The London School of Economics and Political Science

Essays in Applied Microeconomics

Giuseppe Rossitti

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy.
London, May 2020

To my beloved Ioana

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 52,500 words.

Statement of inclusion of previous work

I confirm that Chapter 3 is a substantially revised version of the paper I submitted for the Master of Research (MRes) degree in Economics at the LSE, awarded in July 2015.

Abstract

This thesis is composed of three essays. The first two look at the effect of location on ability sorting in political and judicial institutions, while the third essay focuses on the relationship between religion and innovation.

In chapter 1 I exploit a novel and rich data-set with biographical information of US state legislators to investigate their sorting based on remoteness and attractiveness of the state capital. The main finding of the chapter is that in more remote US state capitals legislators are on average less educated and experienced. The results hold true for several measures of legislators' human capital and are robust to using different indicators of remoteness, based on the spatial distribution of the population, and controlling for other characteristics of the legislatures. The empirical strategy relies on the combination of a within-state analysis with an instrumental variables approach that relies on the proximity of capitals to the state centroids to identify the causal effect of capitals' remoteness across states. Finally, I also find that legislators' level of education and experience affects public good provision and corruption.

Chapter 2 focuses on the spatial variation in the UK criminal courts' performance across the country. I use the differences in real salary generated by a country-wide fixed wage for the category of criminal courts judges as a source of exogenous variation. The main finding of the chapter is that there is a negative relationship between performance, measured as the number of non-overturned trial sentences, and the real salary. In particular, London courts perform better on average than those in the rest of the country. I interpret the results as evidence of the willingness of highly paid professionals, such as judges, to accept a negative wage premium to sort in locations with a higher level of amenities. This evidence aligns with that provided by chapter 1 in showing that the location of institutions - in this case courts - affects their composition and in turn outcomes.

Chapter 3 studies the relationship between innovation, measured using patent filings per capita, and different indicators of religiousness, both cross-country and at US state-level. I use instrumental variables to attempt to pin down the effect of innovation on religiousness. In the cross-country analysis I use patent protection and property rights protection to instrument innovation, whilst in the US state-level analysis I build an instrument based on the clustering of academic institutions. I show the existence of a negative relationship robust to different specifications and using a different measure of innovation based on scientific articles per capita.

Acknowledgements

I want to thank a number of people without whom it would not have been possible to reach the end of this journey.

I am deeply indebted to my supervisor, Daniel Sturm, for the patient guidance, encouragement and advice he has provided at each stage of the process. I want to thank him for all the times I got sidetracked by a new idea and he helped me steer the course of my work.

This thesis would not have been possible without the help of Annie Petersen and Jacob Petterson from Vote Smart, who provided access to the legislators' data. I am deeply grateful to Vishal Sahu for his support in scripting the data collection process. I thank the Ministry of Justice for providing the data on appeals and, above all, Kristina Gray from the Criminal Court Statistics office for many exchanges.

I am grateful to Mark Wilbor, Deborah Adams and Gary Simpson for all the help provided during the duration of my PhD. I also gratefully acknowledge the financial support from the Bank of Italy and the LSE Economics Department.

I want to thank all friends and colleagues with whom I shared this adventure, in particular Alexia, Andrés, Clement, Giulia, Vincenzo. I am appreciative of the time spent with them. I am deeply thankful for all the conversations with Alessandra and Lorenzo, sharing the ups and downs of our respective journeys from afar has constantly been a therapeutic experience. I am very grateful to Chiara for her friendship and support over the last year, I wish we had met sooner. I am greatly indebted to Giulia for her advice, her comments, her friendship and always being a source of inspiration. I am fortunate to have shared most of my PhD with Marcus. One of the things I cherish the most about these years in London and at the LSE is my time spent with him. He has been a colleague, a flat-mate and, above all, the greatest friend.

I am also lucky to have a wonderful family who has consistently supported me. Since I was a child my grandfather, Giuseppe, instilled in me the importance of knowledge. I would have liked to share this journey with my grandmother, Lucia, who left us shortly before it started. I am thankful for having Marcela and Valentin as part of my family. They have always treated me like their son and have been among my greatest cheerleaders. I am grateful for the love from Emanuela and Marco, my siblings, who gift me with lightness and happiness. I am lucky for the care, support and unconditional love my parents,

Angela and Nicola, have given me since the moment I was born and continue to provide. I hope to make them proud.

Finally, none of this would have been possible without Ioana. There are no words to express how fortunate I am to have her by my side. In the last years, she has filled my time with joy, wit and laughter. She has shared all my best moments and relentlessly pulled me out of the worst ones. She is my best friend, my biggest fan, my mentor and the person I admire the most. I am blessed by her endless love, patience and support.

Contents

List of Figures

List of Tables

Chapter 1

US capitals' location and ability sorting of legislators

1.1 Introduction

Good politicians are fundamental to the well-functioning of the state in democracies and non-democracies alike. A large body of literature has sought to identify the determinants of politicians' selection, such as salary, media coverage, political systems and term limits.[1] Since Glaeser, Kolko, and Saiz (2001) a smaller and seemingly unrelated literature in urban economics and economic geography has focused on the role of cities' amenities in attracting highly skilled and educated workers.[2] An important yet unexplored question lies at the intersection between these two literatures: does the location of political institutions matter for their composition? In particular, does the location of the legislative bodies determine the average competence of legislators and, in turn, their performance?

In this paper I use novel and rich data on the biographical characteristics of the whole population of US state legislators to show that members of legislatures located in more remote, less attractive state capitals exhibit, on average, less and lower quality education and experience. I first show at state level that districts further away from the capital elect less qualified legislators. Then, drawing upon the relative centrality of US state capitals I am able to use instrumental variables based on state centroids to identify the causal effect

[1] See Caselli and Morelli (2004); Besley (2004); Snyder Jr and Strömberg (2010); Besley and Reynal-Querol (2011); Smart and Sturm (2013); Dal Bó, Finan, Folke, Persson, and Rickne (2017).

[2] See also Glaeser and Gottlieb (2006); Carlino and Saiz (2019).

of remoteness across states.[3] I interpret the results as evidence that legislators sort based on the attractiveness of the capital city: a remote capital has lower market access and imposes a larger cost of commuting, making a legislative seat less appealing to individuals with a better outside option. I also show that legislators' education and experience affect public good provision and corruption.

The empirical strategy relies on two pillars. The first is the combination of the within- and across-state parts of the analysis that in conjunction provide a preponderance of evidence on the effect of remoteness. The second is the fact that US state capitals are often located close to the centre of the state. While the location of a capital city is endogenously determined,[4] the centroid of a state is a mechanical construct that only depends on the shape of the state's political boundaries.[5] This creates the perfect setting to test my hypothesis as I am able to exploit the exogenous variation in the location of state centroids for identification, as in Campante and Do (2014).

I measure the quality of legislators with proxies for ability that are widely used in the political economy literature: the level of educational attainment and previous experience in local government (Dal Bó, Dal Bó, and Snyder, 2009; Ferraz and Finan, 2009; Besley and Reynal-Querol, 2011; Galasso and Nannicini, 2011). I also devise variables that capture the quality of legislators' education, based on the the academic institution attended and the subject of the degree pursed.

There are two ways in which a remote capital is unattractive to legislators. First, the state's districts are on average relatively further away from a remote capital.[6] Far away districts are less appealing than closer ones as they impose a higher cost of commuting. I find that the distance of a district from the state capital is a negative predictor of the level of education and experience of the legislator as well as the probability of having a Master's degree, a degree in public policy or politics and being a woman. Second, a remote capital is a less attractive city to work or live in. I use different population-based

[3]See Figure 1.1 for a map of all US state capitals.

[4]Campante, Do, and Guimaraes (2019) show that non-democracies tend to set capitals in more isolated locations so that the elite is able to extract larger rents without increasing the risk of a revolution.

[5]Namely, the centroid of a two-dimensional figure is the arithmetic mean position of all points in the shape.

[6]Districts are drawn to represent an equal amount of people within state boundaries (Kurtz, 2010). For a given number of districts, when more people live far away from the capital, the districts are more distant.

indicators of remoteness[7] such as density,[8] population potential and the isolation measure formulated by Campante and Do (2010). I test them all and find that remote capitals are associated with less educated and less experienced legislators who also have less relevant education from lower ranked academic institutions. More precisely, depending on the measure used, a standard deviation decrease in remoteness is associated with a 0.1 to 0.39 standard deviations increase in legislators' educational attainment and a 0.27 to 0.44 standard deviations increase in the quality of education. The effect is significant when estimated using instrumental variables based on state centroids.

Legislators' level of education affects political outcomes. For example, there is an association between a higher level of educational attainment and lower corruption at state level (Glaeser and Saks, 2006).[9] More educated politicians are also associated with faster economic growth (Besley, Montalvo, and Reynal-Querol, 2011) and higher efficiency in the use of government resources (Gagliarducci and Nannicini, 2013). I devise a simple test to check how remoteness and the composition of the legislature in terms of education affect public good provision and the level of corruption in the public sector. My results show that higher level of education and pre-legislative experience are associated with more public good provision and less corruption. The findings also suggest that not all of remoteness' effect on outcomes is carried through the sorting mechanism,[10] and that the type of education[11] received by legislators is relevant as well.

Related literature: This paper contributes to a number of literatures. First of all, it is closely related to Campante and Do (2014) who show that isolated capital cities are associated with higher corruption and lower public good provision. They produce evidence of isolation reducing accountability as state politics receives less news coverage and voters are less interested in it when they live further away from the capital. This paper, however, focuses on a different mechanism: legislators' sorting. More remote capitals

[7] I use density and population potential as measures of remoteness. Higher density and population potential indicate a *less remote* capital.

[8] Population density is a classic measure used in the urban economics and economics geography literature because of its correlation with urban amenities (Glaeser and Gottlieb, 2006; Rappaport, 2008)

[9] One potential explanation is that educated individuals tend to be more civic-minded (Milligan, Moretti, and Oreopoulos, 2004; Besley and Reynal-Querol, 2011).

[10] Plausibly legislators' sorting coexists with the incentive mechanism identified by Campante and Do (2014).

[11] I focus on law degrees and degrees in public policy or administration, political economy or government studies.

attract a different pool of legislators, less educated and less experienced. Accordingly, I employ different measures of remoteness such as density and population potential aimed at capturing capitals' attractiveness rather than the salience of legislative politics within the state.

A vast literature in political economy has discussed different aspects influencing politicians' characteristics and behaviour, such as media coverage (Ferraz and Finan, 2008; Snyder Jr and Strömberg, 2010; Repetto, 2018), salaries (Besley, 2004; Mattozzi and Merlo, 2008; Gagliarducci, Nannicini, and Naticchioni, 2010) and term limits (Ferraz and Finan, 2011; Smart and Sturm, 2013). This paper is closely related to those focusing on legislators' selection based on ability (Dal Bó, Dal Bó, and Snyder, 2009; Ferraz and Finan, 2009; Besley and Reynal-Querol, 2011; Galasso and Nannicini, 2011; Dal Bó, Finan, Folke, Persson, and Rickne, 2017) and, consistently with these works, uses measures of education and previous political experience as proxies for unobserved ability. It differentiates from the rest of the literature by studying the otherwise overlooked role of geographical location in political selection.

Since Carey, Niemi, and Powell (1998), a related strand of literature has focused specifically on US state legislators. A majority of works[12] relies on a data-set of legislators' demographic characteristics deriving from responses to two surveys run in 1995 and 2002 (Carey, Niemi, and Powell, 1998; Carey, Niemi, Powell, and Moncrief, 2006).[13] The composition of state legislatures is studied in relation to electoral competition (Squire, 2000), term limits (Wright, 2007; Carey, Niemi, and Powell, 2009), compensation (Hoffman and Lyons, 2018) and campaign contributions (Powell, 2012). This paper differs from the others by pointing out the importance of capitals' location and using a richer data-set comprising information on all state legislators.[14]

Urban economists studying the spatial distribution of activity, resources and people

[12]An exception is Bratton and Haynie (1999) who collect data from 6 states over a 3-year period. More recently, Kurtz (2015) lead a joint study from NCLS and Pew (NCSL, 2015b) to analyse the demographic composition of state legislators and compare it to population baselines.

[13]The 1995 survey collected information on 3,542 legislators (47% response rate on a fraction of the population), on 2.982 (40% response rate on the whole population). The 2002 survey, in particular, has been the only source of information on state legislators for years and it has been used in Kurtz, Moncrief, Niemi, and Powell (2006); Powell (2012); Powell and Kurtz (2014); Hoffman and Lyons (2018) among others.

[14]Moreover, the data I use is collected by a third party, hence less prone to selection bias due to non responses or miss-reporting.

across cities have overwhelmingly focused on the private sector (Becker, Heblich, and Sturm, 2018). This paper, instead, provides evidence of sorting in the public sector. It contributes to an extensive literature on agglomeration in cities (Combes and Gobillon, 2015), in particular supply-side agglomeration driven by the role of cities as consumption centres (Glaeser, Kolko, and Saiz, 2001). I also focus on a special set of cities, capitals, that has been understudied despite some distinctive characteristics: their *special status* is often embroidered in the country's constitution;[15] they are often the cultural and economic centre of the country, the most populous city[16] and host the government seat.[17] Studying capital cities is complicated by the endogeneity of their location or relocation choices,[18] with only specific exceptions (Becker, Heblich, and Sturm, 2018; Faggio, Schluter, and vom Berge, 2019). In this context, instead, I am able to use the proximity to state centroids to address the endogeneity problem.

Finally, in terms of its methodological contribution, this papers more generally lies with a transversal set of works that addresses the problem of endogenous regressors by using geographical location to build instrumental variables. These could take the form of distances (Acemoglu, Johnson, and Robinson, 2001; Card, 2001; Neumark, Zhang, and Ciccarella, 2008; Becker and Woessmann, 2009; Oster, 2012) or distribution-based measures (Campante and Do, 2014). I, instead, use concentration variables such as population density.

The rest of the paper is organised as follows: Section 1.2 provides some background on the US State Legislatures; Section 1.3 presents the data; Section 1.4 discusses the empirical strategy; Section 1.5 discusses the results; Section 1.6 concludes.

[15] Notable counterexamples are France and India that have no official capital. In the US, instead, moving the federal capital to another state would require amending the Constitution, as well as returning the land of the District of Columbia to the State of Maryland and carving a new federal district out of the receiving state (Bomboy, 2014).

[16] This does not hold true for US state capitals, where in only 17 of the 50 states the capital coincides with the largest city.

[17] There are many exceptions worldwide: for example, the Netherlands where Amsterdam is the capital, but the government sits in the Hague; or Bolivia, where La Paz hosts the government and Sucre is the official capital.

[18] In this spirit, Campante, Do, and Guimaraes (2019) show that in non-democracies there are incentives to move the government seat to an isolated place to reduce the threat of rebellion.

1.2 US State Legislatures

This paper focuses on the US State legislatures and the profiles of current state legislators.[19] Each of the 50 US States has its own legislative, executive and judicial powers that interact with each other in a similar fashion to what happens at national level. The legislatures are composed of two chambers: a Lower House called either House of Representatives, Assembly or House of Delegates and a Upper House called Senate. The exception is Nebraska where the legislature is made of a single (non-partisan) body whose members go by the title of senators.[20] On par with the US Congress structure, states' Lower Houses are more numerous, usually have less restrictive access rules in terms of age and residency, and a shorter term (CSG, 2018). Also, the legislative process closely resembles the national level's one. Bills make their way from inception through legislative committees first, then to both houses to finally be handed over to the state governor whose signature is needed for them to become law (NCSL, 2019). However, these are just highly general characterisations. Since all regulations are set at state level, there is a large variability in terms of institutional design across states: in the structure of the legislature, the size, compensation, staff availability and legislative requirements for legislators, as well as in the rules of the legislative process.[21]

Historically, most state legislatures were born as citizen legislatures. However, over time most have professionalised: salaries have more than doubled in real terms and aggregate full time staff has exploded from 500 units to around 30,000. In 1933 only 5 states had annual sessions, compared to 46 nowadays.[22] In the same time period sessions became longer[23] as well as the amount of constituents per legislators increased, since population

[19]The data captures the composition of state legislatures at the beginning of 2019.

[20]This paper focuses only on the 50 US states. It is worth noting, however, that also the District of Columbia and the territories of Guam and the US Virgin Islands have unicameral legislatures, while the others (American Samoa, the Commonwealth of the Northen Mariana Island and Puerto Rico) follow the canonical bicameral legislature (CSG, 2018).

[21]There are two main authoritative sources reporting frequently updated and detailed information on the plethora of different state regulations: the Council of State Governments who publish the Book of States (CSG, 2018) and the National Conference of State Legislatures (NCSL, 2019), who provide information and documentation to state level legislators and legislative staff. Excellent sources on the history, evolution and functioning of state legislatures are: Montès (2014); Squire (2012); Squire and Moncrief (2015); Moncrief and Squire (2017).

[22]Only Montana, Nevada, North Dakota and Texas still have bi-annual sessions. In 1933, 42 states used to meet bi-annually and Alabama every four years , in 1961 the number had increased to 19 and exploded up to 41 in the 70s following the great wave of legislative reforms in the 60s and early 70s (Kurtz, 2010; NCSL, 2018).

[23]Half of the states with session limits increased them, while the others spend more time in session.

growth in the US was matched by a decrease in the total number of legislators. At the same time, the legislative capacity of legislators is now higher following devolution policies pursued by the Nixon and Reagan administrations (Kurtz, 2010).[24] We can see that the least populated, rural Midwest states are among the ones with lowest pay and workload and the most populous and urban states such as California, New York and Pennsylvania have the most professionalised legislatures. These states, however, entail greater remuneration yet a higher cost of living and a larger workload: despite legislatures being larger and with more staff, individual legislators have more constituents to serve (Kurtz, 2010). Irrespective of the pay and amount of resources, recent studies have found that all legislators tend to spend more time on the job than they expected *ex-ante*, with the highest discrepancies for part-time legislators (Kurtz, Moncrief, Niemi, and Powell, 2006).

There are other dimensions that can influence individuals' ability and willingness to run for state legislative seats. For example, most states offer additional pay for legislative leaders and committee members.[25] The requirements to run for office vary dramatically across the US. The minimum legal age to run ranges from 18 to 25 for the Lower House and 18 to 30 for the Upper House. States also have different rules in terms of residency, US citizenship, voter registration and for how long before the elections these conditions need to be met. Most states require US citizenship, all states but Washington demand district residency, although without restrictive temporal constraints.[26] The state residency condition is sometimes more limiting, going as far as 7 years for the New Hampshire Senate. In some cases, state citizenship is also formally required.[27]

Finally, there is variation in the length of terms and the existence of term limits. Most frequently, a term lasts for two years in the Lower House (with four exceptions: Alabama, Louisiana, Mississippi and North Dakota that have a 4-year term) and 4 years in the

[24] The evolution has not been the same across the US. Figure 1.C1 categorises 5 groups based on the average time spent on the job by legislators, their salary and the size of the staff employed.

[25] All states that provide compensation for committee members also have additional pay for legislative leaders. The compensation is often limited to committee chairs and to specific committees. See Tables 3.11 and 3.12 in the Book of States (CSG, 2018) for state by state specific provisions.

[26] Often the requirement is as little as 30 days before elections.

[27] In Tennessee, for example, a candidate must have been a state citizen for three years to run for the House of Representatives. For more information on electoral requirements see Table 3.5 of the Book of States (CSG, 2018).

Upper House, with 14 exceptions.[28] Term limits, instead, are present in 15 states.[29] In six of these states the limits are lifetime: it is not possible to run for the same office once serving the maximum allowed amount of terms, while for the others the ban is only on consecutive spells (NCSL, 2015a).[30] Six states used to have term limits but repealed them. In two of these cases, it was the legislators' themselves who enacted the repeal.[31] Term limits in state legislatures have been the object of a number of studies that found no particular effect on the composition of the legislature, while they tend to affect the way legislators use their time, the balance between state legislative and executive power and the relevance of legislative leaders and committee chairs (Carey, Niemi, and Powell, 2009). It is important to note that legislators have the capacity to modify some of the regulations that directly affect them, as it happened in some states with term limits. This is also the case for the salary, which is sometimes set by the state constitutions but often requires a much less burdensome process to be amended.[32] All the features discussed in this section are potential determinants of legislators' sorting choices. As such, they need to be controlled for to be able to identify the effect of remoteness alone.

1.2.1 The location of US state capitals

In this analysis, the attractiveness of US state legislatures does not depend on the large variation in the many different legislative features enshrined in law but on one that is rather based on convention. In the US, state capitals often occupy a geographically central position in the state as evidenced by Figure 1.1. This is due to the long-standing tendency of US legislators to place government seats in the middle of the jurisdiction, based on the idea of equal representation of all citizens (Engstrom, Hammond, and Scott, 2013).[33] Often capitals do not correspond with the main centre of economic activity: in

[28]For 12 states the term is two years as it is in the Lower House. Two states, Illinois and New Jersey, are on a 10-year cycle made of two 4-year terms and a 2-year one CSG (2018).

[29]Figure 1.C2 provides an overview.

[30]With consecutive term limits a legislator could, in theory, cycle indefinitely between the two houses.

[31]In the other four cases they were repealed by the respective State Supreme Courts that objected the method of imposition, not the limits *per se* (NCSL, 2015a).

[32]See Table 3.8 in the Book of States CSG (2018) for individual states' methods of setting legislative compensation.

[33]For some of the older states the reason for moving the capital inwards was different (for example, for defence puroposes) than purely choosing a central location in the spirit of equal representation (Squire, 2012; Montès, 2014).

only 17 out of 50 states the capital is also the largest city, only one state capital features among the top 10 largest cities in the US (Phoenix, Arizona) and almost half of all capitals (24/50) have a population of less than 150,000.[34]

The reason why the central location of US capitals is important has to do with the severe endogeneity problem faced by anyone willing to determine the effects of the location of institutions, such as government or legislative bodies. Even if the location of a specific institution is arguably exogenous, the underlying spatial distribution of the population may depend on it. A potential solution would be exploiting the variation generated by changes over time in the location of institutions. However, such events are rare and, above all, the expression of institutional choices, which unless determined by an uncorrelated event, such as Germany's partition at the end of WWII in Becker, Heblich, and Sturm (2018), are endogenous as well. The centrality of US state capitals provides a reasonably exogenous instrument to address the endogeneity issue. The central point in a state is a mere geographical construct that cannot possibly influence the spatial distribution of population, above all since US state borders have been stable over the last 100 years.[35] Later, I use the exogenous variation generated by the location of state centroids as a source of identification of the effect of state capitals on the composition of the legislature.

1.3 Data

This paper uses a newly assembled, rich and unique data-set with biographical information on 7366[36] members of US state Lower and Upper Houses in 2019 from all 50 states. The data-set combines the biographical data with geographical data, including measures of capital city remoteness, and state legislatures' characteristics. The following sections describe the main variables used for the analysis. For a more detailed discussion on the data collection process and the variables' construction see Appendix 1.A.

[34] When considering only the contiguous US States, the capital is the largest city in 16 out of 48 states (excluding Honolulu, Hawaii) and 23 capitals out of 48 (excluding Juneau, Alaska) have a population of less than 150,000. City population data is taken from 2017 population estimates (U.S. Census Bureau, Population Division, 2018).

[35] See Squire (2012); Montès (2014) for the historical administrative evolution of the US.

[36] 7231 once excluding Alaska and Hawaii.

1.3.1 Biographical data

The biographical data was obtained from the Vote Smart (2019) online platform and includes indicators on personal characteristics, family members, religion, education, political career, political affiliation and work addresses.[37] Vote Smart is a free to access platform curated by volunteers who gather and compile information on US candidates and elected officials at all levels from Federal Government to city councils, with the aim of informing voters. Vote Smart staff members are mostly volunteers who collect biographical data, public statements, votes as well as legislative information and input it on the website. The biographical information reported on Vote Smart is collected from legislators' official profiles and copied without being transformed or standardised, thus it can be considered self-reported. The data is not organised in a unified data-set but available in text entries under politicians' profiles, that are free to access through the Vote Smart website. Figure 1.2 provides an example for a newly elected Utah Representative. However, upon purchasing access, it is possible to use an API that queries individual observations and provides them in XML or JSON format.[38]

Personal information: I obtain information on age, gender, relationship status, whether the legislator has any children, the number of children and the gender of the partner from the legislators' personal and family information. Age is computed from the birth-date until Jan, 1st 2019. The remaining variables are extracted using algorithms that analyse the text strings containing information on the family members.

Education: I build two main indicators of educational attainment: the level of education and whether a legislator has obtained a Master's degree or above; and four indicators measuring the quality of education: having a policy or politics-related degree, a continuous score variable based on the ranking of the academic institutions attended, and a dummy indicating whether the legislator has a Master's degree from a top 20 or a top 50 university. Starting from the individual biographical entries, I use an algorithm that first separates degree, field(s) of study and school from the text and then uses the information to attribute the best match in terms of level of education, using the official United States

[37] Matter and Stutzer (2015) use biographical data from Vote Smart on 1413 lawyer legislators in Congress and 383 in 17 state legislatures to study their voting record on tort reforms.

[38] These are later converted into delimited text files that constitute the raw data.

education levels mapping into the UNESCO ISCED 2011 classification,[39] which is based on a 0-8 increasing scale. The second attainment measure is a dummy variable indicating if the highest degree received is a postgraduate one. The median level of education is a Bachelor's degree (ISCED level 6) and the average is 6.3 with more than 25% of the population having completed at least a 2-year Master's degree. However, such measures treat all degrees with the same qualification as equal regardless of the effective difficulty in obtaining them, the type of degree and the quality of education received. To be able to better differentiate legislators I introduce measures of *quality* of education. The idea is that admission and completion of a degree in a prestigious, highly ranked academic institution is a good proxy for talent. First, to control for the type of degree, I build a dummy variable indicating if the person has graduated from law school and another one recording if they received any policy or politics-related 4-year undergraduate or higher degree.[40] As a proxy for education quality I use the score from one of the three most influential international ranking systems of academic institutions: the Times Higher Education (2019a) World University Rankings (THE).[41] I am able to attach a score to all reported institutions,[42] by assigning the lowest score to those not included in the ranking.[43] Also, for each legislator, I record if they have received a postgraduate degree from any academic institution included among the top-20 or top-50 in the ranking.

