie.
Spain	1900	Dirección general del Instituto geográfico y estadístico (1903). Censo de la Poblacion de Espana, según el Empadronamiento hecho en la Península é Islas adyacentes en 31 de diciembre de 1900, Tomo II and Tomo III, Madrid: Imprenta de la Dirección general del Instituto geográfico y estadístico.
United Kingdom	1901	Hechter, M. (1976). U.K. County Data, 1851-1966 [computer file]. Colchester, Essex: UK Data Archive [distributor]. SN: 430, http://dx.doi.org/10.5255/UKDA-SN-430-1. Although all efforts are made to ensure the quality of the materials, neither the original data creators, depositors or copyright holders, the funders of the Data Collections, nor the UK Data Archive bear any responsibility for the accuracy or comprehensiveness of these materials.

Note: Age definitions are as follows: Italy = 6+; Austria, Bosnia-Herzegovina = 7+; Spain = 8+; Albania, Belgium, Bulgaria, France, Greece, Ireland, Portugal, Russian Empire = 10+; Hungary, Romania, Serbia = 11+; Cyprus = 15+; United Kingdom = unavailable (a comparison of the included data for Ireland with the source for Ireland as listed above has revealed similar overall results, so that an age definition of 10+ can reasonably be assumed). Age definitions are either directly given in the publication or have been linearly estimated from available age definitions in order to be as close as possible to the standard definition of ages above 10 years. The various age definitions as calculated here may possibly not significantly affect the final results. For example, the percentage change of using a 5+ instead of a 10+ definition is below 1 % in Ireland.

Data for 1930

Kirk, D. (1946). *Europe's Population in the Interwar Years*, Princeton: Princeton University Press

Data for 1960

Country	Census year	Source
Bulgaria	1956	централно статистическо управление при министерския свет (1960). преброяване на населението в народна република българия на 1. XII. 1956 година, общи ресултати, книга II, софия: държавно издателство "наука и изкуство".
Greece	1961	Royaume de Grèce (1968). Résultats du recensement de la population et des habitations effectué le 19 mars 1961, Vol. III, Athènes: Office nationale de Statistique.
Hungary	1960	Központi Statisztikai Hivatal (1962). 1960. Évi népszámlálás, Budapest, Allami Nyomda.
Italy	1961	Istat (2012). Serie Storiche, Tavola 7.1.1, online, last accessed 3 August 2012, http://seriestoriche.istat.it/fileadmin/allegati/Istruzione/tavole/Tavola_7.1.1.xls.
Poland	1960	Glówny Urzad Statystyczny (1960). Biuletyn statystyczny. Spis powszechny z dnia 6 grudnia 1960 r., Ludnosc. Gospodarstwa domowe, Wyniki ostateczne, Seria "L", various issues, Warszawa.
Portugal	1960	Instituto nacional de Estatística (1960). X Recenseamento Geral Da Populacao, Tomo III, Lisboa: Sociedade Tipográfica.
Romania	1956	Republica Populara Romîna (1961). Recensamîntul Populatiei din 21 Februarie 1956, Rezultate Generale, Bucuresti, Direcția Centrala de Statistica.
Spain	1960	Instituto nacional de Estadística (1969). Censo de la Poblacion y de la Viviendas de Espana, según la Inscripción realizada el 31 de diciembre de 1960, Tomo III, Madrid: I.N.E. Artes graficas.
USSR	1959	Demoscope (2012). Всесоюзная перепись населения 1959 года. Таблица 7. Распределение населения по возрасту и уровню образования. РГАЭ. Ф.1562 Оп. 336 Д.1591-1594, online, last accessed 8 August 2012, http://demoscope.ru/weekly/ssp/rus_edu_59.php. Demoscope (2012). Всесоюзная перепись населения 1959 года. Таблица 2,5. Распределение всего населения и состоящих в браке по полу и возрасту. РГАЭ. Ф.1562 Оп. 336 Д.1535-1548. Российская Государственная библиотека, отдел "литературы

Country	Census year	Source
		ограниченного пользования", online, last accessed 8 August 2012, http://demoscope.ru/weekly/ssp/sng_mar_59_r.php.
Yugoslavia	1961	Statisticni urad Republike Slovenije (2012). Popis prebivalstva 1961, Prebivalstvo, staro 10 let ali več, po spolu, starosti in pismenosti, online, last accessed 8 August 2012, http://www.stat.si/publikacije/popisi/1961/1961_2_40.pdf.

Note: Age definitions are as follows: Italy = 6+; Hungary, Poland, Portugal, Spain = 7+; Bulgaria, Romania = 8+; Greece, USSR, Yugoslavia = 10+. Age definitions are either directly given in the publication or have been linearly estimated from surrounding available age definitions to be as close as possible to the standard definition of ages above 10 years.

Data for 2000 and 2010

All countries except Russia: Eurostat (2011i). Economically active population by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000), online, last accessed 22 June 2012, lfst_r_lfp2acedu; http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Russia: всероссийская перепись населения о россии языком цифр (2012). Население по уровню образования по субъектам Российской Федерации, online, last accessed 12 October 2012, http://www.perepis-2010.ru/results_of_the_census/tab8.xls.

9.6.2 Tables

Table 9.1 Databases on international evolution of human capital in the longer term

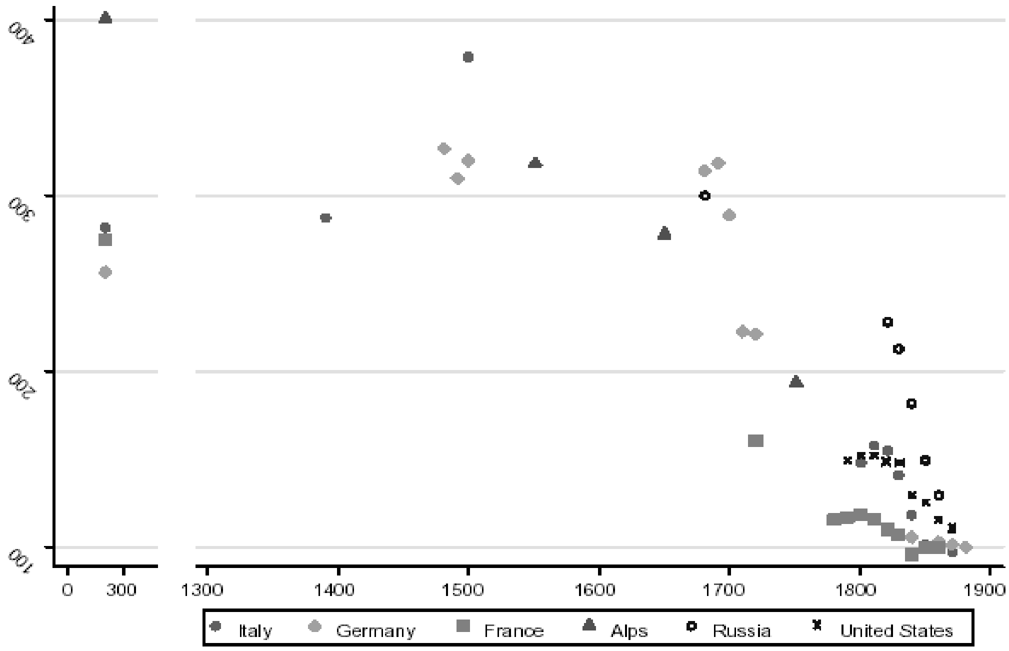
Authors	Time period	Human capital proxies
Banks (1971)	1860-1966	Primary, secondary and tertiary school enrolment; literacy; number of books
Barro and Lee (2001)	1960-2000	Educational attainment, years of schooling
Benavot and Riddle (1988)	1870-1940	Primary enrolment rates
Cohen and Soto (2007)	1960-2010	Educational attainment, years of schooling, enrolment rates
De la Fuente and Doménech (2006)	1960-1995	Educational attainment, years of schooling
Flora (1983)	1810-1977	Enrolment rates, number of pupils, number of teachers
Lindert (2004a, b)	1830-2000	Enrolment rates, years of schooling, teachers
Mitchell (2003)	1830-1919	Primary, secondary and tertiary education (number of pupils and teachers)
Morrisson and Murtin (2009)	1870-2010	Years of schooling

Table 9.2 Descriptive statistics for the human capital indicators

Indicator	year	obs.	mean	sd	min	max
ABCC	1850	304	91.96	11.64	26.38	100.00
Literacy	1900	230	0.57	0.30	0.05	1.00
Literacy	1930	238	0.74	0.21	0.17	1.00
Literacy	1960	168	0.82	0.11	0.59	0.99
Educational attainment	2000	254	0.70	0.16	0.14	0.95
Educational attainment	2010	339	0.77	0.12	0.26	0.97

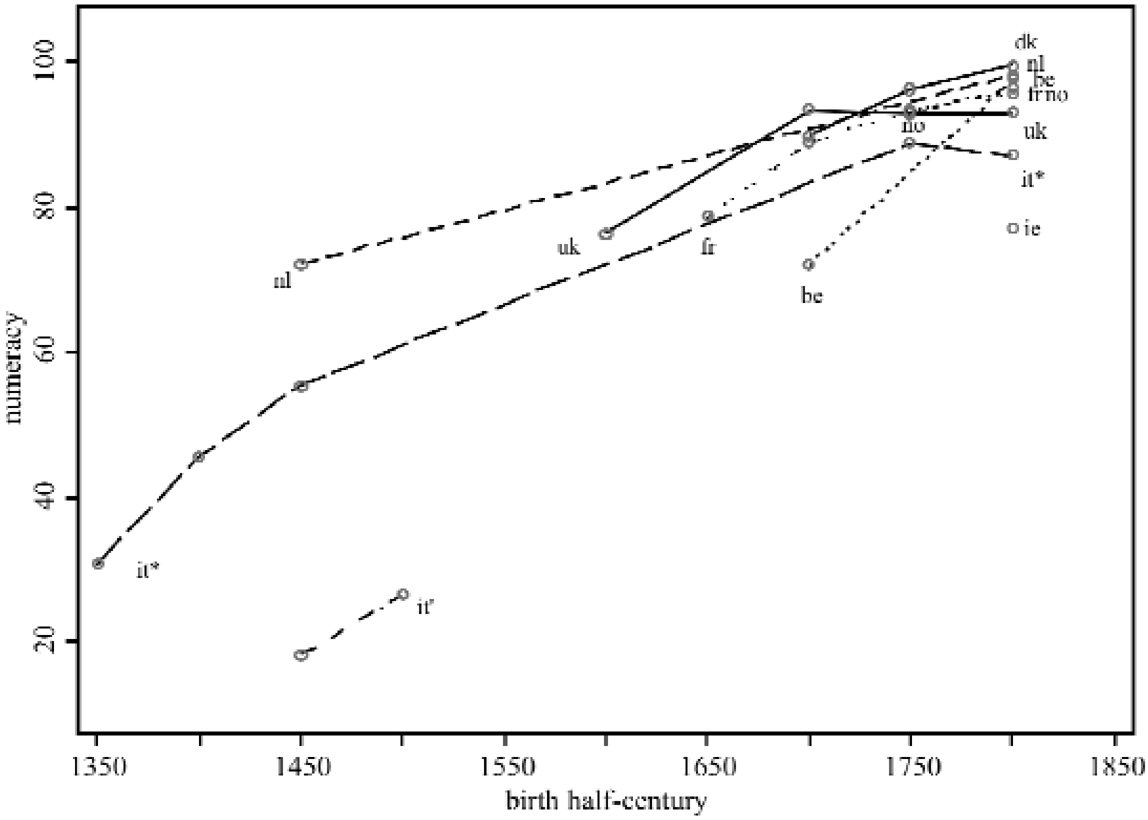
9.6.3 Figures

Figure 9.1 Very long-term evolution of numeracy in Europe



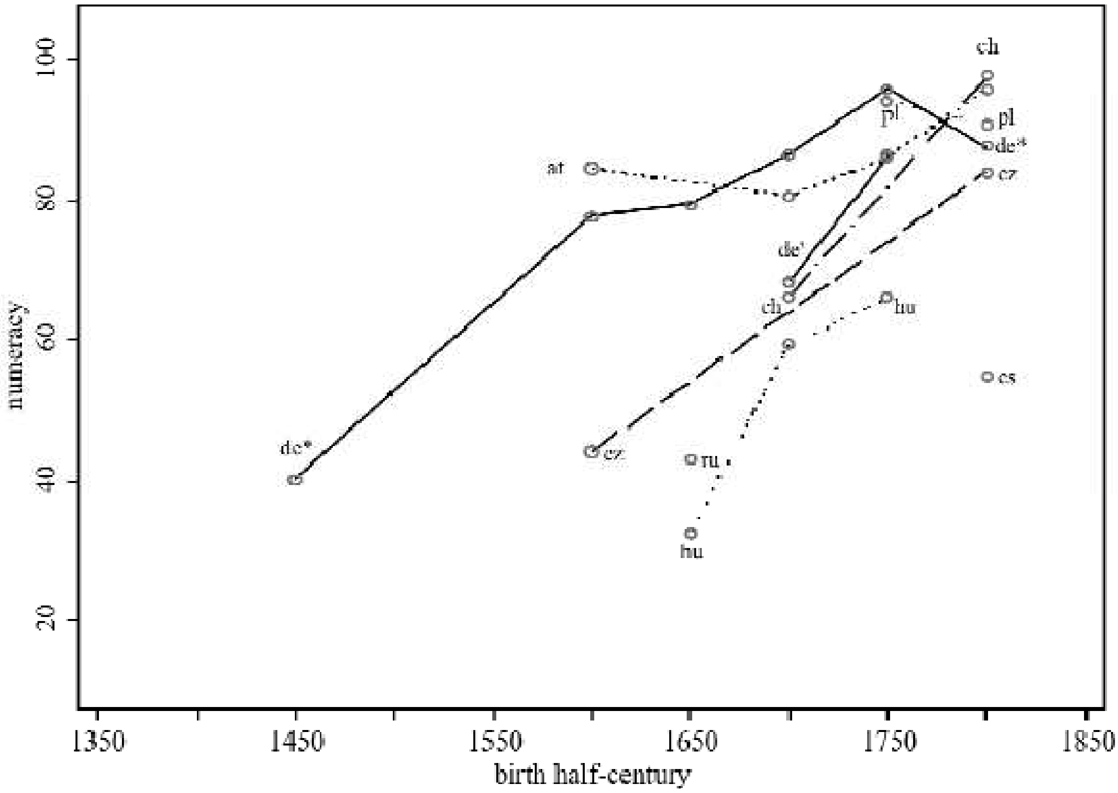
Note: Whipple Index values on y-axis (for orientation: the maximum WI value (500) corresponds to the minimum ABCC value (100) and vice versa (100 and 0)).
Source: A'Hearn *et al.* (2006).

Figure 9.2 Long-term evolution of numeracy in western and northern Europe



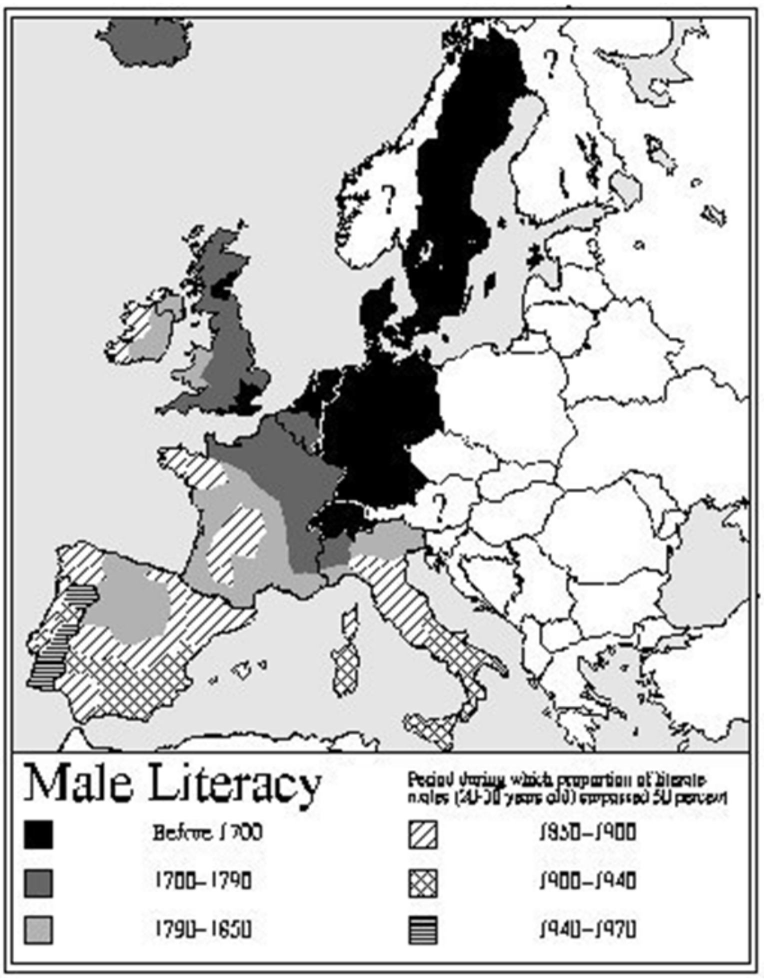
Note: it* refers to northern Italy, it' to southern Italy.
Source: A'Hearn *et al.* (2009).

Figure 9.3 Long-term evolution of numeracy in central and eastern Europe



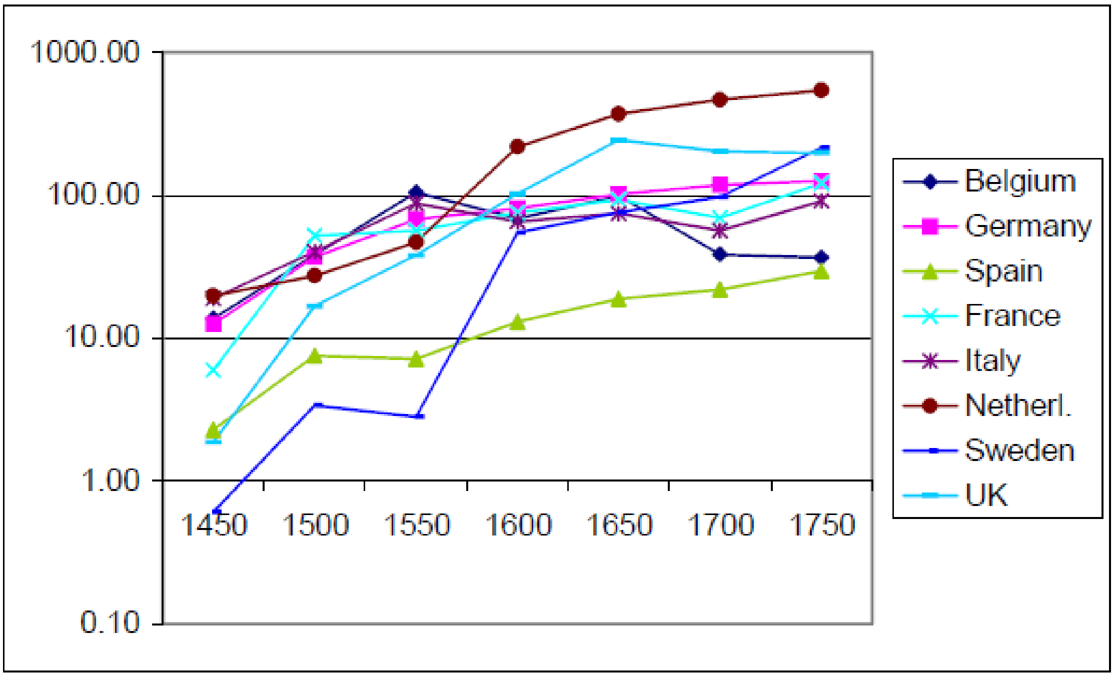
Note: de* refers to Protestant Germany, de' to Catholic Germany.
Source: A'Hearn *et al.* (2009).

Figure 9.4 Evolution of attainment of literacy threshold in the European regions



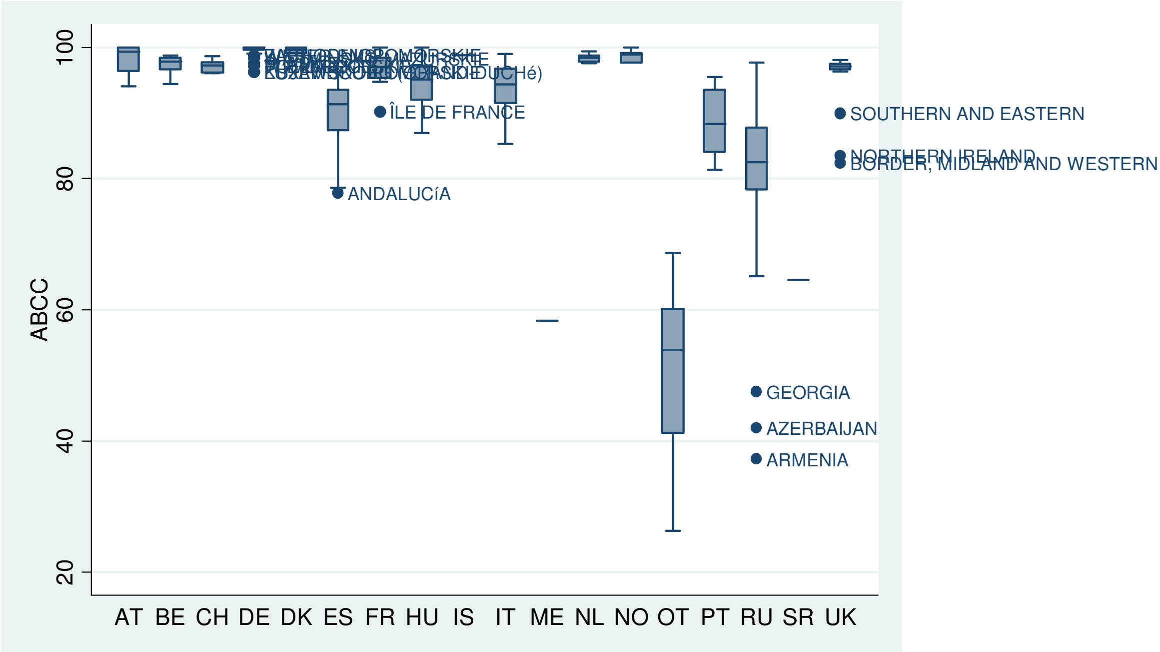
Note: The literacy threshold refers to the time period when the share of literate male individuals between 20 and 50 years surpassed 50 %.
Source: From Houston (2001). *Encyclopedia of European Social History*, 1E. © 2001 Gale, a part of Cengage Learning, Inc. Reproduced by permission. www.cengage.com/permissions.

Figure 9.5 Book production in Europe, 1450/99-1750/99



Note: Data refer to the number of new editions per million inhabitants, log scale on vertical axis.
Source: Springer and the Journal of Economic Growth, 13 (3), 2008, p. 220, Book production and the onset of modern economic growth, Baten, J. and J.-L. van Zanden, Fig. 1; with kind permission from Springer Science and Business Media.

Figure 9.6 Regional variation in ABCC by historical countries, 1850



Note: The Ottoman Empire refers to Albania, Bosnia-Herzegovina, Bulgaria, Cyprus and FYROM. The United Kingdom comprises the main countries of the British Isles, i.e., the United Kingdom and the Republic of Ireland (independence of the Republic of Ireland was declared in 1916 and recognised in 1922). Accordingly, Cyprus is separately considered but note that it was administered by the British Empire since 1878 and finally annexed in 1914. For simplicity reasons, Germany refers to the German Empire in its borders from 1871 onwards and Austria and Hungary to the external and internal borders from 1867 onwards. The same methodology applies to the other countries.

