Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 31

Settore Concorsuale: 13/A1

Settore Scientifico Disciplinare: SECS-P/01

Essays in Long-run Development and Spatial Economics

Presentata da: **Alexander Lehner**

Coordinatore Dottorato **Supervisore**

Prof.ssa Maria Bigoni **Prof. Matteo Cervellati**

Esame finale anno 2019

Essays
in
Long-run Development and Spatial Economics

Alexander Lehner * [a]

[a]Unversity of Bologna

PhD Thesis in Economics

Thesis Abstract

This thesis consists of three papers that are a topically distinct. The common denominator, and thus also the main theme of this thesis, is the exploration of effects and features that the increasing worldwide spatial interconnectedness brought about throughout the last centuries. Two papers aim to contribute to the age-old, but highly enticing and unresolved question of *"Why are some countries rich and others not?"* through the lens of history. They are geographically focussing on Europe and India, and touching upon diverse sub-themes such as urbanisation/agglomeration, trade, migration, fertility, and the roots of gender inequality.
A further paper is a methodological contribution in order to improve an econometric technique called spatial Regression Discontinuity Designs (RDDs), which comes with a full-fledged statistical package written in the programming language **R** (not explicitly part of the thesis but cross-referenced). All of these works are intended to open up the different strands of my future research.

The first work investigates the long-term impact of Portuguese, i.e. Catholic, colonisation in South Asia by looking at the Indian state of Goa, a place which was constantly colonised for the exceptional timespan of 450 years. My research design is built around a historical, now meaningless, border that was abandoned 250 years ago which I am exploiting for a spatial RDD. I find that sharp gaps, measurable in censuses up until 1991 and 2011, in terms of female education (as measured by literacy) and sex-ratios can be traced back to historical experiences. The colonisers, with the crucial help of missionary orders, forbade sati, polygamy, and childhood marriage and thus unintentionally strengthened the position of women in society for the upcoming centuries. I thus implicitly also provide a new narrative for the evolvement of gender norms for which the early granting of property rights to women by the Portuguese were a crucial condition. By exploiting the differences in discontinuities from 1991 to 2011, I find that educational gaps tend to wash out, but the deeply rooted preference for sons seems to stay unchanged. I thus identify outcomes that are different in the way in which they persist, even though the initial "treatment" was the same.

*First and foremost I thank my parents for their unconditional support over the last three decades, and for giving me the opportunity to be the first person of my family to attend university.
I want to express my gratitude towards Matteo Cervellati for being a great teacher and mentor throughout the last four years. For giving me freedom to develop and shape my own ideas, but being back at the right time to give them a check and channel them.
My period as a Visiting Research Fellow at Brown University from January to July 2018 was a crucial and influential period for my PhD studies. I am indebted to David Weil and Oded Galor for kindly hosting me and for the many hours of discussions. I also want to thank the numerous other people at Brown that I had the opportunity to interact and discuss with.
Lastly I am also grateful to all the members of the Department of Economics in Bologna for providing a fruitful and challenging environment for research and learning.

Out of this paper grew chapter two of the thesis, which is a methodological contribution that aims to improve the state-of-the-art spatial RDD machinery. I am proposing a way to report heterogeneous treatment effects alongside the RD cutoff and a way to visualise the results with more meaningful and intuitive plots that employ spatial smoothing techniques. I then point out problems with the way some spatial RDs are commonly estimated. I provide solutions to alleviate what I identify as a main issue: there are many arbitrary researcher choices that can be made on the way to obtaining statistically significant results. Furthermore, I put recent results from the multi-dimensional RD literature into perspective, and claim that we cannot apply the hard, data hungry standards from the classic causal inference literature to these estimations that exploit geographic variation. This is being demonstrated with (spatial) Monte Carlo simulations.

The third chapter, co-authored with Matteo Cervellati and Gianandrea Lanzara, develops a novel quantitative spatial economic model with two differential sectors. The insights from the model are then utilised in order to explain the peculiar evolution of the European urban system on its way to the Industrial Revolution. Contrary to common intuition, we postulate that it is not the decreasing costs to trade manufactured goods. It was rather the increasing possibilities to trade agricultural goods that was facilitated by the peculiar geography of Europe with its long coastline and dense Network of navigable rivers. This allowed to feed more people from early on, even in the absence of an increase in agricultural productivity. Thus opening up the possibility for the unprecedented dense city network to evolve. This was one of the necessary conditions to capitalise on increasing returns that are drivers for innovations but also for technological and institutional improvements. The model is fed with a plethora of data, most importantly the (updated and revised) data on urban population by Bairoch, and the bilateral effective trade costs between all gridcells in Europe. It implies a spatial distribution of the (rural) population and delivers urbanisation/specialisation rates. This endogenous specialisation takes place without the need of non-homothetic preferences, thus uncovering a novel mechanism of structural change.

Table of Contents

450 Years of Portuguese Colonialism in India: Missionaries, Education, and the Roots of Gender Inequality *

Alexander Lehner [†a]

[a]Unversity of Bologna

This version: September 28, 2019

Abstract

This research makes use of a historical quasi-natural experiment to document the persistent effect of Portuguese (catholic) colonialism in a South-Asian context. It then utilises Goa as a "lab" in order to study a set of questions that pertain mostly to the roots of gender inequality and related issues regarding male favouritism (e.g. biased sex-ratios) and educational gender inequality in India. The empirical strategy exploits peculiarities of the history of the colonisation of Goa. In some parts the colonisers forbade satì, polygamy, childhood marriage, and gave women property rights already around 500 years ago. Furthermore Catholic missionaries brought structured education.

The village level analysis achieves identification via a geographic RDD that is carried out in several different ways. I point out the weakness of common estimation approaches and propose a more transparent and rigorous way in which results could be reported and analysis carried out. This is implemented in companion (geo-)statistical package written in **R** .

I provide strong evidence that the early Portuguese colonial activities were the cause of (geographically) sharp within differences that are still measurable today in terms of male biased sex-ratios and the gender gap in education. Looking at differences in discontinuities, I find that the latter converge, but male son preferences do not. I thus identify outcomes that have differential degrees of persistence within the same quasi-experimental setting. Due to my identification device I can trace this effect back to culture, as institutions and other observables are constant across the RD cutoff. In addition to the time component I also find tentative evidence for horizontal diffusion through space: also Hindus and Muslims within the "treated areas" exhibit systematically lower educational gender gaps. For the male son preference I again fail to find such an effect.

Keywords: Colonialism, Roots of Gender Inequality, Male Favouritism, Culture vs. Institutions, India, Portuguese Empire, Geographic Regression Discontinuity Design (GRDD), Spatial RDD.

JEL Codes: F54, I24, I25, J16, N35, O15, O43, Z12.

*Winner of the biennial Young Researcher Award of the Regional Science Association International. Goa, India, June 2018.

†First and foremost I want to thank my academic supervisor, Matteo Cervellati, for his guidance and support. I am especially grateful for detailed comments from David Weil and Klaus Desmet. So far I benefited from extensive conversations with Jeanet Bentzen, Melissa Dell, Alessandra Fogli, Oded Galor, Nicolà Gennaioli, Trevon D. Logan, Mara Squicciarini, Robson Tigre, Rocío Titiunik, Romain Wacziarg, and Andrea Weber. I received very helpful comments from Ankush Asri, Stephen Broadberry, Jeanet Bentzen, Jean-Paul Carvalho, Siddharth Chandra, Rui Esteves, Margherita Fort, Paolo Masella, Giovanni Prarolo, Ferdinand Rauch, Anand Swamy, Uwe Sunde, Eric Tagliacozzo, Christian Ochsner, Felipe Valencia, seminar participants at the University of Oxford (Nuffield, Graduate Seminar Series), the Spring Meeting of Young Economists 2017, Michigan State University, CESifo Munich, EHES 2017 Tübingen, ASREC Europe 2017, the University of Bologna, the Macro Lunch at Brown University, RSAI18 (Goa), WEAI18 (Vancouver), NOeG19 (Graz), and the EEA-ES meeting in Manchester. E-Mail: alexander.lehner@unibo.it

1 Introduction

The Indian state of Goa was colonised at a very early stage and *constantly* under Portuguese rule for 450 years, making it the longest constantly colonised piece of land in younger human history. Once the capital of the mighty Portuguese seaborne empire, due to the terribly bad governing of the territory Goa was essentially the poorest area in India upon its liberation in 1961. Yet, once the Indian government went in and built up infrastructure, a rapid catch-up process began that transformed Goa to what is essentially the richest state of the sub-continent[1]. By examining the long-run effect of Catholic colonisation in a South-Asian context, this research establishes that deep-rooted cultural norms regarding education and gender might interact differentially with modern-day laws and policies and thus have a diverse effect on observable outcomes today. Areas in Goa which had historically distinct norms regarding the position of women in society show a convergence in terms of gender gaps in education. Yet, there seems to be no change when it comes to one of India's most challenging issues: its pronounced favouritism towards sons; measured by the sex-ratio in this study.
Building on previous work that established that human capital transmission across generations is a main driver for prosperity in the long run, I establish that early missionary activities had a long lasting impact on education and prosperity. Another layer is added by analysing a switch in cultural norms towards women that the Portuguese induced by force throughout the 16th century. The burning of widows (satì), childhood marriage, and polygamy were forbidden [2]. Upon conversion to Catholicism women were granted property rights and could thus inherit. These interventions regarding the position of women in society, as well as the schooling network of the missionaries were in effect only in a specific part of contemporary Goa, the so-called *Old Conquests*, that is geographically distinguishable due to a historical quasi-natural experiment. For identification, this study partly employs the strategy that was used by Valencia Caicedo [2019], resting on the exogenous expulsion of the Jesuits by the pope in Rome. In addition to that, I make use of the aforementioned quasi-natural experiment, redraw a now meaningless border that has been abandoned 250 years ago, and create a spatial Regression Discontinuity Design (RDD). At the village level alongside this cutoff I observe robust jumps in terms of the gender gap in education (as measured by literacy rates), and the sex-ratio in 1991 and 2011. These discontinuities explain the respective average gaps almost in their entirety and there is almost no gradient across space observed. From 1991 to 2011 I document a convergence[3] of the areas that did not experience the first wave of colonial "activities" by the Portuguese which were marked by the abovementioned cultural and educational interventions. This convergence, though, I only observe for the gender gap in literacy rates, but the difference in the sex-ratio alongside my narrow RDD cutoff between the so-called *Old & New Conquests* remains unchanged. This implies that the government interventions, which were very effective in Goa and contributed to its rise, were able to address the gender gap in education (on top of the enormous rise of overall literacy), but had seemingly no effect on the deep rooted male child preference in the *New Conquests*, an area which is comparable to

[1] In terms of local GDP and HDI numbers as reported by the Planning Commission of the Government of India and several other indices of social progress.

[2] It has to be noted that these things were obviously not undertaken out of philantropic motives. It was what the Portuguese believed to be a superior way of organising society. Further it should be noted that the early colonial years have been marked by hostility towards the Muslim population, which were preceded by hostility by Adil Shah and the Bahmani Muslims towards the local Hindus prior to the European arrival.

[3] 1991 marks the first census recording after Goa was upgraded from a Union Territory together with Daman and Diu to a state in 1987.

large parts of India in terms of its socio-demographic composition.
This allows me to relate the study towards the pressing problem of male son preferences which result in male skewed sex-ratios [as pointed out by Sen, 1990][4]. These are getting worse in recent years, and research indeed shows that these biased sex-ratios don't improve as countries develop and being poor is insufficient to explain them [see e.g. Jayachandran, 2015].

Through an extensive historical narrative, the analysis of historical census data (showing that the gap was already there in the 19th century and not an artifact of post-colonial policies), and the verification of numerous "placebo outcomes", I provide extensive evidence that the root of these differences lie in the early improvement of the position of women and the early exposition to organised education in the *Old Conquests*. I also argue that these cultural values not only were transmitted vertically throughout generations, but also horizontally within village communities. The latter is elicited by the fact that the distribution across space of the educational gender gap for Hindus and Muslims, the former constituting the main religious group in contemporary Goa, is systematically different in the *Old Conquests* as shown by non-parametric statistical tests. It thus seems that the effect that I am describing is not something that stems from "being Christian", but rather from something cultural that is deeply rooted within the village communities.
Clear advantages of my study are, first of all, that I am investigating cultural variation, holding the institutional setting constant, in a geographically very small area. Thus, at least to comparable literature, I am less likely to suffer from omitted variable problems since first nature geographical features such as climate, agricultural suitability, and the disease environment are virtually constant across space. This very localised and disaggregated nature of my study also demonstrates that inequality potentially has a spatially fine grained dimension: an analysis on an aggregate level would have averaged over all of these and potentially come to a different conclusion.

Goa's general economic success due to its impressive catch-up after the liberation in 1961 might be partly driven by the whole uniform Portuguese institutional setting (the civil code of 1871 - based on the Code Napoleon - which is still in place up until today). It is still the only state so far in India which has a uniform civil code that applies invariably to *all* people, independent of religion or gender. The importance of historical shocks to institutions has been pointed out in the economics literature, most importantly by Dell [2010] in the context of a forced labour system in Latin America.
The novelty of this paper is that it investigates the long-run impact of an early forcefully induced shift of cultural norms regarding women and education, restricted to a specific part - the *Old Conquests*, interacted with a general institutional setting that was implemented later on by the Portuguese without any geographical restrictions in all of Goa. These cultural norms persisted locally and are the reason why educational gender gaps are lower and sex-ratios are higher in the parts that were colonised early. These places experienced the "double-treatment": exposure to schools and structured education, and the early alleviation of the position of women in society.
I am thus contributing to a novel strand that tries to investigate the interaction of culture and institutions, something that is still not well understood by economists [as has been recently argued by Alesina and Giuliano, 2015]. Compared to other, more general studies on

[4]Sen spoke of 100 million missing women. Recent demographic estimates suggest that currently around 60 million women are missing in India. The problem is very similar in China.

the effect of institutions or cultural norms, this research arguably does not suffer from the ubiquitous problem that differences in observable institutions/norms are almost always related to some other unobservable characteristic. Due to the peculiar and "lucky" nature of the territorial expansion by the Portuguese in the late 18th century which had nothing to do with any economic fundamentals, I would claim that the assignment of norms was undertaken in a close-to-random way (even though I have to admit that the *Old Conquests* were under colonial rule for 250 years longer and in addition experienced the heydays of Goa in a more direct way). The institutional set-up is held constant across space and the RD cutoff identifies variation in cultural attitudes towards education and women only. Thus the described effect can be fully ascribed to culture. This confirms results from the literature on the deep roots of male son preferences [Abrevaya, 2009; Almond, Edlund, and Milligan, 2013].

Arguably the most important contribution of this study is towards the literature on the evolvement an importance of female agency [5]. Especially in the context of developing countries it is paramount to understand the roots of gender inequality [see e.g. Jayachandran, 2015, for a recent survey]. In numerous studies that examine the strong correlation between women empowerment and economic development, it is typically not clear in what direction the causation goes and whether the interrelationship is self-sustaining [Duflo, 2012; Doepke, Tertilt, and Voena, 2012]. A common problem is that efforts towards the improvements of the position of females in society are often endogenous in the sense that they are a function of economic growth and commonly implemented features in societies on their way to prosperity. In the history of Goa I identify several early "treatments" regarding the improvement of the position of females in society. These pre-date the experience of economic development by several hundred years and thus allow me to take a much stronger position on the causal link from the improved position of women towards economic prosperity. What is more I can also outrule that there are any channels regarding higher female wages (or comparative labour market advantages) and their positive implications at play [see e.g. Galor and Weil, 1996; Xue, 2018; Qian, 2008]. The villages that are compared on both sides of the RD cutoff are very similar in all fundamental characteristics. Since they are very close to each other and commuting is possible, wage rates should equalise in any case [6]. Channels regarding a higher female labour force participation can also be outruled as the share of women in the workforce was even higher in the parts that did not receive the "gender treatment" (see Section 3.

Contrary to my study, very recent sociological research argues that religion played an ambiguous role in the early-modern onset of Portuguese colonialism, as compared to other cases where religion is highlighted as a strong marker and maker of cultural difference [Henn, 2014]. This stems mostly from the fact that in Goa things seem to be somewhat molten together: Christian converts kept a lot of Hindu practices up until today (e.g. the dowry system, clothing, ...). I am trying to convince the reader, that even though Hindu and Chris-

[5] Starting off more than two decades ago with the theoretical contribution by Galor and Weil [1996]. Important contributions being Doepke and Tertilt [2009], Fernández [2014], Doepke and Tertilt [2018], among others. Tertilt [2005] documents the effect of banning polygamy (polygyny). Heath and Jayachandran [2016] document the effect of increased female education. Dhar, Jain, and Jayachandran [2018] documents the intergenerational persistence of gender attitudes that was also present in the Goan context.

[6] Issues regarding migration will be discussed in the empirical section, but in general a sorting around the border of individuals can be outruled. Migration existed, but it happened from rural villages towards the coastline where the jobs are and not within the bandwith around the RD border. Thus this does not invalidate the identification strategy.

tian culture seems to be somewhat molten together, what mattered most for contemporary economic outcomes was the common historical experience of those families and villages in the *Old Conquests*. My study argues that, albeit having no clear measurable impact during early stages apart from the differential sex-ratios, the historical experience turned out to be important once education became widely available after the liberation in 1961. A certain set of believes towards education and the status of women was conducive for the appreciation of education and made people more likely to send their kids to school. This effect through missionaries on education was already shown to be important in the Latin American context by Valencia Caicedo [2014] and Waldinger [2017].

Broadly speaking I am contributing to the persistence literature that tries to link historical events to contemporary economic outcomes, emphasising the importance of history for economic development, as e.g. summarised by Nunn [2009, 2014b], Spolaore and Wacziarg [2013]Michalopoulos and Papaioannou [2017]. Within this sub-field, the present study tries to assess the long-term impacts of colonialism in the specific context of the Portuguese seaborne empire [the historical seminal study still being Boxer, 1969] and the Indian Ocean trade[7]. Contributions focusing on Asia in this respect almost exclusively study the impact of Britain's influence in India [Banerjee and Iyer, 2005; Iyer, 2010; Gaikwad, 2015]. Recently there is some work that tries to asses the long-term impacts of Protestant missionaries in the territories of the British East India Company [Mantovanelli, 2013; **?**], and of colonial educational investments therein [Chaudhary and Garg, 2015; Castelló-Climent, Chaudhary, and Mukhopadhyay, 2017]. Supposedly one reason why not much attention was paid so far to Lusitanian legacies, was that the Portuguese empire was a "forgotten empire", amongst historians and even more so amongst economic historians up until recently [Marcocci, 2012, p. 33]. Arguments, explanations and a description of the development in the field of Portuguese "overseas history", including a summary of the 2003 e-JPH debate are described in detail by Ferreira [2016].

More specifically I am contributing to the very recent literature on the long-term effects of (Catholic) missionary orders. Studies on Africa and India generally focus on protestant missions: e.g. Cagé and Rueda [2016] point out the positive impact of the printing press that was present in protestant missions in Africa & Nunn [2014a] documents the beneficial impact of protestant missions on women in Africa. Woodberry [2004, 2012] documents the strong correlation between e.g. educational attainment and the presence of Protestant missions. This is the first work that specifically tries to assess the impact of Catholic missionaries in Asia within this highly localised "Goan-setting" that heavily reduces the concern of omitted variable problems. Since I also can clearly distinguish the area of influence of the two main orders, the Franciscans and the Jesuits, before their expulsion at least, this work can be seen along the line of what Valencia Caicedo [2014] and Waldinger [2017] are doing in Latin America by comparing the differential impact of the different orders. Waldinger [2017] claims that the mendicant orders had a more beneficial impact on contemporary economic conditions in Mexico and Valencia Caicedo [2014] underlines the heavily beneficial impact of the Jesuits in the Guaraní area in contemporary Paraguay, Argentina & Brazil. Having a clear geographic distinction between Jesuits and Franciscans, my setup also allows me to draw the tentative conclusion that the Jesuit imprint was felt stronger and lasting longer. The Jesuit areas have slightly lower literacy gaps, the highest sex-ratios, and still the high-

[7] And thus also generally relates to previous work investigating the impact of colonialism and colonial investments as in **?**, Huillery [2009], Bruhn and Gallego [2012].

est shares of Christians in general. This is corroborated by the fact that they became the virtual overlords of Goa throughout the 16th and 17th centuries [Borges, 1994; Velinkar, 1984].

The methodological contribution of this paper is that it implements a more fruitful way to carry out geographic RD's that does no throw away spatial features of our data [as suggested by Keele and Titiunik, 2015]. Also the finite-sample-minimax linear estimator approach by Imbens and Wager [2019], including the visualisation of average weighted treatment effects, is implemented and supports my results. Everything is implemented in a companion (geo-)statistical package for easy replication[8]. I show that some commonly put forward specifications of spatial RDD's may lead to spurious conclusions in the context of this study. At first sight there seems to be also a discontinuity in male literacy rates and female labour force participation, after some more detailed investigations using spatial interpolation techniques and exploring the heterogeneity of the treatment effect alongside the cutoff, these have to be flagged as non-robust. This is discussed in more detail when the results are presented in Section 4.2. Further technical details are provided in the Appendix.
These nonparametric approaches to pin down potential effects also somewhat reduce concerns about the persistence literature that have been raised recently [Kelly, 2019]. No distributional assumptions on parameters are needed for these estimations and thus error terms that could potentially be spatially autocorrelated do not exist.

[8] Written in R. Preliminary version available from Github.

2 Historical Background

<table>
<tr><th colspan="2"><u>Historical Narrative - A Quick Overview</u></th></tr>
<tr><th>Old Conquests</th><th>New Conquests</th></tr>
<tr><td>• Conquered in 1510 (Tiswadi) & 1543 (Bardez, Salcete/Mormugao)</td><td>• Acquired in peculiar ways in different stages around 1780</td></tr>
<tr><td>• Experienced the heydays of the Portuguese Estado da Índia where the so-called ”golden Goa” had supposedly up to 100.000 inhabitants</td><td>• No ”early” economic impetus since the Portuguese thalassocracy was already at the bottom when these parts were acquired</td></tr>
<tr><td>• Network of parishes/schools from early days</td><td>• Missionaries never enter</td></tr>
<tr><td>• College(s) [printing press]</td><td>• Structured education arrives late</td></tr>
<tr><td>• Sati, polygamy, early childhood marriage curbed from early 16th century on</td><td>• Polygamy and early childhood marriage ubiquitous up until the 20th century.</td></tr>
<tr><td>• Women can inherit and remarry already in the 16th century</td><td>• Laws improving the position of women being implemented from late 19th century on</td></tr>
<tr><td colspan="2">• Uniform civil code of 1871 (still in place, makes Goa the only state so far in India which has a uniform civil code, applying to all people across religions, female and male)
• From 1961: uniform investments from Indian government (schools, electricity,...)</td></tr>
</table>

Table 1: An overly simplified sketch of the two historically distinguishable parts of Goa.

2.1 Geography

The contemporary Indian state of Goa, admeasures an area of only around 3700 square kilometres and is located mid-way on the west coast of India. Technically its around 120km long coastline is part of the Konkan Coast, to its south the Malabar Coast begins. Goa stretches out to a width of about 60 kilometres in an east-west direction and extends to a length of about 105 kilometres from north to south. To the east, Goa (and the whole Konkan) are separated from the Deccan highlands of Karnataka, the neighbouring state to the west and south, by the mountain ridge of the western Ghats. Thus from the outset it can already be seen that my study focuses on a geographically very restricted area that allows me to hold constant important features of first nature geography such as climate, soil quality, and the like. Consequently a potential omitted variable bias is less of a concern.

Politically the state is divided into two districts and eleven sub-districts, so-called talukas: Pernem, Bardez, Bicholim, Tiswadi, Sattari, and Ponda being part of North Goa; and Mormugao, Salcete, Sanguem, Quepem, and Canacona being part of South Goa.

Goa enjoys a salubrious, sub-tropical, monsoon type of climate with alternating wet and dry seasons.

Linguistically it belongs to the Konkani-speaking region that reaches from Thane in Maharashtra, the neighbouring state in the north, to Mangalore in Karnataka in the south. The current population comprises about 1.2 million people of whom 65 percent are Hindus, 27 percent Christians, and 6 percent Muslims according to the 2011 Indian Census. It stands out from its neighbours' culture by apparently European, that is, Portuguese, features in its architecture, folklore, and cuisine. Generally this is true mostly for specific parts of Goa, namely in the so-called *Old Conquests*, as will become clearer later.

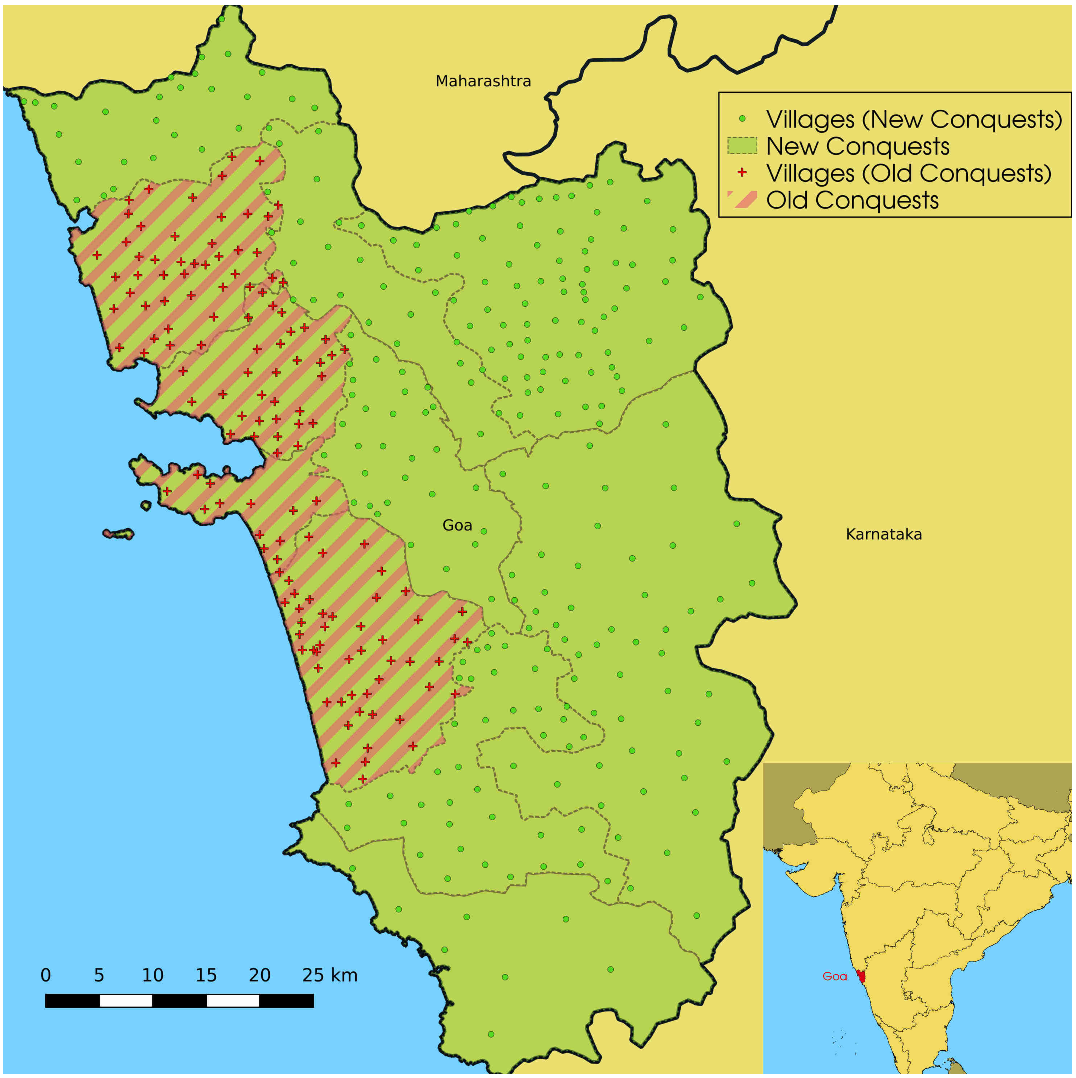


Figure 1: The location of Goa on the Indian west coast.

2.2 Portuguese Conquest

The Portuguese were the first foreign power to arrive and the last one to depart from the sub-continent of India, their 450 year long stay thus marks one of the longest uninterrupted periods of colonisation in history. The lusitanian adventure in the Indian Ocean starts in 1498 when Vasco da Gama lands in the port of Calicut with the famous first sentence *"We are looking for Christians*[9] *and spices!"*. Later, in 1510, Alfonso de Albuquerque captured the Islands of Goa (its territory roughly equivalent to today's taluka of Tiswadi) from the Sultan

[9]Prester John was thought to be in India at that time, plus they had evidence that the early apostles converted Indians already 1500 years ago.

of Bijapur. By 1543, the Portuguese had annexed the adjoining lands of Bardez in the north and Salcete (including today's taluka of Mormugao) in the south. These three territories have been designated as *Velhas Conquistas*, the Old Conquests.
It soon became the capital of the Portuguese *Estado da Índia* and rose to one of the world's most magnificent cities of the 16th century. It became an important trading post that connected China, Japan, the Moluccas (the famous "Spice Islands") and India with Europe. It was also referred to as *Goa Dourada*, the golden Goa, and supposedly was comparable in size to the major cities in Europe back then [Srivastava, 1990]. The extensive prosperity was caused by trade, mainly with spices, and was facilitated by naval superiority. For which the foundation was essentially laid by Henry the Navigator and his successors in the 15th century. They managed to prevent the secrets of the demanding maritime navigation in the Indian Ocean from spreading to other European nations for around 100 years, up until the Dutch traveller and later secretary to the viceroy in Goa, Jan Huyghen van Linschoten, copied all the information and published it in his *Itinerario* in 1596 [see e.g. Russell-Wood, 1992]. Due to this catch-up in knowledge by the ascending Dutch provinces, the Portuguese quickly lost their pre-eminence in the Indian Ocean around 1600. Their entrance with the newly formed East India Company, the so called Vereinigde Oostindische Compagnie (VOC), marked the beginning of a sharp decline of the whole Portuguese seaborne empire in general, and of its capital Goa in particular.
One of the distinct features of the Portuguese colonisation strategy was their zeal to convert the local populace to Christianity, and their enmity towards Muslims. Mass conversion campaigns were flanked by the destruction of temples and mosques. In Goa the result was an exodus of Hindus and Muslims, of which the former were by far the biggest religious group, and a rising number of Christians who soon became the overwhelming majority. The village communities, the so-called *Gaunkaris* (or *Communidades*), around which economic and social life was organised were typically left in tact. The only requirement was that taxes and tributes were collected by the *Gaunkars* and transferred to the viceroy in Goa.
The key feature of the Portuguese colonising "strategy" as opposed to all other European colonisers was, that they always encouraged their men to intermarry with local, in this case Hindu, women. The aim was to generate a local populace which is loyal to the colonial government and thus reduce the potential for uprisings and revolts. Upon marriage with a Portuguese *soldado*, which made him a *casado* then, women were granted property rights. Thus their position in society was strengthened and they could inherit in case their husband died. Prior to that women were typically burned on the pyre of their dead man (satì). This clearly served as an additional incentive for local females to convert.

2.2.1 The Portuguese Trading (Colonising) System

The state government in Goa was a macrocosm of that of the other areas and forts [10]. Each had a captain, usually assisted by a "vedor da fazenda", other minor officials such as clerks, and more important the factor, who supervised the royal trade in the area. There were also various clerics, a judge, and in the larger areas a municipal council. The object of all these forts and captains was to enable the Portuguese to achieve several economic aims. These may be listed as: a monopoly of the spice trade to Europe, a monopoly on the trade between various specified ports within Asia; the control, direction and taxation of all other trade in

[10] In total the Portuguese empire comprised of XX forts in year XXYY [Boxer, 1961]

the Indian Ocean; private trade, done on their own behalf by most Portuguese living in Asia [Pearson, 1988].
Most of the official positions ensured high revenues to their holders. Distinction between public and private funds supposedly must have been blurred. The bulk of the officials' salary came in the form of several privileges and perquisites, such as the right to collect a certain tax [11][Pearson, 1988]. Since they got rewards into their own pockets, they somehow had an incentive to extract. Probably one of the sources for the corruption that is described as endemic in numerous historical records [e.g. Boxer, 1969]. The biggest chunk of the state revenue was derived from customs duties.
Generally speaking, instead of exploiting the comparative advantage of their Cape-route in the East India trade and to supply Europe with as much spices as possible, they focused on taxing the already existing Indian Ocean trade and extracting as much as possible from the local (mostly Muslim) traders that operated between the Moluccas (the so-called "spice islands"), Malacca (close to nowadays Kuala Lumpur) and the Red Sea. Their lack of manpower [12] and the notorious inefficiency of the Portuguese colonial system can be viewed as the root cause of its later downturn.
One important implication of all this for my present study on Goa is that the respective colonial rulers never really were interested in developing the hinterlands or making meaningful public investments, at least not until the eve of the 19th century. Portuguese "interference" happened almost exclusively through the Catholic missionaries, above all the Franciscans and the Jesuits.

2.2.2 Decline of the Seaborne Empire

As already indicated, a sharp decline of the whole seaborne empire set in at around 1600, when the rapidly ascending trading company of the Dutch Provinces, the VOC, entered the Indian ocean. Throughout the 17th and 18th centuries, the Portuguese thalassocracy was deprived of almost all its strongholds. Albeit the Dutch made two serious efforts to capture Goa, it constantly remained under Portuguese control.
Goa's customs revenues fell by almost one half between 1600 and the 1630s. Estimates of the value of her sea trade, based on these customs figures, show a decline from 2,700,000 cruzados in 1600 to 1,800,000 in 1616/17, 1,400,000 in 1635, and a minuscule 500,000 by 1680 [Pearson, 1988]. Goan capitalists vanished as the economy deteriorated. Additionally the Portuguese crown devoted its meagre resources mostly to Brazil for it was generally always higher valued than the possessions in the east.

It seems that there was a rise of country trade in the later sixteenth century as the carreira to Portugal declined.

One of the leading historians of the field, Malyn Newitt, suggests people to take care on the differentiation between the different periods. they have been different and must thus be treated differently Newitt [2005]. I would argue that, when we read the history of Goa, we need to do likewise.

[11] The posts were seen as property from which the holder expected to make a profit, and thus they were willing to pay much more for a lucrative post than for one with few opportunities for pickings, even if the status of the latter was higher than that of the former. In a sale of 1618 more than twice as much was paid for the post of judge of the Goa customs house as for the captaincy of the whole city for example.

[12] The whole *Estado da Índia*, from Mozambique to Nagasaki, never had more than 10.000 soldiers [Boxer, 1969].

2.2.3 The New Conquests of Goa

In 1763 the Portuguese were able to capture Ponda from the Marathas. In 1764 Sanguem, Quepem, and Canacona were placed under Portuguese jurisdiction because of an invasion by Hyder Ali. During hostilities between the Bohsles and the Raja of Kolhapur, the Portuguese took advantage of the situation and captured Pernem, Bicholim and Sattari and annexed them to Goa. By 1788 the modern territorial boundaries of Goa had been chalked out. The new parts were termed as the *Novas Conquistas* (New Conquests). They extended the area of Goa to the north, south and east; the Old Conquests make up about 785 square kilometres, the New Conquests a little under 3000. The latter were little valued at the time, but it was here that large deposits of iron ore were later exploited. By the end of the eighteenth century Portuguese India consisted of this enlarged Goa, and a moribund Diu and Daman [13].
The attitude towards the autochthonous population was by far not as hostile as it was during the 16th century in the Old Conquests. The locals were even guaranteed religious freedom. What is more, the main proponents of the proselytism in the 16th century, that is, the religious orders, were not present anymore: the Jesuits were expelled in 1759 and all other remaining orders were forbidden in 1835 [de Souza, 1990, p 107].
Goa remained Lusitanian until the 19th December of 1961, when the Indian Army drove out the colonisers with Operation Vijay. Thus Goa today marks the territory of the longest-held European colony in all of Asia, if not the world.

2.2.4 Why could the Portuguese rule uninterruptedly for 450 years?

In the early days of Lusitanian presence, their main opponents were the different Indian empires which were far superior by almost any means to the Portuguese (and most certainly to all other European powers at that time as well), except for naval technologies and strategies. All of those Indian empires were land based entities and their rulers didn't care too much about the coastline. The trade in the Arab Sea and the Indian Ocean was carried out by several trading guilds, most of which were of Muslim faith, since hundreds of years prior to the arrival of Europeans. Since the Portuguese Seaborne Empire was meant to be a trading empire with coastal outposts scattered around the world, a so-called thalassocracy, there was not much conflict of interest with the autochthonous ruling classes of the south Asian subcontinent in the first place.
Later on, when the British became the overlords of India, they had little intention to conquer the small remaining Portuguese parts due to the strong ties of Portugal to the English crown. Catherina de Braganza married Charles II (part of her dowry was Mumbai, or Bom Bahia as the Portuguese called it). Later the Methuen treaty makes Portugal essentially a British trading puppet.

2.3 Christianity

From the historical introduction so far it is not hard to tell, that the *Estado português da Índia* in its early days during the 16th and early 17th century[14] was more or less a theocratic entity.

[13]The Portuguese lost the extensive lands and quite prosperous town of Bassein to the Marathas in 1739; the New Conquests were somehow seen as a sort of compensation for the losses in the north

[14]It is important to distinguish between the different, in some respects distinct, periods of the Portuguese thalassocracy when it comes to analysing the respective historical narratives as Newitt [2005] strongly empha-

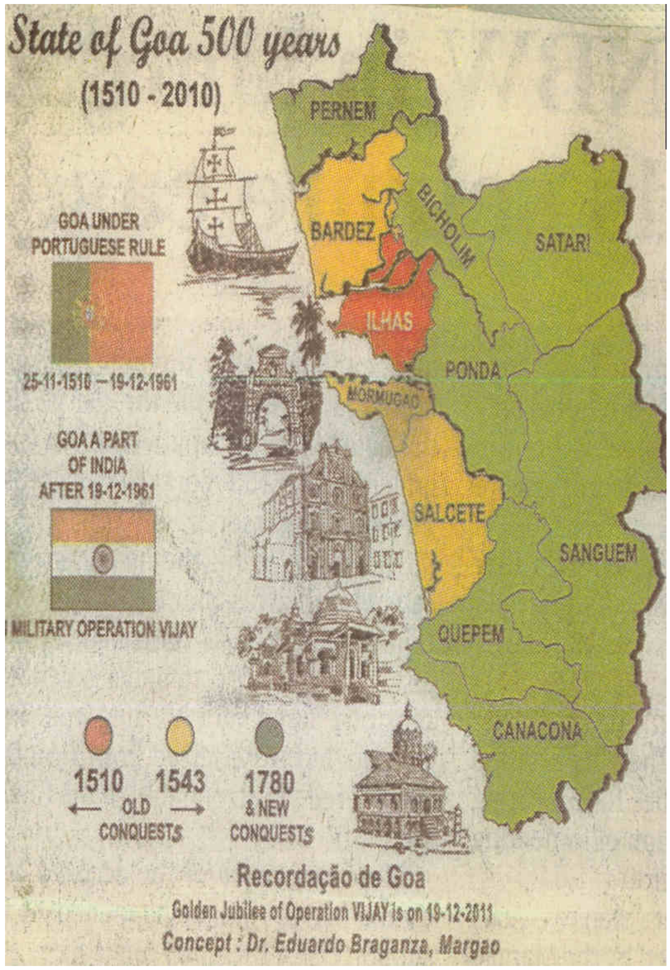


Figure 2: The expansion of Goa

During this time, its state machinery was fully geared towards the needs of evangelisation. Later on, as their seaborne empire and the Portuguese state itself were continuously declining, they lost their zeal for proselytisation. From the 16th century on, Goa, which before had experienced the presence of Hindu and Muslim rulers alike, was now subjected to probably the most dominant themes of Portuguese imperialism: christianisation and acculturation.
Especially the Muslims of Goa were exposed (has tradition in POR, think of Ceuta) to harsh brutality. In the first place, the Hindu majority enjoyed independence in order to remain a certain degree of local support. There is a "conversion-by-conviction-or-by-coercion debate" of how much the christianisation efforts harmed the local population of Goa and how cruel they were [see e.g. Kamat, 1999, p. 42 for a discussion].

2.4 The Role of Women

Generally speaking, the Portuguese tried to change as little as possible when they arrived. They deliberately left the economy in the hinterland, which was organised on a village level, as much in tact as possible in order to e.g. maintain a stable tax base. This was especially

sises.

true for the *New Conquests*, once they were acquired. The colonial government needed these pieces of land for agricultural purposes and thus didn't want to interfere with the Hindu land-owning class. Certainly one of the reasons why people in those parts that were acquired around 1780 were guaranteed complete religious freedom.
The few main "interventions" they tried to make regarding customs in the Goan society actually specifically concerned females.
Discrimination against women was (and still is to some extent, it has to be remarked) apparent in several ways on almost the entire sub-continent of south Asia: early childhood marriages, polygamy, the interdiction of remarriage, the prohibition of property rights (girls could thus not inherit from their families or passed away husbands), sati (a practice where the widow is burned alive on the pyre of her husband) [15], and infanticide [16] (since the 1970's rather feticide). Da Silva Gracias [1996], Chapter 1, draws an extensive picture of the status of women in Indian society at the time of the Portuguese arrival.

Upon his arrival, Afonso the Albuquerque issued the so-called *Politica dos Casamentos*, a mixed marriage policy between his men and native women. He incentivised this by endowing those couples with cash, a house, cattle, and land [see e.g. Da Silva Gracias, 1996; Kamat, 2000]. These Portuguese soldiers were then referred to as *casados*. The main purpose, it is said, was to create a "white" identity in the *Estado* and to develop a community which is loyal[17] toward the Portuguese. Intermarried women, which had to convert to Christianity, thus could suddenly claim all the privileges of a born Portuguese.
Also upon their arrival, the Portuguese immediately forbade practices that supposedly appeared barbaric to them: the tonsuring of widows and the practice of sati. Later on polygamy (1567) and childhood marriages were also forbidden (sat'i was abolished in the *New Conquests* only in 1884). Whether all those laws were effectively serving their purpose at the time they were implemented must be questioned. The historical narrative suggests that people tended to continue with traditional practices. But it has to be acknowledged that there existed at least written laws that tried to curb the discrimination against women at a very early stage. Also in the sixteenth century, women were granted the right to inherit property of their husbands and fathers in the absence of sons. These things generally affected only Christian women since the Portuguese had no real hold on the Hindu village communities. They had influence on the Christians though via the well-established net of churches and parishes that was set up by the Catholic orders.
All those measures of course were designated to serve one purpose: Christianisation. Women were provided with incentives to convert to Christianity. Furthermore Hindu widows and orphans were only entitled to the property of their ancestry in case they converted. Regarding the *New Conquests*, Hindu women were only permitted at the end of the 19th century to opt out of their Hindu law and seek justice under Portuguese law.

Later on the Portuguese civil code of 1867, which was extended to the colonies in 1869,

[15] Even though it must be noted that the ban on sat'i was temporarily lifted in the 16th century due to pressure from the population.

[16] Albeit there are no records on female infanticides in Goa, it is reported that daughters on average were treated worse than sons. They were breastfed for shorter periods and were also later on poorer fed. Male children were also provided with better medical care as well since they were regarded as an asset according to social customs. Thus the disease survival rate of boys was reported to be higher [Da Silva Gracias, 1994].

[17] n.b.: the British, on the other hand, typically discouraged marriage (or even interference) with the local populace.

regulated all the above mentioned things more clearly and formally and further improved the position of (Christian) women. The code was uniform for all citizens, irrespective of caste or sex. It guaranteed equality of the sexes with respect to property, protected the interests of widows and it contained laws concerning the family. Hindus generally were subject to the so called *Codigo dos Usos e Costumes* (in the *New Conquests* it was promulgated in 1853). When Portugal became a republic in 1910, the civil code was further expanded, e.g. by the possibility to annul or divorce a marriage.

Albeit socio-religious customs prevented women to a large extent from taking advantage of all these legislations, it cannot be stressed enough that their sheer existence and partial exertion were a major achievement, brought about at a comparatively very early stage in history.

2.4.1 The Dowry System

Dowry, as practised in India, is an amount of money, property, ornaments and other gifts that the family of the bride has to pay on the occasion of the marriage to the bridegroom or his family. Though dowry is prohibited by law in India, it is widely practised. Since most marriages in India are arranged by the families, dowry is decisive in finding a suitable bridegroom. Moreover, in practice, dowry is not a onetime payment, but the husband and his family continue to expect and demand money and other gifts from the wife's family, the denial of which leads to a lot of tension and violence. Consequences of dowry include female foeticide, discrimination against girls, late marriages for many girls, unsuitable matches for girls, lowering of women's status, breakdown of marriages, increase in mental diseases, increased rate of suicide and impoverishment in poor and middle class families and so on.
It could be argued that the sheer religious differences across the *Old*- and *New-Conquests* would explain the differences in female literacy rates and the educational gender gap that I observe across the old Border which doesn't exist anymore. The implicit driver for the preference for males being the dowry system. This can be rejected since the Christians, albeit partly adopting Portuguese culture, still kept many ancient Hindu practices. One of them being the dowry system [see e.g. Ifeka, 1989; Hickman, 2007; Kumar, Kiran, and Gone, 2013].