Political experience: I introduce two measures of political experience, one is *resume length*, based on the number of different political roles a legislator held in their career;

[39]International Standard Classification of Education (UNESCO Institute for Statistics, 2011). The individual country specific mapping from national qualifications to standardised equivalent international levels can be found at `http://uis.unesco.org/en/isced-mappings`.

[40]They include, among others, Bachelor's and Master's in Public Policy, Public Administration, Political Economy, Government. While policy and politics graduates would appear as the most relevant category, lawyers have been a predominant force in US legislatures since inception as Alexis de Tocqueville noted in 1835 (Tocqueville, 2003). To this day they make up much of the legislative bodies at both state and federal level (Miller, 1995; Matter and Stutzer, 2015). There are 1111 legislators with a postgraduate law degree, approximately 18% of the sample with educational background information. Among those, 445 majored in public policy, politics or government before moving on to law school.

[41]The other two are the Academic Ranking of World Universities by ShanghaiRanking Consultancy (2018) and the QS (2019) World University Rankings. I prefer the Times Higher Education ranking because it provides a cardinal score out of a 100 rather than just a ranking position. I choose to adopt an international ranking as some of the legislators in the data set attended top institutions abroad. The choice of one ranking over another is not of particular importance; the main goal is to distinguish the few most prestigious academic institutions from the majority of colleges and graduate schools.

[42]Times Higher Education ranks the best 1258 universities in the world. I collect the score for all 172 US-based institutions and the international schools in the top 50.

[43]I consider the score of the highest ranked institution, regardless of the level of the degree. For robustness, I use another version of this variable based the score of the highest degree obtained.

the second is length of tenure, measuring the number of years since the start of the political career. I also register in two dummy variables: whether the legislator has held legislative leadership positions in the past[44] and whether they had some role in their party.[45] I am able to categorise every single role held by each legislator so that the resume length and tenure variables can be adjusted to include all political and government roles, including work in public administration, membership of public boards, electoral campaigns and leadership positions in the party and in the legislature or only focus on legislative and executive positions, such as town mayor, county councilor, state representative and state governor. Length of tenure is computed as the numbers of years since the first appointment.

Congressional information: I also collect information on the main party affiliation, whether a legislator is affiliated with more parties and whether they are running for office in 2019.[46]

1.3.2 Population distribution

In the next section I show how I compute different measures of capital city remoteness. To be able to do so, I need to closely approximate the population distribution. For this purpose, I use the American Community Survey 2012-2016 population estimates at block group and census tract level. The block group is the second smallest census unit, falling in between census tract and census blocks. Data at block group level generate a better approximation of the population distribution. However, for computational reasons, I use census tract level data for deriving the isolation variables.

1.3.3 District controls

For each Lower or Upper House district I compute: population density using average population estimates and land area in m^2; the district average per capita income in the last 12 months (averaged over 2012 to 2016) in 2016 inflation-adjusted dollars; educational

[44] For example, president, speaker, majority or minority leader and whip.

[45] One such instance is being a party delegate to the national convention.

[46] There are only 26 such instances in my data set. Among those, 18 legislators are running for the exact same office, while 4 are State House members running for the State Senate, 3 are running for a governor position and 1 for district attorney.

attainment at district level, using the ISCED classification for comparability with legislators' level of educational attainment; the share of 25+ years old population with at least a bachelor's degree.[47] The above variables are useful to control for baseline candidates' characteristics, as legislators must be district residents in almost all states. Moreover, they are also proxies for voter turnout and political knowledge at district level. Education is correlated with political involvement and, in the US, strongly correlated with higher turnout (Milligan, Moretti, and Oreopoulos, 2004). Both education and income levels are also positively correlated with newspaper readership (Snyder Jr and Strömberg, 2010).

1.3.4 State controls

I introduce two state-level sets of control: measures of legislatures' characteristics, such as salary and compensation structure, level of professionalism, terms limits, residence requirements to run for office; and state characteristics such as the size of the state, population, average education level, measures of voters' engagement. I borrow the data on legislators' salaries and length of the legislative session from Bowen and Greene (2014a), who compute *ex-post* 2010-2014 averages. The Book of States publishes every year the information required to compute legislative salaries and the length of the session *ex-ante*.[48] However, the effective number of days in session is impossible to forecast in advance as well as salaries since, in some cases, they depend on the former. Using a lagged average value for salaries has two benefits. First, while legislators in many states cannot predict the exact amount of compensation they will receive for their service, they are aware of what the prevailing salaries are and use those to inform their decision to run for office. Second, since in some states legislators have the power to set their own compensation, the current or future level of remuneration is endogenous. The real salary is computed by dividing the average nominal 2010-2014 compensation by the average gross median rent for the

[47] Data from the American Community Survey (ACS) 2012-2016 average, in U.S. Census Bureau, Geography Program (2016).

[48] Table 3.9, Legislative Compensation and Living Expense Allowances During Sessions, combined with the information from Tables 3.19 and 3.20 on regular and special sessions, since in some states legislators' pay is per session or calendar day (CSG, 2018).

same time period.[49] To measure the level of professionalism of a legislature, I employ the index developed by Squire (2017) that assigns a score from 0 to 1 to state legislatures comparing them across different features (such as pay, length of sessions, size of staff, available resources) to the US Congress.[50] Information on term limits is derived from the NCSL (2019). I record if a state has term limits in place, when they were introduced and if the limit is only on consecutive terms or lifetime. I build additional variables that can influence individuals candidates' choice to run for office. First, I build two dichotomous indicators that report for each house and state if there is a provision for additional pay for legislative leaders and committee members. Then, I construct six different variables capturing the minimum requirements to be able to run for office at state and house level (minimum age, length of citizenship in the US, length of state residency, length of district residency, state citizenship, voter registration).[51]

I introduce a set of variables to control for the characteristics of the pool of candidates and voters in the state. To measure the *ex-ante* level of education of potential legislators, I measure educational attainment at state level, using the ISCED classification for comparability with the legislature's average level of educational attainment. Similarly, I also compute for each state the share of the population over 25 that has obtained at least a 4-year college or university degree (Bachelor's degree or higher).[52] I use two types of variables to proxy voters' engagement. First, I borrow measures of newspaper political coverage from Campante and Do (2014).[53] Then, I compute average voter turnout as a fraction of all voters in the last 3 presidential elections 2008, 2012 and 2016.[54] To capture the size and shape of the state, I compute the state area and the maximum distance between the capital and the state boundaries.

[49] Gross median rents data from ACS 2010-2014 Census Bureau estimates. The salary data from Bowen and Greene (2014a) is expressed in 2010 US dollars, the rent is converted to 2010 US dollars using the ACS 2010-2014 Census Bureau suggested inflation value.

[50] The index is an update of the original version from Squire (1992). It assigns a score of 1 to Congress, which means the index is interpreted as follows: a legislature with a score 0.5 is half as *professional* as the US Congress. Bowen and Greene (2014b) while favouring a multi-dimensional approach over a unidimensional measure, find that the index performs well at capturing the difference between amateur and professional legislatures.

[51] All these variables are derived from the Book of States (CSG, 2018).

[52] Education data from US Census Bureau 2012-2016 American Community Survey estimates of the resident population over 25 per level of education.

[53] The variable is based on principal component analysis of search terms related to state politics from the webiste NewsLibrary.com, see Campante and Do (2014).

[54] Turnout data are taken form CSG (2018).

Finally, to test for the effect of legislators' characteristics on outcomes, I use three different measures of performance. First, from the Book of States (CSG, 2018) I calculate the average number per bills per legislator per day of session.[55] Then from Campante and Do (2014) I borrow a measure of public good expenditure, namely, the share of state expenditure assigned to education, public welfare, health, and hospitals from US Statistical Abstract 2012. Also, I borrow a measure of corruption from Glaeser and Saks (2006): the number of federal convictions for public corruption per capita in the state.[56]

1.4 Empirical strategy

The empirical analysis presented in this paper consists of three parts. In each of the first two, I address one of two mechanisms for the effect of capitals' remoteness on the ability sorting of state legislators: first, the fact that a district far away from the capital is unappealing and then, the unattractiveness of the capital *per se*. In the third, I provide a link between legislators' sorting and some outcome measures.

1.4.1 Remote districts

The first of the two closely related mechanisms studied in this paper relates to the fact that remote capitals are relatively further away from a larger number of districts that are unattractive to potential legislators. Following the principles set by successive Supreme Court rulings on re-districting Kurtz (2010), state districts tend to be drawn to represent roughly the same amount of people. Accordingly, states in which large masses of the population live further away from the capital city have more *remote* districts. I study how the distance between the capital and an electoral district affects the type of candidates being elected in the same district. As briefly explained in Section 1.2, legislators running for office in a district need to reside there. Therefore, the pool of candidates eligible for a seat - or willing to relocate just to run for one - is limited in size. A remote district

[55]Bills data are 2014-2017 averages, to smooth out potential increased activity in a specific year due to idiosyncratic factors. The number of legislators is also taken from CSG (2018). Session length data are from Bowen and Greene (2014a) and refer to 2010-2014.

[56]Average from the period 1976-2002, data are derived from the 1989, 1999 and 2002 issues of the Report to Congress on the activities and operations of the Public Integrity Section, issued by the Department of Justice, see Glaeser and Saks (2006).

may be more unappealing for two main reasons: first, there is a hefty psychological cost of commuting;[57] second, a remote district could mean lower visibility for the legislator. We should then expect that, after controlling for local characteristics, in districts further away from the capital the individuals from the eligible pool with a better outside option are less likely to run for office. This is due to their net benefit being lower compared to equally talented candidates who live closer. To test this hypothesis, I employ the following model:

$$Y_{ids} = \beta_0 + \beta_1 \delta_{ds} + \boldsymbol{\beta_2} \mathbf{W_d} + \boldsymbol{\beta_3} \mathbf{X_i} + \gamma_s + \epsilon_{ids} \tag{1.1}$$

Y_{ids} is a biographical characteristic of legislator i elected in district d of state s. It can be the level of education measured on the ISCED scale or a continuous measure of political experience. δ_{ds} is the distance between state s' capitol building and the centroid of the district d of election.[58] w_d is a vector of district characteristics such as the average level of education of the residents, income per capita, population density $\mathbf{X_i}$ is a vector of individual covariates and γ_s a state fixed effect. For dichotomous legislators' characteristics, such as gender, having obtained a postgraduate degrees or having graduated from law school, I estimate the following model:

$$P(Y_{ids} = 1|X) = \Phi(\beta_0 + \beta_1 \delta_{ds} + \boldsymbol{\beta_2} \mathbf{W_d} + \boldsymbol{\beta_3} \mathbf{X_i} + \gamma_s + \epsilon_{ids}) \tag{1.2}$$

where all the variables on the right-hand side are defined as above. Finally, I use the above model to investigate the effect of remoteness of the gender composition of the legislature, also investigating whether having children alters the results.

[57]The monetary cost is already covered by the legislature. Almost all states have a mileage cost and a session *per diem* allowance, see Table 3.9 in the Book of States (CSG, 2018) for a state by state breakdown.

[58]In all the geographical measures I construct I always define the location of the capital city based on the coordinates of the capital hill as in Galster et al. (2001). From Vote Smart I also collect information on legislators' work addresses. I am able to compute the exact distance between the capitol hill's office address and a district office address for 4,521 legislator. However, I use the distance between the state capitol and the district's centroid to avoid reducing the sample size. I am indeed able to match all 7,350 legislators with their district of election. The measurement error is negligible. Figure 1.C3 shows that the correlation between the two measures is near perfect.

1.4.2 Capital city remoteness and composition of the legislature

The second part of the analysis focuses on how capitals' remoteness affects the composition of the legislature. The mechanism is based on the idea that more remote cities are less attractive places where to relocate, even if only temporarily, during the legislative session. Individuals with a better outside option should find remote capitals less attractive, even after controlling for state characteristics and legislative features, such as salary. Hence, we should expect the average level and quality of legislators' education to be negatively correlated at state level with capitals' remoteness. There are two main ingredients in this part of the empirical analysis. The first is the definition of capital city remoteness and the methods employed to measure it. The second is the identification strategy that, following Campante and Do (2014), relies on the relative centrality of US state capitals to build instrumental variables based on the arguably exogenous location of state centroids, addressing the potential endogeneity of the spatial distribution of the population in and around the state capital. In all specifications I restrict the sample to the 48 contiguous US states.[59] To study the effect of remoteness on the type of legislators, I use the following model:

$$AvgEdu_s = \beta_0 + \beta_1 Remoteness_s + \boldsymbol{\beta_2}\mathbf{Legi_s} + \boldsymbol{\beta_3}\mathbf{State_s} + \epsilon_s \qquad (1.3)$$

where $AvgEdu_s$ is the average type (measure of education attainment, quality of education or political experience) of legislators in state s. $\mathbf{Legi_s}$ is a vector of state legislatures' characteristics such as the real salary and the level of professionality meant to control for the variation in the attractiveness and difficulty of legislative roles across states. When using measures of political experience as outcome variables, I also control for the existence of term limits in the state legislature and whether they are binding or not. $\mathbf{State_s}$ is a vector of state characteristics. It includes the average educational attainment in the state that is an indicator of the level of education in the pool of potential legislators. It also includes measures of political engagement such as the voter turnout in presidential elec-

[59] The results are robust to including Alaska and Hawaii. However, it is more reasonable to exclude the two states because of their peculiar characteristics. Alaska is the largest state but the third least populous, with a considerable proportion of the state entirely uninhabited. Hawaii's capital, Honolulu, due to the mechanics of the measures of remoteness that are explained later, is one of the least remote capitals: 25% of the state's population lives there and roughly two thirds of the state population is concentrated in the small island of Oahu that hosts the state capital.

tions and newspapers' political coverage as well as the geographical size and shape of the state. The main coefficient of interest is β_1, informing on the effect of higher remoteness on the composition of state legislatures. I employ several approaches to measure capitals' remoteness. First, I build three measures of market access using population density in a 30, 50 and 100 km radius around the capital, a commonly used variable in the urban economics literature as it captures well the cities' level of amenities (Glaeser and Gottlieb, 2006; Rappaport, 2008):

$$PopDensity_s = \frac{\sum_j^N pop_{js} \times w_{js}}{area_s} \tag{1.4}$$

Equation 1.4 shows the formula used to compute density around state s' capital. pop_{js} is the population of block group j in the circle around state s capital; $w_{js} \in (0, 1]$ is a weight that attributes a share of block group j's population, if not fully included in the circle, equal to the ratio of the area of the intersection between the block group and the circle and the total area of the block group, $w_{js} = 1$ if the block group is fully included in the circle. $area_s$ is the total amount of land area in the circle around the capital.

The second variable I introduce is a measure of population potential, widely used in the economic geography literature. I compute population potential at state level using a gravity equation of the type:

$$PopPotential_s = \sum_i \frac{pop_i}{d_{si}} \tag{1.5}$$

Where pop_i is the population of census tract i in the same state s as the capital city and d_{si} is the distance between state s capital and the centroid of the census tract.

Finally, I borrow the measures of *isolation* used by Campante and Do (2014), measured as:

$$Isolation_s = 1 - \sum_i \Big(1 - \frac{log(d_{si})}{log(d_{max})}\Big) s_i \tag{1.6}$$

where d_{si} is the distance between the capital city of state s and the centroid of census tract i, d_{max} is equal to the maximum distance between any state capital and a point in the same state[60] and s_i is the share of state population living in census tract i. A

[60]This is equal to roughly 900 km between Sacramento, California's capital, and the southern border

different, *adjusted*, version of the first measure takes into account the size of the state by setting d_{max} for each state as the maximum distance between the capital and a point in that state. A correct formulation of the second variable would be:

$$AdjIsolation_c = 1 - \sum_i \left(1 - \frac{log(d_{si})}{log(d_{max_s})}\right) s_i \tag{1.7}$$

The different measures of isolation have each advantages and drawbacks. Isolation measures use all the spatial information in the state. On the other side, the first version of the variable is affected by state size,[61] while the second measure assumes that the important of distance varies with state size.[62] Both variables do not take population density in the capital into account, but just its *relative* density.[63] The density variables are better suited to describe the level of economic activity in the area. On the other side, they neglect the information on the spatial distribution of the population away from the point of it. Population potential is somewhere in between, by using the whole spatial distribution in the state, but weighing each point by the population mass. I will test all the different measures of remoteness, but I consider population density as the best variable. This is because the attractiveness of a capital for legislators, and the consequent decision to work there, does not necessarily depend on how the whole state population is concentrated relative to that city but just on what is happening in the immediate proximity of the capital. As a result, if the objective is to capture capitals' attractiveness as closely as possible, it may be better to discard the information on the rest of the spatial distribution of the population. Population potential and isolation, however, may be better suited to capture other mechanisms such as the legislators' accountability, since they could more closely approximate the salience of state politics within the state.

I also shortly investigate three alternative *blunt* measures of capital cities' remoteness.[64] The first is the percentage of state population living in the capital city. The

with Arizona.

[61]However, it is enough to control for it as I will do in all models.

[62]In other words, a population mass 150 km away from the capital is *relatively further away* in New Hampshire than in Texas because the latter is a much larger state.

[63]Two states with different concentrations of population around the capital but the same relative spatial distribution within the state are assigned the same level of isolation. See Appendix 1.B for a simple graphical example.

[64]From the US Census Bureau 2017 population estimates.U.S. Census Bureau, Population Division (2018) for city level estimates and U.S. Census Bureau, Population Division (2017) for state level values.

second is the ratio between the population of the capital and that of the largest city in the state. The above two variables are highly correlated (0.92) and represent measures of capital primacy, pioneered by Jefferson (1989) and widely used in the economic geography literature, for example in Henderson (2003). The last one is a dichotomous variable indicating if the capital is the largest city in the state. These three variables are appealing because of their intuitiveness but, compared to the ones above, they discard almost all of the information on the population distribution.

1.4.3 Identification

To identify the causal effect of capital city remoteness on different measures of education and experience, I employ instrumental variables based on the spatial distribution of the population around the centroid of the state. Indeed, the model from Equation 1.3 suffers from endogeneity: the spatial distribution of the population around the capital city may be affected by the type of the legislators. For each one of the three measures of remoteness discussed in the previous section, I build an instrument constructed in the same way but using the coordinates of the state centroid *in lieu* of those of the capital.[65] The model I estimate is:

$$Remoteness_s = \gamma_0 + \gamma_1 CenRemoteness_s + \boldsymbol{\gamma_2}\mathbf{Legi_s} + \boldsymbol{\gamma_3}\mathbf{State_s} + \xi_s \tag{1.8}$$

$$AvgEdu_s = \alpha_0 + \alpha_1 \widehat{Remoteness}_s + \boldsymbol{\alpha_2}\mathbf{Legi_s} + \boldsymbol{\alpha_3}\mathbf{State_s} + \epsilon_s \tag{1.9}$$

Where $\mathbf{Legi_s}$ and $\mathbf{State_s}$ are defined as in Section 1.4.2 and $CenRemote_s$, the remoteness of the centroid, is the excluded instrument. The relevance of the instrumental variables relies on the relative proximity of capital cities to the centroid of the state. Figure 1.3 shows that all of the 48 contiguous US states but Wyoming have capitals less than 300 km away from the centroid of the state, with a median value lower than 100 km. There is an inherent trade-off between the precision and the relevance of the instrument. The narrower measures, such as density in a 30 or 50 km radius are better at capturing the

[65] For example, population density in a 100 km radius around the state capital is instrumented by population density in a 100 km radius around the state centroid. Similarly, isolation from Equation 1.6 is computed with the same formula where d_{si} is the distance between the centroid of state s and the centroid of census tract i in state s; d_{max} is the maximum distance across the continental US between any state centroid and the centroid of a census tract in the same state.

attractiveness of capitals, whilst the broader measures ensure a larger spatial overlap between the endogenous regressor and the instrument, hence the instruments are more powerful.[66]

In terms of the instruments' exogeneity, given that the centroid of the state is a geographical construct, its location is convincingly exogenous. One may suspect that state boundaries, which mechanically determine the centroid of the state, could change due to factors correlated with the outcome variables. However, contiguous US states' boundaries have not changed in the last century (Squire, 2012).

We could still worry, however, that while the location of the centroid is exogenous, the population distribution around it may still depend on the endogenous distribution of the population around the capital. I argue this is not a significant issue in this case for two reasons. First, I mainly focus on variables that discard most of the information on the population distribution away from the capital. Second, the outcome variables, legislators' education and experience, are unlikely to be correlated with factors that affect the population distribution around the capital.

However, there are other factors that can lead to the violation of the exclusion restriction. The central location of the capital could be correlated with other variables that affect legislators' characteristics. For example, one reason why some capitals are in central and more remote locations could depend on the attitude of legislators towards rent extraction.[67] Accordingly, Campante and Do (2014) show that more isolated US capitals display higher level of politicians' corruption. We may then worry that the correct mechanism is that more corrupt legislatures attract less educated and experienced legislators. There is a number of factors mitigating this concern. First, as already noted state capitals' location has not changed in more than a century. For it to be endogenous, we would need to believe the same applies to the legislators' characteristics that affected the definitive location. Second, in the US there is a general, country-wide tendency to place government seats at the centre of the administrative unit of interest (Engstrom

[66]Figure 1.C4 plots the first stage F stats for each measure of remoteness using state level average educational attainment and state size measures as included instruments. As expected, the centroid density in 30 km measure is the weakest instrument, while centroid density in 100 km is the strongest. Isolation measures do not fare as well as, despite taking the whole state into account, they heavily discount information further away from the point of interest.

[67]A remote capital city would impose less constraints on legislators' ability to extract rents.

et al., 2013). Finally, we can look at the within-state analysis, discussed later, where we learn that further away districts are represented on average by legislators with a lower level of education and experience, suggesting the existence of an effect of remoteness on the characteristics of elected legislators. Sewing together the two parts of the analysis pointing to the same conclusion provides stronger evidence that the correct mechanism affecting the composition of state legislature is indeed the remoteness of the capital.

1.4.4 Sorting and incentives

In the last part of the empirical analysis I employ a simple test to check if the differences in legislators' education levels across states affect legislative outcomes. The model I estimate is:

$$Y_s = \beta_0 + \beta_1 Remoteness_s + \beta_2 AvgEdu_s + \boldsymbol{\beta_3}\mathbf{Legi_s} + \boldsymbol{\beta_4}\mathbf{State_s} + \epsilon_s \tag{1.10}$$

where Y_s is an outcome variable such as the intensity of legislative activity, the provision of public good or corruption in state s and all the right-hand side regressors include the same variables presented in the earlier sections. The vector of state level controls $\mathbf{State_s}$ is enriched with the average state income per capita[68] and dummies for macro-regions of the US. I place both remoteness and the education of legislators on the right hand side to check if there is any residual effect of remoteness on outcome variables. If that is the case, then legislators' sorting is the mechanism through which remoteness affects outcomes. Otherwise, if we detect a residual effect of remoteness, that could be interpreted as evidence for another mechanism in place, for example accountability.

1.5 Results

I start this section by presenting some stylised facts about state legislatures that can inform about the determinants of their composition, which are further investigated later. The great amount of variation in terms of salary, staff, benefits, time on the job, term limits would suggest that there is little room for an effect of ability sorting based on remoteness. In Figure 1.4 we can see that full-time legislatures in large states like Califor-

[68] Also from Census Bureau ACS 2012-2016 averages.

nia and New York have higher salaries and, on average, more highly educated legislators. However, when we look at the rest of the distribution, there is no evidence of a strong correlation between professionalism and salaries on one hand and educational attainment on the other. Some full-time legislatures, Wisconsin for example, fare way worse than less professional ones in terms of education.[69] Figure 1.5 shows the correlation between different measures of remoteness and the residuals from regressing the average legislators' educational attainment on the average state educational attainment measured on the same scale, legislators' real salary and the level of legislative professionalism.[70] We can see that there is a clear positive relationship (negative for isolation) between all measures of remoteness and the unexplained variation in education levels. The strongest correlation is that with the narrower variable, density in 50 km (Figure 1.5a), while the weakest is that when remoteness is measured using isolation (Figure 1.5d). This is consistent with the conjecture presented in Section 1.4.2: density measures are better at capturing the attractiveness of a city.

1.5.1 Districts' distance from the capital

I proceed to discuss the evidence for the relationship between a district's distance from the capital and the characteristics of the legislator(s) elected. Table 1.1 shows that different measures of educational attainment and quality of education are negatively correlated with election in a remote districts. For example, columns 1-3 of Panel A show that a standard deviation (σ) increase in the logarithm of the distance to the capital decreases educational attainment by 0.036 to 0.076 σ. The magnitude of the effect is quite small, yet significant in all models. Similarly, columns 4-6 of Panel A show that being further away from the capital has a significant negative impact on the probability of having obtained a Master's degree.[71] The same negative effect exists when the dependent variable

[69] Figure 1.C5 shows a similar pattern using the percentage of legislators with a postgraduate degree as measure of education, instead of educational attainment on the ISCED scale.

[70] The regression used to compute the residuals in Figure 1.5d also includes the logarithm of state population.

[71] The magnitude of the effect is harder to interpret for probit models given the log transformation of the explanatory variable of interest. We can say that a 1.7-fold increase in the distance from the capital increases the probability by 2.2 percentage point in the richest specification.

is the score of the highest ranked university attended[72] or the probability of having a postgraduate law degree. Panel A of Table 1.2 shows that the same relationship exists for two different measures of legislators' previous experience: number of relevant non-state legislative positions held and length of tenure. The point estimates are consistent across specifications, however not statistically significant in columns 1-3. Unsurprisingly, the district's average educational attainment is a strong positive and statistically significant predictor of legislators' quality in almost all specifications. On one hand educational attainment is a measure of the quality of legislators' pool. On the other more highly educated voters have a lower cost of acquiring information and may select better candidates. Including controls for district level of education, income per capita, population density and type of office (legislator or senator) dampens the effect of distance on the measures of education.[73] The states fixed effects capture the impact of different term limits and levels of remuneration across states. The findings are consistent with the original hypothesis. A district far away from the capital is unappealing for two main reasons. First, commuting is costly. Even if legislators are compensated for mileage and sitting days away from home, there are significant non-monetary costs of commuting (Small et al., 2005; Van Ommeren and Fosgerau, 2009). Second, electors in far away districts tend to be less involved and knowledgeable about politics (Campante and Do, 2014). Given that roughly half of Congress is made of former state legislators (Kurtz, 2010), indicating that a political career is a clear motive to run for the state house, we could also conjecture that legislators from more remote districts could have lower visibility, weakening their political career prospects.