Figure 9.7 Regional variation in ABCC by current countries, 1850

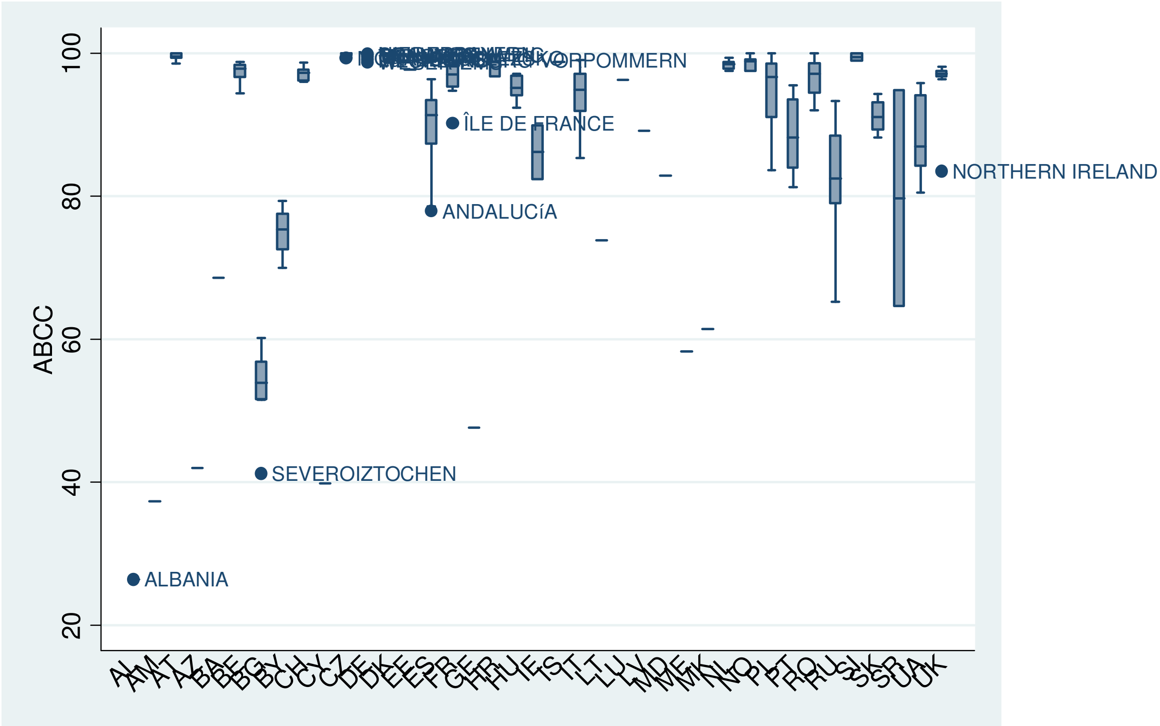
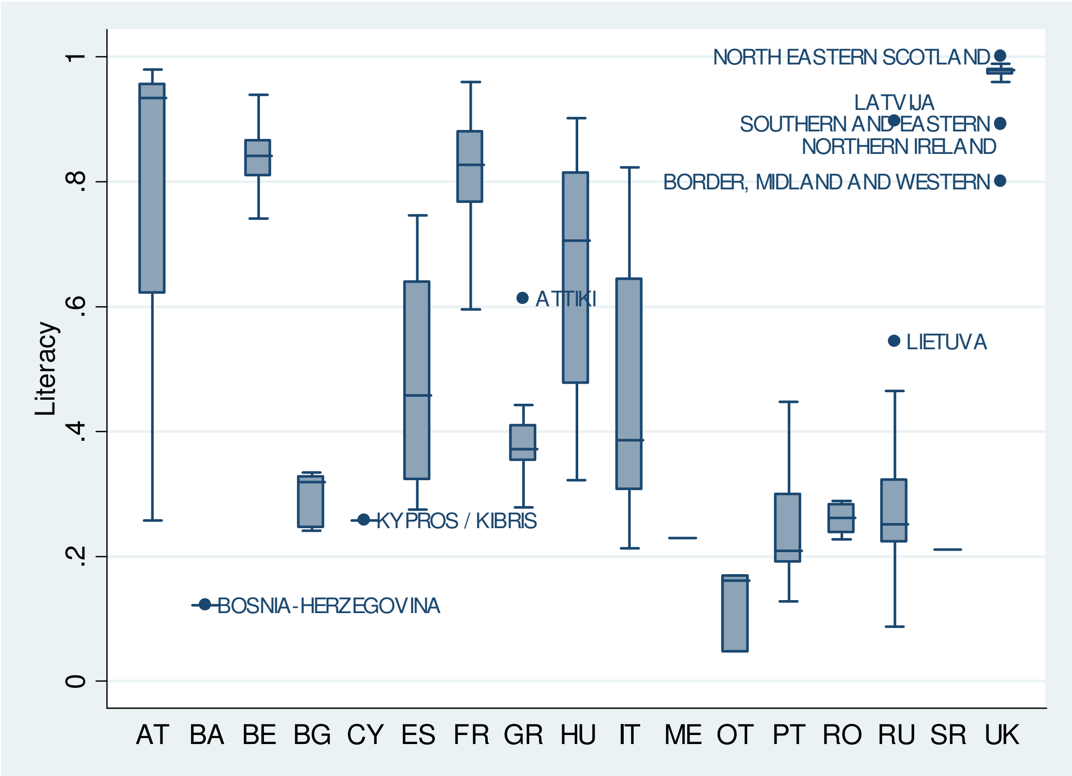


Figure 9.8 Regional variation in literacy by historical countries, 1900



Note: Bosnia-Herzegovina is considered as a separate country (note that it was under Austrian-Hungarian jurisdiction since 1878 and finally annexed in 1908). The Ottoman Empire refers to Albania, Kosovo and FYROM. The United Kingdom comprises the main countries of the British Isles, i.e., the United Kingdom and the Republic of Ireland (independence of the Republic of Ireland was declared in 1916 and recognised in 1922). Accordingly, Cyprus is separately considered but note that it was administered by the British Empire since 1878 and finally annexed in 1914.

Figure 9.9 Regional variation in literacy by current countries, 1900

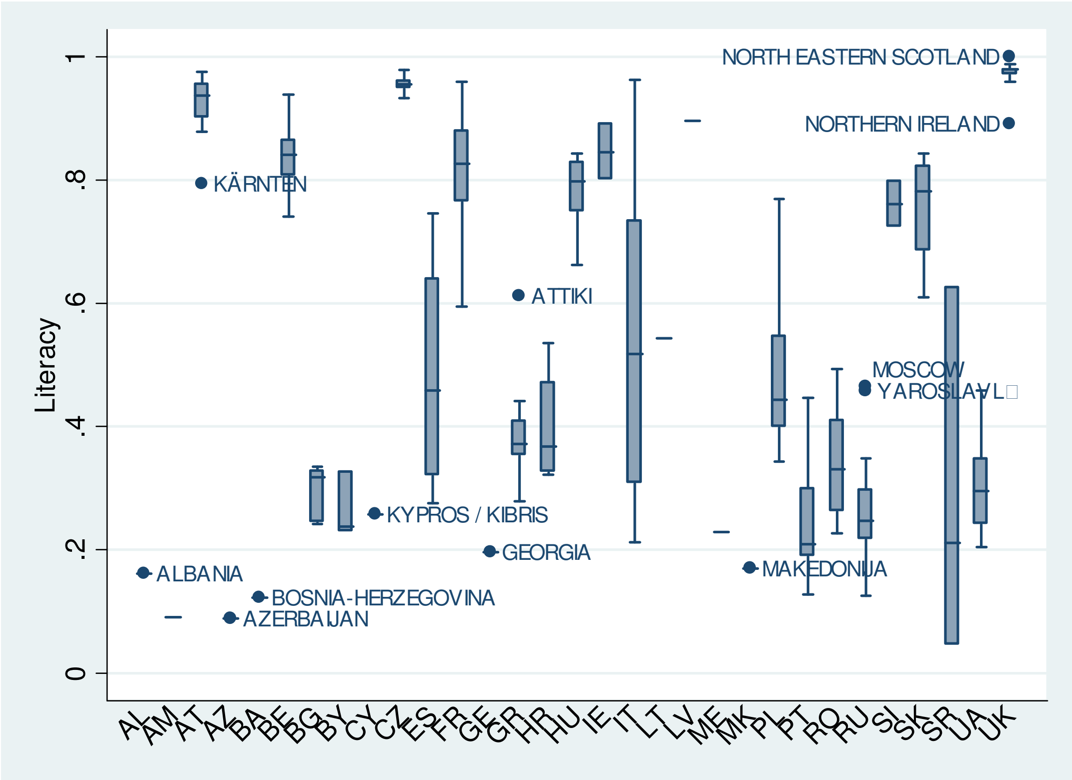
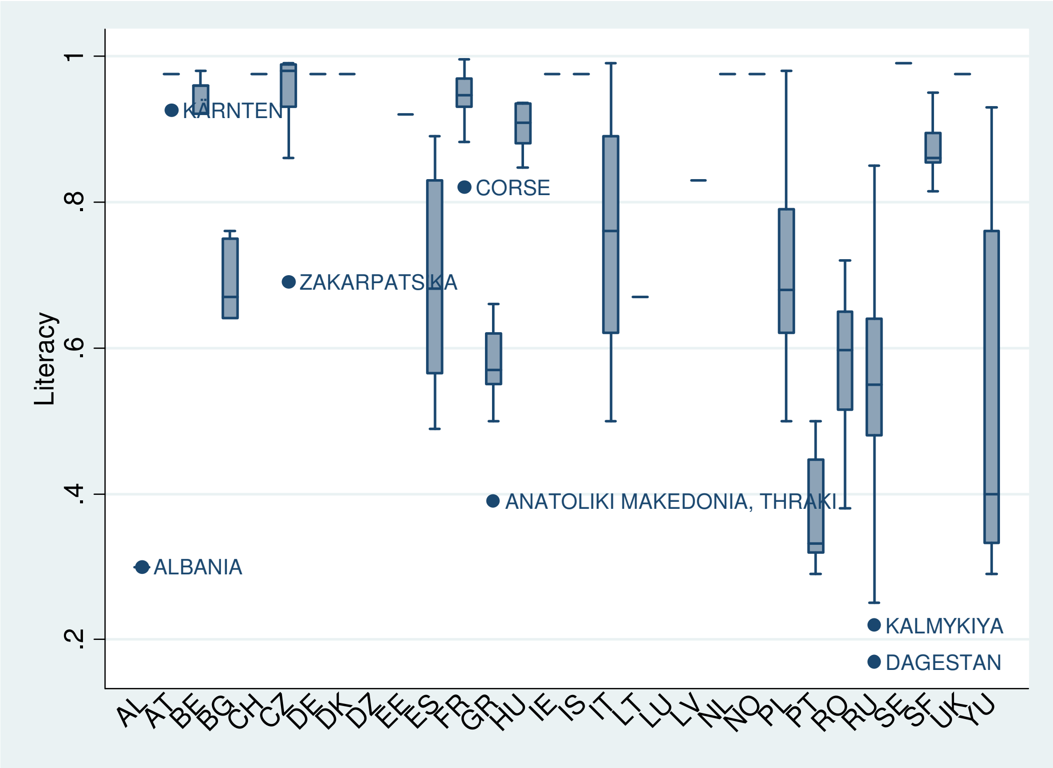


Figure 9.10 Regional variation in literacy by historical countries, 1930



Note: RU refers to the European part of the Soviet Union, CZ refers to Czechoslovakia.

Figure 9.11 Regional variation in literacy by current countries, 1930

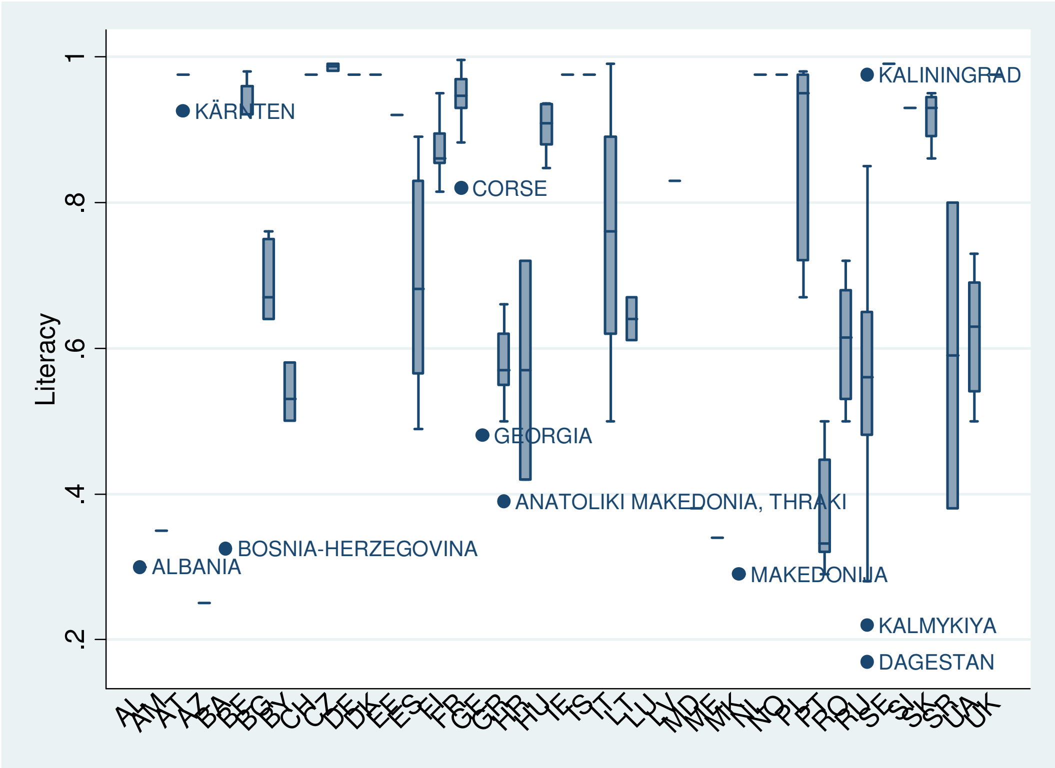
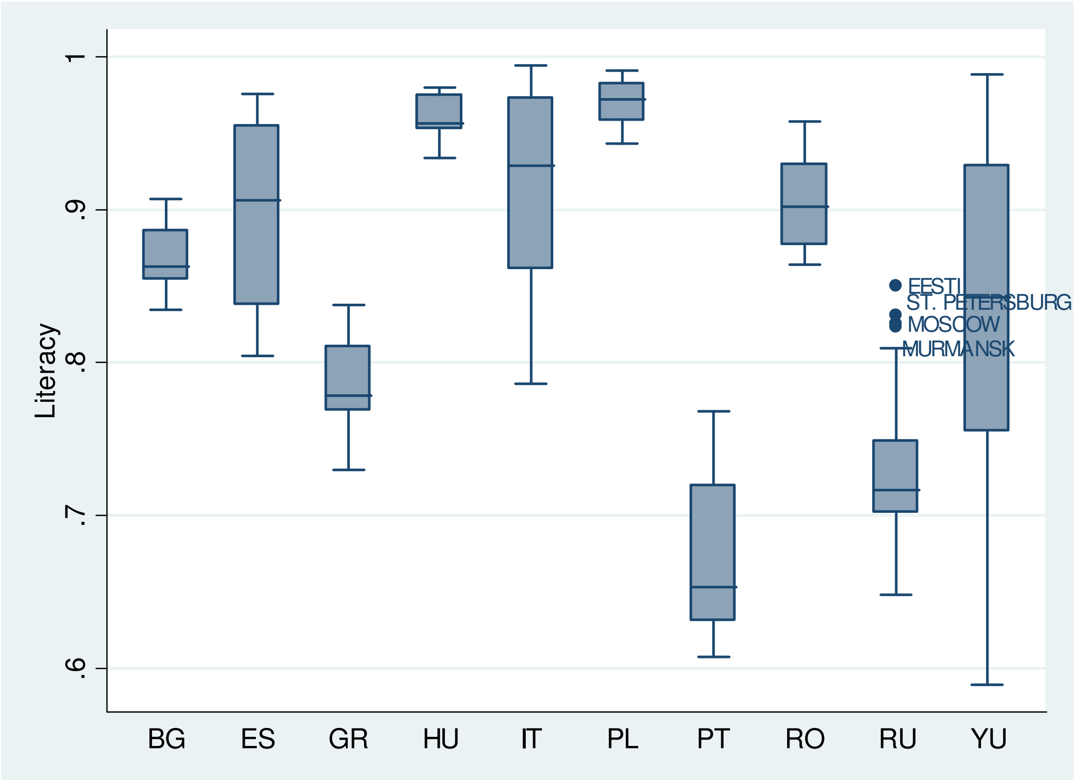


Figure 9.12 Regional variation in literacy by historical countries, 1960



Note: RU refers to the European part of the Soviet Union.

Figure 9.13 Regional variation in literacy by current countries, 1960

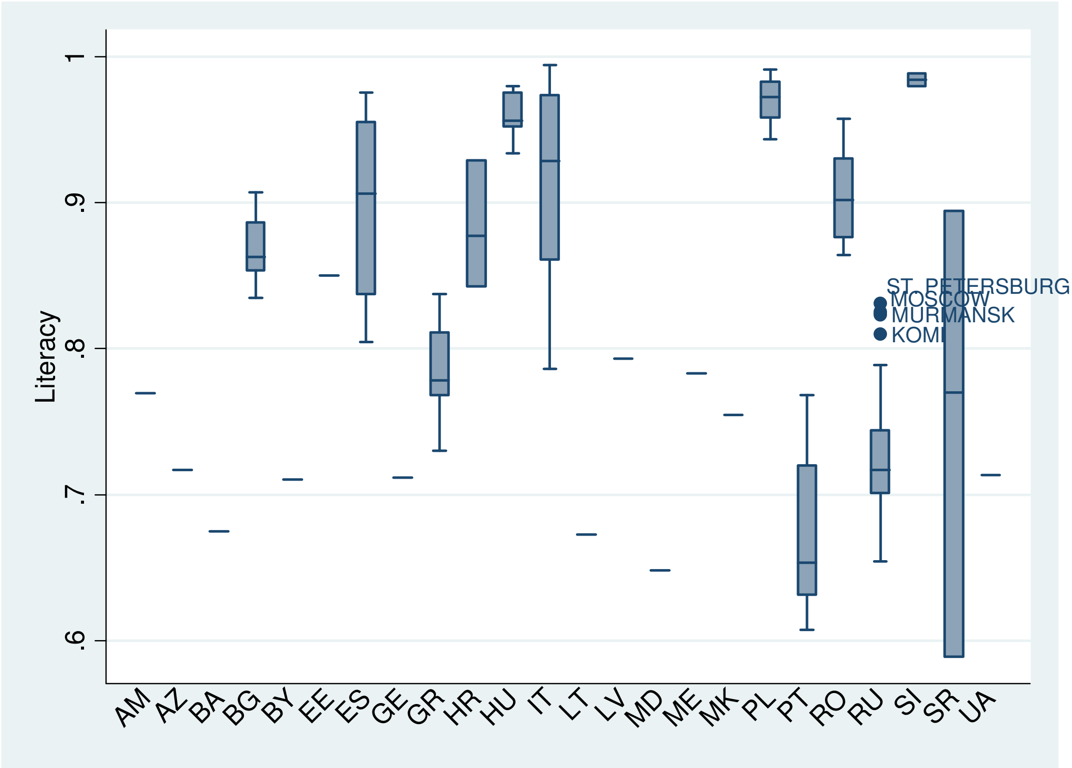


Figure 9.14 Regional variation in educational attainment by current countries, 2000

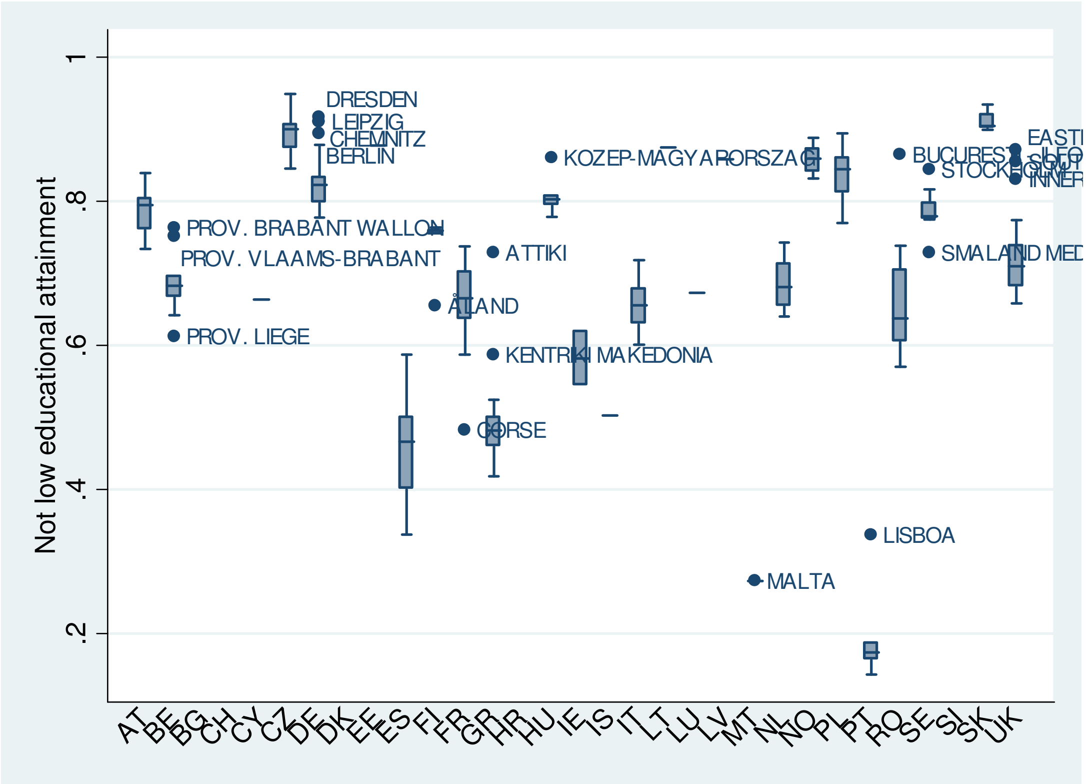


Figure 9.15 Regional variation in educational attainment by current countries, 2010