2.4.2 Education of Women

It has to be mentioned that in early times the education of women was generally neglected. Only some girls of upper strata were exposed to home schooling. If females benefited at all at early stages from Portuguese and missionary's interventions from an educational point of view, then the historical narrative again strongly suggests that this exclusively was the case for the *Old Conquests*. In this process the Catholic church played a significant role, although the precise motives and channels being somewhat opaque, scholars agree on the importance of the church in this respect [Neill, 1985; Xavier, 1993; Da Silva Gracias, 1996; Emma Maria, 2002]

2.4.3 Women and the Church

As already mentioned, Catholic missionaries played a significant role in the process of raising the status of women in the Goan society. Numerous letters and decrees display the concern of curch officials towards the plight of women. Through their ability to influence the government substantially, several achievements have to be ascribed to their efforts. Especially when it

comes to the ban of polygamy and the encouragement of widow remarriages.
Around the turn towards the 17th century, the curch started to set up several homes that served as shelters for women. At around the same time the first nunnery was started in the city of Goa Colonial government forbade the *satì*. They also passed laws to allow Hindu widows to marry again. They also requested the king to impose strict punishments on those who violated the rules of monogamy.

Goan women in general could (officially) opt out of an unhappy or undesirable marriage from 1910 on [Da Silva Gracias, 1996, p. 144].

> "*It must be admitted that unlike other religions, Christianity gave their women due respect and position which was an impetus for others to reform their society. The Christian were neither treated as chattels nor were they treated as properties. Whey were not treated as as door-mats but as human beings with rights and privileges. They enjoyed the proprietary rights, they were consulted in all matters of importance, they attended all functions and so on. The position which the Christian women enjoyed was in fact a matter for envy for non Christian women. Even the Hindu reformists in the later years, became fervent adcovates to critisice the disparaged position of the Hindu women. The Portuguese themselves were instrumental to improve the position of non-Christian women through several state laws in the 17th and 18th centuries.*" [Xavier, 1993]

> "*Christian missionaries were the first to put women on the agenda of Indian social reform and drew attention to the low social status of women. They felt that education alone wouldc help them to oppose things like sati, female infanticide, child marriage, and enforced widowhood.*" [Basu, 1993]

> "*Historically, the Portuguese have displayed a deep concern for women's rights and their egalitarian sense has reflected itself in the people of Goa. One can see this in the equalaccess to education and the resultant freedom to choose a full time profession, the increase in the age of marriage and the Portuguese Uniform Civil Code, later called the Uniform Civil Code which gives the daughter an equal right to her father's inheritance and property.*" [D'Costa, 2007]

> "*The State and the Curch played a significant role in upgrading their [the Women's] position. The Portuguese rule seemed to have made a difference to the status of women. As a result the conditions of women in Goa were far better than their counterparts elsewhere in India.*" [Da Silva Gracias, 1996]

Generally speaking, scholars agree that there are several root conditions that need to be met in order to begin with the empowerment of women in India: the abandonment of polygamy, early childhood marriage, sat'i and the permittance for women to possess property and be able to remarry as widows [see e.g. Khanday, Shah, Mir, and Rasool, 2015]. All of these conditions have been met for several generations within the *Old Conquests* of Goa.

2.5 Goan Catch-up & Development after 1961

Since the Portuguese government was not willing to give up on Goa, the Indian government decided to take it over by force. Operation Vijay, the military takeover of Goa by the Indian army, can be described as a success in the sense that the number of casualities was very low

and the later transition towards ”normality” took place comparatively smooth. Especially when one takes into account that the territory was ruled for 450 years by a European power. Fact is, that Goa back then can be described as one of the poorest regions in India: the number of schools was very low, and education was tailored towards an elite that was close to the colonial government, only less than 5% of the villages had electricity, and so forth.
Once the Indians took over, transfers in form of infrastructural investments were flowing from the government in Delhi. This was especially true for the construction of primary and secondary schools [Varde, 1977; Malvankar, 2015]. As the later census data suggests, these investments were highly successful and were one of the reasons for Goa's immense catchup. One thing that has to be thought of here is, that this is not the usual way how things unfold when we look back at the history of development aid and infrastructural investments in structurally weak areas. The effect of decades of foreign aid in numerous countries across all continents has to be described as mild at best. This holds also true for different parts of India, mostly in its North. So why could Goa be so different?
This study shows that the aggregates in terms of educational improvement (as measured by literacy rates, both for males and females) were in the beginning driven by the *Old Conquests* of Goa, and that the other parts only started to catch-up in the 1980's and 1990's. I try to convince the reader that one of the main reasons for this phenomenon lies in Goa's differential history. Due to the penetration of those four districts that are ”Missionary Goa” by monks and priests for centuries, even in remote villages, people became familiar with the concept of structured education and potentially also saw the returns to it, even without being able to participate by themselves. This is what I would describe as a ”taste for education”, that becomes important once schools become widely available and accessible for children of all backgrounds. Out of a similar reasoning this also differentially contributed to female education. Since the position of females were alleviated by a bundle of ”rules” early on (regarding satì, polygamy, childhood marriage) and these diffused intergenerationally throughout time, families were much more likely to also send their daughters to schools, once the Indian government made them available.
It should be mentioned here already that the government did not preferentially treat the *Old Conquests*. This is crucial to point out, otherwise one might conclude that the effect that I am describing stems from differential infrastructural investments that eventually made those parts more prosperous. Rather the opposite was true: it was obviously known that the *New Conquests* did worse on average, especially the parts in the hinterland, thus it was aimed to harmonise the regions and the *New Conquests* received more rather than less investments from the time on when the Maharashtrawadi Gomantak Party, representing lower castes and social classes, came into power in 1962.

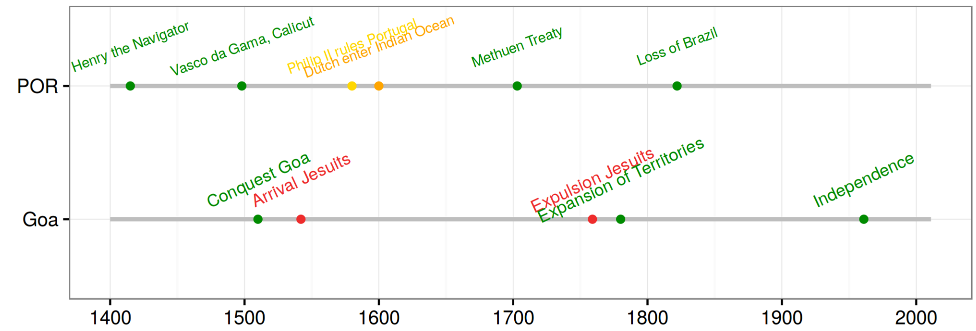


Figure 3: Cornerstones in the history of the Portuguese thalassocracy.

3 Data

The main arguments of the present study are based on 1991 and 2011 Indian census data. The respective Village Census Abstracts contain the number of males/females for each village/town, whether they are literate, and if they participate in the labour force. This allows me also to compute a sex-ratio for each unit of observation, not by age-cohorts though. The religious composition cannot be observed for villages, and is only available at the (statutory) town, taluka (sub-district), and state level.
From the so called District Census Handbook (DHCB) I obtain village/town-level data on infrastructure and the like. This I use as control variables in some specifications (number of doctors and nurses, number of primary schools).
In the sub-district of Tiswadi, I separate the 6 non-urban wards of the capital city Panaji. These are reported together with the city proper for administrative reasons but are clearly separate village entitities and are also reported as such in the census[18].

My sample then consists of 70 towns and 335 villages in contemporary Goa. In a next step I geolocalise all villages and towns via a tool called India Place Finder [Mizushima Laboratory, 2013]. The town of Murda is not listed as an entity in VillageFinder. After several crosschecks it was verified that it was excluded from the administration of the captial Panaji only from the last census period. It is closer to Panaji than most of the above mentioned non-urban wards that I treat separately, which reassures me in my decision to do so. I thus assigned longitude and latitude by hand as obtained from OpenstreetMap. The same was done for the abovementioned 6 rural wards of Panaji.

For robustness checks I employ nightlight data from the DMSP-OLS program which have been shown to correlate highly with regional economic activity [Croft, 1978; Elvidge, Baugh,

[18]The names are Panelim, Morambi-o-Grande, Renovadi, Morambi-O-Pequeno, Cujira, Taleigao, Durgawadi (these are the wards 31 to 37, respectively, of Panaji in the 2011 census). See the picture snapshot that I made from the census handbook. The results are not at all contingent on this step, as these villages are around 30km away from the RD border and thus not included in any of the RD specifications. Earlier versions of the paper had a slightly smaller sample size because I did not make this distinction, the conclusions and point estimates were the same though. As a unique identifier for each of those observations I added the so-called "Rural MDDS CODE" that Census India provides for these types of rural wards.

Kihn, Kroehl, and Davis, 1997; Chen and Nordhaus, 2010; Henderson, Storeygard, and Weil, 2012]. Geographic information (i.e. the respective geopackages/GeoJSON's/shapefiles) on the location of borders and (sub-)districts has been obtained from the Data{Meet} [2016] which was cross-checked with exactly georeferenced maps and numerous historical sources, all of which are cited throughout the paper. An important note of caution for the implementation of Spatial RDD's in general is, that the widely used data on sub-national units and their border from the GADM project [GADM, 2012] or from the widely known (and very good) website naturalearth.com are not precise enough for a full fledged GRD design.

Historical information on the location of churches and parishes in the 18th century (before the expulsion of the missionaries) come partly from Borges [1994] and Gomes [2003]. A map of the location of Jesuit "sites" in Salcete and contemporary Mormugao & Tiswadi from Borges [1994].

For descriptive statistics (e.g. shares of religious groups), I use data from the so called A-Series from the 2011 census. I unfortunately cannot observe the literacy broken down for each religious group for villages, only for towns. The latter information is then used for the non-parametric statistical tests in table 12.

3.1 Descriptives & Average Effects

	Dependent variable:							
	lit. rate	lit. gap	sex ratio	fem. lit. rate	lit. rate	lit. gap	sex ratio	fem. lit. rate
	[2011] Full Sample				[2011] Within 5km of RD border			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{I}$(Village in Old Goa)	0.062***	−0.050***	0.054***	0.087***	0.031***	−0.024***	0.070***	0.043***
	(0.006)	(0.005)	(0.012)	(0.007)	(0.007)	(0.007)	(0.016)	(0.010)
Constant	0.755***	0.105***	0.974***	0.702***	0.783***	0.084***	0.986***	0.741***
	(0.004)	(0.003)	(0.007)	(0.004)	(0.005)	(0.005)	(0.011)	(0.007)
Observations	402	402	402	402	137	137	137	137
R^2	0.225	0.220	0.047	0.307	0.111	0.079	0.128	0.133
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 2: Variables of interest and their differences between "Missionary Goa" and the *New Conquests* in 2011.

Table 3 and Table 2 provide the descriptive statistics for the main outcome variables of interest in 1991 and 2011. These are just univariate regressions without any controls, reporting the averages inside and outside of the cutoff, which is the border betweew the *New Conquests* and the *Old Conquests* that was abandoned 250 years ago. As there are no other control variables on the right hand side, the constant shows the average in the "non-treated" areas. The units of observations here are already the towns and villages that are then also used for the RDD; 141 of which are in "Missionary Goa" and 246 are located in the *New Conquests*. The right part of the tables also report the averages within an arbitrary 5km bandwidth around the border in order to also give a feeling what (part of) the subset of the data that later is used when we move towards causality looks like. As described in detail already before, the latter were not exposed to the Christian monks and priests and women were formally made equal to men only with the Portuguese civil code of 1871. Given that

	Dependent variable:							
	lit. rate	lit. gap	sex ratio	fem. lit. rate	lit. rate	lit. gap	sex ratio	fem. lit. rate
	[1991] Full Sample				[1991] Within 5km of RD border			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{I}$(Village in Old Goa)	0.115***	−0.064***	0.048***	0.149***	0.072***	−0.050***	0.075***	0.100***
	(0.013)	(0.007)	(0.014)	(0.014)	(0.021)	(0.011)	(0.023)	(0.022)
Constant	0.571***	0.191***	0.973***	0.473***	0.607***	0.185***	0.974***	0.513***
	(0.008)	(0.004)	(0.008)	(0.009)	(0.014)	(0.007)	(0.016)	(0.015)
Observations	358	358	358	358	123	123	123	123
R^2	0.177	0.193	0.033	0.237	0.092	0.153	0.083	0.147
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 3: Variables of interest and their differences between "Missionary Goa" and the *New Conquests* in 1991.

Goa up until today is the only state in all of India with such a legal institution that equalises all people - irrespective of gender or religion - the late 19th century could still be described as remarkably early.

In general we can observe that all variables significantly differ by a few percentage points between the two areas. Column 1 in Table 2 shows that the villages in the *Old Conquests* have on average a 6 percentage points higher literacy rate in 2011. The intercept reveals that the average outside is around 75%.
Looking at the literacy gap, i.e. the female literacy rate subtracted from the male literacy rate, the observed gender discrepancy is approximately half the size in "Missionary Goa". Those parts of Goa also have on average 54 females per 1000 males more as. In the appendix some further descriptive tables are provided.

3.2 Disaggregate Religious Estimates

As the religious composition is not reported by Census India[19], and religion clearly plays a crucial role in this study, I am reporting religious classifications inferred from surnames based on an algorithm by Susewind [2015]. The algorithm was developed for all of India, and it seems to perform very well in Goa. Only when it comes to classifying the gender from the given name it performs poorly, mostly for Muslims and Hindus, for Christians it seems to do well also in this respect. It delivers estimates of unrealistically disproportionate gender ratios (sometimes 70 or 80 percent for one gender) in areas where I know from the censuses that it should be fairly balanced. I thus discard the gender classifications and only use the religious groups. Based on the surnames that are observable for all voters at the polling-booth level for the 2014 elections, said algorithm infers whether a person is Christian, Hindu, or Muslim [Susewind, 2017].
The current population comprises about 1.2 million people of whom 65 percent are Hindus, 27 percent Christians, and 6 percent Muslims according to the 2011 Indian Census. The figures in Table 4 and Figure 4 are perfectly in line with these aggregate numbers (note that

[19]Only for statutory towns. This data will be used in Section 6 to look at the differences in literacy gaps and sex-ratios across the three major religious groups.

the polling station sizes are not constant across space, so the numbers are not expected to add up exactly). The map visualises the religious spatial distribution, and demonstrates that the identified RD border is actually meaningful. The religious shares at the sub-district level inferred from the 2011 census are reported in the Appendix.

	Dependent variable:					
	Christian	Hindu	Muslim	Christian	Hindu	Muslim
		Full Sample			Within 5km of RD border	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{I}$(Village in Old Goa)	19.400***	−22.251***	0.738**	21.701***	−25.030***	0.161
	(1.011)	(1.034)	(0.358)	(1.668)	(1.768)	(0.516)
Constant	9.130***	56.460***	3.927***	10.908***	57.404***	3.247***
	(0.757)	(0.774)	(0.268)	(1.209)	(1.281)	(0.374)
Observations	1,613	1,613	1,613	653	653	653
R^2	0.186	0.223	0.003	0.206	0.235	0.0001
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4: Religious shares inferred from the universe of surnames of people eligible to vote across all 1613 polling stations in Goa. (2014 elections)

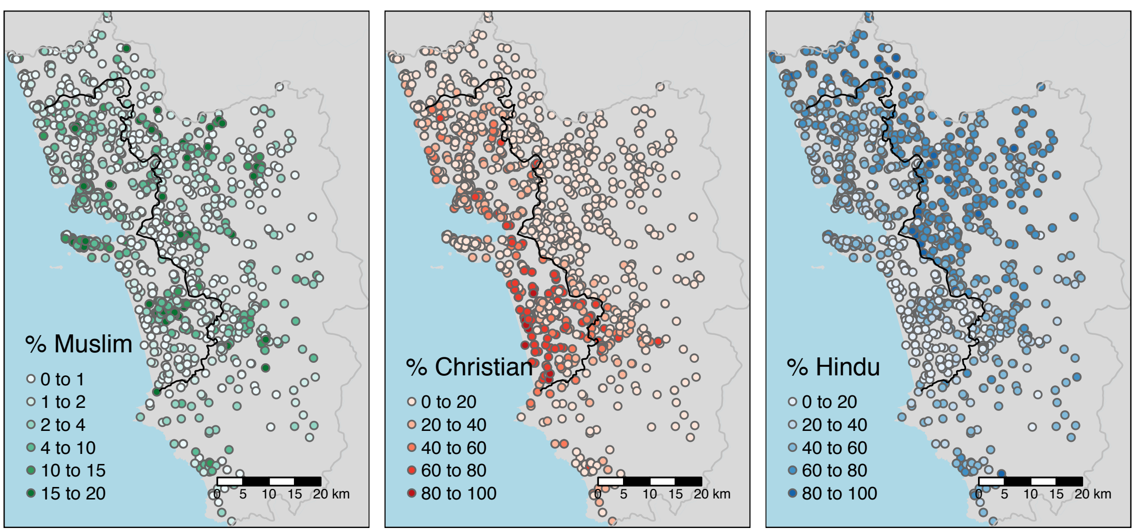


Figure 4: Estimates from surnames for all polling stations in Goa that confirm the historical narrative of a sharp cultural discontinuity at my identified RD cutoff. The "historical" Christian and Hindu groups align alongside this border. Muslims that have been migrating in only since a few decades are in the urban centres.

4 Econometric Specifications

Generally speaking, the living standards in Goa are higher than in its neighbouring states[20]. This average "Goa-effect" is e.g. observable in higher literacy rates for both males and females. Digging a bit deeper into the data, it is also easy to identify average differences between the *Old Conquests* and the *New Conquests*, suggesting that there is also something like a "within-Goa-effect". The historical narrative, which was outlined in detail in Chapter 2 and summarised in Table 1, strongly suggests that these differences are due to the different degrees of Portuguese colonial penetration and missionary influence.

In order to to move towards causality, the historical quasi-experiment is utilised to apply a specific type of a *Regression Discontinuity Design* (RDD): a spatial- or geographic-RDD (GRDD). In general, the "classical" RDD type of research design has tremendously gained popularity throughout the last years [as e.g. summarised by Lee and Lemieux, 2010]. This is largely due to the potentially high credibility that is attributed to it. One of the strengths of RD designs is that a known treatment assignment rule exists and is enforced.

The spatial RDD differs from a standard RDD when it comes to the cutoff that splits units into treatment- and control-groups. It is argued that such a geographic cutoff, usually a border of any kind, occurs in an as-if random fashion. GRD designs, such as the one in the present study, fall short of some of the premises of an ideal RD design. Since geo-referenced individual data are virtually impossible to obtain, inter alia due to privacy reasons, there are tipically very little observations arbitrarily close to the boundary in these type of designs. Thus spatial RDDs in general rely on "ignorability assumptions" within a certain band around the border. Thus many, if not all, will fall short of the RD ideals that were outlined in the "classical" RD literature. Nevertheless it has been demonstrated that such a set-up can lend itself especially well to provide convincing evidence about causal effects when carried out carefully by the researcher [Keele and Titiunik, 2015, 2016; Keele, Lorch, Passarella, Small, and Titiunik, 2017]. In this study I will make use of both parametric and nonparametric spatial RD specifications in order to show robustness and point out differences and potential pitfalls. The tedious geographic tasks that are necessary and the estimations are carried out in R with the `SpatialRDD` package [Lehner, 2019]. Furthermore, I will use spatial interpolation techniques in visualise the spatial dimension of the effect

For identification it has to be guaranteed that all relevant factors besides treatment vary smoothly at the RD border, otherwise we cannot be sure that villages on the opposite side are appropriate counterfactuals. This will be demonstrated in the following section. The reason why in this study an RD is the ideal tool to isolate causal effects is the geographical heterogneity of Goa in the East-West direction in terms of market access. The villages (and towns) on the coast are on average more prosperous due to tourism and trade, whereas the villages in the hinterland suffer from their remoteness. The villages alongside the abandoned border that I use as my RD identifcation device on the other hand are highly compareable in fundamentals as the balancing checks in Table 5 show. By just looking at villages very close to each other on either side of this border, jumps in terms of literacy and sex-ratios cannot be explained by observables and thus need further investigation. As was already indicated,

[20]Summary statistics are reported in the Appendix. I want to emphasise, though, that there is *no discontinuity* to be found from Goa towards Maharashtra in the North and Karanataka in the South, even though these were binding borders for more than 200 years. Further highliting the peculiarity of the within-Goa discontinuity.

	Dependent variable:		
	Area	Village Population	Households
	(1)	(2)	(3)
Village in Old Goa	−128.319	−96.405	22.835
	(131.624)	(728.531)	(171.350)
Constant	671.759***	3,559.673***	817.653***
	(89.337)	(491.721)	(115.653)
Observations	89	90	90
R^2	0.011	0.0002	0.0002
Note:			*p<0.1; **p<0.05; ***p<0.01
			These results are based on a 3km bandwidth.

Table 5: Balancing Checks showing the comparability of the villages alongside the cutoff: there are no jumps in population, household sizes, and the size of the villages.

I am tracing these discontinuities back to the long-term exposure of differential historical "fundamentals".

4.1 Balance Tests and Pre-Treatment

One important condition that needs to be satisfied when it comes to RD estimation in general is that before the actual "treatment" happened, there was no effect observable across the considered response variables. In my setting this turns out to be literally impossible since during the 15th century no data was recorded and detailed quantitative historical records in general are non-existent. I thus have to rely on historical narratives. The point is best made by simply presenting a quote from a Goan history book:

> "*Their country had never enjoyed the same unity that it has had since the Portuguese conquest in 1510. Before that date, the different districts belonged to different kings and different kingdoms, so that Goa, as we know it now, was never one country. Its various parts have been welded into one whole, [...]*" [Saldanha1957]

Due to the constant conflicts between the Hindu Vijanaghara empire and the Muslim Bahmani sultanate, the border changed several times in non-systematic ways within the territory of present day Goa. It is thus outruled that anything of significance happened right across or alongside the abandoned border that I consider for my GRDD design.

Table 5 shows that the villages I am juxtaposing on both sides of the border are highly comparable in size, population, and the number of households. It is thus outruled that any observed effects are spuriously driven by comparing big with small or urban with rural

villages. From the comparable household size it can also be inferred that the observations are approximately comparable when it comes to socio economic factors.
What is more, so-called features of first nature geography such as climate, rainfall or the quality of soil are equally comparable. A quantitative analysis is not reported due to the fact that there is no variation within the data. The border only stretches only around 40 km north-south, and those values are typically reported on gridded datasets with a larger scale, thus any estimation will be insignificant by construction.
When it comes to the balancedness of "fundamentals", other than outcome variables, the placebo checks from Table 9 give further evidence that there is no underlying discontinuity. The number of doctors, nurses, and primary schools per capita is virtually the same across the cutoff. This also demonstrates that there is no preferential treatment by the government for either side.

4.2 RD Estimates

We now turn towards the actual estimations which will be carried out in three different ways:

1. A spatial RDD in "naive" distances, ideally estimated non-parametrically through a local linear regression, identical to classic RD's but with distance to a cutoff as assignment variable (examples: [Michalopoulos and Papaioannou, 2014]).

2. A parametric specification that includes a polynomial in the X- & Y-coordinates to control for space. Typically estimated via OLS and reporting a single homogeneous treatment effect (put forward by Dell [2010]).

3. A nonparametric specification with two running variables, estimated nonparametrically at every point of the cutoff [labelled as GRD design by Keele and Titiunik, 2015]. This approach is then also extended by the finite-sample-minimax linear estimator approach put forward by Imbens and Wager [2019], which nicely complements the GRDD series visualisation that I develop in this paper, because it allows to draw a weighted conditional average treatment effect (CATE) alongside the cutoff on a map.

In the Economics literature, especially the sub-field that tries to link historical events to contemporary economic outcomes, the second specification in which the outcome variable is regressed on a dummy variable whether the unit was treated or not and a set of control variables (including polynomials in the X- and Y-coordinates that control for the position), seems to be the most popular one.
The methodological difficulties that arise in those spatial settings were clarified by Keele and Titiunik [2015]. They further show, building on the multidimensional RD literature [Papay, Willett, and Murname, 2011; Imbens and Zajonc, 2011], that their ideal GRD design leads to identification of local treatment effects at the cutoff under a two-dimensional continuity assumption [thus generalising Hahn, Todd, and van der Klaauw, 2001b]. They also point out the potential problems arising with GRD designs and other spatial RD set-ups in general. An essential difference to "classical" RD's with two forcing variables, individual points at the boundaries of GRDD's have a clear interpretation.
I second their critique and take their approach one step furhter and propose a way to consistently visualise the heterogenous treatment effect alongside any RD-border (this is what I call a GRDDseries). Furthermore I argue that all investigated outcome variables should also

be visualised by simple spatial interpolation techniques such as Kriging. This will give the reader a feeling of the spatial dimension of the data generating process. The two techniques combined, I assert, would allow us to identify false positives that any technique that reports one homogenous treatment effect for the full border might deliver. The detailed discussion of these approaches is postponed to the appendix and the documentation/vignette of the `SpatialRDD` **R** -package [Lehner, 2019].

4.2.1 Naive Distances

This simpler version of a spatial RD uses the euclidean distance of each observation to the boundary as its score for the estimation. This geographically "naive" measure of distance ignores how the units are spatially distributed since the shortest distance towards the border does not determine the exact location in the two-dimensional space [as e.g. pointed out in Keele and Titiunik, 2015, p. 137].
This specification allows to estimate an average effect which will mask considerable heterogeneity along the border. Yet, as we will see, this type of analysis is capable of delivering a quick intuition of what the data can potentially tell. Especially when it comes to the visualisation of this set-up with a "standard" RD-plot.

In this setting units around a narrow band around a border are assumed to be valid counterfactuals. This design may mask underlying heterogeneity and thus may not allow to evaluate the plausibility of the needed identification assumptions. A naive GRD design is still appropriate in some circumstances, e.g. when the boundary of interest is short and defines a homogeneous region. Technically it is only fully valid when treatment effects are constant at all boundary points $\mathbf{b} \varepsilon \mathcal{B}$.

This specification is measured non-parametrically via a local linear regression as is state-of-the art in "classical", ideally with data driven robust confidence intervals as suggested by Calonico, Cattaneo, and Titiunik [2014]. As already mentioned before, such estimations are capable of delivering information about an average effect alongside the full cutoff, which can be very misleading about the actual effect that is subject to study as I will demonstrate in 4.2.

	Left	Right
Number of Obs	141	246
Eff. Number of Obs	55	52
Order Loc Poly (p)	1	1
Order Bias (q)	2	2
BW Loc Poly (h)	3827.8325	3827.8325
BW Bias (b)	6672.3411	6672.3411
rho (h/b)	0.5737	0.5737

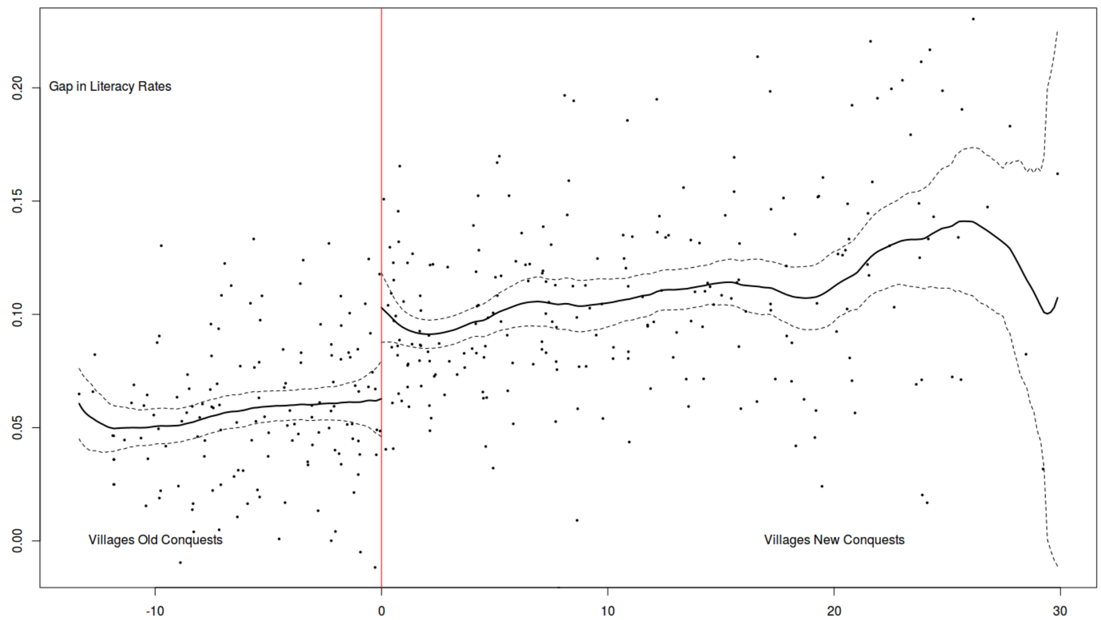


Figure 5: The gender gap in literacy rates visualised: all villages in "naive" distances to the cutoff (local linear regression as put forward by Hahn et al. [2001a], 95% confidence intervals)

	Coeff	Std. Err.	P>|z|	CI Lower	CI Upper
Conventional	0.0405	0.0143	0.0045	0.0126	0.0685
Bias-Corrected		0.0143	0.0019	0.0163	0.0722
Robust		0.0168	0.0085	0.0113	0.0772

Table 6: The results of the local linear regression show a discontinuous rise of around 4 percentage points in the literacy gap once the cutoff towards the "untreated" New Conquests is passed. This amounts to almost 100% of the average gender gap between the two parts of Goa (see e.g. Table 14).

4.2.2 Parametric Specifications

This specification, put forward by Dell [2010], is usually preferred within the Economics literature. It is well suited to report an average treatment effect alongside a given border. But as we will see later with the female labour force participation and male literacy rates, it could be misleading in some circumstances and produce false positives. As the more detailed visualisation in form of the GRDDseries shows, this is due to strong effects that occur locally and drive the average effect. Even masking points where we obtain statistically significant estimates with opposite sign (this is typically a sign that the effect is not robust, in addition to confidence intervals that are strongly overlapping with zero). The spatial interpolation exercises then visually confirm that there seems to be in fact no "action" in terms of this variable at the RD border.

$$y_i = \alpha + \beta\ MissionaryGoa_i + \sum_{s=i}^{S} \gamma_j\ SEGMENT_{ji} + \delta'_i \mathbf{X} + f(\mathsf{geolocation_i}) + \varepsilon_i, \qquad (1)$$

where y_{ib} is the outcome variable of interest for village i. $MissionaryGoa_i$ denotes a dummy variable that represents the "treatment status", equalling 1 if the village is inside the old, non-existing border and was exposed to early Portuguese colonial rule and missionary influence. $f(\mathsf{geolocation_i})$ represents the RD polynomial in X- and Y-coordinates, which is supposed to control for smooth functions of geographic location and is going to take on varying forms across different specifications. Finally, matrix $\mathbf{X}$ contains a set of control variables and $SEGMENT_{ji}$ represents a dummy, equalling 1 if village i has segment j as its closest segment. Regression just produces the weighted average over all segments. These are the equivalent to a set of boundary segment fixed effects and are meant to capture geographic heterogeneity and alleviate omitted variable problems by only exploiting within segment variation. This might be desirable in some settings, but the obvious drawback is that this approach masks the heterogeneity that potentially is capable of delivering deeper insights into the problem at hand.

Another drawback is that there are no clear formalised suggestions for the bandwidth selection in such an estimation procedure. For parametric procedures, on the other hand, a lot of guidance has been put forward [Imbens and Kalyanaraman, 2012; Calonico et al., 2014].

	Dependent variable (2011):					
	lit_gap	lit_rate	male_lit_ra	fem_lit_ra	fem_lab_pa	sex_ratio
	(1)	(2)	(3)	(4)	(5)	(6)
Village in Old Goa	-0.029^{***}	0.032^{**}	0.019^{*}	0.048^{***}	-0.061^{***}	0.082^{***}
	(0.010)	(0.012)	(0.011)	(0.015)	(0.018)	(0.025)
Segment FE	YES	YES	YES	YES	YES	YES
Poly. lat/long	YES	YES	YES	YES	YES	YES
Controls (Edu,Medi)	YES	YES	YES	YES	YES	YES
Observations	77	77	77	77	77	77
R^2	0.569	0.470	0.433	0.518	0.367	0.463
Adjusted R^2	0.435	0.306	0.258	0.368	0.170	0.297
Residual Std. Error (df = 58)	0.030	0.038	0.034	0.045	0.057	0.077
F Statistic (df = 18; 58)	4.256^{***}	2.858^{***}	2.466^{***}	3.457^{***}	1.866^{**}	2.782^{***}

Note: *p<0.1; **p<0.05; ***p<0.01
These results are based on a 3km bandwidth.

Table 7: Polynomial specifications with controls & border segment FEs, 2011 village level census. Results are robust towards all bandwidth extents.

Table 7 shows the average effects for all considered outcome variables from the parametric estimation of Equation 1. The estimations are robust towards all potential bandwidths which can also be inferred visually from the spatial interpolation plots in Section 4.2.5. The literacy gap is on average 3 percentage points lower, which as we are going to see with the parametric specifications, seems to be a slight underestimation of the effect. As can be seen from column

6, the number of women per 1000 men is higher by a number of 82, on average, in the *Old Conquests*. This table also shows significant coefficients for two outcomes that we are later going to flag as potential "false positives" after the investigation of spatial interpolation plots and nonparametric specifications: male literacy rate and the female labour force participation.

4.2.3 Nonparametric Specifications (2011), the Keele and Titiunik [2015] way

In this section I employ the technique proposed by Keele and Titiunik [2015] which measures the RD in space by considering a two-dimensional score as running variable. I push this one step further by carrying out the estimation for each point of the border and then report these in a so-called GRDDseries. The figures should be read as follows: the plotted lines represent each point on the discretised RD border (500m distances inbetween) and its point estimate including a 90% confidence interval. The colour indicates whether the estimate is significant at the 10% level or not. The version of the GRDDseries below displays the estimation of the identical specification, only with a forced selection radius of approximately 20km. This is just to increase the precision of the point estimates by forcing more observations into the estimation. Since a triangular kernel which weighs each observation in the regression by its distance to the respective boundarypoint is used, this should not be much of a concern. Because the bandwidth selection of the CTT algorithm is not very meaningful in some cases when one has to deal with sparse observations, I decide to drop each estimation where less than 10 observations were chosen on each side of the border. This is arbitrary, but estimations with 20+ observations arguable make more sense than ones with five or ten.
These resulting GRDDseries provide the reader with a transparent visualisation of the heterogeneous treatment effect across the whole border. All the results are obtained via local linear regressions [as in Hahn et al., 2001a], estimated with Calonico, Cattaneo, and Titiunik [2015].
Arguably one of the biggest advantages of such an approach - including the one in the next chapter following Imbens and Wager [2019] - is that there are no distributional assumptions on parameters needed. Thus error terms that potentially could exhibit spatial autocorrelation are non-existent, therefore alleviating concerns that have been recently raised about the "persistence literature" [Kelly, 2019].

Figure 6 shows the main specification under scrutiny. The signs of the point estimates essentially never change. Due to the small sample size the confidence intervals do overlap with zero quite often, but the majority of the point estimates at least gets significant once the algorithm is forced to select more observations. Thus it seems fair to conclude that the significance of the literacy gap is ensured when this very "tough" estimation method is applied. A bit of a "weaker picture" is exhibited by the series for the sex ratio, even though the sign of the point estimates never significantly changes (except for the fringe with very few observations). But looking at the extended bandwith of 20km, and especially in combination with the estimates from the next section, we would conclude that this constitutes a somewhat robust result.

The female labour force participation, shown in Figure 8, on the other hand shows more severe heterogeneity alongside the cutoff. It appears to be one of the more robust outcomes when looking at the average treatment effect in Table 7 with a quite substantial coefficient. But the point estimates in the GRDDseries are significant with different signs and thus I abstain from taking a very strong stand on this outcome variable.

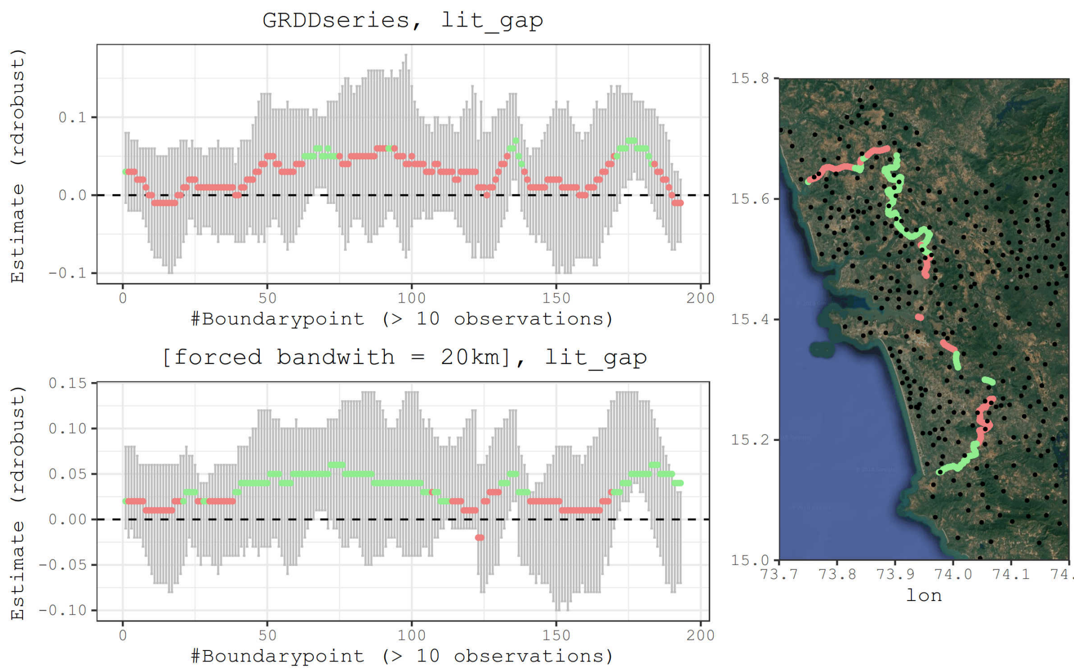


Figure 6: GRDDseries and spatial visualisation of the literacy gap in 2011

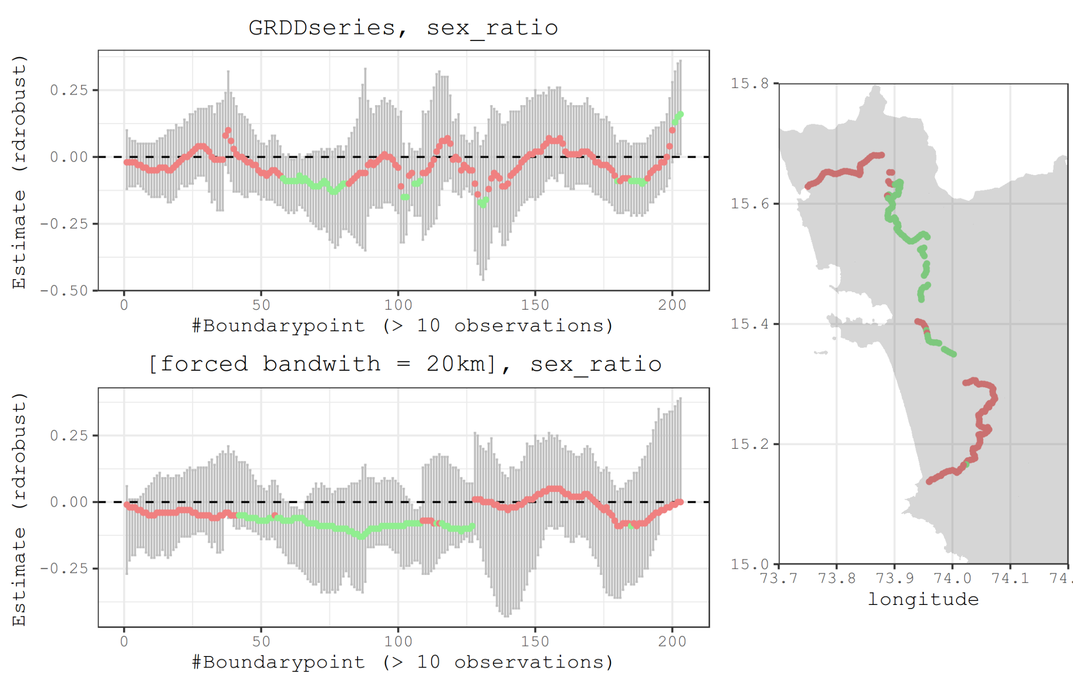


Figure 7: Heterogeneous treatment effect for the sex-ratio as measured by the number of females per 1000 males.

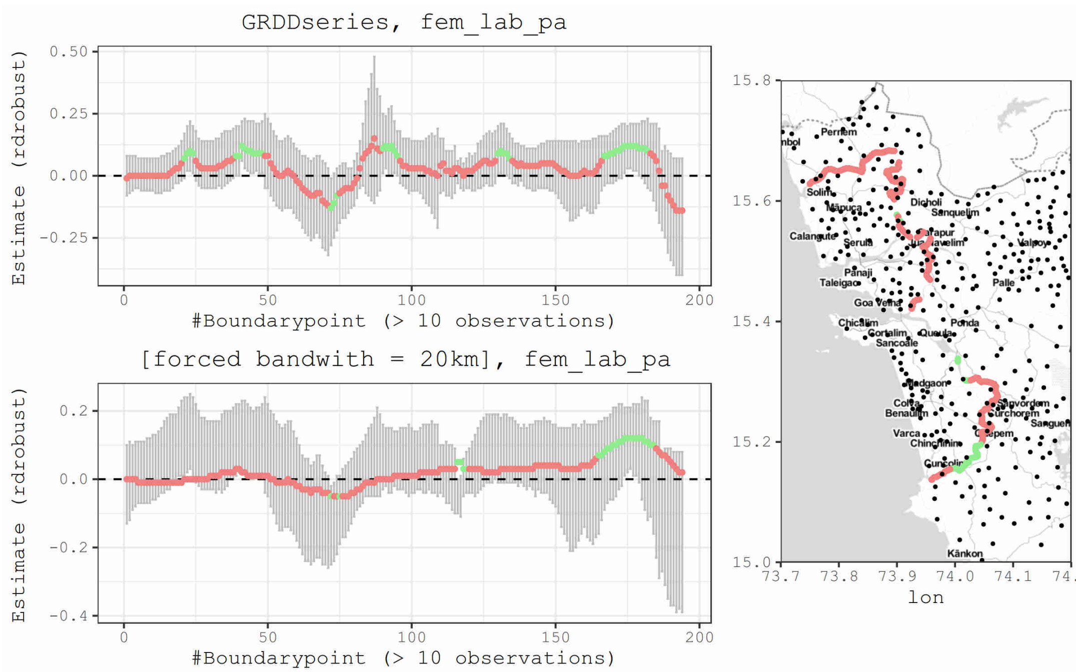


Figure 8: The seemingly non-robust heterogeneous treatment effect for female labour force participation.

4.2.4 Nonparametric Specifications (2011), the Imbens and Wager [2019] way

Instead of relying on estimation via local linear regression, Imbens and Wager [2019] use a minimax linear estimator to estimate both single- and multi-dimensional RDs. One of the niceties when it comes to spatial RDDs is that it allows to estimate a weighted average treatment effect (WATE). Their estimator is well defined regardless of the shape of the treatment region, thus immune to cases where the running variable is not continuous, which happens quite often in economic applications of spatial RDDs due to data sparsity. In such cases the identifying assumption for "classic" RDs is violated.

Table 8: The estimated weighted average treatment effects alongside the full RD cutoff by the "optimized approach" of Imbens and Wager [2019]. Their min-max estimation confirms the results from before: female labour force participation and male literacy seem to be less robust.

		outcome	point estimate	conf interval	max bias	sampling std err
1	WATE	sex_ratio	0.056	(0.011, 0.102)	0.012	0.020
2		lit_gap	-0.026	(-0.052, -0.001)	0.006	0.011
3		male_lit_rate	0.009	(-0.013, 0.031)	0.007	0.009
4		lit_rate	0.021	(0.006, 0.036)	0.004	0.007
5		fem_lit_rate	0.028	(0.004, 0.053)	0.006	0.011
6		fem_lab_part	-0.014	(-0.042, 0.014)	0.009	0.012

From Table 8 it can be seen that the results from before are more or less confirmed, lending

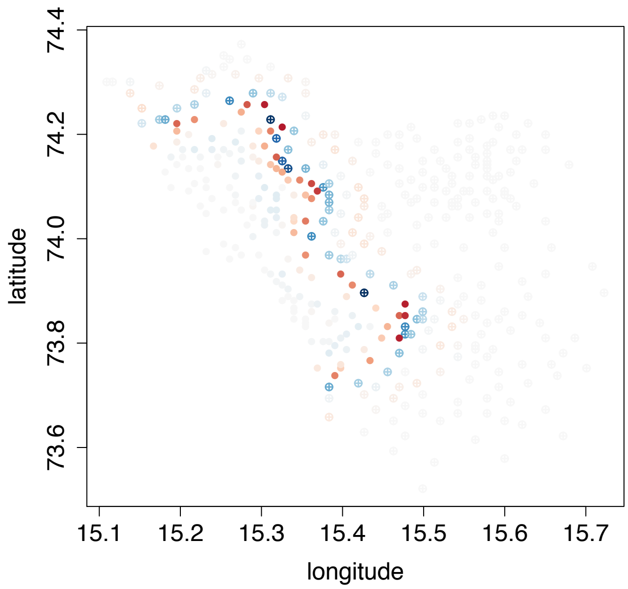


Figure 9: The Imbens and Wager [2019] approach allows to estimate a weighted effect which is clearly identifiable alongside my cutoff. (literacy gap in 2011, colour intensity resembles effect strength)

further credibility to the study design. The point estimates are also somewhat in the ballpark of the parametric specification from before. With the notable exceptions of the female labour force participation and the male literacy rate. Thus, similar to the GRDDseries approach from before, this WATE seemingly does a good job in uncovering "false positives" that the parametric specifications were masking.