Panel B of Table 1.2 focuses on gender. In particular, I use dichotomous a indicators indicating for a legislator being a woman and a woman with children. The results in columns 1-3 show quite clearly that a greater distance between the district and the capital is associated with a lower probability of a female legislator. An approximately two and a

[72]More precisely, the score of the highest ranked university that awarded a Bachelor's or a higher level degree to the legislator.

[73]I do not include party affiliation in the regression because of multicollinearity. Table 1.D1 in the Appendix shows the results from the same models controlling for it. The coefficients are not statistically significant but the negative relationship between distance from the capital and legislator's education or experience is always preserved. In the same table I also report the estimates obtained by using the smaller sample of legislators for whom birthplace data is available and controlling for being born in the same district of election. Once again, the direction of the effect is consistently negative.

half-fold increase in the distance from the capital reduces the probability of the legislator being a woman by 2.8 percentage points in column 2 (2.3 percentage points increase for the probability of being a woman who has children in column 5). In columns 3 and 6 I also include age and relationship status on the right hand side of the regression.[74] This, however, more than halves the sample size due to lack of date of birth reporting. Nonetheless, the point estimates are robust to introducing the additional controls. The marginal effects in the models that include age for the same approximately two and a half-fold increase in distance from the capital is a 2 percentage points reduction in both the probability of being a woman and the probability of being a woman with children. The results are consistent with Clark, Huang, and Withers (2003)' suggestion that women tend to commute less than men. One explanation is that the psychological cost of commuting is larger, even if the occupation is the same, because women take the burden of household responsibilities such as childcare and housework, as shown by Roberts, Hodgson, and Dolan (2011). Age is a positive predictor of a female legislator (see Panel B), indicating that women tend to be on average older the their male peers. My interpretation of this finding is that some of the burden attributed to women by gender norms decreases as they grow older (for example as the children become adults).

1.5.2 Capital cities' remoteness

This section presents the results of the main part of the empirical analysis, focusing on how the level of remoteness of the capital, a proxy for its attractiveness, affects the average composition of the legislature. Table 1.3 presents a series of *naive* regressions of all measures of remoteness on educational attainment. All regressions control for the size of the state.[75] The coefficients are significant and the sign is as expected for all the measures of isolation (negative) and market potential (positive). The variables that do use the least amount of information on population distribution, columns 7-9, still display the correct sign for the coefficients although the estimates are not significant at a 10% level. In Tables 1.4 and 1.5 I introduce a richer set of control variables.

[74]It is worth noting that being married also displays a statistically significant negative relationship with the probability of being a female legislator.

[75]The regressions in columns 1-2 and 7-9, using measures that are independent from the population size, also control for the log of the state's population.

Table 1.4 uses the density and population potential variables as measures of remoteness and three different educational outcomes variables: average educational attainment of legislators, the percentage with a postgraduate degree and a measure of educational quality based on the score assigned to academic institutions by the Times Higher Education (2019a) ranking. We can see that the coefficients have the expected (positive) sign and indicate the existence of a strong relationship between a capital city's remoteness and the measures of education. The coefficients are significant in almost all specifications and the point estimates are relatively stable upon the addition of control variables.[76] A σ increase in population density in a 100 km radius around the capital city is associated, in the richest model, with: a 0.39σ increase in educational attainment; a 3.45 percentage point increase in the proportion of legislators with a postgraduate degree; and a 0.27σ increase in the average quality of education. The magnitude of the effects is similar when using the *narrower* measures of density (respectively in a 30 and 50 km radius): for example, a σ change in 50 km density brings a 0.39σ change in educational attainment in the same direction; a σ increase in 30 km density is associated with a 0.29σ increase in educational attainment. The coefficient estimates are reported in Panel C. Panel B instead documents the results for the same regressions using population potential as a measure of remoteness. Focusing again on the richest specification, a σ positive variation in population potential is related to: a 0.26σ increase in educational attainment; a percentage point increase in the proportion of legislators with a postgraduate degree; and a 0.44σ increase in the average quality of education. The estimates are robust to using different versions of the same outcome variables.[77] When using density as a measure of remoteness, they are also robust to controlling for other state legislature's regulations, such as the presence of term limits and the restrictions in terms of residence and citizenship on the pool of eligible candidates.[78] Table 1.6 replicates the models from Table 1.4 Panel A using measures of relevance of education and political experience as dependent variables.

[76] I perform the same regressions with other, weaker, measures of education quality and quantity. The estimates are reported in Table 1.D2 for the richest specification. The coefficients have all the expected sign, but those for the share with a top 20 or top 50 university Master's degree and population potential as remoteness.

[77] See Table 1.D3. For an explanation on the construction of the original variables and the alternative versions, see Appendix 1.A.

[78] Table 1.D4 shows a set of additional regressions where density in 100 km is used as measure of remoteness and the set of control variables is modified.

The results show that a σ increase in density in 100 km around the state capital leads to a 3 percentage points increase in the proportion of legislators with a post-graduate law degree; a 0.26σ increase in experience measured as the number of different public sector and political appointments; and an extra year of political/public sector experience. Among the other covariates, the average level of education in the state is positively correlated with the average level of education of state legislators. The professionalism of the legislature is positively correlated with education and political experience. More professionalised legislatures are composed of better educated and more experienced politicians. The coefficient for the real salary is, however, negative in some specifications. This could be loosely interpreted as evidence of monetary incentives crowding out some of the best candidates for a given level of professionalism of the legislature. Table 1.5 performs the same regressions using the isolation measures for capitals' remoteness.[79] The effect of isolation and adjusted isolation on educational attainment and the quality of education is negative, as expected. However, the estimates are not precise. Panel A of Table 1.5 shows that, in the richer specifications from columns 2, 4 and 6, a standard deviation (σ) change in isolation is associated with a 0.16σ drop in educational attainment and about 0.38σ drop in quality of education. When the percentage of legislators with a postgraduate degree is used as the dependent variable the estimated coefficient does not have the expected negative sign but it is not statistically different from 0. Similarly, when looking at the adjusted measure of isolation in Panel B, a σ change in isolation is associated with a 0.12σ drop in educational attainment and about 0.29σ drop in the quality of education. Again, we are unable to detect a meaningful effect on the percentage with a postgraduate degree.

I interpret the results as evidence of legislators' sorting based on the attractiveness of US state capitals. While the effect is present with all the different measures of remoteness, it is stronger and more significant when discarding the irrelevant information on the distribution further away from the capital. I also estimate the same model introducing different measures of remoteness on the right-hand side. As expected, when I use a density and an isolation measure at the same time, all the variation in legislators' characteristics

[79] The models are the same as in Table 1.4 with the addition of the logarithm of state population on the right-hand side to control for the fact that isolation variables are invariant to it as mentioned in Section 1.4 and explained in Appendix 1.B.

is explained by the narrower variable.[80]

In Table 1.7 I estimate the model from Equation 1.9 where the potentially endogenous measure of a state capital's remoteness is instrumented by the state centroid's remoteness. Panel A shows the estimates from models with density and population potential as measures of remoteness, Panel B with isolation. The point estimates of the 2SLS are consistent with those of the OLS models. Population density and population potential, however, show a strong positive relationship as per the OLS estimates, indicative of the presence of a sorting mechanism. Isolation has a weak negative relationship with educational attainment, while it is positively correlated with years in politics and the public sector. There is no meaningful relationship between the measures of isolation and the percentage of legislators with a postgraduate degree or the number of political and public sector appointments.

1.5.3 Effect on productivity, public good and corruption

In Table 1.8 I study jointly the effect of remoteness and state legislators' average level of education and experience on three outcome measures. These are the number of bills initiated per legislator per day of session, a measure of public good provision and a measure of corruption in the public sector. I focus on two measures of remoteness: the density in 100 km around the capital and non-adjusted measure of isolation from Equation 1.6. All coefficients, when statistically significant, display the expected signs in all models. More density is associated with a larger number of bills initiated per legislator per session, more public good provision and less corruption. Similarly, states with more isolated capitals display lower public good provision and more corruption. Density is strongly associated with bill production while its effect on public good provision and corruption is weaker and vanishes when controlling for legislators' characteristics. This is expected as density is a strong predictor of the average level of education and experience in a legislature. Isolation is positively associated with bill production but the effect is not statistically significant. In columns 3-6, the coefficient for isolation is significant and robust to controlling for legislators' characteristics. The fact that legislators' education and experience do not

[80]See Table 1.D5 in Appendix 1.D.

affect remoteness measured as isolation is somewhat expected given the weak predictive power of the latter. This evidence is in line with the incentive mechanisms uncovered by Campante and Do (2014) who use the same outcome variables.[81]

There are two potential interpretations of these findings. The first one is that variables such as public good provision and public officers' corruption are imperfect measures of legislative performance as they depend on a much wider set of actors, whose behaviour or characteristics are correlated with the capitals' isolation. Hence, they are strongly associated with the latter, while there is not a comparable effect of density. The other interpretation is that there is indeed an incentive mechanism which is better captured by the measure of isolation that closely describes the salience of legislative affairs across the state. Switching the focus to legislators' characteristics, experience seems to be more important than the level of education. I also investigate the impact of the type of education, looking at law graduates and those with a policy or politics-related degree.[82] It is reasonable to assume that a legislator with extensive knowledge of the law or a Master's degree in Public Policy could perform better than one with an O.D. degree[83] or a PhD in Theoretical Physics, despite being *less educated*. This does not only necessarily depend on the subject knowledge acquired during their studies but also on other characteristics that are correlated with the type and overall duration of a person's education as explained by Dal Bó, Finan, Folke, Persson, and Rickne (2017).[84] While a higher percentage of law graduates is associated to more bills initiated, surprisingly it is also linked to less public good provision and more corruption. A possible interpretation is that law graduates are more likely to be career politicians who are less accountable to voters as they rely on strong party support for re-election.[85] However, controlling for experience should, at least partially, capture this effect. Similarly, in their pursuit to advance their name,

[81] The robustness of the effect of isolation in Panel B is somewhat surprising given the poor correlation between my measure of isolation and that used by Campante and Do (2014). See Figure 1.C6 and the different set of control variables used.

[82] See Table 1.D6 in Appendix 1.D.

[83] Doctor of Optometry.

[84] One of the points Dal Bó, Finan, Folke, Persson, and Rickne (2017) make is that academics have substantially higher levels of education, perform better in cognitive tests, but lack leadership and would make worse politicians.

[85] Table 1.D7 compares law school graduates to the rest of the sample. They are more highly educated, have studied in more prestigious institutions and have held more often state legislative leadership positions. However, they don't have significantly more pre-legislative experience and are less likely to be party officials.

they could focus more on individual performance markers (bills production) rather than collective ones. This is consistent with some empirical evidence in the literature on lawyer legislators being more ambitious and pursuing their own interests (Eakins, 2006; Matter and Stutzer, 2015). Another potential explanation is that law graduates are more efficient at using budgetary resources and more capable of rents appropriation. Education in policy or politics is instead strongly associated with more public good provision and less corruption. On one hand this could be the effect of better subject knowledge that allows them to perform better. On the other we could be witnessing the impact of legislators' *intrinsic motivation*. It is indeed plausible that individuals with a strong sense of public purpose would be more likely to choose policy-related studies.

1.5.4 Discussion

The results of the state level analysis indicate a strong impact of density-based measures on the average characteristics of the legislature, while there is a weaker relationship with the measures of isolation. This is consistent with the hypothesis formulated earlier, the better the measure of capital city attractiveness, the stronger the positive correlation with legislators' education and experience. Isolation variables and population potential include more information on the population distribution further away from the capital that should not be related to the sorting choices of legislators.

The paper suggests that the effect of remoteness is driven by demand factors, in particularly amenities that are captured by the density indicators. While the analysis controls for the differences in real salaries and workload across legislatures, we may also have variation in the cost of commuting across states that affects legislators' location choices. However, the within-state analysis finds the existence of an effect of remoteness after controlling for state individual effects, that should be able to capture those potential differences in commuting cost. Yet, there are still supply side factors that may be causing the observed effect of remoteness. For example, we may worry that remote capitals are associated with a pool of less educated legislators. There are factors mitigating this concern. Even if the local supply of candidates is poorer in the proximity of a remote capital, legislators reside in their districts of election all over the state. There could still be substantial differences in terms of availability of education across states but those are accounted for by controlling for the state average educational attainment. On the other side, the analysis does not control for the variation in access to *quality* education. Indeed, some of the North-Eastern states, where there is a large concentration of highly-ranked academic institutions, are among those with the least remote capitals. The regressions where the outcome variable is the quality of education measured using the scores from the Times Higher Education (2019a) ranking could be capturing the effect of proximity to the best schools. However, the highest rated universities are highly selective institutions where students from all over the country (and the world) compete for entry. It is therefore unlikely that proximity could play a big role in determining who ultimately has access to such places. Moreover, in the within-state analysis we witness the existence of an effect of remoteness even after controlling for the average educational attainment in the district (a more precise measure of the actual pool of legislators).

An alternative mechanism that could explain the effect of remoteness is legislators moving to denser places as they advance in their political careers. The results of both the within- and across-state analyses could be due to legislators with higher human capital (more highly educated, graduated in more prestigious institutions and experienced) moving towards more attractive densely populated districts, including those in non-remote capitals. However, in virtually all states candidates need to be resident in the district where they seek to be elected. In presence of a high fixed cost of relocation, this narrative is unconvincing. I further try to address this point by looking at legislators' birthplaces. This part of the analysis is severely limited by lack of data as information on legislators' birthplaces is available for less than half of the initial sample. Nonetheless, we can conjecture that the sub-sample is plausibly self-selected as legislators who advertise their birthplace are more likely to be local. Figure 1.6 shows that more than half of legislators for whom birthplace information is available are *district local*: they were born in the same district where they have been elected. These districts are also on average further away from the capital. Moreover, Figure 1.7 shows that local legislators are less educated, graduated in less prestigious institutions and, on average, are elected in districts with lower education and income per capita. The partial picture we have at this point is consistent with the original hypothesis: low human capital legislators stay in their districts forever while *better* ones move away. However, the rest of the data seems to lead to a different conclusion. If more qualified career politicians move within the state we would expect *state local* legislators (born in the same state but not in the same district) to possess considerably more human capital. Instead, while *state local* legislators have a higher educational attainment, they are less experienced.[86] Moreover, they are also consistently worse across all dimensions then legislators who are born out-of-state or abroad. My interpretation is that legislators *born and raised* in the district are able to leverage their connections to be elected and have often held public interest positions in the same area, while out-of-state legislators need to be more highly educated to be elected. However, only tracking individual legislators' career in the future will allow to provide a more definitive answer.

Finally, the coefficients derived from estimating the model from Equation 1.10 suggest an effect of legislators' human capital on outcome variables, indicating that remoteness does not only affect the composition of legislature but, indirectly, legislators' productivity, public good provision and corruption, through the sorting mechanism. Remote capitals attract less educated and experienced legislators who are more prone to be corrupt and provide less public good. However, these findings show a richer picture since, after controlling for education, there is still a residual effect of the isolation measure of remoteness. This can be interpreted as evidence of the coexistence of the sorting mechanism with the accountability mechanism theorised by Campante and Do (2014): isolation also affects legislators' behaviour by providing weak incentives. More remote capital cities are associated with a lower level of media scrutiny and citizens tend to be less engaged when they are further away from the capital, thus reducing accountability.[87] On one hand

[86] One may worry that the lower experience of career *state local* politicians is due to them moving to higher roles such as federal legislative or executive, while *district local* ones spend a longer time in the same position. In this context, it is worth reminding that experience does not account for state legislative tenure but only captures the relevant public policy or local government positions held before entering the state legislative arena.

[87] It is worth mentioning the existence of a different potential *selection* mechanism. On one hand, if media coverage influences the average talent of politicians (Ferraz and Finan, 2008; Strömberg, 2015) and remote capitals are subject to lower media scrutiny (Campante and Do, 2014), voters would select less qualified politicians. This is however hard to distinguish from candidates' *self-selection* as voters choose among the options presented to them. Even if more informed voters are better at choosing more qualified candidates (Pande, 2011), the average quality depends on the pool of candidates. Similarly, parties could make less effort in candidate selection when state politics is less salient. However, the benefit of deeper screening is likely small and the average candidate quality depends on *self-selection*.

population density around the capital and population potential are better at capturing market access and the attractiveness of the capital city. On the other isolation, based on the distribution of population within the state, can be interpreted as a proxy of state politics salience within the state. Therefore, it is reasonable to interpret that measures of city attractiveness such as density are strongly correlated with legislators' education and political experience, while measures of political salience are not as much. Being able to fully separate the effects of incentives and sorting on outcome variables, however, would require tracking legislators' behaviour over time, as in Snyder Jr and Strömberg (2010).

1.6 Conclusion

This paper leverages a novel data-set on US state legislators' biographies to show that more remote capitals attract individuals with lower and less relevant education as well as less previous experience in the public sector. I first show that districts further away from the capital tend to be represented by less educated and experienced legislators and less often by women. I then proceed to test at state level the relationship between capitals' remoteness and the average level and quality of education as well as experience, finding that remoteness negatively affects all such variables. I build different measures of remoteness and find that those more closely capturing cities' attractiveness show the stronger effect on the average legislators' type. I proceed to use instrumental variables based on the location of the states' centroids, exploiting the fact that US state capitals are often centrally located in the state. Finally, I introduce outcome measures such as the frequency of legislative initiative, public good provision and corruption and provide suggestive evidence that they are affected by capital cities' remoteness through legislators' ability sorting.

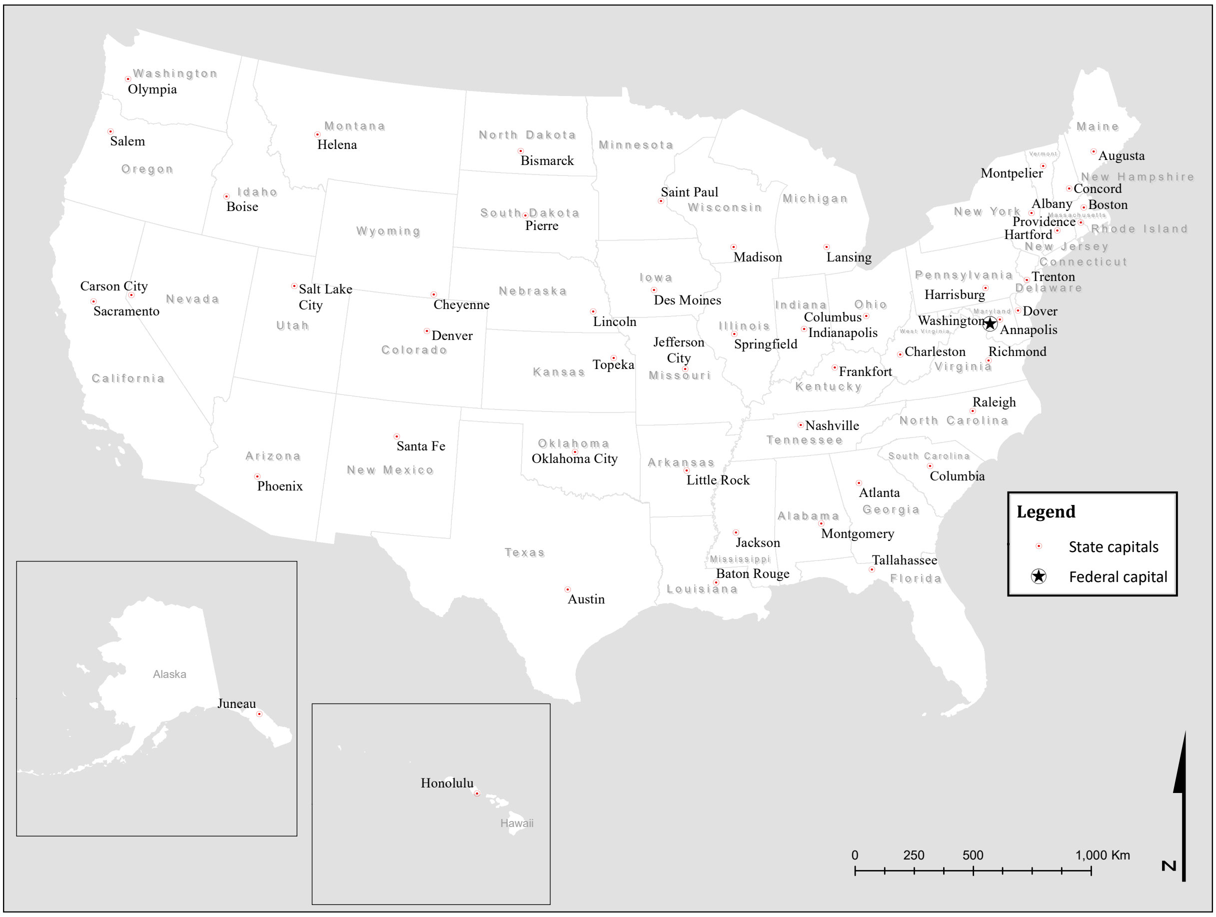
The evidence proposed in this paper should inform policymakers at a time where in different parts of the world there are multiple advocates, in and out of the political sphere, proposing the relocation of capital cities towards the geographical centre of the country for better representation. Countries across the world have seen and will keep seeing capital cities moving inwards. This was the case from Brazil and Nigeria while Indonesia has just started the process of relocating the capital city away from Jakarta.[88] We cannot conclude that moving capital cities and, more generally, governmental institutions away from large, congested capitals is surely detrimental. However, the potential benefits of reducing congestion in the current capitals and diminishing inequality across the country[89] should be weighted against the reduction in the quality of policy-making or service provision. Such trade-off should be well kept in mind before making relocation decisions.

[88]Tan (2019). While the official reason for such moves is usually excessive congestion and the new location tends to be closer to the centroid of the county, Campante, Do, and Guimaraes (2019) show that this may have little to do with overcrowding and equal representation and more with reducing the threat of revolt.

[89]There is, however, evidence that the relocation will have only a modest impact on private sector activity (Becker, Heblich, and Sturm, 2018).

1.7 Figures

Figure 1.1: US state capitals

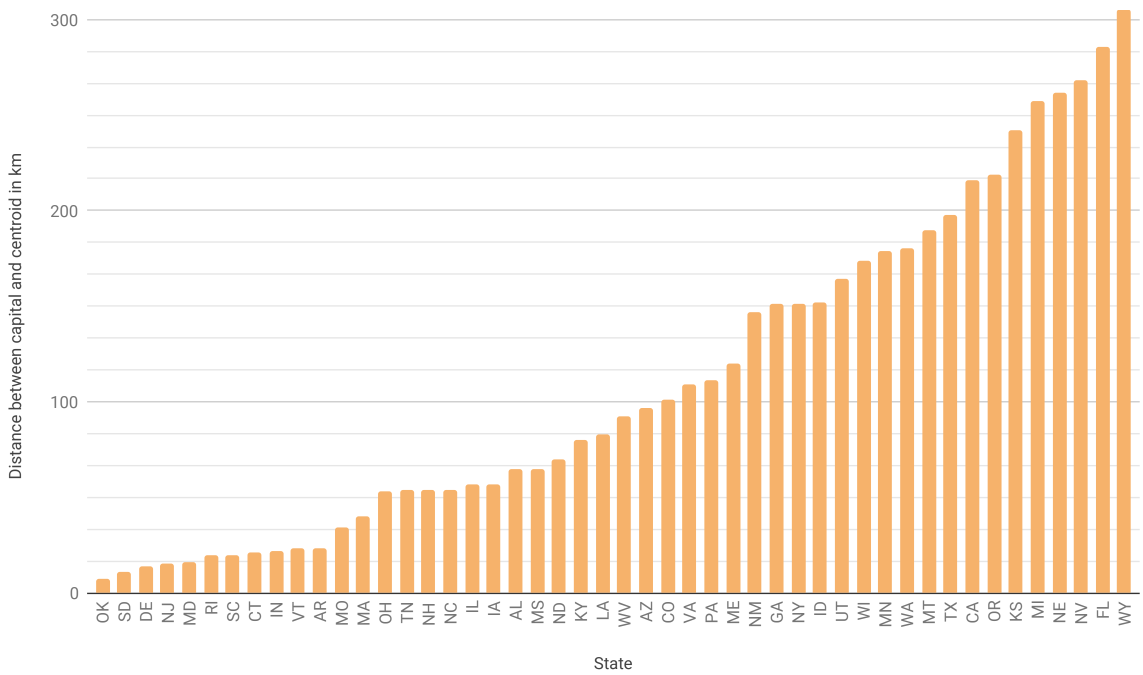


Location of US state capitals. Alaska and Hawaii in bottom left corner.

Figure 1.2: Example of Vote Smart Biography

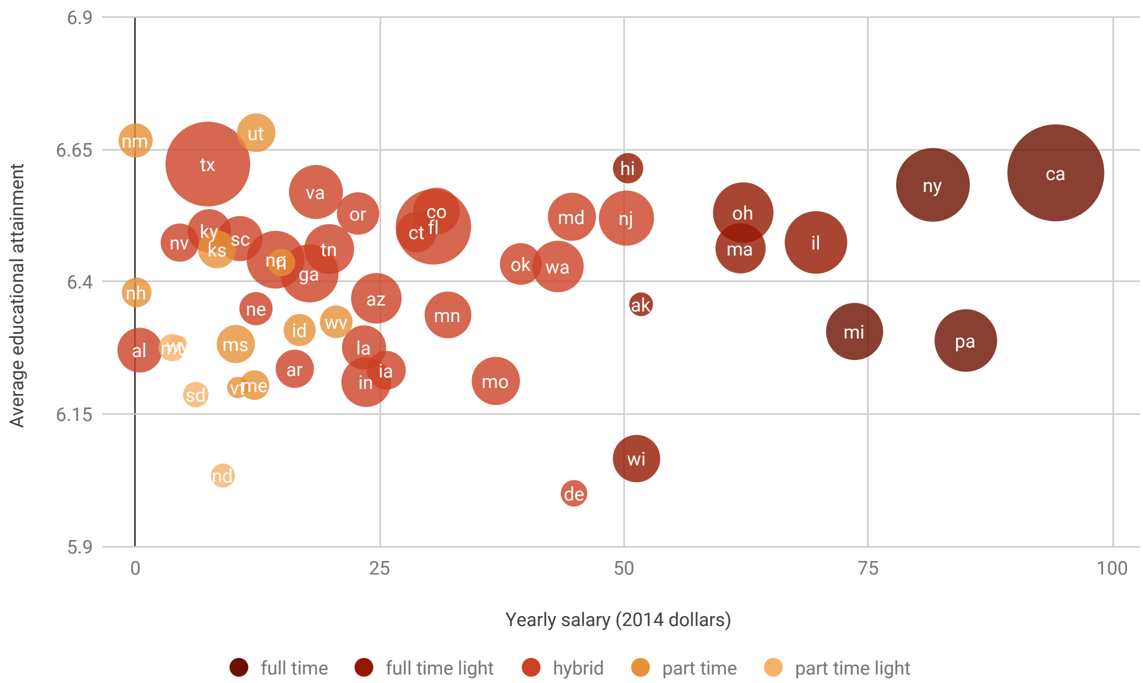
Biography from Vote Smart (2019) for Utah Representative Suzanne Harrison. The Personal, Education and Political Experience sections, containing the data used in this paper, are expanded.