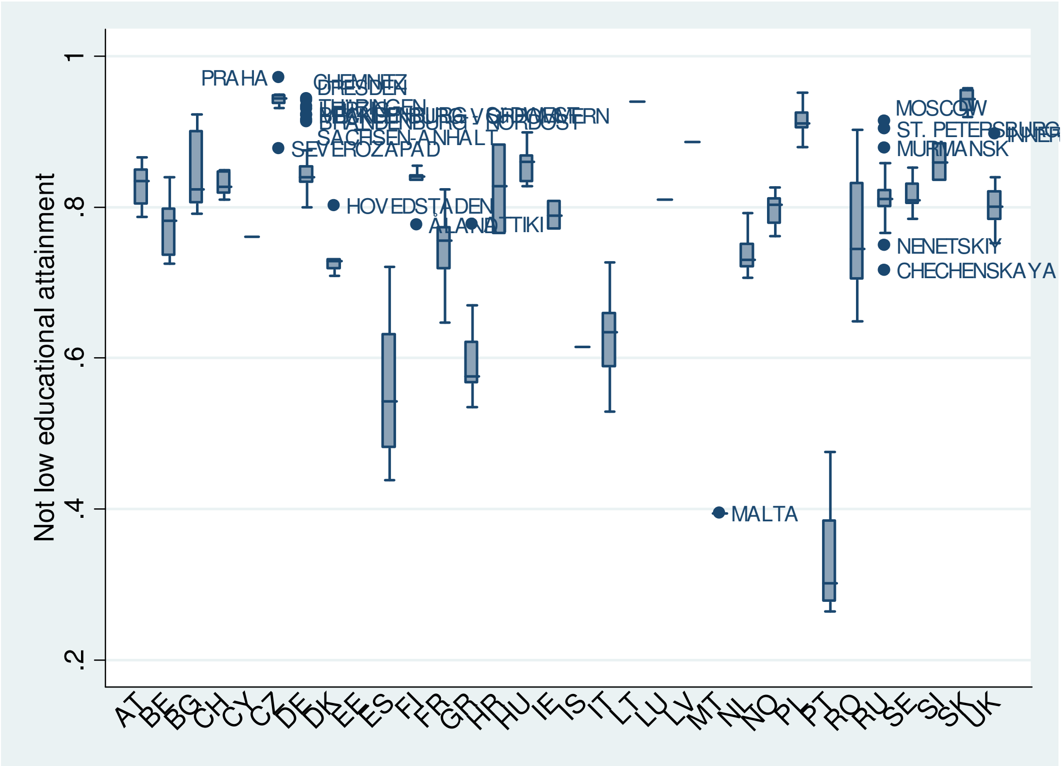


Figure 9.16 CVs over time for the European countries, 1850

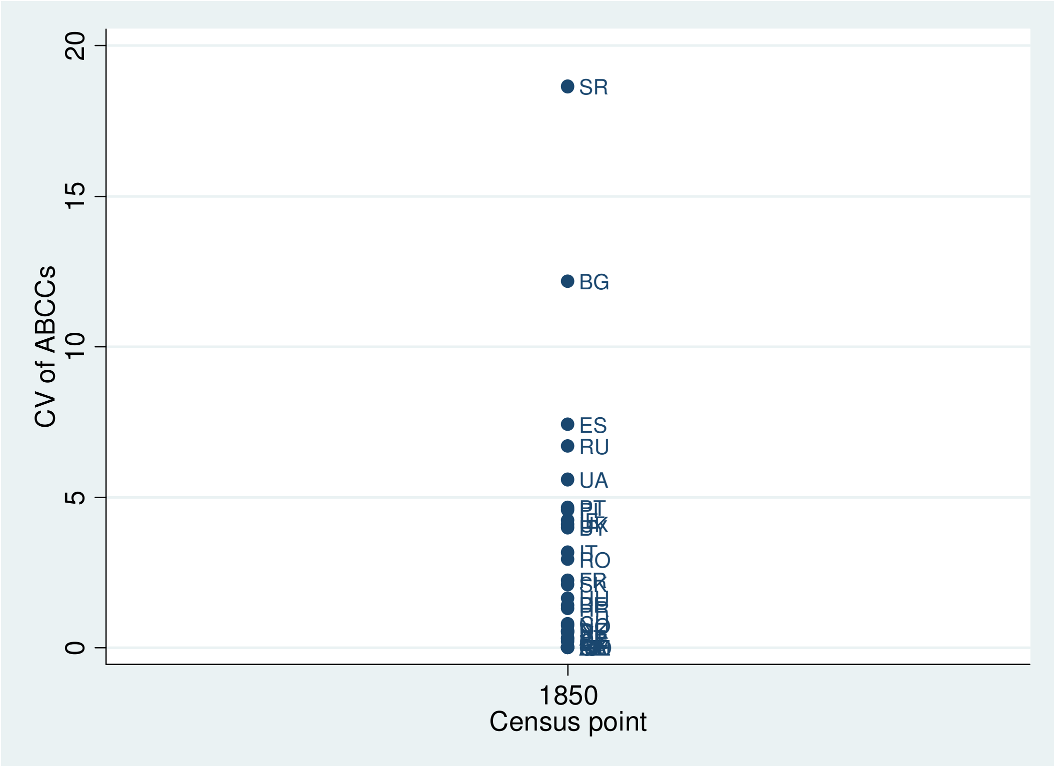


Figure 9.17 CVs over time for the European countries, 1900-1960

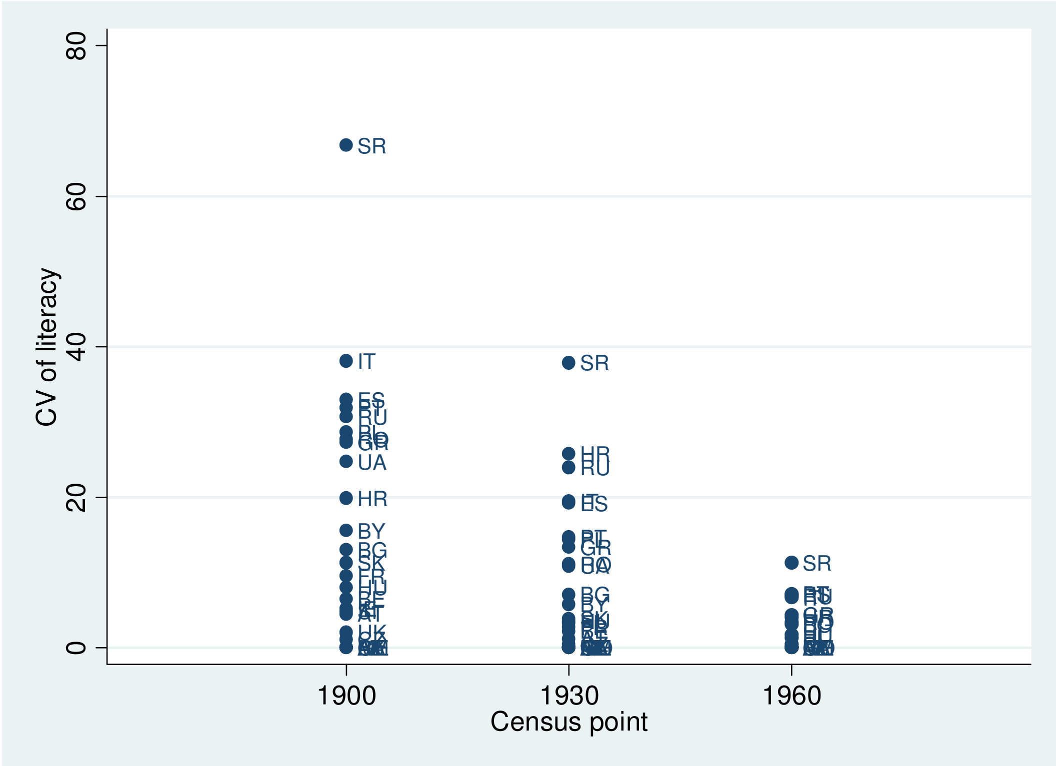
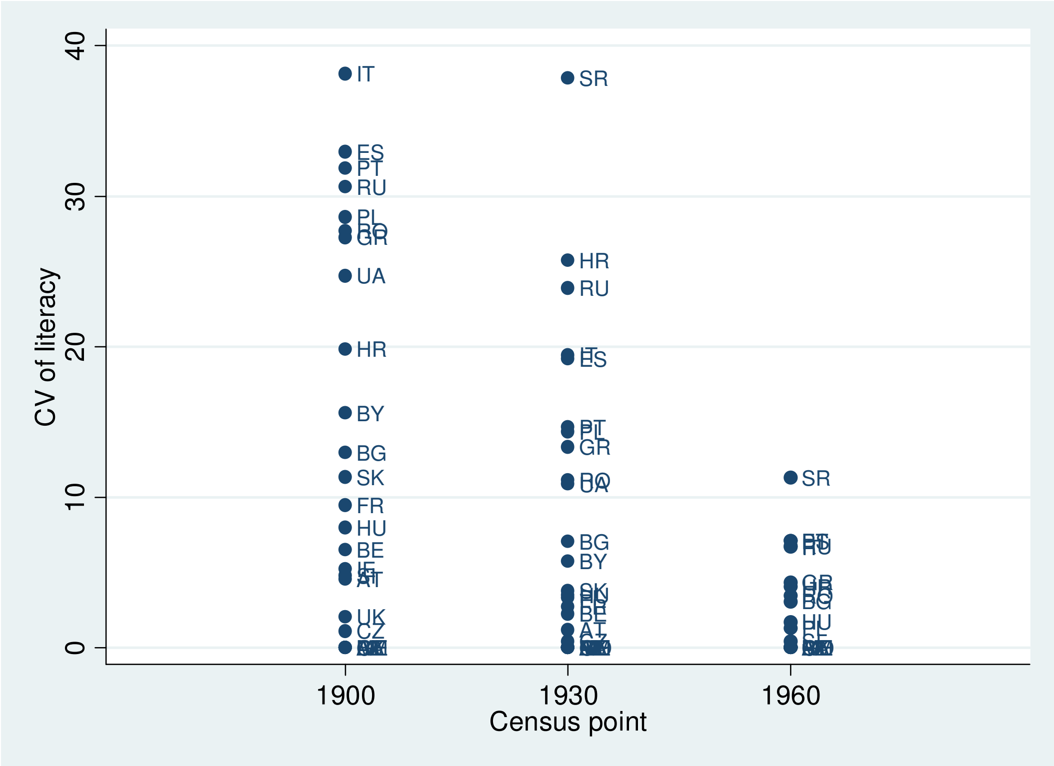


Figure 9.18 CVs over time for the European countries, 1900-1960 without outlier



Note: The concerned outlier is Serbia in 1900.

Figure 9.19 CVs over time for the European countries, 2000-2010

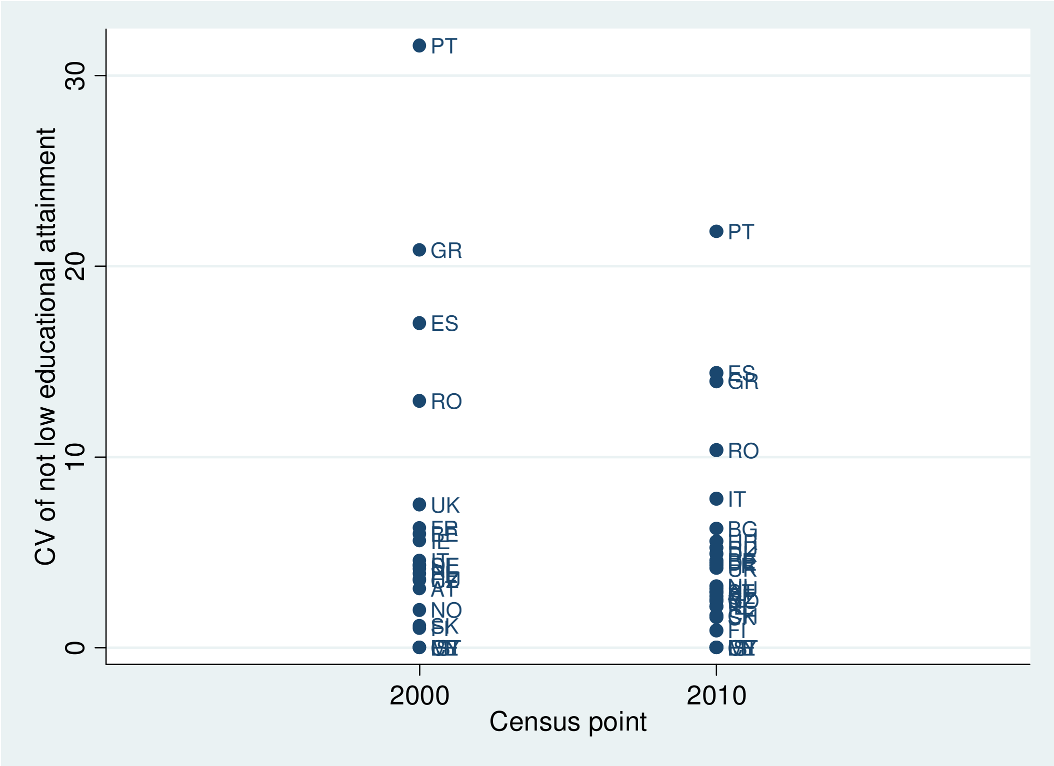


Figure 9.20 Gini coefficients over time for the European countries, 1850

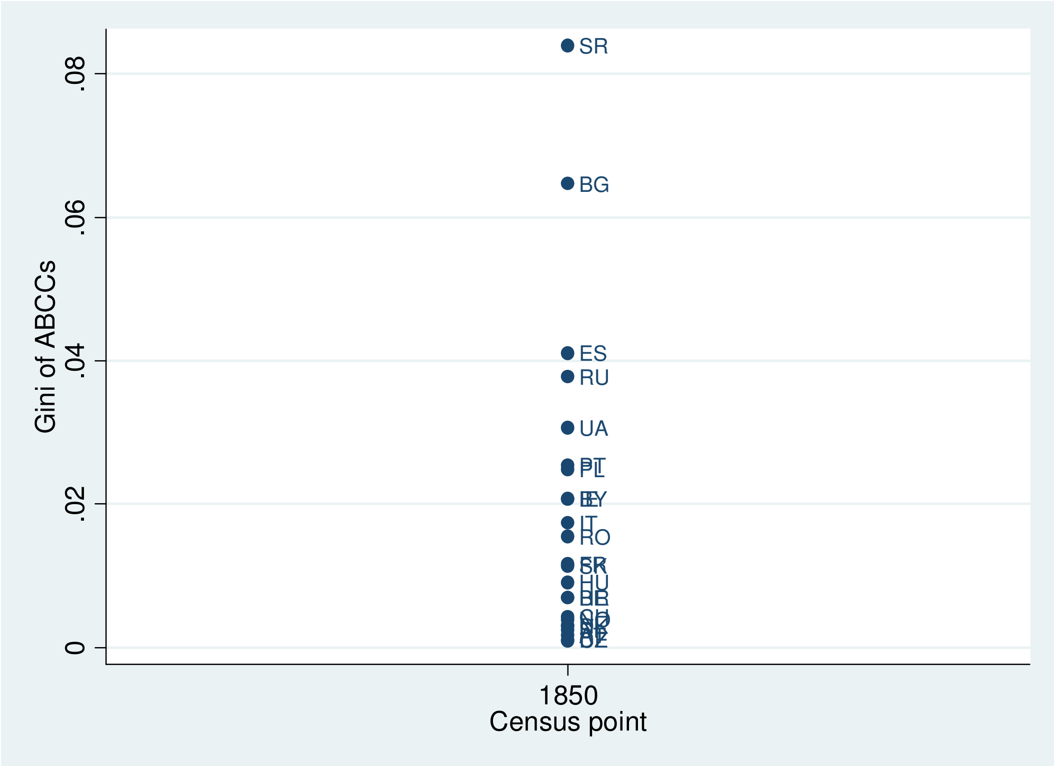


Figure 9.21 Gini coefficients over time for the European countries, 1900-1960

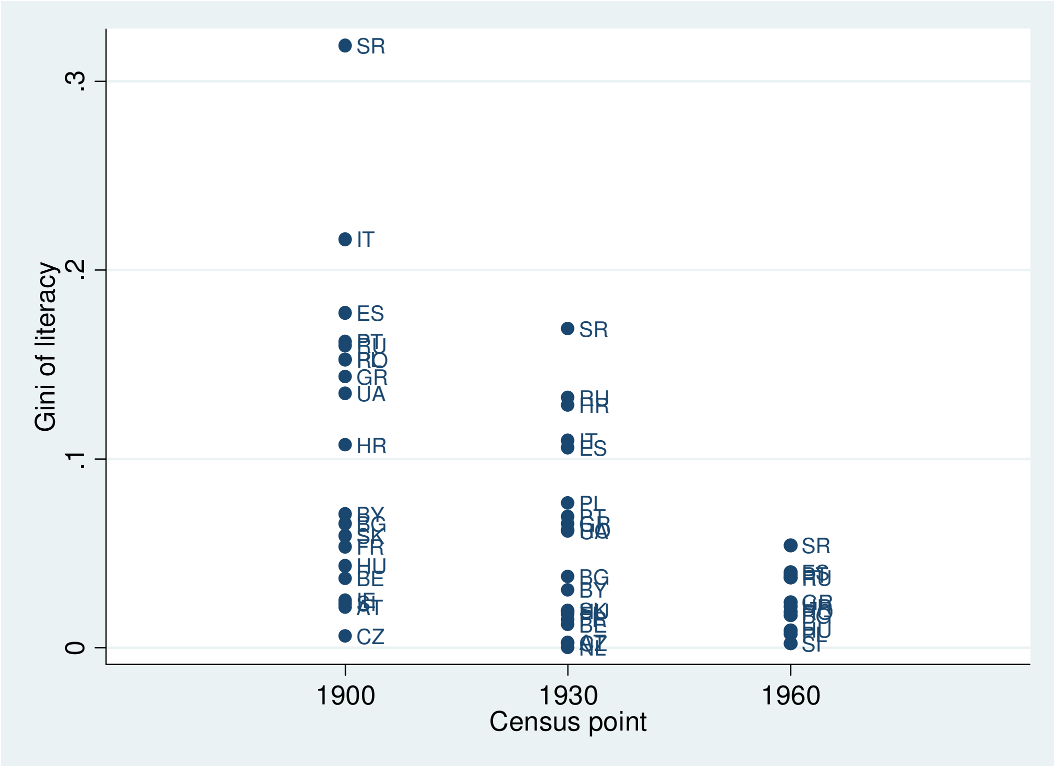
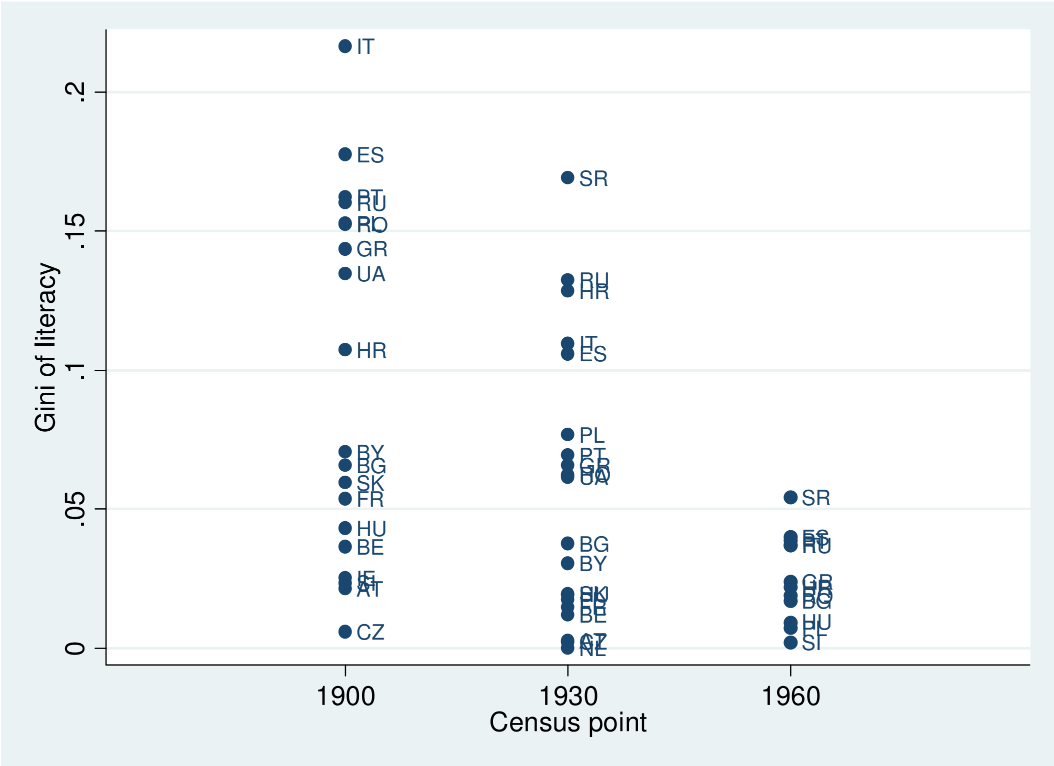
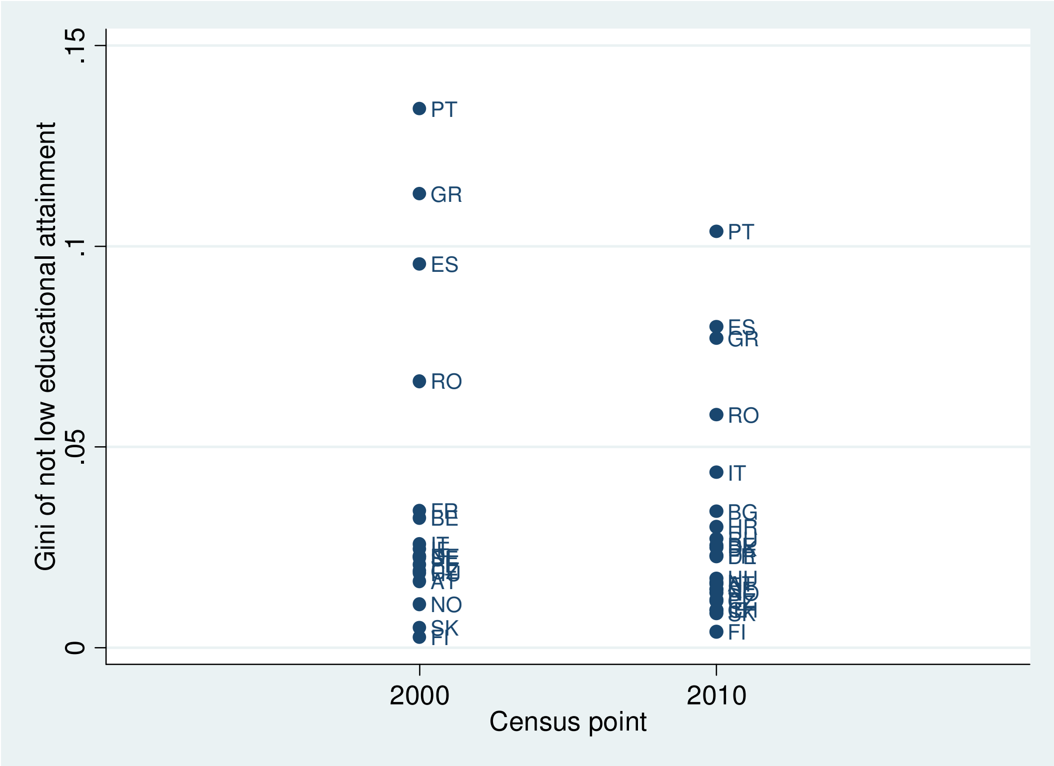


Figure 9.22 Gini coefficients over time for the European countries, 1900-1960 without outlier



Note: The concerned outlier is Serbia in 1900.

Figure 9.23 Gini coefficients over time for the European countries, 2000-2010



10. The long-run impact of human capital on innovation and economic growth in the regions of Europe

Human capital is an important factor for innovation and economic growth. This statement is often made in current academic research and policy debates. However, the long-run impact of human capital on current innovation and economic development has mostly not been taken into consideration, let alone at the regional level. This paper fills this gap in the literature and makes the link between the past and the present. Using a large dataset on regional literacy and numeracy in the 19th and 20th century, we find that historical regional human capital is a key factor explaining regional disparities in innovation and GDP per capita even today.

This chapter is based on a working paper which has been submitted to an academic journal.

10.1 Introduction

Economic growth is one of the predominant research areas in economics. Many theories have been developed to better understand the causes and consequences of economic growth. For example, some of the most important factors for long-run growth are the quality of institutions (e.g., North 1981, Acemoglu *et al.* 2005b, Fourie and von Fintel 2010), geography and naturally given geographical conditions (e.g., Diamond 1997, Engerman and Sokoloff 2000), income inequality (e.g., Alesina and Rodrick 1994, Persson and Tabellini 1994) and land inequality (e.g., Galor *et al.* 2009).

Another important factor is human capital accumulation, as has been put forward by prominent authors such as Glaeser (Glaeser *et al.* 2004) or Galor (Galor and Moav 2002, Galor 2005a, Galor and Moav 2006, Galor 2012). However, human capital is not only essential for economic growth as such but also for the creation of innovations. For instance, an increase in human capital may induce a rise in the number of innovative entrepreneurs and products, thus indirectly spurring economic growth through the channel of innovation (e.g., Dakhli and de Clercq 2004). In fact, the crucial role of innovation for economic growth has been underlined by a large literature in this area (e.g., Solow 1957, Romer 1986, Lucas 1988, Crosby 2000, Capello and Lenzi 2012, see also WIPO 2012). For these reasons, education and innovation are also important pillars of the EU's long-term strategy "Europe 2020" to boost economic growth in Europe (European Commission 2010a, b). Nevertheless, the long-run implications of human capital on innovation and economic growth need further research, even though this issue has been touched upon in different contexts (e.g., Baten and van Zanden 2008, Baten and Jaeger 2009). Therefore, the question remains whether pre-existing human capital is important for the creation of long-run growth.

Thus far, most of the studies in this area only take a national perspective by focusing on countries. However, regional differences in human capital may be even more important than nationals ones (e.g., Cipolla 1969). Considering regions allows to overcome the inherent problems of cross-country analyses and may explain why some regions are richer than others. Therefore, in this paper we analyse the long-run impact of human capital on innovation and economic growth at the regional level in Europe. To this end, we use the large database assembled by Hippe (2012a) which includes data on human capital levels between 1850 and 2010 for many European regions and countries. Furthermore, we add relevant data on innovation and economic growth and additional control variables that stem from a large range of sources. Regions are coded according to the NUTS classification throughout time. In total, we have up to 200 NUTS 2 (or corresponding) regions in our database. In this way, we are able to analyse the relationship between human capital, innovation and economic growth in a regional and long-run perspective.