4.2.5 Spatial Interpolation - Kriging

The patterns that were described above can be visually inferred from the spatially interpolated values shown in this section. This also somewhat makes looking at placebo borders obsolete since it can be seen in the plots where exactly in space the effect is happening. Figure 4.2.5 shows that the jump in the literacy gaps actually happens at the border that doesn't exist anymore since over 250 years and lends further credibility to the research design and identification device that is being applied in this study.
Figure 10 confirms results from the GRDDseries above that the female labour force participation doesn't appear to be a robust outcome at the cutoff.

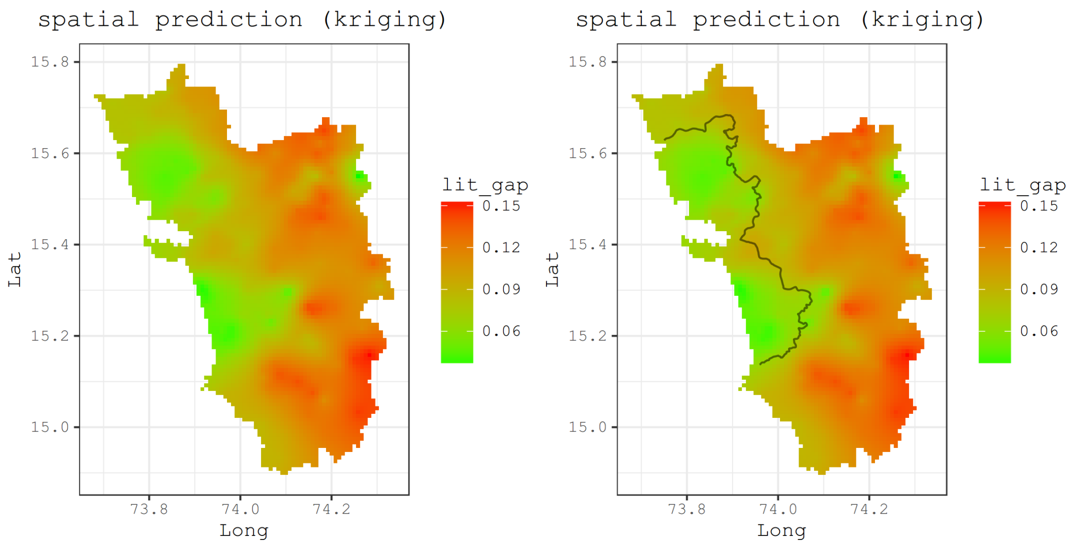
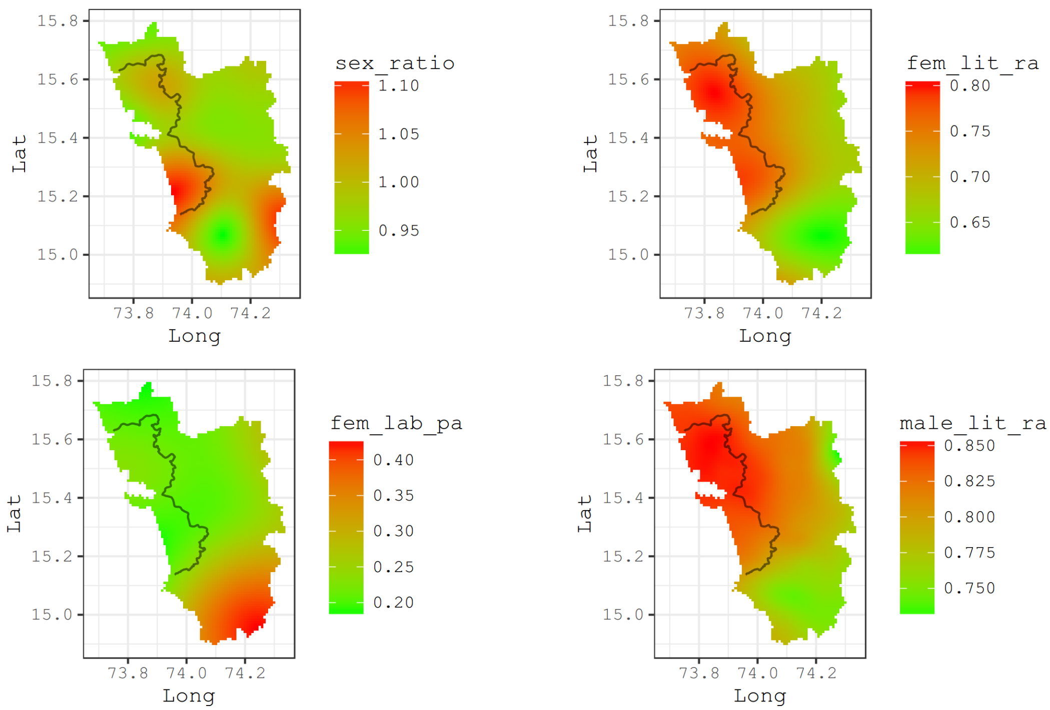


Figure 10: Spatial interpolation, village level 2011



4.3 RDD identification

The border was quasi-random

Removement of treatment was quasi-random

It is not driven by income

sorting right around the cutoff is not an issue

4.4 Falsification Tests

In this subsection I try to convince the reader of the robustness of my results and provide deeper analyses of whether the results so far can be interpreted as a genuine effect of the historical Portuguese presence on (female) education and sex-ratios in Goa. They aim to rule out that the identified effect reflects other things such as selective expansion, in which case the results may reflect pre-existing differences.

4.4.1 Different Bandwiths

All parametric specifications are robust towards different bandwidths ranging from 1 to 10km. This can also be visually inferred from the pictures of the spatial interpolation exercises. For the GRDDseries specifications this is not relevenat in any case, since the algorithm by Calonico et al. [2014] does data-driven bandwidth selection by itself.

4.4.2 Pseudo Borders

When the border is moved in both directions and the estimation carried out on those fictional lines, all of the relevant specifications loose significance. This can also be inferred visually from the spatially interpolated pictures.

4.4.3 Compound Treatment Effects

A typical problem with RDs is that sometimes several differential treatments happen within the exact same geographic area. In may case this is not much of a concern, as the border is meaningless since two and a half centuries and virtually no other border that is meaningful in terms of public administration, schooling, etc. coincides with the presented RD cutoff. This can also be seen in the following section on placebo outcomes.

4.4.4 Intermediate Step: Colonial Census of 1851

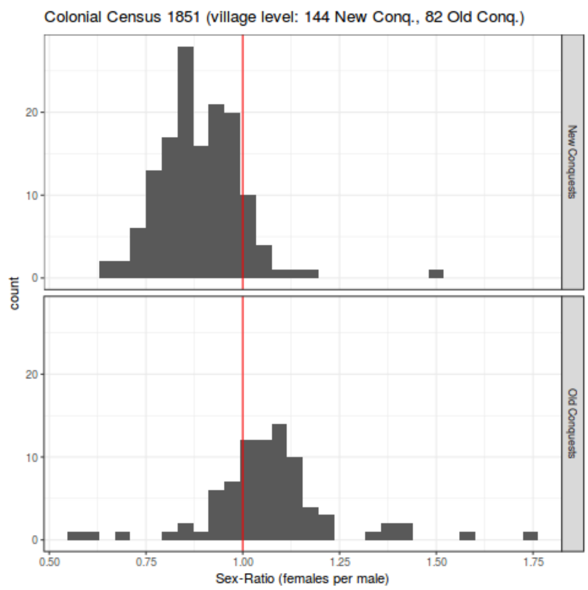


Figure 11: Sex-ratios at the village level in 1851.

One might argue that the observed effects are an artefact of something that happened after the Portuguese had left, or was due to something that was not caused by their presence. I thus analyse the recently digitised Portuguese colonial censuses starting from 1776 which were recently digitised by a group of economic historians [de Matos, 2013, 2016]. The first one which is reported on a village level and is thus compatible with how I carry out the analysis is from 1851. For all of those data-points I know in which district they are, and thus whether they are within the *Old-* or *New Conquests*. The exact location of the villages is almost impossible to obtain since the names changed multiple times over the years and some villages were merged with each other. An RDD on this dataset is thus not feasible. But as figure 11 demonstrates, in order to make the point it suffices to report a simple descriptive statistic. It plots the histogram of the sex-ratio for each village and demonstrates firmly that the ones in the "treated" districts are far off from the ratios in the other villages. It is thus safe to argue that the effect I was talking about was already in place before the Indian government took over and even before Portugal moved away from being a monarchy. Furthermore it has to be noted that this should be seen as something like a lower bound since the older parts of Goa back then were already attracting labour migrants which are typically male. It is true that the numerous servants were mostly women, but to the best of my knowledge only a minority of them was recorded in the cencuses. On top of that, the Portuguese a substantial amount of soldiers for Goa's defense. Those were almost exclusively stationed close to the capital in the *Old Conquests* and obviously all of the were men, thus dragging down the sex-ratios.

4.5 Placebo Outcomes

Typically one part of any study employing Regression Discontinuity Designs has to show that there is no jump in other variables across the cutoff that might potentially be the drivers behind the observed discontinuities in outcomes. As the information provided by the Indian census reported at the level of my units of observation, the village and town level, is very sparse - especially for the former - I can only rely on the three variables that are shown in Table 9. It is shown that there are no jumps when it comes to the availability of education (as measured by primary school since I proxy for education by literacy rates) and the provision of medical coverage, both governmental and non-governmental.

Table 9: Placebo outcomes with the polynomial in longitude/latitude specification

	Dependent variable:		
	PrimSchools	nonGovmedicoverage	medicoverage
	(1)	(2)	(3)
Village in Old Goa	−0.179	−0.744	−0.107
	(0.273)	(0.862)	(0.503)
Segment FE	YES	YES	YES
Poly. lat/long	YES	YES	YES
Observations	77	77	77
R^2	0.114	0.122	0.071
Adjusted R^2	−0.021	−0.010	−0.070
Residual Std. Error (df = 66)	0.990	3.123	1.823
F Statistic (df = 10; 66)	0.845	0.921	0.506

Note: *p<0.1; **p<0.05; ***p<0.01
These results are based on a 3km bandwidth.

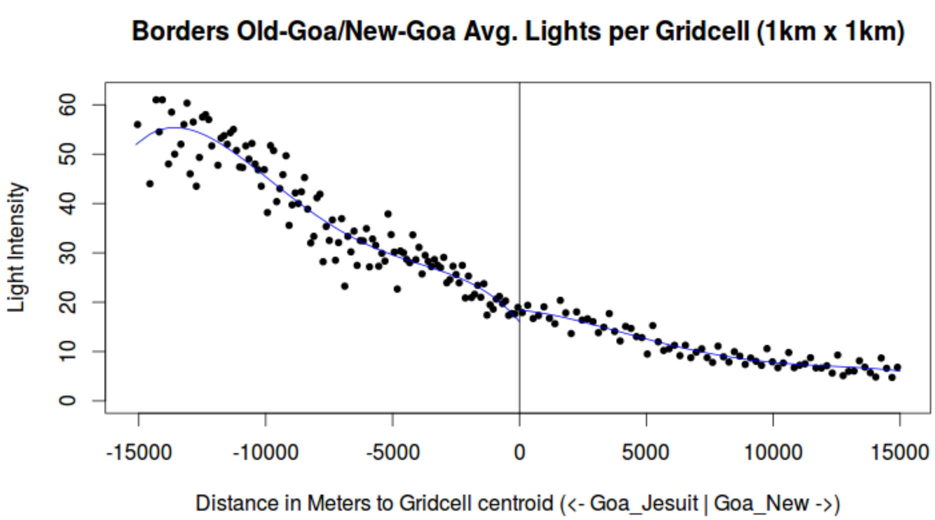


Figure 12: The effects do not seem to be driven by economic activity

5 Documenting the evolution from 1991 to 2011

The exercise described above is repeated in the same manner on the census from 1991, which is the first one after Goa became a full state of India. Before it was organised as a so-called Union Territory together with the other former Portuguese colonies Daman & Diu and the censuses were reported differently. Only from 1991 the reporting was carried out on the village level that I need for my RD estimations.

The estimations reveal that the observed jumps were around two times the size in 1991 as compared to 2011, which indicates that there was a convergence process taking place that seemingly harmonised the regions within Goa. This process, though, is essentially restricted to educational outcomes but does not apply to the sex-ratio bias. This gives strong support for culture being the strong driver behind this effect. It seems that, conditional on having a girl, it is possible to convince families to educate them and governmental efforts appear to have had an effect. These efforts it seems do not have an effect whatsoever when it comes to the skewed sex-ratios.

5.1 Average effects at cutoff in 1991

Table 10 again displays the RDD specification put forward by Dell [2010], including control variables and border segment fixed effects. It can be seen that the gaps in literacy rates and the gender gape were essentially twice as big, documenting the catch-up process in terms of education that the *New Conquests* underwent. The estimates suggest that the literacy gap is 8 percentage points lower in the "treated" areas. They also on average have around 80 women more for every 1000 men.

	Dependent variable:				
	lit_gap91	lit_rate91	fem_lit_ra91	fem_lab_pa91	sex_ratio91
	(1)	(2)	(3)	(4)	(5)
Village in Old Goa	−0.079***	0.073***	0.116***	−0.046**	0.079**
	(0.021)	(0.026)	(0.032)	(0.019)	(0.034)
Segment FE	YES	YES	YES	YES	YES
Poly. lat/long	YES	YES	YES	YES	YES
Controls (Edu,Medi)	YES	YES	YES	YES	YES
Observations	70	70	70	70	70
R^2	0.391	0.576	0.537	0.347	0.275
Adjusted R^2	0.176	0.427	0.374	0.117	0.019
Residual Std. Error (df = 51)	0.062	0.077	0.096	0.056	0.102
F Statistic (df = 18; 51)	1.818**	3.851***	3.290***	1.506	1.075

Note: *p<0.1; **p<0.05; ***p<0.01
These results are based on a 3km bandwidth.

Table 10: Polynomial specifications with controls & border segment FEs (ATE in 1991)

5.1.1 Nonparametric specifications (1991)

The GRDDseries confirm the average effects from Table 10 above. Figure 13 and Figure 5.1.1 show considerable heterogeneity but the point estimates never significantly change their sign and thus appear to "survive" this very demanding specification.

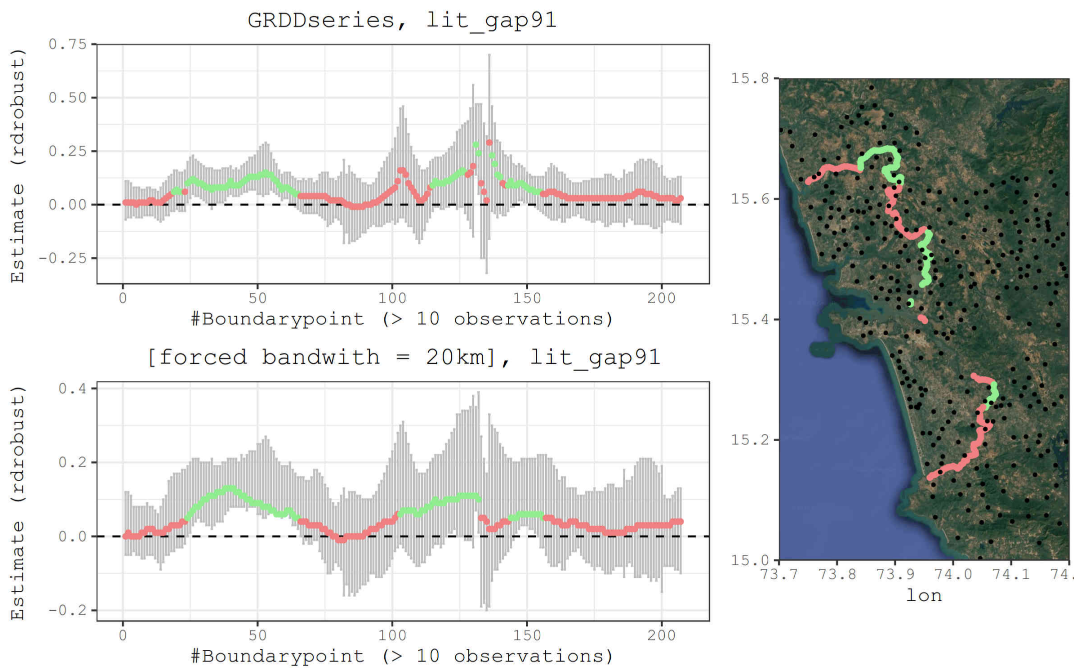
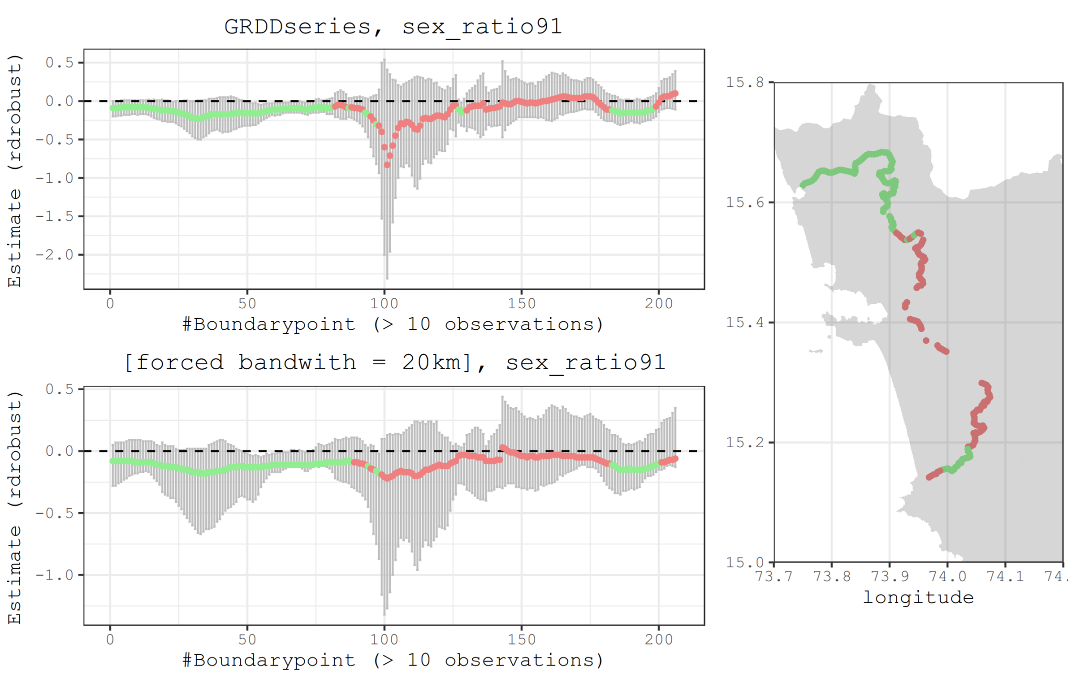
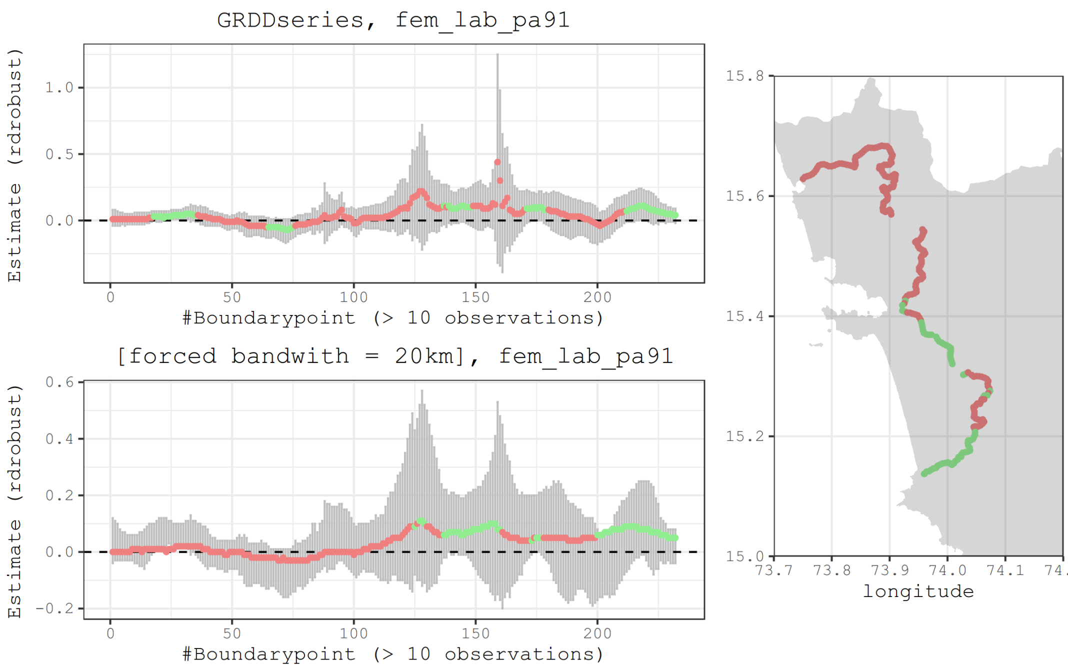


Figure 13: GRDDseries and spatial visualisation of the literacy gap in 1991



When it comes to the female labour force participation the picture is similar than for 2011: the point estimates are significant with different signs and there is even more heterogeneity alongside the cutoff than 20 years later. I thus have to conclude that effect regarding this outcome variable is not as strong and consistent as with the other ones.

Figure 5.1.1 again displays the spatial interpolation for the outcome variables under



scrutiny and somewhat confirms what was shown already for their 2011 equivalents. The female literacy rate and the sex-ratio seem to display the expected jump at the cutoff whereas the male literacy rate and the female labour force participation seem to be inconclusive.

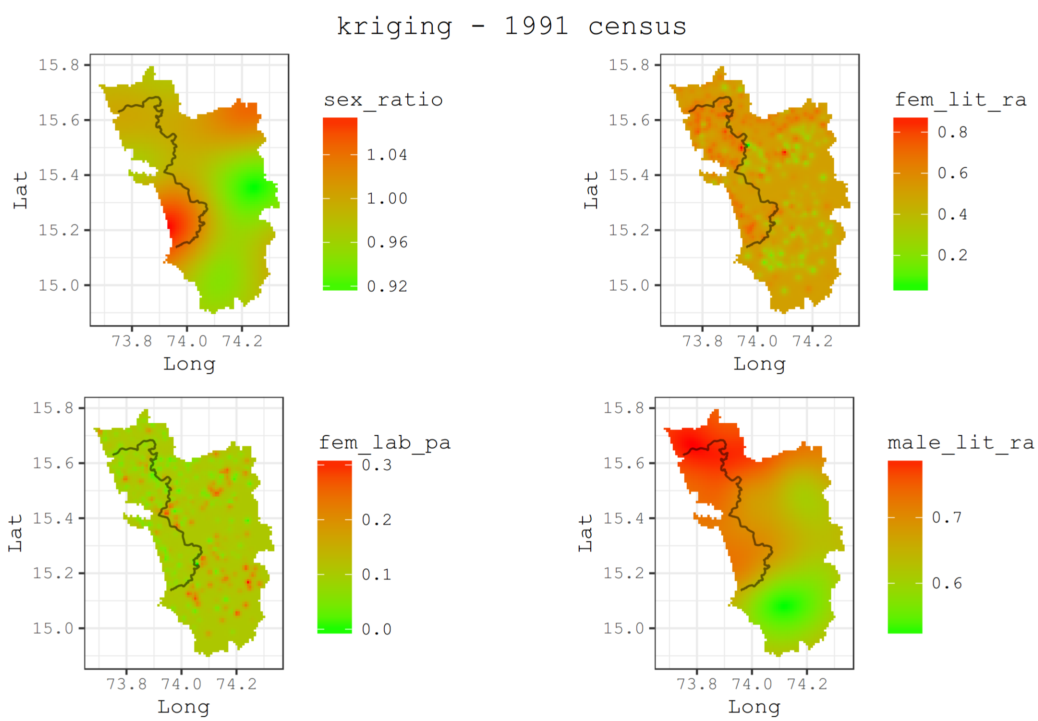


Table 11: Distance to Jesuit historical sites predicts literacy rates in 2011.

	Literacy rate 2011	
	Salcete	Salcete
	(1)	(2)
Distance to Jesuit Site	-0.010^{**}	-0.008^{*}
	(0.004)	(0.005)
Distance to Coast		-0.002
		(0.001)
Observations	62	62
R^2	0.081	0.113
Adjusted R^2	0.066	0.083
Residual Std. Error	0.035 (df = 60)	0.034 (df = 59)
Note:		*p<0.1; **p<0.05; ***p<0.01

6 The Persistent Effects: Mechanisms

Figure 14: Geo-localised map of historical Jesuit sites.

To further show the importance of the educational channel I show that there is an effect on the intensive margin. Figure 14 shows a geo-localised historical map of the accurate locations of historical Jesuit parishes [taken from Borges, 1994]. In Table 11 it is then shown that for every kilometre away from these sites, the literacy rate in villages decays by around 0.1 percentage points. As the typical distance of villages in the 2011 census to those sites is in the range of 5-10 kilometres, this is a sizeable effect which further hints at the historical imprint the structured education that was brought by the religious orders still has.

This is in line with studies on religious orders by Valencia Caicedo [2019] and Waldinger [2017], but also with stories on human capital persistence in general [e.g. Rocha, Ferraz, and Soares, 2017].

Table 12 shows the indicative evidence that, in addition to the documented effect across time, there seems to be also horizontal diffusion through space at play. Hindus and Muslims residing in the *Old Conquests* show significantly different literacy gaps than their respective counterparts in the *New Conquests* which did no receive the early ”treatment” regarding the role of women in society. This potentially hints at an explanation that the effect that I am describing is not entirely driven by religion, but by something cultural which has been ”imprinted” in those *Old Conquests*. Consistent with the deep-rooted cultural explanation for the male son preferences, I cannot find this pattern when it comes to the sex-ratio.
Unfortunately I cannot show this result on the full set of all villages that was used throughout the paper because Census India is reporting the religious composition only for statutory towns (supposedly due to privacy reasons as some of the smaller villages in the countryside e.g. only have one or two Muslim families).

Var	Group	mean(New)	mean(Old)	t-test	Wilcox rank-sum	Kruskal-Wallis
lit_gap	Muslim	0.0890	0.0494	**0.1026**	0.1911	0.1891
	Hindu	0.0775	0.0646	**0.0501**	0.0296	0.0292
	Christ	0.0444	0.0378	0.5477	0.2075	0.2054
fem_lit	Muslim	0.6778	0.6814	0.8415	0.7964	0.7919
	Hindu	0.7701	0.7758	0.4645	0.2579	0.2554
	Christ	0.8072	0.8243	0.3016	0.4253	0.4219
sex_ratio	Muslim	0.8825	0.8776	0.9111	0.5696	0.5656
	Hindu	0.9445	0.8870	0.1013	0.0150	0.0147
	Christ	1.0290	1.1318	0.0001	0.0000	0.0000

Table 12: The dependent variables broken down by religion (for the 44/26 statutory towns)

7 Conclusion

This paper has put forward a deep rooted explanation for the emergence of gender roles by examining the effects of the long-term presence of Portuguese colonisers in the Indian state of Goa. They, inter alia, significantly altered the position of women in society at a very early stage in history. In addition they interfered culturally by introducing what I call a ”taste for education”: their accompanying missionaries set up a network of churches, parishes with schools and even set up a college and brought the printing press. After the sharp downturn of the Portuguese thalassocracy in the 17th century, also the Goan economy was in a continual decline up until its liberation. Interventions by the late Salazar dictatorship to boost economic growth only masked deep structural problems and appeased the urban upper classes. When the Indian government took over in 1961, only a few villages had electricity and the number of schools was low. Once schools and a broader infrastructure became widely available throughout all of Goa, the catchup process, initially driven by iron-ore mining in remote areas and the port in Mormugao, began.

A more disaggregate analysis reveals that this process was experienced differently by parts of Goa that are referred to as the *Old Conquests*. These territories were colonised in a different period and experienced an entirely different colonial reign: Catholic missionaries spread the word of Christ, built schools within their network of parishes even in the most remote places, and thus induced something that I would describe as a ”taste for education”. A further cultural intervention in the 16th century regarded the position of women in society: sati, polygamy, and early childhood marriage were forbidden. Additionally females received property rights upon conversion to Christianity (in the 19th century women of all religions formally received property rights due to the Portuguese civil code which was based on the Code Napoleon).

The heavy aggregate increase in male and female literacy rates from 1961 was driven by the four districts of the *Old Conquests*. Once education became available, boys and girls were sent almost equally to schools. In the *New Conquests* female literacy rates began to increase significantly only during the last two or three decades. This mechanism also manifests itself in the sex-ratios, which were historically always heavily male biased in the *New Conquests* (which is e.g. shown by the colonial census of 1851 in figure 11). In the *Old Conquests* they were always significantly above 1 and only began to fall slowly from 1961 onwards, mostly probably driven by migration[21]. The discontinuity at the cutoff still persists though.

In order to isolate causal effects and to show that the observed patterns were actually driven by the early Portuguese ”interventions”, I make use of the spatial interpretation of a technique from causal inference called Regression Discontinuity Design (RDD). This allows me to compare villages and towns on either side of this border which has no meaning any more since 250 years in a close-to-random manner: as the balancing checks showed, they are highly similar in size, population, household composition, and features of first nature geography such as rainfall, soil quality and the like. Also anything that regards the infrastructural treatment by the Indian government from 1961 on was shown to be comparable when it came to the placebo checks (number of schools, doctors/nurses, medical coverage).

The last concern could be that those outcomes are driven by economic activity or the prosperity of villages in general. The only way to verify this at the disaggregate level is to look at

[21] See e.g. the Goa Development Report 2011 by the Planing Commission of India

light emissions measured from outer space by satellites which has been shown to proxy very well with local GDP [Henderson et al., 2012]. These light data exhibit no jump alongside my RD-border. What is more there is also no observable difference in the slope of the respective fitted lines, thus also a kink RD design does no exhibit significant coefficients[22].
Having applied the method of elimination on all thinkable observables, I am thus confident to conclude that the observed differences were driven by the only thing in which the villages and towns differ(ed): their differential "cultural" experiences in the distant past which were induced by the Portuguese presence.

The alleviation of the position of women in society and especially within the households throughout the 16th century led to a cultural shift and to the emergence of a different set of gender norms on a sub-continent which historically was always plagued by discrimination towards females. Albeit the jump in the observed gaps in education (as measured by the male & female literacy rate differences) between the two territories shrank from around eight percentage points in 1991 down to 4 percentage points in 2011, they are still observable today. The sex ratios on the other hand do not exhibit such a convergence pattern. The villages just inside the old border that I use as my identification device have around 80 women more for each 1000 men.
That I am not describing something that is driven by the distance of the villages to the coastline can also be visually inferred from the spatial prediction techniques (kriging) that I used in subsubsection 4.2.5: the literacy gap is four to five percentage points almost uniformly across the *Old Conquests* without any observable gradient.

Aside from the spatial RD estimations I find evidence that there was diffusion of attitudes towards women within communities (as summarised in table 12): Hindus and Muslims in the towns of the *Old Conquests* - on aggregate - exhibit systematically different educational gaps than their counterparts in the *New Conquests*. The effect that I am describing throughout this paper thus does not seem to be entirely driven be "being Christian".
Relating to the literature in development economics that tries to understand how to overcome deep rooted gender inequalities [e.g. Jayachandran, 2015], I conclude that it seems to be possible to overcome gender gaps in education comparatively easy. When it comes to the male son preferences and the concomitant male skewed sex-ratios, conventional governmental efforts seems to be fruitless.

All in all, the results lead me to tentatively conclude that the early colonial experience in the form of a cultural treatment seemed to be more important than the uniform institutional and legal treatment that all Goans - independent of gender and religion - experienced from the 1871 civil code on. Compatible with this would be the interpretation that the infrastructural investment, pretty much uniformly received from the Indian government, was just more effective in the *Old Conquests* in the long-run. In order to semantically frame it in terms of the current frontier of the long-run development literature: it thus could be the case that what matters most is the interaction of "culture" with institutions and infrastructure, and that none of them alone are sufficient conditions for economic prosperity [Alesina and

[22] The latter might be more appropriate in general since light does not exhibit jumps by nature but fades out continually. If there was something systematically different on both sides of the border, we should observe different first derivatives, i.e. slopes, on either side of the border.

Giuliano, 2015].
From a methodological point of view I showed that the geographic RD representation by Dell [2010] might lead us to spurious conclusions under certain circumstances, albeit it should still the preferred way to report average treatment effects alongside a cutoff. In order to alleviate this problem, I argue that we just have to stop to discard a clearly identified and visualiseable feature within our data: space. Just by making use of simple spatial interpolation techniques such as kriging we could present the analysed data in a way that puts the reader immediately in a position to infer whether there is an actual jump alongside some specified border or not. I then put forward a simple way in which one could visualise the methodology proposed by Keele and Titiunik [2015]. The two-dimensional RD estimations for every point alongside the discretisised border can be both plotted in form of a "GRDDseries" and at their respective position in space at a map; visualised next to each other. In case the RD estimates change signs multiple times and the coefficients are insignificant most of the time, the researcher then probably has to conclude that the response variable under inspection does not exhibit a jump at the cutoff. Of course these procedures demand a lot from the data, but if we want to meet the high standards of the so-called "credibility revolution in applied Economics" also for spatial RD estimations, we should start to implement them.

References

Abrevaya, J. (2009): "Are There Missing Girls in the United States? Evidence from Birth Data," *American Economic Journal: Applied Economics*, 1, 1–34.

Alesina, A. and P. Giuliano (2015): "Culture and Institutions," *Journal of Economic Literature*, 53, 898–944.

Almond, D., L. Edlund, and K. Milligan (2013): "Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada," *Population and Development Review*, 39, 75–95.

Banerjee, A. and L. Iyer (2005): "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review*, 95, 1190–1213.

Basu, A. (1993): *Women and Missions: Past and Present*, Berg Publishers, chap. Mary Ann Cooke to Mother Teresa: Christian Missionary Women and the Indian Response, 187–208.

Borges, C. J. (1994): *The Economics of the Goa Jesuits, 1542-1759: An Explanation of Their Rise and Fall*, Concept Publishing Company, New Delhi.

Boxer, C. R. (1961): *Four centuries of Portuguese expansion, 1415-1825: a succinct survey*, Witwaterstrand University Press, Johannesburg.

——— (1969): *The Portuguese Seaborne Empire, 1415-1825*, The History of Human Society, Hutchinson, London.

Bruhn, M. and F. A. Gallego (2012): "Good, Bad, and Ugly Colonial Activities: Do they Matter for Economic Development?" *Review of Economics and Statistics*, 94, 433–461.

Cagé, J. and V. Rueda (2016): "The Long-Term Effects of the Printing Press in Sub-Saharan Africa," *American Economic Journal: Applied Economics*, 8, 69–99.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

——— (2015): "Rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs," *R Journal*, 7, 38–51.

Castelló-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2017): "Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India," *The Economic Journal*.

Chaudhary, L. and M. Garg (2015): "Does history matter? Colonial education investments in India," *The Economic History Review*, 68, 937–961.

Chen, X. and W. Nordhaus (2010): "The Value of Luminosity Data as a Proxy for Economic Statistics," NBER Working Papers 16317, National Bureau of Economic Research, Inc.

Croft, T. A. (1978): "Nighttime images of the earth from space," *Scientific American*, 239, 86–98.

Da Silva Gracias, F. (1994): *Health and Hygiene in Colonial Goa, 1510-1961*, Concept Publishing Company, New Delhi.

——— (1996): *Kaleidoscope of Women in Goa 1510-1961*, Concept Publishing Company, New Delhi.

Data{Meet} (2016): "Indian Village Boundaries Project," Tech. rep., http://projects.datameet.org/indian_village_boundaries/.

D'Costa, L. (2007): "Women in post-1961 Goa: Problems and Challenges," *Campus Social - Revista Lusófona de Ciências Sociais*, 249–255.

de Matos, P. T. (2013): "Population Censuses in the Portuguese Empire, 1750- 1800. Research notes." *Romanian Journal of Population Studies*, 7, 5–25.

——— (2016): "Counting Portuguese colonial populations 1776 – 1875: a research note," *The History of the Family*, 21, 267–280.

de Souza, T. R. (1990): *Goa through the Ages - Volume II: An Economic History*, Concept Publishing Company, New Delhi, chap. Rural Economy and Life, 78–116, no. 6 in Goa University Publications Series.

Dell, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863–1903.

Dhar, D., T. Jain, and S. Jayachandran (2018): "Intergenerational Transmission of Gender Attitudes: Evidence from India," *The Journal of Development Studies*, 0, 1–21.

Doepke, M. and M. Tertilt (2009): "Women's Liberation: What's in It for Men?" *The Quarterly Journal of Economics*, 124, 1541–1591.

——— (2018): "Women's Empowerment, the Gender Gap in Desired Fertility, and Fertility Outcomes in Developing Countries," *AEA Papers and Proceedings*, 108, 358–62.

Doepke, M., M. Tertilt, and A. Voena (2012): "The Economics and Politics of Women's Rights," *Annual Review of Economics*, 4, 339–372.

Duflo, E. (2012): "Women Empowerment and Economic Development," *Journal of Economic Literature*, 50, 1051–1079.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis (1997): "Mapping city lights with nighttime data from the DMSP Operational Linescan System," *Photogrammetric Engineering and Remote Sensing*, 63, 727–734.

Emma Maria, S. A. (2002): *Women in Portuguese Goa, 1510-1835*, IRISH, Tellicherry.

Fernández, R. (2014): "Women's Rights and Development," *Journal of Economic Growth*, 19, 37–80.

Ferreira, R. (2016): "Taking Stock: Portuguese Imperial Historiography Twelve Years after the e-JPH Debate," *e-journal of Portuguese History*, 14, 54–70.

GADM (2012): "GADM database of Global Administrative Areas, version 2.0," Tech. rep., http://gadm.org/.

Gaikwad, N. (2015): "East India Companies and Long-Term Economic Change in India," *mimeo, Yale.*

Galor, O. and D. Weil (1996): "The Gender Gap, Fertility, and Growth," *American Economic Review*, 86, 374–87.

Gomes, O. J. (2003): *The Religious Orders in Goa (16th - 17th Centuries)*, Konkani Sorospot Prakashan.

Hahn, J., P. Todd, and W. V. der Klaauw (2001a): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Hahn, J., P. Todd, and W. van der Klaauw (2001b): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Heath, R. and S. Jayachandran (2016): "The Causes and Consequences of Increased Female Education and Labor Force Participation in Developing Countries," Tech. Rep. w22766, National Bureau of Economic Research, Cambridge, MA.

Henderson, J. V., A. Storeygard, and D. Weil (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102, 994–1028.

Henn, A. (2014): *Hindu Catholic Encounters in Goa: Religion, Colonialism, and Modernity*, Indiana University Press.

Hickman, L. A. (2007): "Exploring the Impact of the Media on the Arranged Marriage and Dowry System in Goa, India," Ph.D. thesis, Wilfrid Laurier University.

Huillery, E. (2009): "History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa," *American Economic Journal: Applied Economics*, 1, 176–215.

Ifeka, C. (1989): "Hierarchical Woman: The 'Dowry' System and its Implications among Christians in Goa, India," *Contributions to Indian Sociology*, 23, 261–284.

Imbens, G. and S. Wager (2019): "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 1–15.

Imbens, G. W. and K. Kalyanaraman (2012): "Optimal Bandwith Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 127–155.

Imbens, G. W. and T. Zajonc (2011): "Regression Discontinuity Design with Multiple Forcing Variables," *mimeo.*

Iyer, L. (2010): "Direct versus indirect colonial rule in India: Long-term consequences," *The Review of Economics and Statistics*, 92, 693–713.

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

KAMAT, P. (1999): *Farar Far: Local Resistance to Colonial Hegemony in Goa, 1510-1912*, Institute Menezes Braganza.

——— (2000): *Goa and Portugal, History and Development*, Concept Publishing Company, New Delhi, chap. Peasantry and the colonial state in Goa (1946-1961), 133–158, no. 10 in XCHR Studies Series.

KEELE, L. J., S. LORCH, M. PASSARELLA, D. SMALL, AND R. TITIUNIK (2017): "An Overview of Geographically Discontinuous Treatment Assigments With an Application to Children's Health Insurance," in *Advances in Econometrics*, ed. by M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, vol. 38.

KEELE, L. J. AND R. TITIUNIK (2015): "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

——— (2016): "Natural Experiment Based on Geography," *Political Science Research and Methods*, 4, 65–95.

KELLY, M. (2019): "The Standard Errors of Persistence," *mimeograph*.

KHANDAY, M. I., B. A. SHAH, P. A. MIR, AND P. RASOOL (2015): "Empowerment of Women in India - Historical Perspective," *European Academic Research*, 2, 14494–14505.

KUMAR, A. V., B. KIRAN, AND J. T. GONE (2013): "The Impact of the Dowry System In Christian Communities," Tech. rep., Dharmaram Vidya Kshetram, http://www.catholicethics.com/forum-submissions/the-impact-of-the-dowry-system-in-christian-communities.

LEE, D. S. AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

LEHNER, A. (2019): "SpatialRDD: An R-Package for Spatial Regression Discontinuity Designs," .

MALVANKAR, A. (2015): "Education and society in Goa under Portuguese rule," *Journal of Social and Economic Development*, 17, 298–311.

MANTOVANELLI, F. G. (2013): "The Protestant Legacy: Missions and Literacy in India," *Job Market Paper, Boston College*.

MARCOCCI, G. (2012): *A Consciência de um Império: Portugal e seu Mundo (sécs. XV-XVII)*, Imprensa da Universidade de Coimbra, Coimbra.

MICHALOPOULOS, S. AND E. PAPAIOANNOU (2014): "National Institutions and Subnational Development in Africa," *Quarterly Journal of Economics (2014), 151-213*.

MICHALOPOULOS, S. AND E. PAPAIOANNOU, eds. (2017): *The Long Economic and Political Shadow of History, Volume I. A Global View*, CEPR Press.

MIZUSHIMA LABORATORY (2013): "India Place Finder," Tech. rep., Department of Oriental History, Graduate School of Humanities and Sociology, The University of Tokyo, http://india.csis.u-tokyo.ac.jp/.

Neill, S. (1985): *A History of Christianity in India 1701-1858*, vol. 2, Cambridge University Press.

Newitt, M. D. D. (2005): *A History of Portuguese Overseas Expansion: 1400 - 1668*, Routledge.

Nunn, N. (2009): "The Importance of History for Economic Development," *Annual Review of Economics*, 1, 65–92.

——— (2014a): "Gender and missionary influence in colonial Africa." in *Africa's Development in Historical Perspective*, ed. by E. Akyeampong, R. Bates, N. Nunn, and J. A. Robinson, Cambridge University Press, 489–512.

——— (2014b): "Historical Development," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, North-Holland, vol. 2, 347–402.

Papay, J. P., J. B. Willett, and R. J. Murname (2011): "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables," *Journal of Econometrics*, 161, 203–207.

Pearson, M. N. (1988): *The Portuguese in India*, vol. 1 of *The New Cambridge History of India*, Cambridge University Press.

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

Rocha, R., C. Ferraz, and R. R. Soares (2017): "Human Capital Persistence and Development," *American Economic Journal: Applied Economics*, forthcoming.

Russell-Wood, A. J. R. (1992): *A World on the Move: The Portuguese in Africa, Asia, and America, 1415-1808*, Manchester: Carcanet Press Limited.

Sen, A. (1990): "More than 100 Million Women Are Missing," *The New York Review of Books*.

Spolaore, E. and R. Wacziarg (2013): "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51, 325–369.

Srivastava, H. C. (1990): *Goa through the Ages - Volume II: An Economic History*, Concept Publishing Company, New Delhi, chap. Demographic History and Human Resources, 55–77, no. 6 in Goa University Publications Series.

Susewind, R. (2015): "What's in a Name? Probabilistic Inference of Religious Community from South Asian Names," *Field Methods*, 27, 319–332.

——— (2017): "India-Religion-Politics: Data on Religion and Politics in India," .

Tertilt, M. (2005): "Polygyny, Fertility, and Savings," *Journal of Political Economy*, 113, 1341–1371.