Figure 1.3: Distance between capital cities and centroids



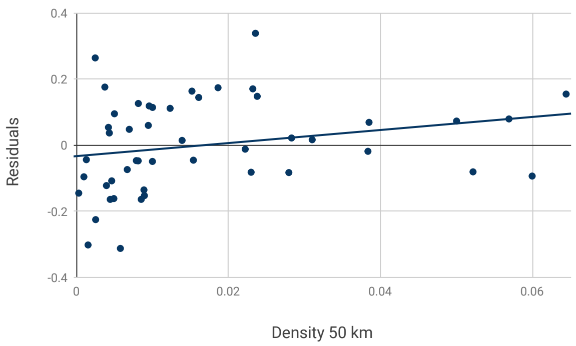
US states ranked according to the geodesic distance in km between the state capital (using the coordinates of the legislature' hall) and the centroid of the state.

Figure 1.4: Education level by legislature

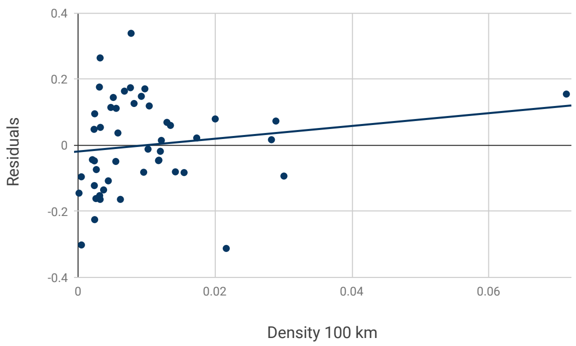


On the horizontal axis: average 2010-2014 legislators' yearly salary measured in 2014 US dollars from Bowen and Greene (2014a). On the vertical axis: average educational attainment using UNESCO ISCED ranking (UNESCO Institute for Statistics, 2011). The size of the bubble represents the US Census Bureau 2017 population estimates of the state from U.S. Census Bureau, Population Division (2017). States are ranked using 5 categories in terms of legislators' time on the job from NCSL (2015b).

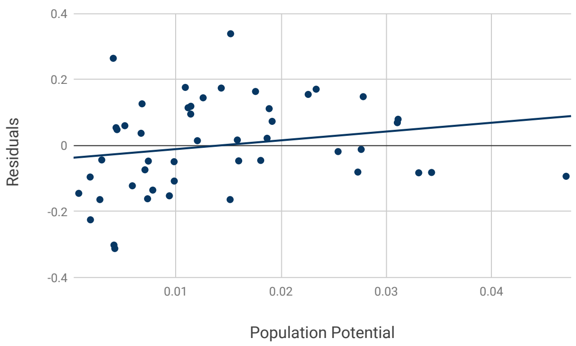
Figure 1.5: Remoteness and average legislators' education



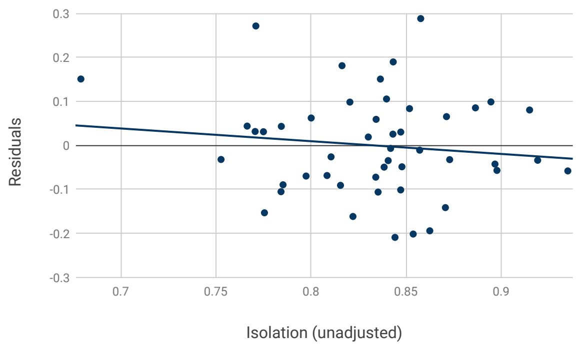
(a) Remoteness = Density 50 km



(b) Remoteness = Density 100 km



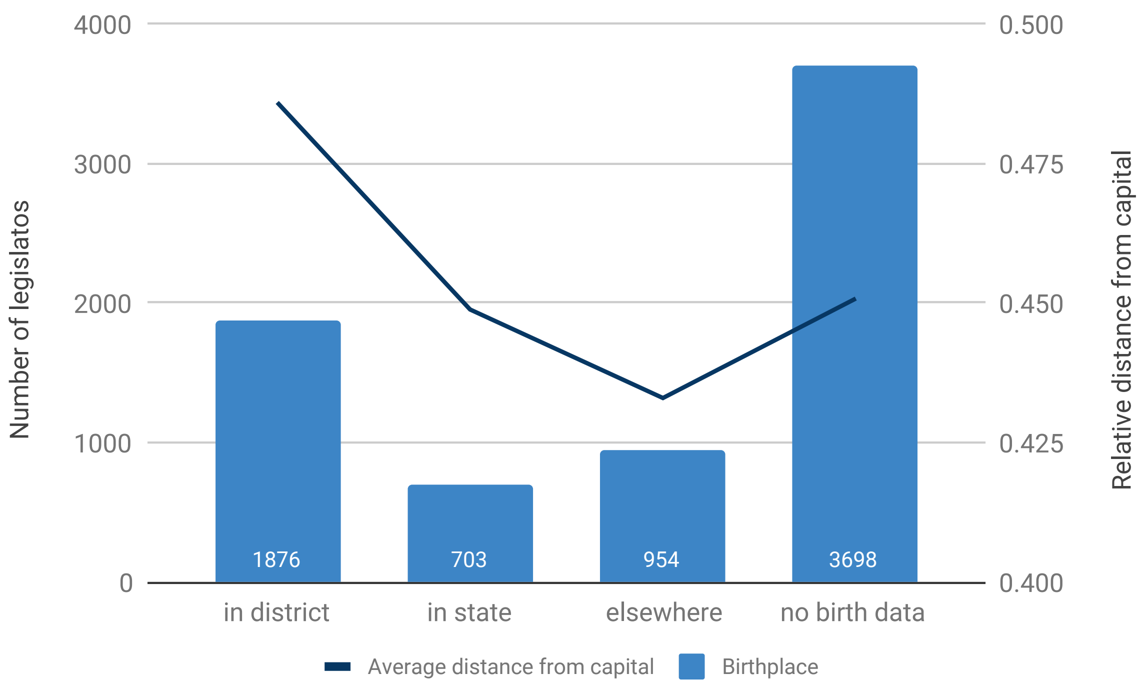
(c) Remoteness = Population Potential



(d) Remoteness = Isolation

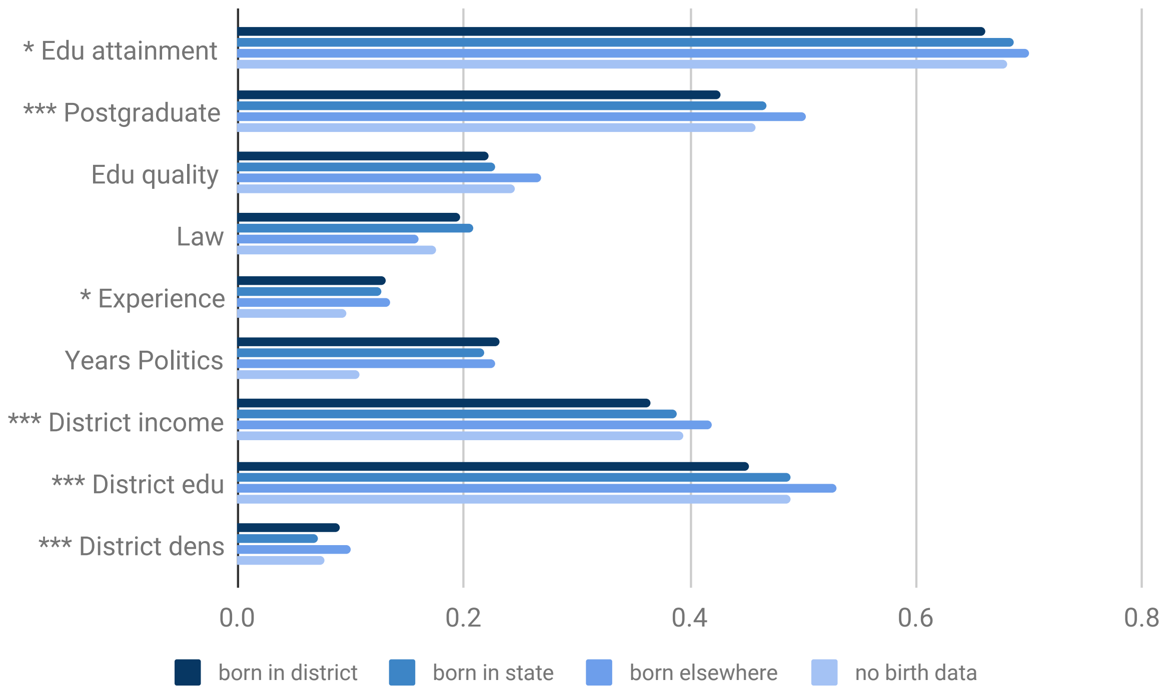
On the vertical axis: residuals from OLS regressions of average state legislators' educational attainment measured on UNESCO Institute for Statistics (2011) ISCED 1-8 scale on the average level of education in the state on the same scale, legislators' real salary and Squire (2017) index of legislative professionalism. On the horizontal axis: different measures of remoteness. Density measures indicate the population density around the state capital in a 50 or 100 km radius (measured in 10,000 people per km^2). Population potential is computed as in Equation 1.5 and measured in 10,000 people per km. Isolation is calculated as described in Equation 1.6 and measured on 0-1 scale.

Figure 1.6: Legislators' birthplaces



The bars show the amount of legislators in the sample that are born in the district of election, in the state of election (but not in the district), elsewhere or have no reported birth data. For each category the line shows the average of the ratio of the distance between the district of election and the state capital and the maximum distance between the capital and any point in the state. The difference in the average relative distance between legislators born in the district and in the state (the first two groups) is statistically significant in a two-sided t-test with $p = 0.0050$. Among the 954 legislators born out of state, 64 were born in US territories (Puerto Rico, Virgin Islands) or abroad.

Figure 1.7: Legislators' characteristics by birthplace



The figure displays the average legislators' characteristics by birthplace. Each category reports the result from a two-sample two-sided t-test comparing legislators born in their district to those born only in the same state with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The categories are from top to bottom: *Edu attainment:* educational attainment measured on ISCED scale converted to a 0-1 scale, *Postgraduate:* share of legislators with a postgraduate degree, *Edu quality:* average legislators' best graduate or postgraduate institution score converted to a 0-1 scale, *Law:* share of legislator has obtained a postgraduate law degree, *Experience:* average number of different public sector and political appointments excluding those as state legislator converted to a 0-1 scale, *Years Politics:* years from the first appointment in the public sector or as elected official, *District income:* average log income per capita in the district of election converted to a 0-1 scale, *District edu:* average educational attainment on ISCED scale in the district of election converted to a 0-1 scale, *District dens:* average population density in the district of election converted to a non-normalised 0-1 scale.

1.8 Tables

Table 1.1: Legislators' education and district remoteness

	Dependent variable:					
		ISCED			Postgrad	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: level						
Log distance	-0.0601***	-0.0353***	-0.0288***	-0.102***	-0.0693***	-0.0587***
	(0.011)	(0.011)	(0.011)	(0.018)	(0.018)	(0.018)
District education		0.263***	0.443***		0.356***	0.576***
		(0.020)	(0.041)		(0.034)	(0.068)
Observations	6063	6063	6063	6063	6063	6063
		Edu quality			Law	
Panel B: quality						
Log distance	-1.626***	-1.002***	-0.924***	-0.0917***	-0.0615***	-0.0603***
	(0.31)	(0.31)	(0.31)	(0.021)	(0.022)	(0.022)
District education		6.462***	5.279***		0.312***	0.289***
		(0.59)	(1.04)		(0.039)	(0.077)
Observations	5720	5720	5720	6069	6069	6069
State FE	✓	✓	✓	✓	✓	✓
Controls			✓			✓

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. In Panel A OLS with dependent variable=educational attainment measured on ISCED scale (1-3), Probit with dep. var. dummy=1 if legislator has obtained a Master's degree (4-6). In Panel B OLS with dep. var.=legislators' best graduate or postgraduate institution score (1-3), Probit with dep. var. dummy=1 if legislator has obtained postgraduate education in law (4-6). Log distance is the logarithm of the distance in km between the centroid of the district of election and the legislative hall. District education is educational attainment in the district of election measured on ISCED scale. *Controls* include the logarithm of the 2012-16 average district income per capita measured in 2016 inflation adjusted dollars, population density in the district of election and a dichotomous indicator for being a state senator. All models include state fixed effects.

Table 1.2: Experience and gender in state legislatures

	Dependent variable:					
	Experience			Years politics		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: experience						
Log distance	-0.0165	-0.0124	-0.0159	-0.375***	-0.412***	-0.423***
	(0.016)	(0.017)	(0.017)	(0.14)	(0.14)	(0.14)
District education		0.0429	0.184***		-0.391	1.346***
		(0.033)	(0.059)		(0.24)	(0.45)
Observations	7205	7205	7205	7205	7205	7205
	Woman			Woman with children		
Panel B: gender						
Log distance	-0.124***	-0.0865***	-0.0779**	-0.115***	-0.0810***	-0.0869**
	(0.017)	(0.018)	(0.033)	(0.020)	(0.021)	(0.034)
District education		0.585***	0.532***		0.569***	0.485***
		(0.064)	(0.11)		(0.074)	(0.11)
Age			0.0124***			0.0138***
			(0.0027)			(0.0028)
Observations	6998	6998	2491	5401	5401	2473
State FE	✓	✓	✓	✓	✓	✓
Controls		/ ✓	✓/ ✓		/ ✓	✓/ ✓
Relationship			/ ✓			/ ✓

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. In Panel A OLS with dependent variable=number of different public sector and political appointments excluding those as state legislator (1-3), OLS with dep. var.=years from the first appointment in the public sector or as elected official (4-6). In Panel B Probit regressions with dep. var. dummy=1 if the legislator is female (1-3), or dep. var. dummy=1 if the legislator is female and has children (4-6). Log distance is the logarithm of the distance in km between the centroid of the district of election and the legislative hall. District education is educational attainment in the district of election measured on ISCED scale. *Controls* include the logarithm of the 2012-16 average district income per capita measured in 2016 inflation adjusted dollars, population density in the district of election and a dichotomous indicator for being a state senator. Controls are included in columns 3 and 6 of Panel A and columns 2-3, 5-6 of Panel B. Age is measured in calendar years on Jan, 1st 2019. *Relationship* is a set of dummies to control for relationship status: *partnership*, *married*, *single*, *widowed*, *divorced* or *no status*. Age and Relationship are only included in Panel B columns 3 and 6. All models include state fixed effects.

Table 1.3: Education and different measures of remoteness

	Dependent variable: Educational attainment ISCED scale								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Isolation	-1.074*								
	(0.54)								
Isolation (adj.)		-1.048**							
		(0.50)							
Density 30km			2.212***						
			(0.71)						
Density 50km				4.225***					
				(1.12)					
Density 100km					6.051***				
					(1.63)				
Pop potential						5.802***			
						(1.78)			
Capital Large							0.0359		
							(0.041)		
Capital/Largest								0.0276	
								(0.050)	
Capital share									0.155
									(0.29)
R^2	0.367	0.371	0.259	0.304	0.257	0.262	0.323	0.316	0.315
Observations	48	48	48	48	48	48	48	48	48

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on UNESCO Institute for Statistics (2011) ISCED 1-8 scale. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area. columns 1-2 also include the log of state population from average 2012-2016 ACS Census Bureau estimates. Each model uses a different measure of remoteness. Isolation measures are built as described in Equations 1.6 and 1.7. Density measures indicate the population density around the state capital in a 30, 50 or 100 km radius (measured in 10,000 people per km^2). Population potential is computed as in Equation 1.5 (measured in 10,000 people per km). Capital Large is a dummy=1 if the capital is the largest city. Capital/Largest is the ration of state capital population over the population of the largest city. Capital share is the fraction of the state population living in the state capital (from Census Bureau 2017 population estimates).

Table 1.4: Education and capital city attractiveness

	Dependent variable:					
	Edu attainm		Postgrad		Edu quality	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Density 100km	5.847***	5.316***	3.334***	2.953***	28.53*	19.72*
	(1.71)	(1.12)	(0.71)	(0.69)	(15.5)	(10.4)
State education	0.0474	0.0555	0.0212	0.0406	0.809	0.694
	(0.099)	(0.11)	(0.058)	(0.059)	(0.49)	(0.53)
Real salary		-3.696**		-2.383**		-7.686
		(1.54)		(0.94)		(5.91)
Professionality		1.079***		0.719***		2.155
		(0.39)		(0.24)		(2.09)
Panel B						
Pop potential	5.603***	4.070*	2.456**	0.983	37.43***	35.12**
	(1.89)	(2.26)	(1.06)	(1.20)	(12.8)	(16.1)
State education	0.0572	0.0752	0.0356	0.0649	0.737	0.558
	(0.098)	(0.11)	(0.059)	(0.058)	(0.47)	(0.55)
Real salary		-3.615**		-2.397**		-6.471
		(1.61)		(0.97)		(6.40)
Professionality		0.978**		0.741***		0.557
		(0.43)		(0.26)		(2.53)
Panel C						
Density 50km	4.167***	3.451***	1.798**	1.297	23.86**	18.86**
	(1.20)	(1.21)	(0.81)	(0.90)	(10.9)	(8.91)
Density 30km	2.138***	1.701**	0.757*	0.424	14.53**	11.96*
	(0.74)	(0.82)	(0.44)	(0.47)	(6.60)	(6.11)
Accountability		✓		✓		✓
State size	✓	✓	✓	✓	✓	✓
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on ISCED 1-8 scale (1-2), the share of legislators with a postgraduate degree (3-4), the average legislators' best graduate or postgraduate institution score (5-6) from Times Higher Education (2019a) ranking. In Panel A and C density indicates the population density around the state capital respectively in a 30, 50, 100 km radius (measured in 10,000 people per km^2). In Panel B population potential is computed as in Equation 1.5 (measured in 10,000 people per km). All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area. Models 2, 4 and 6 also include the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). In 2, 4 and 6 Montana is excluded as there is no political newspaper coverage data.

Table 1.5: Education and capital city isolation

	Dependent variable:					
	Edu attainm		Postgrad		Edu quality	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Isolation	-0.856	-0.522	0.00338	0.134	-5.670	-6.615*
	(0.53)	(0.59)	(0.33)	(0.38)	(5.12)	(3.86)
State education	0.103	0.141	0.0897*	0.114*	0.921*	0.596
	(0.080)	(0.11)	(0.051)	(0.063)	(0.53)	(0.50)
Real salary		-4.006**		-2.451**		-9.825
		(1.56)		(0.98)		(5.87)
Professionality		0.631		0.519*		1.191
		(0.42)		(0.28)		(2.23)
Panel B						
Isolation (adj.)	-0.850*	-0.490	-0.00706	0.137	-5.259	-6.304*
	(0.50)	(0.56)	(0.31)	(0.36)	(4.83)	(3.67)
State education	0.0997	0.142	0.0888*	0.115*	0.933*	0.597
	(0.080)	(0.11)	(0.051)	(0.063)	(0.52)	(0.50)
Real salary		-4.001**		-2.449**		-9.778
		(1.56)		(0.98)		(5.87)
Professionality		0.630		0.518*		1.191
		(0.42)		(0.28)		(2.23)
Accountability		✓		✓		✓
State pop & size	✓	✓	✓	✓	✓	✓
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on ISCED 1-8 scale (1-2), the share of legislators with a postgraduate degree (3-4), the average legislators' best graduate or postgraduate institution score (5-6) from Times Higher Education (2019a) ranking. In Panel A remoteness is measured as isolation on a 0-1 scale from Equation 1.6. In Panel B remoteness is adjusted isolation on a 0-1 scale from Equation 1.7. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, the log of state population from average 2012-2016 ACS estimates. Models 2, 4 and 6 also include the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). In 2, 4 and 6 Montana is excluded as there is no political newspaper coverage data.

Table 1.6: Political experience and capital city remoteness

	Dependent variable:					
	Law		Experience		Years politics	
	(1)	(2)	(3)	(4)	(5)	(6)
Density 100km	3.138***	2.675***	22.29***	15.77***	114.3***	91.08***
	(0.81)	(0.81)	(3.69)	(3.75)	(20.7)	(28.8)
State education	-0.126**	-0.112*	0.733**	0.774**	-1.673	-3.683**
	(0.054)	(0.055)	(0.34)	(0.35)	(2.06)	(1.76)
Real salary		-1.878***		2.624		-12.45
		(0.66)		(4.38)		(16.6)
Professionality		0.745***		1.226		9.290*
		(0.14)		(1.17)		(4.66)
Accountability		✓		✓		✓
Term limits				✓		✓
State size	✓	✓	✓	✓	✓	✓
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is the share of legislators with a postgraduate degree in law (1-2), the number of different public sector and political appointments excluding those as state legislator (3-4), the years from the first appointment in the public sector or as elected official (5-6). The explanatory variable of interest is population density around the state capital respectively in a 100 km radius (measured in 10,000 people per km^2). All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area. Models 2, 4 and 6 also include the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). Models 4 and 6 also include a dummy=1 if the state has term limits, a dummy=1 if the state has binding term limits and the interaction between the two. In 2, 4 and 6 Montana is excluded as there is no political newspaper coverage data.

Table 1.7: 2SLS Centroid remoteness as instrument for capital city remoteness

	Dependent variable:							
	Edu attainm		Postgrad		Experience		Years politics	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Density 100km	5.890***		3.563***		20.85***		106.9***	
	(1.23)		(0.49)		(3.78)		(17.8)	
Pop potential		11.50***		4.944**		30.69***		104.6
		(3.90)		(2.07)		(11.6)		(66.0)
Panel B								
Isolation	-1.497		-0.155		-0.590		26.23**	
	(0.94)		(0.62)		(2.65)		(11.9)	
Isolation (adj.)		-1.430		-0.0905		0.678		10.79
		(1.10)		(0.77)		(2.97)		(13.9)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	47	47	47	47	47	47	47	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. 2SLS regressions where the dependent variable is the average state legislators' educational attainment measured on ISCED 1-8 scale (1-2), the share of legislators with a postgraduate degree (3-4), the number of different public sector and political appointments excluding those as state legislator (5-6), the years from the first appointment in the public sector or as elected official (7-8). In Panel A remoteness is measured as population density around the state capital respectively in a 100 km radius (measured in 10,000 people per km^2) and population potential computed as in Equation 1.5 (measured in 10,000 people per km). In Panel B remoteness is measured as isolation on a 0-1 scale from Equation 1.6 and adjusted isolation on a 0-1 scale from Equation 1.7. Each measure of remoteness is instrumented in the first stage by the equivalent variable built with respect to the centroid of the state. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). Models 5-8 also include a dummy=1 if the state has term limits, a dummy=1 if the state has binding term limits and the interaction between the two. Panel B also includes the log of state populations in all regression. In 2, 4 and 6 Montana is excluded as there is no political newspaper coverage data. Olea and Pflueger (2013) first stage robust F-stat is 60.299 for Density 100km, 14.130 for Pop potential, 22.083 for Isolation, 14.130 for Isolation (adj.).

Table 1.8: Remoteness and legislative outcomes

	Dependent variable:					
	Bills initiated		Public good		Corruption	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Density 100km	14.34***	14.15***	1.214**	-0.302	-1.482	-0.603
	(4.12)	(4.44)	(0.54)	(0.63)	(1.42)	(1.83)
Edu attainment		0.275		0.151**		0.149
		(0.23)		(0.058)		(0.12)
Experience		-0.0731		0.0580*		-0.108**
		(0.056)		(0.029)		(0.050)
Panel B						
Isolation	0.540	0.767	-0.552*	-0.534*	1.238**	1.467***
	(0.71)	(0.68)	(0.28)	(0.28)	(0.54)	(0.49)
Edu attainment		0.378		0.0812		0.290**
		(0.24)		(0.077)		(0.11)
Experience		-0.0110		0.0487		-0.100*
		(0.10)		(0.035)		(0.054)
Controls	✓	✓	✓	✓	✓	✓
Observations	48	48	48	48	48	48

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is bills per legislator per day of session (1-2), public good expenditure from Campante and Do (2014) (3-4), Glaeser and Saks (2006) measure of state level corruption (5-6). In Panel A remoteness is measured as population density around the state capital respectively in a 100 km radius (measured in 10,000 people per km^2). In Panel B remoteness is measured as isolation on a 0-1 scale from Equation 1.6. *Edu attainm* is the average state legislators' educational attainment measured on ISCED 1-8 scale. *Experience* is the average legislators' number of different public sector and political appointments excluding those as state legislator. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, real salary, Squire (2017) index of legislature's professionalism, log of state income, voter turnout from in the last 3 presidential elections from CSG (2018), census region dummies. Panel B also includes the log of state populations in all regression.

1.A Data

Gender and age: Almost for all legislators gender is reported. Only 2784, in the final data-set excluding Alaska and Hawaii, report the date of birth. I compute the age on January 1st, 2019, so that I am able to include also those who report only the year of birth.

Children: The number of children was extracted from a text string using an algorithm using keywords such as *child(ren)*, *daughter(s)*, *son(s)* to identify the position in the string. If the singular form was used a number of children equal to one is assumed. In the few cases when the plural is used and the number was not specified I assume a number of children equal to two, which coincides to the mean and mode of the distribution. In any case when the number of children was larger than zero, I record in a dummy variable that the legislator has children.