The results show that historical human capital is a significant determinant of today's regional levels of innovation and economic development in Europe. In particular, literacy in 1930 has a significant influence on current patent applications per capita and GDP per capita. In addition, supplementary results for 1850, 1900 and 1960 confirm this finding. Therefore, our results suggest that historical human capital has important persisting effects on economic development until today.

The paper is structured in the following way. First, we present the relevant literature on the relationship between human capital, innovation and economic growth in Europe. Then, we discuss the employed methodology, the underlying data and our econometric strategy. Finally, we show the current relationship between human capital, innovation and economic growth and analyse the long-run relationship between historical human capital, current innovation and economic growth using OLS regression models. The last section concludes.

10.2 Literature

Human capital may directly affect economic growth or indirectly, in particular through the generation of technology. According to Acemoglu and Autor (2012), there are several channels through which human capital may affect technological progress. Firstly, they stress that the individuals with the highest talents may contribute to technological progress by the use of their human capital if they have the necessary access to educational facilities. These individuals probably influence most importantly technological progress. Secondly, the workforce in more general terms may affect technology, first, due to the externalities derived from human capital and, second, because human capital alters and increases the incentives to invest more in technological progress. For example, it is possible that a technology is only sufficiently profitable if there are enough workers who have the necessary skills. Finally, technological progress may be influenced by the workforce's mix of skills and human capital.

In general, the literature on human capital, innovation and economic growth as such is very large. Human capital was already considered in early works by Smith and Marshall (see Demeulemeester and Diebolt 2011). However, it took much longer for human capital to emerge as a key factor for economic growth. In fact, the most important contributions were developed from the middle of the 20th century onwards. In particular, Becker (e.g., see Becker 1964) is widely acknowledged as a founder of human capital theory, stressing that human capital increases the productivity of workers. Similarly, Arrow (1962) highlights the effect of experience on technical change. In addition, Nelson and Phelps (1966) emphasise that human capital is also important for implementing and adopting new technologies. Later on, Schultz (1975) argues that workers are better able to cope with changes in the economic structure and handle new technologies if they have more human capital.

Around the beginning of the 1990s emerged new theoretical advances. An extension of the original Solow growth model was presented by Mankiw *et al.* (1992), the so-called human-capital augmented Solow model. It explicitly included human capital as a factor in the Cobb-Douglas production function. Another kind of growth models, the endogenous growth models, was initiated by Romer (1986) and Lucas (1988). The former focuses on human capital accumulation and the latter on technological change. The aim is to endogenise the various factors which may lead to economic growth in the model. In general, these models consider human capital to be an important driver for economic growth.

Finally, the newest contribution in the area of human capital and economic growth are the Unified Growth models (e.g., Galor and Weil 2000, Galor and Moav 2002, Galor 2012). Their aim is to explain economic development in the (very) long run. In these models, human capital is attributed a crucial role for the creation of economic growth. They also incorporate the notion of a quantity-quality trade-off which originally stems from Becker (1981). According to this theory, parents face a trade-off between the quantity (number) and the quality (education) of their children. Whereas the quantity of children prevailed during most of human history, parents began to prioritise child quality in the course of development. The increased investment in human capital spurred technological progress and economic growth. Ultimately, more child quality meant less quantity of children, reducing the number of children, leading to lower fertility rates and causing the demographic transition. Modern growth, therefore, is characterised by the negative relationship between GDP per capita and fertility rates (see also Galor and Weil 1999, Galor and Weil 2000).

All in all, these different theories show that human capital in an important driver for economic growth. Still, there has been some controversy about this issue over the last decades. In fact, Demeulemeester and Diebolt (2011) refer to several alternating waves of optimism and scepticism on the relevance of human capital to generate growth since the Second World War. The contributions by authors such as Solow (1956), Mincer (1958), Schultz (1961) and Becker (1964) led to the consensus in the 1950s and 1960s that education makes an important contribution to economic growth. In contrast, the 1970s where more marked by scepticism in a time of economic downturn. The new important theoretical contributions of the 1990s (Lucas 1988, Romer 1990) reinvigorated once again the case for human capital. These optimistic ideas were supported by different empirical studies (e.g., Barro 1991, Mankiw *et al.* 1992, Barro and Lee 1993) but also more critical voices appeared such as Benhabib and Spiegel (1994) and Pritchett (2001) (Demeulemeester and Diebolt 2011). Measurement error may account for some of these results (Krueger and Lindahl 2001). Thus, Sianesi and Van Reenen conclude in their literature survey in 2003 that "as a whole we feel confident that there are important effects of education on growth" (Sianesi and Van Reenen 2003, p. 197). Moreover, the more recent studies by, e.g., De La Fuente and Doménech (2006), Cohen and Soto (2007) and Ciccone and Papaioannou (2009) show the crucial impact of human capital on growth. The key contribution of cognitive skills (including numeracy and literacy skills) is further highlighted by Hanushek and Woessmann (2008).

In addition, Goldin and Katz (2008) analyse the role of human capital for the creation of economic growth and world leadership in the United States. They come to the conclusion that the United States led the world in terms of education in the 20^{th} century, enabling the US to become the economic leader in the world. Moreover, both human capital and technological progress affect inequality in the society. Whereas technological progress tends to widen the differences between different skills groups (i.e., low skill and high skill groups), investing in human capital may counter this evolution which otherwise would lead to more inequality (see also Galor and Weil 2000). In other words, skills are assumed to be more demanded when technological progress advances and this new demand may be set off by additional investments in human capital. Thus, inequality increases if the demand for skills is higher than its supply and decreases in the opposite case. Education is thus shown to be the central element for the explanation of economic growth in the US. Furthermore, by enlarging the analyses of the approach taken by Goldin and Katz, Acemoglu and Autor (2012) find that human capital is still even more important for economic development.

The literature on the impact of human capital and innovation on economic growth in the European regions is also quite large (e.g., Fagerberg *et al.* 1997, Cappelen *et al.* 1999, Badinger and Tondl 2003, Rodríguez-Pose and Crescenzi 2008, Sterlacchini 2008, Ljungberg and Nilsson 2009, Cuaresma *et al.* 2012). For example, Badinger and Tondl (2003) investigate whether human capital and innovation (as measured by patent applications) have a significant impact on the growth rates of Gross Value-Added per capita in 128 regions between 1993 and 2000. Both the relative patent applications and higher education variables are shown to have a significant impact. However, medium levels of education are not significant which highlights that economic growth in Europe's 'knowledge-driven' economies is boosted by the highest form of educational attainment. Moreover, Sterlacchini (2008) finds that human capital (in the form of higher education) and a region's knowledge base have a significant and positive impact on economic growth in twelve EU15 countries between 1995 and 2002. Finally, Cuaresma *et al.* (2012) use a data set including 255 EU regions to analyse which of their 48 potential determinants are significantly explaining economic growth between 1995 and 2005. Two of their most important results are that capital regions grow faster than other regions and that human capital (i.e., higher education) is a robust determinant of economic growth. Thus, these different studies show that human capital is an important determinant of economic growth in the European regions today.

But what do we know about its long-term impact in the world in general and in Europe in particular? There have some been studies which shed some light on the question whether historical human capital and technology matter for today's economies. For instance, Comin *et al.* (2010) take a long-run perspective and show that there is a strong relationship between technology in 1500 AD and current GDP per capita as well as technology adoption in the world. Moreover, Baten and Jaeger (2009) find in their sample that human capital and patents in 1910 have a strong and significant effect on GDP per capita in 1960. These findings suggest that historical factors may be important for the explanation of current or recent economic levels.

We advance this line of research by focusing on regions instead of countries in a European perspective. Using regions instead of countries considerably sharpens the picture because countries may be composed of regions which are very different to each other. Regional differences may thus be very high. However, this information is lost in employing average for countries. Unfortunately, regional data for innovation proxies such as regional patent applications are not available in Europe for the past. For this reason, it is impossible to identify the interconnection between human capital and patent applications at the regional level in the past in all of Europe. Still, it is possible to show the long-run effects of human capital on innovation and economic growth, using regional historical human capital data and current innovation and economic growth data. Can we thus find a persisting effect of historical human capital on today's economic growth and innovation?

10.3 Methodology and data

In our study it is very important to clarify first the underlying concepts. Human capital, innovation and economic growth are rather large and vague ideas whose measurement has to be specified in greater detail.

The human capital data used in this study come from different sources. First, we employ the new and large database created by Hippe (2012a) which traces human capital between 1850 and 2010. From this database, we use the years 1850, 1900, 1930 and 1960 to follow the evolution of human capital. Human capital is proxied by numeracy (ABCC) in 1850 and by literacy (ability to read and write) in 1900, 1930 and 1960. Additional data on educational attainment have been added from Eurostat (2012).

Both numeracy and literacy indicators may be considered appropriate for their respective time period. Before 1900, literacy data are not available for many European countries. Even in 1900 a range of countries do not consider literacy in their censuses. This is the case for e.g. Scandinavian countries such as Denmark or Sweden but also for Germany, Switzerland or the Netherlands. In general, these are countries where basic reading and writing skills can be considered almost universal. They had their own specific reasons to refrain from this question in the census. For example, the Swiss administration considered that a sufficient literacy level was already attained in 1860, as the corresponding 1860 census documents highlight (Statistisches Bureau 1862). According to the census materials, military data had shown that 93 % of recruits were able to read and write in the Bern region and even 100 % of recruits were literate in the Solothurn region already at the middle of the 19th century. Similarly, the Netherlands had already very high literacy levels if one considers recruitment data: only 15 % of recruits were illiterate (not or only unsatisfactorily able to read and write) in 1857/1858 (Statistisches Bureau 1862).

These examples highlight the very high levels in literacy which existed in (probably all of) the countries where literacy was not asked in the census anymore at the end of the 19th century. For this reason, it appears more suitable to use another indicator for the earliest point in time. Numeracy as proxied by the age heaping method is the appropriate choice because, first, it is – as literacy data later on – directly derived from censuses. Second, it refers broadly to the same population (the entire population, excluding certain age groups). This allows a better comparison of both indicators. Taking military data from recruits would not allow to take the major parts of the population into account but only a very small selected group: men, in military service, of rather younger age and limited to a defined small age range. Moreover, regional data are often not available.

In consequence, numeracy is the appropriate indicator which is also available for almost all European regions around 1850. Numeracy is measured by the age heaping method which has been used in an increasing number of recent publications (A'Hearn *et al.* 2009, Manzel and Baten 2009, Crayen and Baten 2010a, Hippe and Baten 2012a). The method takes advantage of the fact that in historical censuses (and also still in part in censuses in developing countries, see Hippe 2012b) there is a heaping phenomenon on ages particularly ending on 0 and 5. One can show that individuals were not able to calculate their own age, so that they did not report their exact age but only a rounded age.

The deviation from the ideal age distribution (where all ages are represented by the same share) can be employed to create an index measuring numeracy. This index has originally been the Whipple index (WI) but has recently been improved by the ABCC Index (see A'Hearn *et al.* 2009). This index has the same value range as literacy (0 to 100 percentage points or simply points) which makes comparisons much easier.

Therefore, we employ the ABCC Index also in this study. It is defined as

$$ABCC_{jt} = 125 - 125 \times \left(\sum_{i=5}^{14} n_{5i,jt} \Big/ \sum_{i=23}^{72} n_{i,jt} \right), \qquad (10.1)$$

where i is the number of years, j is a region, t is the point in time and n is the number of individuals.[114]

Second, literacy was the standard education variable around the turn of the 20th century and the first half of the 20th century in many European countries. Illiteracy had to be eradicated – this was a common tenor in all European countries. Success, however, was quite different not only in these countries but also *within* these countries. For this reason, a completely literate population was not achieved in many European countries in 1900 and still in 1960 illiterates were more or less common in many European countries. This fact underlines our methodology to use literacy as our human capital indicator for the period. After 1960 one may presume that the ability to read and write is more or less attained by the entire population so that other education variables have to be used. We define literacy as

$$Literacy_{jt} = \sum_{i=10}^{N} rw_{i,jt} \Big/ \sum_{i=10}^{N} n_{i,jt}, \qquad (10.2)$$

where *rw* is the ability to read and write and *N* is the total number of years. The age definition is the standard contemporary definition and has also been used by Kirk (1946), whose literacy data we use for 1930. However, the employed literacy definition is not always used by all countries in 1900, 1930 and 1960. Kirk (1946) states that differences due to the age definition are minor in 1930. For the data in 1900 and 1960, the data have been collected in such a way to correspond the most exactly possible to the standard definition, thus avoiding potential biases.

---

[114] See Hippe (2012a) for more details.

Finally, educational attainment is of the standard human capital indicators in research for the present. Educational attainment can include different educational levels, i.e., primary, secondary and tertiary education. Depending on the context and the interest of the researcher, educational attainment measures the share of individuals which have attained primary, secondary or tertiary education.[115] In our study, we define educational attainment as

$$Eduatt_{jt} = \sum_{i=20}^{64} st_{i,jt} \Bigg/ \sum_{i=20}^{64} n_{i,jt}, \qquad (10.3)$$

where *st* is the number of individuals with upper secondary or tertiary education as their highest level of education.[116] The data in our database refer to 2000 and 2008.

Furthermore, innovation is a very large concept and is difficult to be measured statistically. One standard way is to take the number of patent applications or grants (e.g., Acs *et al.* 2002, Diebolt and Pellier 2009a, b, c, Diebolt and Pellier 2011, Diebolt and Pellier 2012). Historically, intellectual property rights in the form of patents go back to the 15th century. The first predecessors of patent rights were granted in the trade hotspot of Venice (the so-called *Parte Veneziana*) (Machlup and Penrose 1950, Diebolt and Pellier 2009b). In this way, inventors had privileges over their inventions during ten years. Subsequently, the second important law in the history of patents is the 'Statute of Monopolies' in England in 1623. This statute gave monopoly rights to inventors which had created new inventions and outlawed monopoly privileges in all other cases. Afterwards, the first modern law on patents was passed in the United States in 1790, quickly followed by the first respective French law in 1791. Thus, the English, American and French patent

[115] Years of schooling is an alternative expression based on the similar idea, only that not the share of individuals at a certain level is taken account of but the total number of years.
[116] These are the levels 3 to 6 according to the International Standard Classification of Education (ISCED-1997) (UNESCO 1997).

laws inspired laws in the other European countries during the 19th century (Machlup and Penrose 1950, Diebolt and Pellier 2009b). An overview on the first patent laws in the different European countries is given in Table 10.1. According to the data presented, the first countries with modern patent laws were France, the Netherlands, Austria, Russia, Sweden and Spain. In some countries patent laws were only passed after World War I, such as Greece. Overall, it took more than a century after the French revolution for all European countries to pass modern patent laws.

In addition to patent applications, other variables that are used to measure innovation include investments in R&D (e.g., Cohen and Levinthal 1989), changes in productivity (David 1990, Von Tunzelmann 2000), bibliometrics (Andersen 2001) and data on (international) expositions and fairs (Moser 2005). Patent statistics have certain setbacks; for example, organisational changes or know-how cannot be patented and not all patented products become innovations (Griliches 1990). Nevertheless, patents are generally considered to be the best indicator (e.g., Cantwell 1989, Anderson 2001) and are most frequently employed (Diebolt and Pellier 2009b), in particular for the past. Therefore, we use patent applications per million inhabitants to the European Patent Office (EPO) as our indicator of innovation. The regional data come from Eurostat (2012).

Lastly, economic growth is measured in a standard way by GDP per capita (in PPS) as presented by Eurostat (2012).

We use scatter plots and regression models to analyse the relationship of regional human capital, innovation and economic growth. For the influence of historical human capital on current innovation and economic development, we employ standard OLS regression frameworks which are formulated in the following way:

$$ln(Patents/c_j) = \beta_0 + \beta_1 H_j + X_j + \varepsilon_j \qquad (10.4)$$

$$ln(GDP/c_j) \qquad = \beta_0 + \beta_1 H_j \ + X_j + \varepsilon_j, \qquad (10.5)$$

where *ln(Patents/c)* is the number of patents per million inhabitants (in logarithmic terms), *ln(GDP/c)* is GDP per capita (in PPS and in logarithmic terms), H is the human capital indicator, X are other explanatory variables, j is a region and ε are the unexplained residuals.

X is composed of different variables which may have an influence on economic development. In particular, in 1930 these are total fertility, marital status, population density (in logarithmic terms), the share of individuals not dependent on agriculture, infant mortality, a dummy for capital regions, a dummy for the newer EU regions and country dummies. There is a large literature showing that (total) fertility can have an important effect on growth (e.g., Barro and Becker 1989, Becker *et al.* 1990, Galor and Weil 1996, Galor and Weil 2000, Galor 2012). In our case, we use the total fertility data provided by the famous Princeton European Fertility Project, which defines total fertility as “a measure of the fertility of all women in the population” (Coale and Treadway 1986, p. 154).

Moreover, marital status comes from the same source and is “the ratio of the number of births produced by married women in […] a population to to the number that would be produced if all women were married” (Coale and Treadway 1986, p. 154). In other words, according to Watkins this measure represents “the proportions married at each age” (Watkins 1986, p. 315) and can thus be used as a proxy for nuptiality. There have been important nuptiality differences in Europe in the past, as has most famously been put forward by Hajnal (1965). According to this author, western Europe was characterised in the past by a specific European Marriage Pattern (EMP) which was unique in the world. It consisted basically in much lower average ages at marriage in western Europe than in other parts of the world, including eastern Europe. Thus, differences in the average age at marriage may also explain differences in economic development (de Moor and van Zanden 2010, Foreman-Peck 2011).

The supplementary data, which are only available for 1930, come from Kirk (1946). Population density is measured (in logarithmic terms) as the number of individuals per square kilometre. Population density, as Klasen and Nestmann (2006, p. 623) point out, "generates the linkages, the infrastructure, the demand and the effective market size for technological innovations". In this way, it may foster innovations and economic development in the long run.

The next variable is the share of the total population which is not dependent on agriculture. This share roughly proxies the regional economic development in 1930. Shares of agriculture or industry have been used in different historical publications where GDP per capita estimates are not available (e.g., Good 1994, Hatton and Williamson 1994, Becker and Woessmann 2009). Indeed, although we cannot show the relationship for historical GDP per capita estimates due to lack of data, Figure 10.1 shows that there is a relationship between this historical share and current GDP per capita. Some outliers are apparent, such as Luxembourg (LU00) and Brussels (BE10) but the general pattern clearly holds.

In addition, infant mortality may represent a variable related to health and may promote economic growth through different channels such as population growth (Kalemli-Ozcan 2002).

Moreover, the capital region dummy has been introduced because capital regions have often specific characteristics due to their administrative functions. The dummy for the newer EU regions captures the fact that these countries joined the EU later on and have had different historical and economic experiences in the past, having mostly been part of the Communist bloc before the fall of the Soviet Union. More specifically, these regions come from the newest 12 EU members (Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Malta, Poland, Romania, Slovenia and Slovakia). For this reason, West Germany is also part of the ‘old’ member states, while (former Communist) East Germany is considered as part of the ‘new’ states even though it was already reunified with West Germany in 1990. Finally, there may be different inherent characteristics of countries (e.g., institutions) which may bias the results. Therefore, the inclusion of country dummies allows to control for these country fixed effects.

Most variables are available for 1930, which is why we focus in our analysis on this year. Additionally, we supplement our evidence for 1930 with data for 1850, 1900 and 1960. This enables us to show the differences over time when explaining current regional economic disparities.

Descriptive statistics on all variables are shown in Table 10.2. We have up to almost 200 regions in our dataset. However, we have to consider the question how a region is defined in this paper. Clearly, the regions in 1930 and at other points in time are not the same as today, at least for most European countries.[117] For this reason, the historical regions have been adapted to the NUTS classification of the European Union (see also related work by e.g., Hippe 2012a, b). More precisely, we use NUTS 2 regions as our standard regional classification, which is also done in the relevant literature in regional economics (e.g., Badinger *et al.* 2004, Herwartz and Niebuhr 2011, Scherngell and Barber 2011).

Moreover, note that the availability of the data can be quite different at each time period. In particular, the Eurostat data for the current period (2000 and 2008) refer only to

[117] For example, Spain and France have preserved almost the same regions and regional boundaries until today.

countries of EU27, EFTA and some Candidate countries. For this reason, the corresponding regressions only consider these regions.

On the other hand, whereas the ABCC data for 1850 consider most of the European regions in the larger sense, the literacy data for 1900 and 1930 only refer to those countries where literacy was still measured. Still, most countries can be included in this study. In contrast, literacy in 1960 is only available for a reduced number of countries (see appendix). For this reason, the results for the data for 1960 are less comparable than for the other time periods. Still, they allow to get some additional insights for the respective regions at the beginning of the second half of the 20th century.

10.4 Results

10.4.1 Regional economic development, innovation and human capital in the European regions today

Before analysing the long-run impact of human capital on innovation and economic growth, we consider the relationship of the latter two. Figure 10.2 and Figure 10.3 show this relationship for 2000 and 2008. In summary, both figures highlight a general positive relationship between GDP per capita and patent applications per million inhabitants to the EPO in Europe in the present.[118] However, the 'new' member countries did virtually not apply to the EPO at all in 2000. This phenomenon has become less striking in 2008 when more and more of these regions begin to increase innovation. The most important outliers are Inner London (UKI1), Brussels (BE10), Luxembourg (LU00) which had much higher GDP per capita levels than their relative number of patent applications would suggest. On the other hand, Germany's core industrial zones in the greater region around Munich (Oberbayern, DE21) and Stuttgart (DE11) apply much more often than any other regions in

[118] Note that data are available for more regions in 2008 than in 2000.

2000. In 2008 they are still in the leading group, although their extraordinary number has strikingly decreased. Finally, the lowest GDP per capita are found in particular in the two newest member states, Bulgaria and Romania.

In addition, Figure 10.4 and Figure 10.5 depict the relationship between GDP per capita and this time the share of persons that have obtained a higher form of education. The differences between the western and the (ex-Communist) eastern European regions are quite evident. In both 2000 and 2008, there is a positive relationship between the two variables. In the new member states, the relationship is still clearer than in the 'old' regions, having GDP per capita values mostly on a lower level. Consequently, the highest shares of higher education are found in East Germany, Prague (CZ01) and Bratislava (SK01).

10.4.2 The influence of historical human capital on regional economic development and innovation today in the European regions

In the next step, we use standard OLS regression models to dig deeper into the relationship between human capital and innovation on the one hand and between human capital and economic growth on the other hand. We start by taking a look at current indicators (Table 10.3) We only include those control variables for which we have also similar historical data later on. The regressions suggest that the employed human capital indicator (i.e., educational attainment) has a positive and significant impact on GDP per capita and patent applications per capita in 2008. Fertility does not appear to be important in the European regions today for economic development, also because the European regions have already accomplished the demographic transition.

The results for historical human capital indicators and other explanatory variables are highlighted in Table 10.4. Note that we always include country dummies to control for country fixed effects. We begin this part with the relationship between patent applications per million inhabitants and historical economic and social indicators for 1930 because we have the highest number of available explanatory variables at this point in time.

In each case, human capital is a significant positive explanatory variable of current patents per million inhabitants, mostly even at the 1 % level. In fact, literacy is the only significant explanatory variable in 1930 including all other available explanatory variables (column 1). When only those variables where data are also available for the other time periods are considered, the dummy for the newer EU regions turns significant (column 2). This means that the newer EU regions have a lower number of patent applications per capita than the older EU regions (*ceteris paribus*). This negative sign in all cases corresponds to the descriptive evidence shown in the figures above.

Considering the data for 1960 (column 3), literacy is again highly significant. This time the human capital variable is joined by total fertility which is negatively significant. This means that regions with higher total fertility in 1960 have lower patent applications per capita in 2008. This finding fits well with the literature on the negative economic impact of higher fertility. However, note that the data include fewer countries in 1960 than in 1930 (and in 1900 and in 1850), i.e., for those where literacy data are still available.[119] For this reason, the results are not directly comparable to those in 1930.