Valencia Caicedo, F. (2014): "The Mission: Human Capital Transmission, Persistence and Culture in South America," *Job Market Paper, UPF*.

——— (2019): "The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America," *The Quarterly Journal of Economics*, 134, 507–556.

Varde, P. S. (1977): *History of Education in Goa, from 1510 to the Present Day*, Goa Vidya Pratishthan.

Velinkar, J. (1984): "Jesuit Educational Style in Sixteenth Century Goa," *Indica*, 21, 17–27.

Waldinger, M. (2017): "The Long-Run Effects of Missionary Orders in Mexico," *Journal of Development Economics*, forthcoming.

Woodberry, R. D. (2004): "The Shadow of the Empire: Christian Missions, Colonial Policy, and Democracy in PostColoPost Societies," Ph.D. thesis, University of North Carolina at Chapel Hill.

——— (2012): "The Missionary Roots of Liberal Democracy," *American Political Science Review.*

Xavier, P. D. (1993): *Goa: A Social History*, Rajhauns Vitaran.

Xue, M. M. (2018): "High-Value Work and the Rise of Women:The Cotton Revolution and Gender Equality in China," *mimeo.*

8 Appendix: Spatial RDDs

Moved to [Lehner, 2019].

9 Data Appendix

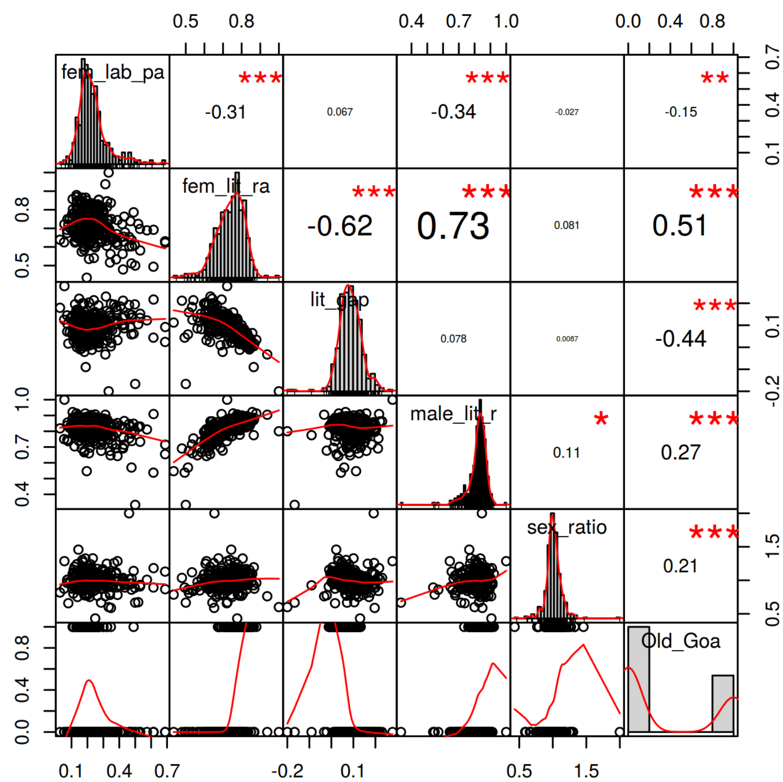


Figure 15: Correlation matrix of all considered outcome variables at the village level (2011 census)

9.1 Further Descriptives

9.2 Religion by district

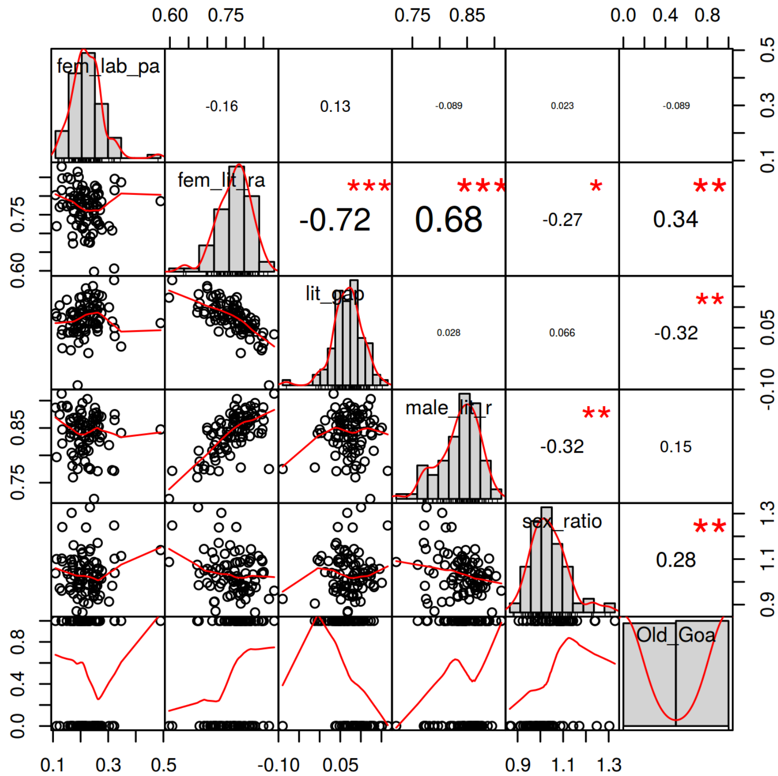


Figure 16: Correlation matrix of all considered outcome variables at the village level, within a 5km bandwidth of the RD border (2011 census)

Table 13: Average effects (OLS with dummy)

	Dependent variable:			
	Literacy rate 2011			
	(1)	(2)	(3)	(4)
$\mathbb{I}$(Village in Old Goa)	0.062***	0.046***	0.036***	0.022***
	(0.006)	(0.008)	(0.006)	(0.007)
Dist. to Coast		−0.001***		−0.001***
		(0.0003)		(0.0003)
Dist. to hist. Goa			−0.002***	−0.002***
			(0.0002)	(0.0002)
Constant	0.755***	0.776***	0.813***	0.831***
	(0.004)	(0.008)	(0.007)	(0.009)
Observations	402	402	402	402
R^2	0.225	0.243	0.365	0.380
Adjusted R^2	0.223	0.239	0.362	0.375
Residual Std. Error	0.056 (df = 400)	0.055 (df = 399)	0.050 (df = 399)	0.050 (df = 398)
F Statistic	116.055*** (df = 1; 400)	64.011*** (df = 2; 399)	114.657*** (df = 2; 399)	81.239*** (df = 3; 398)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 14: Average effects (OLS with dummy)

	Dependent variable:			
	Literacy gap 2011			
	(1)	(2)	(3)	(4)
$\mathbb{I}$(Village in Old Goa)	-0.050^{***}	-0.037^{***}	-0.047^{***}	-0.035^{***}
	(0.005)	(0.006)	(0.005)	(0.007)
Dist. to Coast		0.001^{***}		0.001^{***}
		(0.0003)		(0.0003)
Dist. to hist. Goa			0.0002	0.0002
			(0.0002)	(0.0002)
Constant	0.105^{***}	0.087^{***}	0.098^{***}	0.082^{***}
	(0.003)	(0.006)	(0.006)	(0.008)
Observations	402	402	402	402
R^2	0.220	0.238	0.223	0.240
Adjusted R^2	0.218	0.234	0.219	0.234
Residual Std. Error	0.046 (df = 400)	0.045 (df = 399)	0.046 (df = 399)	0.045 (df = 398)
F Statistic	112.803^{***} (df = 1; 400)	62.234^{***} (df = 2; 399)	57.100^{***} (df = 2; 399)	41.871^{***} (df = 3; 398)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 15: Average effects (OLS with dummy)

	Dependent variable:			
	Sex Ratio 2011			
	(1)	(2)	(3)	(4)
$\mathbb{I}$(Village in Old Goa)	0.054^{***}	0.067^{***}	0.076^{***}	0.088^{***}
	(0.012)	(0.016)	(0.013)	(0.017)
Dist. to Coast		0.001		0.001
		(0.001)		(0.001)
Dist. to hist. Goa			0.002^{***}	0.002^{***}
			(0.0005)	(0.0005)
Constant	0.974^{***}	0.956^{***}	0.924^{***}	0.908^{***}
	(0.007)	(0.017)	(0.016)	(0.021)
Observations	402	402	402	402
R^2	0.047	0.050	0.077	0.080
Adjusted R^2	0.044	0.046	0.072	0.073
Residual Std. Error	0.117 (df = 400)	0.117 (df = 399)	0.116 (df = 399)	0.116 (df = 398)
F Statistic	19.667^{***} (df = 1; 400)	10.589^{***} (df = 2; 399)	16.557^{***} (df = 2; 399)	11.460^{***} (df = 3; 398)
Note:				*p<0.1; **p<0.05; ***p<0.01

	Name	lit	m_lit	f_lit	lit_gap
1	Pernem	0.797	0.838	0.755	0.083
2	Bardez	0.869	0.884	0.856	0.028
3	Tiswadi	0.856	0.871	0.842	0.029
4	Bicholim	0.864	0.889	0.838	0.051
5	Satari	0.848	0.873	0.823	0.049
6	Ponda	0.825	0.855	0.797	0.058
7	Mormugao	0.817	0.85	0.785	0.065
8	Salcete	0.807	0.834	0.785	0.049
9	Quepem	0.73	0.77	0.694	0.076
10	Sanguem	0.763	0.804	0.726	0.078
11	Canacona	0.815	0.84	0.794	0.046

Table 16: Shares only for Christians

	Name	lit	m_lit	f_lit	lit_gap
1	Pernem	0.707	0.732	0.678	0.054
2	Bardez	0.714	0.744	0.678	0.066
3	Tiswadi	0.704	0.724	0.683	0.041
4	Bicholim	0.758	0.781	0.732	0.05
5	Satari	0.825	0.843	0.807	0.035
6	Ponda	0.721	0.755	0.682	0.073
7	Mormugao	0.749	0.784	0.712	0.072
8	Salcete	0.728	0.752	0.702	0.05
9	Quepem	0.708	0.739	0.676	0.063
10	Sanguem	0.714	0.756	0.669	0.087
11	Canacona	0.726	0.765	0.681	0.084

Table 17: Shares only for Muslims

	Name	lit	m_lit	f_lit	lit_gap
1	Pernem	0.802	0.844	0.758	0.086
2	Bardez	0.821	0.849	0.791	0.058
3	Tiswadi	0.819	0.853	0.781	0.071
4	Bicholim	0.805	0.846	0.762	0.084
5	Satari	0.761	0.819	0.699	0.12
6	Ponda	0.813	0.855	0.769	0.086
7	Mormugao	0.799	0.837	0.754	0.083
8	Salcete	0.817	0.844	0.787	0.057
9	Quepem	0.753	0.793	0.711	0.082
10	Sanguem	0.744	0.802	0.685	0.117
11	Canacona	0.748	0.799	0.696	0.103

Table 18: Shares only for Hindus

A Note on Spatial Regression Discontinuity Designs

Alexander Lehner *[a]

[a]Unversity of Bologna

This version: June 17, 2019

Abstract

This note aims to improve the state-of-the-art spatial regression discontinuity design (RDD) machinery. It proposes a way to report heterogeneous treatment effects alongside the RD cutoff and a way to visualise the results with more meaningful and intuitive plots that employ spatial smoothing techniques. A companion R-package, `SpatialRDD`, implements the nonparametric estimation techniques by both Keele and Titiunik [2015] and Imbens and Wager [2019]. I then point out problems with the way some parametric spatial RDs are commonly estimated. I provide solutions to alleviate what I identify as a main issue, which is that there are many arbitrary researcher choices - especially with the combination of fixed effects and cluster level choices - that can be made on the way to obtaining statistically significant results. Furthermore, I put recent results from the multi-dimensional RD literature into perspective, and claim that we cannot apply the hard, data hungry standards from the classic causal inference literature to these estimations that exploit geographic variation. This is being demonstrated with (spatial) Monte Carlo simulations. I partly second recent critique raised by Keele and Titiunik [2015] and Imbens and Wager [2019] when it come to parametric estimation and even add some elements to it. But in the end I also conclude that, with enough caution and additional checks, parametric estimations are still a valuable way to report average treatment effects. The problems that I address are also alleviating concerns that have been recently raised when it comes "matching spatial noise" in the "persistence literature" [Kelly, 2019].
`SpatialRDD` is also capable of setting up border segment fixed effects, shift RD cutoffs in space in order to create placebo borders, and finally also assign the treatment status for these robustness exercises. All of which are usually very tedious tasks in GIS software that also do not guarantee easy replicability.

Keywords: Geographic Regression Discontinuity Design (GRDD), Spatial RDD, Causal Inference, R.

JEL Codes: xx

*So far I benefited from comments & conversations with Melissa Dell, Margherita Fort, Nicolà Gennaioli, Trevon D. Logan, Robson Tigre, Rocío Titiunik, Andrea Weber, David Weil, seminar participants at the University of Oxford (Nuffield, Graduate Seminar Series), the University of Bologna, the Macro Lunch at Brown University.
A shoutout also goes to the attendants of the PhD course in advanced spatial analysis at the University of Bologna (A.A.18/19) who were to first ones to test-run the `SpatialRDD` package and gave useful suggestions.
A development version of the package can be found [here] (description in the "vignettes" folder).
E-Mail: `alexander.lehner@unibo.it`

Contents

1 Introduction

In a "classic" regression discontinuity design (RDD), initially developed in the 1960's [Thistlethwaite and Campbell, 1960], units are assigned treatment and control conditions solely based on a single cutoff score on a continuous assignment variable. Under certain identifying assumptions, the observed jump of the dependent variable at the cutoff then represents the average treatment effect [Hahn, Todd, and van der Klaauw, 2001; Imbens and Lemieux, 2008]. Recently there has been a growing interest in applications that extend this approach to two or more assignemnt variables. In Economics and Political Science this is especially true for a subset of these: geographic RDDs that use a line in space as discontinuity. Spatial interpretations of RDDs are rising in popularity since the well known application by Dell [2010]. Before her seminal paper on the peruvian mining mita, there was virtually just a handful [e.g. Black, 1999; Bayer, Ferreira, and McMillan, 2007]. In the last decade, though, the number of such applications was rising immensely. This seems mostly driven by the large scale availability of GI software and the rising number of geo-referenced datasets. The equivalent of an arbitrary cut-off along the real line in these applications is formed by a line in space (e.g. (abandoned) borders, ethnic boundaries, media or schooling boundaries, etc.). The methodological difficulties that arise in those spatial settings were clarified by Keele and Titiunik [2015], building on the existing multidimensional RD literature [Papay, Willett, and Murname, 2011; Reardon and Robinson, 2012; Zajonc, 2012; Wong, Steiner, and Cook, 2013]. Similar to a recent contribution by Imbens and Wager [2019], they implicitly cast doubt on existing parametric implementations and propose nonparametric estimations.

By utilising a companion R-package, `SpatialRDD`, I will first juxtapose parametric and nonparametric ways of estimation on simulated data with a known, discontinuous, data generating process (DGP). I will present a novel way of reporting heterogeneous treatment effects alongside the cutoff that are estimated on any point of the border. By constructing "corner examples", will see that parametric estimation, reporting one homogeneous treatment effect via OLS, is more error prone.

Afterwards I will demonstrate with spatial Monte Carlo simulations that the performance of nonparametric estimates on small sample sizes is fairly well.

In the last section of this note I will replicate the RD application of Keele and Titiunik [2015], analysing the effect of TV advertisement on 2008 electoral turnout alongside a media market boundary. This has also been used by Imbens and Wager [2019] in order to illustrate their "optimized approach" to multi-dimensional RDDs. Both conclude that there is no significant effect at the RD cutoff. As I will show, this is both casting doubt on parametric estimations in the spirit of Dell [2010], but also generally puts standards on spatial RDs that are slightly too high. Once sparse boundary points on the fringes are removed, I will demonstrate by nonparametrically estimating effects on numerous points (KT15 only do this for three) that there in fact seems to exist a significant effect at the cutoff. Nevertheless I do agree with their overall conclusion that there was no causal effect on advertising on 2008 turnout. This is because the balancing checks suggest that there was already a "jump" in important covariates before the treatment.

The solutions that are presented are also somewhat alleviating concerns that have been recently raised about the "persistence literature" [Kelly, 2019]. The preferred nonparametric specifications do not require parametric assumptions, and thus there are no error terms that could be spatially correlated.

The R-package `SpatialRDD` also unifies previously very tedious tasks in one framework. A

general complication so far was that people relied on external GIS APIS such as QGIS or GeoDA, which makes reproducibility and the easy creation of numerous robustness checks very cumbersome and time consuming.
Some technical issues will be clarified before we move to the first simulations in section 3.

2 Spatial RDDs

As a geographic RD design is just a special case of an RD with multiple running variables, we can draw on the existing literature that developed identification conditions for these cases [Papay et al., 2011; Reardon and Robinson, 2012; Zajonc, 2012; Wong et al., 2013], typically using some form of local linear regression for estimation [as in Hahn et al., 2001; Porter, 2003]. The equivalence between these designs and geographic RDs has been shown by Keele and Titiunik [2015], pointing out additional problems that are specific to geographic settings. A binary treatment, T, is assumed to be a function of known covariates, $\mathbf{S} = (S_1, S_2)$, extending the dimension of the running/forcing variable, or score, from the standard RD framework by one dimension. The 2-tuple $\mathbf{S}_i$ contains the X- and Y-coordinates for each observation i, (S_{i1}, S_{i2}), and thus uniquely pins down the location in space for each observation[1]. Given treated ($\mathcal{A}^t$) and control ($\mathcal{A}^c$) areas that are non-overlapping and adjacent (and obviously subset of some set of coordinate pairs that represent space, $\mathcal{A}^{c,t} \subset R^2$), the treatment status of observation $i = 1, ..., n$ is assigned as $T_i = \mathbf{1}(\mathbf{S}_i \,\epsilon\, \mathcal{A}^t)$. In practice the assignment of the "treated dummy" is achieved by a simple spatial intersection of any i with the spatial polygon that resembles the treatment area. The function `assign_treated()` alleviates this task. Eventually any RD study is then interested in a potential treatment effect of unit i, $\tau_i = Y_{i1} - Y_{i0}$, but as we cannot observe both outcomes Y at the same time, it is not possible to recover this individual effect. Thus, under a set of assumptions, we will use a spatial RD setup in order to learn about local averages of τ_i.

This essay follows a three-way classification that is partly inspired by Keele and Titiunik [2015], but making an important distinction between parametric and nonparametric specifications[2]:

1. A spatial RDD in "naive" distances, ideally estimated non-parametrically through a local linear regression, identical to classic RD's but with distance to a cutoff as assignment variable (examples: [**?**]).

2. A parametric specification that includes a polynomial in the X- & Y-coordinates to control for space. Typically estimated via OLS and reporting a single homogeneous treatment effect (put forward by Dell [2010]).

3. A nonparametric specification with two running variables, estimated nonparametrically at <u>every</u> point of the cutoff [labelled as GRD design by Keele and Titiunik, 2015]. This approach is then also extended by the finite-sample-minimax linear estimator approach

[1] In case the coordinate reference system (CRS) being used is EPSG:4326, this would be (longitude, latitude). As is noted in Lehner [2019b], a local coordinate system would be preferable for spatial RDDs, as this would allow us to do calculations in metres rather than degrees (in the case of the common 4326) and additionally would give us more precision.

[2] The discussion is restricted to the - generally more credible - subset of studies wherein experiments are based on <u>adjacent</u> areas.

put forward by Imbens and Wager [2019], which nicely complements the GRDD series visualisation Lehner [2019a] because it allows to draw a weighted conditional average treatment effect (CATE) alongside the cutoff on a map.

In the Economics literature, especially the sub-field that tries to link historical events to contemporary economic outcomes, the second specification in which the outcome variable is regressed on a dummy variable whether the unit was treated or not and a set of control variables (including polynomials in the X- and Y-coordinates that control for the position), seems to be the most popular one.
We will see in section 3 on a sample of (spatially) simulated data with an embedded discontinuity that all these approaches are almost equivalent when it comes to determining an average treatment effect. With complicated real world data, repeating the replication of Keele and Titiunik [2015] as in Imbens and Wager [2019], we are going to see in section 5 how convoluted things can get. An approach in naive distances proves to be of no use, and parametric specifications may lead to inconclusive and confusing results, depending on the set of choices made by the researcher when it comes to determining cluster levels, border segment lengths, and bandwidths. This resembles a classic problem of what Gelman and Loken [2013] call "researcher degrees of freedom".
The classification also somewhat depicts the evolution of the spatial RDD literature, being mostly restricted to Economics and Political Science, on its path towards more credible empirical analysis of geographically heterogeneous effects. Less surprisingly, the conclusion is going to be that nonparametric approaches are more desireable in most cases. These are sometimes impossible to implement, though, because quite often either the sample size is too small for such data-demanding procedures or the continuous treatment assignment assumption is violated due to restrictions in the geolocalisations (e.g. due to privacy, reporting guidelines,...). The latter case is touched upon in subsection 2.4.

The methodological difficulties that arise in those spatial settings were clarified by Keele and Titiunik [2015]. They further show that their ideal GRD designs lead to identification of local treatment effects at the cutoff under a two-dimensional continuity assumption [thus generalising Hahn et al., 2001; Porter, 2003]. They also point out the potential problems arising with GRD designs and other spatial RD set-ups in general. An essential difference to "classical" RD's with two forcing variables is, that individual points at the boundaries of GRDD's have a clear interpretation.
I second their critique and take their approach one step further and propose a way to consistently visualise the heterogenous treatment effect alongside any RD-border (this is what I call a GRDDseries). Furthermore I argue that all investigated outcome variables should also be visualised by simple spatial interpolation techniques such as Kriging. This will give the reader a feeling of the spatial dimension of the data generating process.

2.1 Specification in Naive Distances

This simpler version of a spatial RD uses the euclidean distance of each observation to the boundary as its score for the estimation. This geographically "naive" measure of distance ignores how the units are spatially distributed since the shortest distance towards the border does not determine the exact location in the two-dimensional space [as e.g. pointed out in Keele and Titiunik, 2015, p. 137].

This specification allows to estimate an average effect which will mask considerable heterogeneity along the border. Yet, as has been shown by Lehner [2019a] and in the documentation of the `SpatialRDD` package, this type of analysis is capable of delivering a quick intuition of what the data can potentially tell. Especially the visualisation of this set-up with a "standard" RD-plot can be a quite informative first check. Also Dell and Olken [forthcoming] use this approach to visually summarise results for their numerous RD borders in one single plot.

In this setting units around a narrow band around a border are assumed to be valid counterfactuals. As this design may mask underlying heterogeneity, it may not allow to evaluate the plausibility of the needed identification assumptions. This problem gets arguably more severe the longer the RD border is. A naive GRD design is still appropriate in some circumstances, e.g. when the boundary of interest is short and defines a homogeneous region. Technically it is only fully valid when treatment effects are constant at all boundary points $\mathbf{b} \varepsilon \mathcal{B}$.

Ideally such specifications are measured non-parametrically via local linear regressions as is state-of-the art in "classic" RDs, ideally with data driven robust confidence intervals as suggested by Calonico, Cattaneo, and Titiunik [2014].

2.2 Parametric Specifications

This is what Keele and Titiunik [2015] call nongeographic RD designs, Keele and Titiunik [2016] classify under strategies that rely on a "Conditional Local Geographic Treatment Ignorability Assumption", and Keele, Lorch, Passarella, Small, and Titiunik [2017] subsume under "Geographic Quasi Experiment" (GQE). Put forward by Dell [2010], this specification is usually preferred within the Economics literature. It is well suited to report an average treatment effect alongside a given border. But as we will see later it could be misleading in some circumstances and produce false positives. As the more detailed visualisation in form of the GRDDseries shows, this is due to strong effects that occur locally and drive the average effect. Even masking points where we obtain statistically significant estimates with opposite sign (this is typically a sign that the effect is not robust, in addition to confidence intervals that are strongly overlapping with zero).

$$y_i = \alpha + \beta\ T_i + \sum_{s=i}^{S} \gamma_j\ SEGMENT_{ji} + \delta'_i \mathbf{X} + f(\mathsf{geolocation_i}) + \varepsilon_i, \tag{1}$$

where y_{ib} is the outcome variable of interest for location i. T_i denotes a dummy variable that represents the "treatment status", equalling 1 if the location intersects with the treated polygon. $f(\mathsf{geolocation_i})$ represents the RD polynomial in X- and Y-coordinates, which is supposed to control for smooth functions of geographic location and is going to take on varying forms across different specifications. Finally, matrix $\mathbf{X}$ contains a set of control variables and $SEGMENT_{ji}$ represents a dummy, equalling 1 if location i has segment j as its closest segment. Regression just produces the weighted average over all segments [see e.g. Angrist and Pischke, 2008]. These are the equivalent to a set of boundary segment fixed effects and are meant to capture geographic heterogeneity and alleviate omitted variable problems by only exploiting within segment variation. This might be desirable in some settings, but the obvious drawback is that this approach masks the heterogeneity that potentially is capable

of delivering deeper insights into the problem at hand.
The inclusion of segment dummies rather than fixed effects is also desirable for interpretability reasons, as the conditional independence assumption, if it would hold, would mean that results might already be causal since

$$E[Y_{0i}|T_i, SEGMENT_i,\ \mathbf{X}] = E[Y_{0i}|SEGMENT_i,\ \mathbf{X}],$$

where Y_{0i} is the dependent variable in the absence of treatment (note the treatment indicator T_i dropping out, given CIA being true).

Another drawback of this approach is that there are no clear formalised suggestions for the bandwidth selection in such an estimation procedure as were put forward for non-parametric methods [Imbens and Kalyanaraman, 2012; Calonico et al., 2014].

2.3 Nonparametric Specifications (two running Variables)

The two approaches that are being presented here are aiming to determine an average treatment effect for any point $\mathbf{b}\epsilon\mathcal{B}$, where $\mathcal{B}$ represents the set of all points on the cutoff.

2.3.1 The Keele and Titiunik [2015] way

The spatial RD design, if estimated nonparametrically by the use of two running variables, enables us to potentially identify a larger family of treatment effect functionals. These were brought into a geographic context by Keele and Titiunik [2015]. Therein they develop in detail their so-called GRD. The difference, as compared to the method from Dell [2010], is that the comparability of treated and control units need not occur in a certain band around the geographic discontinuity. Given a focal point c, they first measure Euclidean distance of each observation to it, $D_i = \|X_i - c\|$, and then use D_i as a running variable in a univariate regression discontinuity analysis. Consistency is established via the standard local regression argument following Hahn et al. [2001]. Imbens and Wager [2019] remark that the resulting estimator has no guarantees in terms of optimality.
The effect in this type of setting is measured at *any* point of the border. One of the huge advancements is therefore, that one can obtain heterogeneous treatment effects alongside the border. The estimates are obtained with "standard" RD techniques [e.g. Calonico, Cattaneo, and Titiunik, 2015], using a triangular kernel, so that points farther away from the cutoff receive less weights in the estimates. This local linear estimation has also the advantage that there are no distributional assumptions on parameters needed, and thus spatial autocorrelation is less of a concern.

Drawbacks. All the motivating theory for local linear regression relies on the running variable having a continuous distribution. However, in practice, this running variable often has a discrete distribution with a modest number of points of support. When the running variable is discrete there is no compelling reason to expect local linear regression to be particularly effective in estimating the causal effect of interest. In spite of these limitations, we will see that local linear regression performs fairly well. Another pitfall when it comes to the data-driven selection procedures that choose specific bandwidths for the estimation [Imbens and Kalyanaraman, 2012; Calonico et al., 2014] is, that they in some cases are not capable of dealing adequately with sparse boundary points where there are no observations close to the border.

To sum up, the effect identified in a GRDD is not a point estimate but a line of treatment effects along the whole cut-off, displaying the magnitude of the heterogeneous effect at each point. As compared to "classical" RDs with two forcing variables, here the heterogeneity has a clear interpretation. In essence this can be highly valuable since geographic patterns in the results would potentially affect the interpretation of the results.

2.3.2 The Imbens and Wager [2019] way

In a recent publication, Imbens and Wager [2019] are not quite satisfied with the approach by Keele and Titiunik [2015] as it incorporates a multivariate identification strategy but then reduce the problem to a univariate regression discontinuity problem for estimation. This is a valid concern, but I will show later that all estimation methods are more or less equivalent.

Motivated by the large literature on minimax linear estimation, Imbens and Wager [2019] study an alternative approach based on directly minimizing finite sample error bounds via numerical optimization, under an assumption that the second derivative of the response surface is bounded away from the boundary of the treatment region. Their estimator is well defined regardless of the shape of the treatment region. More importantly for spatial RDDs, their inference is asymptotically valid for both discrete and continuous running variables.
An additional plus of this way of nonparametrically estimating treatment effects is, that it allows for a very intuitive visualisation of a weighted average effect alongside the full RD border. We will explore this further in the following chapters.

2.4 Discussion of Identification Issues

Sorting around the cutoff One potential problem is that the continuity assumptions needed for identification will hold less often in geographical settings because agents may sort around the boundaries and thus undermine the design's validity. This is less of a concern in historical applications where gains from moving just a few kilometres across a border in the form of tax cuts or monetary benefits were virtually non existing. And if they existed such a border typically would fail to meet important assumptions, as in these cases a cutoff would also coincide with other important boundaries. Albeit humans were historically always highly mobile, migrations happened towards urban centers that presented occupational opportunities, and not e.g. across a river from one rural setting to another where important characteristics (soil quality, climate,...) are uniform. This is the argument that Lehner [2019a] uses to alleviate sorting concerns. Dell and Olken [forthcoming], for example, convince the reader by making the point that "historical high productivity sorters" would have been actually disadvantaged by a forced labour system. The concerns of outwards migration of low productivity individuals is alleviated by showing today's population density differentials.

Compound Treatment Effects Another issue could be so called compound treatments. In the context of historical applications, this could be additional events or policies that affected the region that is subject to study in exactly the same geographical dimension. A classic example that suffers from such a problem would be a political- or country-border. In such a setting it would be impossible to isolate the effect of one single treatment, because many different policies and laws are geographically bounded at the same border.

Pre-Treatment Discontinuities Quite often it could be the case that that a given RD cutoff was also a "demarcation line" in space for other reasons. This is why it is necessary to show that other important covariates exhibit no jump across the threshold. As we are going to see in section 5, this is the case for several important variables in the household data for the replication of Keele and Titiunik [2015]: units right across the border differ in terms of age, income, and race. These could basically be seen as balancing checks, even though the latter typically focus more on geographical features and climate.

Non-availability of exact locations A classic problem that arises is when the exact geolocation of each unit of observation is not available. When it comes to micro data this might be for privacy reasons (i.e. only the coordinates of the respective cluster in which the unit is located is known), or in more macro applications it might be the case that only the centroid of a certain polygon, e.g. a geographic area, is known. Nonparametric estimations with local linear regressions are not feasible any more then, as they require continuous running variables in order for identification to be valid. For these cases there exist techniques that are tailored towards discrete running variables [see Cattaneo, Idrobo, and Titiunik, forthcoming, for a textbook treatment]. The nonparametric "optimized approach" by Imbens and Wager [2019] is immune towards this problem, given the other assumptions are met and the sample size is large enough. The latter problem typically being one of the main issues that arises when it comes to geographic RD applications.
Dell [2010] solves both the discrete running variable problem and sample size issues with a parametric approach, controlling for space, that obtains estimates via OLS.
In section 4 we will see that the sample size restrictions for nonparametric estimation are not as strict as one might think. Just a few hundred observations seem to be enough to get a good approximation of an average treatment effect.

2.5 Spatial RD Robustness Checks

In some occasions it might be appropriate, depending what sort of entity the units of observation represent, to adopt some of the statistical tests from the "classic" RD literature. The test by McCrary [2008] tries to infer whether there is sorting around the cutoff. In case the units of observations are villages or geographic units, as is quite often the case in applications in Economics, this might not be sensible because such geographic entities are obviously not able to move. It rather gives information on whether the geographic extents of the units of observations are comparable across the cutoff. This might potentially be more sensibly tested by standard balancing checks via regressing the geographic area on a treated dummy. The McCrary test implements a local linear density estimator and, simply put, fails to reject the null-hypothesis that there is no sorting if the units of observations are equally spaced on both sides of the threshhold.
A more recent test by Canay and Kamat [2018] provides a way to infer whether covariates other than the outcome variable are displaying a jump across the threshhold. The test is only implemented for one dimensional tressholds and is conceptually very similar to the balancing checks of baseline variables that are a standard requirement in any spatial RDD analysis.

Especially the latter is computationally demanding because it draws on permutation methods and thus might not be desireable if numerous RD point estimates are obtained alongside the border. Both of these tests are implemented alongside the estimation in the companion `SpatialRDD` package [Lehner, 2019b] and can be turned on and off via the respective parameters in the `SpatialRD()` and `printspatialrd()` functions.
The important robustness checks that have to be undertaken are

1. estimation on shifted placebo borders
2. estimation on the actual cutoff but with different outcomes, showing that there is no jump in other variables
3. expansion and reduction of the estimation bandwidth

furthermore the robustness towards inclusion of different segment dummies, cluster levels, and additional covariates has to be shown.
Along the way the workhorse functions of the package will be explained to carry out these tasks. Saving the researcher a lot of time and making the analysis highly replicable.

2.6 Further Spatial RD Enhancements

2.6.1 Spatial Interpolation

An additional element that, complementary to the econometric analysis described above, should be implemented in the toolbox of applied social scientists using spatial boundaries as identification devices are spatial interpolation techniques. These range from very simple approaches, like predicting the value of a neighbouring cell just by using a certain percentage of the value of the actual cell, to highly sophisticated techniques that use state of the art methods from statistics and machine learning in order to spatially predict outcome variables. In a way these interpolations preclude some of the necessary robustness checks, as one can visually infer from the spatial patterns whether there is actually some discontinuity identifiable alongside the identified cutoff. Especially in the simulated "false positive" examples in the next chapter this is becoming evident.
Several of these visualisations are shown in the appendix.

2.6.2 Spatial Matching

In some occasion it might be worthwile to enhance a given spatial RD design by applying matching methods within a certain RD treshold alongside a cutoff. This has been mapped into a geographic RD setting by Keele, Titiunik, and Zubizarreta [2015], deploying the matching methods from Zubizarreta [2012]. The goal is to match units on opposite sides of the RD cutoff that are similar with respect to a given set of covariates. The additional element of the units being also geographically very close to each other, allows them to analyse the matched outcomes as if they came from a randomized experiment. For approaches like this one, however, fairly large sample sizes are required in order to have enough statistical power. A detailed treatment is beyond the scope of this paper, but the matching approach can be easily implemented with the available functions from the `SpatialRDD` package in combination with `mipmatch`, the R implementation of Zubizarreta [2012].

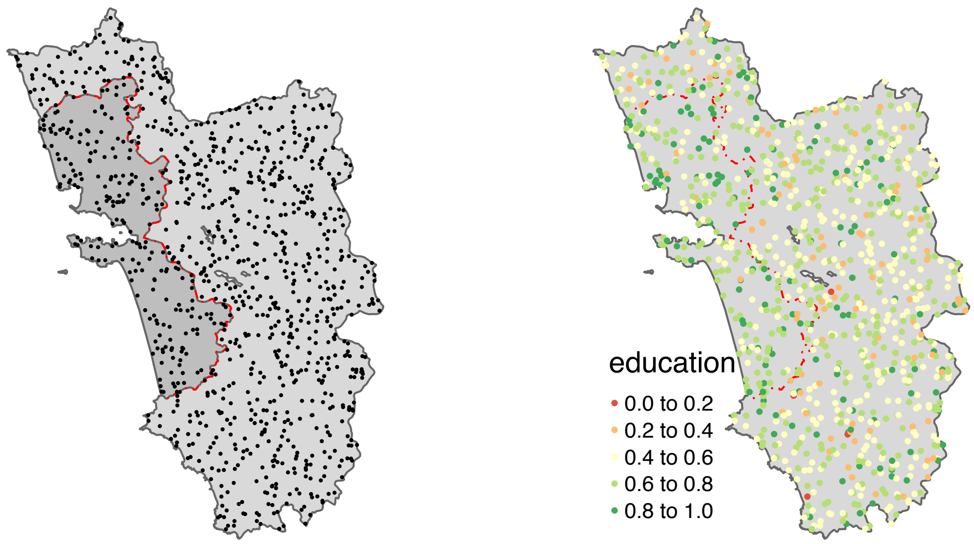


Figure 1: All c. 1000 (spatially) simulated observations and the values for the dependent variable. The red dashed line depicts the discontiuity in space.

3 A Simulation Study

The geographic polygon data used is from Lehner [2019a], representing the Indian state of Goa, and projected in the local UTM Zone43 projection system (EPSG32643). For the purpose of this exercise the projection system is not important, but in general it has to be noted that working in a local system (rather than the ubiquitious EPSG4326 that is defined in degrees (longitude, latitude)) is preferable as it is more precise and it allows us to work in meters rather then degrees. The dataset contains a "treated" polygon, a polygon for the full extent that also defines the bounding box for our study, and a line that represents our discontinuity in space. It comes together with the `SpatialRDD` package.
The cutoff that describes the discontinuity in space is 129 kilometres long and the treatment area is 802 square kilometres large.

Figure 1 shows all the observations that were simulated within the geographies. The data generating process for the outcome variable of interest, education as measured in literacy rates, looks as follows:

```
points_samp.sf$education[treated == 1] <- 0.7 + rnorm(NTr, mean = 0, sd = .15)
points_samp.sf$education[treated == 0] <- 0.6 + rnorm(NCo, mean = 0, sd = .15)
```

The treated observations were assigned 70% by default and received a random draw from a normal distribution with mean zero and a standard deviation of 0.15. The few observations that ended up having values above 1 were assigned 1.00 by hand.
Before this assignment, we had to make use of the `assign_treated()` function of the `SpatialRDD` package which assigns each point in space its treatment status. Figure 2 visualises this assignment next to the five border segment categories that were created using the `border_segment()` function. These are lateron going to be used as dummies in the regressions of the parametric specifications which thus are the equivalent to a "within estimator" using these categories

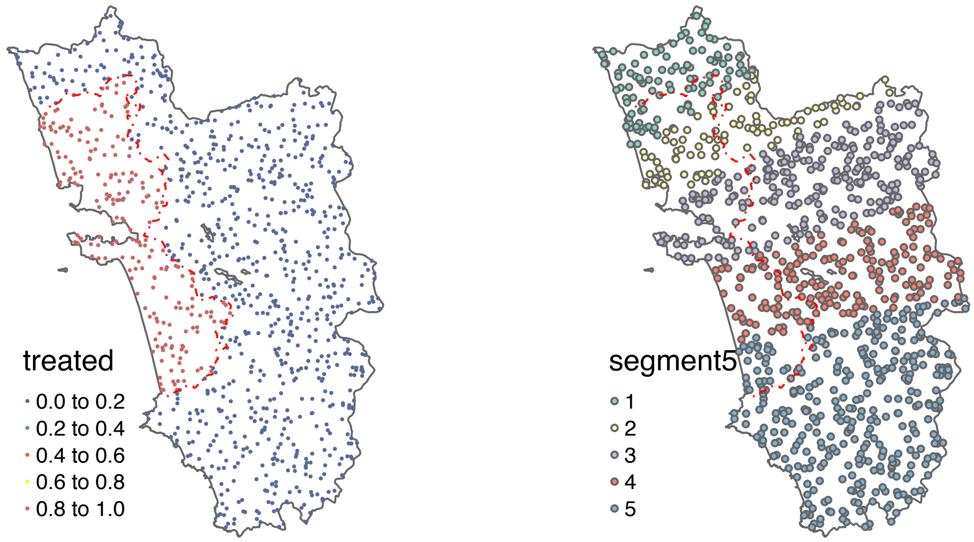


Figure 2: Visualisation of the assigned 'treated' categories and the created border segments.

as fixed effects levels. The analysis of only the variation within the segments is generally alleviating omitted variable problems. In our simulation study, though, both the segment dummies/fixed-effects and polynomials in X- and Y-coordinates to control for smooth functions of geographic location [as proposed by Dell, 2010] are not going to impact the point estimate for our treated dummy. This is simply due to the fact that we introduced a data generating process (DGP) that is spatially uniform in each of the two areas. In this simulation section we on purpose create data with such a homogeneous and very straightforward structure in order to illustrate the functioning and performance of the different spatial RDD specifications that we discussed in section 2.

The border segments have a length of approximately 26 kilometres each and are displayed in Figure 2. These are more or less arbitrarily chosen by the researcher and thus give room for what Gelman and Loken [2014] call "researcher degrees of freedom". The volitality that is induced by the choice of different border segments is illustrated in detail in section 5, when we replicate the study from Keele and Titiunik [2015]. This problem is aggravated when one considers all the different combinations that can be made with different clustering levels for the standard errors [even though being perfectly consistent with what e.g. Abadie, Athey, Imbens, and Wooldridge, 2017, suggest].

3.1 Parametric Specification

As a first step we naturally run the parametric specification on both the full, and a restricted sample that only includes observations within a certain bandwith of the RD border. Showing the robustness towards this arbitrary choice for the bandwith is of course redundant, as we know that it is not going to change a lot since we induced a uniform DGP. Table 1 shows the expected effects that correspond closely to what we would expect. In the treated area the literacy rate is around 10 percentage points higher and from the two univariate columns we can see from the constant that the average in the non-treated area is around 60 percent. All of the coefficients are highly significant.

Table 1: Point estimates for the parametric specification.

	Dependent variable:							
	education							
	Full Sample				3km Bandwidth			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treated	0.115	0.116	0.127	0.126	0.137	0.136	0.123	0.123
	(0.012)	(0.012)	(0.018)	(0.018)	(0.026)	(0.026)	(0.031)	(0.032)
Constant	0.590	0.594	0.587	0.641	0.582	0.571	0.608	0.577
	(0.005)	(0.015)	(0.008)	(0.053)	(0.017)	(0.028)	(0.048)	(0.148)
Segment FE	NO	YES	NO	YES	NO	YES	NO	YES
Polynomial	NO	NO	YES	YES	NO	NO	YES	YES
Observations	1,000	1,000	1,000	1,000	162	162	162	162

Note: Columns (1) and (5) show that education in the non-treated areas is around 0.6. Knowing our DGP, this is exactly what we should expect. Regardless of the inclusion of a polynomial that controls for the X and Y coordinates in space, and the inclusion of segment dummies/fixed-effects, the treated areas have on average around 10 percentage points more education. Since the data was simulated to be distributed evenly across space, this is what we should expect.

This is the most commonly used specification, especially in Economics, and was put forward by Dell [2010]. As we see here, it is perfectly able to depict one homogenous treatment effect by averageing over the full RD border[3]. This is also typically the thing that the researcher is most interested in. Furthermore it can be easily estimated by OLS.

As has been remarked by e.g. [Keele and Titiunik, 2015], and recently demonstrated by Lehner [2019a], this specification masks any heterogeneity in terms of the treatment intensity alongside the border. Something the researcher might be highly interested in, especially when it comes to determining certain segments that are better suited to estimate an average treatment effect - as an additional step after a set of balancing checks on covariates [see e.g. Keele and Titiunik, 2016]. Therefore this specification might in some cases also produce false positives, i.e. significant results when there are none, for example when there are certain segments with a very strong positive effect even though most of the segments have a negative effect. One such example is constructed in subsection 3.5. We will explore this further in section 5. Another drawback is induced by the fact that we need distributional assumptions on the parameters, something that is not the case for the non-parametric specifications and thus eliminating concerns of spatially correlated errors [e.g. very recently raised by Kelly, 2019].

For these reasons we propose to rely on other metrics, in addition to these parametric specifications, and most importantly to visualise the data properly with spatial interpolations. This also somewhat precludes the need for large sets of robustness checks.

[3]In fact due to the inclusion of the segment dummies, regression produces a weighted average over all segments, conditional on control variables [see e.g. Angrist and Pischke, 2008].

3.2 Nonparametric Specification: KT15

In this section we are going to explore deeper the properties and performance of the non-parametric version of a spatial RDD that was put forward by Keele and Titiunik [2015]. We arbitrarily chose to demonstrate this with 30 border segments, for each of which we will obtain a point estimate and a confidence interval. In the spirit of a "moving window" alongside the cutoff, we then obtain a set of point estimates that potentially allow us to infer something about a treatment heterogeneity[4]. All this is done on a set of points that we have to extract from the cutoff with the function `discretise_border()`. These boundarypoints we then can feed into the `SpatialRD()` function together with the data and some additional parameters. As described in section 2, the point estimates are obtained via a local linear regression with a two-dimensional score. We use a data driven bandwidth selection and compute both conventional and robust confidence intervals Calonico et al. [2014]. Through the use of a triangular kernel to weigh observations, those points that are closer to the RD border receive more importance than the ones who are farther away when it comes to determine the point estimates. As can be seen from the results in Table 2, and the corresponding visualisation in Figure 3, some of the confidence intervals are very broad and every now and then a point estimate even changes its sign. This is of course mostly due to the small sample size, at least as compared to the "classic" RD literature. The main point of this spatial RD exercise on a generated set of data with a known jump at the border is, that just because some of the point estimates don't seem to confirm a discontinuity, this doesn't prove that there is no discontinuity at all. Further checks, aside from the inspection of spatially interpolated plots, are needed in order to gain more information.