Relationship: I build an algorithm to extract the type of relationship from the text string with information on the family. I am able to categorise if a legislators is: *divorced*, *married*, in a *partnership*, including those who report themselves as engaged, *single*, *widowed* or *no status*. I determine if the legislator is in a same-sex couple by comparing their gender to that of the partner. In case there was no specific indication such as *wife* or *husband*, I used the following strategy to trace the gender of the partner of all those that reported being married, engaged or having a partner. I used the Social Security Administration top 1000 names for babies born in 1961.[90] The chosen year is the median year of birth for the population of interest. If the official reports the name of the partner and such name coincides to a *top female name* they are reported as female, the opposite for males. If the gender of the partner coincides with that of the candidate, the latter is reported to be in a same-sex relationship. There are cases in which the partner's name is on both the male and female lists. In such instances, I record how high the name is ranked on both list and assigned the gender with the highest rank. There are only 56 legislators who according to the variable I built are in a same-sex relationship, not offering enough variation to be able to study the effect of remoteness on it.

Birthplace: Information on the birthplace is available for only 3533 legislators and it is quite noisy. As for most variables, I manipulate the text string to extract a machine readable location name and then geocode the output to obtain the best approximation of the coordinates of the location. I then match the coordinates with all Upper House and Lower House districts in a 15 km radius. I choose the above buffer to better deal with cases where the data report a large city that is divided in multiple small districts. If one of the matched districts is the same as the district of election I record the legislator

[90]The sample is taken from Social Security card application data (Social Security Administration, 2019).

as *district local*. If not, I check if the state in which the legislator was born is the same as that of the district of election and record them accordingly as *state local*. Because of the 15 km buffer there are 28 cases of legislators that I consider *district local* but are actually born in the same metropolitan area but in a different state (for example born in Washington but elected in a neighbouring Virginia district or, similarly, born in New York and elected in New Jersey). There are 64 legislators born outside of the US states: 16 in US territories (15 in Puerto Rico and 1 in the Virgin Islands), the rest abroad (Germany, Jamaica and Vietnam are the most common origins).

Education: The education data-set contains 12209 entries for 6194 legislators, close to an average of 2 per person, in text strings containing information on the degree, the field and the name of the institution attended, which I extract and separate using algorithms. I start by eliminating the entries that constitute on-the-job training or honorary positions or just professional entry exams. For example, residency for nurses and doctors. This does not generate any data loss since the education section of the biography would also include the relevant degree - respectively categorised as Master of Science (in nursing) and Medicinae Doctor. The same holds true for licenses (law, accounting), fellowships and honorary degrees. The above process eliminates 51 observations, representing approximately 0.42% of the total. In certifications (ISCED level 4) I include all sorts of professional certifications and military academies that are not colleges. For some legislators the text did not explicitly indicate whether the degree pursued was obtained or not. If it is a college or a university I assume a Bachelor's degree (ISCED level 6), the median degree. That potentially introduces some measurement error as some of such candidates could actually have a Master's degree (ISCED level 7) or have not finished their degree or obtained an Associate's degree. Given that the number of legislators with a Master's degree is larger than that of those with an Associate's degree, the bias is likely to be upwards. For robustness, I generate two other versions of the ISCED education level variable that correct for this. In the first, I attribute to all such individuals the next lower level of education. The underlying assumption is that if someone has attended an undergraduate degree without completing it they must have at least obtained a high school diploma. In the second, I discard from the sample all such cases. The results are unaffected by using any of the three measures, suggesting that the non-reporting of the degree is not correlated with remoteness. I keep in the data-set short programmes such as summer school or executive programmes or leadership programmes. These short-term degrees are not accounted for when building the level of education variables as well as the academic institution is not considered for the quality of education variables. I use them to build dummy variables. However, very few legislators have obtained such degrees - only 34 attended an executive course - not providing enough variation for this variables to be used in the analysis. I create a dummy variable indicating if the legislator has attended a Bachelor's degree or higher in Public Policy, Public Administration, Policy, Politics, Government. Joint degrees are also taken into account. I also record in a dummy variable if a legislator has graduated from law school. I am able to match 182 US schools[91] with their Times Higher Education (2019a), abbreviated THE, ranking plus the international institutions in the top 50 of the ranking.[92] 35% of the 11930 reported college or university degrees have been obtained in one of the institutions included in the THE ranking, with an average score of 53.6.out of 100.[93] For the analysis I approximate the scores to integers. I also attach a score to all non-ranked universities, equal to the lower-bound of the last score band.[94] I also record if the legislator has received a Master's degree or higher from one of the top 20 or top 50 universities in the ranking. There are respectively 447 and 648 such cases, representing 7.8% and 11.3% of those with a reported academic institution. We can expect these numbers to be upward biased as legislators attending prestigious institutions are more likely to advertise it. I attach to each legislator the score and ranking of the best university they attend. I create another set of measures in which I only look at the university where they obtained the highest degree. I only consider Bachelor's degree (4-year undergraduate degree) or higher.

Political career: The paper uses two types of variables that measure quality in terms of political experience. One is focused on the *lenght of the resume* - looking at how many roles a politician has held in their career. The second looks at the number of years the politician has been involved in politics. The process of building such variables presents an inherent difficulty given by the fact that the data collection results is a series of strings prone to typos and inconsistencies. Moreover, an additional difficulty is given by the fact that different US States use different terminologies for the same *official*'s levels, for example the Lower House at State Level, usually House of Representative as at Federal Level, in some states is called State Assembly (California and New York), in others House of Delegates (Maryland and Virginia). Similarly, county level *officials* are most commonly referred to as *supervisors* but can also be *selectmen* (New England) or *freeholders* (New Jersey) or simply members of the county council. In some states, such as New York, groups of counties use different nomenclatures. As a result, it becomes harder to identify classes of *officials* with the same level of responsibility. Moreover, there is some ambiguity deriving from the overlap of some definitions: an *officials* declaring his job title to be simply *member* could be an Assembly Member (state legislator), a county council member, a city/town council member or just a member of a legislative committee or a local political association. Similarly, a *delegate* could be a state legislator from a State where the Lower House is called House of Delegates or a delegate at the Democratic or Republican National

[91] A total of 1,258 US universities are ranked.

[92] In the data-set at least one alumnus form each of the following: Cambridge University, Oxford U, the London School of Economics and the U of Edinburgh in the UK; U of Toronto, U of British Columbia and McGill U in Canada; Peking U; Hong Kong U and Heidelberg U.

[93] The score is based on 5 performance indicators: teaching; research volume, income and reputation; research influence; international outlook; and industry income. See Times Higher Education (2019b) for an explanation of the methodology and the data collection process.

[94] Universities ranked 200 and lower are attached score and rank bands rather than an exact number and position. For all those included in the ranking, I assume a score equal to the mean of the band.

Convention. As a result, job title *per se* is not sufficient to properly classify the different political resume entries (a total of 21,897 observations, an average of 3 per legislator). The task required coupling data manipulation algorithms using job title and organisation information derived from the text strings with careful human observation of the class of data entries. The (self-reported) political career data also includes some events such as the date when someone won a special election or was appointed to the house or senate. I eliminate these as they do not constitute political experience *per se*. Eliminating such entries does not lead to data loss as the biography also includes the relevant position. For a few representatives (7, representing the 0.08% of the total population), the summary also includes the date when they switched allegiance from the Democratic party to the Republican or the opposite. The political biography also includes positions that do not seem immediately relevant to a political career. This could be due to errors in compiling the Vote Smart profile or observable characteristics of the specific jobs that made them worthy to be included under the political career section rather than the professional career one. I identify 65 such cases (0.3% of total observations) such as: teacher, reporter, police officer, paralegal, solicitor, pastor, served in armed forces, business owner and several cases of employment in the private sector. I discard such entries. The biography also includes positions held during the legislature such as being the majority or minority leader, whip or speaker, party delegate at the national convention or president/chair of county or city level councils as well as roles in legislative committees and public administration boards. These roles are undoubtedly source of additional experience and exposure. Moreover, having been selected for any of the above positions could be a signal of relative stature in the party that may be correlated with quality as well as a measure of the size of the politician's own network and political capital. I exclude from experience the roles held as state legislators or senators as I want to capture the *ex-ante* experience. I count only different positions held: holding the same role for many years it will be counted as one. The other measure instead counts the years since the first appointment in a political role.

Party data: I record the party affiliation of legislators. In some states, like New York, it is common to have more than one party affiliation. I record all of them and build a dummy indicating if a legislator is affiliated with multiple parties. I then also record the main party affiliation. Excluding the North Dakota NPL and Minnesota Democratic Farmer and Labor party that are effectively branches of the Democratic party (I record them as such), there are only 73 with a different primary party affiliation, amongst whom 51 have no affiliation at all. There are 128 legislators with multiple party associations. Almost all of them (125) are in the state of New York.

State district: I use an algorithm to extract from the political information the name of the districts of election and re-code them to match them to the Census Bureau Lower and Upper House districts IDs to be able to merge them with the Census Bureau Tiger Line geographical data to calculate distances in ArcGIS. The state of New Hampshire

Table 1.A1: Main biographical variables, summary statistics

Variable	Mean	SD	Min	Max	N Obs
Age	58.19	12.34	19	92	2748
Female	0.284	0.451	0	1	6998
Has children	0.849	0.358	0	1	5424
ISCED level (1-8)	6.39	0.863	3	8	6063
Has a postgraduate degree	0.457	0.498	0	1	6063
University score	37.48	22.59	19	96	5720
Has a law postgraduate degree	0.183	0.387	0	1	6069
Experience (nr different political roles)	1.702	1.370	0	15	7206
Years since first appointment	10.35	9.909	0	62	7206
Democrat	0.466	0.500	0	1	7215

uses floterial districts in their lower-chamber (SLDL) plan. Floterial districts are overlay districts made up of two or more discrete districts.[95] The districts making up floterial districts are close but not necessarily contiguous. In this case I just use the first district (by district name) among those making up the floterial district to calculate the distance to the capital.

Legislatures' characteristics: Variables capturing legislatures' regulations come from a number of sources. Real salary is the ratio between Bowen and Greene (2014a) 2010-2014 average nominal legislative salary and gross median rents data from ACS 2010-2014 Census Bureau estimates, measured in 2010 dollars. Sessions lenght is taken from Bowen and Greene (2014a) 2010-2014 average. Legislative professionalism is Squire (2017) index of professionalism on 0-1 scale, where 0 is non-professional and 1 is as professional as the US Congress. The information on term limits is manually collected from NCSL (2019). It includes a dummy for term limits being in place, when they were in introduced and if they were repealed. The information on requirements to run for office is manually collected from the Book of States (CSG, 2018). It includes required minimum age, length of citizenship in the US, length of state residency, length of district residency and dummies for the presence of a requirement of state citizenship and voter registration.

State and district level variables: Data on educational attainment of the population over 25 and average GDP per capita at state and district level is taken from the American Community Survey 2012-2016 averages (U.S. Census Bureau, Geography Program, 2016). Population data at state, district, census tract and block group level, used to build the remoteness measures also derives from the American Community Survey 2012-2016 averages. Newspaper coverage is taken from Campante and Do (2014) and is the first principal component of search terms related to state politics from the webiste NewsLibrary.com.

[95] A listing of the floterial districts and their component districts is available as a report. See `https://www2.census.gov/programs-surveys/decennial/rdo/mapping-files/2012/2012-state-legislative-bef/nh-2012-floterial-list.pdf`.

Voter engagement is measured using the average voter turnout as a fraction of all voters in the last 3 presidential elections 2008, 2012 and 2016 form CSG (2018).

Remoteness: Population density is measured as:

$$PopDensity_s = \frac{\sum_j^N pop_{js} \times w_{js}}{area_s}$$

pop_{js} is the population of block group j in the circle around state s capital; $w_{js} \in (0, 1]$ is a weight that attributes a share of block group j's population, if not fully included in the circle, equal to the ratio between the size of the intersection between the block group and the circle and the total area of the block group, $w_{js} = 1$ if the block group is fully included in the circle. $area_s$ is the total amount of land area in the circle around the capital. Population potential is measured as:

$$PopPotential_s = \sum_i \frac{pop_i}{d_{si}}$$

pop_i is the population of census tract i in the same state s as the capital city and d_{si} is the distance between state s capital and the centroid of the census tract. Both population density and potential are respectively scaled to 10,000 people per km^2 and 10,000 people per km.

The isolation index is computed as $1 - GCISC$ where the GCISC is the convex version of the Centered Index of Spatial Concentration, developed by Campante and Do (2010).[96] There are two variants of the CISC - a convex one, also known as Gravity-based CISC or GCISC, and a linear one, or LCISC. They both result in a concentration index on a 0 to 1 scale. Following Campante and Do (2014), I focus on the GISC based measure as it assigns more weight to population concentration closer to the point of interest, the capital in this case.[97] $Isolation$ is $1 - Concentration$, the GISC[98] based formulas are:

$$Isolation_c = 1 - \sum_i \Big(1 - \frac{log(d_{ci})}{log(d_{max})}\Big) s_i \qquad AdjIsolation_c = 1 - \sum_i \Big(1 - \frac{log(d_{ci})}{log(d_{max_c})}\Big) s_i$$

[96] Campante and Do (2010) argue that the CISC is superior to other measures of spatial concentration because it focuses on concentration around a point rather than over space and guarantees the following three properties: *subgroup consistency* - the overall distribution becomes more concentrated around a point when a subgroup of the population becomes more concentrated around that point; *monotonicity* - the distribution around a point becomes more concentrated if some people move closer to the point; *rank invariance* - difference units of measure for the distance yield the same measure. The *spatial GINI index*, for example, would be unaffected by moving a whole mass of population from one area to another inhabited area of equal size closer to the point of interest, violating monotonicity.

[97] Campante and Do (2014) claim that another benefit of the GISC is reducing the measurement error induced by approximating the exact spatial distribution of the population. This is not a concern in my case since I use population data at a smaller level - census tract - compared to counties, the smallest administrative unit available for their whole period of interest, 1920 to 2000. They also show that the two resulting measures of isolation are highly correlated and produce the same results.

[98] The linear measure is exactly the same without taking the logarithm of the distance.

d_{ci} is computed using ArcGIS as the geodesic distance between the capitol hill and the centroid of a census tract. To measure the maximum distance in each state, I use ArcGIS to fragment state borders in a series of points, one for each vertex, for a total of roughly 900,000 state vertices. I then compute the distance from each capital to each vertex and find the maximum to use as d_{max} for the unadjusted measure and the maximum within each state to use as d_{max_c} in the state-size adjusted measure. The instrumental variables using the centroids are computed in the exact same way substituting the coordinates of the state capitol with the coordinates of the state centroid. I exclude Alaska and Hawaii. Figure 1.C6 plots my measure of isolation, computed using ACS 2012-2016 population estimates at census tract level, versus that from Campante and Do (2014), that use the average GCISC measure from 1920 to 2000, up to 1970 in their main specifications, computed using population census data at county level. The correlation is poor and the distribution of capitals' isolation has shifted to the right with time. I interpret this as increased agglomeration that is happening faster away from capital cities, other cities are agglomerating more relative to capital cities, that become more isolated with respect to the population distribution.

Table 1.A2: Remoteness, summary statistics

Variable	Mean	SD	Min	Max	N Obs
Density 30km	0.0281	0.0271	0.0007	0.1022	48
Density 50km	0.0168	0.0168	0.0003	0.0644	48
Density 100km	0.0100	0.0117	0.0001	0.0713	48
Population potential	0.0142	0.0104	0.0008	0.0471	48
Isolation	0.8329	0.0480	0.6789	0.9349	48
Isolation (adjusted)	0.8873	0.0387	0.8073	0.9584	48

Bills introduced: I only use bills introduced to measure productivity as some states do not record resolutions. I use an average to avoid yearly fluctuations. I exclude the 2017 record for Georgia and Wyoming because bills and resolutions are mixed. The data on bills is derived from Table 3.19, Bills and Resolutions Introductions and Enactments from the 2015, 2016, 2017, 2018 and 2019 editions of the Book of States. Each table provides information on the previous year's session. I compute productivity as the average bills introduced per legislator per day of session. The number of days in session I use Bowen and Greene (2014a) 2010-2014 average. The number of legislators is manually collected from the Book of States (CSG, 2018).

Public good provision: I use Campante and Do (2014)' measure of public good expenditure, the share of state expenditure assigned to education, public welfare, health, and hospitals from US Statistical Abstract 2012.

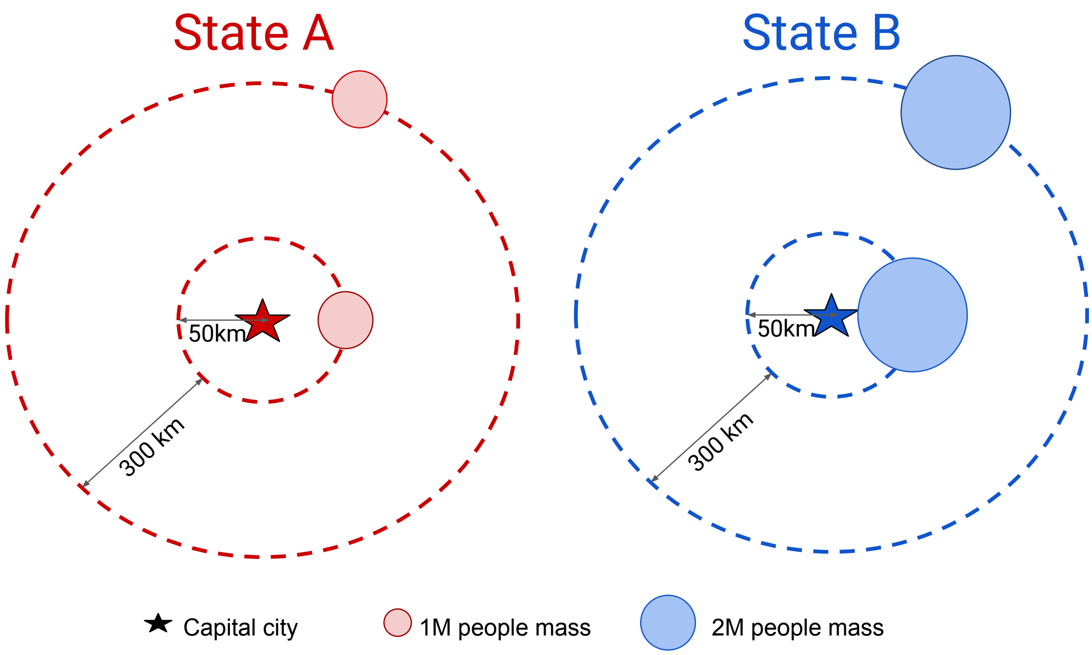
Corruption: I use the measure of corruption built by Glaeser and Saks (2006) and also used by Campante and Do (2014), the average number of federal convictions for public corruption per capita in the state from the 1989, 1999 and 2002 issues of the Report to Congress on the activities and operations of the Public Integrity Section, issued by the Department of Justice.

Table 1.A3: Outcome variables, summary statistics

Variable	Mean	SD	Min	Max	N Obs
Bills introduced	0.182	0.194	0.028	1.287	48
Public good provision	0.660	0.063	0.507	0.766	48
Corruption	0.268	0.125	0.073	0.603	48

1.B Measures of isolation and population size

Figure 1.B1: GCISC based isolation and population size



Fictional states with 2 population masses at the same distance from the capital city in each state.

Figure 1.B1 shows two states of the same size, each with too masses of population at the same distance from the capital city. Let us assume for simplicity that the capital city is empty - no people leave there. Let us also assume that the maximum distance from the capital city is 350 km in each state. Both states have the same configuration: a mass of population i close to the capital city and one at the boundary. The size of the population masses is identical within the state, while in State B they are two times as large. Therefore each population mass represents a share $s_i = 0.5$ of total state population. Using the formula to compute Isolation from Equation 1.6:

$$
\begin{aligned}
Isolation_A &= 1 - \sum_i \left(1 - \frac{log(d_{Ai})}{log(d_{max})}\right) s_i = \\
&= 1 - \left(1 - \frac{log(50)}{log(350)}\right) \times 0.5 - \left(1 - \frac{log(300)}{log(350)}\right) \times 0.5 = \\
&= \frac{log(50)}{log(350)} \approx 0.67
\end{aligned}
$$

It is easy to recognise that the level of isolation for State B would be exactly the same, despite the population density around the capital city being double. Instead, if we compute population potential for State A would be:

$$PopPotential_A = \sum_i \frac{pop_i}{d_{Ai}} = \frac{1M}{50} + \frac{1M}{300} = K$$

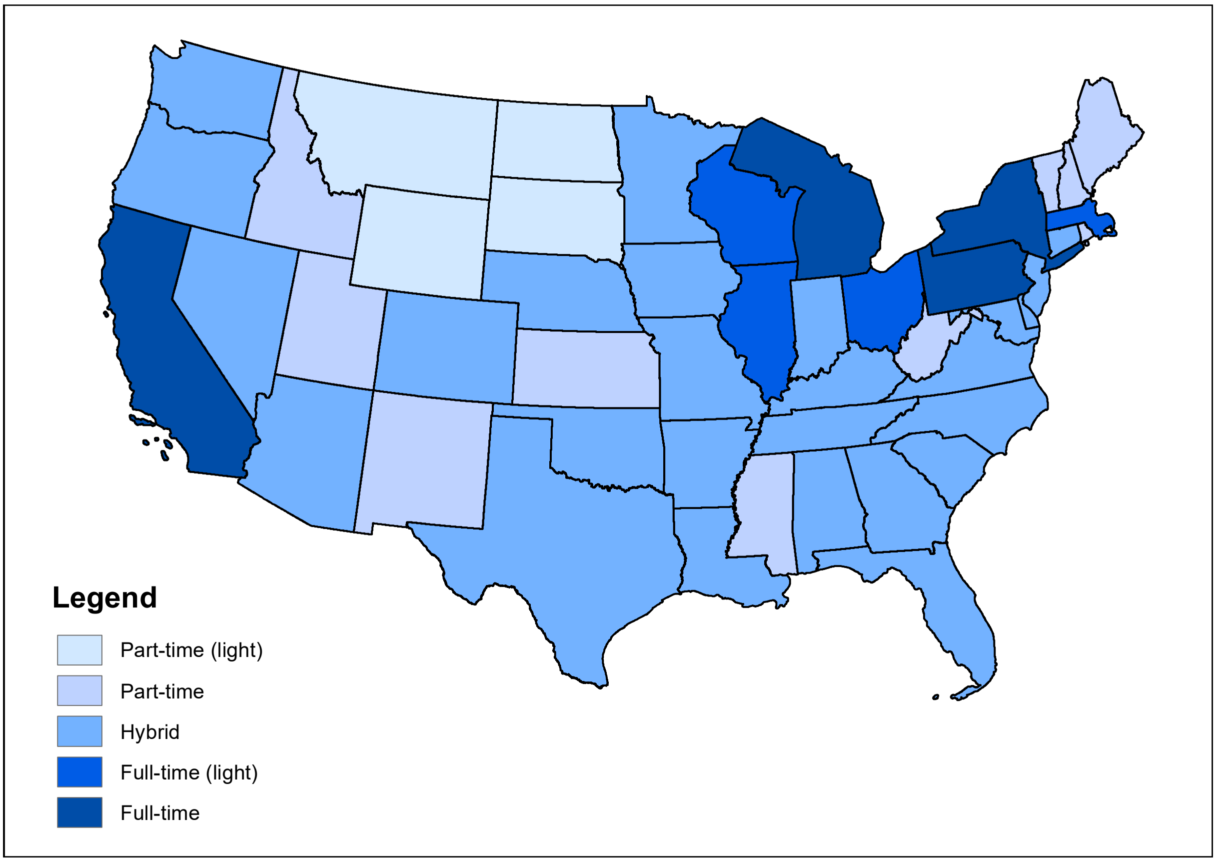
where $K := \frac{1M}{50} + \frac{1M}{300}$. Using the same formula for State B:

$$PopPotential_B = \sum_i \frac{pop_i}{d_{Bi}} = \frac{2M}{50} + \frac{2M}{300} = 2K$$

Population potential in State B is exactly double, reflecting the higher number of people living closer to the capital. The same applies for measures of population density in a 50 and 100 km radius around the capital, they would be exactly double for State B.

1.C Additional Figures

Figure 1.C1: US state legislatures' time on the job

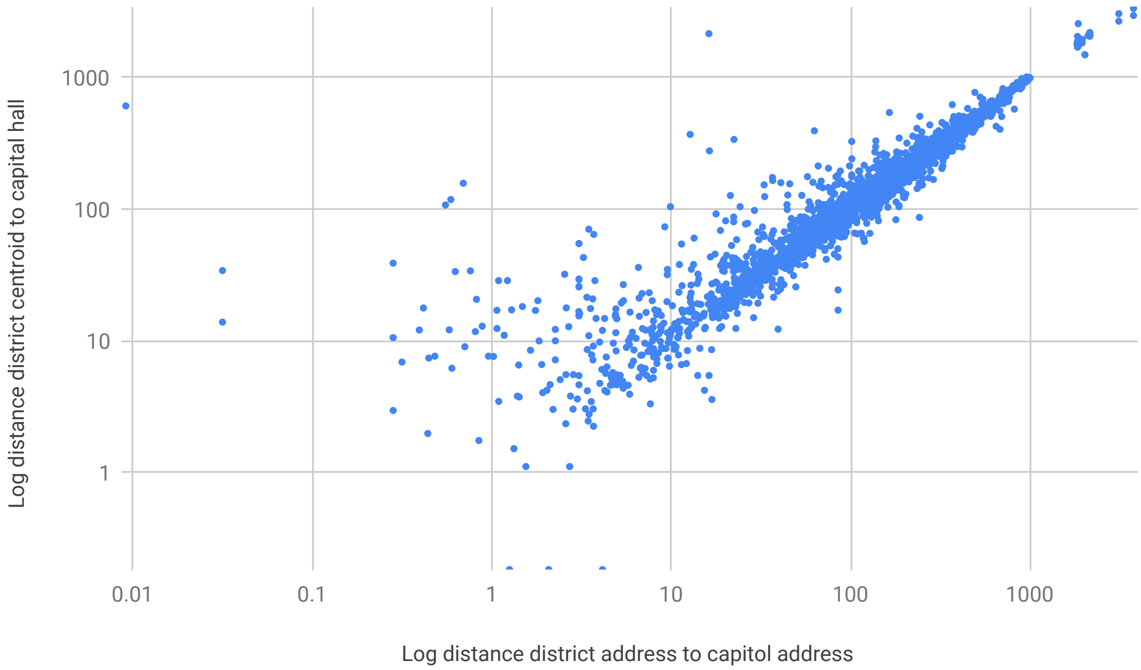


The 5-way classification ranking is taken from NCSL (2017). Legislatures are ranked in terms of the time spent on the job by legislators. The full-time light and part-time light legislatures differ from the full-time and part-time ones by having a lower remuneration and staff size. The estimated average time in 2014 on the job was 84% for the full-time legislatures, 74% for the hybrid and only 57% for the part-time ones. The average salary measured in 2014 US dollars, was respectively around 82k, 41k and 18k; the average size of staff 1250, 469 and 160 (NCSL, 2015b).