Finally, the data for 1900 (column 4) and 1850 (column 5) further highlight the importance of human capital on current economic regional disparities in innovative activities in Europe; the human capital indicator is each time positively significant. All in all, these results confirm the positive long-run impact of regional human capital. Is this also true for GDP per capita?

[119] These countries are Bulgaria, Greece, Hungary, Italy, Poland, Portugal, Romania, Soviet Union, Spain and Yugoslavia.

In Table 10.5, we reproduce the results with GDP per capita as the dependent variable. Globally, the results are similar to those previously shown. Column (1) shows that literacy is a significant and positive determinant of GDP per capita in 2008 at the 1 % significance level. This confirms the hypothesis that historical human capital is important for economic development in the long run. Marital status is negatively significant, meaning that those regions where couples married on average earlier to that time have lower current GDP per capita. This appears to be in line with the assumption that early marriage may have negative consequences on economic development. Moreover, the capital regions have significantly higher and the newer EU regions lower GDP per capita. The other explanatory variables do not have any significant effect. Column (2) shows the same regression but only includes those explanatory variables for which data are available at the other historical data points. This change affects several variables. Fertility in 1930 is now significantly negatively related to GDP per capita (instead of marital status) and the newer EU regions do not have significant lower GDP per capita anymore. In turn, the exclusion does not affect the significance of literacy, underlining the robustness of this finding. For this reason, we can conclude that the level of literacy in 1930 significantly predicts current levels of economic development.

We run the same regressions also for other years in the past. The conclusions are quite analogous in 1960 to those in 1930 because literacy in 1960 is once again positively significant and fertility negatively.

Going back to the beginning of the 20th century and earlier, we obtain once again similar results for the impact of human capital on current economic development. Column (4) depicts the results for literacy in 1900 and column (5) for numeracy (ABCC) in

1850.[120] In both cases, the human capital indicator is highly positively significant, whereas marital status is negatively significant. Capital and new EU regions are also significant.

These results for human capital and other indicators at different historical points in time suggest that those regions that had a higher endowment in human capital in the past, that is even more than one hundred years ago, have higher GDP per capita levels today than those regions which lagged behind. Moreover, capital regions are more prosperous than other regions.

To conclude, we find a positive and significant influence of historical human capital on current innovation and economic development. Human capital appears to be the most important factor contributing to today's innovation and economic growth in our analysis. This suggests that human capital formation in Europe at the regional level is an important driver of economic development in the long run.

10.5 Conclusion

This paper has focused on the relationship between human capital, innovation and economic growth in the European regions in a long-term perspective. There already exists a large literature on the effects of human capital on economy growth (e.g., Demeulemeester and Diebolt 2011). However, the existing evidence for the regional level in Europe remains scarce. Therefore, by using a large and new dataset we analyse the relationship between historical human capital and current economic indicators in the European regions.

We have employed different indicators of human capital, innovation and economic growth. These proxies are numeracy, literacy and educational attainment for human capital, patent applications per million inhabitants for innovation and GDP per

[120] Note that total fertility and marital status data are not available for 1850 but only for 1870. Therefore, we take this point in time as the closest proxy for 1850.

capita (in PPS) for economic growth. Regions have been defined according to the NUTS classification system set up by the European Union to allow a maximum of comparability throughout time. Human capital data are available for countries which still reported literacy or allowed the calculation of numeracy estimates in 1850, 1900, 1930 and 1960. In contrast, countries such as the Sweden, the Netherlands, Germany or the UK did not include information on literacy in their official censuses. Kirk (1946) considers that these countries had (almost) achieved full literacy. In consequence, these countries are not part of our analysis.

The results show that human capital is the most significant historical factor to explain current patent applications per capita and current GDP per capita. Other explanatory variables that have been considered are population density, share of individuals not dependent on agriculture, infant mortality, total fertility, marital status, a dummy for newer EU regions and a capital dummy. Note that all regressions include country dummies to account for country fixed effects. Independent of the point in time considered between 1850 and 1960, human capital appears to be a significant determinant of current regional economic disparities.

Therefore, our analysis suggests that historical human capital formation is important to explain current economic prosperity in the European regions. For this reason, it appears crucial not to neglect long-term evolutions that have key implications for today's economic development. For this reason, still more advanced studies on long-run human capital formation in the European regions and other parts in the world appear to be essential to better understand economic development in the past, present and the future.

10.6 Appendix

10.6.1 Data

Regions of the following countries are included in the regressions at each point in time:

1850 (plus GDP/c and patents/c):

Austria, Belgium, Germany, Denmark, Estonia, Spain, France, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Poland, Portugal, Romania, Slovakia, United Kingdom

1900:

Austria, Belgium, Bulgaria, Spain, France, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovakia, United Kingdom

1930:

Austria, Belgium, Bulgaria, Estonia, Spain, Finland, France, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Poland, Portugal, Romania, Slovakia

1960:

Bulgaria, Estonia, Spain, Greece, Hungary, Italy, Lithuania, Latvia, Poland, Portugal, Romania

2008 (except GDP/c and patents/c):

Austria, Bulgaria, Czech Republic, Germany, Denmark, Finland, Greece, Hungary, Netherlands, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia

10.6.2 Tables

Table 10.1 First patent laws in the European countries

Country	First patent law	Note
Austria	1810	
Belgium	1855	Before: application of patent laws of France and Netherlands
Bulgaria	1921	
Czechoslovakia	1919	
Denmark	1894	
Finland	1898	
France	1791	
Germany	1877	Before: earlier separate patent laws in the various states
Great Britain	1852	
Greece	1920	
Hungary	1894	First separate patent law from Austria
Ireland	1927	
Italy	1859	Before: earlier separate patent laws in the various states
Luxembourg	1880	Before: application of Dutch 1817 patent law
Netherlands	1809	
Norway	1885	
Poland	1919	
Portugal	1852	
Romania	1906	
Russia / USSR	1812	
Spain	1826	
Sweden	1819	
Switzerland	1888	
Turkey	1880	
Yugoslavia	1921	

Note: Countries refer to those existing in 1964.
Source: Based on Federico (1964).

Table 10.2 Descriptive statistics

Variable	obs.	mean	sd	min	max
Capital	198	0.06	0.24	0.00	1.00
Newer EU regions	198	0.20	0.40	0.00	1.00
ABCC 1850	198	0.96	0.05	0.74	1.00
Total fertility 1870	198	0.37	0.06	0.23	0.51
Marital status 1870	198	0.51	0.09	0.28	0.75
Literacy 1900	147	0.65	0.27	0.13	1.00
Total fertility 1900	147	0.35	0.09	0.20	0.68
Marital status 1900	147	0.54	0.11	0.31	0.77
Literacy 1930	131	0.82	0.18	0.32	1.00
Total fertility 1930	131	0.25	0.07	0.05	0.44
Marital status 1930	131	0.55	0.09	0.32	0.80
ln(Pop. density 1930)	131	4.43	0.90	0.69	8.82
Infant mortality 1930	131	0.11	0.04	0.04	0.20
Not dep. on agr. 1930	131	0.56	0.20	0.14	0.99
Literacy 1960	94	0.88	0.10	0.61	0.99
Total fertility 1960	94	0.22	0.05	0.12	0.33
Marital status 1960	94	0.62	0.08	0.48	0.82
Higher edu. attain. 2008	147	0.77	0.15	0.18	0.97
Total fertility 2008	147	1.51	0.23	1.13	2.27
ln(GDP/c 2008)	198	10.06	0.35	8.88	11.15
ln(Patents/c 2008)	197	3.80	1.53	-0.77	6.26

Table 10.3 Relationship between regional GDP per capita, patent applications per capita and higher education in 2008

	(1)	(2)
Dependent variable	ln(GDP/c 2008)	ln(Patents/c 2008)
Higher edu. attain. 2008	0.98**	5.35***
	(0.038)	(0.004)
Total fertility 2008	-0.09	0.46
	(0.544)	(0.473)
Capital	0.47***	0.66***
	(0.000)	(0.010)
Newer EU regions	-0.46***	-1.66***
	(0.000)	(0.000)
Constant	9.82***	0.10
	(0.000)	(0.956)
Country FE	Included	Included
Observations	147	140
R-squared	0.86	0.90

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. GDP, patents, higher education and fertility data are derived from Eurostat (2012) datafiles.

Table 10.4 Relationship between regional patent applications per capita in 2008 and historical variables

	(1)	(2)	(3)	(4)	(5)
Dependent variable	ln(Patents/c 2008)				
Literacy 1930	5.55***	4.56***			
	(0.001)	(0.000)			
Total fertility 1930	1.70	0.49			
	(0.542)	(0.792)			
Marital status 1930	1.01	0.82			
	(0.656)	(0.687)			
ln(Pop. density) 1930	0.23				
	(0.185)				
Infant mortality 1930	1.32				
	(0.801)				
Not dep. on agr. 1930	-1.25				
	(0.185)				
Capital	0.36	0.36	0.22	0.42	0.68***
	(0.265)	(0.131)	(0.403)	(0.103)	(0.002)
Newer EU regions	-0.73	-1.67***	-1.19**	-1.52***	-1.13***
	(0.127)	(0.000)	(0.010)	(0.000)	(0.000)
Literacy 1960			9.99***		
			(0.000)		
Total fertility 1960			-8.14***		
			(0.000)		
Marital status 1960			0.16		
			(0.945)		
Literacy 1900				3.73***	
				(0.000)	
Total fertility 1900				-2.08	
				(0.282)	
Marital status 1900				0.88	
				(0.545)	
ABCC 1850					4.49**
					(0.015)
Total fertility 1870					0.28
					(0.865)
Marital status 1870					-1.30
					(0.266)
Constant	-2.77	0.01	-3.05	1.21	-0.19
	(0.319)	(0.996)	(0.144)	(0.141)	(0.908)
Country FE	Included	Included	Included	Included	Included
Observations	126	142	87	142	197
R-squared	0.87	0.85	0.81	0.82	0.81

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Patents/c refers to patent applications to the EPO per million inhabitants.

Table 10.5 Relationship between regional GDP per capita in 2008 and historical variables

	(1)	(2)	(3)	(4)	(5)
Dependent variable			ln(GDP/c 2008)		
Literacy 1930	0.85***	0.83***			
	(0.001)	(0.000)			
Total fertility 1930	-0.11	-0.51*			
	(0.802)	(0.072)			
Marital status 1930	-0.55**	-0.39			
	(0.047)	(0.110)			
ln(Pop. density) 1930	0.01				
	(0.629)				
Infant mortality 1930	-0.66				
	(0.415)				
Not dep. on agr. 1930	0.16				
	(0.307)				
Capital	0.34***	0.34***	0.41***	0.49***	0.48***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Newer EU regions	-1.11***	-0.30***	-0.56***	-0.46***	-0.36***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Literacy 1960			1.89***		
			(0.000)		
Total fertility 1960			-1.69***		
			(0.000)		
Marital status 1960			0.02		
			(0.970)		
Literacy 1900				0.70***	
				(0.000)	
Total fertility 1900				0.21	
				(0.557)	
Marital status 1900				-0.53*	
				(0.061)	
ABCC 1850					1.29***
					(0.001)
Total fertility 1870					0.26
					(0.499)
Marital status 1870					-0.58**
					(0.019)
Constant	10.50***	9.58***	8.98***	9.58***	8.95***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country FE	Included	Included	Included	Included	Included
Observations	131	148	94	147	198
R-squared	0.91	0.90	0.88	0.83	0.78

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

10.6.3 Figures

Figure 10.1 Relationship between share of agriculture, 1930 and GPD/c, 2008

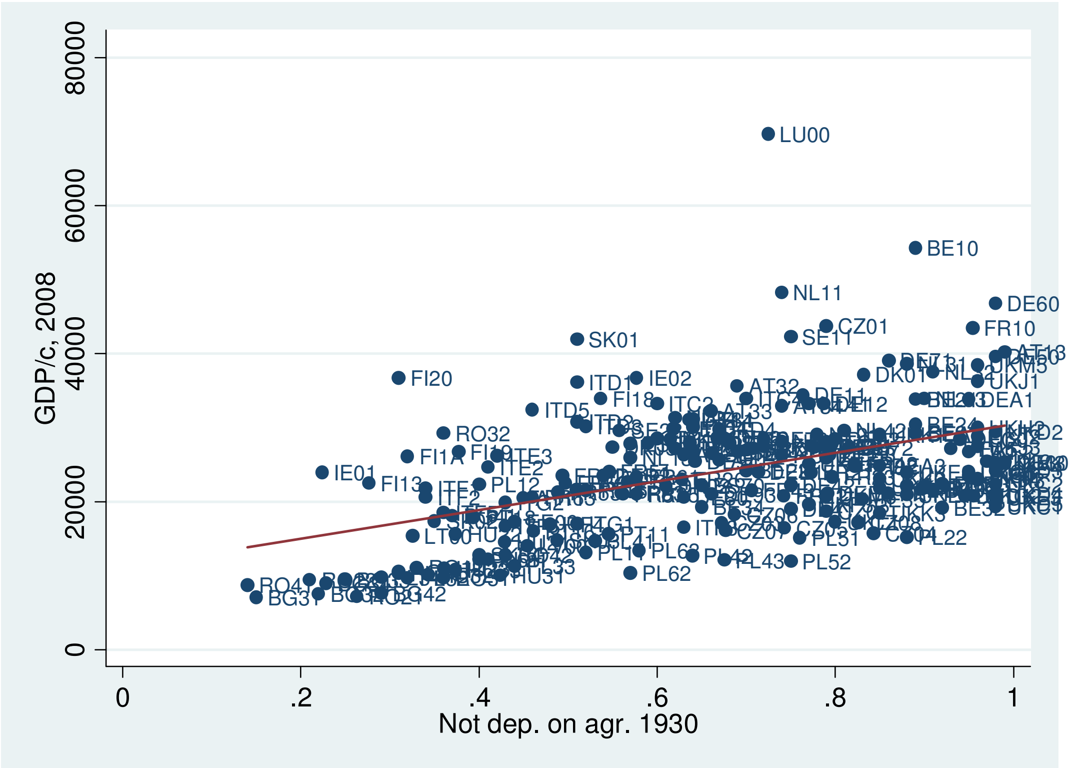
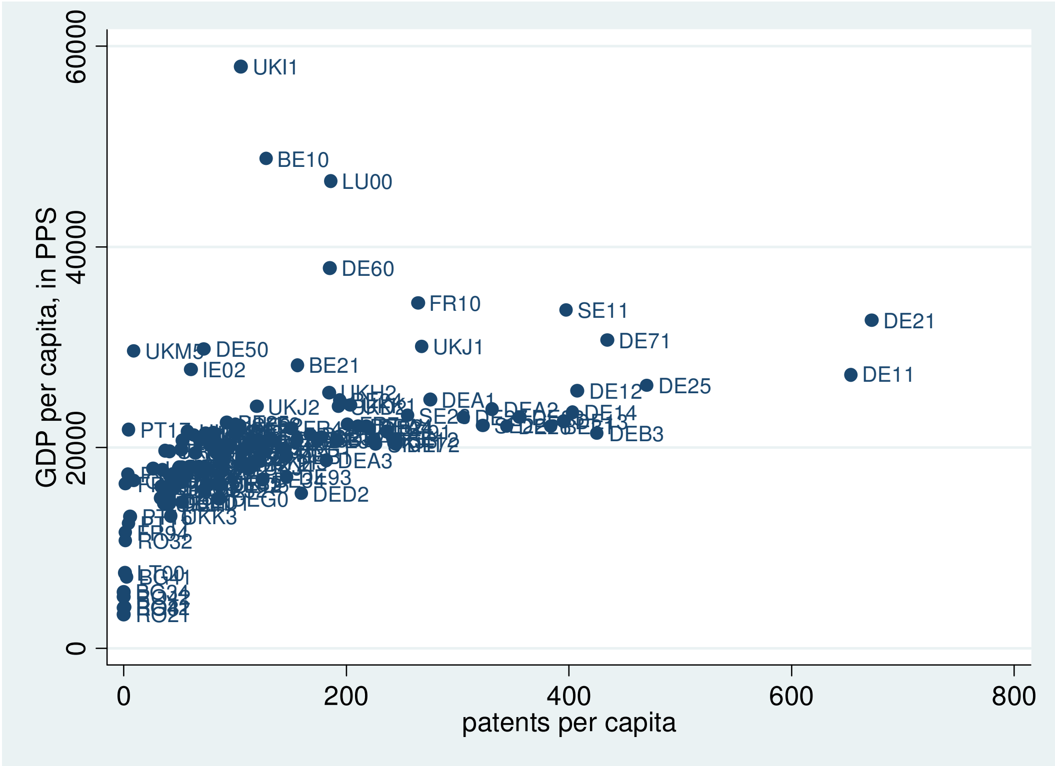
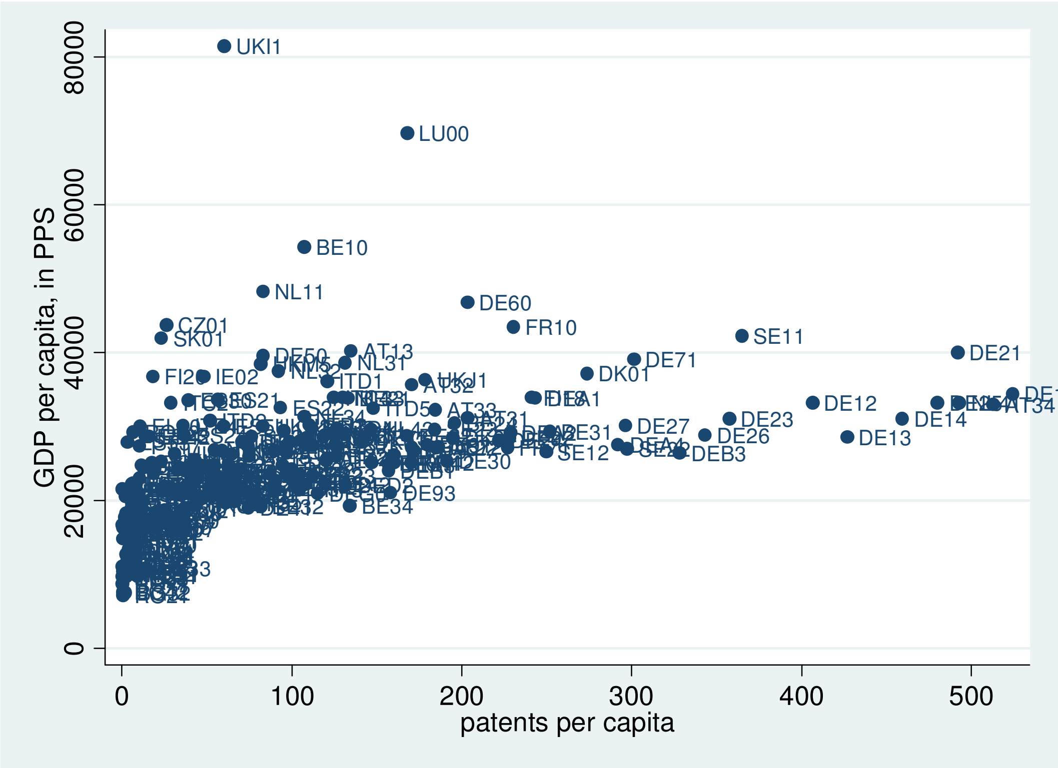


Figure 10.2 Regional per capita GDP and patent applications to the EPO in Europe, 2000



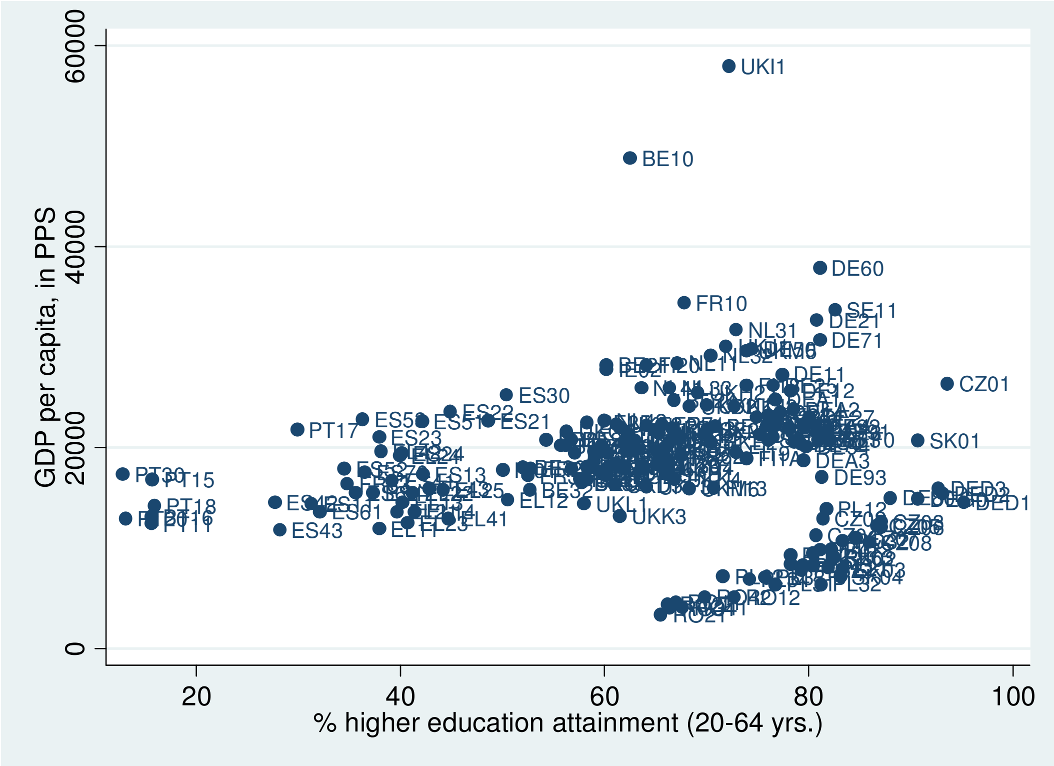
Note: Patent applications per capita are defined as patent applications to the EPO per million of inhabitants.
Source: Data provided by Eurostat (2012).

Figure 10.3 Regional per capita GDP and patent applications to the EPO in Europe, 2008



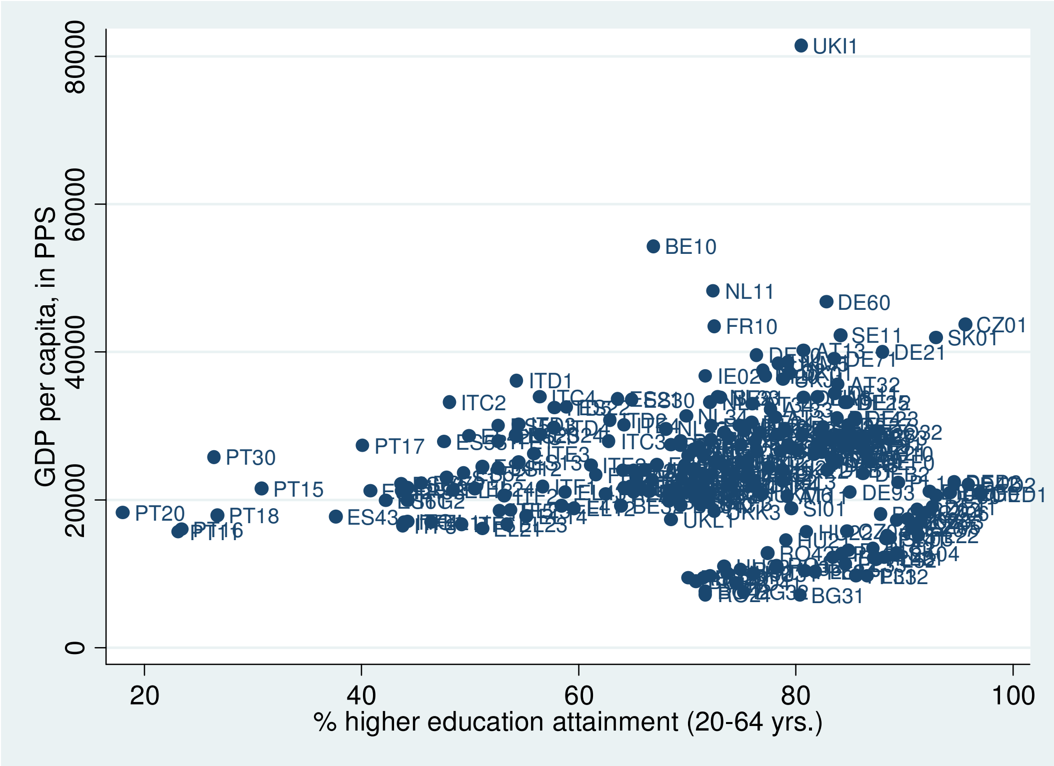
Note: Patent applications per capita are defined as patent applications to the EPO per million of inhabitants.
Source: Data provided by Eurostat (2012).