Table 2: Nonparametric point estimates on simulated data with flexible bandwidth (bw in km).

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.14	90	84	20.60	0.00	0.29	-0.03	0.33
2	0.24	57	53	13.10	0.03	0.45	-0.02	0.50
3	0.25	92	58	14.80	0.06	0.43	0.04	0.49
4	0.22	123	87	22.30	0.08	0.37	0.07	0.41
5	0.23	96	39	16.90	0.02	0.44	-0.02	0.50
6	0.22	138	81	19.70	0.06	0.38	0.04	0.43
7	0.20	145	101	20.10	0.08	0.33	0.06	0.38
8	0.24	90	51	14.90	0.10	0.38	0.09	0.43
9	0.22	64	49	12.60	0.08	0.35	0.07	0.41
10	0.19	62	88	13.90	0.03	0.35	0.01	0.41
11	0.18	39	58	11.30	-0.03	0.39	-0.07	0.45

[4]As this estimation is only carried out on one single dependent variable, just on different subsets of the data, and not on numerous different outcomes, we don't run into multiple testing problems here. Conceptionally the local estimation that we also use to obtain the point estimates works in a similar fashion, moving along the data with a pre-determined window and deliver a smooth curve. This is further explored when we do Monte Carlo simulations. It will be shown that the effect curve that we obtain is closely approximated within just a few iterations, and that a Bonferroni correction would be way to punishing and eliminate any significance, even though we know from the DGP that the effect is there.

Table 2: Nonparametric point estimates on simulated data with flexible bandwidth (bw in km). *(continued)*

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
12	-0.23	40	24	8.90	-0.75	0.30	-0.82	0.38
13	-0.14	59	22	10.20	-0.77	0.50	-0.81	0.58
14	-0.02	55	26	10.50	-0.30	0.26	-0.33	0.33
15	-0.16	26	20	8.40	-0.44	0.12	-0.53	0.12
16	0.49	18	26	8.10	0.03	0.95	0.03	1.09
17	0.45	18	60	10.40	0.01	0.89	0.00	1.01
18	0.15	17	41	9.30	-0.11	0.40	-0.13	0.47
19	0.07	60	72	14.10	-0.10	0.24	-0.15	0.26
20	0.15	51	47	11.20	-0.02	0.32	-0.07	0.35
21	0.05	65	57	12.40	-0.08	0.19	-0.12	0.22
22	0.11	147	84	17.50	-0.03	0.25	-0.07	0.27
23	0.11	299	94	22.90	-0.02	0.23	-0.06	0.26
24	0.09	271	83	21.30	-0.04	0.22	-0.08	0.24
25	0.09	201	79	19.30	-0.07	0.24	-0.12	0.27
26	0.06	200	71	18.60	-0.12	0.24	-0.17	0.27
27	0.10	142	62	16.00	-0.08	0.28	-0.15	0.29
28	0.21	118	57	15.30	0.05	0.36	0.03	0.42
29	0.15	230	74	23.50	0.02	0.29	0.01	0.33
30	0.12	105	59	18.40	-0.02	0.26	-0.05	0.30
Mean	0.14	104	60	15.22	-0.08	0.36	-0.11	0.41

Note:
Ntr/Nco represent the number of treated/control observations on each side. Conventional and robust confidence intervals reported (lower/upper).

One obvious way is to just force the bandwith for each RD estimate to be of a certain large enough size, for we see in Table 2 that the confidence intervals that overlap with zero, and the negative point estimates, are obtain from comparatively small samples. This ”brute force bandwith” approach is of course ad-hoc, but Table 3 and the corresponding visualisation in Figure 4, demonstrate that we obtain estimates that are much closer to the true DGP. In addition we are in any case going to run other specifications, such as the nonparametric approach by Imbens and Wager [2019], in order to draw a bigger and more complete picture. The takeaway from this is, that when we fix the bandwith, the insignificant results are reduced in number because the data driven bandwith selection algorithms are not designed for these applications on small samples. Because such a forced widening of the bandwith would not have the effect of producing false positives on data where there in fact is no discontinuity, this approach seems useful to ”enhance” the credibiliyt of a nonparametric spatial RDD.

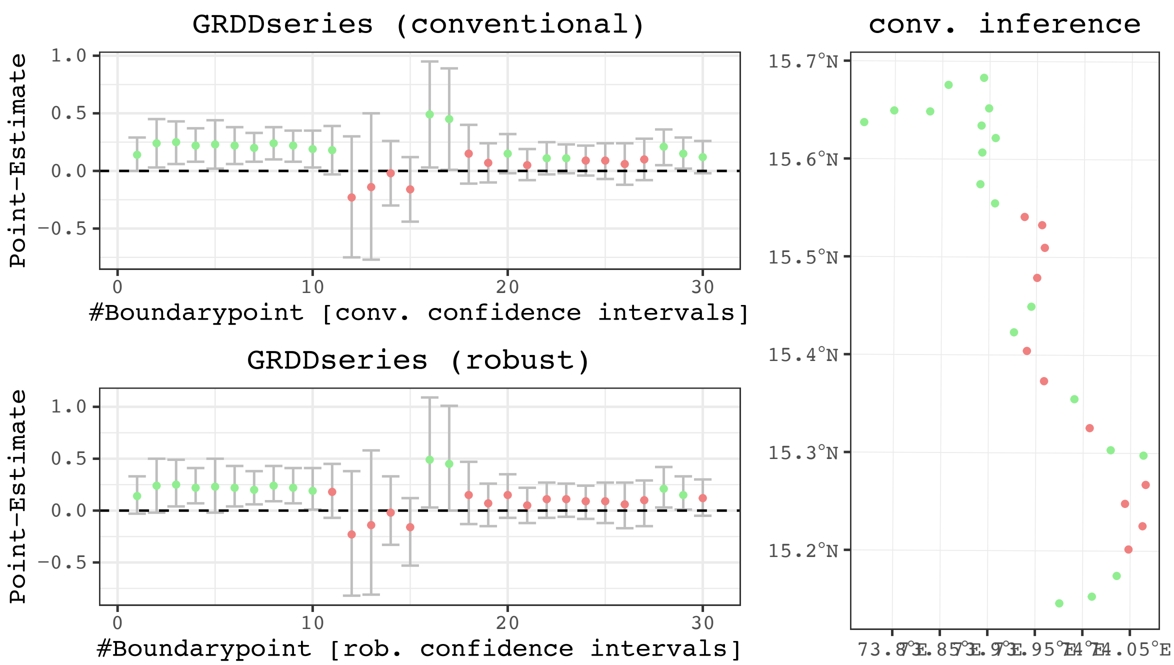


Figure 3: GRD results with data driven bandwith selection at each point. Some point estimates even dip into the negative, but it is visually inferable that there appears to be a discontinuity.

Table 3: Nonparametric point estimates on simulated data with forced bandwidth of 20km.

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.14	87	80	20	-0.01	0.29	-0.10	0.49
2	0.18	101	81	20	0.05	0.32	0.01	0.60
3	0.21	112	92	20	0.06	0.36	-0.02	0.50
4	0.24	112	75	20	0.08	0.40	0.00	0.55
5	0.24	118	65	20	0.06	0.42	-0.15	0.51
6	0.22	142	85	20	0.06	0.37	-0.08	0.51
7	0.20	143	100	20	0.08	0.33	0.02	0.47
8	0.22	153	102	20	0.10	0.33	0.06	0.47
9	0.19	149	108	20	0.08	0.30	0.09	0.46
10	0.14	146	116	20	0.01	0.26	-0.01	0.46
11	0.11	149	124	20	-0.01	0.22	-0.10	0.40
12	0.13	168	114	20	0.01	0.25	-0.26	0.46
13	0.15	176	110	20	0.01	0.29	-0.28	0.64
14	0.13	178	117	20	0.02	0.25	-0.11	0.42
15	0.12	164	120	20	0.00	0.24	-0.15	0.30
16	0.13	151	121	20	0.00	0.26	-0.04	0.59

Table 3: Nonparametric point estimates on simulated data with forced bandwidth of 20km. *(continued)*

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
17	0.10	126	128	20	-0.04	0.25	-0.03	0.71
18	0.09	128	130	20	-0.03	0.22	-0.10	0.36
19	0.08	127	128	20	-0.04	0.20	-0.15	0.29
20	0.10	158	120	20	0.00	0.21	-0.10	0.28
21	0.11	173	106	20	0.01	0.20	-0.10	0.24
22	0.11	197	93	20	-0.02	0.23	-0.20	0.31
23	0.10	227	84	20	-0.03	0.24	-0.11	0.32
24	0.09	241	80	20	-0.05	0.23	-0.09	0.35
25	0.09	213	82	20	-0.06	0.24	-0.20	0.35
26	0.07	237	74	20	-0.10	0.23	-0.31	0.39
27	0.10	220	73	20	-0.05	0.24	-0.26	0.36
28	0.15	203	66	20	0.02	0.28	0.01	0.45
29	0.18	163	62	20	0.02	0.33	0.04	0.66
30	0.12	117	63	20	-0.01	0.26	-0.11	0.32
Mean	0.14	159	97	20	0.01	0.28	-0.09	0.44

Note:
Ntr/Nco represent the number of treated/control observations on each side. Conventional and robust confidence intervals reported (lower/upper).

The tests for sorting around the cutoff [McCrary, 2008] (the null is "no sorting") and for the continuity of baseline covariates Canay and Kamat [2018] (the null is "continuity", which is what since we strive for "balancing") seem to be somewhat informative, even though the McCrary test doesn't deliver results for every point. The problem with the permutation test by Canay and Kamat [2018] is that it is computationally very demanding and thus imposes a big burden when we run such a test for each of our dozens of RD estimates.
Finally it has to be noted that the mean over all point estimates, which is 0.14 for the baseline GRDDseries and 0.14 for the one with a forced bandwith of 20km, does a very good job of approximating the (conditional) average treatment effect in the DGP. Thus lending credibility to this approach and further demonstrating that a handful of "outlying point estimates" are not proof for the non-existence of a discontinuity in the data. These point estimate means are also close to the treated coefficients of the nonparametric specification in the spirit of Dell [2010].

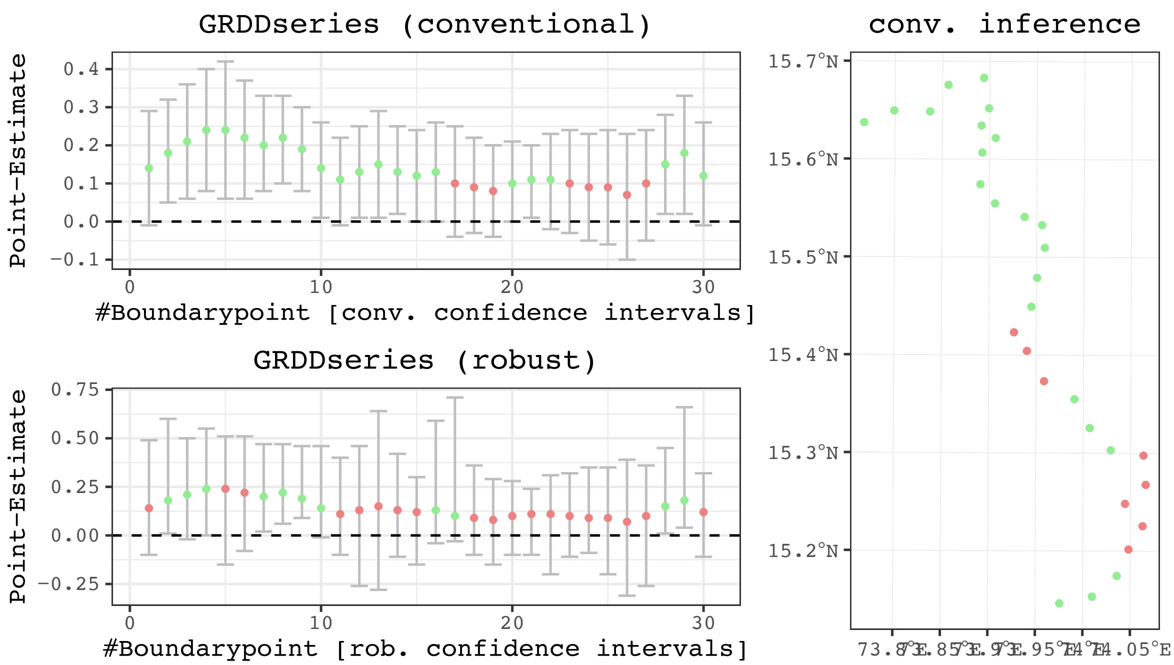


Figure 4: GRD results with a fixed bandwith of 20km leading to point estimates that are closer to the DGP and oddly behaving robust confidence intervals.

3.3 Nonparametric Specification: IW19

Table 4 confirms the average treatment effect (ATE) that is consistent with the DGP and the two previous specifications that we saw. The single boundary point that we chose for estimation is slightly insignificant. Having in mind the GRDDseries visualisation from before, this should not be much of a concern.
Aside from the fact that the minimax linear estimator of Imbens and Wager [2019] is well defined even for cases when the running variable is not continuous, another useful feature in a spatial RDD setting is the possibility to obtain a weighted average treatment effect. As has been shown in Lehner [2019a], this provides a way to spot supposedly "weaker" outcome variables. An additional approach to the one that we saw in the last chapter.

	V1	outcome	point estimate	conf interval	max bias	sampling std err
1	WATE	education	0.141	(0.080, 0.202)	0.014	0.028
2	CATE	education	0.150	(-0.057, 0.357)	0.098	0.066

Table 4: Weighted ATE and Conditional ATE for the optimized approach by Imbens and Wager [2019]

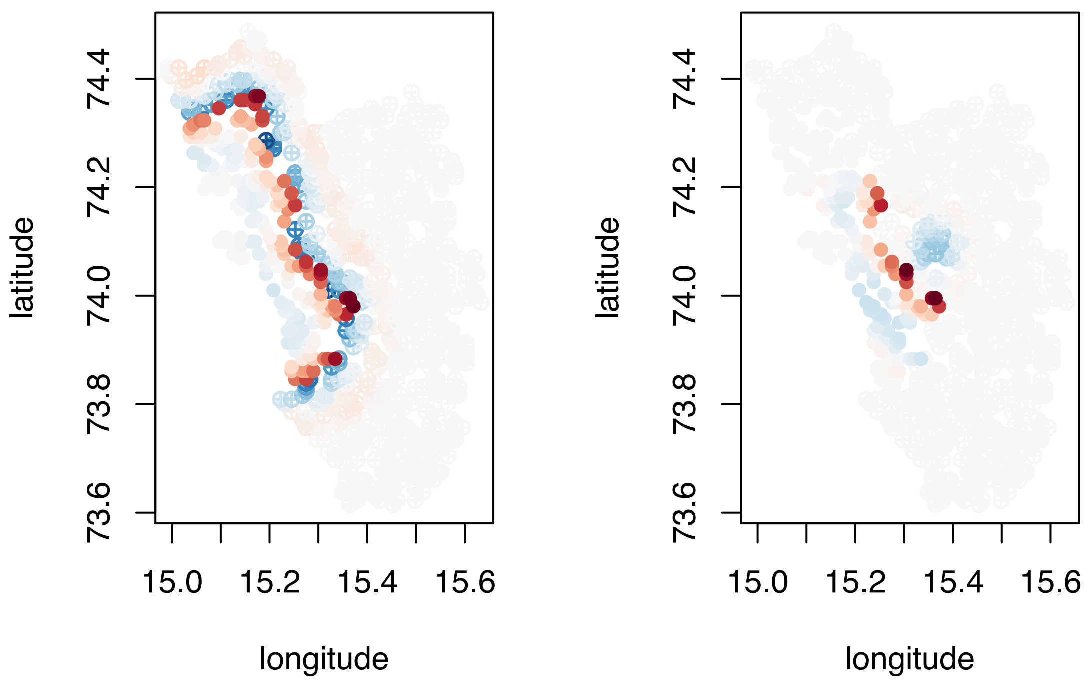


Figure 5: IW19 results visualised. The weighted effect visualises nicely the discontinuity (left). One specific borderpoint (right)

3.4 Robustness

As we discussed earlier already several times, due to the spatially homogenous nature of our simulated data, it is redundant to do most of the usual robustness checks for spatial RDD's. For the parametric specification the inclusion of additional controls, different bandwiths, or the use of "placebo outcomes" as dependent variables are obviously not going to show any effect. Thus we restrict the exercises in this section to arguably the most important robustness checks: shifting the border in several directions and show that there is no "jump" in the data.

3.4.1 Placebo Borders

For the parametric specification, researchers typically just add or substract a certain distance for each unit of observation from the "distance to cutoff" variable, and then re-assign the treatment status for observations where the sign changed. This is provides usually sufficently accurate results, but is somewhat a blackbox because it is not possible to visualise the new border with this approach. Furthermore, with this "add/substract approach" it is not possible to create placebo borders to do robustness checks with nonparametric specifications. Thus the approach that is outline in the following lines has to be undergone in any case.
`SpatialRDD` implements "placebo bordering" with a textbook affine transformation of the type $f(x) = x\mathbf{A} + b$, where the matrix $\mathbf{A}$ is the projection matrix to shift, (re-)scale, or rotate the cutoff, where the latter is achieved via a standard rotation matrix:

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

In this example our original cutoff was simply shifted along the X- and Y- axis and then scaled the following way:

```
placebo_border(cutoff,
               operation = c("shift", "scale"), shift = c(-7000, -3000), scale = .85)
placebo_border(cutoff,
               operation = c("shift", "scale"), shift = c(4000, 2000), scale = 1.1)
```

This is one of the applications where it now pays off that we are working in a local UTM projection system, as the shifting of borders can be done in meters and calculations are more accurate. The new treatment status is assigned via the creation of a polygon with `cutoff2polygon()`, one for each scenario, which is visualised in Figure 6. More information can be found in the `SpatialRDD` documentation [Lehner, 2019b].

For the first placebo border, the inward shift, we have 106 treated observations which means that 105 have gone to "non-treated". For the outward shift we have 320 treated observations, meaning that 109 went from untreated in the baseline scenario to treated.

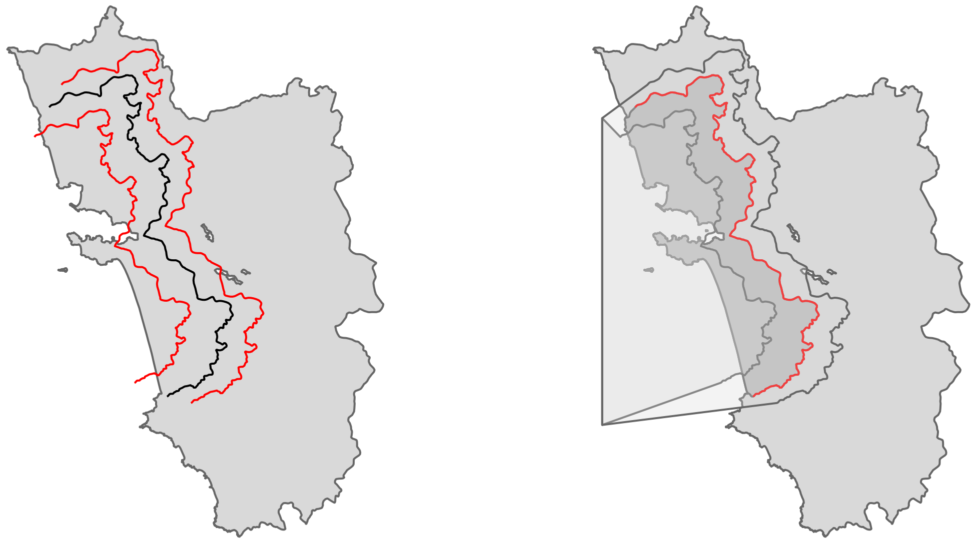


Figure 6: Illustration of the shift and scale operations (left) and the newly created polygons to re-assign the treatment status (right).

3.5 Examples

In this section we are going to see some artificially set up simulations for our outcome variable, education as measured by literacy rates. These are obviously unrealistic, but suffice to demonstrate some of the problems that one might run into. Robustness checks are omitted in this section for brevity. Needless to say that it is easy to also construct "robustness check proof" examples of e.g. "false positives", referring to situations where the estimates suggest there is an effect, even though there is none.

3.5.1 Parametric: False Positives

Another thing that has to be noted here is that, looking at columns (6) and (7), the polynomial in the X- and Y-coordinates seemingly does a better job than the border segment dummy/fixed-effect in "catching" that there might be something odd going on here. This is due to the fact that we created our positive outliers all within one segment category, and thus regression just averages over it.

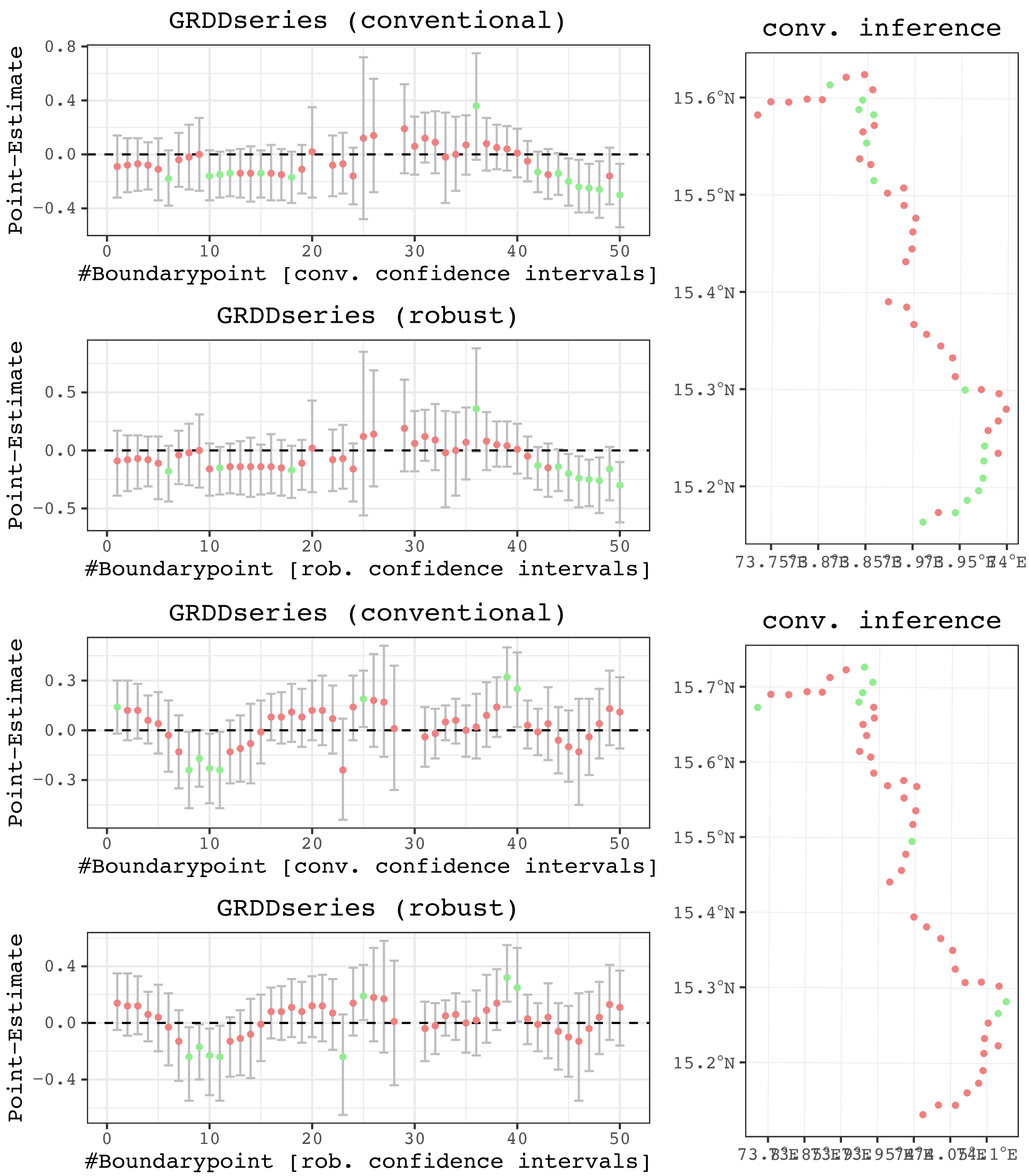


Figure 7: No effect on both placebo borders, thus allowing to visually infer robustness.

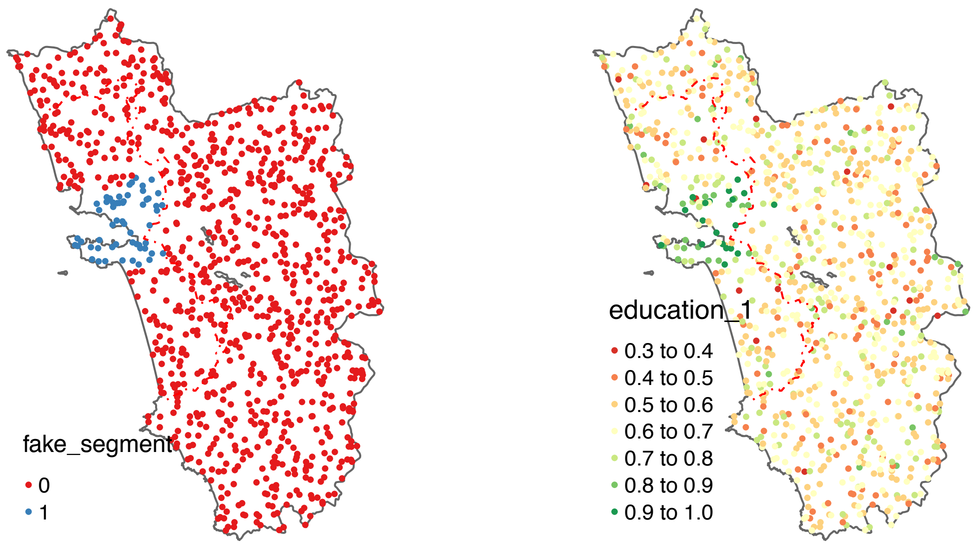


Figure 8: The newly simulated data for our "false positive example" with a mean of 0.8 with respect to 0.6 anywhere else.

Table 5: Point estimates for the parametric specification ("false positive example").

	Dependent variable:							
	education_1							
	Full Sample				3km Bandwidth			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treated	0.066	0.068	0.058	0.052	0.075	0.068	0.050	0.062
	(0.009)	(0.009)	(0.013)	(0.013)	(0.020)	(0.019)	(0.023)	(0.023)
Constant	0.600	0.579	0.601	0.523	0.604	0.574	0.655	0.598
	(0.004)	(0.011)	(0.006)	(0.037)	(0.013)	(0.020)	(0.036)	(0.107)
Segment FE	NO	YES	NO	YES	NO	YES	NO	YES
Polynomial	NO	NO	YES	YES	NO	NO	YES	YES
Observations	1,000	1,000	1,000	1,000	162	162	162	162

Note: The parametric specification is not able to "detect" that there is no true effect alongside the cutoff. Looking at column (7), the inclusion of a polynomial that controls for the X and Y coordinates in space seemingly does the best job in detecting an odd spatial pattern.

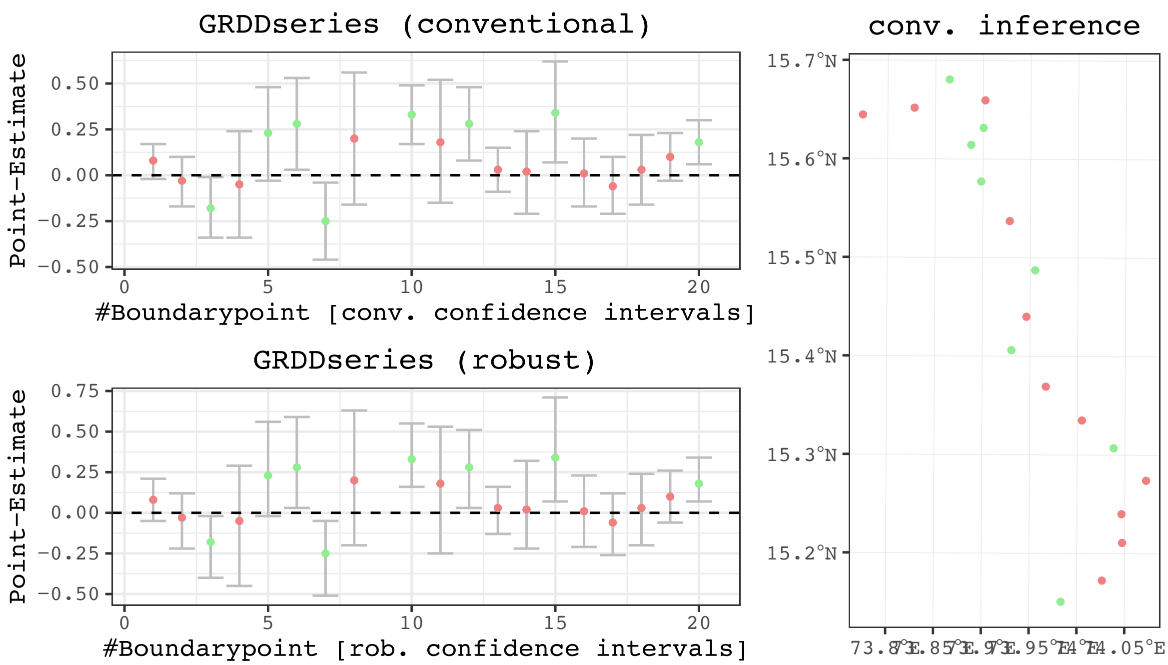


Figure 9: From the plots of the nonparametric estimates for each boundarypoint we can infer that there in fact seems to be no discontinuity, confirming the "false positive" already visually.

3.5.2 Parametric: False Negatives

As was already noted in chapter 2, it seems to be the case that we have to be careful with the type of polynomial we include in order to control for the position in space as suggested by Dell [2010]. In accordance with the logic of Gelman and Imbens [forthcoming], it seems to be warranted to only include second order polynomials also in the parametric specifications for spatial RDDs. They show for "classic" RDDs, for example, that high-order polynomials may give large influence to observations far from the boundary, and thus may lead to unreliable performance. One could construct an example where including a third order polynomial instead of a second order polynomial might create an insignificant point estimate. This is not to say that showing robustness towards different polynomial degrees is important, on the contrary, but that additional approaches should be consulted

As a last simulated example, let us set up a "false negative situation" where there is actually a discontinuity alongside the majority of the border, but there is one small part of the study area where the outcomes are uniformly distributed with a lower mean in the outcome, say 0.4. The DGP for the rest of the area is exactly the same as in our very first example as can bee seen in Figure 10. From this example we can conclude, that the polynomial spatial RD specifications are not very prone to run into "false negative" problems, i.e. not detecting an effect even though it is there. It is interesting to note, that from column (7) we can see that the polyonmial in X- and Y- coordinates seemingly does the best job in controlling for for the new lower average within the bandwith.

Another thing that has to be noted here is that, looking at columns (6) and (7), the polynomial in the X and Y coordinates seemingly does a better job than the border segment dummy/fixed-effect in "catching" that there might be something odd going on here. This is

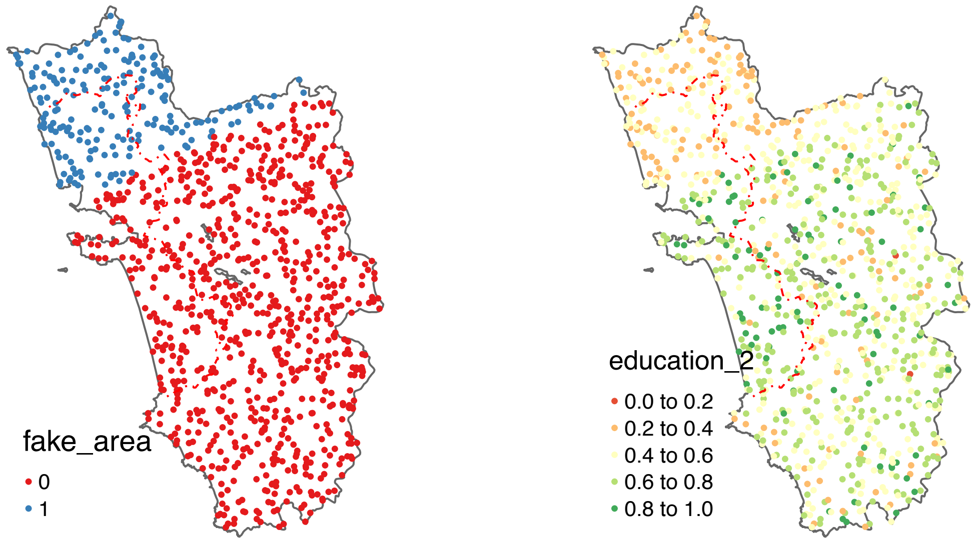


Figure 10: The newly simulated data for our "false negative example" with a mean of 0.4 in the marked area.

again not too surprising, though, as all the data that was generated with a uniformly lower value for education is all location with two specific segments. Thus regression just drags down the point estimate as it produces a weighted average over all segments.

Table 6: Point estimates for the parametric specification ("false negative example").

	Dependent variable:							
	education_2							
		Full Sample			3km Bandwidth			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treated	0.030	0.076	0.084	0.100	0.105	0.091	0.071	0.080
	(0.013)	(0.011)	(0.017)	(0.016)	(0.025)	(0.019)	(0.024)	(0.023)
Constant	0.563	0.379	0.554	0.339	0.519	0.381	0.542	0.347
	(0.006)	(0.014)	(0.007)	(0.047)	(0.017)	(0.021)	(0.037)	(0.109)
Segment FE	NO	YES	NO	YES	NO	YES	NO	YES
Polynomial	NO	NO	YES	YES	NO	NO	YES	YES
Observations	1,000	1,000	1,000	1,000	162	162	162	162

Note: As can be seen from column (1), univariate OLS almost cannot detect the effect anymore. Once the spatial control and/or the segment dummies for the exploitation of within variation are introduced, the effect is perfectly visible again. From this we can see that the parametric spatial RDD specification is not very prone to run into "false negative" issues.

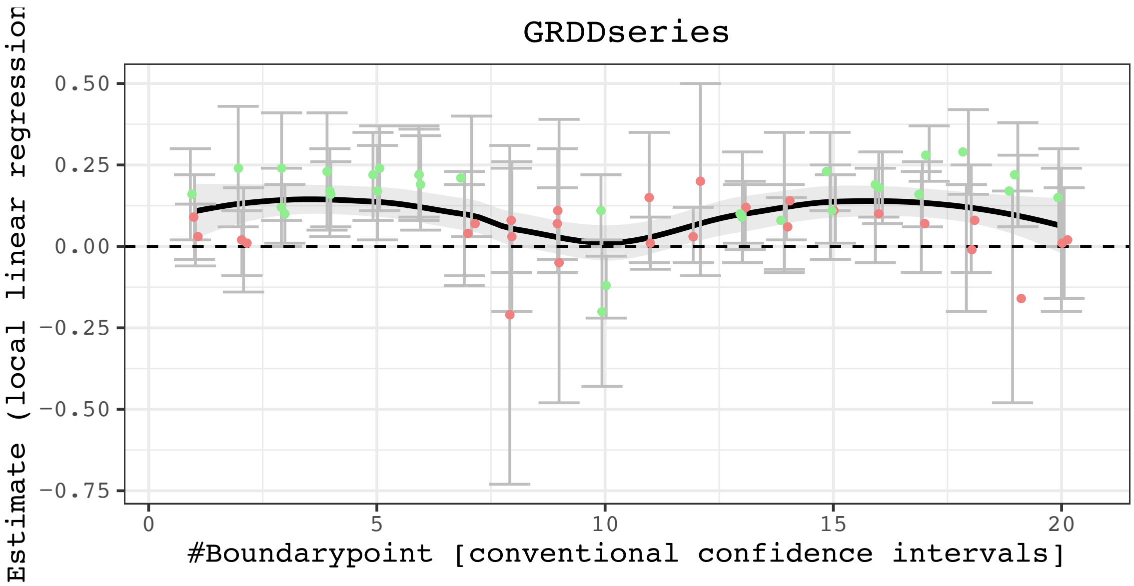


Figure 11: All points out of three simulated runs with jittered confidence intervals. A local linear fit across all nonparametric estimates approximates the true 'treatment effect curve' already quite well. Even though it have been only three runs of simulations.

4 A (Spatial) Monte Carlo Exercise

In this simulation exercise we are simply going to use the same data generating process from above, constructing a discontinuity with a ten percentage point jump in literacy at the cutoff. We then iterate over these simulations numerous times and report the averages. In fact it turns out that only few simulation runs are needed in order to approximate the treatment effect curve alongside the RD cutoff. The plotted series of just three runs can be seen in Figure 11. It is not unusual that the confidence intervals are overlapping with zero, or sometimes even dip into the negative, even though the "true DGP" has a discontinuity baked in. In Table 7, where we report the averages for each point, it can be seen that this imprecision is of course due to the small sample size. For some borderpoints we on average have even less than 50 observations on each side. Obviously the nonparametric local linear estimation techniques, include the identification assumptions, were designed for very large sample sizes that are in the thousands. Nevertheless it can be inferred from these exercises that the nonparametric spatial RD estimations perform fairly well in identifying the discontinuity alongside the cutoff.

Summing over all point estimates in Table 7 we obtain an average effect of 0.11, which is very close to the true means.

Table 7: Averages of all obtained results over all runs per every point. Simulations with only three runs already approximate the true DGP well. Some point estimates turn negative and many confidence intervals overlap with zero.

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.09	78.00	76.67	18.57	-0.03	0.22	-0.06	0.24

2	0.09	68.67	47.33	12.67	-0.06	0.24	-0.09	0.26
3	0.15	113.00	73.00	19.93	0.03	0.28	0.01	0.31
4	0.19	92.33	46.33	15.80	0.05	0.32	0.03	0.37
5	0.21	84.33	57.33	14.17	0.07	0.34	0.06	0.39
6	0.21	56.33	49.67	11.77	0.07	0.36	0.07	0.41
7	0.11	40.67	63.67	11.50	-0.06	0.27	-0.09	0.31
8	-0.03	37.33	44.33	9.43	-0.34	0.27	-0.39	0.30
9	0.04	63.00	46.00	11.23	-0.20	0.29	-0.24	0.33
10	-0.07	33.00	23.67	8.27	-0.22	0.07	-0.26	0.07
11	0.08	30.00	37.00	9.00	-0.06	0.22	-0.09	0.24
12	0.12	20.50	51.50	9.95	-0.07	0.31	-0.09	0.36
13	0.10	47.00	57.00	11.87	-0.02	0.23	-0.04	0.26
14	0.09	126.67	77.67	15.53	-0.04	0.23	-0.09	0.25
15	0.15	144.67	78.33	16.77	0.03	0.27	0.00	0.31
16	0.16	213.00	71.00	18.57	0.04	0.27	0.01	0.30
17	0.17	154.00	70.00	16.73	0.06	0.29	0.04	0.32
18	0.12	90.33	46.67	12.67	-0.04	0.29	-0.08	0.31
19	0.08	102.00	41.33	13.90	-0.12	0.28	-0.16	0.31
20	0.06	60.00	36.33	12.97	-0.12	0.24	-0.17	0.26

In Figure 12 we plot the simulations of 15 runs of simulations. It shows that, as the number of iterations goes to infinity, the treatment effect curve approximates a straight line.

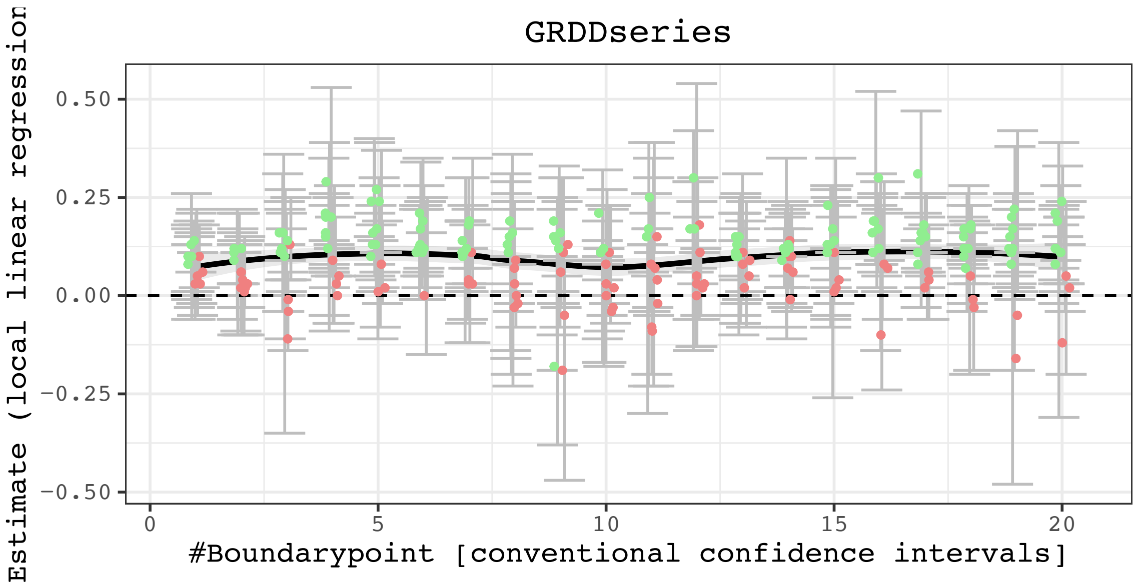


Figure 12: 15 runs of simulations are already enough to approximate the true treatment effect curve very well.

5 Replication of KT15

In order to illustrate all the points that have been raised so far, we will replicate the example of Keele and Titiunik [2015] that has also been used by Imbens and Wager [2019] in order to illustrate their "optimized approach" to multi-dimensional RDD's.
A school district in New Jersey is used to investigate the effect of televions advertising on voter turnout in the 2008 presidential elections. Half of this district belongs to the Philadelphia media market and the other half to the New York media market. The former was subject to heavy campain advertising before the 2008 elections whereas the New York half was not. Since no other major boundary coincides with this media market boundary, thus not creating a compound treatment effect problem, this could be seen as a natural experiment and thus lends itself for an identification strategy that relief on a spatial RDD, using the media boundary as cutoff.

Both papers conclude that there is no observable robust effect. This study agrees with their general result that there is no strong evidence for an effect of advertising on turnout for the 2008 elections. The conclusion, though, is achieve by different means. We argue that the estimated effect does at least appear to be "borderline significant" but the balancing checks of observable variables such as age, race, and income, do suggest that there is some underlying omitted variable that is driving the observed coefficients, and not the television advertising. In light of what we found out by running the extended spatial RDD arsenal on our simulated data, we argue that it cannot be concluded solely from the obtained point estimates that there was no effect of advertising on 2008 electoral turnout in the study area.

In the last part of this chapter we are going to illustrate some of the troubles researchers might run into when there is a large set of segment lengths interacted with different cluster levels to choose from. This constitutes what Gelman and Loken [2014] call "researcher degrees of freedom".

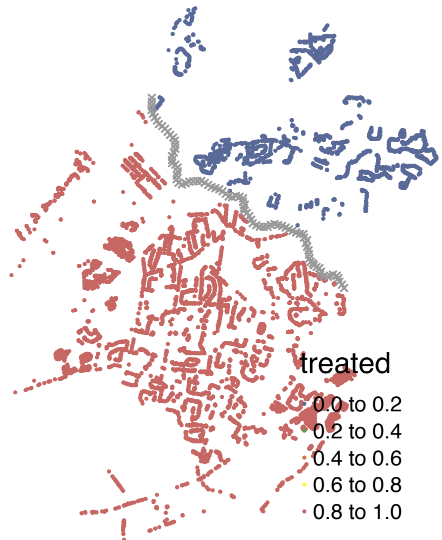


Figure 13: The dataset of Keele, Titiunik (2015) visualised.

5.1 Balancing Checks

The following four tables of balancing checks on important covariates in the data show significant jumps within narrow bandwiths aroun the RD treshhold. The OLS estimations are done via the simple parametric specification from Dell [2010]. The number of females and political affilation (dem standing for democrats) seems to be rather balanced. Age, education, income, and race paint a different picture. Households in the treated areas are proportionally earning more, having higher income, and are more educated. Admittedly, the point estimates go down as we move closer to the border (still being statistically significant). This is typically seen as an indication that the sample is more comparable around the RD cutoff than in the whole sample. Something that would an argument in favour of carrying out a spatial RDD. Nevertheless the jumps seem to be too significant as to conclude that any effect that we might find can be causally attributed to tv advertising alone.