Figure 1.C2: US state legislatures' term limits

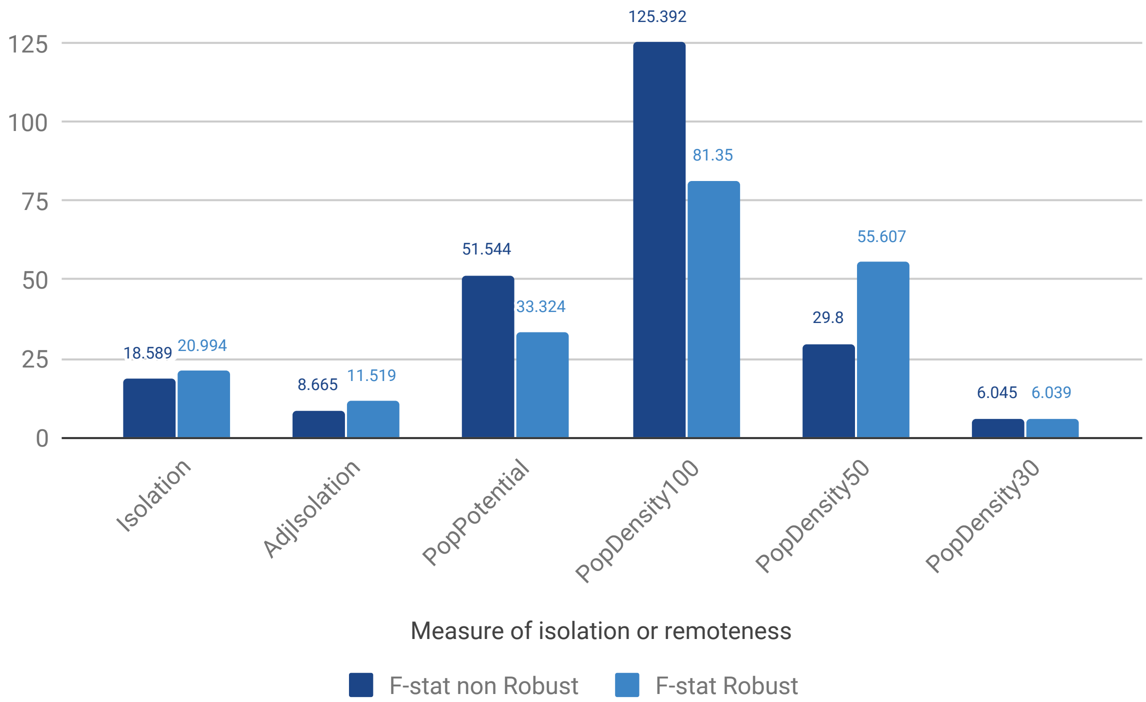
The information on term limits in different states is taken from NCSL (2015a). States with a lifetime ban do not allow running for the same position once the term limit has been hit. All the other states with term limits ban only consecutive legislative spells. Term limits repealed indicates states that had term limits which have later been repealed by the legislature or the State Supreme Court.

Figure 1.C3: Address distance vs district distance



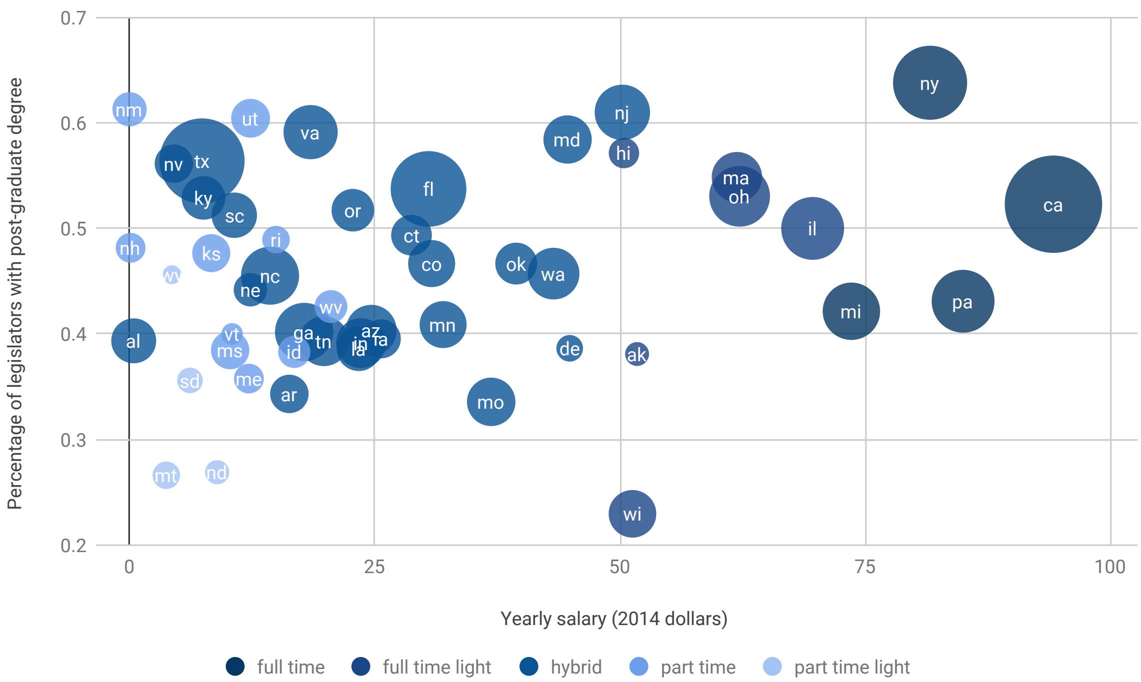
On the horizontal axis: distance between district address and capitol address as recorded in Vote Smart (2019). On the vertical axis: distance between the centroid of the district of election and the coordinates of the capitol building in the state capital.

Figure 1.C4: Relevance of centroid based instruments



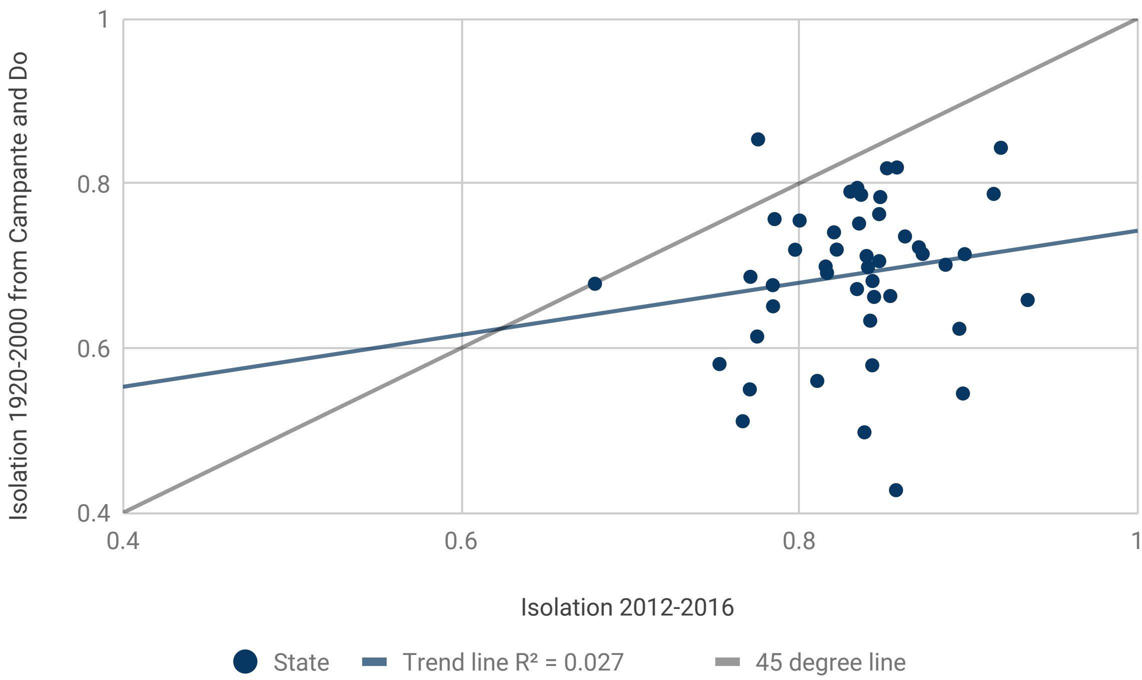
Left bar is Cragg and Donald (1993) Wald F statistics from 2SLS regression of average state legislators' educational attainment measured using ISCED classification (UNESCO Institute for Statistics, 2011) on capital city remoteness, state average educational attainment from Census Bureau ACS 2012-2016 average, logarithm of state land area and logarithm of maximum distance between the capital and any point in the state. Regressions for isolation and adjusted isolation also include log state population on the right-hand side. Capital city remoteness is instrumented using the equivalent measure of remoteness for the centroid of the state. The right bar is the robust Kleibergen and Paap (2006) Wald F statistics for the same specification. It also coincides with Olea and Pflueger (2013) efficient F statistics. The 10% Stock and Yogo (2005) critical test value is 16.38, the 15% critical value is 8.96. The critical values for the Olea and Pflueger (2013) efficient F statistics are 23.109 for 10% of worst case bias and 15.062 for 20%.

Figure 1.C5: Percentage with postgraduate degree by legislature



On the horizontal axis: average 2010-2014 legislators' yearly salary in 2014 US dollars from Bowen and Greene (2014a). On the vertical axis: the percentage of legislators' with a postgraduate degree (Master's or above). The size of the bubble represents the US Census Bureau 2017 population estimates of the state from U.S. Census Bureau, Population Division (2017). States are ranked using 5 categories in terms of legislators' time on the job from NCSL (2015b).

Figure 1.C6: Isolation evolution over time



Scatter plot of unadjusted isolation measured as 1.6 using the maximum distance in contiguous US states for normalisation and population data from ACS estimates 2012-2016 vs the unadjusted isolation measure from Campante and Do (2014) based on 1920-2000 US Census population data.

1.D Additional Tables

Table 1.D1: Legislators and district remoteness, robustness

	Dependent variable:					
	ISCED (1)	Postgrad (2)	Edu quality (3)	Law (4)	Experience (5)	Years Politics (6)
Panel A: party						
Log distance	-0.00797	-0.0225	-0.575*	-0.0355	-0.0205	-0.380***
	(0.011)	(0.019)	(0.32)	(0.022)	(0.017)	(0.14)
Observations	6063	6063	5720	6069	7205	7205
Panel B: birth						
Log distance	-0.0386***	-0.0699***	-0.690	-0.0574*	-0.0224	-0.478**
	(0.014)	(0.025)	(0.43)	(0.029)	(0.024)	(0.19)
Observations	3365	3365	3165	3370	3529	3529
Panel C: grad						
Log distance	-0.0232**	-0.0509***	-0.809***	-0.0511**	-0.0150	-0.411***
	(0.011)	(0.018)	(0.31)	(0.022)	(0.017)	(0.14)
Observations	6063	6063	5720	6069	7205	7205
State FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS with dependent variable=educational attainment measured on ISCED scale (1), Probit with dep. var. dummy=1 if legislator has obtained a Master's degree (2), OLS with dep. var.=legislators' best graduate or postgraduate institution score (3), Probit with dep. var. dummy=1 if legislator has obtained postgraduate education in law (4), OLS with dependent variable=number of different public sector and political appointments excluding those as state legislator (5), OLS with dep. var.=years from the first appointment in the public sector or as elected official (6). Log distance is the logarithm of the distance in km between the centroid of the district of election and the legislative hall. Panel A includes a dummy for being a Democrat, Panel B a dummy for being born in the district of election. *Controls* include the average educational attainment in the district of election measured on ISCED scale, the logarithm of the 2012-16 average district income per capita measured in 2016 inflation adjusted dollars, population density in the district of election and a dichotomous indicator for being a state senator. In Panel C instead of educational attainment in the district is replaced by the share of district's population above 25 with at least a bachelor's degree. All models include state fixed effects.

Table 1.D2: Education and capital city remoteness, other variables

	Dependent variable:					
	Edu politics (1)	Top 20 (2)	Top 50 (3)	Uni rank (4)	Graduate (5)	Associate (6)
Panel A						
Density 100km	1.842***	0.0360	0.428	-2963.5	1.029***	0.450
	(0.48)	(0.38)	(0.47)	(2352.0)	(0.37)	(0.27)
Panel B						
Pop potential	1.101	-0.472	-0.587	-6014.2*	1.431**	0.839*
	(0.81)	(0.49)	(0.72)	(3094.0)	(0.60)	(0.43)
All controls	✓	✓	✓	✓	✓	✓
Observations	47	47	47	47	47	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is the share of legislators with a graduate or postgraduate degree in public policy/administration or politics (1), the share of legislators with a postgraduate degree from a top 20 university according to the Times Higher Education (2019a), THE, ranking (2), the share of legislators with a postgraduate degree from a top 50 university according to the THE ranking (3), the average legislators' graduate of postgraduate academic institution rank from the THE ranking (4), the share of legislators with a Bachelor's degree or higher (5), the share of legislators with an Associate's degree or higher (6). In Panel A density indicates the population density around the state capital in a 100 km radius (measured in 10,000 people per km^2). In Panel B population potential is computed as in Equation 1.5 (measured in 10,000 people per km). All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). Montana is excluded as there is no political newspaper coverage data.

Table 1.D3: Education, experience and capital city remoteness, robustness

	Dependent variable:			
	Edu attainm v2	Edu attainm v3	Quality v2	Experience v2
	(1)	(2)	(3)	(4)
Panel A				
Density 100km	5.932***	5.289***	171.7*	3.586***
	(1.17)	(1.20)	(94.3)	(0.85)
Panel B				
Pop potential	4.851*	3.813	322.2**	3.581***
	(2.48)	(2.32)	(142.3)	(1.19)
All controls	✓	✓	✓	✓
Observations	47	47	47	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on ISCED 1-8 scale assuming the lower level whenever the degree is not reported (1), average state legislators' educational attainment measured on ISCED 1-8 scale excluding legislators whose degree is not reported (2), the average legislators' highest academic degree institution score from the Times Higher Education (2019a) ranking (3), the number of different political appointments in local government, excluding state level (4). In Panel A density indicates the population density around the state capital in a 100 km radius (measured in 10,000 people per km^2). In Panel B population potential is computed as in Equation 1.5 (measured in 10,000 people per km). All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). Montana is excluded as there is no political newspaper coverage data.

Table 1.D4: Education and capital city remoteness, robustness

	Dependent variable:					
	Edu attainm			Postgrad		
	(1)	(2)	(3)	(4)	(5)	(6)
Density 100km	4.613***	5.339***	5.179***	2.648***	2.950***	2.554**
	(1.22)	(1.13)	(1.46)	(0.79)	(0.69)	(0.99)
All controls	✓	✓	✓	✓	✓	✓
Observations	47	47	46	47	47	46

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on ISCED 1-8 scale (1-3), the share of legislators with a postgraduate degree (4-6). Density indicates the population density around the state capita in a 100 km radius. All regressions include the following controls: the log of the maximum distance between the capital and a point in the state, the log of the state land area, the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). In columns 1 and 4 state education is measured as the share of the population over 25 in the state with a graduate degree, in columns 2-3 and 5-6 as the average level of education in the state of population over 25. In columns 2 and 5 I use a different measure of newspaper political coverage from Campante and Do (2014). In columns 3 and 6 I control for a large set of legislatures' characteristics: existence of term limits and lifetime term limits, time required to reside in district, state, country to run for office, requirement of state citizenship, average minimum age to run for office, average share of Democrats in the two houses. Montana is excluded as there is no political newspaper coverage data. In columns 3 and 6, when controlling for share of Democrats in the houses, Nebraska is also excluded because the state legislature is composed of a single non-partisan house.

Table 1.D5: Density vs isolation

	Dependent variable:							
	Edu attainm				Postgrad			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Density 100km	5.229***		5.230***		3.201***		3.205***	
	(1.01)		(1.00)		(0.52)		(0.51)	
Density 50km		3.883***		3.905***		1.882**		1.908**
		(1.16)		(1.16)		(0.91)		(0.91)
Isolation	-0.126	0.438			0.363	0.594		
	(0.72)	(0.74)			(0.40)	(0.44)		
Isolation (adj.)			-0.120	0.435			0.353	0.584
			(0.69)	(0.71)			(0.39)	(0.42)
All controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	47	47	47	47	47	47	47	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is average state legislators' educational attainment measured on ISCED 1-8 scale (1-4), the share of legislators with a postgraduate degree (5-8). Density indicates the population density around the state capital respectively in a 50 and 100 km radius (measured in 10,000 people per km^2). Isolation is measured on a 0-1 scale following Equation 1.6. Adjusted isolation is measured on a 0-1 scale following Equation 1.7. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, the real salary, Squire (2017) index of legislature's professionalism, newspaper political coverage from Campante and Do (2014), voter turnout from in the last 3 presidential elections from CSG (2018). Montana is excluded as there is no political newspaper coverage data.

Table 1.D6: Remoteness, type of education and legislative outcomes

	Dependent variable:					
	Bills initiated		Public good		Corruption	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Density 100km	13.15***	14.73***	0.311	-0.151	-1.232	0.584
	(3.89)	(4.19)	(0.69)	(0.60)	(1.60)	(1.63)
Law	0.807**		-0.0144		0.472*	
	(0.38)		(0.12)		(0.26)	
Edu Politics		0.648		0.586***		-0.851*
		(0.65)		(0.20)		(0.44)
Experience	-0.0451	-0.0839	0.0653**	0.0434	-0.0925*	-0.0688
	(0.050)	(0.069)	(0.030)	(0.035)	(0.048)	(0.051)
Panel B						
Isolation	0.319	0.384	-0.629**	-0.599**	1.285**	1.242**
	(0.60)	(0.66)	(0.27)	(0.23)	(0.53)	(0.52)
Law	1.351**		-0.00163		0.421*	
	(0.64)		(0.11)		(0.22)	
Edu Politics		1.215*		0.558***		-0.783*
		(0.70)		(0.18)		(0.39)
Experience	0.0747	0.0287	0.0683**	0.0434	-0.104**	-0.0667
	(0.11)	(0.12)	(0.030)	(0.035)	(0.039)	(0.046)
Controls	✓	✓	✓	✓	✓	✓
Observations	48	48	48	48	48	48

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS regressions where the dependent variable is bills per legislator per day of session (1-2), public good expenditure from Campante and Do (2014) (3-4), Glaeser and Saks (2006) measure of state level corruption (5-6). In Panel A remoteness is measured as population density around the state capital respectively in a 100 km radius (measured in 10,000 people per km^2). In Panel B remoteness is measured as isolation on a 0-1 scale from Equation 1.6. *Law* is the share of legislators with a postgraduate degree in law. *Edu politics* is the share of legislators with a graduate or postgraduate degree in public policy/administration or politics. All regressions include as controls the average level of education in the state, the log of the maximum distance between the capital and a point in the state, the log of the state land area, real salary, Squire (2017) index of legislature's professionalism, log of state income, voter turnout from in the last 3 presidential elections from CSG (2018), census region dummies. Panel B also includes the log of state populations in all regression.

Table 1.D7: Lawyer legislators' characteristics

	Law=0		Law=1		
	mean	n. obs.	mean	n. obs.	diff.
Edu attainm	6.251	4952	7.009	1111	0.758***
Edu quality	35.562	4609	45.430	1111	9.869***
Experience	1.79	4957	1.814	1111	0.017
District local	0.498	2732	0.522	638	0.023
Exp leader	0.187	4957	0.221	1111	0.035**
Party official	0.032	4957	0.028	1111	-0.005

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$***. Two-sample two-sided t-test comparing legislators with and without a postgraduate law degree OLS regressions. *Edu attainm:* educational attainment measured on ISCED 1-8 scale, *Edu quality:* the best graduate or postgraduate institution score, *Experience:* the number of different public sector and political appointments excluding those as state legislator, *District local:* dummy=1 if the legislator is born in the same district where they have been elected, *Exp leader:* dummy=1 if the legislator has held state legislative leadership positions in the past (president, speaker, majority or minority leader and whip), *Party official:* dummy=1 if the legislator has held some leadership or coordination role in their party.

Chapter 2

The role of location as a determinant of judicial performance

2.1 Introduction

Since Henderson (1974), a vast literature has focused on agglomeration forces as the driver of higher wages in larger cities. High-skilled workers benefit the most from productivity spillovers and move to large cities to enjoy a higher wage premium.[1] However, there is a different mechanism based on preferences. If high-earning high-skilled agents have a stronger taste for consumption variety,[2] they may be drawn to large cities by the abundance of consumption choices and urban amenities (Glaeser, Kolko, and Saiz, 2001). The question that arises is: are high-skilled workers drawn to larger cities only because of higher wage prospects or also by urban amenities and opportunities of consumption? As large cities offer both, it is difficult to separate the two mechanisms.

In this paper I employ data from the UK Ministry of Justice (MoJ) on the Crown Courts,[3] to show that criminal judges accept a negative wage premium to move to locations with a higher level of amenities and consumption choices. I use the spatial differentials in real salary, generated by a predetermined rigid wage structure, as a source of

[1] There is a wealth of sorting models that account for heterogeneous agents. See, for example, Davis and Dingel (2019); Behrens, Duranton, and Robert-Nicoud (2014); Eeckhout, Pinheiro, and Schmidheiny (2014). A different mechanism for agglomeration benefiting high-skilled workers is human capital externalities (Moretti, 2004).

[2] This is true, for example, if preferences are non-homotethic, as shown by Fajgelbaum, Grossman, and Helpman (2011); Handbury and Weinstein (2015).

[3] The Crown Court is the highest criminal court of first instance in the UK.

exogenous variation. To proxy judges' ability, I build a measure of performance based on how often sentences are overturned or appealed. I find the existence of a negative relationship between real wage and judges' performance, while measures of urban amenities positively predict performance. This constitutes evidence of supply-driven agglomeration. I also find that London courts tend to perform better than those in the rest of the country, while the performance improvement is half as much for the other metropolitan areas.

There are a number of crucial advantages in focusing on the population of criminal judges. First, the limited size of the criminal sentences industry allows for a partial equilibrium framework, where land cost and goods prices are exogenous relative to the location choices of judges. Second, the nationwide regulated wage generates an exogenous variation in real wages that can be used as a source of identification. Third, justice is a non-tradable good and courts are evenly spread across the country. Fourth, the tasks faced by criminal courts' judges are the same everywhere. Finally, there is little room for cases' selection by judges, prosecutors and defendants alike.

To guide the empirical investigation, I use the theoretical framework from Lee (2010) modified to account for the inelastic demand for judges and the regulated remuneration. I show that judges face a trade-off when choosing their preferred location: larger market areas offer more consumption choices but lower real wages due to a higher cost of housing. Judges' decision whether to apply for a job in a specific location ultimately depends on their outside option, which increases with ability. Therefore, better locations will attract more talented candidates. A crucial assumption, based on empirical evidence (Handbury, 2019),[4] is that high-earners, such as criminal court judges, have a stronger taste for different consumption varieties.

I use Ministry of Justice court-level data on the success and frequency of appeals against criminal judges' sentences to build hard performance measures that proxy ability. I show that there is a strongly significant negative relationship between the real wages and the performance measures employed. Talented judges accept a negative wage premium to live in areas offering more consumption choices and amenities. I also build proxies for market access and urban amenities based on population density and the presence of

[4] Higher earners also prefer higher quality goods that are mostly found in larger cities (Bils and Klenow, 2001; Berry and Waldfogel, 2010).

recreational activities such as restaurants and museums, and find that they positively correlate to criminal courts' performance. I then conduct a placebo test using a measure of performance independent from judges' behaviour and find that the effect disappears.

Finally, I employ an instrumental variable analysis using a geographical instrument, namely the distance from Charing Cross, the notional centre of London.[5] The estimates are consistent with the previous findings. All the evidence gathered supports the consumer-city, supply-driven sorting of high-skilled workers. The results also seem to suggest that judicial quality is higher in London compared to the rest of the country.

In addition to providing insights into courts' performance differentials and the spatial sorting of judges, another contribution of this paper is introducing spatial variation as a determinant of institutional performance. The role of institutions as a driver of economic growth and prosperity (North, 1991; Acemoglu et al., 2005) has long been established in Economics. The judiciary, in particular, is a core institution for the well functioning of a state as the legal system is necessary to sustain state capacity (Besley and Persson, 2009) and property rights enforcement.[6] One of the key roles of courts and judges is to provide the independent arbitration required for property rights protection. Hence, differential judicial performance across regions in the UK would imply unequal representation and rights protections for people living in different areas.

Related literature: First and foremost, this paper contributes to a vast literature on spatial sorting[7] and, in particular, supply-side agglomeration (Tabuchi and Yoshida, 2000; Glaeser et al., 2001). It is particularly close to Lee (2010), who shows the existence of a negative wage premium for the highest skilled workers in the health care sector in the US. Just like criminal judges, health-care workers produce a non-tradable good and are evenly spread across the country.[8] This paper differs in two crucial ways: through the presence of a rigid national pay structure that provides a quasi-experimental design; and through the use of an objective performance measure to proxy ability. The fixed wage,

[5] Charing Cross is a main junction in the City of Westminster, in west-central London, that is used to mark the center of London. It is not the geographical centre but a notional one based on the original site of a memorial cross for Queen Eleanor erected after her sudden death in 1290 (Lovell, 1892).

[6] Property rights' enforcement has been shown to be a key determinant of development, see Besley and Ghatak (2010).