Figure 10.4 Regional per capita GDP and higher education in Europe, 2000



Note: Higher education attainment refers to the share of "persons aged 25-64 and 20-24 with upper secondary or tertiary education attainment" (Eurostat 2012).
Source: Data provided by Eurostat (2012).

Figure 10.5 Regional per capita GDP and higher education in Europe, 2008



Note: Higher education attainment refers to the share of "persons aged 25-64 and 20-24 with upper secondary or tertiary education attainment" (Eurostat 2012).
Source: Data provided by Eurostat (2012).

11. Summary, policy recommendations and directions for future research

This thesis has analysed the formation of human capital in Europe at the regional level and its implications for economic growth.

Several broad conclusions emerge from its results. First, numeracy as measured by the age heaping method is closely related to other human capital proxies, particularly to literacy. It is thus a complementary approach and broadens existing knowledge on human capital formation in Europe and in other regions of the world. In particular, its characteristics allow the use of a long-run, regional approach to human capital that is unmatched by other indicators.

Second, human capital formation as measured by numeracy, literacy and educational attainment at the regional level in Europe from the 19th to the 21st century is a story of success: human capital levels in Europe are in each case increasing throughout time. Nevertheless, important regional differences within countries characterise many European countries. These intranational disparities can be even larger than international ones, highlighting the appropriateness of the regional level as the unit of analysis.

Third, land inequality played an important role in numeracy formation at the end of the 19th century in Europe. Corresponding to the predictions of a theoretical model related to Unified Growth Theory, it appears that greater land inequality decelerated numeracy formation in the less industrialised regions of Europe. This result also implies that land inequality was also an important hurdle for economic growth.

Fourth, the geographical location of a region is a determinant of its past numeracy or literacy level. Regions that belonged to the core of Europe formed a cluster that was characterised by very high levels of human capital, whereas most of the peripheral regions in western, southern and eastern Europe could not escape the influence of remoteness. This finding was illustrated through Exploratory Spatial Data Analysis and by econometrically testing a New Economic Geography model that focuses on the impact of market access on human capital.

Fifth, historical human capital is related to current economic disparities. More specifically, historical human capital levels in the 19th and 20th centuries can significantly explain current regional differences in patent applications per capita to the European Patent Office and GDP per capita. Therefore, human capital is an important factor that has to be considered to explain regional growth processes.

These results may imply different policy recommendations. First, the effects of human capital are long lasting, and human capital is important for generating economic growth. Therefore, increasing the human capital of the population should be a key pillar of policy actions focusing on economic development. Policy makers need to focus on providing an environment that produces incentives for individuals to increase their human capital in keeping with economic conditions, particularly in periphery countries. Land reforms may be a tool for achieving greater prosperity in the least developed countries in Europe and on other continents where agriculture still plays an important role. In more advanced countries, policy actions designed, for example, to provide improved infrastructure may support human capital formation by integrating regional and national markets and offsetting the disadvantage produced by remoteness.

The important discrepancies that characterised many European countries throughout history until the present may further indicate the need to take policy decisions that explicitly address the inherent characteristics of each region. Many policy discussions continue to focus exclusively on the national level. The present economic, financial and debt crisis in Europe presents a further illustration of this. However, national ‘one size fits all’ policies may not always be appropriate for every region. The proper application of the principle of subsidiarity appears to be essential to combine the necessary harmonisation, standardisation and integration of policies at the national and European levels with sufficient margins for individual responsibility, initiative and policy actions at the regional and local levels.

Furthermore, there are several areas that may deserve additional attention in future research. On the one hand, these are related to empirics and statistics. It may be possible to trace the long-term evolution of human capital further back in time by taking advantage of existing or new sources. In this way, it may be possible to construct a detailed history of human capital that encompasses existing evidence. The age heaping method is a particularly advantageous and fruitful means of advancing this knowledge. This long-term approach is especially important because it makes it possible to validate theoretical models and hypotheses concerning the (very) long run, such as Unified Growth Theory.

Similarly, a more detailed examination of the regional level may help to understand differences at the local level. Therefore, it may be interesting to extend the methodology from the NUTS level to the local level. Many regional trends are the product of local processes, and the disaggregation of these regions permits further insights in the formation of human capital and its economic effects. Adopting a European perspective may further extend research that is too often limited by national borders. Moreover, this thesis has employed technology that uses geographical information systems. GIS are advancing in many areas of daily life and are also gradually taking hold in economics. The possibilities that GIS offer are substantial, particularly for the spatial analysis of many relevant topics in economics. Future research should further exploit the opportunities offered by GIS and spatial econometric methods. These methods also make further tools available to test models derived from theories such as New Economic Geography.

Finally, it may also be important to enrich the analysis of human capital by accounting for indicators that comprise different levels and forms of human capital. In this thesis, the focus has been on rather basic human capital. Clearly, there are also much more advanced forms of human capital. Their simultaneous consideration may improve the approximation of human capital in a region. Further indicators may be added and new proxies may be developed that enable an improved measurement of human capital.

On the other hand, research not only has to advance empirically but also theoretically. The validation of theory appears to be a necessary condition to better analyse the past and the present and improve or develop new models. More and better data can improve theory but are not a necessary condition for the elaboration of better theories. These theories, however, are important to solve the many riddles that academics and policy makers consistently face. In the present context, this means that unveiling the fundamentals of the process of economic growth and focussing on its long-run and spatial aspects to a greater extent has the potential to change the way we conceive of economic development in time and space.

In these ways, it will be possible to better understand and explain human capital formation and the impact of human capital on socio-economic outcomes in the past and the present. Therefore, incorporating the dimensions of both time and space in the analysis appears to be essential because history and geography matter.

Human capital will also be crucial for shaping the world of tomorrow. This is also emphasised by the Council of the European Union, stressing that "[e]ducation and training have a crucial role to play in meeting the many socio-economic, demographic, environmental and technological challenges facing Europe and its citizens today and in the years ahead" (Council of the European Union 2009, C 119/2). Never forget that, as one of the major initiators of human capital theory and Nobel laureate, Gary S. Becker, reminds us, we are currently living in the "Age of Human Capital" (Becker 2002, p. 3).

12. References

A'Hearn, B., Crayen, D. and J. Baten (2006). Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital, Working Paper Version 19 Nov., 2006.

A'Hearn, B., Crayen, D. and J. Baten (2009). Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital, *Journal of Economic History*, Cambridge University Press, 68 (3): 783-808.

Acemoglu, D. and D. Autor (2012). What Does Human Capital Do? A Review of Golding and Katz's The Race between Education and Technology, *Journal of Economic Literature*, 50 (2): 426-463.

Acemoglu, D., Johnson, S. and J. A. Robinson (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution, *Quarterly Journal of Economics*, 117 (4): 1231-94.

Acemoglu, D., Johnson, S. and J. A. Robinson (2005a). The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth, *American Economic Review*, 95 (3): 546-579.

Acemoglu, D., Johnson, S. and J. A. Robinson (2005b). Institutions as the Fundamental Cause of Long-Run Growth, in: Aghion, P. and S. Durlauf (eds.). Handbook of Economic Growth, Amsterdam: Elsevier, 385-472.

Acs, Z. S., Anselin, L. and A. Varga (2002). Patents and innovation counts as measures of regional production of new knowledge, *Research Policy*, 31: 1069-1085.

Adick, C. (2003). Globale Trends weltweiter Schulentwicklung: Empirische Befunde und theoretische Erklärungen, *Zeitschrift für Erziehungswissenschaft*, 6 (2): 173-187.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction, *Econometrica*, 60: 323-351.

Aghion, P. and P. Howitt (1998). *Endogenous growth theory*, Cambridge: MIT Press.

ALARM (2012). ALARM observed and scenario climate data, online, last accessed 12 June 2012, http://www.alarmproject.net/climate/climate/.

Alesina, A. and D. Rodrik (1994). Distributive Politics and Economic Growth, *Quarterly Journal of Economics*, 109: 465-490.

Allen, R. C. (2001). The Great Divergence in European Wages and Prices from the Middle Ages to the First World War, *Explorations in Economic History*, 38: 411-447.

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*, Cambridge: Cambridge University Press.

Allen, R. C. (2012). Data: Wage and Price History, online, last accessed 25 June 2012, http://www.nuffield.ox.ac.uk/General/Members/allen.aspx.

Allen, R. C. and R. W. Unger (2012). Allen – Unger Global Commodity Price Database, online, last accessed 25 June 2012, http://www.gcpdb.info/.

Andersen, B. (2001). *Technological Change and the Evolution of Corporate Innovation. The Structure of Patenting, 1880-1990*, Cheltenham/Northampton: Edward Edgar

Ang, J. B. and J. B. Madsen (2011). Can Second-Generation Endogenous Growth Models Explain the Productivity Trends and Knowledge Production in the Asian Miracle Economies?, *Review of Economics and Statistics*, 93 (4): 1360-1373.

Annuaire Statistique et Historique Belge (1857). *Quatrième année*, Brussels/Leipzig/Bonn: Aug. Schnée et Cie./A.D. Marcus.

Anselin, L. (1995). Local indicators of spatial association-LISA, *Geographical Analysis*, 27: 93-115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association, in: Fisher, M., Scholten, H.J. and D. Unwin (eds.). Spatial analytical perspectives on GIS, London: Taylor&Francis, 111-126.

Anselin, L. (1998a). Interactive techniques and exploratory spatial data analysis, in: Longley, P. A., Goodchild, M. F., Maguire, D. J. and D. W. Wind (eds.). Geographical information systems: principles, techniques, management and applications, New York: Wiley, 251-264.

Anselin, L. (1998b). Exploratory spatial data analysis in a geocomputational environment, in: Longley, P. A., Brooks, S. M., McDonnell, R. and B. Macmillan (eds.). Geocomputation, a primer, New York: Wiley, 77-94.

Anselin, L. (2001). Spatial econometrics, in: Baltagi, B. H. (ed.). Companion to econometrics. Oxford: Basil Blackwell, 310-330.

Anselin, L. and S. Bao (1997). Exploratory spatial data analysis linking SpaceStat and ArcView, in: Fisher, M. and A. Getis (eds.). Recent developments in spatial analysis, Berlin: Springer, 35-59.

Anselin, L., Varga, A. and Z. S. Acs (2000). Geographical Spillovers and University Research: A Spatial Econometric Perspective, *Growth and Change*, 31: 501-515.

Arrow, K. J. (1962). The Economic Implications of Learning by Doing, *Review of Economic Studies*, 29 (3): 155-173.

Azomahou, T., Diebolt, C. and T. Mishra (2009). Spatial persistence of demographic shocks and economic growth, *Journal of Macroeconomics*, 31: 98-127.

Badinger, H. and G. Tondl (2003). Trade, human capital and innovation: the engines of European regional growth in the 1990s, in: Fingleton, B. (ed.). European Regional Growth, Heidelberg, New York: Springer, 215-239.

Badinger, H., Müller, W. and G. Tondl (2004). Regional Convergence in the European Union, 1985-1999: A Spatial Dynamic Panel Analysis, *Regional Studies*, 38 (3): 241-253.

Bailey, T. and A. C. Gatrell (1995). *Interactive spatial data analysis*, Harlow: Longman.

Bailey, T. J. and J. B. M. Schick (2009). Historical GIS: Enabling the Collision of History and Geography, *Social Science Computer Review*, 27: 291-296

Bairoch, P., Batou, J. and P. Chèvre (1988). *La population des villes européennes : 800-1850 : banque de données et analyse sommaire des résultats*, Geneva: Droz.

Banerjee A. and Iyer L. (2005). History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India, *American Economic Review*, 95: 1190-1213.

Banks, A. S. (1971). Cross-Polity Time-Series Data, Cambridge: MIT Press.

Barro, R. J and G. S. Becker (1989). Fertility Choice in a Model of Economic Growth, *Econometrica*, 57 (2): 481-501.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries, *Quarterly Journal of Economics*, 106 (2): 407-443.

Barro, R. J. (2001). Human Capital and Growth, *American Economic Review, Papers and Proceedings*, 91 (2): 12-17.

Barro, R. J. and J.-W. Lee (1993). International Comparisons of Educational Attainment, *Journal of Monetary Economics*, 32: 363-394.

Barro, R. J. and J.-W. Lee (1996). International Measures of Schooling Years and Schooling Quality, *American Economic Review*, Papers and Proceedings, 86 (2): 218-223.

Barro, R. J. and J.-W. Lee (2001). International Data on Educational Attainment: Updates and Implications, *Oxford Economic Papers*, 53 (3): 541-563.

Barro, R. J. and X. Sala-i-Martin (1995). *Economic Growth*, New York: McGraw-Hill.

Baten, J. and J. L. van Zanden (2008). Book production and the onset of modern economic growth, *Journal of Economic Growth*, 13 (3): 217-235.

Baten, J. and K. Jaeger (2009). On the persistence of human capital and patent effects around 1900 on per capita income levels in the 1960s, *Brussels Economic Review – Cahiers économiques de Bruxelles*, 52 (3/4): 289-304.

Baten, J., Ma, D. Morgan, S. and Q. Wang (2010). Evolution of living standards and human capital in China in the 18–20th centuries: Evidences from realwages, age-heaping, and anthropometrics, *Explorations in Economic History*, 47 (3): 347-359.

Baumol, W. J. and A. S. Blinder (1991). *Economics. Principles and Policy*, 5th edition, New York: Harcourt Brace Jovanovich Inc.

Baumont, C., Combes, P.-P., Derycke, P.-H. and H. Jayet (2000). *Economie géographique*, Paris: Economica.

Baumont, C., Ertur, C. and J. Le Gallo (2003). Spatial convergence clubs and the European regional growth process, 1980–1995, in: Fingleton, B. (ed). European regional growth. Berlin: Springer, 131-158.

Beach, M. J. (2009). A Critique of Human Capital Formation in the U.S. and the Economic Returns to Sub-Baccalaureate Credentials, *Educational Studies: A Journal of the American Educational Studies*, 45(1): 24-38.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Chicago: University of Chicago Press.

Becker, G. S. (1981). *A Treatise on the Family*, Cambridge: Harvard University Press.

Becker, G. S. (2002). The Age of Human Capital, in: Lazear, E. P. (ed.). Education In The Twenty-First Century, Stanford: Hoover Institution Press, 3-8.

Becker, G. S., Murphy, K. M. and R. Tamura (1990). Human Capital, Fertility, and Economic Growth, *Journal of Political Economy*, 98 (5-2): S12-S37.

Becker, S. O. and L. Woessmann (2009). Was Weber wrong? A Human Capital History of Protestant Economic History, *Quarterly Journal of Economics*, 124 (2): 531-596.

Becker, S. O., Cinnirella, F. and L. Woessmann (2010). The trade-off between fertility and education, evidence from before the demographic transition, *Journal of Economic Growth*, 15: 177-204.

Becker, S. O., Cinnirella, F. and L. Woessmann (2012). The effect of investment in children's education on fertility in 1816 Prussia, *Cliometrica*, 6: 29-44.

Becker, S. O., Hornung, E. and L. Woessmann (2011). Education and Catch-up in the Industrial Revolution, *American Economic Journal: Macroeconomics*, 3 (3): 92-126.

Benavot, A. and P. Riddle (1988). The Expansion of Primary Education, 1870-1940: Trends and Issues, *Sociology of Education*, 61 (3): 191-210.

Benhabib, J. and M. M. Spiegel (1994). The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country and Regional U.S. Data, *Journal of Monetary Economics*, 34 (2): 143-173.

Berend, I. (1985). Agriculture, in: Kaser M. and F. Radice (eds.). The Economic History of Eastern Europe 1919-1975, Vol. I, Oxford: Claredon Press, 152-162.

Betran, C. (2011). Regional specialisation and industry location in the long run: Spain in the US mirror (1856-2002), *Cliometrica*, 5: 259-290.

BICSE (Board on International Comparative Studies in Education), National Research Council (NCR) (1995). *International Comparative Studies in Education: Descriptions of Selected Large-Scale Assessments and Case Studies*, Washington, D.C.: The National Academies Press, Winter 1994-95.

Bleakley, H. and F. Lange (2009). Chronic disease burden and the interaction of education, fertility, and growth, *Review of Economics and Statistics,* 91 (1): 52-65.

Boucekkine, R., de la Croix, D. and D. Peeters (2007). Early literacy achievements, population density, and the transition to modern growth, *Journal of the European Economic Association*, 5 (1): 183-226.

Bowles, S. (1978). Capitalist Development and Educational Structure, *World Development*, 6: 783-796.

Braudel, F. (1979). *Civilisation materielle, économic el capitalisme, XVe-XVIIIe siècle. Le temps du monde*, Paris: Armand Colin.

Braudel, F. (1992). *Civilization and Capitalism, 15th-18th century: The perspective of the world*, Berkeley: University of California Press.

Breinlich, H. (2006). The spatial income structure in the European Union: what role for economic geography?, *Journal of Economic Geography*, 6 (5): 593-617.

Bretagnolle, A., Pumain, D. and C. Rozenblat (1997). Space-time Contraction and the Dynamics of Urban Systems, *Cybergeo : European Journal of Geography*, [En ligne], Dossiers, 10ème Colloque Européen de Géographie Théorique et Quantitative, Rostock, Allemagne, 6-11 septembre 1997, article 61, mis en ligne le 10 juillet 1998, consulté le 06 juillet 2012. URL : http://cybergeo.revues.org/373 ; DOI : 10.4000/cybergeo.373.

Brint, S. G. (2006). *Schools and societies*, Stanford: Stanford University Press.

Brunet, R. (2002). Lignes de force de l'espace européen, *Mappe Monde*, 66,2002.2, http://www.mgm.fr/PUB/Mappemonde/M202/Brunet.pdf.

Bulcock, J. W. and M. J. Beebe (1981). Some Common Causes of Literacy and Numeracy, *Canadian Journal of Education*, 6 (3): 19-44.

Canals, V., Diebolt, C. and M. Jaoul (2003). Convergence et disparités régionales du poids de l'enseignement supérieur en France: 1964-2000, *Revue d'Economie Régionale et Urbaine*, 4: 649-669.

Cantillon, R. (1755/1952). *Essay sur la Nature du Commerce en Général*, Reprint, Paris: Institut National d'Etudes Démographiques.

Cantwell, J. (1989). *Technological Innovation and Multinational Corporations*, Oxford/Cambridge: Basil Blackwell.

Capello, R. and C. Lenzi (2012). Knowledge, innovation and economic growth: spatial heterogeneity in Europe, *Growth and Change*, 43 (4): 697-698.

Cappelen, A., Fagerberg, J. and B. Verspagen (1999). Lack of regional convergence, in: Fagerberg, J., Guerrieri, P. and B. Verspagen (eds.). The Economic Challenge for Europe: Adapting to Innovation Based Growth, Cheltenham: Edward Elgar, 130-148.

Central Advisory Council for Education (England) (1959). *A report of the Central Advisory Council for Education (England), Crowther Report*, London: HMSO.

Chiswick, B. R., Lee, Y. L. and P. W. Miller (2003). Schooling, Literacy, Numeracy and Labour Market Success, *Economic Record*, 79: 165-181.

Christaller, W. (1933). *Die Zentralen Orte in Süddeutschland*, Jena: Gustav Fischer Verlag.

Ciccone, A. and E. Papaioannou (2009). Human Capital, the Structure of Production, and Growth, *Review of Economics and Statistics*, 91 (1): 66-82.

Cinnirella, F. and E. Hornung (2011). Landownership Concentration and the Expansion of Education, EHES Working Papers in Economic History, No. 10, October 2011.

Cipolla, C. M. (1969). *Literacy and development in the West*, Baltimore: Penguin Books.

Cliff, A. and J. Ord (1981). *Spatial Processes: Models and Applications*, London: Pion.

Coale, A. J. (1973). *The demographic transition*, in: Proceedings of the International Population Conference, Vol. 1, Liège: International Union for the Scientific Study of Population, 53-72.

Coale, A. J. and R. Treadway (1986). A Summary of the Changing Distribution of Overall Fertility, Marital Fertility, and the Proportion Married in the Provinces of Europe, in: Coale, A. J. and S. C. Watkins (ed.). The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project, Princeton: Princeton University Press, 31-181.

Coale, A. J. and S. C. Watkins (ed.) (1986). *The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project*, Princeton: Princeton University Press.

Cohen, D. and M. Soto (2007). Growth and Human Capital: Good Data, Good Results, *Journal of Economic Growth*, 12: 51-76.

Cohen, I. B. (2005). *The Triumph of Numbers*, New York: Norton & Company.

Cohen, P. C. (1982). *A calculating People: The Spread of Numeracy in Early America*, Chicago: University of Chicago Press.

Cohen, W. M. and D. A. Levinthal (1989). Innovation and Learning: The Two Faces of R&D, *Economic Journal*, 99: 569-596.

Combes, P.-P., Duranton, D. and L. Gobillon (2008). Le rôle des marchés locaux du travail dans la concentration spatiale des activités économiques, *Revue de l'OFCE 2008/1*, 104: 141-177.

Combes, P.-P., Lafourcade, M., Thisse, J.-F. and J.-C. Toutain (2011). The rise and fall of spatial inequalities in France: A long-run perspective, *Explorations in Economic History*, 48: 243-271.

Combes, P.-P., Mayer, T. and J.-F. Thisse (2006). *Economie géographique. L'intégration des régions et des nations*, Paris: Economica.

Comin, D., Easterly, W. and E. Gong (2010). Was the Wealth of Nations Determined in 1000 BC?, *American Economic Journal: Macroeconomics*, 2: 65-97.

Costa, D., Demeulemeester, J.-L. and C. Diebolt (2007). What is 'Cliometrica'?, *Cliometrica*, 1: 1-6.

Council of the European Union (2003). Council conclusions of 5 May 2003 on reference levels of European average performance in education and training (Benchmarks) (2003/C 134/02), C 134/4, online, last accessed on 22 June 2012, http://www.cedefop.europa.eu/EN/Files/Benchmarks.pdf.

Council of the European Union (2009). Council conclusions of 12 May 2009 on a strategic framework for European cooperation in education and training ('ET 2020'), C119/2, online, last accessed on 22 June 2012, http://www.cedefop.europa.eu/EN/Files/ET_2020.pdf.

Council of the European Union (2011). Council Recommendation of 28 June 2011 on policies to reduce early school leaving, 2011/C 191/01, online, last accessed on 22 June 2012, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:191:0001:0006:EN:PDF.

Crampton, R. J. (2007). *Bulgaria*, Oxford: Oxford University Press.

Crawford, R. (1991). *In the Era of Human Capital*, New York: Harpercollins.

Crayen, D. and J. Baten (2010a). Global Trends in Numeracy 1820-1949 and its Implications for Long-Run Growth, *Explorations in Economic History*, 47: 82-99.

Crayen, D. and J. Baten (2010b). New evidence and new methods to measure human capital inequality before and during the industrial revolution: France and the US in the seventeenth to nineteenth centuries, *Economic History Review*, 63 (2): 452-478.

Cressy, D. (1980). *Literacy & the Social Order*, New York: Cambridge University Press.

Crosby, M. (2000). Patents, Innovation and Growth, *Economic Record*, 76 (234): 255-262.

Cuaresma, J. C., Doppelhofer, G. and M. Feldkircher (2012). The Determinants of Economic Growth in European Regions, *Regional Studies*, DOI:10.1080/ 00343404.2012.678824.

Dakhli, M. and D. de Clercq (2004). Human capital, social capital, and innovation: a multi-country study, *Entrepreneurship & regional development*, 16: 107-128.

Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989-1999: An exploratory spatial data analysis, *Annals of Regional Science*, 39: 121-148.

Daskalova, K. (1996). *Literacy and Reading in 19th Century Bulgaria*, Washington: Donald W. Treadgold Papers.