Table 8: Balancing Checks

	Dependent variable:											
	white				age				educ			
	full	3km	2km	1km	full	3km	2km	1km	full	3km	2km	1km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
treated	0.087	0.116	0.127	0.318	2.161	1.382	−0.880	1.278	0.258	0.424	0.320	0.222
	(0.006)	(0.008)	(0.011)	(0.017)	(0.216)	(0.269)	(0.388)	(0.547)	(0.007)	(0.008)	(0.011)	(0.017)
Constant	0.577	0.567	0.591	0.395	46.946	47.451	49.578	46.721	15.466	15.340	15.491	15.574
	(0.005)	(0.006)	(0.008)	(0.013)	(0.169)	(0.196)	(0.284)	(0.418)	(0.005)	(0.006)	(0.008)	(0.013)
Segment FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
No. treated	15000	7888	3997	1853	15000	7888	3997	1853	15000	7888	3997	1853
Observations	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183

Note:

Table 9: Balancing Checks

	Dependent variable:											
	female				dem				income			
	full	3km	2km	1km	full	3km	2km	1km	full	3km	2km	1km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
treated	−0.003	−0.012	−0.001	0.013	−0.013	−0.012	−0.021	−0.024	0.036	0.039	0.021	0.007
	(0.007)	(0.008)	(0.012)	(0.018)	(0.006)	(0.008)	(0.011)	(0.017)	(0.0004)	(0.0004)	(0.001)	(0.001)
Constant	0.486	0.491	0.479	0.453	0.338	0.339	0.346	0.339	0.106	0.100	0.121	0.139
	(0.005)	(0.006)	(0.009)	(0.014)	(0.005)	(0.006)	(0.008)	(0.013)	(0.0003)	(0.0003)	(0.0004)	(0.001)
Segment FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
No. treated	15000	7888	3997	1853	15000	7888	3997	1853	15000	7888	3997	1853
Observations	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183

Note:

Table 10: Balancing Checks FE

	Dependent variable:											
	white				age				educ			
	full	3km	2km	1km	full	3km	2km	1km	full	3km	2km	1km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
treated	0.122	0.137	0.248	0.406	1.053	0.765	2.477	3.652	0.278	0.387	0.294	0.064
	(0.010)	(0.010)	(0.013)	(0.019)	(0.336)	(0.353)	(0.478)	(0.635)	(0.010)	(0.010)	(0.012)	(0.018)
Constant	0.708	0.713	0.776	0.737	52.874	53.145	55.825	50.832	15.807	15.757	15.830	15.742
	(0.012)	(0.012)	(0.013)	(0.028)	(0.402)	(0.412)	(0.486)	(0.941)	(0.012)	(0.011)	(0.013)	(0.027)
Segment FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
No. treated	15000	7888	3997	1853	15000	7888	3997	1853	15000	7888	3997	1853
Observations	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183

Note:

Table 11: Balancing Checks FE

	Dependent variable:											
	female				dem				income			
	full	3km	2km	1km	full	3km	2km	1km	full	3km	2km	1km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
treated	−0.004	−0.012	0.014	0.023	−0.022	−0.023	0.003	0.017	0.021	0.026	0.011	−0.007
	(0.010)	(0.011)	(0.015)	(0.021)	(0.010)	(0.010)	(0.014)	(0.020)	(0.0005)	(0.001)	(0.001)	(0.001)
Constant	0.508	0.510	0.513	0.536	0.390	0.393	0.390	0.386	0.115	0.111	0.122	0.123
	(0.012)	(0.013)	(0.015)	(0.031)	(0.012)	(0.012)	(0.014)	(0.029)	(0.001)	(0.001)	(0.001)	(0.001)
Segment FE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
No. treated	15000	7888	3997	1853	15000	7888	3997	1853	15000	7888	3997	1853
Observations	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183	24,460	14,860	7,446	3,183

Note:

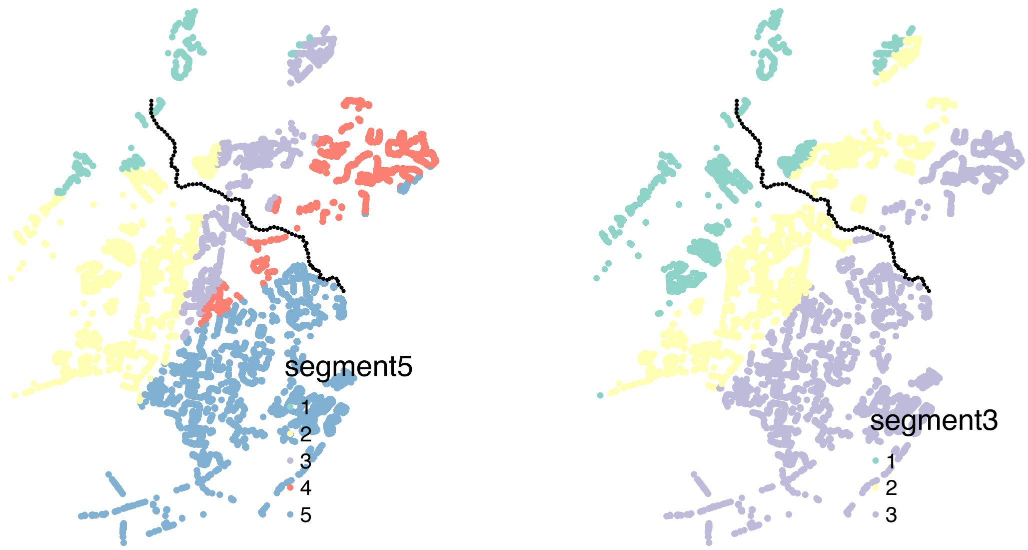


Figure 14: Two sets of border segment categories juxtaposed. 3 and 5 seem both reasonable choices. With a higher amount of segment - even though the distance still is quite long - the groups that are compared within segments take quite odd shapes.

5.2 Nonparametric SpatialRD

Similar to to our previous section we are going to report the point estimates, obtained over many boundarypoints across the media market boundary, both in terms of a GRDDseries and in the form of tables where every of these estimates is described in more detail.
From the maps we already saw that to the right side of the border we have no observations close to the cutoff in the control area to the North. This is problematic because due to the fact that we use a triangular kernel that weighs the impact of units by their distance to the border, this leads to imprecision in the point estimates. In fact once we control for that, by excluding points to the far right - and later also forcing the bandwith for each estimate to be five kilometres - we find that the discontinuity at this media market boundary seems to be much more robust than it has been claimed by Keele and Titiunik [2015] and Imbens and Wager [2019], carrying the analysis out on the exact same set of data. Most of the estimates don't even overlap with zero, and the ones that do are on the fringes. Having in mind the simulation exercises from above on "discontinuous data", these results seem to suggest that the outcome variable of 2008 election turnout seems to constitute a robust jump at the media market boundary.
Before we move to the parametric estimations of this study, presented in subsection 5.4, we first show two placebo border exercises that will show us that there is no jump to be found below or above the actual treshhold.

Table 12: GRD borderpoints for Keele and Titiunik [2015].

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.07	4951	5023	5.10	-0.01	0.15	-0.07	0.24

Table 12: GRD borderpoints for Keele and Titiunik [2015].
(continued)

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
2	0.12	2719	2880	3.60	0.01	0.22	-0.02	0.33
3	0.31	1478	452	2.00	-0.01	0.62	0.04	0.73
4	0.30	2057	1803	2.50	0.18	0.42	0.21	0.51
5	0.22	1075	1437	2.00	0.08	0.37	0.10	0.44
6	0.23	1081	1964	2.00	0.09	0.37	0.08	0.44
7	0.21	1325	1843	2.10	0.06	0.36	0.08	0.42
8	0.19	1323	1335	1.80	0.04	0.33	-0.01	0.34
9	0.14	1593	1072	1.70	-0.01	0.30	-0.05	0.35
10	0.01	1870	1072	1.80	-0.14	0.16	-0.19	0.14
11	-0.03	2063	2378	2.20	-0.16	0.09	-0.22	0.06
12	-0.09	1366	1925	2.00	-0.23	0.04	-0.29	0.02
13	-0.24	1338	1430	2.00	-0.37	-0.10	-0.46	-0.11
14	-0.24	1354	1511	2.10	-0.37	-0.10	-0.46	-0.11
Mean	0.09	1828	1866	2.35	-0.06	0.23	-0.09	0.27

Note:
NA

Table 12

Table 13: GRD borderpoints for Keele and Titiunik [2015], fixed bandwidth.

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.07	4915	4893	5	-0.01	0.15	-0.16	0.27
2	0.09	4981	5736	5	0.02	0.15	-0.05	0.27
3	0.13	5680	6890	5	0.07	0.19	0.13	0.43
4	0.16	6176	7711	5	0.10	0.21	0.21	0.46
5	0.15	6133	8999	5	0.10	0.20	0.23	0.45
6	0.15	6852	10540	5	0.10	0.20	0.24	0.44
7	0.16	7870	10384	5	0.11	0.21	0.23	0.42
8	0.15	8806	10615	5	0.11	0.20	0.17	0.35
9	0.14	9328	10677	5	0.10	0.18	0.02	0.20
10	0.13	9456	10944	5	0.08	0.17	-0.11	0.08
11	0.09	9356	12017	5	0.05	0.14	-0.22	-0.01
12	0.06	9329	12027	5	0.01	0.10	-0.28	-0.07
13	0.02	9258	11062	5	-0.02	0.07	-0.33	-0.13
14	-0.02	8622	10749	5	-0.07	0.03	-0.34	-0.12
Mean	0.11	7626	9517	5	0.05	0.16	-0.02	0.22

Note:
NA

Table 13
Figure 15
Figure 17

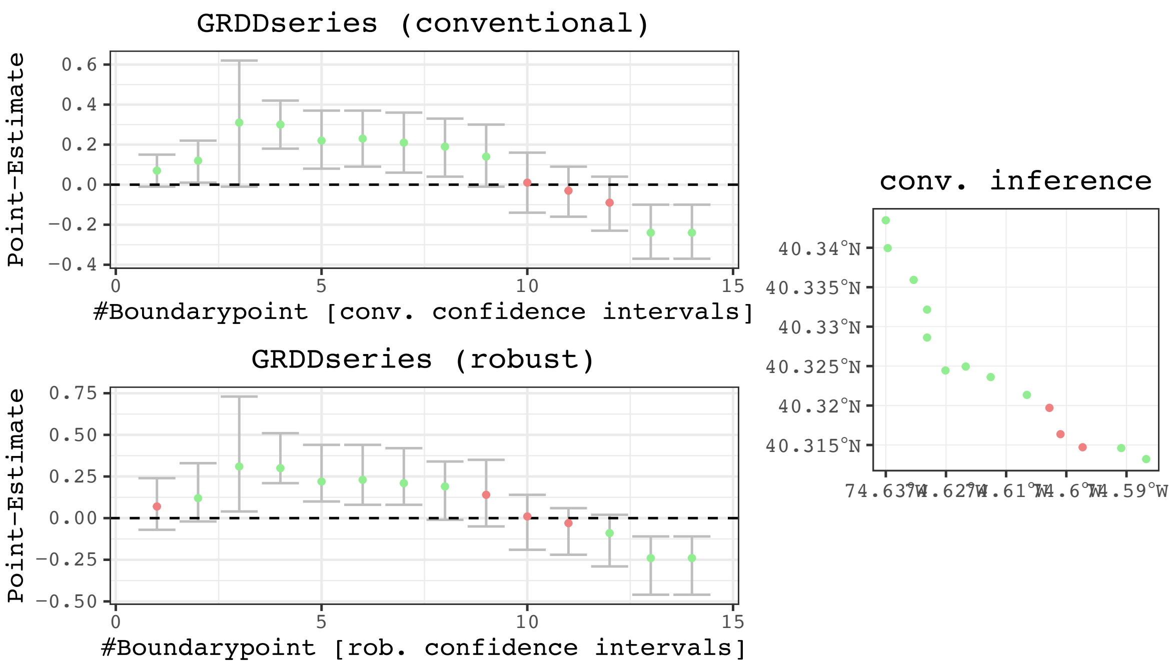


Figure 15: GRD results with data driven bandwith selection at each point. Some point estimates even dip into the negative, but it is somewhat visually inferable that there appears to be a discontinuity. To the right of the border we have fewer observations that are also quite far away from the border and thus receive low weights, leading to imprecision.

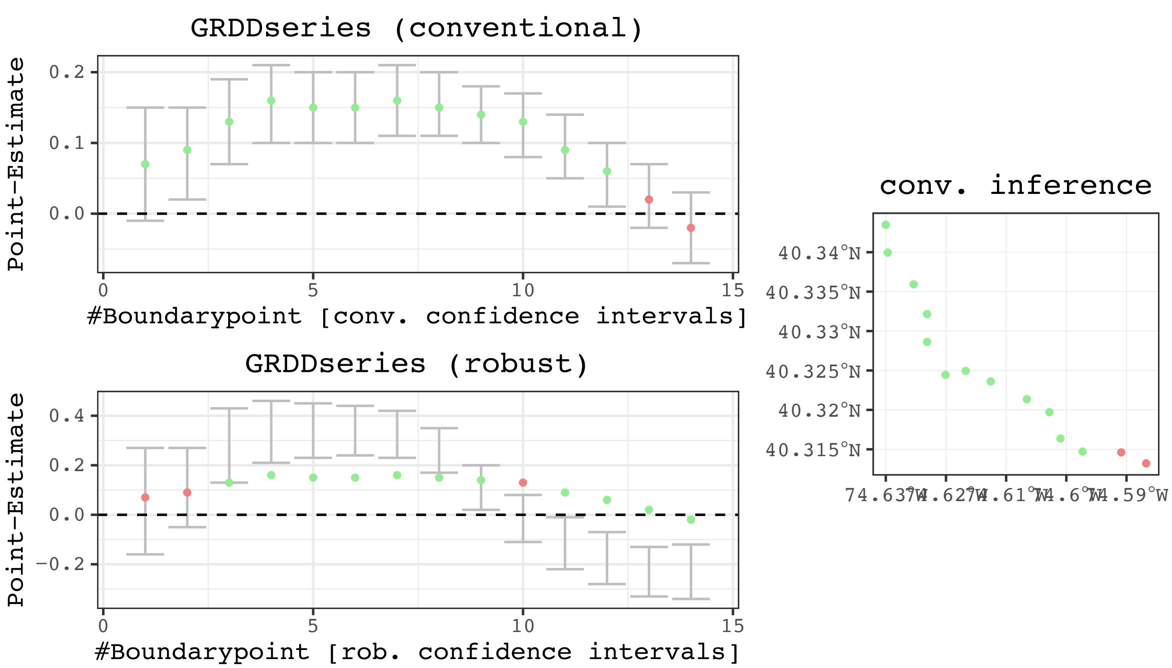


Figure 16: Once we force the bandwidth to be 5km, it becomes clear that there appears to be a discontinuity. Now even the right fringe with very few observations moves the upper end of its confidence intervals above zero.

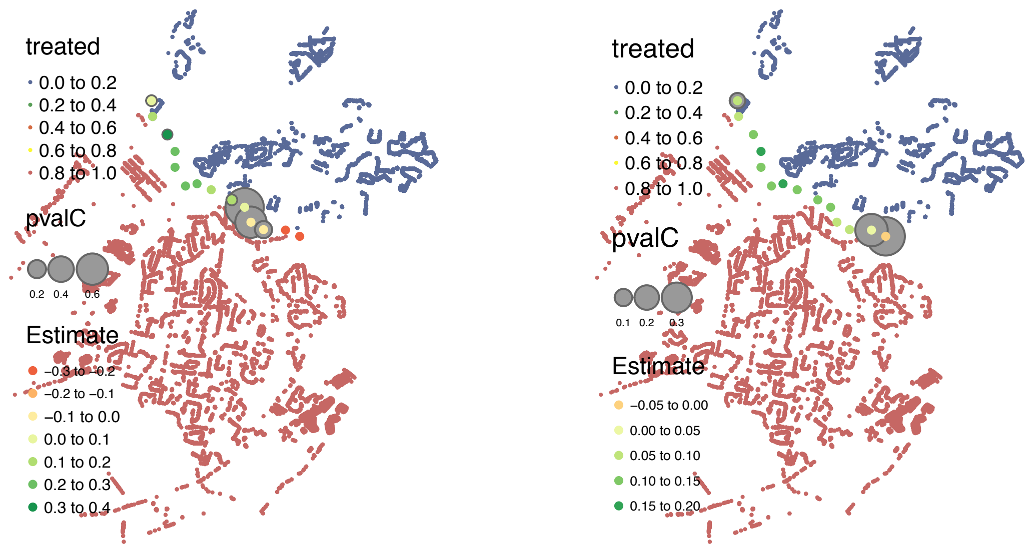


Figure 17: The heterogeneous effects alongside the border visualised. Data driven bandwith selection (left) and forced bandwith of 5km (right), pvalues for each point estimate are represented by the grey circle. Contrary to what Keele and Titiunik [2015] and Imbens and Wager [2019] suggest, there appears to be a positive and significant discontinuity a vast part of the boundary (despite from the fringes). The unbalancedness of important covariates across the cutoff, however, should make us skeptical of the causal effect of advertising in 2008.

5.2.1 Placebo Borders KT15

The full results for this robustness exercise on the KT15 data is relegated to subsection 7.4. All the results turn insignificant and constitute a good example for what a GRDDseries in such a case looks like.
The original cutoff was simply shifted along the X- and Y- axis by 0.01 degrees (as the this data comes in EPSG:4326, longitude - latitude and I did not reproject them on purpose as this constitutes a replication study) each in both directions by an affine transformation. This is carried out with the `placebo_border()` function. The `discretise_border()` function then extracts the relevant borderpoints for estimation and the new treatment status is assigned via the creation of a polygon with `cutoff2polygon()`, one for each scenario.

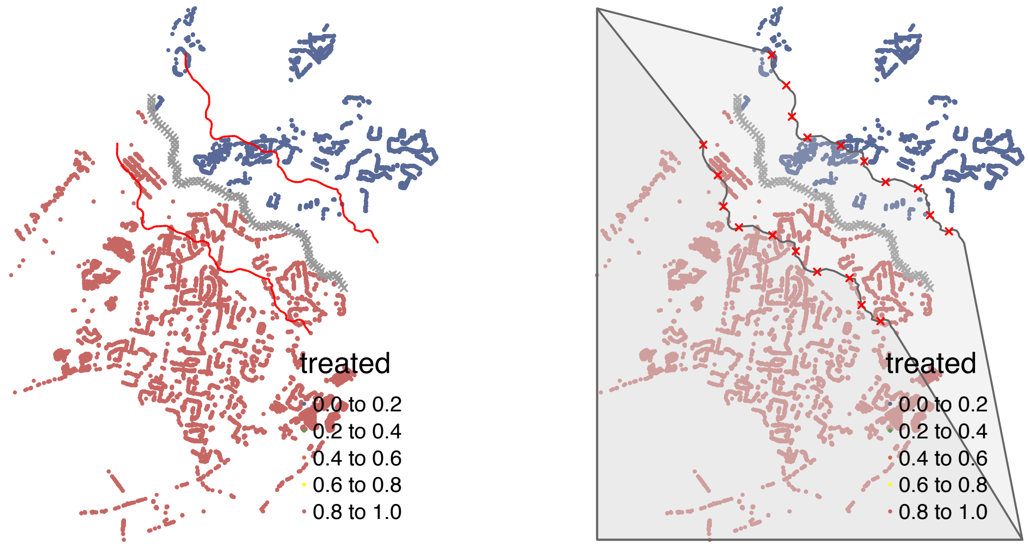


Figure 18: Illustration of the shift and scale operations (left) and the newly created polygons to re-assign the treatment status (right). The original cutoff is simply shifted along the X- and Y- axis by 0.01 degrees each in both directions by an affine transformation.

5.3 Nonparametric, the IW19 way

For brevity this is omitted, as it has been shown in Imbens and Wager [2019]. They obtain a weighted estimate of 0.045 for 2008 voter turnout, but due to the confidence interval of (-0.107, 0.197) they conclude that there is no robust discontinuity to be observed. Once the border is corrected for the fringes to the right where there are no observations in the "control area" to the North, the confidence interval almost doesn't overlap with 0 anymore and I thus conclude that the estimate is not as insignificant as Imbens and Wager [2019] (and also Keele and Titiunik [2015]) claim it is to be.

5.4 The weakness of parametric specifications

We now finally arrive at the section where we show a huge amount of possible combinations of segment dummies/fixed effects [7] (that have been created with the `border_segment()` function), cluster levels [5], bandwiths, and control variables. Three of the cluster levels are perfectly reasonable choices from the KT15 data: city level, census block level, and zip code level. Two of them have been simulated with created rasters of .005 by .005 degrees (resulting in 276 clusters), and .0005 by .0005 degrees (resulting in 3446 clusters). These would basically amount to proximity clusters that resemble squared neighbourhoods. The following tables show all the potential combinations estimated parametrically via OLS, whereby only the coefficient reported is the treated dummy and the first standard error below it is a robust (HC3) standard error which is not clustered.

The main takeaway from this exercise is that, as we move towards narrower bandwiths and approach samplesizes of below 1000 on each side, the point estimates get more and more responsive towards the differential inclusion of border segment lengths. The same holds true for the standard errors: in these "borderline scenarios" there seems quite often to be a fixed effects - cluster level combination that turns an insignificant point estimate into a significant one and vice versa.

It is thus paramout to consult additional metrics, such as nonparametric estimations or simple spatial interpolations, in order to draw conclusions.

Table 14: Demonstration of the volatility of results (only FE, no controls).

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
NO	treated_NO	**0.048**	**0.056**	**0.031**	**0.035**	**0.023**	**0.084**
		(0.006)	(0.008)	(0.011)	(0.017)	(0.03)	(0.047)
		(0.025)	(0.033)	(0.032)	(0.031)	(0.021)	(0.022)
		(0.019)	(0.023)	(0.025)	(0.027)	(0.036)	(0.057)
		(0.008)	(0.01)	(0.015)	(0.022)	(0.04)	(0.057)
		(0.017)	(0.018)	(0.023)	(0.028)	(0.028)	(0.047)
		(0.009)	(0.012)	(0.017)	(0.022)	(0.037)	(0.049)
3	treated_3	**0.049**	**0.056**	**0.056**	**0.073**	**0.075**	**0.158**
		(0.006)	(0.008)	(0.012)	(0.019)	(0.038)	(0.07)
		(0.026)	(0.033)	(0.024)	(0.018)	(0.015)	(0.071)
		(0.019)	(0.021)	(0.024)	(0.03)	(0.033)	(0.066)
		(0.008)	(0.01)	(0.015)	(0.026)	(0.053)	(0.079)
		(0.016)	(0.016)	(0.022)	(0.029)	(0.029)	(0.076)
		(0.009)	(0.011)	(0.016)	(0.027)	(0.049)	(0.079)
4	treated_4	**0.053**	**0.055**	**0.06**	**0.077**	**0.052**	**0.198**
		(0.006)	(0.008)	(0.012)	(0.021)	(0.045)	(0.096)
		(0.022)	(0.028)	(0.019)	(0.014)	(0.035)	(0.128)
		(0.016)	(0.017)	(0.021)	(0.024)	(0.045)	(0.098)
		(0.008)	(0.01)	(0.015)	(0.026)	(0.066)	(0.095)
		(0.015)	(0.015)	(0.02)	(0.026)	(0.031)	(0.099)
		(0.009)	(0.01)	(0.016)	(0.028)	(0.058)	(0.097)
5	treated_5	**0.035**	**0.045**	**0.079**	**0.093**	**0.075**	**0.121**
		(0.01)	(0.01)	(0.014)	(0.02)	(0.042)	(0.082)
		(0.02)	(0.028)	(0.014)	(0.028)	(0.012)	(0.085)
		(0.019)	(0.022)	(0.021)	(0.024)	(0.037)	(0.079)
		(0.012)	(0.013)	(0.018)	(0.026)	(0.056)	(0.091)
		(0.015)	(0.016)	(0.021)	(0.026)	(0.035)	(0.085)
		(0.012)	(0.013)	(0.018)	(0.027)	(0.052)	(0.098)
10	treated_10	**0.036**	**0.056**	**0.093**	**0.041**	**0.039**	**0.211**
		(0.013)	(0.014)	(0.017)	(0.032)	(0.058)	(0.125)
		(0.039)	(0.053)	(0.02)	(0.021)	(0.025)	(0.191)
		(0.026)	(0.031)	(0.026)	(0.043)	(0.058)	(0.125)
		(0.016)	(0.018)	(0.021)	(0.043)	(0.078)	(0.136)
		(0.023)	(0.026)	(0.025)	(0.041)	(0.044)	(0.138)
		(0.018)	(0.018)	(0.021)	(0.04)	(0.072)	(0.136)
		0.04	**0.037**	**0.034**	**0.03**	**0.03**	**0.233**

Table 14: Demonstration of the volatility of results (only FE, no controls). *(continued)*

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
15	treated_15						
		(0.03)	(0.03)	(0.031)	(0.037)	(0.064)	(0.122)
		(0.023)	(0.02)	(0.011)	(0.013)	(0.039)	(0.103)
		(0.042)	(0.044)	(0.044)	(0.041)	(0.041)	(0.098)
		(0.037)	(0.038)	(0.039)	(0.047)	(0.092)	(0.103)
		(0.041)	(0.042)	(0.043)	(0.045)	(0.042)	(0.098)
		(0.037)	(0.038)	(0.038)	(0.047)	(0.083)	(0.136)
20	treated_20	**0.097**	**0.105**	**0.105**	**0.11**	**0**	**0.307**
		(0.025)	(0.027)	(0.027)	(0.038)	(0.067)	(0.144)
		(0.006)	(0.006)	(0.007)	(0.008)	(0.052)	(0.121)
		(0.03)	(0.029)	(0.029)	(0.045)	(0.061)	(0.096)
		(0.026)	(0.028)	(0.028)	(0.041)	(0.067)	(0.102)
		(0.024)	(0.025)	(0.026)	(0.044)	(0.058)	(0.091)
		(0.026)	(0.028)	(0.028)	(0.04)	(0.077)	(0.117)
NA							
	Treated	15000	7888	3997	1853	713	301
	Nobs.	24460	14860	7446	3183	1077	430

Note:
This table summarises the point estimates for the treated dummy over all different border segment fixed-effects and bandwith combinations on the data from Keele, Titiunik [2015]. It demonstrates the volatility of the point estimates and standard errors depending on fixed effect and cluster choices. Especially for the narrower bandwidths (last three columns), there are always sets of clustered standard errors that make the point estimates significant, and such that do the opposite. 20 border segments are admittedly too extreme, but any choice between three and ten or fifteen segments seems fairly justifiable.

[a] The rows below the point estimates (in parenthesis) represent standard errors in the following order: robust (HC3), cluster city [5], cluster census block [165], cluster zip code [2714], cluster simulation 1 [276], cluster simulation 2 [3446].

Table 15: Demonstration of the volatility of results (FE + polynomial in X- and Y- coordinates as controls).

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
NO	treated_NO	**0.095**	**0.075**	**0.116**	**0.121**	**0.068**	**0.103**
		(0.017)	(0.022)	(0.028)	(0.044)	(0.077)	(0.111)
		(0.013)	(0.044)	(0.057)	(0.038)	(0.042)	(0.027)
		(0.029)	(0.046)	(0.048)	(0.058)	(0.095)	(0.142)
		(0.022)	(0.03)	(0.038)	(0.056)	(0.091)	(0.132)
		(0.028)	(0.04)	(0.046)	(0.053)	(0.097)	(0.174)
		(0.021)	(0.03)	(0.039)	(0.054)	(0.093)	(0.119)
3	treated_3	**0.097**	**0.065**	**0.091**	**0.105**	**0.108**	**0.258**
		(0.017)	(0.022)	(0.029)	(0.046)	(0.084)	(0.136)
		(0.019)	(0.056)	(0.07)	(0.046)	(0.032)	(0.095)
		(0.028)	(0.042)	(0.047)	(0.051)	(0.077)	(0.128)
		(0.021)	(0.03)	(0.04)	(0.059)	(0.093)	(0.16)
		(0.027)	(0.039)	(0.045)	(0.052)	(0.083)	(0.115)
		(0.021)	(0.03)	(0.04)	(0.056)	(0.097)	(0.146)
4	treated_4	**0.06**	**0.046**	**0.106**	**0.111**	**0.087**	**0.318**
		(0.019)	(0.024)	(0.032)	(0.048)	(0.104)	(0.165)
		(0.018)	(0.04)	(0.076)	(0.054)	(0.062)	(0.196)
		(0.039)	(0.046)	(0.058)	(0.066)	(0.109)	(0.156)
		(0.025)	(0.033)	(0.044)	(0.059)	(0.117)	(0.149)
		(0.035)	(0.044)	(0.052)	(0.057)	(0.113)	(0.153)
		(0.024)	(0.032)	(0.043)	(0.058)	(0.116)	(0.161)
5	treated_5	**0.093**	**0.07**	**0.113**	**0.09**	**0.207**	**0.338**
		(0.017)	(0.023)	(0.03)	(0.05)	(0.119)	(0.169)
		(0.012)	(0.054)	(0.075)	(0.073)	(0.085)	(0.115)
		(0.032)	(0.047)	(0.051)	(0.068)	(0.107)	(0.146)
		(0.022)	(0.031)	(0.041)	(0.064)	(0.137)	(0.172)
		(0.029)	(0.043)	(0.046)	(0.063)	(0.134)	(0.128)
		(0.022)	(0.031)	(0.041)	(0.061)	(0.132)	(0.173)
10	treated_10	**0.111**	**0.109**	**0.14**	**0.068**	**0.243**	**0.651**
		(0.02)	(0.025)	(0.033)	(0.055)	(0.125)	(0.196)
		(0.007)	(0.033)	(0.071)	(0.09)	(0.075)	(0.065)
		(0.035)	(0.05)	(0.063)	(0.067)	(0.094)	(0.117)
		(0.025)	(0.033)	(0.045)	(0.068)	(0.121)	(0.167)
		(0.033)	(0.046)	(0.055)	(0.064)	(0.117)	(0.089)
		(0.025)	(0.033)	(0.044)	(0.068)	(0.131)	(0.158)
		0.061	**0.024**	**0.069**	**0.087**	**0.246**	**0.391**

Table 15: Demonstration of the volatility of results (FE + polynomial in X- and Y- coordinates as controls). *(continued)*

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
15	treated_15						
		(0.031)	(0.034)	(0.039)	(0.057)	(0.125)	(0.189)
		(0.019)	(0.035)	(0.055)	(0.068)	(0.075)	(0.116)
		(0.047)	(0.055)	(0.066)	(0.06)	(0.09)	(0.188)
		(0.04)	(0.044)	(0.053)	(0.072)	(0.122)	(0.204)
		(0.046)	(0.053)	(0.062)	(0.055)	(0.104)	(0.148)
		(0.039)	(0.044)	(0.052)	(0.07)	(0.131)	(0.199)
20	treated_20	**0.127**	**0.095**	**0.161**	**0.13**	**0.267**	**0.55**
		(0.028)	(0.031)	(0.038)	(0.062)	(0.131)	(0.226)
		(0.019)	(0.038)	(0.051)	(0.083)	(0.116)	(0.09)
		(0.038)	(0.048)	(0.053)	(0.079)	(0.135)	(0.181)
		(0.032)	(0.036)	(0.047)	(0.071)	(0.136)	(0.209)
		(0.035)	(0.043)	(0.054)	(0.072)	(0.153)	(0.113)
		(0.031)	(0.036)	(0.047)	(0.072)	(0.14)	(0.187)
NA							
	Treated	15000	7888	3997	1853	713	301
	Nobs.	24460	14860	7446	3183	1077	430

Note:
This table summarises the point estimates for the treated dummy over all different border segment fixed-effects and bandwith combinations on the data from Keele, Titiunik [2015]. It demonstrates the volatility of the point estimates and standard errors depending on fixed effect and cluster choices. Especially for the narrower bandwidths (last three columns), there are always sets of clustered standard errors that make the point estimates significant, and such that do the opposite. 20 border segments are admittedly too extreme, but any choice between three and ten or fifteen segments seems fairly justifiable.

[a] The rows below the point estimates (in parenthesis) represent standard errors in the following order: robust (HC3), cluster city [5], cluster census block [165], cluster zip code [2714], cluster simulation 1 [276], cluster simulation 2 [3446].

Table 16: Demonstration of the volatility of results (FE + X- and Y-coordinate polynomials + X (white, income, female)).

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
NO	treated_NO	**0.033**	**0.013**	**0.062**	**0.07**	**0.028**	**0.08**
		(0.018)	(0.022)	(0.029)	(0.046)	(0.08)	(0.116)
		(0.018)	(0.045)	(0.062)	(0.045)	(0.059)	(0.043)
		(0.031)	(0.042)	(0.048)	(0.057)	(0.09)	(0.147)
		(0.022)	(0.029)	(0.039)	(0.056)	(0.095)	(0.135)
		(0.028)	(0.038)	(0.047)	(0.054)	(0.095)	(0.18)
		(0.022)	(0.03)	(0.04)	(0.055)	(0.096)	(0.123)
3	treated_3	**0.031**	**0.01**	**0.047**	**0.061**	**0.067**	**0.224**
		(0.018)	(0.023)	(0.03)	(0.048)	(0.085)	(0.139)
		(0.026)	(0.06)	(0.073)	(0.06)	(0.047)	(0.097)
		(0.032)	(0.043)	(0.048)	(0.054)	(0.078)	(0.135)
		(0.022)	(0.03)	(0.041)	(0.06)	(0.096)	(0.161)
		(0.027)	(0.04)	(0.048)	(0.055)	(0.085)	(0.126)
		(0.022)	(0.031)	(0.041)	(0.058)	(0.099)	(0.147)
4	treated_4	**0.011**	**-0.004**	**0.049**	**0.06**	**0.046**	**0.314**
		(0.019)	(0.024)	(0.032)	(0.05)	(0.106)	(0.165)
		(0.011)	(0.044)	(0.088)	(0.056)	(0.068)	(0.171)
		(0.038)	(0.046)	(0.061)	(0.064)	(0.111)	(0.151)
		(0.025)	(0.033)	(0.044)	(0.059)	(0.119)	(0.147)
		(0.033)	(0.043)	(0.054)	(0.058)	(0.114)	(0.137)
		(0.025)	(0.033)	(0.044)	(0.06)	(0.119)	(0.164)
5	treated_5	**0.034**	**0.011**	**0.061**	**0.053**	**0.183**	**0.332**
		(0.018)	(0.023)	(0.031)	(0.053)	(0.122)	(0.169)
		(0.018)	(0.049)	(0.079)	(0.083)	(0.125)	(0.146)
		(0.033)	(0.046)	(0.053)	(0.07)	(0.123)	(0.146)
		(0.023)	(0.031)	(0.042)	(0.066)	(0.14)	(0.176)
		(0.029)	(0.043)	(0.049)	(0.066)	(0.142)	(0.111)
		(0.022)	(0.032)	(0.042)	(0.064)	(0.139)	(0.176)
10	treated_10	**0.06**	**0.05**	**0.085**	**0.037**	**0.233**	**0.639**
		(0.02)	(0.025)	(0.033)	(0.059)	(0.128)	(0.199)
		(0.014)	(0.034)	(0.078)	(0.11)	(0.085)	(0.104)
		(0.035)	(0.049)	(0.065)	(0.072)	(0.107)	(0.127)
		(0.025)	(0.033)	(0.045)	(0.07)	(0.126)	(0.174)
		(0.033)	(0.045)	(0.056)	(0.068)	(0.121)	(0.103)
		(0.025)	(0.033)	(0.044)	(0.071)	(0.137)	(0.168)
		-0.019	**-0.036**	**0.014**	**0.057**	**0.236**	**0.366**

Table 16: Demonstration of the volatility of results (FE + X- and Y-coordinate polynomials + X (white, income, female)). *(continued)*

FE	dep.var	full	3km	2km	1km	0.5km	0.3km
15	treated_15						
		(0.032)	(0.035)	(0.04)	(0.063)	(0.128)	(0.188)
		(0.035)	(0.037)	(0.069)	(0.1)	(0.094)	(0.134)
		(0.06)	(0.067)	(0.074)	(0.071)	(0.105)	(0.185)
		(0.044)	(0.048)	(0.055)	(0.073)	(0.122)	(0.204)
		(0.057)	(0.063)	(0.068)	(0.064)	(0.102)	(0.135)
		(0.042)	(0.046)	(0.053)	(0.074)	(0.138)	(0.2)
20	treated_20	**0.075**	**0.058**	**0.121**	**0.09**	**0.238**	**0.512**
		(0.028)	(0.032)	(0.039)	(0.064)	(0.134)	(0.228)
		(0.028)	(0.037)	(0.055)	(0.105)	(0.137)	(0.102)
		(0.038)	(0.05)	(0.059)	(0.082)	(0.15)	(0.188)
		(0.033)	(0.038)	(0.049)	(0.076)	(0.14)	(0.215)
		(0.036)	(0.047)	(0.056)	(0.075)	(0.159)	(0.109)
		(0.032)	(0.037)	(0.048)	(0.075)	(0.145)	(0.192)
NA							
	Treated	15000	7888	3997	1853	713	301
	Nobs.	24460	14860	7446	3183	1077	430

Note:
This table summarises the point estimates for the treated dummy over all different border segment fixed-effects and bandwith combinations on the data from Keele, Titiunik [2015]. Here we add the control variables that were showing a significant jump for the balancing checks before. Especially for the narrower bandwidths (last three columns), there are always sets of clustered standard errors that make the point estimates significant, and such that do the opposite. 20 border segments are admittedly too extreme, but any choice between three and ten or fifteen segments seems fairly justifiable.

[a] The rows below the point estimates (in parenthesis) represent standard errors in the following order: robust (HC3), cluster city [5], cluster census block [165], cluster zip code [2714], cluster simulation 1 [276], cluster simulation 2 [3446].

6 Conclusion & Outlook

This note has shown that in theory both parametric and nonparametric specifications are equally well suited to deliver appropriate average treatment effects for spatial Regression Discontinuity Designs. Generally we have to be a bit more careful with parametric estimations, though. First of all, as has been demonstrated on simulated data, they are more prone towards producing "false positive" results. Secondly, there is less guidance for researchers on how many border segments to choose to exploit "within variation", the level for standard error clustering, and how certain robustness checks, placebo border shifts for instance, should take place. This is what Gelman and Loken [2014] describe as "researcher degrees of freedom".

Nonparametric estimation or spatial interpolation are an easy way to alleviate these problems and provide more transparency.
It has also been demonstrated that recent critique by Keele and Titiunik [2015] and Imbens and Wager [2019] is to hard, as it implictly rejects parametric estimations for spatial RDDs. This is demonstrated by the replication of the results in Keele and Titiunik [2015], which say that the effect of televions advertising on voter turnout in the 2008 presidential elections alongside a part of the New York - Philadelphia media boundary is non-robust. We would conclude with this result, but solely on the grounds of failed balancing checks that seem to suggest that fundamental covariates such as race and income vary in non-systematic ways across the RD cutoff. The obtained nonparametric estimates, once sparse borderpoints with no observations in their vicinity are removed, seem to be somewhat robust. This is contrary to what Keele and Titiunik [2015] and Imbens and Wager [2019] find. Our conclusion is substantiated by using (spatial) Monte Carlo simulations in order to show that a single non-robust borderpoint estimate should not lead us to reject an average effect, and that such behaviour is perfectly rationalisable with a DGP that in fact exhibits a spatial discontinuity.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017): "When Should You Adjust Standard Errors for Clustering?" *arXiv:1710.02926 [econ, math, stat].*

Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press, 1 edition ed.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115, 588–638.

Black, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *The Quarterly Journal of Economics*, 114, 577–599.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

——— (2015): "Rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs," *R Journal*, 7, 38–51.

Canay, I. A. and V. Kamat (2018): "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design," *The Review of Economic Studies*, 85, 1577–1608.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (forthcoming): *A Practical Introduction to Regression Discontinuity Designs: Volume II*, Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press.

Dell, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863–1903.

Dell, M. and B. A. Olken (forthcoming): "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java," *Review of Economic Studies.*

Gelman, A. and G. Imbens (forthcoming): "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 0, 1–10.

Gelman, A. and E. Loken (2013): "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "p-Hacking" and the Research Hypothesis Was Posited Ahead of Time," *mimeo, Columbia.*

——— (2014): "The Statistical Crisis in Science," *American Scientist*, 102, 460.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Imbens, G. and S. Wager (2019): "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 101, 264–278.

Imbens, G. W. and K. Kalyanaraman (2012): "Optimal Bandwith Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 127–155.

Imbens, G. W. and T. Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635.

Keele, L. J., S. Lorch, M. Passarella, D. Small, and R. Titiunik (2017): "An Overview of Geographically Discontinuous Treatment Assigments With an Application to Children's Health Insurance," in *Advances in Econometrics*, ed. by M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, vol. 38.

Keele, L. J. and R. Titiunik (2015): "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

——— (2016): "Natural Experiments Based on Geography," *Political Science Research and Methods*, 4, 65–95.

Keele, L. J., R. Titiunik, and J. R. Zubizarreta (2015): "Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives," *Journal of the Royal Statistical Soociety: Series A*, 178, 223–239.

Kelly, M. (2019): "The Standard Errors of Persistence," *mimeograph.*

Lehner, A. (2019a): "450 Years of Portuguese Colonialism in India: Missionaries, Education, and the Roots of Gender Inequality," *mimeograph, University of Bologna.*

——— (2019b): "SpatialRDD: An R-Package for Spatial Regression Discontinuity Designs," .

McCrary, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714.

Papay, J. P., J. B. Willett, and R. J. Murname (2011): "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables," *Journal of Econometrics*, 161, 203–207.

Porter, J. (2003): "Estimation in the Regression Discontinuity Model," .

Reardon, S. F. and J. P. Robinson (2012): "Regression Discontinuity Designs With Multiple Rating-Score Variables," *Journal of Research on Educational Effectiveness*, 5, 83–104.

Thistlethwaite, D. L. and D. T. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51, 309–317.

Wong, V. C., P. M. Steiner, and T. D. Cook (2013): "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods," *Journal of Educational and Behavioral Statistics*, 38, 107–141.

Zajonc, T. (2012): "Essays on Causal Inference for Public Policy," .

Zubizarreta, J. R. (2012): "Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery," *Journal of the American Statistical Association*, 107, 1360–1371.

7 Appendices

7.1 Spatial Interpolations

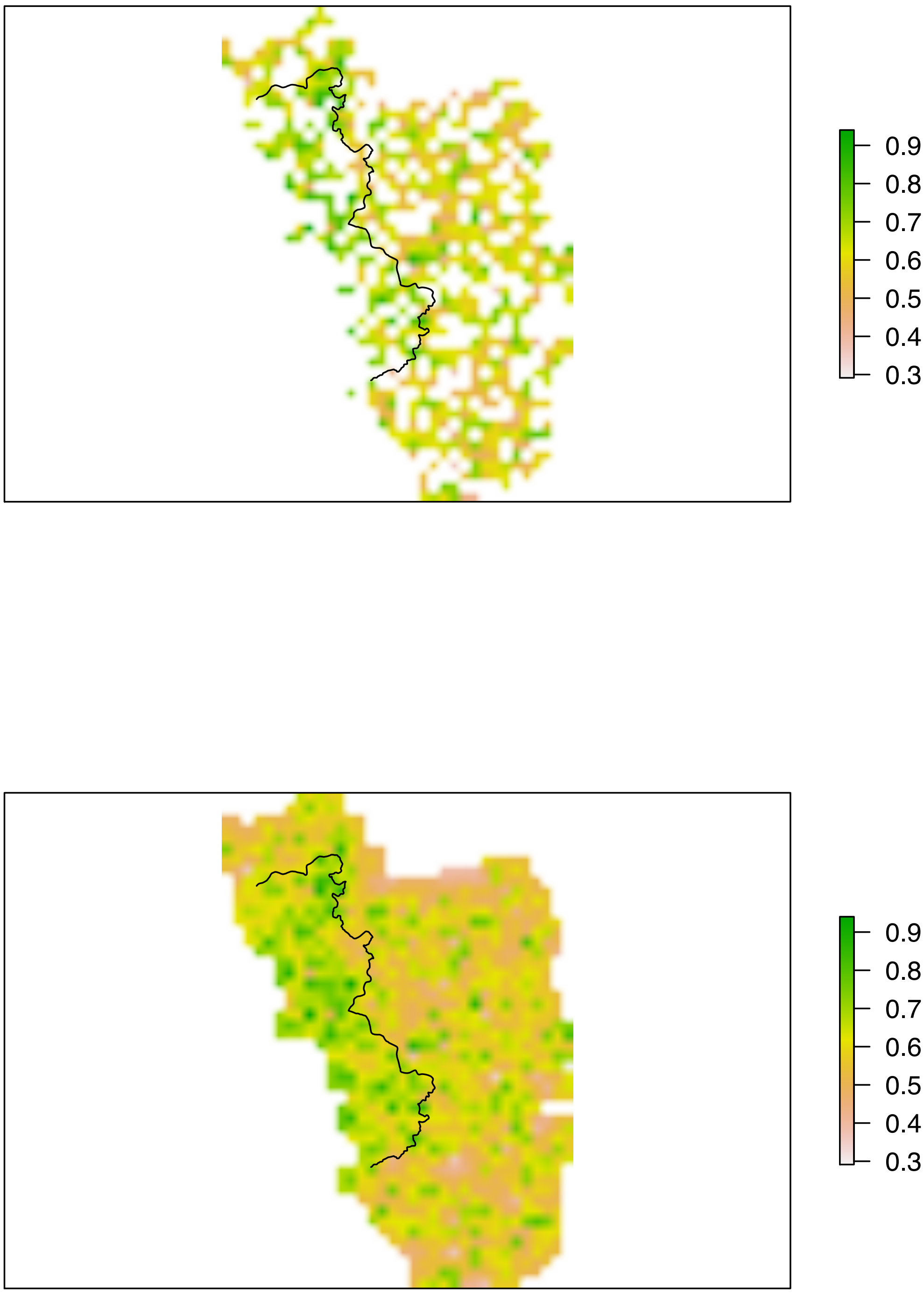


Figure 19: The most basic ways to spatially interpolate shown on the simulated data on education.