[7] See Behrens and Robert-Nicoud (2015) and Combes and Gobillon (2015) for a comprehensive review.

[8] Lee (2010) also finds that lawyers, another category of high skilled professionals, have high urban wage premia, increasing in the skill level. This is because, unlike for health-care professionals, lawyers' demand varies across cities and less lawyers are needed in smaller cities.

in particular, simplifies the exercise of disentangling the supply-side from demand-side agglomeration effect by eliminating the demand effect.

Second, this paper relates to the research focused on the effects of exogenous variation in real wages because of rigid pay structures that generate a wedge between inside and outside wages (Cappelli and Chauvin, 1991; Propper and Van Reenen, 2010). This literature documents that the estimated impact of pay regulation is lower quality and productivity when in presence of a stronger external labour market. Lower-skilled agents that don't have the same preference for consumption varieties will choose to locate where the real wage is higher, generating a shortage of workers in areas with better outside options and higher prices. This paper, instead, focuses on wage regulation for a category of high-skilled, high earning professionals such as judges and show that the opposite is true. Large cities have a stronger external labour market as well as more consumption varieties and urban amenities, thus attracting the highest skilled workers.

This paper also contributes to the Economics literature studying the impact of institutional design on judiciary outcomes. It is one of few works focused on the UK judiciary alongside Malleson (2006); Blanes i Vidal and Leaver (2011, 2015) and provides insights into the workings of Crown Courts. It differs from other works by proposing a new determinant of judiciary performance: location, while the core body of the literature has instead focused on judges' accountability (Maskin and Tirole, 2004; Malleson, 2006; Blanes i Vidal and Leaver, 2015).[9]

Finally, this paper can also be related to works focusing on the mobility of top earners (Young and Varner, 2011; Kleven et al., 2013, 2014; Moretti and Wilson, 2017).[10] Instead of relying on variation in tax rates, it focuses on the rigidity of the centralised wage setting as a source of exogenous variation.

The rest of the paper is structured as follows: Section 2.2 provides some background on the functioning of UK courts and, in particular, Crown Courts. Section 2.3 presents a simple theoretical model describing location choices in the presence of an exogenous fixed

[9] In particular, a small empirical literature has focused the variation across US states in appointment rules (direct appointments, partisan and non-partisan elections) to investigate the effect of accountability on judiciary outcomes. See, for example, Tabarrok and Helland (1999); Besley and Payne (2003); Huber and Gordon (2004); Gordon and Huber (2007); Berdejó and Yuchtman (2013); Ash and MacLeod (2016).

[10] Young and Varner (2011); Kleven et al. (2014) look at all *high-earners* while Kleven et al. (2013); Moretti and Wilson (2017) focus on a specific category, respectively footballers playing in Europe and top scientists.

nominal wage and preferences for consumption variety. Section 2.4 describes the data-set. Section 2.5 presents the empirical strategy. Section 2.6 discusses the results. Section 2.7 concludes.

2.2 Institutional Background

In this section I provide an overview of the organisation of the Courts and Tribunals system in the UK, the role of Crown Courts and the criminal judges' appointment process.

2.2.1 The Courts and Tribunals system

The current structure of UK justice is the result of centuries of evolution that resulted in a complicated system involving different courts and tribunals (Judiciary UK, 2018b). It is, indeed, referred to as a *courts and tribunals system* rather than a *courts system*.[11] The court structure covers England and Wales while the tribunals system covers England, Wales, and in some cases Northern Ireland and Scotland. Tribunal proceedings are generally more informal compared to the courts and people often appear without legal representation. Similar to the tribunals system, different cases are dealt within specific courts across the court system:[12]

- Civil cases are either heard in the Magistrates' Court or in a County Court. Appeals go to the High Court and then to the Court of Appeal.

- Criminal cases always start in the Magistrates' Court, but the ones involving serious offences are sent to the Crown Court. Magistrates' Court appeals are dealt with in the Crown Court. Appeals from the Crown Court go to the the Court of Appeal.[13]

This paper focuses on the Crown Court because it fulfils three requirements that are fundamental to the viability of the empirical analysis: 1) the sample is large enough

[11]Figure 2.B1 shows a simplified version of the routs taken by different cases in the tribunals system. For a more detailed representation see the tribunal Structure Chart at Judiciary UK (2018b).

[12]Figure 2.B2 shows the routes taken by different cases as they move through the courts system, and the judges who deal with each.

[13]However, a smaller fraction of criminal cases' appeals, only on points of law, go to the High Court, Queen's Bench division. The tribunals system, instead, has its own internal appeals route, as shown in Figure 2.B1. However, decisions from the Upper Tribunal and the Employment Appeals Tribunal may also go to the Court of Appeal.

to confer adequate power and there is enough variation in the location of courts over the country; 2) the population of judges sitting in the Crown Court is homogeneous and comprises highly-educated professionals with large salaries; 3) the tasks performed in a Crown Court and the type of cases dealt with are fairly standardised compared to other courts. Moreover, there is an objective measure of judicial performance based on the success rate and frequency of appeals against Crown Courts' decisions. Clearly, the Higher Courts (High Court, Court of Appeal and Supreme Court) do not respect the first requirement, as they employ a small number of judges and operate as single courts. The tribunals, on the other hand, because of their complex structure and high level of specialisation, do not fully respect the fourth criterion. Also, tribunal judges are often paid on fee basis and are assisted by non-legally trained members, subject matter experts such as physicians or engineers, in solving disputes, which means the second condition is partially unfulfilled as well. The same holds true for the Magistrates' Courts, where the large majority of cases[14] is presided by a group of three magistrates (or Justices of the Peace), members of the local community who volunteer their services. Magistrates receive training, and are advised in court on matters of law practice and procedure, but do not have formal legal qualifications. Finally, the County Court fares better than all the aforementioned courts in all dimensions. The majority of trials is dealt with by district judges. District judges are full-time judges who handle most cases in the County Courts. Their work involves civil and family cases ranging from landlord damage claims, injunctions and insolvencies to divorces and domestic violence. The selection process of district judges is similar to that of circuit judges, who sit in the Crown Courts, which I will discuss later. District judges as a category fit well the requirements of the ideal population for this study.

There are two reasons, however, why I focus on Crown Courts rather than County Courts. First of all, compared to the Crown Court, the types of sitting judges are more mixed[15]. District judges are indeed those with the largest number of sitting days in County Courts, but in 2016 more than 50% of time was served by other types of judges. Among these there are deputy district judges, whose qualifications are the same of district judges

[14]The roughly 330 Magistrates' Courts in England and Wales are presided by approximately 16,000 magistrates, 140 district judges and 170 deputy district judges.

[15]See Table 2.C1

but are paid on fee basis. Also, other types of judges, including very different professional figures such as lay assessors, legal advisors and employment judges, oversee a sizeable amount of cases. Given the unavailability of trial data at individual level, the excessive variation in the type of judges at court level would hinder the reliability of the analysis. Second, despite the existence of different types of crimes, there is much larger variation in the task faced by civil judges as they have to deal with very different areas of dispute such as civil, family matters, bankruptcy and insolvency (Judiciary UK, 2018a).

2.2.2 The Crown Court

The Crown Court is part of the Senior Courts of England and Wales[16] and represents the highest court of first instance for criminal cases. Unlike Magistrates' Courts, it acts as a single entity sitting in 77 centres across the country. The Crown Court deals only with the most serious criminal offences which constitute around 2-3% of all criminal cases. It also deals with sentencing of offenders convicted in the Magistrates' Court whenever the magistrates find their maximum sentencing power (six months imprisonment) is insufficient and it hears appeals against sentences and convictions by magistrates. The Crown Court is administratively divided in six circuits overlapping with the old Assize Court, which were replaced by the Crown Court through the Courts Act (1971), and the Bar regions: Midland, Northern, North Eastern, South Eastern, Wales and Chester, and Western.

The most serious cases in Crown Courts are heard by a High Court judge, while the great majority of cases are dealt with by a circuit judge or a recorder. In case of a "not guilty" plea, a trial takes place with a judge and a jury of 12 randomly selected members of the public registered on the electoral roll. The jury, after considering all evidence, is ultimately responsible for the conviction. The judge, in case of a conviction, will pass sentence.[17] However, the judge is not only responsible for sentencing but also for dealing with all the procedural and technical aspects of the trial, as well as pre-trial (Gibson and Cavadino, 2008). Circuit judges (henceforth CJs) account for the bulk of the days sat by any judge in the Crown Court, for example 73.5% in 2016.[18] Once we add recorders to the

[16]Senior Courts Act (1981).

[17]In determining the sentence, the judge will weigh a *pre-sentence report* and, if applicable, medical and psychiatric assessments of the defendant (Gibson and Cavadino, 2008).

[18]See Table 2.C1

figure, the above percentage increases to 96.3%. Recorders are circuit judges *in pectore*. Their jurisdiction is largely overlapping to that of CJs, but they usually deal with less complex cases. They also receive training in the form of courses and by sitting together with a CJ. They can count training days as *sitting days*, which means that 73.5% is an underestimate of the caseload dealt with by CJs in 2016.[19]

CJs sitting in the Crown Court are the population of interest in this paper. They are high-skilled, educated professionals. The prerequisite to be a circuit judge is to have held the *right of audience* for at least 10 years.[20] Moreover, a newly appointed circuit judge has often had to work on criminal cases as a recorder or on civil cases as a district judge before appointment. There are over 600 CJs - precisely 635 in April 2017, 591 of whom employed full-time. The population of CJs is quite homogeneous: the typical CJ is a white male over 50 who was a barrister before becoming a judge.[21] Their tasks, as per the official job advert, are relatively standardised. Their responsibilities involve exercising control over the proceedings, the pre-trial stage and presiding over the trial. CJs deal with a broad range of criminal cases in the Crown Court and are required to balance the right of the defendant to a fair trial with the interests of the victims and witnesses. They also have to maintain order in the courtroom, and ensure the jury understand the case. Finally, the judges have to pass sentence when the defendant pleads guilty or is convicted by the jury.[22]

2.2.3 Circuit judges' appointment

CJs are formally appointed by the Queen based on the recommendation of the Lord Chancellor. However, the appointment choices are made by the Judicial Appointment Commission (JAC) following a *fair and open* competition. Job adverts are posted on the JAC's website and involve different vacancies in different locations and jurisdictions, for example, family, civil or crime. An example of a job advert from the JAC website can be seen in Figure 2.1. Applicants have to indicate their preference of location and can apply

[19]There were 1035 recorders in post in 2016 as per the Judicial Diversity Statistics 2016, Table 1.1. That is consistent with the requirement of Records to sit for 30 days a year (Judiciary UK, 2018a).

[20]A right of audience is the right to appear before and address the court as an advocate.

[21]Figure 2.B3 shows the composition of circuit judges with data from the Judicial Diversity Statistics (Judiciary, 2017).

[22]Judicial Appointments Commission (2017).

to more than one jurisdiction, but need to provide a statement indicating their suitability for each post. The whole process spans over several months and involves: the application with a self-assessment and a suitability statement; an independent assessment; a second self-assessment; and a situational questioning and interviews in a selection day held in London. The selection is based uniquely on merit and candidates are offered only one post.[23]

CJs are officially deployed by the Lord Chief Justice[24] but the allocation to a specific court is decided by senior leadership judges. CJs appointed in the crime jurisdiction are assigned a base court, while those in the civil jurisdiction, in some cases, may be asked to sit across two different County Courts. In other cases, due to business needs, the Presiding Judge may ask CJs to temporarily sit at a different Crown Court (Ministry of Justice, 2018b).[25]

2.3 Theoretical Framework

In this section I present a simple spatial sorting model with an inelastic demand for judges and a fixed nominal wage. The model is inspired by Lee (2010). The objective of the model is to highlight the trade-off between real wage and higher availability of consumption varieties faced by judges.

2.3.1 Judiciary

The demand for judges is inelastic as there is a fixed number of vacancies to be filled in each court. Candidates to judgeship differ in ability t. Ability is better interpreted as innate talent or *experience* rather than level of education since all judges need to meet the

[23]However, the JAC has an equal merit provision, which means that if the Commission considers two or more candidates of equal merit, they may choose a specific candidate who meets the objective of increasing diversity (Judicial Appointments Commission, 2017).

[24]The Head of the Judiciary of England and Wales and the President of the Courts of England and Wales. As of 2018 the role is occupied by The Right Honourable The Lord Burnett of Maldon. Before the Judicial Reform Act in 2005 such functions were held by the Lord Chancellor (Judiciary UK, 2018a).

[25]This phenomenon, if frequent, would constitute an issue as court level statistics would not be representative of the judges sitting in a specific court. Luckily, the Ministry of Justice (2018c) claims this does not happen very often. It is impossible to know the exact incidence as the MoJ does not keep a record due to the absence of legal or business requirement to do so (Ministry of Justice, 2018c).

same minimum requirements in terms of educational attainment.[26] Given that I assume t to be observable, I prefer the experience interpretation.[27] The hiring process is modelled as follows:

1. vacancies are posted;
2. candidates apply stating their preferences;
3. the JAC observes t and ranks candidates depending on their ability;
4. vacancies are filled with candidates starting with the one with the highest ability.

As a result, the distribution of talent depends only on candidates' applications, hence on the judge's ability sorting.

2.3.2 Judges

In a framework *a la* Roback (1982) and Glaeser and Gottlieb (2009), the utility a judge j derives from locating in city c depends on: the (rigid nationally regulated) wage w, the city-specific average rental price r and city-specific level of amenities a:

$$U_c = u(w, r_c, a_c) \tag{2.1}$$

I assume that there is a continuum of cities that differ only in population size n. The exogenous[28] rental price can be rewritten as $r(n)$. Let the amenity value of a city represent the level of consumption variety that, just as rents, depends on city size $a = \nu(n)$.[29] We can think about that, for example, as the number of museums, cinemas, parks, sport venues, restaurants. The total number of varieties and rents are increasing in city size:

[26]That consists in a undergraduate LLB, a Graduate Diploma in Law (GDL), followed by the Legal Practice Course (LPC) or Bar Professional Training Course (BPTC) (Sexton, 2018).

[27]Other options are publications, success record as a lawyer, reputation in the legal word.

[28]The total number of judges is a tiny fraction of the total population. It is therefore safe to assume that their location choice has no general equilibrium effects on rents.

[29]A key assumption is that high-skilled workers have different preferences compared to the lower-skilled counterparts. Such preference profiles can be obtained using a non-homotethic utility function. In Fajgelbaum, Grossman, and Helpman (2011) non-homotethicity is modelled as a stronger taste for differentiated goods by high earners. For simplicity, I follow Lee (2010) in considering the taste for variety as a luxury good.

$\nu'(n) > 0$ and $r'(n) > 0$. As a result (2.1) can be rewritten as:

$$U(n) = u(w, r(n), \nu(n)) \tag{2.2}$$

$U(n)$ is the indirect utility a judge receives from locating in a city of size n. It depends only on city size because the wage is exogenous.

If a judge chooses to locate in city n, they will rent one unit of housing at price $r(n)$ and consume a bundle of varieties $C(x)$ where $x \in [0, \nu(n)]$.[30] The judges' preferences take the following CES form:

$$\left(\int_0^{\nu(n)} C(x)^{\gamma} dx \right)^{\frac{1}{\gamma}} \tag{2.3}$$

where $0 < \gamma < 1$ generates the *taste for variety.* There is an infinite supply of identical judges who differ only in their ability t. The timing is as follows:

1. a candidate chooses the preferred cities where to work;
2. the JAC makes an offer to take a position in city n;
3. the candidate decides whether to accept it or take an outside option $\mu(t)$, for example, keep working as a barrister;
4. if the offer is accepted, the judge moves to city n. They rent one unit of housing at $r(n)$ and use the residual income $w - r(n)$ to buy local varieties $\nu(n)$ of goods.

The indirect utility of a judge living in city of size n, keeping the price for all goods and locations constant and equal to the numeraire[31] is then:

$$\begin{aligned} U(n) = &\max_{C(x)} \left(\int_0^{\nu(n)} C(x)^{\gamma} dx \right)^{\frac{1}{\gamma}} \\ &\text{s.t.} \quad \left(\int_0^{\nu(n)} C(x) dx \right) = w - r(n) \end{aligned} \tag{2.4}$$

[30] Following Lee (2010), I impose inelastic demand for housing that implies homothetic preferences. This is a simplification; the model can be generalised to non-homothetic housing preferences.

[31] I assume equal prices across locations for simplicity. NEG models (Helpman, 1998; Ottaviano et al., 2002; Behrens et al., 2014) predict that tradable goods are cheaper in larger cities. This prediction is however at odds with empirical evidence (DuMond et al., 1999; Tabuchi, 2001). Handbury and Weinstein (2015) solve the puzzle by attributing higher CPI prices in larger cities to the combination of three factors: 1. variety bias; 2. heterogeneity bias; 3. non-traded land price component. They show that once rental prices and goods' variety are accounted for, the price of tradable goods slightly falls with city size.

The visible effect of wage regulation is that the indirect utility depends only on the city size and not on the ability parameter t. The latter, however, is still relevant to determine whether the judgeship position is accepted. Assuming that judges can perfectly move across the country, the outside option, $\mu(t)$, is an increasing function of the level of ability, $\mu'(t) > 0$. The total indirect utility function, that depends on both the level of ability and city size, is then:

$$V(n,t) = \max_{n}\{U(n), \mu(t)\} \tag{2.5}$$

where $U(n)$ is as defined above in (2.2).

The solution of the optimisation problem in (2.5) is straightforward, since all goods have the same price and the same utility is attached to each variety. In equilibrium a judge consumes the same quantity of each good, or $C(x) = C \ \ \forall x \in [0, \nu(n)]$. As a result, the indirect utility for a judge living in city n becomes:

$$U(n) = \nu(n)^{\frac{1-\gamma}{\gamma}}(w - r(n)) = \nu(n)^{\frac{1-\gamma}{\gamma}}\omega(n) \tag{2.6}$$

where $\omega(n)$ can be interpreted as the real wage, $\omega'(n) < 0$ since $r'(n) > 0$ and w is fixed.

There are two components of the indirect utility: the first one is the *love of variety* which is increasing in city size n as larger cities offer a greater range of consumption possibilities. The second component, the real wage, is decreasing in city size as the nominal wage is fixed and rental prices are higher in larger cities. As a result, the effect of city size on the preferences of a given candidate is ambiguous: they will prefer to locate in larger city if the utility gain from additional consumption varieties is larger than the negative income effect arising from a lower real wage.

2.3.3 Discussion

The model crucially departs from the standard notion of spatial equilibrium. In a canonical general equilibrium with homogeneous preferences for amenities, the marginal worker is indifferent between two cities of different sizes and level of amenities at the prevailing private sector real wages.[32] Judges' wages, however, are set in a non-equilibrium way, due to the rigid wage structure that does not compensate for the variation in amenities. Given that the private sector urban nominal wage premium is larger than that of criminal court judges, they would be worse off in relocating to a urban area with higher rental cost even after accounting for the higher level of amenities. In this context, however, the equilibrium result is not valid as following Lee (2010) I assume that preferences for amenities are heterogeneous and vary with skills levels: judges have a stronger taste for amenities compared to individuals further down on the skills' distribution.[33] Therefore, our population of interest faces a different trade-off between real wage and amenities compared to the marginal worker.

Proposition *A candidate accepts to work as a judge in city n only if their ability* $t \leq \bar{t}$ *where* $\bar{t} = \mu^{-1}(\nu(n)^{\frac{1-\gamma}{\gamma}}\omega(n))$.

Proof Refer to problem (2.5): a candidate prefers an offer to the outside option only if $U(n) \geq \mu(t)$, then the above Proposition just follows from (2.6).

The ability of the best candidate applying for the job will then depend on the real wage and amenities level in the location of the Crown Court. Given the hiring process described in 2.3.1, the Judiciary Appointment commission will receive applications for a court located in city n from candidates with ability: $t \in [\underline{t}, \tau(n)]$, where $\underline{t}$ is the minimum ability in the pool of potential judges and $\tau(n)$ is the endogenous level of ability of the best candidate applying for the job:

$$\tau(n) = \bar{t} = \mu^{-1}(\nu(n)^{\frac{1-\gamma}{\gamma}}\omega(n)) \tag{2.7}$$

The most talented candidate will be offered the post. Therefore, the ability of a judge hired in a court located in city n will depend on the love of variety and real wage components that vary as city size increases.

[32]In equilibrium, the real wage will be higher in a less attractive city to compensate for the scarcity of amenities.

[33]Lee (2010) shows that private sector wage premia are decreasing in skills and become negative at the top of the skills' distribution. This is consistent with the idea of high-skill workers having a stronger preference for city amenities compared to lower skilled counterparts.

2.4 Data

I built a novel data-set with court level indicators, administrative and spatial inputs for 74 Crown Courts in the UK and Wales spanning the years 2010 to 2016 for whom *trial outcomes* data is available.[34] The data was obtained from a number of judiciary publications, Freedom of Information requests and administrative data. The data sources are briefly mentioned below and discussed in more detail in the Appendix.

Judiciary performance: One of the main issues in the literature looking at judiciary performance is devising a reliable measure of performance. The most common example in the literature is the number of citations, the output of Supreme Court judges (Cross and Spriggs, 2010). I overcome the challenge by constructing a *hard* indicator of performance based on the number of appeals against decisions of the Crown Courts. The data on appeals has been provided by the UK Ministry of Justice in response to two Freedom of Information requests.[35] It includes court-level indicators of: number of appeals (allowed and dismissed) heard against Crown Court sentences from 1998 to 2016 and the number of appeals (allowed and dismissed) heard against Crown Court convictions from 1998 to 2016. The main performance measure I use is equal to 1 minus the ratio between the number of allowed sentence appeals over the number of *effective trials*.[36] If an appeal has been allowed, then the Crown Court sentence has been changed. Alternatively, I use the number of all sentence appeals against Crown Courts' rulings as a percentage of the number of effective trials. I also look at the percentage of allowed conviction appeals against Crown Courts' rulings to use in a placebo test. The conviction of a defendant is decided by the jury and, although the judge has the authority to overturn it if certain conditions apply, the occurrence is extremely rare. Figure 2.2 shows the distribution of

[34] Yearly trial outcomes for Crown Courts are reported at court level per quarter in the Criminal Court statistics (Ministry of Justice, 2017a).

[35] Ministry of Justice, Data Access & Compliance Unit (2017a,b).

[36] An *effective trial* is a trial where the jury is sworn in and a verdict is reached. A trial can alternatively be *ineffective* if it does not go ahead and needs to be re-listed; *cracked* if an outcome is reached without a trial when an acceptable plea is entered by the defendant or the prosecution produces no evidence; *vacated* if the trial is removed by the list before it starts (Ministry of Justice, 2017b). In the UK, defendants can appeal against sentences after pleading guilty. Despite that, focusing on effective trials still seems to be the better choice as guilty pleas are often the product of extensive *plea bargaining* and result in sizeable discounts on sentences (Baldwin and McConville, 1978) and, as such, much less likely to be appealed. However, in the robustness checks I control for the number of cases that go to trial, to account for the possibility that the number of guilty pleas varies systematically across the country.

Crown Courts over the country. We can notice that performance is, on average, higher in the London area and lower in the north of the country. Also, high performance courts further away from London are sometimes located outside large metropolitan areas.[37]

Real wage and amenity: The first key variable from the stylised model introduced in Section 2.3 is the real wage. I build a *naive* estimate of the real wage equal to the ratio between circuit judges' nominal wage and the cost of housing, averaged from 2012 to 2016. The wage data is derived from the official Ministry of Justice judicial salaries.[38] The cost of housing is computed as the average house price, from the UK House Price Index, in a 15 km radius around the court.[39]

My strategy for building a proxy for the amenities offered by a city is to compute the population density in a 30 and 50 km radius around the location of each Crown Court. The above variables are a measure of market access widely used in the literature studying agglomeration starting from Ciccone and Hall (1996).[40] The population data is obtained from the 2011 UK census at output area level.[41] For each Crown Court I generate a circle of 30 km radius centred in the location of the court and intersect it with the output area polygons as shown in Figure 2.B4 for Lincoln Crown Court. For each polygon i inside the 30 km radius around a Crown Court c, I compute the weighted population based on the amount of the total output area that intersects the circle around the Crown Court: $intersect_{ic}$. I then sum all the weighted population values of each polygon i in the circle around and divide that by the total area of the circle:

$$popdens_c = \frac{\sum_i^N pop_{ic} \times \frac{intersect_{ic}}{outputarea_{ic}}}{area_c} \tag{2.8}$$

I build two alternative measures of amenity using business density per m^2 derived

[37]OECD (2016) functional urban areas with population above 500,000.

[38]Ministry of Justice (2015, 2016). The nominal wage is adjusted upwards every year. See Table 2.1 for the prevailing salaries for circuit judges between 2012 and 2016.

[39]HM Land Registry (2018). See Appendix 2.4 for further details. The 15 km radius was chosen based on the average commute distance in the UK in 2015 (Le Vine, Polak, and Humphrey, 2017).

[40]Density has been shown to be correlated with measures of city amenity (Glaeser and Gottlieb, 2006; Rappaport, 2008). Also, this specific variable, density in a circle around a point of interest, is easy to compute and has the beneficial property of reducing the modifiable area unit problem: the size and shape of the spatial units of interest, for example census areas when computing population density, could introduce bias (Briant et al., 2010).

[41]ONS (2012). The output area is the smallest census area. The size goes from as little as 153 m^2 to 201 km^2, with a UK average, including Northern Ireland and Scotland, of 0.833 km^2.

from the UK Business Counts.[42] In particular, I focus on businesses related to leisure and retail consumption, computing averages for the same years the judiciary data is available, 2010-2016.[43] I build two such measures, a narrower one focusing purely on leisure and hospitality businesses,[44] the second also including health and education-related businesses as well as retail consumption ones. I employ the same strategy used to compute population density. The final amenity indicators are the first principal components of respectively two and four business density variables.

Figure 2.4 displays a positive correlation between the logarithm of the first proxy for amenities, density in a 30 km radius around the Crown Court, and the logarithm of the house price index in the relevant local area.[45] The courts located in or around London which, compared to the average, display a considerably higher level of amenities and cost of housing.