David, P. A. (1990). The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox, *American Economic Review Proceedings*, 80: 355-361.

De La Fuente, A. and R. Doménech (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?, *Journal of the European Economics Association*, 4: 1-36.

De Long, J. B. and A. Shleifer (1993). Princes and Merchants: European City Growth before the Industrial Revolution, *Journal of Law and Economics*, 36 (2): 671-702.

De Moor, T. and J. L. van Zanden (2008). Uit fouten kun je leren. Een kritische benadering van de mogelijkheden van 'leeftijdstapelen' voor sociaal-economisch historisch onderzoek naar ecijferdheid in het pre-industriële Vlaanderen en Nederland, *Tijdschrift voor Economische en Sociale Geschiedenis*, 5 (4): 55-86.

De Moor, T. and J. L. van Zanden (2010). Girl power: the European marriage pattern and labour markets in the North Sea in the late medieval and early modern period, *Economic History Review*, 63 (1): 1-33.

De Vries, J. (1984). *European Urbanization, 1500-1800*, London: Taylor and Francis.

Deichmann, U. and L. Eklundh (1991). Global digital data sets for land degradation studies: a GIS approach, GRID Case Study Series No. 4; UNEP/GEMS and GRID; Nairobi, Kenya; 103 pages (see pp. 24-27).

Deininger K. and Squire D. (1998). New ways of looking at old issues: inequality and growth, *Journal of Development Economics*, 57: 259-287.

Del Barrio-Castro, T. and J. García-Quevedo (2005). Effects of University Research on the Geography of Innovation, *Regional Studies*, 39 (9): 1217-1229.

Demeulemeester, J.-L. and C. Diebolt (2011). Education and Growth: What Links for Which Policy?, *Historical Social Research*, 36 (4): 323-346.

Denison, E. F. (1967). *Why growth rates differ: postwar experience in nine Western countries*, Washington, D.C.: Brookings Institution.

Diamond, J. (1997). *Guns, Germs and Steel: The Fates of Human Societies*, New York/London: W. W. Norton.

Diebolt, C. and K. Pellier (2009a). La convergence des activités innovantes en Europe : les enseignements de l'économétrie spatiale appliquée à l'histoire du temps présent, *Economies et sociétés*, 43 (5): 805-831.

Diebolt, C. and K. Pellier (2009b). Vers une nouvelle histoire économique des brevets, *Brussels Economic Review – Cahiers economies de Bruxelles*, 52 (3/4): 204-214.

Diebolt, C. and K. Pellier (2009c). Analyse spectral de l'évolution de longue période des brevets en France, en Allemagne, en Grande-Bretagne, aux Etats-Unis et au Japon (17ème – 20ème siècles), *Brussels Economic Review – Cahiers economies de Bruxelles*, 52 (3/4): 341-354.

Diebolt, C. and K. Pellier (2011). Measuring the “Ideas”: Evidence from a New International Patent Database, Working Papers 11-04, Association Française de Cliométrie (AFC).

Diebolt, C. and K. Pellier (2012). 400 ans de protection par les brevets, *Revue économique*, 63 (3): 611-621.

Diebolt, C., Jaoul, M. and G. San Martino (2005). Le mythe de Ferry : une analyse cliométrique, *Revue d’économie politique*, 115 (4): 471-497.

Dobado R. and H. García (2010). Neither so low nor so short! Wages and heights in eighteenth and early nineteenth centuries colonial Hispanic America, Universidad Complutense Madrid Working Paper, March 2010.

Dovring, F. (1965). *Land and labor in Europe in the twentieth century*, The Hague: Martinus Nijhoff.

Easterly, W. (2007). Inequality Does Cause Underdevelopment: Insights from a New Instrument, *Journal of Development Economics*, 84: 755-776.

Eastwood, R. and M. Lipton (2011). Demographic transition in sub-Saharan Africa: How big will the economic dividend be?, *Population Studies: A Journal of Demography*, 65 (1): 9-35.

Eckey, H. F. and R. Kosfeld (2004). New Economic Geography, Volkswirtschaftliche Diskussionsbeiträge, Fachbereich Wirtschaftswissenschaften, Universität Kassel, Nr. 65/04.

Eddie, S. M. (1967). The Changing Pattern of Landownership in Hungary, 1867-1914, *The Economic History Review*, 20 (2): 293-310.

Elazar, D. S. (1996). Agrarian Relations and Class Hegemony: A Comparative Analysis of Landlord, Social and Political Power – Italy 1861-1920, *The British Journal of Sociology*, 47 (2): 232-254.

Emigh, R. J. (2002). Numeracy or Enumeration? The uses of numbers by states and societies, *Social Science History*, 26 (4): 653-698.

Engelen, T. and P. Puschmann (2011). How unique is the Western European marriage pattern? A comparison of nuptiality in historical Europe and the contemporary Arab world, *The History of the Family*, 16 (4): 387-400.

Engerman, S. and K. L. Sokoloff (2000). Institutions, Factor Endowments, and Paths of Development in the New World, *Journal of Economic Perspectives*, 14: 217-232.

Erickson L. and D. Vollrath (2004). Dimensions of Land Inequality and Economic Development, IMF Working Paper WP/04/158, Washington D.C: IMF.

European Commission (2008). Working for the regions, European Commission Directorate-General for Regional Policy, Luxembourg: Office for Official Publications of the European Communities, online, last accessed 12 June 2012, http://ec.europa.eu/regional_policy/sources/docgener/presenta/working2008/work_en.pdf.

European Commission (2010a). Europe 2020. Flagship Initiative Innovation Union SEC(2010) 1161, online, last accessed 2 October 2012, http://ec.europa.eu/research/innovation-union/pdf/innovation-union-communication_en.pdf.

European Commission (2010b). Europe 2020. A European strategy for smart, sustainable and inclusive growth, online, last accessed 2 October 2012, http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf.

European Commission (2011). Cohesion Policy 2014-2020. Investing in growth and jobs, Luxembourg: Publications Office of the European Union, online, last accessed on 22 June 2012, http://ec.europa.eu/regional_policy/sources/docoffic/official/regulation/pdf/2014/proposals/regulation2014_leaflet_en.pdf.

European Commission (2012a). COM(2012) 669 final, Rethinking Education: Investing in skills for better socio-economic outcomes, online, last accessed 10 January 2013, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0669:FIN:EN:PDF.

European Commission (2012b). What is the INSPIRE Directive?, online, last accessed on 22 June 2012, http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48.

Eurostat (2009). Eurostat regional yearbook 2009, Luxembourg: Publications Office of the European Union, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-HA-09-001-01/EN/KS-HA-09-001-01-EN.PDF.

Eurostat (2010). European Union, candidate countries and EFTA countries, accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/cache/GISCO/yearbook2010/RYB-Full-NUTS2-2010-EN.pdf.

Eurostat (2011a). Introduction, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction.

Eurostat (2011b). National structures (EU), online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/correspondence_tables/national_structures_eu.

Eurostat (2011c). National structures (non-EU), online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/correspondence_tables/national_structures_non_eu.

Eurostat (2011d). Principles and characteristics, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/principles_characteristics.

Eurostat (2011e). Regions in the European Union, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-020/EN/KS-RA-07-020-EN.PDF.

Eurostat (2011f). Gross domestic product (GDP) per inhabitant, in purchasing power standards (PPS), by NUTS 2 regions, 2008, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/statistics_explained/images/8/8a/Gross_domestic_product_%28GDP%29_per_inhabitant%2C_in_purchasing_power_standard_%28PPS%29%2C_by_NUTS_2.PNG.

Eurostat (2011g). Gross domestic product (GDP) per inhabitant, in purchasing power standards (PPS), highest and lowest NUTS 2 region within each country, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/statistics_explained/images/8/80/Gross_domestic_product_%28GDP%29_per_inhabitant%2C_in_purchasing_power_standard_%28PPS%29%2C_highest_and_lowest_NUTS_2_region_within_each_country%2C_2008.PNG.

Eurostat (2011h). Change of gross domestic product (GDP) per inhabitant, in purchasing power standards (PPS), by NUTS 2 regions, 2008 as compared with 2000, online, last accessed on 22 June 2012, http://epp.eurostat.ec.europa.eu/statistics_explained/images/4/41/Change_of_gross_domestic_product_%28GDP%29_per_inhabitant%2C_in_purchasing_power_standard_%28PPS%29%2C_by_NUTS_2_regions%2C_2008_as_compared_with_2000.PNG.

Eurostat (2011i). Economically active population by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000), data online, last accessed 22 June 2012, lfst_r_lfp2acedu; http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Eurostat (2012). Database. Regional Statistics, online, last accessed on 2 October 2012, extractions from data sources "demo_r_frate2", "edat_lfse_13", "nama_r_e2gdp" and "pat_ep_rtot"; http://epp.eurostat.ec.europa.eu, © European Union, 1995-2013.

Fagerberg, J., Verspagen, B. and M. Caniels (1997). Technology, growth and unemployment across European regions, *Regional Studies*, 31: 457-466.

FAZ (2005). Das "Unwort des Jahres": „Humankapital", Frankfurter Allgemeine Zeitung, 18.01.2005, online, last accessed on 22 June 2012, http://www.faz.net/aktuell/feuilleton/sprache-das-unwort-des-jahres-humankapital-1210498.html.

Federico, G. (2009). Agriculture and economic growth in Italy, 1870-1939, in: Lains, P. and V. Pinilla (eds.). Agriculture and Economic Development in Europe since 1870, London and New York: Routledge, 234-254.

Federico, P. J. (1964). Historical Patent Statistics 1791 – 1961, *Journal of the Patent Office Society*, 46 (2): 89-171.

Felice, E. (2012). Regional convergence in Italy, 1891-2011: testing human and social capital, *Cliometrica*, 6 (3): 267-306.

Fleury, M. and P. Valmary (1957). *Les Progrès de l'instruction élémentaire de Louis XIV à Napoléon III d'après l'enquête de Louis Maggiolo (1877-1879)*, Paris: INED.

Flora, P. (1975). *Indikatoren der Modernisierung – Ein historisches Datenhandbuch*, Opladen: Westdeutscher Verlag.

Flora, P. (1983). *State, Economy, and Society in Western Europe 1815-1975, Vol. I*, Frankfurt: Campus Verlag.

Földvari, P. and B. van Leeuwen (2010). Capital accumulation and growth in Central Europe, 1920–2006, University of Debrecen, Hungary.

Folloni, G. (2010). Human capital measurement: a survey, *Journal of Economic Surveys*, 24 (2): 248-279.

Foreman-Peck, J. (2011). The Western European marriage pattern and economic development, *Explorations in Economic History*, 48: 292-309.

Foreman-Peck, J. and P. Lains (2000). European economic development. The core and the southern periphery, 1870-1910, in: Pamuk, S. and J. G. Williamson (eds.). The Mediterranean response to globalization before 1950, London: Routledge, 76-106.

Fourie, J. and D. von Fintel (2010). The dynamics of inequality in a newly settled, pre-industrial society: the case of the Cape Colony, *Cliometrica*, 4: 229-267.

Friesen, J., Baten, J. and V. Prayon (2012). Women count. Gender (In-)Equalities in the Human Capital Development in Asia, 1900-60, University of Tübingen Working Papers in Economics and Finance, No. 29.

Fronzek, S., Carter, T. R. and K. Jylhä (2012). Representing two centuries of past and future climate for assessing risks to biodiversity in Europe, *Global Ecology and Biogeography*, 21 (1): 19-35, doi:10.1111/j.1466-8238.2011.00695.x.

Fujita, M. and J.-F. Thisse (1997). Économie géographique, Problèmes anciens et nouvelles perspectives, *Annales d'Économie et de Statistique*, 45: 37-87.

Fujita, M. and P. Krugman (2004). The new economic geography. Past, present, and the future, *Papers in Regional Science*, 83: 139-164.

Fujita, M., Krugman, P. and A. J. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*, Cambridge: MIT Press.

Furet, F. and J. Ozouf (1977). *Lire et écire. L'alphabétisation des Francais de Calvin à Jules Ferry*, Paris: Les editions de minuit.

Gal, I. (ed.) (2000). *Adult Numeracy Development: Theory, research, practice*, New Jersey: Hampton Press.

Galor, O. (2005a). From Stagnation to Growth: Unified Growth Theory, in: Aghion, P. and S. N. Durlauf (eds.). Handbook of Economic Growth, Volume 1A, Amsterdam: North Holland, 171-293.

Galor, O. (2005b). The demographic transition and the emergence of sustained economic growth, *Journal of the European Economic Association*, 3 (2-3): 494-504.

Galor, O. (2012). The demographic transition: causes and consequences, *Cliometrica*, 6: 1-28.

Galor, O. and D. N. Weil (1996). The Gender Gap, Fertility and Growth, *American Economic Review*, 86 (3): 374-387.

Galor, O. and D. N. Weil (1999). From Malthusian Stagnation to Modern Growth, *American Economic Review*, 89 (2): 150-154.

Galor, O. and D. N. Weil (2000). Population, Technology and Growth: From the Malthusian Regime to the Demographic Transition, *American Economic Review*, 90 (4): 806-828.

Galor, O. and O. Moav (2002). Natural Selection and the Origin of Economic Growth, *Quarterly Journal of Economics*, 117: 1133-1192.

Galor, O. and O. Moav (2006). Das Human-Kapital: A Theory of the Demise of the Class Structure, *Review of Economic Studies*, 73 (1): 85-117.

Galor, O., Moav, O. and D. Vollrath (2009). Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence, *Review of Economic Studies*, 76: 143-179.

Gans, J. S. and G. B. Shepherd (1994). How are the mighty fallen: Rejected classic articles by leading economists, *Journal of Economic Perspectives*, 8: 165-179.

Garretsen, H. and R. Martin (2010). Rethinking (New) Economic Geography Models: Taking Geography and History More Seriously, *Spatial Economic Analysis*, 5 (2): 127-160.

Giedraitiene, E., Kiliuvene, D. and S. Brauckmann (2007). Lithuania, in: Hörner, W., Döbert, H., von Kopp, B. and W. Mitter (eds.). The Education Systems of Europe, Dordrecht: Springer, 451-469.

Glaeser, E. L., La Porta, R., López-de-Silanes, F. and A. Shleifer (2004). Do Institutions Cause Growth?, *Journal of Economic Growth*, 9: 271-303.

Goldin, C. D. and L. F. Katz (2008). The Race between Education and Technology, Cambridge: Belknap Press.

Good, D. F. (1994). The Economic Lag of Central and Eastern Europe: Income Estimates for the Habsburg Successor States, 1870-1910, *Journal of Economic History*, 54 (4): 869-891.

Good, D. F. and T. Ma (1999). The economic growth of Central and Eastern Europe in comparative perspective, 1870-1989, *European Review of Economic History*, 2: 103-137.

Goodwin, B. K. and T. J. Grennes. (1998). Tsarist Russia and the Word Wheat Market, *Explorations in Economic History*, 35, 405-430.

Goyer, D. S. and G. E. Draaijer (1992). *The Handbook of National Population Censuses. Europe*, Westport: Greenwood Press.

Graff, H. J. (1987). *The Legacies of Literacy*, Bloomington: Indiana University Press.

Graff, H. J. (1991). *The Literacy myth. Cultural Integration and Social Structure in the Nineteenth Century*, New Brunswick: Transaction Publishers.

Graff, H. J. (2009). Themes and Questions, in: Graff, H. J., Mackinnon, A., Sandin B. and I. Winchester (eds.). Understanding literacy in its historical contexts: Social cultural-history and the Legacy of Egil Johansson, Lund: Nordic Academic Press, 243-264.

Graff, H. J., MacKinnon, A, Sandin, B. and I. Winchester (eds.) (2009). *Understanding literacy in its historical contexts: Social cultural-history and the Legacy of Egil Johansson*, Lund: Nordic Academic Press.

Green, A. (1990). *Education and State Formation: The Rise of Education Systems in England, France and the USA*, Basingstoke: Palgrave.

Griliches, Z. (1990). Patent Statistics as Economic Indicators, *Journal of Economic Literature*, 28: 1661-1707.

Grossman, G. M. and E. Helpman (1991). *Innovation and Growth in the Global Economy*, Cambridge: MIT Press.

Guinnane, T. W. (2008). *The historical fertility transition and theories of long-run growth: A guide for economists*, Mimeo: Yale University.

Guttormsson, L. (1990). The Development of Popular Religious Literacy in the Seventeenth and Eighteenth Centuries, *Scandinavian Journal of History*, 15: 7-35.

Haber, S. (2012). Rainfall and Democracy: Climate, Technology, and the Evolution of Economic and Political Institutions, Stanford University and NBER Working Paper, Draft of August 24, 2012.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*, Cambridge: Cambridge University Press.

Hajnal, J. (1965). European Marriage Pattern in Historical Perspective, in: Glass, D. V. and D. E. C. Eversley (eds.). Population in History, Chicago: Aldine Publishing Company, 101-143.

Hajnal, J. (1982). Two Kinds of Preindustrial Household Formation System, *Population and Development Review*, 8 (3): 449-94.

Hanushek, E. A. and L. Woessmann (2008). The Role of Cognitive Skills in Economic Development, *Journal of Economic Literature*, 46 (3): 607-668.

Harper, D. A. (2008). A bioeconomic study of numeracy and economic calculation, *Journal of Bioeconomics*, 10: 101-126.

Hatton, T. J. and J. G. Williamson (1994). What Drove the Mass Migrations from Europe in the Late Nineteenth Century?, *Population and Development Review*, 20 (3): 533-559.

Hauser, M. and E. S. Spelke (2004). Evolutionary and developmental foundations of human knowledge: A case study of mathematics, in: Gazzaniga, M. S. (ed.). The cognitive neurosciences III, Cambridge: MIT Press, 853-864.

Head, K. and T. Mayer (2004). The Empirics of Agglomeration and Trade, in: Henderson, J. V. and J.-F. Thisse (eds.). Handbook of regional and urban economics, Vol. 4, Oxford: Elsevier, 2609-2669.

Head, K. and T. Mayer (2011). Gravity, market potential and economic development, *Journal of Economic Geography*, 11: 281-294.

Heidenreich, M. (1998). The changing system of European cities and regions, *European Planning Studies*, 6 (3): 315-332.

Herlihy, P. (1986). *Odessa: A History, 1794-1914*, Cambridge: Harvard University Press.

Herwartz, H. and A. Niebuhr (2011). Growth, unemployment and labour market institutions: evidence from a cross-section of EU regions, *Applied Economics*, 43 (30): 4663-4676.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and A. Jarvis (2005). Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology*, 25: 1965-1978.

Hippe, R. (2012a). Regional human capital formation in Europe in the long run, 1850-2010, University of Strasbourg Working Paper.

Hippe, R. (2012b). How to measure human capital? The relationship between numeracy and literacy, *Economies et Sociétés*, 45 (8): 1527-1554.

Hippe, R. (2012c). Spatial clustering of human capital in the European regions, University of Strasbourg Working Paper.

Hippe, R. and C. I. Barmeyer (2009). Banane bleue et Banane rouge : L'exemple de l'Alsace dans les relations économiques, *Documents. Revue du dialogue franco-allemand*, 65 (2): 14-18.

Hippe, R. and J. Baten (2012a). Regional Inequality in Human Capital Formation in Europe, 1790 – 1880, *Scandinavian Economic History Review*, 60 (3): 254-289. See also former working paper version: AFC Working Papers 7/11.

Hippe, R. and J. Baten (2012b). 'Keep them ignorant.' Did inequality in land distribution delay regional numeracy formation?, University of Tuebingen Working Paper.

Hodder, J. (1671). *Hodder's Decimal Arithmetick*, sig. A4.

Hofflander, A. E. (1966). The human life value: an historical perspective, *Journal of Risk and Insurance*, 33 (3): 381-391.

Houston, R. A. (1985). *Scottish Literacy and the Scottish Identity: Illiteracy and Society in Scotland and Northern England, 1600-1800*, Cambridge: Cambridge University Press.

Houston, R. A. (1987). The literacy campaign in Scotland, 1560-1803, in: Arnove, R. F. and H. J. Graff (eds.). National literacy campaigns historical and comparative perspectives, New York: Plenum Press, 49-64.

Houston, R. A. (1999). Literacy in Europe, in: Wagner, D. A., Venezky, R. L. and B. V. Street (eds). Literacy: An International Handbook, Boulder: Westview Press, 385-390.

Houston, R. A. (2001). Literacy, in: Stearns, P. N. (ed.). Encyclopedia of European Social History, Vol. 5, Detroit: Charles Scribner's Sons, 391-406.

Houston, R. A. (2002). *Literacy in Early Modern Europe*, Harlow: Pearson Education Limited.

Huber, E. and F. Safford (eds.) (1995). *Agrarian Structure & Political Power. Landlord & Peasant in the Making of Latin America*, Pittsburgh: University of Pittsburgh Press.

IEA (2011). Brief history of IEA: 50 Years of Educational Research, online, last accessed on 25 June 2012, http://www.iea.nl/brief_history.html.

ILO (1982). Resolution concerning statistics of the economically active population, employment, unemployment and underemployment, adopted by the Thirteenth International Conference of Labour Statisticians, online, last accessed 11 November 2012, http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087481.pdf.

Inforegio (2008). EU Cohesion Policy 1988-2008: Investing in Europe's future, No. 26, June 2008, online, last accessed on 22 June 2012, http://ec.europa.eu/regional_policy/sources/docgener/panorama/pdf/mag26/mag26_en.pdf.

Ivanov, M. and A. Tooze (2007). Convergence or Decline on Europe's Southeastern Periphery? Agriculture, Population, and GNP in Bulgaria, 1892-1945, *Journal of Economic History*, 67 (3): 672-704.

Jacks, D. (2005). Intra- and International Commodity Market Integration in the Atlantic Economy, 1800-1913, *Explorations in Economic History*, 42 (3): 381-413.

Jacks, D. (2006). What Drove Nineteenth Century Commodity Market Integration?, *Explorations in Economic History*, 43 (3): 383-412.

Jaffe, A. B., Trajtenberg, M. and R. Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics*, 63: 577-598.

Janelle, D. G. (1969). Spatial reorganisation: a model and concept, *Annals of the Association of American Geographers*, 343-368.

Jenkins, S. P. (1999/2010). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup, Statistical Software Components S366002, Boston College Department of Economics, revised 19 Apr 2001, online, last accessed 12 October 2012, available at http://ideas.repec.org/c/boc/bocode/s366007.html.

Johansson, E. (2009). The History of Literacy in Sweden, in: Graff, H. J., MacKinnon, A, Sandin, B. and I. Winchester (eds.). Understanding literacy in its historical contexts, Lund: Nordic Academic Press, 28-59.

Jones, C. I. (1995). R&D based models of economic growth, *Journal of Political Economy*, 103: 759-784.

Jones, C. I. (2002). Sources of U.S. economic growth in a world of ideas. *American Economic Review*, 92: 220-239.

Jorgenson, D. W. and Z. Griliches (1967). The explanation of productivity change, *Review of Economic Studies*, 34: 249-280.

Juif, D. and J. Baten (2013). On the Human Capital of Inca Indios before and after the Spanish Conquest. Was there a "Pre-Colonial Legacy"?, *Explorations in Economic History*, in press, http://dx.doi.org/10.1016/j.eeh.2012.12.002.

Juillard, E. (1972). Espace et temps dans l'évolution des cadres régionaux, in: Etudes de géographie tropicale offertes à P. Gourou, Paris: Mouton, 29-43.

Kalemli-Ozcan, S. (2002). Does the Mortality Decline Promote Economic Growth?, *Journal of Economic Growth*, 7: 411-439.

Kirk, D. (1946). *Europe's Population in the Interwar Years*, Princeton: Princeton University Press.

Klasen, S. and T. Nestmann (2006). Population, population density and technological change, *Journal of Population Economics*, 19: 611-626.

Kleinewefers, H. (2010). Standort und Raumstruktur. Einige Grundlegende Modelle der Raumwirtschaftstheorie, *Jahrbuch für Wirtschaftswissenschaften / Review of Economics*, 61: 125-155.

Knowles, A. K. (2008). *Placing History: How GIS Is Changing Historical Scholarship*, Redlands: ESRI Press.