7.2 Appendix to Chapter 3, Simulation

7.3 Appendix to Chapter 4, Monte Carlo

7.4 Appendix to Chapter 5, KT15 replication

Here we explore the results of the nonparametric estimation on two placebo borders with the data from Keele and Titiunik [2015]. The original cutoff is simply shifted along the X- and Y- axis by 0.01 degrees each in both directions by an affine transformation. The results are highly insignificant and show what the estimates alongside a border where is no effect look like.

Table 17: Nonparametric point estimates for placebo border 1 (shifted to North) for Keele and Titiunik [2015].

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	0.03	1242	1159	3.00	-0.06	0.12	-0.05	0.15
2	0.39	788	276	1.60	0.15	0.62	0.18	0.69
3	-0.17	3676	4333	3.80	-0.24	-0.10	-0.36	-0.15
4	-0.02	5617	6350	4.20	-0.08	0.04	-0.13	0.11
5	0.05	628	1001	1.40	-0.07	0.18	-0.09	0.21
6	0.35	649	395	1.00	0.14	0.56	0.13	0.63
7	0.12	2054	792	1.60	-0.02	0.27	-0.03	0.33
8	0.11	6018	1668	2.70	0.03	0.20	-0.02	0.19
9	-0.25	1083	245	1.40	-0.55	0.04	-0.67	0.03
10	-0.04	5790	1920	3.10	-0.16	0.08	-0.31	0.04
Mean	0.06	2754	1814	2.38	-0.09	0.20	-0.14	0.22

Note:
NA

Table 12

Table 18: Nonparametric point estimates for placebo border 1 (shifted to North) for Keele and Titiunik [2015], fixed bandwidth.

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	-0.05	3846	4376	5	-0.12	0.02	-0.07	0.13
2	-0.10	5542	5660	5	-0.15	-0.04	-0.22	0.02
3	-0.10	6387	7682	5	-0.15	-0.04	-0.50	-0.23
4	-0.02	7492	8895	5	-0.07	0.03	-0.14	0.12
5	0.09	7774	8482	5	0.04	0.13	-0.01	0.15
6	0.14	7774	8491	5	0.09	0.18	0.03	0.18
7	0.15	7757	8951	5	0.10	0.19	0.01	0.19
8	0.13	7618	7791	5	0.08	0.17	0.05	0.23
9	0.08	6919	8368	5	0.03	0.14	-0.14	0.10

Table 18: Nonparametric point estimates for placebo border 1 (shifted to North) for Keele and Titiunik [2015], fixed bandwidth. *(continued)*

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
10	0.02	6906	8235	5	-0.04	0.07	-0.29	0.02
Mean	0.03	6802	7693	5	-0.02	0.08	-0.13	0.09

Note:
NA

Table 13
Figure 15

Table 19: Nonparametric point estimates for placebo border 2 (shifted to South) for Keele and Titiunik [2015].

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	-0.05	1063	1252	2.50	-0.20	0.09	-0.20	0.16
2	-0.08	2730	3209	3.10	-0.17	0.01	-0.23	0.13
3	-0.17	963	2339	2.10	-0.29	-0.05	-0.34	-0.05
4	-0.13	2442	5529	3.10	-0.21	-0.05	-0.26	-0.05
5	0.08	699	1418	1.40	-0.03	0.19	-0.03	0.23
6	-0.15	860	1303	1.30	-0.30	0.00	-0.35	-0.02
7	-0.11	1227	2206	1.80	-0.26	0.03	-0.32	0.02
8	0.13	1018	816	1.40	-0.02	0.28	0.00	0.33
9	-0.24	599	532	1.00	-0.49	0.00	-0.56	-0.02
10	-0.31	505	430	0.90	-0.51	-0.10	-0.56	-0.12
Mean	-0.10	1211	1903	1.86	-0.25	0.04	-0.29	0.06

Note:
NA

Table 12

Table 20: Nonparametric point estimates for placebo border 2 (shifted to South) for Keele and Titiunik [2015], fixed bandwidth.

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
1	-0.08	4781	5295	5	-0.15	-0.01	-0.12	0.16
2	-0.09	5422	7137	5	-0.14	-0.03	-0.14	0.09
3	-0.10	5647	8956	5	-0.15	-0.05	-0.28	-0.08
4	-0.08	6057	9814	5	-0.12	-0.03	-0.28	-0.07
5	-0.01	7402	10364	5	-0.06	0.03	-0.05	0.10

Table 20: Nonparametric point estimates for placebo border 2 (shifted to South) for Keele and Titiunik [2015], fixed bandwidth. *(continued)*

Point	Estimate	Ntr	Nco	bw	CI_Conv_l	CI_Conv_u	CI_Rob_l	CI_Rob_u
6	-0.01	8332	11579	5	-0.05	0.03	-0.05	0.10
7	-0.02	8688	11805	5	-0.07	0.02	-0.07	0.10
8	-0.01	9629	10703	5	-0.05	0.03	-0.07	0.10
9	-0.02	9060	10755	5	-0.06	0.02	-0.15	0.00
10	-0.03	8692	10203	5	-0.07	0.02	-0.19	-0.03
Mean	-0.04	7371	9661	5	-0.09	0.00	-0.14	0.05

Note:
NA

Table 13
Figure 15
Figure 22

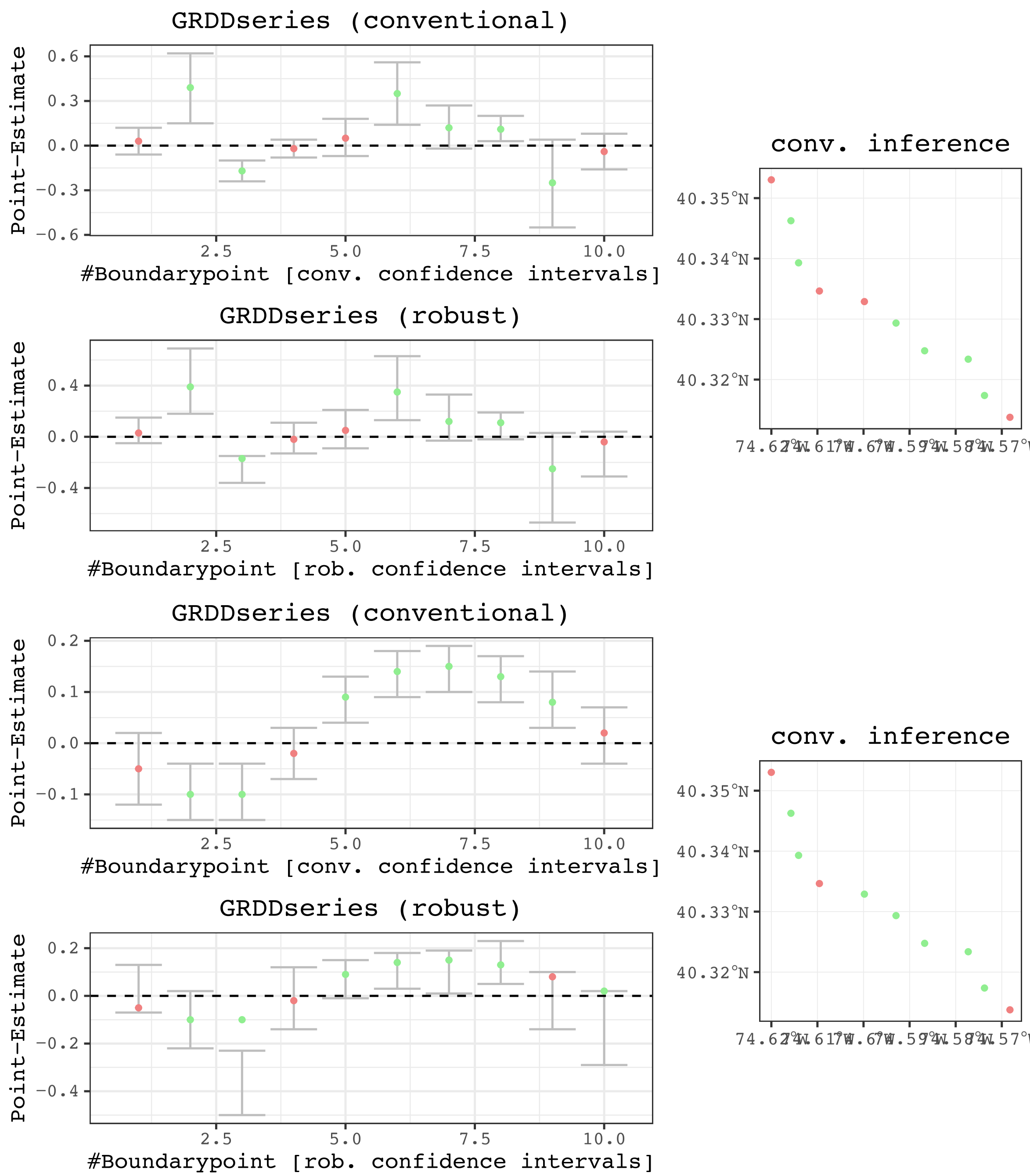


Figure 20: Visualised results for both the flexible and fixed bandwith estimation for placebo border 1. This resembles a highly insignificant result, the sign changes significantly multiple times.

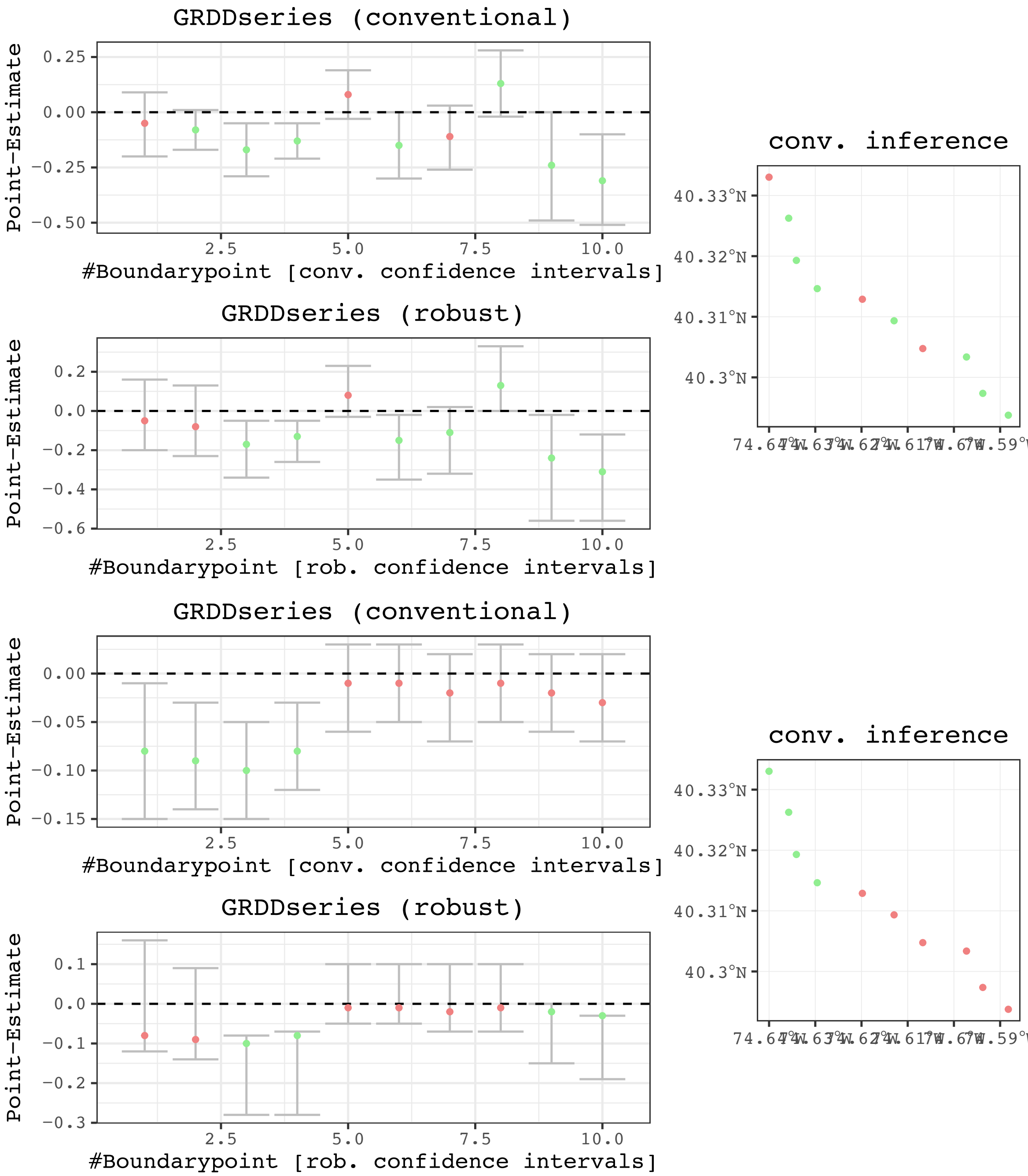


Figure 21: Visualised results for both the flexible and fixed bandwith estimation for placebo border 2 (southshift). This resembles a highly insignificant result, the sign changes significantly multiple times. Interestingly the sign for the fixed bw scenario switched in the opposite direction.

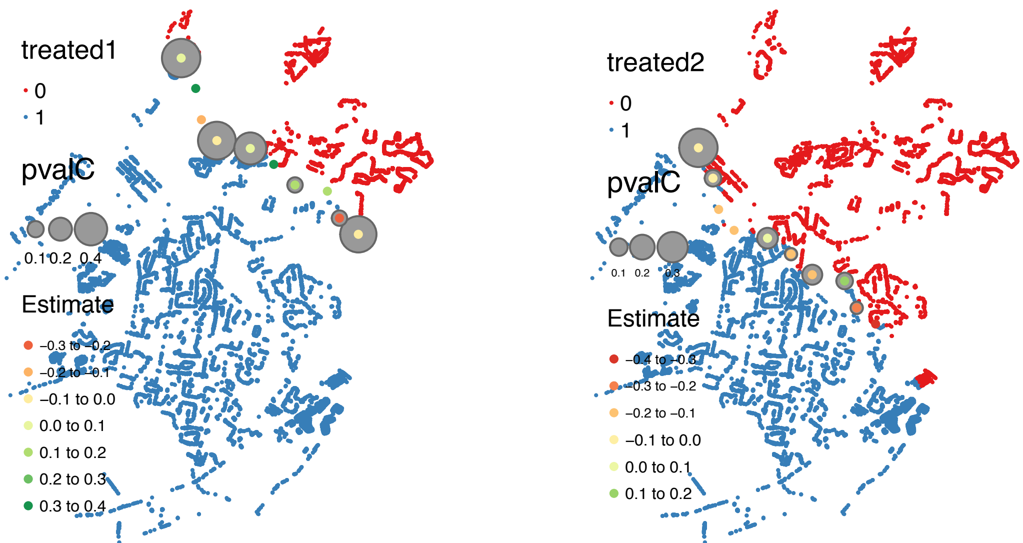


Figure 22: The estimates on the placebo borders visualised. Data driven bandwith selection (left) and forced bandwith of 5km (right), pvalues for each point estimate are represented by the grey circle. This is what a truly insignificant result looks like, even though this is easier to infer from the GRDDseries. The point estimates change sign multiple times and are significant in both directions.

Gravitating towards modern Economic Growth: Trade Costs, Agglomeration, and Specialisation

Matteo Cervellati[a,d,e,f], Gianandrea Lanzara[a], and Alexander Lehner[a]

[a]Unversity of Bologna
[b]CEPR London
[c]CESifo Munich
[d]IZA Bonn

This version: September 30, 2019

Abstract

This paper studies the pattern of European historical development in the pre-industrial era (1000-1800) by the means of a novel quantitative spatial economic model with two sectors of production. Economies of scale and bilateral trade costs are sector-specific. The insights from the model are then used to explain the evolution of the European urban system on its way to the Industrial Revolution. We show that the increasing possibilities to trade agricultural goods can explain the emergence of large urban agglomerations specialized in manufacturing production, even in the absence of an increase in agricultural productivity. We quantify the model using a plethora of historical data at the grid-cell level, most importantly the (updated and revised) data on urban population by Bairoch, and the bilateral effective trade costs between all grid cells in Europe. As a by-product, our model delivers a spatial distribution of the rural population and urbanisation/specialisation rates that vary across time and space.

Keywords: Quantitative Spatial Economics, Economic Geography, Urbanisation, Agglomeration, European Economic History, Little Divergence.

JEL Codes:

1 Introduction

"The study of industrialisation in any given European country will remain incomplete unless it incorporates a European dimension."

—Pollard (1973)

Around the year 1000 Europe had around two million inhabitants which lived in cities above 5000 inhabitants. By the year 1800, around the time when major parts of Europe started to industrialise, this number had increased to 20 million. For the first time the world saw a dense network of sizeable cities evolving. This meant that the city size distribution became more and more right tailed over time, eventually approximating Zipf's law, and, most importantly, a number of urban giants emerged. Over the same period, the continent experienced a shift of its economic centre of gravity from the Mediterranean towards the North-West. This so-called "little divergence" has been at the centre of discussion amongst Historians and Economists alike. Some scholars suggest that understanding this process is a key element in order to further our understanding of why the North-Sea region of Europe, in particular the Dutch Provinces and Britain, was the first area in the world to break out of the Malthusian trap and undergo sustained economic growth, eventually industrialising and paving the way to an unprecedented period of human prosperity.

Building on recent developments in spatial economics [Allen and Arkolakis, 2014; Redding, 2016; Redding and Rossi-Hansberg, 2017], this paper proposes a quantitative economic geography model that is able to rationalize these patterns, thus shedding light on the roots of pre-industrial development in Europe. In particular, we highlight the role of falling trade costs for agricultural goods, in combination with economies of scale in the manufacturing sector, as a source of urban growth and structural transformation. When agricultural goods must be imported from the surrounding countryside, urban growth is limited by the increasing cost of attracting farmers from longer and longer distances, as well as from the presence of other cities competing for the same land. However, as inter-city trade costs for agricultural goods fall, cities may potentially sustain large populations even in the absence of a rural hinterland. As a result, some cities will specialize in manufacturing, giving rise to large urban agglomerations, whereas other cities, as well as their surrounding regions, will specialize in agricultural, though export-oriented, production. Economies of scale reinforce this effect, dragging more and more workers to the expanding manufacturing cities.

The importance of an increase in agricultural productivity has been shown to be a key element for urbanisation and industrialisation, both in the historical [e.g. Jones, 1968; Schultz, 1968; Overton, 1996] and economics literature [e.g. Matsuyama, 1992; Desmet and Parente, 2012]. Our framework suggests that in addition to that, another crucial condition was the possibility to easily ship around food in order to feed non-agricultural, urban people. The transport of the bulky grains over land was only viable for distances of up to 30-40 kilometres due to their low value. The per hectare output per year, given there was no crisis, was only a few hundred kilos. Thus the only way to sustain larger cities was to import grains from farther away [Abel, 1980; Pounds, 1973, and many others]. The peculiar shape of Europe, with its long coastline and dense network of navigable rivers, naturally offered many city locations that provided favourable market access and made this type of trade possible. This suggests that geography of Europe was crucial in another important way than typically put forward by the literature [Diamond, 1997; Morris, 2010].

The first historical examples are the Dutch cities, which were able to grow due to imported grains, mostly rye, from Eastern Europe via the Baltic Sea [van Tielhof, 1995; Tielhof, 2002]. As one person needs around 250-300 kilograms of grain per year in order to be properly fed [Abel, 1981], the amount of ships and journeys needed is very high. At least juxtaposed to the typical pattern of trade with manufactured goods and spices from Asia via the Levant during the commercial revolution of the 12th and 13th centuries. One convoy of five to ten Venetian ships full of goods, escorted by several warships, were enough to feed the appetite of almost all of Europe for one year. Furthermore the relative importance of grain trade with respect to expensive consumptions goods is substantiated by the fact that less than 1% of the available Dutch shipping tonnage was sailing to Asia for the Dutch East India Company [Kossman, 1970]

Scholars have pondered over the importance of cities and their role on a society's path to prosperity since centuries. Early Italian thinkers such as Botero [1588] or Serra [1613] already described very accurately the paramount role that urban agglomerations have to play when it comes to inquiries into the nature and causes of the wealth of nations. Smith [1776] himself devoted quite a substantial part of book three of his magnum opus to the crucial role of cities and towns. Something that Pirenne [1927] further expanded and elaborated on. Urban centres always were well understood as points of attraction and economic powerhouses that were catalysts for economic development and (technical) progress.

Historians further broadened and developed those ideas and put them in the context of different periods [e.g. de Vries, 1984; Bairoch, 1988; Braudel, 1993; Hohenberg and Lees, 1995]. All these contributions were largely descriptive case studies with no formal framework on how to think about cities and networks on a larger scale and their changing role throughout time.

To us, some major and crucial broader questions remain unanswered. Why is it that we historically observe several clusters of smaller cities, relatively well connected and geographically close to each other (N-Italy, Dutch Provinces), and areas where only one primate city is existing, seemingly sucking in people and goods from large surroundings (such as London or Paris). What were their respective roles and importance when it comes to long-run development? How and why did that change over time? How did this system evolve the way it did? What were the factors that changed the structure of the European network, and can exogenous events such as the Columbian exchange explain this? The spatial general equilbrium model that we put forward potentially lets us contribute to some of those issues. It suggests that conventional explanations need to be augmented by taking into account how cities were fed and how the tightly knit urban network was possible to arise.

As already indicated, in the early historical treatments, interactions between cities and points of agglomeration in a network sense received little to no attention. Let alone how the distribution evolved and why and how it changed over the course of Europe's history.

The early theoretical contributions in Economics could provide answers for why cities form [e.g. Marshall, 1890; Henderson, 1974; Krugman, 1991], but not where they form - let alone when they do. One of the reasons precisely being the absence of (first nature) geography in those models. It also has been well established that geography itself matters [e.g. Krugman, 1999; Henderson, Shalizi, and Venables, 2001]. Outside of Spatial Economics and Economic Geography this is also well accepted amongst growth economists [e.g. Gallup, Sachs, and Mellinger, 1999; Weil, 2012].

In our model we treat the location of cities as exogenous. There is a larger literature, starting with early work from geographers [Smith, 1978; Perpillou, 1977], that informs our understanding of where urban agglomeration formed under what circumstances. Arguably the first one who tried to understand in what exact locations cities formed in a quantitative and structured way was Stelder [2005]. He was pioneering the use of a grid-cell level world in order to simulate a simple model that made predictions on size and location of urban agglomerations in Europe. Only recently Bosker and Buringh [2017] followed up on this by thoroughly analysing the locational fundamentals of cities with a mainly empirical approach. They find that first nature geography is the prime determinant of city location throughout time. Being close to water, a river, or a transport hub matters most but also the agricultural potential plays a decisive role. They assess the impact of second nature geography by using an established index for urban- or market potential, find that it matters less than first nature, but identify a non-linear effect in distance to already existing cities which is consistent with economic geography theory. They further establish that the importance of trade increases throughout time. This is somewhat confirmed by a more timely but conceptually similar study by Henderson, Squires, Storeygard, and Weil [2018] who try to understand the characteristics that determine the global spatial distribution of economic activity, and thus implicitly people. Their striking finding is that for "early developers" agriculture mattered at lot whereas for "late developers" variables related to trade mattered much more due to the fact that the fall in global transport costs pre-dated structural transformation.

Taking into account this knowledge, we could easily add an endogenous city formation framework to our model. But using Occam's razor we deliberately choose not to do so in order focus what is most important to this research: understanding the change of the spatial pattern of economic activity in pre-industrial Europe.

The importance of trade costs has also been recently emphasised by Delventhal [2018] within a quantitative spatial framework. He asserts that changes in transport costs benefited some locations more than others and can account quantitatively for a number of key patterns in global population and income growth. To account better for this natural heterogeneity in the importance of transport costs he uses an approach similar to Donaldson and Hornbeck [2016], calculating distances to trading partners and determining the value of the trading connection using a general equilibrium model. This is similar to what our model is aiming to achieve.

In terms of theory we relate to the growing field of Quantitative Spatial Economics [see e.g. Redding and Rossi-Hansberg, 2017, for an outstanding summary], initiated by Allen and Arkolakis [2014] and Redding [2016], which models the economy over a rich set of geographical features and links it to quantitative trade models with labour mobility to examine the spatial distribution of economic activity. These models are capable of accommodating an arbitrary number of locations, heterogeneous in their factor endowments, productivity and shipping costs.

Most similar to our paper is the framework developed by Nagy [2017], which takes aggregate population and technology growth in the 19th century United States as given, and seeks to explain their distribution across space in the decades leading up to 1860.

Like Nagy [2017] we are also interested not only in the spatial distribution of economic activity, and thus implicitly where people live, but also its evolution throughout time. Hence we also relate to the small but growing number of papers that try to investigate the evolution of productivity, taking space into account [Michaels, Rauch, and Redding, 2012; Desmet and Rossi-Hansberg, 2014; Desmet, Nagy, and Rossi-Hansberg, 2016].

After a quick historical introduction in the next section we will present the model and then finally run some simulation studies in order to illustrate its insights.

2 (Geographic) Evolution of the European Economic System and the "Little Divergence" Debate

Going back in history, there was nothing that would have predicted the rise of Europe. Our species originated in Africa, Agriculture was invented in the Middle East and China, and the first towns emerged alongside the Euphrates, Tigris, and the Nile Rivers. Yet, it was Europe that overtook Asia and the Middle-East, experiencing episodes of substantial per capita growth, and eventually being the first area in human history to industrialise. This Industrial Revolution was one of the key events in history, serving as a gateway towards a period of unprecedented levels of welfare and human prosperity. One pre-condition for this process within Europe was the so-called little divergence, a gradual shift of the center of economic gravity from the Mediterranean towards the North Sea, specifically the Dutch Provinces and England. At the same time we see an unprecedented densification of the urban network in Europe and, for the first time, the evolution of a heavily right-tailed distribution of cities that is so characteristic for the world today.
This process of divergence started only roughly a century after the first outbreaks of the black death around 1350. During the major boom of Europe from 900 to 1300, most probably induced by favourable climate change [Campbell, 2016], growth equally occurred on a pan-continental scale. This lead to the geographic expansion of economic activity to wide parts of Europe [see e.g. Bartlett, 1993]. Trade increased substantially and a commercial revolution took place [Lopez, 1976; Epstein, 2009]. Technological innovations were substantial [e.g. White, 1962]. Some scholars even speak of a medieval industrial revolution [Gimpel, 1976].
At some point around 1500 the North-Sea region started to grow on a per capita basis [van Zanden and van Leeuwen, 2012] and started to overtake the major economic areas of the Mediterranean [Malanima, 2013]. This marked the beginning of the little divergence, leading to early institutional improvements in the Dutch Provinces de Vries and van der Woude [1997] and eventually the industrialisation of England in the 18th century.

3 Literature

Our paper contributes to several strands of literature.

First, we contribute to the literature on quantitative economic models - see Redding and Rossi-Hansberg [2017] for a review, and especially to the papers who take a dynamic perspective (Desmet and Rossi-Hansberg [2014]; Desmet et al. [2016]; Nagy [2017]; Allen and Donaldson [2018]). Two papers in this field, Allen and Arkolakis [2014] and Nagy [2017], are most closely related. In their Appendix 5, Allen and Arkolakis [2014] add a second traded sector to their baseline one-sector model, and characterize the properties of the equilibrium under the assumption of no spillovers in productivity or amenities. Nagy [2017] retains the one-sector model for goods traded across cities, but introduces a second sector, whose goods are produced in the countryside and shipped to one - and only one - urban market. His model thus features agricultural market areas that endogenously emerge around cities.

Our model incorporates elements of both papers. As in Nagy [2017], agricultural goods are produced in the countryside and traded in cities; however, unlike in Nagy [2017], and similarly to Allen and Arkolakis [2014][Appendix 5.2], agricultural goods are also exchanged across cities. This is crucial for our results: when agricultural goods are costly to trade across cities, urban growth is limited by the productivity of the surrounding rural area; in turn, rural productivity is bounded by diminishing returns to agricultural production *and* by *rural-urban* shipping costs[1]. In contrast, when agricultural trade costs fall, cities can substitute local production with imports of agricultural goods, thus relaxing the agricultural constraint and triggering the boom of the manufacturing sector.

Our paper is also related to Fajgelbaum and Redding [2018] [2], who investigate the link between falling trade frictions and structural transformation in Argentina from 1869 to 1914. There are, however, two major differences. First, In Fajgelbaum and Redding [2018] traded goods are homogenous, so that trading locations fully specialize in one production sector, according to their comparative advantage; moreover, there may be a set of remote locations that do not open up to trade and produce all goods in autarky. In contrast, we assume differentiated traded goods in combination with CES preferences, with two implications: first, cities (with their surrounding rural areas) never fully specialize in one sector, because urban workers always need to consume at least a small amount of locally-produced agricultural goods (and viceversa); second, there exist no autarkic locations, since consumers always demand a positive amount of goods produced elsewhere. The second difference is that Argentina is treated as a small open economy, such that manufacturing prices are exogenously set on world markets. In contrast, we model Europe in pre-Industrial times as a closed economy, such that all prices are determined in general equilibrium.

Second, our paper is related to the literature on structural transformation and growth. Two (non-exclusive) mechanisms have been proposed to explain the secular shift from agriculture to manufacturing: first, an income elasticity of demand for farm goods lower than 1 (CaselliColeman2001); second, an elasticity of substitution across consumption goods lower than 1, combined with faster TFP growth in the sector whose labor share is declining NgaiPissarides2007 [3]. Our model features none of these two mechanisms, given that we assume Cobb-Douglas functional forms both on the demand and on the supply side; still, lowering agricultural trade costs yields an aggregate increasing in the manufacturing (urban) labor share.

The role of falling trade costs had been discussed by the New Economic Geography literature. In Krugman [1991], falling trade costs, combined with economies of scale in manufacturing, are associated with the emergence of manufacturing agglomerations, since manufacturing goods can be shipped more conveniently to immobile farmers; notably farm goods are freely traded. In contrast, Helpman [1998] finds that falling trade costs are associated with greater dispersion, the reason being that the other consumption good is non-traded and therefore subject to congestion forces; with low trade costs, agents can disperse to mitigate congestion forces and at the same time import manufacturing goods at a low price. Allen and Arkolakis [2014] show that increasing trade costs in one sector will reduce the agglomeration only if trade costs in the other sector is sufficiently low.

[1] Both elements are necessary: absent diminishing returns, all agricultural production could be concentrated inside the city, with no need to exploit new and more distant plots of land; in turn, absent shipping costs, cities could escape diminishing returns by exploiting more and more distant plots of land (although in this case the rural area would still be bounded by other cities' rural areas.

[2] See also [Coşar and Fajgelbaum, 2016].

[3] Relatedly, Nagy [2017] assumes that farm goods are used as an intermediate input in manufacturing production, in conjunction with labor, with an elasticity of substitution lower than 1. In this case, an increase in farm-good augmenting TFP is associated with an increase in the manufacturing labor share.

4 Model

In this section we present our gravity model of trade with two goods (manufacturing & agriculture), bilateral trade frictions, and labour mobility. In this grid-cell world, where the location and number of cities in each century is determined by the historical data of Bairoch, Batou, and Chèvre [1988], Farmers decide to which city to carry their goods. The model exhibits external economies of scale in manufacturing (taking place in cities exclusively), and congestion due to land use in agriculture.

4.1 Geography and endowments

The economy consists of a discrete set of cells $\mathcal{X}$, which we label rural cells. In a subset $\mathcal{Y}_t \subset \mathcal{X}$, there is a city. Elements of $\mathcal{X}$ are denoted by r or s, whereas the elements of $\mathcal{Y}$ are denoted by i or j. Note that, while the set of rural cells is fixed, the set of cities varies over time. We take the set of cities, as well as their location, as given in each period. Time is discrete and indexed by t, where $t = 1000, 1100, ..., 1800$. In the following, except when required to avoid confusion, we suppress the current time subscript t, and retain only the previous period subscript $t-1$.

There are two sectors in the economy. Cities produce a manufacturing (or urban) good, whereas rural locations produce an agricultural (or rural) good. We index sectors by $k = M, A$.

As in Nagy [2017], cities are trading places. That is, goods can only be exchanged in cities. Farmers commute to cities to sell agricultural goods and purchase manufacturing goods. We denote the set of rural cells trading with city i by Ω_i; we refer to Ω_i as the rural area around city i.

Rural cells differ in terms of agricultural productivity, in the amount of land available for agriculture, and in geographic position. All these fundamentals are taken to be time-invariant.[4] Urban cells further differ in terms of manufacturing productivity, which in turn is allowed to evolve over time.

Agents are *ex-ante* identical. All of them are endowed with one unit of labor that is supplied inelastically to the market. In equilibrium, they are employed in the urban sector (workers) or in the rural sector (farmers). Farmers own an equal share of land at their rural location r. Thus, urban income equals the urban wage rate, whereas rural income equals farm revenues per capita.

Labor is freely mobile across sectors and across locations. The total urban population in the economy, denoted by $\bar{L}_t^M$, is exogenously given at each time $t = 1000, 1100, ..., 1800$.

4.2 Production

The farm good is produced using labor l^A and land h under conditions of constant returns to scale. Farm output in r is given by:

$$y_{r,t}^A = \phi_r^A (l_{r,t}^A)^{1-\beta} (h_r)^\beta, \quad 0 < \beta < 1 \tag{1}$$

where ϕ_r^A is agricultural productivity, $l_{r,t}^A$ is the rural population in $r \in \mathcal{X}$, h_r is land area, and β is the the Cobb-Douglas parameter. The above expression makes clear that agricultural output may vary over time only as a consequence of varying population density in a given

[4] In the empirical application, cells differ in land area because of the earth curvature and because the coastline cuts through some of them.

rural cell. Land use in agriculture is the source of the congestion in the economy, given that land is available in fixed supply.

Manufacturing output in city s at time t given by:

$$Y_{s,t}^{M} = \phi_{s,t}^{M} L_{s,t}^{M}, \tag{2}$$

where $L_{s,t}^{M}$ is the labor force employed in manufacturing. Manufacturing TFP, $\phi_{s,t}^{M}$, is the source of agglomeration spillovers in the economy. We assume that spillovers are both static and dynamic. That is, we allow both current and past population to influence the current productivity in a given city. Following Allen and Donaldson (2018), we use the following specification:

$$\phi_{s}^{M} = \bar{\phi}^{M} (L_{s}^{M})^{\gamma_1} (L_{s,t-1}^{M})^{\gamma_2}, \tag{3}$$

where $\bar{\phi}^{M}$, the exogenous TFP component, is assumed to be constant across cities. According to (3), differences in manufacturing productivity across European cities are due to differences in current and in previous period urban population. The parameter γ_1 captures the strength of contemporaneous externalities, while γ_2 captures the strength of dynamic externalities. Since the TFP scale doesn't affect the population distribution, we can normalize $\bar{\phi}^{M} = 1$.

Firms take ϕ_{s}^{M} at the time of making their choices. Given perfect competition and constant returns to scale, the off-the-shelf price in city s equals the marginal cost:

$$p_{s}^{M} = \frac{w_s}{\phi_{s}^{M}}. \tag{4}$$

Finally, firms make zero profit in equilibrium, so

$$p_{i}^{M} Y_{i}^{M} = w_i L_{i}^{M}. \tag{5}$$

4.3 Trade structure

Agricultural goods are shipped from the countryside to the urban market. Then, from cities, both goods can be traded to other cities. We model trade as Armington (1961). Thus each rural area Ω_i produces a unique variety of manufacturing and agricultural goods, and consumers wish to consume all varieties. Note that farm good varieties acquire their identity, to the consumer's eyes, at the trading location, rather than at the production location.

Shipping goods from one location to another is costly. We assume that trade costs take the iceberg form: it takes $D(r,i)$ units of the farm good to ship 1 unit from rural cell $r \in \mathcal{X}$ to city $i \in \mathcal{Y}$. Similarly, it takes $T^{M}(i,j)$ and $T^{A}(i,j)$ units of, respectively, manufacturing good and agricultural good, to ship one unit from one city to another.

Given no arbitrage conditions, delivery prices from city j to city i are given by: $p_{ij}^{M} = T^{M}(i,j) p_{j}^{M}$ and $p_{ij}^{A} = T^{A}(i,j) p_{j}^{A}$, for manufacturing and agricultural goods respectively.

Finally, we parametrize trade costs to be an exponential function of distance $d(r,s)$, with $r, s \in \mathcal{X}$. Thus, we have $D(r,i) = e^{\delta d(r,s)}$, $T^{M}(i,j) = e^{\tau^{M} d(i,j)}$, and $T^{A}(r,i) = e^{\tau^{A} d(i,j)}$.

4.4 Consumer's problem

Agents order consumption baskets according to Cobb-Douglas preferences. We follow Nagy (2018) and assume that all agents, including farmers, consume their goods at the trading place.[5]. Then, all agents inside a rural area Ω_i will face the same consumption prices, and, as a consequence, they will also have the same nominal income v_i, since at the spatial equilibrium welfare equalizes between farmers and workers. We first solve the consumer's problem for a representative agent living in rural area Ω_i, taking Ω_i as given; then, deal with the farmer's choice of where to ship her goods.

An agent in Ω_i solves:

$$\max_{\{c_{ij}^M, c_{ij}^A\}_{j\in\mathcal{Y}}} \quad (C_i^M)^\alpha (C_i^A)^{1-\alpha}$$
$$\text{s.t.} \quad \sum_{j\in\mathcal{Y}} p_{ij}^M c_{ij}^M + \sum_{j\in\mathcal{Y}} p_{ij}^A c_{ij}^A = v_i,$$

where C^M and C^A are CES bundle of the good varieties imported from all cities:

$$C_i^M = \left(\sum_{j\in\mathcal{Y}} (c_{ij}^M)^{\frac{\sigma^M-1}{\sigma^M}}\right)^{\frac{\sigma^M}{\sigma^M-1}},$$
$$C_i^A = \left(\sum_{j\in\mathcal{Y}} (c_{ij}^A)^{\frac{\sigma^A-1}{\sigma^A}}\right)^{\frac{\sigma^A}{\sigma^A-1}},$$

p_{ij}^M and p_{ij}^A are the prices in i of, respectively, the manufacturing and agricultural variety imported from j , and α is the share of expenditure devoted to manufacturing goods. The parameter α, together with β, determine the strength of congestion forces in the economy, since they capture how much fixed land supply bites on consumer's welfare. Nominal income v_i is equal to the wage rate, w_i, for urban workers, and to farm revenues per capita, $(p_i^A/D(r,i))\times(y_r^A/l_r^A)$, for farmers. Therefore, the indirect utilities, V_i^M and $V_{r,i}^A$ respectively, are given by:

$$V_i^M = \frac{w_i}{(P_i^M)^\alpha (P_i^A)^{1-\alpha}}, \tag{6}$$
$$V_{r,i}^A = \frac{(p_i^A/D(r,i))\times(y_r^A/l_r^A)}{(P_i^M)^\alpha (P_i^A)^{1-\alpha}}, \tag{7}$$

where P^M and P^A are the standard dual price indices for CES demands; their expressions are given by:

$$P_i^M = \left(\sum_{j\in\mathcal{Y}} (T_{ij}^M)^{1-\sigma^M} (p_j^M)^{1-\sigma^M}\right)^{\frac{1}{1-\sigma^M}}, \tag{8}$$
$$P_i^A = \left(\sum_{j\in\mathcal{Y}} (T_{ij}^A)^{1-\sigma^A} (p_j^A)^{1-\sigma^A}\right)^{\frac{1}{1-\sigma^A}}. \tag{9}$$

[5] Nagy(2018) also solves the model under the more plausible assumption that farmers consume their goods at their home place, and thus incur in different shipping costs; he finds that it doesn't make a big difference. Nevertheless, we might consider working on this assumption

4.5 Gravity

Solving for consumer's demands, the total value of city i's imports from city j is given by:

$$X_{ij}^M = \alpha (T_{ij}^M)^{1-\sigma^M} (p_j^M)^{1-\sigma^M} (P_i^M)^{\sigma^M - 1} w_i L_i, \quad (10)$$

$$X_{ij}^A = (1-\alpha)(T_{ij}^A)^{1-\sigma^A} (p_j^A)^{1-\sigma^A} (P_i^A)^{\sigma^A - 1} w_i L_i. \quad (11)$$

These are the standard gravity equations from the international trade literature.

4.6 Welfare equalization

First, all farmers in a rural area must receive the same welfare in equilibrium: $V_{r,i}^A = V_{s,i}^A$, for all $r, s \in \Omega_i$ and all $i \in \mathcal{Y}$. Using this condition together with the production function (1), we derive an expression for the rural population in each rural area Ω_i, where again we take the spatial tessellation as given:

$$l_r^A = \frac{\left(\phi_r^A / D(r,i)\right)^{\frac{1}{1-\beta}} h_r}{\sum_{s \in \Omega_i} \left[\left(\phi_s^A / D(r,i)\right)^{\frac{1}{1-\beta}} h_s\right]} \sum_{s \in \Omega_i} l_s^A. \quad (12)$$

Rural population in r is a fraction of the total rural population in the corresponding rural area; the term $\left(\phi_r^A / D_{r,i}\right)^{\frac{1}{1-\beta}} h_r$ represents the effective agricultural productivity of rural cell r, net of shipping costs. Let us define: $L_i^A = \sum_{s \in \Omega_i} l_s^A$ and

$$Y_i^A = \sum_{s \in \Omega_i} \left[\left(\phi_s^A / D(r,i)\right)^{\frac{1}{1-\beta}} h_s\right]. \quad (13)$$

Thus, L_i^A is the total rural population in rural Ω_i, and Y_i^A is the total effective agricultural productivity in the rural area. Note that the only unknown term in Y_i^A is the set of rural cells Ω_i.

Secondly, welfare equalization between farmers and workers within a rural area implies

$$\frac{p_i^A}{D_{r,i}} \frac{y_r^A}{l_r^A} = w_i \quad \text{for all } r \in \Omega_i, \text{ and all } i \in \mathcal{Y}. \quad (14)$$

Summing over all cells in Ω_i and isolating p_i^A, we obtain:

$$p_i^A = \frac{L_i^A}{\sum_{s \in \Omega_i} \frac{y_r^A}{D(r,i)}} w_i.$$

Using the production function (1) and equation (12), we obtain:

$$p_i^A = \left(\frac{L_i^A}{Y_i^A}\right)^{1-\beta} w_i. \quad (15)$$

Equations (4) and (15) define the "city-gate prices" of manufacturing and agricultural varities in city i, before trade frictions are incurred.

Finally, welfare equalizes for urban workers living in different cities:

$$V_i^M = V, \quad \text{for all } i \in \mathcal{Y}. \quad (16)$$

4.7 Market clearing

We require all markets to clear. Therefore, total revenues of a sector in each location $i \in Y$ must equal the total value of exports to all other locations. We have:

$$p_i^M Y_i^M = \sum_{j \in \mathcal{Y}} X_{ji}^M$$

in the manufacturing sector, and

$$p_i^A \sum_{r \in \Omega_i} \frac{Y_r^A}{D_{r,i}} = \sum_{j \in \mathcal{Y}} X_{ji}^A$$

in the agricultural sector. Given that manufacturing firms make zero profits - equation (5) - and that welfare equalizes between the city and the countryside, we can rewrite the market clearing conditions as:

$$w_i L_i^M = \sum_{j \in \mathcal{Y}} X_{ji}^M, \tag{17}$$

$$w_i L_i^A = \sum_{j \in \mathcal{Y}} X_{ji}^A \tag{18}$$

4.8 Rural areas

The previous discussion holds for a given tessellation of Europe in rural areas: $\{\Omega_i\}_{i \in \mathcal{Y}}$. We now show how to determine the rural areas in equilibrium given the choices of farmers. Using (16), we rewrite the indirect utility of a farmer who lives in r and trades with s, given in equation (6), as:

$$V_{r,i}^A = \frac{p_i^A}{D_{r,i} w_i} \frac{y_r^A}{l_r^A} V. \tag{19}$$

This equation shows that farmer's welfare depends on i through: first, the ratio of the agricultural price to the wage rate, p_i^A / w_i; and, second, travel distance to all cities. Farmers in r solve:

$$\max_{i \in \mathcal{Y}} V_{r,i}.$$

Therefore we can write the rural area around city i in the following way:

$$\Omega_i = \{r : i = \operatorname*{argmax}_{j \in \mathcal{Y}} V_{r,i}\}. \tag{20}$$

4.9 Equilibrium

Definition 1. *A competitive equilibrium in this economy is a set of price vectors: $\{P^A, P^M, p^A, p^M, w\}$, population distribution across regions and sectors $\{L, L^A, L^M\}$, and a common welfare level V, such that:*

1. *the markets for manufacturing and agricultural goods clear, (17) and (18);*

2. *the price indexes are given by (8) and (9);*
3. *welfare equalizes across cities (16), and within rural areas (14);*
4. *local labor markets clear $L_s^A + L_s^M = L_s$, and the aggregate urban population constraint holds: $\sum_{i \in \mathcal{Y}} L_s^M = \bar{L}^M$;*
5. *rural areas are constructed according to (20);*

and where, furthermore: bilateral trade expenditures are given by (10) and (11), factory prices are given by (4) and (15), and manufacturing TFP is given by (3).