Kopsidis, M. (2009). Agricultural development and impeded growth: the case of Hungary, 1870-1973, in: Lains, P. and V. Pinilla (eds.). Agricultural and Economic Development in Europe since 1870, Abingdon: Routledge, 286-310.

Kostovicova, D. (2005). *Kosovo: the politics of identity and space*, New York: Routledge.

Krueger, A. B. and M. Lindahl (2001). Education for Growth: Why and For Whom?, *Journal of Economic Literature*, 39 (4): 1101-1136.

Krugman, P. R. (1991a). Increasing returns and economic geography, *Journal of Political Economy*, 99 (3): 483-499.

Krugman, P. R. (1991b). *Geography and Trade*, Cambridge: MIT Press.

Krugman, P. R. (1999). The Role of Geography in Development, *International Regional Science Review*, 22: 142-161.

Kuznets, S. (1955). Economic growth and income inequality, *American Economic Review,* 45: 1-28.

Kwon, D.-B. (2009). Human capital and its measurement, The 3rd OECD Forum on “Statistics, Knowledge and Policy”, Busan, Korea – 27-30 October 2009, online, last accessed on 25 June 2012, http://www.oecd.org/dataoecd/31/52/44109779.pdf.

Le Gallo, J. and C. Ertur (2003a). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980–1995, *Papers in Regional Science*, 82: 175-201.

Le Gallo, J. and C. Ertur (2003b). An Exploratory Spatial Data Analysis of European Regional Disparities, 1980–1995, in: Fingleton, B. (ed.). European Regional Growth, Berlin: Springer, 55-98.

Levine, R. E. and D. Renelt (1992). A Sensitivity Analysis of Cross-Country Growth Regressions, *American Economic Review*, 82 (4): 942-963.

Levy, H. (1911). *Large and small holdings. A study of English agricultural economics*, Cambridge: Cambridge University Press.

Lindert, P. (2004a). *Growing Public. Social Spending and Economic Growth Since the Eighteenth Century*, Cambridge: Cambridge University Press, Vol. I.

Lindert, P. (2004b). *Growing Public. Social Spending and Economic Growth Since the Eighteenth Century*, Cambridge: Cambridge University Press, Vol. II.

Ljungberg, J. and A. Nilsson (2009). Human capital and economic growth: Sweden, 1870-2000, *Cliometrica*, 3: 71-95.

Lockridge, K. A. (1974). *Literacy in Colonial New England*, New York: Norton.

López-Bazo, E., Vayá, E., Mora, A. and J. Surinach (1999). Regional economic dynamics and convergence in the European union, *Annals of Regional Science*, 33: 343-370.

López-Rodríguez, J. (2008). Regional Convergence in the European Union: Results from a Panel Data Model, *Economics Bulletin*, 18 (2): 1-7.

López-Rodríguez, J. and A. Faíña (2006). EU enlargement, European Spatial Development Perspective and Regional Policy: Lessons from Population Potentials, *Investigaciones Regionales*, 9: 3-21.

López-Rodríguez, J., Faíña, A. and A. García-Lorenzo (2007). The Geographic Concentration of Population and Income in Europe: Results for the Period 1984-1999, *Economics Bulletin*, 18 (7): 1-7.

López-Rodríguez, J., Faíña, J. A. and J. López-Rodríguez (2005). New Economic Geography And Educational Attainment Levels In The European Union, *International Business & Economics Research Journal*, 4 (8): 63-74.

López-Rodríguez, J., Faíña, J. A. and J. López-Rodríguez (2007). Human Capital Accumulation and Geography: Empirical Evidence from the European Union, *Regional Studies*, 41 (2): 217-234.

Lösch, A. (1940). *Die Räumliche Ordnung der Wirtschaft*, Jena: Gustav Fischer Verlag.

Lucas, R. E. (1988). On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22 (1): 3-42.

Machlup, F. and E. Penrose (1950). The Patent Controversy in the Nineteenth Century, *Journal of Economic History*, 10 (1): 1-29.

Madsen, J. B. (2010). The anatomy of growth in the OECD since 1870, *Journal of Monetary Economics*, 57 (6): 753-767.

Madsen, J. B., Saxena, S. and J. B. Ang (2010). The Indian growth miracle and endogenous growth, *Journal of Development Economics*, 93 (1): 37-48.

Manaster Ramer, A., Michalove, P. A., Baertsch, K. S. and K. L. Adams (1998). Exploring the Nostratic hypothesis, in: Salmons, J. C. and B. D. Joseph (eds.). Nostratic: Sifting the evidence, Amsterdam: John Benjamins Publishing Company, 61-84.

Mankiw, N. G., Romer, D. and D. N. Weil (1992). A Contribution to the Empirics of Growth, *Quarterly Journal of Economics*, 107 (2): 408-437.

Manzel, K. and J. Baten (2009). Gender Equality and Inequality in Numeracy: The Case of Latin America and the Caribbean, 1880-1949, *Revista de Historia Económica*, 27 (1): 37-73.

Manzel, K., Baten, J. and Y. Stolz (2012). Convergence and divergence of numeracy: the development of age heaping in Latin America from the seventeenth to the twentieth century, *Economic History Review*, 65 (3): 932-960.

Marshall, A. (1890/1920). *Principles of Economics*, London: Macmillan and Co. Ltd.

Martí-Henneberg, J. (2005). Empirical Evidence of Regional Population Concentration in Europe, 1870-2000, *Population, Space and Place*, 11: 269-281.

Martí-Henneberg, J. (2011). Geographical Information Systems and the Study of History, *Journal of Interdisciplinary History*, 42 (1): 1-13.

McMahon, W. W. (1999). *Education and Development: Measuring the Social Benefits*, New York: Oxford University Press.

Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G. and S. E. Tolnay (1999). The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis, *Journal of Quantitative Criminology*, 15: 423-450.

Middleton, J. (1807). *A general view of the agriculture of Middlesex*, London: McMillan.

Mill, J. S. (1848/1926). *Principles of Political Economy, with Some of Their Applications to Social Philosophy*, Reprint, London: Longmans, Green.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66: 281-302.

Ministère des Affaires Etrangères (2006). *La politique d'aménagement et de développement durable du territoire en France*, Paris: Mediatys/Groupe Corlet.

Minnesota Population Center (2011). *Integrated Public Use Microdata Series*, International: Version 6.1 [Machine-readable database], Minneapolis: University of Minnesota.

Mironov, B. N. (1991). The Development of Literacy in Russia and the USSR from the Tenth to the Twentieth Centuries, *History of Education Quarterly*, 31 (2): 229-252.

Mishkova, D. (1994). Literacy and Nation-Building in Bulgaria, 1878-1912, *East European Quarterly*, 29(1): 63-93.

Mitch, D. (1993). The Role of Human Capital in the First Industrial Revolution, in: Mokyr, J. (ed.). The British Industrial Revolution: An economic perspective, Boulder: Westview Press, 267-307.

Mitchell, B. R. (2003). *International Historical Statistics: Europe 1750-1993*, New York: M. Stockton Press.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. and M. New (2004). A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100), Tyndall Centre Working Paper 55, Tyndall Centre for Climate Change Research, University of East Anglia.

Mokyr, J. (1983). *Why Ireland Starved*, London: George Allen & Unwin.

Mokyr, J. (2002). *The Gifts of Athena. Historical Origins of the Knowledge Society*, New Jersey: Princeton University Press.

Moriconi-Ebrard, F. (1994). *Géopolis, pour comparer les villes du monde*, Economica-Anthropos, Paris: Coll. Villes.

Morillas-Torné, M. (2012). Creation of a Geo-Spatial Database to Analyse Railways in Europe (1830-2010). A Historical GIS Approach, *Journal of Geographic Information System*, 4: 176-187.

Morrisson, C. and F. Murtin (2009). The Century of Education, *Journal of Human Capital*, 3 (1): 1-42.

Moser, P. (2005). How do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs, *American Economic Review*, 95 (4): 1214-1236.

Murtin, F. and M. Viarengo (2010). American education in the age of mass migrations, 1870-1930, *Cliometrica*, 4: 113-139.

NCES (2005). Highlights From the 2003 International Adult Literacy and Lifeskills Survey (ALL), online, last accessed on 25 June 2012, http://nces.ed.gov/pubs2005/2005117.pdf.

Nelson, R. R. and E. S. Phelps (1966). Investment in humans, technological diffusion, and economic growth, *American Economic Association Papers and Proceedings*, 56 (1-2): 69-75.

Netz, R. (2002). Counter culture: Towards a history of Greek numeracy, *History of Science*, 40: 321-352.

North, D. C. (1981). *Structure and Change*, New York: W.W. Norton Company Inc.

Núñez, C. E. (1992). *La fuente de la riqueza: educación y desarrollo económico en la Espana contemporánea*, Madrid: Alianza.

O'Donoghue, J. (2002). Numeracy and Mathematics, *Irish Mathematical Society Bulletin*, 48: 47-55.

O'Rourke, K. and J. G. Williamson (1997). Around the European Periphery: Globalization, Schooling and Growth, *European Review of Economic History*, 1: 153-190.

OECD (1998). *Human Capital Investment: An International Comparison*, Paris: Centre For Educational Research and Innovation.

OECD (2010a). The OECD programme for the international assessment of adult competencies (PIAAC), online, last accessed 25 June 2012, http://www.oecd.org/dataoecd/13/45/41690983.pdf.

OECD (2010b). PISA 2009 at a Glance, OECD Publishing, online, last accessed on 25 June 2012, http://dx.doi.org/10.1787/9789264095298-en.

OECD (2012). OECD Teaching and Learning International Survey (TALIS) FAQ, online, last accessed on 25 June 2012, http://www.oecd.org/document/22/0,3746,en_2649_39263231_47792150_1_1_1_1,00.html.

Pace, R. K. and R. Barry (1997). Quick computation of spatial autoregressive estimators, *Geographical Analysis*, 29: 232-246.

Paci, R. and S. Usai (2009). Knowledge flows across European regions, *Annals of Regional Science*, 43: 669-690.

Peretto, P. (1998). Technological Change and Population Growth, *Journal of Economic Growth*, 3 (4): 283-311.

Perotti, R. (1996). Growth, Income Distribution, and Democracy: What the Data Say, *Journal of Economic Growth*, 1: 149-187.

Persson, T. and G. Tabellini (1994). Is Inequality Harmful for Growth?, *American Economic Review*, 84: 600-621.

Petty, W. (1690/1899). Political Arithmetik, or a Discourse Concerning the Extent and Value of Lands, People, Buildings, reprinted in: Hull, C. H. (1899). The

Economic Writings of Sir William Petty, Vol. I, Cambridge: Cambridge University Press.

Pidal, J. C. and J. R. Rosés (2011). Was land reform necessary? Access to Land in Spain, 1860 to 1931, Universidad Carlos III de Madrid Working Papers 11-01.

Pinkse, J. and E. Slade (1998). Contracting in space: an application of spatial statistics to discrete-choice models, *Journal of Econometrics*, 85: 125-154.

Pomeranz, K. (2000). *The Great Divergence. China, Europe and the Making of the Modern World Economy*, Princeton: Princeton University Press.

Porter, M. E. (1994). The role of location in competition, *Journal of the Economics of Business*, 1: 35-39.

Prados de la Escosura, L. and J. R. Rosés (2010). Capital Accumulation in the Long Run: the Case of Spain, 1850-2000, in: Clark, G. and W. A. Sundstrom (eds.). Research in Economic History, Volume 27, Bingley: Emerald Group Publishing Limited, 141-200.

Pritchett, L. (2001). Where has All the Education Gone?, *World Bank Economic Review*, 15: 367-391.

Ramcharan, R. (2010). Inequality and redistribution: evidence from U.S. counties and states, 1890-1930, *Review of Economics and Statistics*, 92 (4): 729-744.

Raun, T. O. (1986). The Latvian and Estonian National Movements, 1860-1914, *The Slavonic and East European Review*, 64 (1): 66-80.

Redding, S. and A. J. Venables (2004). Economic geography and international inequality, *Journal of International Economics*, 62: 53-82.

Redding, S. and P. Schott (2003). Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?, *Journal of Development Economics*, 72 (2): 515-41.

Reis, J. (2005). Economic Growth, Human Capital Formation and Consumption in Western Europe before 1800, in: Allen, R. C., Bengtsson, T. and M. Dribe (eds.). Living Standards in the Past, Oxford: Oxford University Press, 195-225.

Report of the International Commission (1914). *Report of the International Commission To Inquire into the Causes and Conduct of the Balkan Wars*, Publications No. 4, Washington D.C: Carnegie Endowment for International Peace.

Rodríguez-Pose, A. and R. Crescenzi (2008). Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe, *Regional studies*, 42 (1): 51-67.

Rodríguez-Pose, A. and V. Tselios (2011). Mapping the European regional educational distribution, *European Urban and Regional Studies*, 18 (4): 358-374.

Romer, D. (2001). *Advanced macroeconomics*, New York: McGraw-Hill.

Romer, P. M. (1986). Increasing Returns and Long-run Growth, *Journal of Political Economy*, 94 (5): 1002-1037.

Romer, P. M. (1990). Endogenous Technological Change, *Journal of Political Economy*, 99 (5): 71-102.

Rüegg, W. (ed.) (2004). *A History of the University in Europe, Vol. 3, Universities in the Nineteenth and Early Twentieth Centuries, (1800-1945)*, Cambridge: Cambridge University Press.

Sahai, H. (1988). What is a census?, *International Journal of Mathematical Education in Science and Technology*, 19 (1): 119-124.

Samuelson, P. A. (1952). Spatial price equilibrium and linear programming, *American Economic Review*, 42: 283-303.

Samuelson, P. A. (1983). Thünen at two hundred, *Journal of Economic Literature*, 21: 1468-1488.

Sandberg, L. G. (1979). The case of the impoverished sophisticate: human capital and Swedish economic growth before World War I, *Journal of Economic History*, 39: 225-241.

Scherngell, T. and M. J. Barber (2011). Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU framework, *Annals of Regional Science*, 46: 247-266.

Schmandt-Besserat, D. (1992). *Before writing: From counting to cuneiform*, Austin: University of Texas Press.

Schneider, R. (1982). Die Bildungsentwicklung in den westeuropäischen Staaten 1870-1875, *Zeitschrift für Soziologie*, 11: 207-226.

Schofield, R. S. (1968). The measurement of literacy in pre-industrial England, in: Goddy, J. (ed.). Literacy in traditional societies, Cambridge: Cambridge University Press, 311-325.

Schofield, R. S. (1981). Dimensions of illiteracy in England 1750-1850, in: Graff, H. J. (ed.). Literacy and social development in the west: a reader, Cambridge: Cambridge University Press, 201-213.

Schultz, T. W. (1961). Investment in Human Capital, *American Economic Review*, 51: 1-17.

Schultz, T. W. (1971). *Investment in Human Capital*, New York: Free Press.

Schultz, T. W. (1972). Human Capital: Policy Issues And Research Opportunities, in: Schultz, T. W. (ed.). Economic Research: Retrospect and Prospect Vol 6: Human Resources, UMI, online, last accessed 22 June 2012, http://www.nber.org/books/schu72-1.

Schultz, T. W. (1975). The Value of the Ability to Deal with Disequilibria, *Journal of Economic Literature*, 13 (3): 827-846.

Schütt, F. (2003). *The Importance of Human Capital for Economic Growth*, in: Knorr, A., Lemper, A., Sell, A. and K. Wohlmuth (eds.). Materialien des Wissenschaftsschwerpunktes „Globalisierung der Weltwirtschaft“, Vol. 27, Bremen: Institute for World Economics and International Management.

Segerstrom, P. S., Anant, A. C. T. and E. Dinopoulos (1990). A Schumpeterian Model of the Product Life Cycle, *American Economic Review*, 80 (5): 1077-1091.

Sen, A. (1999). *Development as Freedom*, New York: Anchor Books.

Sheets-Johnstone, M. (1990). *The roots of thinking*, Philadelphia: Temple University.

Sianesi, B. and J. Van Reenen (2003). The Returns to Education: Macroeconomics, *Journal of Economic Surveys*, 17 (2): 157-200.

Sicherman, N. and O. Galor (1990). A Theory of Career Mobility, *Journal of Political Economy*, 98 (1): 169-192.

Smith, A. (1776/1976). An Inquiry into the Nature and Causes of the Wealth of Nations, in: Campbell, R. H., Skinner, A. S. and W. B. Todd (eds.) (1976). The Glasgow Edition of the Works and Correspondence of Adam Smith (Vol. II), Oxford: Oxford University Press.

Snowdon, B. (2008). Towards a Unified Theory of Economic Growth. Oded Galor on the transition from Malthusian stagnation to modern economic growth, *World Economics*, 9 (2): 97-151.

Solow, R. M. (1956). A contribution to the theory of economic growth, *Quarterly Journal of Economics*, 70 (1): 69-94.

Solow, R. M. (1957). Technical change and the aggregate production function, *Review of Economics and Statistics*, 39 (3): 312-320.

Soysal, Y. N. and D. Strang (1989). Construction of the first mass education systems in nineteenth-century Europe, *Sociology of Education*, 62: 277-288.

Sroka, W. (2007). Belarus, in: Hörner, W., Döbert, H., von Kopp, B. and W. Mitter (eds.). The Education Systems of Europe, Dordrecht: Springer, 85-103.

Statistisches Bureau (1862). *Schweizerische Statistik, Bevölkerung. Eidgenössische Volkszählung vom 10. December 1860. I. Lieferung, Statistisches Bureau des eidg. Departement des Innern*, Bern: Walder & Schiller.

Steen, L. A. (1997). *Why Numbers Count: Quantitative Literacy for Tomorrow's America (Literacy Series)*, New York: The College Board.

Sterlacchini, A. (2008). R&D, higher education and regional growth: Uneven linkages among European regions, *Research Policy*, 37 (6): 1096-1107.

Stolz, Y. and J. Baten (2012). Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity?, *Explorations in Economic History*, 49: 205-220.

Stolz, Y., Baten, J. and T. Botelho (2013). Growth effects of nineteenth-century mass migrations: "Fome zero" for Brazil?, *European Review of Economic History*, 17 (1): 95-121.

Stone, L. (1969). Literacy and Education in England, 1640-1900, *Past and Present*, 2: 69-169.

Strong, A. L. (2004). Property Rights in the 21st century: Righting past wrongs, in: Jacobs, H. M. (ed.). Private Property in the 21st century: the future of an American ideal, Northampton: Edgar Elgar Publishings, 143-170.

Swan, T. W. (1956). Economic Growth and Capital Accumulation, *Economic Record*, 32: 334-361.

Swinnen, J. F. M. (2002). Political reforms, rural crises, and land tenure in western Europe, *Food Policy*, 27: 371-394.

Temple, J. (1999). The New Growth Evidence, *Journal of Economic Literature*, 37 (1): 112-156.

Therborn, G. (1995). *European Modernity and Beyond. The Trajectory of European Societies 1945-2000*, London: Sage.

Thomas, K. (1987). Numeracy in Early Modern England, The Prothero Lecture, *Transactions of the Royal Historical Society*, 37: 103-132.

Tilly, C. (1999). Power – top down and bottom up, *Journal of Political Philosophy*, 7: 330-352.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46: 234-240.

Tollnek, F. and J. Baten (2011). The Farmer's U: Which Occupational Group Inherited Human Capital in Early Modern Europe and Latin America?, Working Paper University of Tuebingen.

Tortella, G. (1994). Patterns of economic retardation and recovery in south-western Europe in the nineteenth and twentieth centuries, *Economic History Review*, 47: 1-21.

UNESCO (1953). *Progress of literacy in various countries*, Paris: Firmin Didot et Cie.

UNESCO (1997). *International Standard Classification of Education: ISCED1997*, Paris: UNESCO.

UNESCO (2005). *EFA Global Monitoring Report 2006*, Paris: Graphoprint.

United Nations (2002). Resolution 56/116 on United Nations Literacy Decade, online, last accessed on 25 June 2012, http://portal.unesco.org/education/en/file_download.php/7a7d1ec1686411521f7840d7f41b2e9cResolution+English.pdf.

University of Texas Libraries (2012a). Belarus, online, last accessed on 25 June 2012, http://www.lib.utexas.edu/maps/commonwealth/belarus_pol_97.jpg, Courtesy of the University of Texas Libraries, The University of Texas at Austin.

University of Texas Libraries (2012b). Russia: Adminstrative Divisions, online, last accessed on 25 June 2012, http://www.lib.utexas.edu/maps/commonwealth/russiaaddivisions.jpg, Courtesy of the University of Texas Libraries, The University of Texas at Austin.

University of Texas Libraries (2012c). Ukraine: Administrative Divisions, online, last accessed on 25 June 2012, http://www.lib.utexas.edu/maps/europe/ukraine_adm93.jpg, Courtesy of the University of Texas Libraries, The University of Texas at Austin.

Vanhanen, T. (2003). Democratization and Power Resources 1850-2000 [computer file]. FSD1216, version 1.0 (2003-03-10) Tampere: Finnish Social Science Data Archive [distributor].

Vincent, D. (2000). *The Rise of Mass Literacy: Reading and Writing in Modern Europe*, Cambridge: Polity.

Vinokur, A. D., Schul, Y., Vuori, J. and R. H. Price (2000). Two Years after a Job Loss: Long-term Impact of the JOBS Program on Reemployment and Mental Health, *Journal of Occupational Health Psychology*, 5: 32-47.

Voigtländer, N. and H.-J. Voth (2009). Malthusian Dynamism and the Rise of Europe: Make War, Not Love, *American Economic Review*, 99 (2): 248-254.

Von Franges, O. (1937). *Die sozialökonomische Struktur der jugoslawischen Landwirtschaft*, Berlin: Weidmannsche Verlagsbuchhandlung.

Von Thünen, J. H. (1826/2008). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Reprint, Berlin: Wiegandt, Hempel & Parey.

Von Tunzelmann, G. N. (2000). Technology Generation, Technology Use and Economic Growth, *European Review of Economic History*, 4: 121-146.

Watkins, S. C. (1986). Regional Patterns of Nuptiality in Western Europe, 1870-1960, in: Coale, A. J. and S. C. Watkins (ed.). The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project, Princeton: Princeton University Press, 315-336.

Webb, R. K. (1955). *The British Working Class Reader: 1790-1848; literacy and social tension*, London: Allen & Unwin.

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism*, New York: Charles Scribner's Sons.

Wegenast, T. (2009). The Legacy of Landlords: Educational Distribution and Development in a Comparative Perspective, *Zeitschrift für vergleichende Politikwissenschaft*, 3: 81-107.

Weiss-Bartenstein, W. (1917). *Bulgariens volkswirtschaftliche Verhältnisse*, Berlin: Carl Heymann Verlag.

Wikimedia Commons (2007). Europe, online, last accessed 4 November 2013, http://upload.wikimedia.org/wikipedia/commons/thumb/f/ff/Europe.jpg/680px-Europe.jpg.

Williamson, J. G. (2010). Five Centuries of Latin American Income Inequality, *Revista de Historia Económica*, 28: 227-252.

Winter, W. (1992). Some thoughts about Indo-European numerals, in: Gvozdanovic, J. (ed.). Indo-European numerals, Berlin: Mouton De Gruyter, 11-28.

WIPO (2012). Economic development and patents, online, online, last accessed 20 October 2012, http://www.wipo.int/patent-law/en/developments/economic.html.

Woessmann, L. (2003). Specifying Human Capital, *Journal of Economic Surveys*, 17 (3): 239-270.

Woods, M. and M. B. Woods (2000). *Ancient computing: from counting to calendars*, Minneapolis: Runestone Press.

Xu, B., Chaudhry, S. S. and Y. Li (2009). Factors of production: Historical Theories and New Developments, *Systems Research and Behavioral Science*, 26: 219-224.

Young, A. (1998). Growth without Scale Effects, *Journal of Political Economy*, 106 (1): 41-63.

Zamagni, V. (1978). Istruzione e sviluppo económico. II caso italiano, 1861-1913, in: Toniolo, G. (ed.). L'economia italiana, 1861-1940, Bari: Laterza, 137-78.

Zündorf, L. (1997). Wirtschaftliche Schwerpunktbildungen in Europa. Eine langzeitliche und sozialräumliche Erklärungsskizze, in: Lange, E. and H. Voelzkow (eds.). Räumliche Arbeitsteilung im Wandel, Marburg: Metropolis, 215-248.