After some algebra, we can write the equilibrium as a system of $7 \times |\mathcal{Y}| + 1$ equations in terms of the same number of unknowns: the $|\mathcal{Y}|$-dimensional vectors: $w, P^M, P^A, L^M, L^A, L, \lambda$, and the welfare scalar V. If rural areas were given, this would be the end of the story. With endogenous rural areas, we must also consider the set of $|\mathcal{Y}|$ expressions used to construct the sets $\Omega_i, i \in \mathcal{Y}$.

In sum, we have:

1. the market clearing condition for the manufacturing good:
$$w_s^{\sigma^M} (L_s^M)^{1-(\sigma^M-1)\gamma_1} = \alpha \sum_{i \in S} (\tau_{s,i}^M)^{1-\sigma^M} (P_i^M)^{\sigma^M-1} (\hat{\phi}_s^M)^{\sigma^M-1} w_i L_i, \tag{21}$$
2. the expression for the manufacturing price index:
$$(P_s^M)^{1-\sigma^M} = \sum_{i \in S} (\tau_{s,i}^M)^{1-\sigma^M} (\hat{\phi}_i^M)^{\sigma^M-1} w_i^{1-\sigma^M} (L_i^M)^{(\sigma^M-1)\gamma_1}, \tag{22}$$
3. the market clearing condition for the farm good:
$$w_s^{\sigma^A} (L_s^A)^{1+(\sigma^A-1)(1-\beta)} = (1-\alpha) \sum_{i \in S} (\tau_{s,i}^A)^{1-\sigma^A} (P_i^A)^{\sigma^A-1} B_s^{(\sigma^A-1)(1-\beta)} w_i L_i, \tag{23}$$
4. the expression for the agricultural price index:
$$(P_s^A)^{1-\sigma^A} = \sum_{i \in S} (\tau_{s,i}^A)^{1-\sigma^A} B_i^{(\sigma^A-1)(1-\beta)} w_i^{1-\sigma^A} (L_i^A)^{-(\sigma^A-1)(1-\beta)}, \tag{24}$$
5. welfare equalization across cities:
$$w_s = V (P_s^M)^{\alpha} (P_s^A)^{1-\alpha}, \quad \forall s \in S, \tag{25}$$
6. the local population constraint:
$$L_s = L_s^M + L_s^A, \quad \forall s \in S, \tag{26}$$
7. the aggregate population constraint:
$$\sum_{i \in S} L_s = \bar{L}, \tag{27}$$

with the understanding that the tessellation is built according to (20).

4.10 Connecting the Model with Historical Context

Consumption We assume a constant consumption basket across time. This is not decisive, we could allow for varying/increasing share of manufacturing consumption. In fact the share of manufacturing did not change decisively in our study period. This is warranted by the fact that there was just not a lot of choice when it came to manufacturing. It is mostly cloth, candles, oil, soap [e.g. Allen, 2001; Malanima, 2009]. And furthermore a typical reaction to rising wages by the pre-industrial populace was to reduce the working time because preferences were such that once the desired consumer basket was achieved, leisure was preferred. We don't allow for leisure in our consumption function but this feature is reflected by the fact that the consumption shares stay constant over time. On top of that, there is plenty of evidence that the substitution that took place happened between goods categories and not across. There was usually a desire to find grain substitutes for bread and also to increase the share of meat [Abel, 1981]. All of these are classified as "agricultural goods" for us.

Agriculture This contains everything that is food. Also processed food. Barley for ale, wheat/rye for bread, oats for horses and oxen [see e.g. Malanima, 2009]. Later there were summergrains like buckwheat.
New world food such as the potato do not interfer with this in our setting, as it was adopted on a large scale only after 1800 [see e.g. Pounds, 1979] (except for very few places it was mostly just used as a substitute and in the early periods exclusively used as animal fodder, as people refused to eat it even during famine)

One simplifying assumption that we make for now is that agricultural TFP is constant. We could allow for it to vary (increase) over time, and this would arguably be "on our side" for it will improve the fit to the data.

Free mobility & Migration Another assumption that seems ad-hoc is the free mobility of people. Taking into account novel historical evidence such as Lucassen and Lucassen [2009] or Schäfer [2013], disproving the old conjecture by Zielinsky, this assumption does not seem as unrealistic anymore. The fact that we do not want to track individual people but are interested in aggregate movements further justifies our assumption.

5 Data

The core empirical element of our model is the well known data on urban settlements by Bairoch [1988]; Bairoch et al. [1988] reported in centurywise increments. We supplement this data with recent corrections by Bosker, Buringh, and van Zanden [2013]. The second main ingredient is the agricultural TFP that enters the production function of our model. We make use of the caloric suitability index by Galor and Özak [2016] which is reported on a 5 arcminute by 5 arcminute level - thus finer than our quater degree grid-cell world - and based on the widely used suitability index by Ramankutty, Foley, Norman, and McSweeney [2002].

The only other data that we take into consideration which is not related to features of first nature geography is the Roman road network for the computation of the transport costs. Since this road network was an important feature in European development but was set up way before the start of our study period we have to exogenously take it into account. The data comes from McCormick, Zambotti, Grigoli, and Gibson [2007] and Åhlfeldt [2015].

First Nature Geography Data on navigable rivers and elevation are from the Natural Earth database.

5.1 Trade Costs

Following Allen and Arkolakis [2014] we create an instantaneous cost function $\tau : S \rightarrow \mathbb{R}^+$ that assigns values based on first nature geography to each of of our observations, i.e. each gridcell. For land transport we take into account the ruggedness[6] and elevation of each gridcell, the higher those values are, the more costly it is to pass this gridcell. When there is a roman road present in the gridcell, the respective τ decreases, ruggedness still taken into account. Water transport was always cheaper by several orders of magnitude and cells that have a navigable river in it are cheaper to pass. In order to avoid having to handselect which rivers were navigable or not we also account here for elevation and ruggedness. The fastest mode of transport is via the open sea. For the time being we do not let the transport costs vary over time and thus abstract from technical improvements since they arguably did not affect intra European trade substantially during our study period from 1000 until 1800. Furthermore we abstract from different transport costs for agricultural and manufacturing goods for the time being. Figure 2 shows the instantaneous tradecosts that have to be incurred in order to pass the respective gridcell.
For any origin $i \in S$ and destination $j \in S$ pair we apply the FMM algorithm[7] in order to determine the bilateral transport costs between all of our gridcells. The resulting matrix is then of dimension $|S| \times |S|$. Figure 3 illustrates the gradient of transport costs for the origin cell which contains the city of Bologna: one can see that open sea transport matters quite a lot and that the Alps to the North are a quite prohibitive barrier. Transporting goods for instance to the interior of Southern Italy or Sicily is costlier than to, say, Barcelona.

[6] Nunn and Puga [2012] based on Riley, DeGloria, and Elliot [1999]

[7] Implemented in python, outperforming the standard Dijkstra algorithm by a tremendous magnitude. The compution of the transport cost matrix on our network would have taken more than 10 years on a fairly strong personal computer. The FMM delivered a solution within around one minute on the same machine.

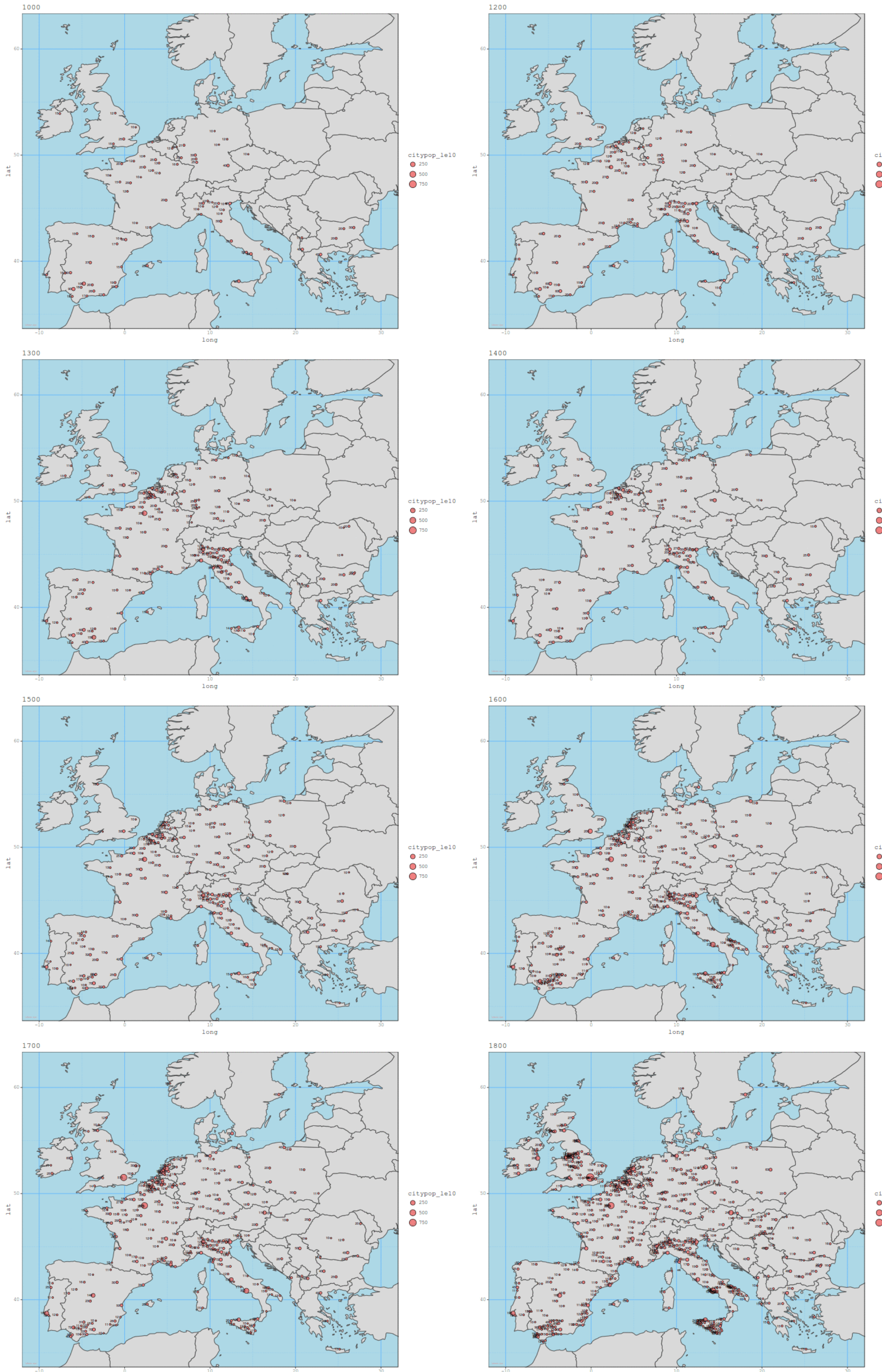


Figure 1: Visualisation of the Bairoch data throughout time.

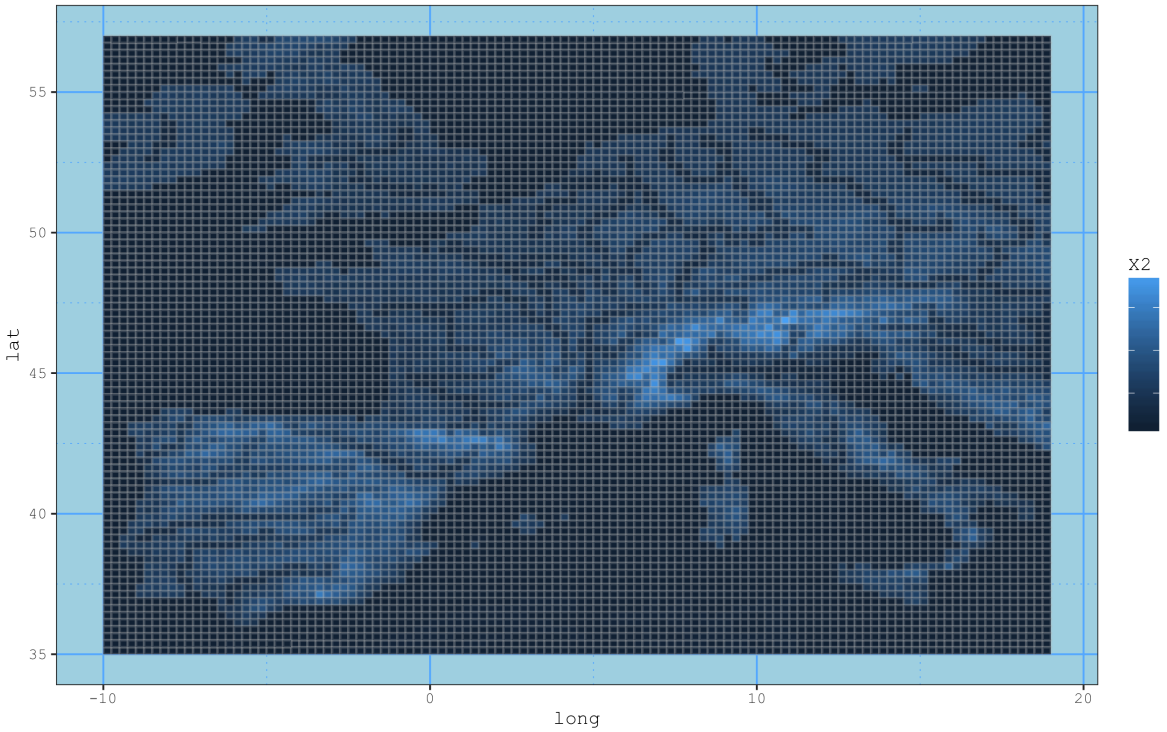


Figure 2: Instantaneous tradecosts across europe.

5.2 Calibration

- Share of manufacturing consumption
  $\alpha = 0.3$ [Allen, 2001]

- Elasticity of substitution between manufacturing varieties
  $\sigma^A = \sigma^M = 4$ (calibrated from CES literature)

- Labor share in agricultural production
  $\beta = 0.7$ [Grigg, 1980, 1992]

- Manufacturing parameters (persistence)
  $\gamma_1 = 0.144$, $\gamma_2 = 0.033$ [Allen and Donaldson, 2018, Table 2]

The parameters for trade (taus) we will then pin down with our simulation exercises.

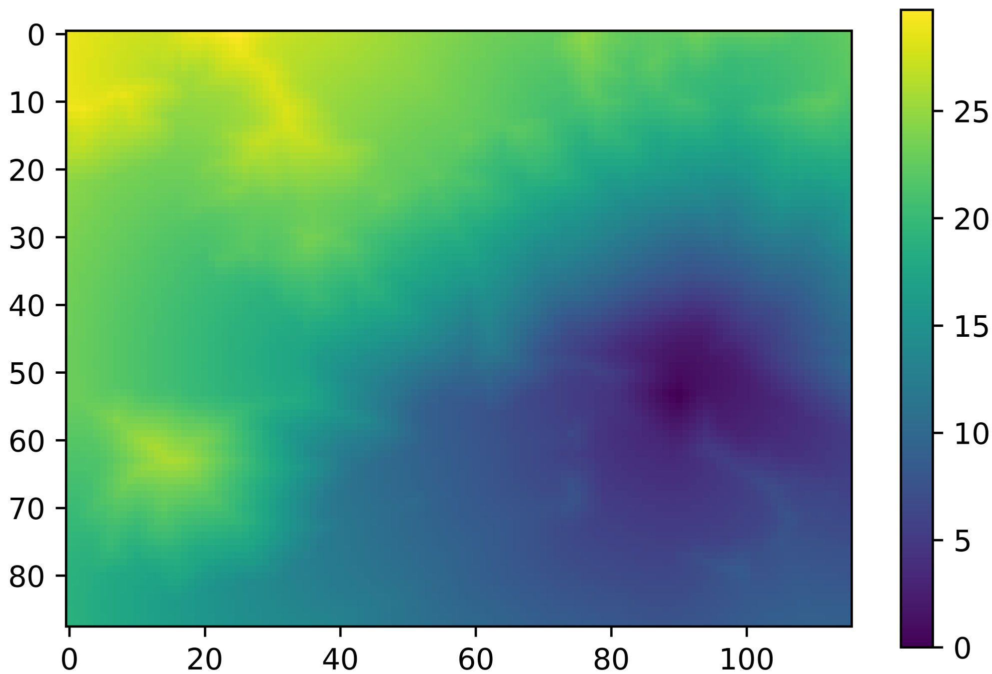


Figure 3: For illustrative purposes this plot shows all $|S|$ (i.e. 10208 for our 0.25 degree cell level) iceberg transport costs that one has to incur travelling from Bologna to all the other grids in our sample. The dark-blue epicenter is Northern Italy, and one can spot the shape of the peninunsula and the coastline in the West towards the Iberian peninsula.

6 Simulations

In this section we will show the intuition of the model and try to pin down the time evolution of both trade parameters in order to show that it was in fact the reduction in (effective) agriculutral trade costs that allowed the urban giants in Europe to form.

Measures of Fit Eventually we are interested in fitting the observed city size distribution in the data which is getting more and more right skewed from century to century. We thus compare the second to fourth moments, century by century, of both the distribution of the Bairoch data and the one which is implied by our model. Standard deviation, skewness, and kurtosis are reported as fractions, data divided by model, thus 1 representing a perfect fit. To capture the right-tail we also look at the distance between mean and median (”M-M”), where, as usual, high values indicate the existence of a right tail. For completeness we report the plain correlation between model and data in addition to the mean squared error (MSE) as a measure of fit. An equivalent procedure would be to use a Kolmogorov-Smirnoff test [see e.g. Conover, 1999] or a 2-sample Anderson-Darling test [Scholz and Stephens, 1987] in order to obtain a single, unified indicator of fit.
To illustrate that we also match the shift of urban/economic gravity from the South to the North of Europe, we add columns that show the share of the urban mass that is in the North of Europe.
In order to visualise the specialisation mechanism of our framework, we report two coeffi-

cients of a regression of the agricultural population ("LA") and the urban/manufacturing population ("LM") for each city and its hinterland. The former represents the full sample, whereas the latter ("LALMbig") is the same regression carried out on the sample of towns above 50.000 inhabitants. Once we account for the proper reduction in agricultural trade costs, we observe that bigger cities now require less workers in the agricultural sector, thus a smaller hinterland, because grain can be shipped in.

A very conservative first run ("autarky") For now we arbitrarily choose the set of parameters as follows:

- Share of manufacturing consumption

$$\alpha = 0.3$$

- Elasticity of substitution between manufacturing varieties

$$\sigma^A = \sigma^M = 4$$

- Labor share in agricultural production

$$\beta = 0.7$$

- Manufacturing parameters (persistence)

$$\gamma_1 = 0.1 \quad \boldsymbol{\gamma_2 = 0.2}$$

- Trade cost parameters

$$\boldsymbol{\tau^{A} = 50 \quad \tau^{M} = 10 \quad \delta = 30}$$

We then simulate the model over the full period of 1000 until 1800 and compare the model outcome to the data in each century. The results are then reported in Table 1. As can be seen from this and the mapplots, we do a very bad job in matching the right tail of the distribution. The model wants to produce something bell shaped, wheras the actual urban data looks pareto shaped with a long right tail.

Table 1: Data and model results moments compared. SD, Skew, Kurt are presented as simulation divided by data. Thus one being a perfect moment fit. M-M resembles the distance of Median to Mean in data and model.

	year	Cor	MSE	LALM	LALMbig	N/S-CLL	N/S-dat	SD	Skew	Kurt	M-M	M-M
1	X10_5kLM	0.47	215.12	2.33	NaN	0.22	0.25	0.49	0.09	0.06	0.90	5.35
2	X11_5kLM	0.52	158.53	2.33	E	0.20	0.27	0.46	0.09	0.06	0.79	3.08
3	X12_5kLM	0.55	190.64	2.33	2.34	0.21	0.32	0.60	0.14	0.10	0.89	5.84
4	X13_5kLM	0.54	375.24	2.33	2.33	0.24	0.38	0.32	0.05	0.01	4.06	7.69
5	X14_5kLM	0.56	312.52	2.33	2.33	0.31	0.46	0.40	0.08	0.03	5.20	7.62
6	X15_5kLM	0.58	234.09	2.33	2.33	0.28	0.43	0.41	0.10	0.03	3.31	5.96
7	X16_5kLM	0.54	499.74	2.33	2.1	0.23	0.41	0.22	0.03	0.01	3.50	7.62
8	X17_5kLM	0.49	1302.70	2.32	2.1	0.26	0.54	0.11	0.01	0.00	3.41	9.26
9	X17.5_5kLM	0.57	1355.05	2.33	2.34	0.29	0.56	0.11	0.01	0.00	3.94	10.43
10	X18_5kLM	0.53	1376.57	2.28	1.94	0.27	0.63	0.10	0.01	0.00	2.94	9.19

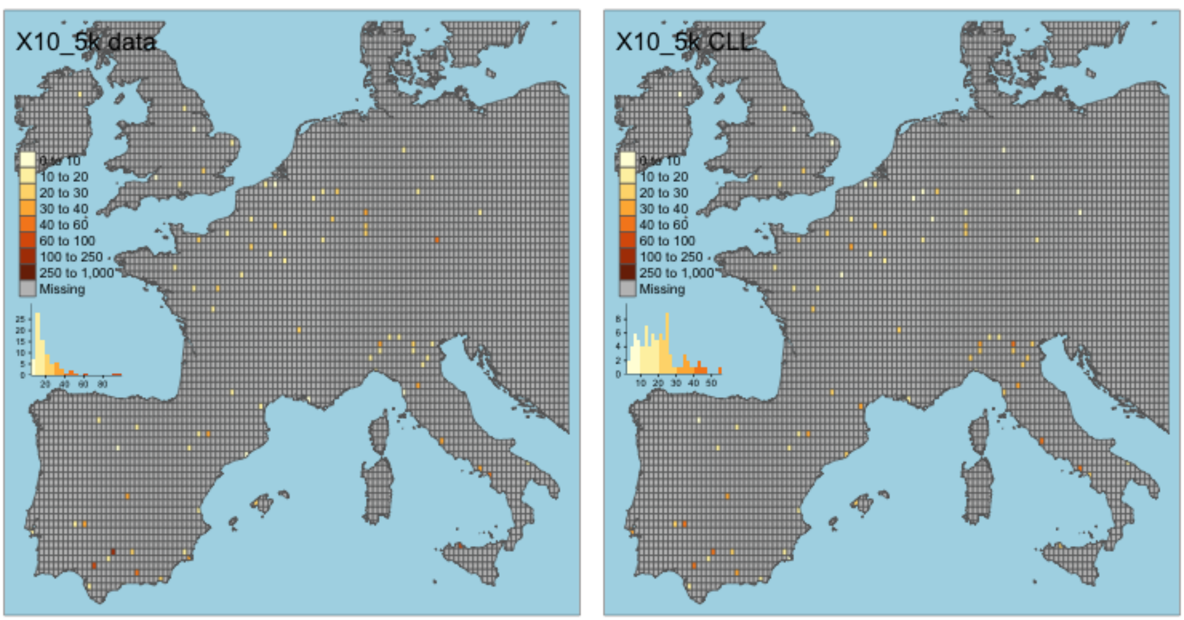


Figure 4: We do a bad job with $\tau^{\mathbf{A}} = 50$

Another run with reduced $\tau^{\mathbf{A}}$ The following exercise is going to illustrate the main logic of the model. We will leave everything else constant, but reduce (effective) agricultural trade costs. This allows the model to create the very big cities and we do a good job in matching the right tails. Getting very close to real-world city distributions.

- Share of manufacturing consumption

$$\alpha = 0.3$$

- Elasticity of substitution between manufacturing varieties

$$\sigma^A = \sigma^M = 4$$

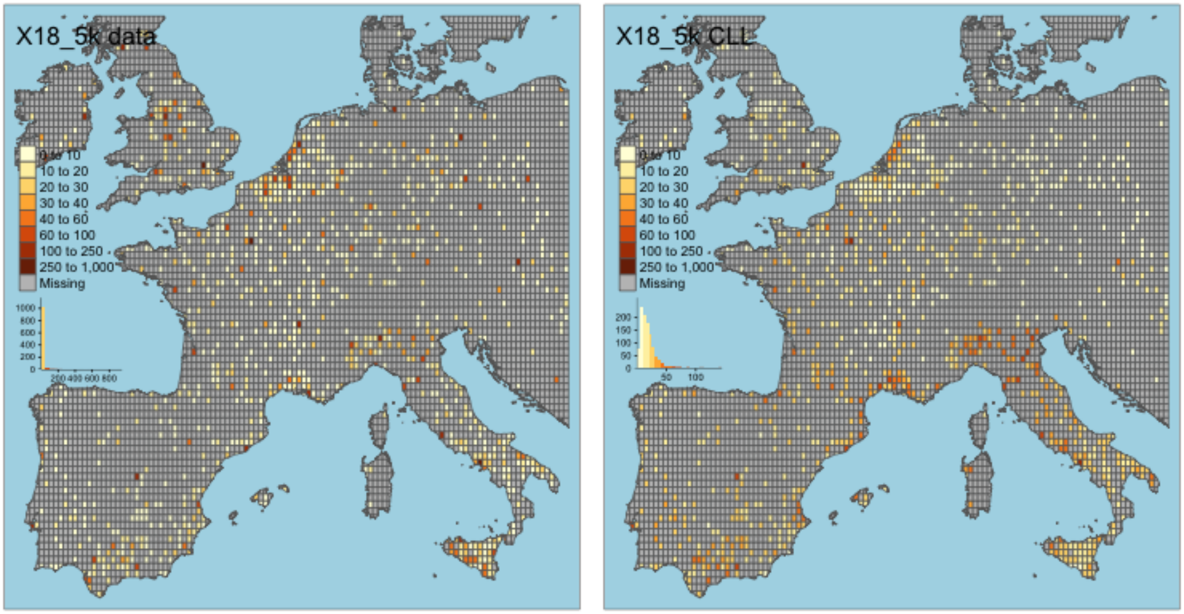


Figure 5: We do a bad job with $\tau^{\mathbf{A}} = 50$

- Labor share in agricultural production

$$\beta = 0.7$$

- Manufacturing parameters (persistence)

$$\gamma_1 = 0.1 \qquad \gamma_2 = 0.1$$

- Trade cost parameters

$$\tau^{\mathbf{A}} = \mathbf{10} \qquad \tau^{\mathbf{M}} = \mathbf{10} \qquad \delta = \mathbf{30}$$

Table 2: Data and model results moments compared. SD, Skew, Kurt are presented as simulation divided by data. Thus one being a perfect moment fit. M-M resembles the distance of Median to Mean in data and model.

	year	Cor	MSE	LALM	LALMbig	N/S-CLL	N/S-dat	SD	Skew	Kurt	M-M	M-M
1	X10_5kLM	0.66	155.84	1.71	-0.47	0.21	0.25	0.75	0.43	0.35	2.66	5.35
2	X11_5kLM	0.66	150.92	1.31	0.51	0.18	0.27	1.11	2.31	4.49	2.18	3.08
3	X12_5kLM	0.61	264.94	1.21	0.32	0.19	0.32	1.57	4.04	9.21	2.40	5.84
4	X13_5kLM	0.54	460.14	1.20	0.45	0.24	0.38	0.86	0.64	0.49	6.21	7.69
5	X14_5kLM	0.54	352.70	1.57	0.22	0.31	0.46	0.66	0.42	0.34	5.40	7.62
6	X15_5kLM	0.51	406.26	1.01	0.27	0.34	0.43	1.36	1.65	1.80	6.15	5.96
7	X16_5kLM	0.25	4126.17	0.27	0.15	0.46	0.41	6.07	4.88	27.79	7.86	7.62
8	X17_5kLM	0.18	12964.10	0.20	0.14	0.74	0.54	7.79	4.74	25.42	10.50	9.26
9	X17.5_5kLM	0.23	9364.92	0.26	0.19	0.64	0.56	5.09	2.11	8.49	10.52	10.43
10	X18_5kLM	0.48	5864.20	0.30	0.24	0.54	0.63	4.26	1.16	2.05	10.89	9.19

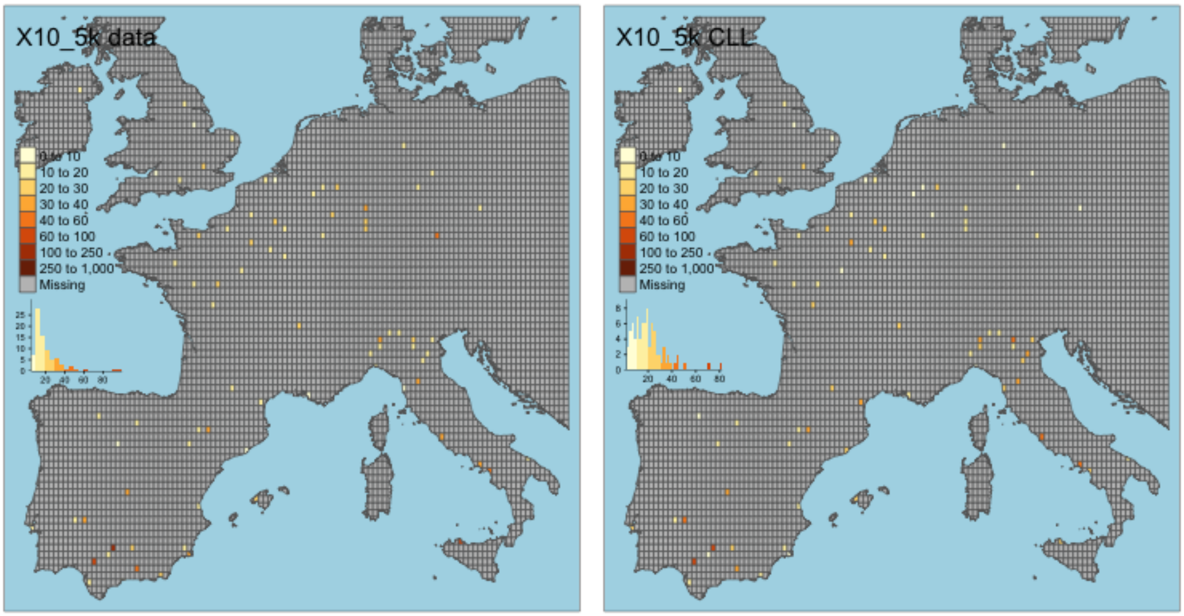


Figure 6: We seem to get the tails right with $\tau^{\mathbf{A}} = 10$

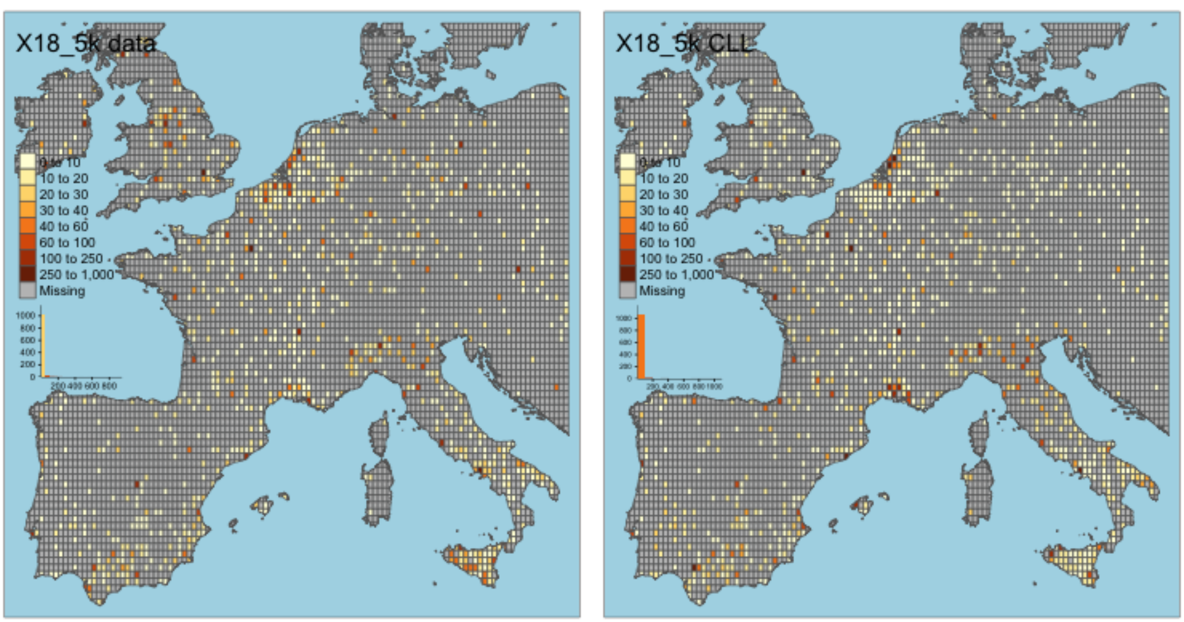


Figure 7: We seem to get the tails right with $\tau^{\mathbf{A}} = 10$

7 Conclusion

This paper proposes a quantitative economic geography model with two sectors and tradable goods. We utilise it to study the pre-conditions for industrialisation and the evolution of the European economic system, taking into account the whole continent and all its spatial features. Our framework highlights the role of falling trade costs for agricultural goods, in combination with economies of scale in the manufacturing sector, as a source of urban growth and structural transformation. The model suggests that the observed increase in agricultural productivity alone was not enough for the emergence of large urban agglomerations in Europe. We show that the market access for agricultural goods, decisively shaped by the peculiar geography of Europe with its long coastline and network of navigable rivers, was much more important than previously thought. It allowed European cities to overcome the natural constraint of needing a fertile hinterland in order to feed the urban population at a very early stage. On aggregate the result was the emergence of the tightly knit urban network, allowing more and more people to move out of agriculture, fostering trade, and the flow of ideas and people. These were necessary pre-conditions for the onset of early per capita growth, the overtaking of the historically much richer regions of South- and East-Asia, and, eventually, for the onset of the first industrial revolution.

References

ABEL, W. (1980): *Agricultural Fluctuations in Europe: From the Thirteenth to the Twentieth Centuries*, Taylor & Francis.

——— (1981): *Stufen der Ernährung: Eine historische Skizze*, Göttingen: Vandenhoeck & Ruprecht.

ÅHLFELDT, J. (2015): "Digital Atlas of the Roman Empire," Http://dare.ht.lu.se/.

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, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129, 1085–1140.

ALLEN, T. AND D. DONALDSON (2018): "The Geography of Path Dependence," *mimeo*, 48.

BAIROCH, P. (1988): *Cities and Economic Development*, University of Chicago Press.

BAIROCH, P., J. BATOU, AND P. CHÈVRE (1988): *La Population des villes européennes de 800 à 1850*, Geneve: Droz.

BARTLETT, R. (1993): *The Making of Europe: Conquest, Colonization and Cultural Change 950 - 1350*, Allen Lane.

BOSKER, M. AND E. BURINGH (2017): "City Seeds: Geography and the Origins of the European City System," *Journal of Urban Economics*, 98, 139–157.

BOSKER, M., E. BURINGH, AND J. L. VAN ZANDEN (2013): "From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa, 800–1800," *The Review of Economics and Statistics*, 95, 1418–1437.

BOTERO, G. (1588): *On the Causes of the Greatness and Magnificence of Cities*, delle Cause della Grandezza delle Città.

BRAUDEL, F. (1993): *A History of Civilizations*, London: Penguin Books, grammaire des Civilisations (1987).

CAMPBELL, B. M. S. (2016): *The Great Transition: Climate, Disease and Society in the Late-Medieval World*, 2013 Ellen Mcarthur Lectures, Cambridge University Press.

CONOVER, W. J. (1999): *Practical Nonparametric Statistics*, New York: John Wiley & Sons, 3rd edition ed.

COŞAR, A. K. AND P. D. FAJGELBAUM (2016): "Internal Geography, International Trade, and Regional Specialization," *American Economic Journal: Microeconomics*, 8, 24–56.

DE VRIES, J. (1984): *European Urbanization, 1500-1800*, London: Routledge.

DE VRIES, J. AND A. VAN DER WOUDE (1997): *The First Modern Economy: Success, Failure, and Perseverance of the Dutch Economy, 1500-1815*, Cambridge University Press.

Delventhal, M. (2018): "The Globe as a Network: Geography and the Origins of the World Income Distribution," *JMP*.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2016): "The Geography of Development," *mimeograph*.

Desmet, K. and S. L. Parente (2012): "The Evolution of Markets and the Revolution of Industry: A Unified Theory of Growth," *Journal of Economic Growth*, 17, 205–234.

Desmet, K. and E. Rossi-Hansberg (2014): "Spatial Development," *The American Economic Review*, 104, 1211–1243.

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

Donaldson, D. and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 131, 799–858.

Epstein, S. A. (2009): *An Economic and Social History of Later Medieval Europe, 1000-1500*, Cambridge University Press.

Fajgelbaum, P. D. and S. J. Redding (2018): "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914," *mimeo*.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999): "Geography and Economic Development," *International Regional Science Review*, 22, 179–232.

Galor, O. and Ö. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review*, 106, 3064–3103, https://ozak.github.io/Caloric-Suitability-Index/.

Gimpel, J. (1976): *The Medieval Machine: The Industrial Revolution of the Middle Ages*, Penguin Books.

Grigg, D. (1992): *The Transformation of Agriculture in the West*, Oxford, UK ; Cambridge, USA: Blackwell Pub, first edition edition ed.

Grigg, D. B. (1980): *Population Growth and Agrarian Change: An Historical Perspective*, Cambridge Eng. ; New York: Cambridge University Press.

Helpman, E. (1998): "The Size of Regions," in *Topics in Public Economics: Theoretical and Applied Analysis*, ed. by D. Pines, E. Sadka, and I. Zilcha, Cambridge: Cambridge University Press, 33–54.

Henderson, J. V. (1974): "The Sizes and Types of Cities," *American Economic Review*, 64, 640–56.

Henderson, J. V., Z. Shalizi, and A. J. Venables (2001): "Geography and Development," *Journal of Economic Geography*, 1, 81–105.

Henderson, V., T. Squires, A. Storeygard, and D. Weil (2018): "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade," *The Quarterly Journal of Economics*, 133, 357–406.

HOHENBERG, P. M. (2004): "The Historical Geography of European Cities: An Interpretive Essay," in *Handbook of Regional and Urban Economics*, Elsevier, vol. 4, 3021–3052.

HOHENBERG, P. M. AND L. H. LEES (1995): *The Making of Urban Europe, 1000-1994*, Harvard University Press.

JONES, E. L. (1968): "Agricultural Origins of Industry," *Past & Present*, 58–71.

KOSSMAN, E. H. (1970): "The Low Countries," in *The New Cambridge Modern History: The Decline of Spain and the Thirty Years War, 1609-48*, ed. by J. P. Cooper, Cambridge University Press, vol. 4.

KRUGMAN, P. (1991): *Geography and Trade*, Leuven Belgium ;Cambridge Mass. ;;London: Leuven University Press ;;The MIT Press.

——— (1999): "The Role of Geography in Development," *International regional science review*, 22, 142–161.

LOPEZ, R. S. (1976): *The Commercial Revolution of the Middle Ages, 950-1350*, Cambridge University Press.

LUCASSEN, J. AND L. LUCASSEN (2009): "The Mobility Transition Revisited, 1500–1900: What the Case of Europe Can Offer to Global History," *Journal of Global History*, 4, 347–377.

MALANIMA, P. (2009): *Pre-Modern European Economy: One Thousand Years (10th-19th Centuries)*, Boston, NETHERLANDS: BRILL.

——— (2013): "When Did England Overtake Italy? Medieval and Early Modern Divergence in Prices and Wages," *European Review of Economic History*, 17, 45–70.

MARSHALL, A. (1890): *Principles of Economics*, London: Macmillan.

MATSUYAMA, K. (1992): "Agricultural Productivity, Comparative Advantage, and Economic Growth," *Journal of Economic Theory*, 58, 317–334.

MCCORMICK, M., G. ZAMBOTTI, L. GRIGOLI, AND K. GIBSON (2007): *The Digital Atlas of Roman and Medieval Civilizations (DARMC)*, Cambridge (Massachusetts): Harvard University, https://darmc.harvard.edu/.

MICHAELS, G., F. RAUCH, AND S. REDDING (2012): "Urbanization and Structural Transformation," *The Quarterly Journal of Economics*, 127, 535–586.

MORRIS, I. (2010): *Why the West Rules-For Now: The Patterns of History, and What They Reveal About the Future*, New York: Farrar Straus & Giroux.

NAGY, D. K. (2017): "City Location and Economic Development," *JMP*.

NUNN, N. AND D. PUGA (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94, 20–36.

OVERTON, M. (1996): *Agricultural Revolution in England: The Transformation of the Agrarian Economy 1500-1850*, Cambridge ; New York: Cambridge University Press.

PERPILLOU, A. V. (1977): *Human Geography*, London: Longman Publishing Group.

PIRENNE, H. (1927): *Medieval Cities: Their Origins and the Revival of Trade.*

POUNDS, N. J. G. (1973): *An Historical Geography of Europe 450 B.C.-A.D. 1330*, CUP Archive.

——— (1979): *An Historical Geography of Europe, 1500-1840*, Cambridge University Press.

RAMANKUTTY, N., J. A. FOLEY, J. NORMAN, AND K. MCSWEENEY (2002): "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change," *Global Ecology and Biogeography*, 11, 377–392.

REDDING, S. AND E. ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

REDDING, S. J. (2016): "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148–167.

RILEY, S. J., S. D. DEGLORIA, AND R. ELLIOT (1999): "A Terrain Index That Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 5, 23–27.

SCHÄFER, R. (2013): "Rural–Urban Migrations, Medieval Era," in *The Encyclopedia of Global Human Migration*, American Cancer Society.

SCHOLZ, F. W. AND M. A. STEPHENS (1987): "K-Sample Anderson–Darling Tests," *Journal of the American Statistical Association*, 82, 918–924.

SCHULTZ, T. W. (1968): *Economic Growth and Agriculture*, McGraw-Hill.

SERRA, A. (1613): *A 'Short Treatise' on the Wealth and Poverty of Nations*, Anthem Press, breve trattato delle cause che possono far abbondare li regni d'oro e argento dove non sono miniere, con applicazione al Regno di Napoli.

SMITH, A. (1776): *An Inquiry into the Nature and Causes of the Wealth of Nations*, London: W. Strahan and T. Cadell.

SMITH, C. T. (1978): *An Historical Geography of Western Europe before 1800*, Geographies for Advanced Study, London: Longman Pub Group, revised ed., first edition 1967.

STELDER, D. (2005): "Where Do Cities Form? A Geographical Agglomeration Model for Europe," *Journal of Regional Science*, 45, 657–679.

TIELHOF, M. (2002): *The Mother of All Trades: The Baltic Grain Trade in Amsterdam from the Late 16th to the Early 19th Century*, Leiden ; Boston: Brill.

VAN TIELHOF, M. (1995): *De Hollandse Graanhandel, 1470-1570: Koren Op de Amsterdamse Molen*, Den Haag: Stichting Hollandse Historische Reeks.

VAN ZANDEN, J. L. AND B. VAN LEEUWEN (2012): "Persistent but Not Consistent: The Growth of National Income in Holland 1347–1807," *Explorations in Economic History*, 49, 119–130.

WEIL, D. (2012): *Economic Growth*, Boston, MA: Routledge, 3 edition ed.

WHITE, L. J. (1962): *Medieval Technology and Social Change*, Oxford University Press.

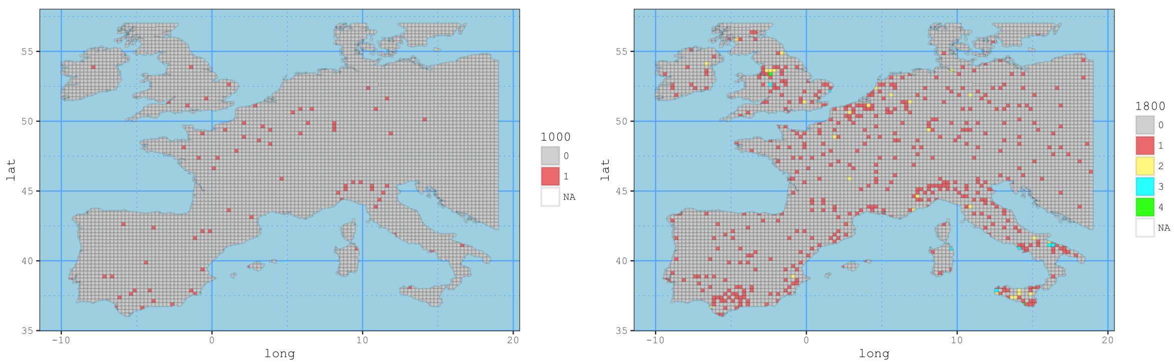


Figure 8: The extensive margin of the Bairoch data at the beginning and end of our study period for our 0.25 degree grid-cell world (number of Bairoch cities in our grid-cells).

8 Appendix

Malanima [2009]

Hohenberg [2004] Hohenberg and Lees [1995] Braudel [1993]