Court level controls: I also collect information on the number of cases received by a Crown Court.[46] Since judges do not only try cases but also oversee the work of the court and pass sentence for those cases (the majority) that never go to trial (Gibson and Cavadino, 2008), if a court is overloaded with work the judges could be more prone to making mistakes or less careful in overseeing court proceedings. These could result in a larger number of sentences with grounds for appeal. To control for this, I derive a measure of *caseload*: the number of cases received per judges.[47] Another factor that could skew the performance result is the type of cases received by a Crown Court. It is reasonable to expect that different types of criminal cases have different levels of difficulty and, potentially, a different likelihood to be appealed. If a court receives a disproportionate amount of such cases, its performance measured in terms of appeals will be negatively affected. Similarly, if a court overwhelmingly focuses on a specific offence category it may become

[42] Office for National Statistics (2018).

[43] Glaeser and Gottlieb (2006) claim that higher income and more educated individuals demand more museums, restaurants, concerts.

[44] It includes, for example, creative, arts, entertainment activities, libraries, archives, museums and other cultural activities.

[45] Figure 2.5 replicates the same exercise using a different proxy for amenity based on the first principal component of respectively two (hospitality, leisure) and four (health, hospitality, leisure and retail) business density variables.

[46] From the Ministry of Justice quarterly statistics Ministry of Justice (2017a).

[47] The number of judges in each Crown Court was obtained in response to a Freedom of Information request (Ministry of Justice, 2018c).

more specialised and its decisions will be successfully appealed less often. For this reason, I collect the percentage of received cases for each offence group to be able to control for the case-mix, from the Ministry of Justice (2018a). Figure 2.3 compares the case composition in London courts with the rest of the country. We can see that despite the existence of a statistically significant difference in the amount of cases received for some offence categories, the type of cases dealt with by individual courts across the country seems reasonably balanced. I also employ other variables to control for potential differences across courts in the type and behaviour of defendants and prosecutor. Specifically, I include a measure of trial length, equal to the logarithm of the days since a case is first listed to its completion. We can think about this variable as a proxy for the difficulty of cases to try. Finally, I also introduce a measure of *litigiousness* based on the fraction of all cases that end up in trial rather than being dismissed or ending with a guilty plea.

2.5 Empirical strategy

The empirical strategy exploits the quasi-natural experiment generated by the wage regulation. The location of the Crown Courts is exogenous and not affected by the conditions of the local labour and house markets. At the same time, as explained in Section 2.3, the location choices of judges cannot affect the local markets equilibria given the small size of the pool of potential judges compared to the whole population. We can then think about the real wage as being exogenous since it only depends on the house prices in the area. I estimate a reduced-form version of Equation 2.7:

$$performance_c = \beta_0 + \beta_1 w_c + \beta_2 a_c + \boldsymbol{\beta_3} \mathbf{X_c} + \epsilon_c \tag{2.9}$$

where $performance_c$ is the average performance between 2010 and 2016 of court c, w_c is the average real wage in the same period in the corresponding local authority, a_c is a measure of amenity and $\mathbf{X_c}$ is a vector of court-level controls. I use average values for the 2010-2016 to smooth out yearly variations in the amount and type of cases dealt by individual courts. The key elements of the analysis are: the centralised setting of the wage which generates exogenous variation at court level, the existence of a quality marker defined as the performance measures presented in Section 2.4 and the standardisation of tasks faced by judges which are the same across the country. The analysis controls for potential regional variation in the type of crimes across the country using the individual courts' case-mix. Moreover, there is no selection issue: defendants cannot freely choose where to be tried, the location depends on where the crime was committed. At the same time, there is limited scope for courts to choose the defendants. Prosecutors are in charge of initiating a case and bringing it in front of a court. The latter's function is that of an arbitrator. Courts can choose to dismiss a case but the decision must only be based on its merit. It is true however, that prosecutors or defendants may differ systematically across the country. Measuring the number of appeals and allowed sentence appeals could be capturing different levels of effort or talent of prosecutors and defendants' lawyers alike. For example, if defendants' wages are higher in areas where house prices are also higher, they may be able to hire better representation and, in turn, manage to find grounds to appeal judges' decisions more often. The estimates would then be capturing the effect of better representation as well and would underestimate the impact of real wage and amenities. The opposite could also be true. For example, in poorer, less dense areas, where the real wage is higher, defendants may be more likely to get *bad* representation and later appeal on these grounds. For this reason, I devise a test to control for such possibility. As explained in Section 2.4 judges have little control over convictions which are decided by the jury. I use such discrepancy to run a placebo regression. If the correct mechanism is the one proposed in this paper, the results should not hold true for appeals against convictions as these do not depend on judges' ability. On the other hand, if the variation across the country depends on the behaviour of defendants or prosecutors, we should expect to find the same relationship between conviction appeals, real wage and amenities as that existing with sentence appeals.

The validity of estimates discussed earlier rely on the assumption of exogeneity of the wage and amenities explanatory variables. This is certainly plausible considered the partial equilibrium framework: it is highly unlikely that the type of judges sorting in a court could affect house prices or the level of amenities in the city where the court is located. While reverse causality is certainly not an issue, it is possible that the explanatory variables are correlated with omitted variables. A further step to achieve identification consists in using a geographical indicator as an instrument for the explanatory variables. I then estimate the following model:

$$\begin{aligned} w_c &= \beta_0 + \beta_1 d_c + \boldsymbol{\beta_3}\mathbf{X_c} + \epsilon_c \qquad (2.10) \\ performance_c &= \gamma_0 + \gamma_1 \hat{w}_c + \boldsymbol{\gamma_3}\mathbf{X_c} + \xi_c \end{aligned}$$

in which the real wage is instrumented by a geographical variable d_c, the distance from Charing Cross in London. ξ_c is correlated with the amenities' level, that is omitted in this model, and will cause the estimated coefficient $\hat{\gamma}_1$ to be biased. However, I use the instrument as an indirect way of aggregating the effect of real wage and amenities and interpret the resulting estimate as their combined effect.

2.6 Results

In this section I discuss the results of the empirical analysis presented above. I first focus on the effect of real wage and various proxies for amenities' level on my two measures of performance. Then I run the placebo test. After, I look at the difference between London and other metropolitan areas in the country. Finally, I use the distance to London as an instrument for the real wage.

2.6.1 Real wage and performance

I investigate the relation between real wage and performance and find a significant negative effect. Column 4 of Table 2.2 shows that a 10% increase in the real wage is associated with approximately a 0.2 percentage point fall in performance measured as 1 minus the percentage of allowed sentence appeals. When we look at performance considering all sentence appeals in column 4 of Table 2.3, the estimated effect of a 10% increase in the real wage is 0.37 percentage point fall and it is still strongly statistically significant. In both cases, I consider the model where I control for the type and quantity of cases received by the courts. This does not mean that a higher wage will attract worse candidates. However, a higher real wage, lower house prices, is a proxy for other local characteristics that make a certain location less attractive to a high earner.

Similarly, the different measures of amenities are a predictor of higher performance. In Table 2.2 column 5, when controlling for case-mix and case-load, we see that a standard deviation increase (σ) in urban amenities' level measured by hospitality and leisure business density increases performance by approximately 1 percentage point, and in Table 2.3, with performance computed using all sentence appeals, by 1.27 percentage points. A 10% increase in population density yields a 0.07 percentage point increase in the first measure of performance in Table 2.2, and a 0.06 percentage point increase in the second measure of performance in Table 2.3. The signs of the estimated coefficients are robust to using different measures of amenities. First, I use business density in a 50 km, rather than 30, radius to measure urban amenities. Second, I also use population density in a 50 km radius to capture market access. Finally, I include health and retail businesses with hospitality and leisure in the construction of the amenities variable.[48]

For all variables, the point estimates are consistently larger in the models with controls (columns 4-6). This indicates that omitting the type and number of cases received biases the estimates downward, suggesting that the workings of courts in areas with lower real wages or higher amenities are somewhat more complex and affect performance negatively. It is also consistent with the evidence provided by Figure 2.3 and discussed in Section 2.4 pointing to the fact that the average case composition is similar across courts. In spite of some offence categories such as public order and fraud offences being more common in London courts, the highest density area in the country, no court displays such dominance of a case type to lead us to think that the differences in performance are driven by courts' specialisation. When introducing both the real wage and a measure of amenities as explanatory variables, as in Tables 2.2 and 2.3 columns 3 and 6, we can see the effect of collinearity. Figure 2.4 shows the correlation between house prices and density. The real wage is just the reciprocal of the house prices given the fixed wage structure. Therefore, the two variables, real wage and density, are describing a similar variation in the attractiveness of the location of the Crown Courts. The wage captures all the effect of the attractiveness of location in these specifications. As a result, the coefficients on urban amenities and density become smaller and not statistically significant. The only

[48] See Tables 2.C2 and 2.C3 in Appendix 2.C for the results.

exception is in Panel B of Table 2.2, which I attribute to the fact that log density is less correlated with the log of the real wage (-0.52) compared to the urban amenities proxy (-0.73).

2.6.2 Placebo test and robustness

Table 2.4 uses a measure of performance equal to 1 minus the percentage of allowed conviction appeals. The coefficients show no meaningful relationship between the dependant variable and real wage or amenities, as expected. The coefficient on urban amenities and density are actually negative in most models. The estimated coefficients in columns 3, 5 and 6 are statistically significant. I interpret this as evidence that it is the behaviour of the judges driving the results discussed earlier. Indeed, when using a measure of performance that does not directly depend on their actions, the effect in the results disappears.[49] For additional robustness checks, I control for other variables that proxy the behaviour of defendants and prosecutors across the country as well as the difficulties of local trials. In particular, I add the average time for a case to be completed and a measure of litigiousness based on how often a case goes to trial. Introducing these variables does not affect the statistically significant relationship between real wage and urban amenities on one side, and measures of performance on the other.[50]

2.6.3 The role of London

As seen in Figure 2.4, London courts represent outliers compared to the rest of the population, with much larger amenity levels measured as population density in a 30km radius compared to the rest of the country. London is a mega-city, with a total Gross Value Added of £408.5bn in 2016,[51] representing 23.4% of total UK GVA, nearly 6.5 times as much as the second highest region, Greater Manchester, and a GVA per capita more than double the second highest region, again Greater Manchester. Looking at all local authorities in the UK ranked by GVA per capital, the top 5 (and 6 of the top 10) are located in the Greater London area, with the 4 exceptions being two state capitals, Belfast and

[49]See Table 2.C4 for the same placebo test with amenities and market access indicators computed using a 50 km radius around the courts.

[50]See Tables 2.C5 and 2.C6 in Appendix 2.4.

[51]ONS (2017), regional gross value added.

Edinburgh, and two local authorities adjacent to the Greater London area, Berkshire and Milton Keynes. Gardiner, Martin, Sunley, and Tyler (2013) suggest that such a large imbalance between London and the rest of the country, and more generally between the South and North, is potentially detrimental to the British economy.[52] Such disparities also exist when we look at judicial performance. Figure 2.6a shows the relationship between courts' performance and the distance from Charing Cross. We can see how there is a clear negative trend as we move away from London. However, the figure could just be depicting a North-South divide as London is in the south of the country. Figure 2.6b excludes London courts from the sample and looks at the correlation between courts' performance and the distance from a metropolitan area. We can see that in this case there is a weak negative relationship.

The questions that arises is whether the results are driven by London courts or capturing a nationwide trend. Table 2.5 columns 1-2 show the results from regressing the first performance measure on a dummy variable indication if the court is in London. When controlling for case-mix and case-load, London courts are on average 2.12 percentage points less likely to have their sentences overturned. Columns 3-4 show that, when using a dummy indicating if the court is in metropolitan area [53] (excluding London courts from the exercise) there is weaker (half as much) and non-significant positive difference in performance when comparing metropolitan areas to the rest of the country. The above result is consistent with the idea of a location trade-off: living in a high density area implies greater access to different consumption varieties and, in general, a higher amenities' value. That, however, comes at a cost: a lower real wage. The evidence proposed suggests that the smaller metropolitan areas in the UK, compared to London, offer a lower level amenities that does not compensate enough for the higher rental prices. A better characterisation is provided by Figure 2.7 in which I divide courts in five groups based on the quintiles of the distribution of population density in a 30 km radius around the court. We can see how the normalised performance indicators do not linearly increase with density but are mostly flat, decrease for the third quintile, and then jump for the last

[52] The spatial inequality in the UK is at the centre of a heated public debate nowadays with proponents of a capital move towards the north as a possible solution (Jenkins, 2017; Greaves et al., 2017).

[53] Level 3 metropolitan areas are defined by OECD (2016) as *functional urban areas* with more than a 250,000 population size.

group which includes London courts. Plotting on the same graph a measure of amenity provides the perfect graphic depiction of the narrative above. As density increases but it is not matched by an increase in amenities performance falls (as we can see by comparing the first two quintiles to the third). In London where the availability of amenities is much larger, performance is also considerably higher.

2.6.4 Location as an instrument

In Table 2.6 I show reduced form IV estimates for the model in Equation 2.10. In the richest specifications we can see that being 10% closer to Charing Cross is associated with a 0.06 percentage point reduction in the first measure of performance and a 0.12 percentage point reduction in the second measure.

Table 2.7 reports the results from 2SLS regressions. The point estimates are actually smaller than in the OLS: in the richer specification, a 10% increase in the real wage is associated with a 0.2 percentage point fall in overturned sentences and a 0.3 percentage point fall in the amount of appeals. In all specifications the effect of the real wage on performance is negative and significant. This is evidence that judges are indeed willing to accept a negative wage premium to sort in locations offering a higher level of amenities.

2.6.5 Alternative mechanisms

This paper argues that the differences in performance of UK Crown Courts are due to the location choice of judges who are attracted by amenities in higher density areas, London in particular. There are, however, potential alternative mechanisms that would explain these findings. First of all, as already mentioned, differences in performance could be due to higher specialisation in London courts. Nonetheless, the evidence seems to point against such conclusion. Figure 2.3 shows that despite systematic differences between London courts and the others in terms of offence categories of cases received, the overall case composition is fairly balanced. Moreover, in the empirical analysis controlling for the case-mix actually strengthens the relationship between amenities and performance. A different possibility is that judges are ex-ante identical but being in London increases their ability due to human capital externalities correlated with amenities.[54] This is surely a possibility but the profile of sitting judges seems to suggest that most candidates join after having already acquired sizeable experience in the private sector. Clearly, the human capital built working as a lawyer or in other positions in the legal sector is not necessarily a substitute for that acquired on the job. There are also other factors rather than amenities that could determine the location choices of talented judges. In particular, the preference for London could be locked in by a previous career choice occurred at a younger age in presence of a fixed cost of relocation. Also, the location preference could be driven by a joint decision with the spouse as theorised by Costa and Kahn (2000).[55] Career concerns could also be a factor. While for many being a circuit judge is an end of career job (Schultz and Shaw, 2003), a few progress to higher courts.[56] From one side, being exposed to higher profile cases could increase the chances of being recommended for a position in a higher court. From the other, as the High Court and the Court of Appeal, the usual career avenues, are located in London, being in the same city could allow them to build a network that will ease career advancement. Yet the fact that courts in Bristol and Manchester as well as those in the outskirts of London are high performers seems to partially refute the latter. The above hypotheses would all require individual level data on judges to be formally tested. Finally, we can also conjecture the existence of a *job amenity* deriving from being in London. Even though the case composition does not vary drastically across the country, it is possible that cases in London are more interesting to try for judges.

2.7 Conclusion

In this paper I exploit the rigidity of nationally regulated wages for criminal court judges to show the existence of a negative relationship between real wages and performance. At the same time, I show that performance is higher when courts are located in areas with

[54]See Moretti (2004) for a discussion on human capital externalities in cities.

[55]It must be noted that one of the reasons Costa and Kahn (2000) cite as a determinant of power couples' choice to relocate to large cities is amenities. Conversely, Compton and Pollak (2007) suggest that it is only the characteristics of the husband that affect joint location choices, while power couples are formed *inside* large cities.

[56]The most common route to high courts is through the Queen's Counsel (Schultz and Shaw, 2003).

more urban amenities. The results hold true with two different measures of performance and several measures of urban amenities and market density. I interpret these findings as evidence that high-skilled, high-earning workers sort into large cities because they are attracted by urban amenities and consumption varieties. This differs from standard agglomeration theory predicting that the concentration of high-skilled workers in larger cities is driven by the higher wages and other forms of externalities.

A second positive implication of the paper is outlining another source of inequality between the city of London and the rest of the United Kingdom. Courts in London, indeed, perform better on average than the ones in the rest of the country. If the performance of judges differs among courts, defendants may receive a different treatment depending on the court in which they are tried. Given the role of a judge as a mediator between public interest and defendants' rights, these findings deserve scrutiny by policymakers. A normative implication would be that wages should vary across the country depending on the attractiveness of a location.

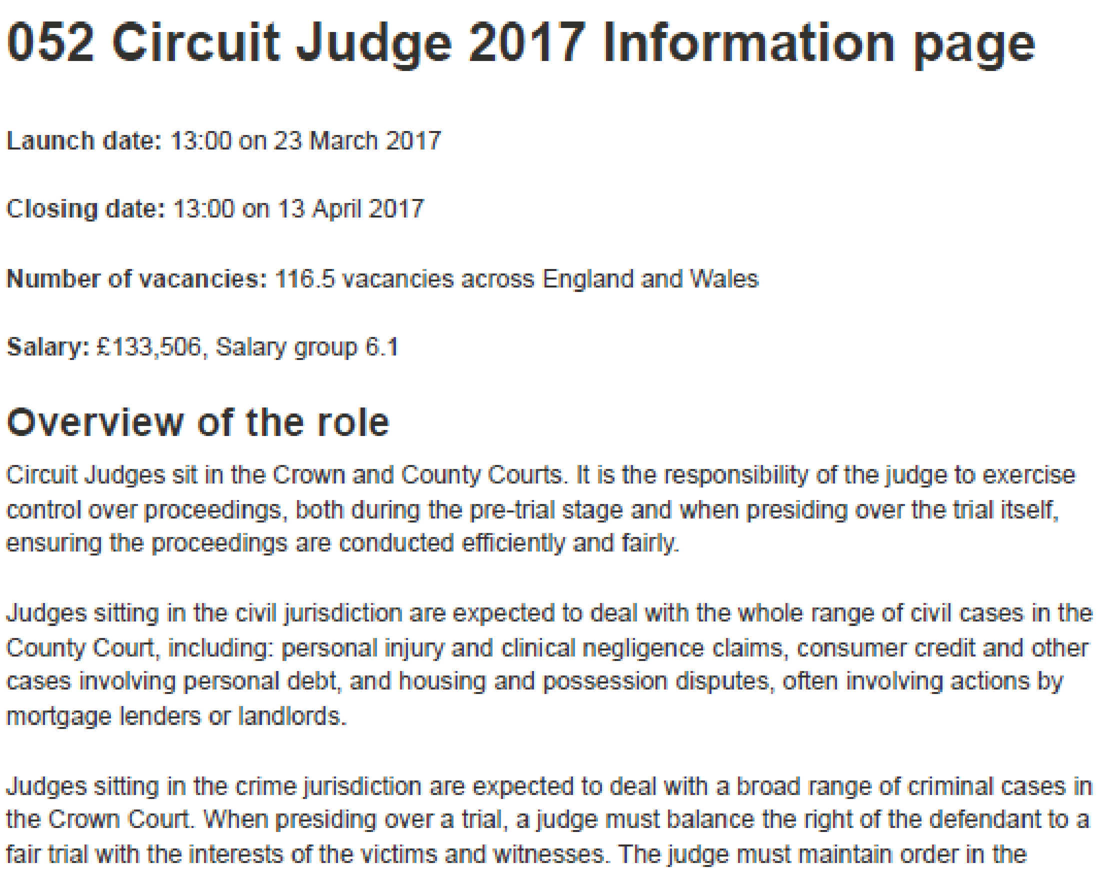
More generally, we learn that governments willing to relocate institutions serving the public interest away from congested cities with high rental cost, should be aware of a trade-off between cost efficiency and the quality of civil servants that depends on their skill level and remuneration.[57]

Moreover, the findings of this paper are particularly relevant at a time when the Ministry of Justice and the Home Office are piloting virtual hearings (Home Office, 2017) to increase the efficiency of the justice process. Despite criticism that the *court gravitas* is a necessary component of criminal trials as well as the emotions of defendants and witnesses (The Telegraph, 2015), a future where the judiciary is centralised and all functions are performed by a single large court is not unimaginable. If that is the case, this paper suggests that the location of the court should be carefully chosen to make sure to attract the most talented candidates.

[57]For example, consistent with the evidence presented in this paper, the recently conducted independent review of UK economics statistics (Bean, 2016) concluded that the 2007 relocation of the Office for National Statistics from London to Newport, South Wales, has significantly reduced the quality of its output.

2.8 Figures

Figure 2.1: Example of a circuit judge Job Advert



052 Circuit Judge 2017 Information page

Launch date: 13:00 on 23 March 2017

Closing date: 13:00 on 13 April 2017

Number of vacancies: 116.5 vacancies across England and Wales

Salary: £133,506, Salary group 6.1

Overview of the role

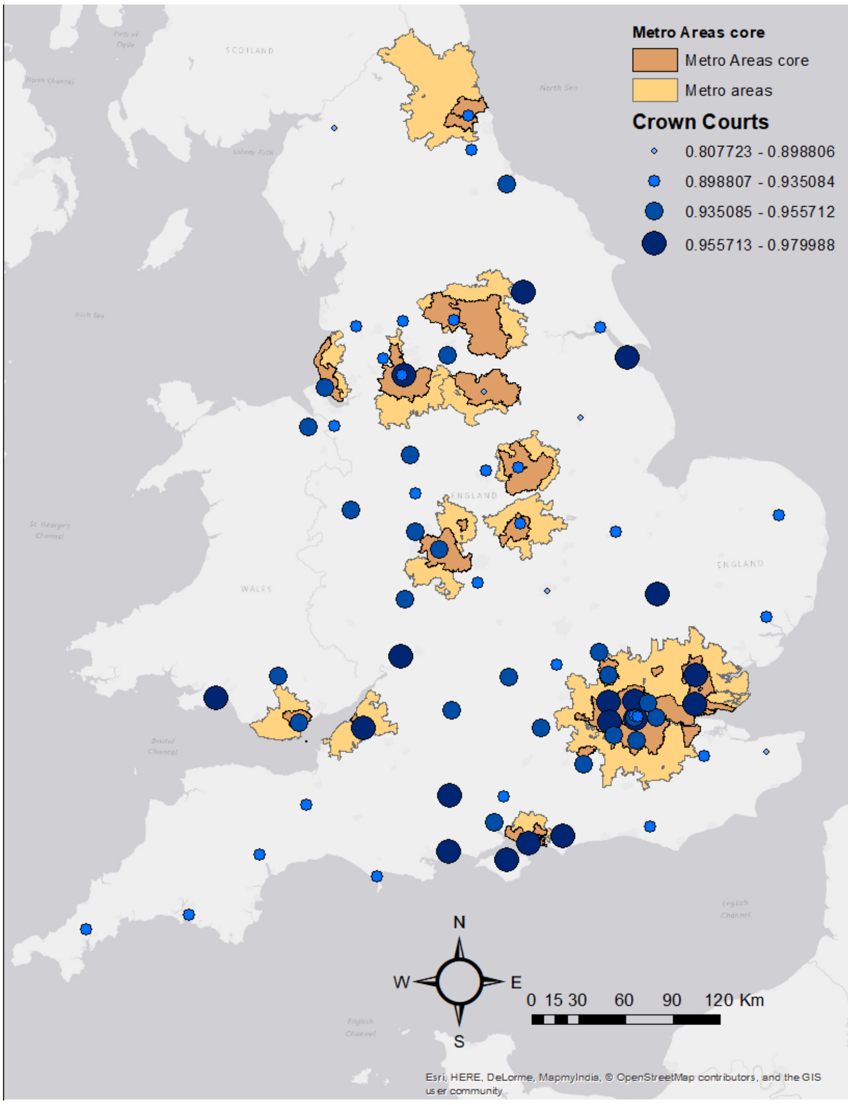
Circuit Judges sit in the Crown and County Courts. It is the responsibility of the judge to exercise control over proceedings, both during the pre-trial stage and when presiding over the trial itself, ensuring the proceedings are conducted efficiently and fairly.

Judges sitting in the civil jurisdiction are expected to deal with the whole range of civil cases in the County Court, including: personal injury and clinical negligence claims, consumer credit and other cases involving personal debt, and housing and possession disputes, often involving actions by mortgage lenders or landlords.

Judges sitting in the crime jurisdiction are expected to deal with a broad range of criminal cases in the Crown Court. When presiding over a trial, a judge must balance the right of the defendant to a fair trial with the interests of the victims and witnesses. The judge must maintain order in the

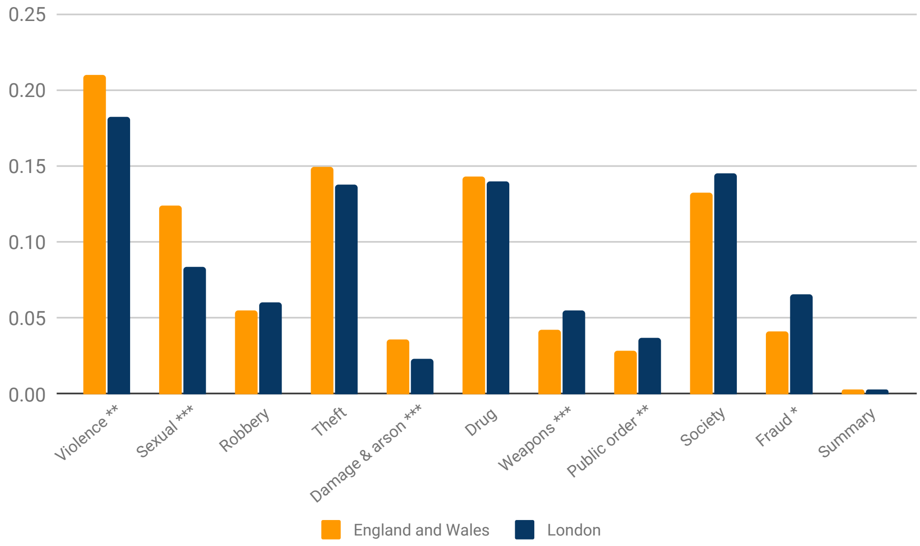
Source: Judicial Appointments Commission (2017)

Figure 2.2: Crown Court performance

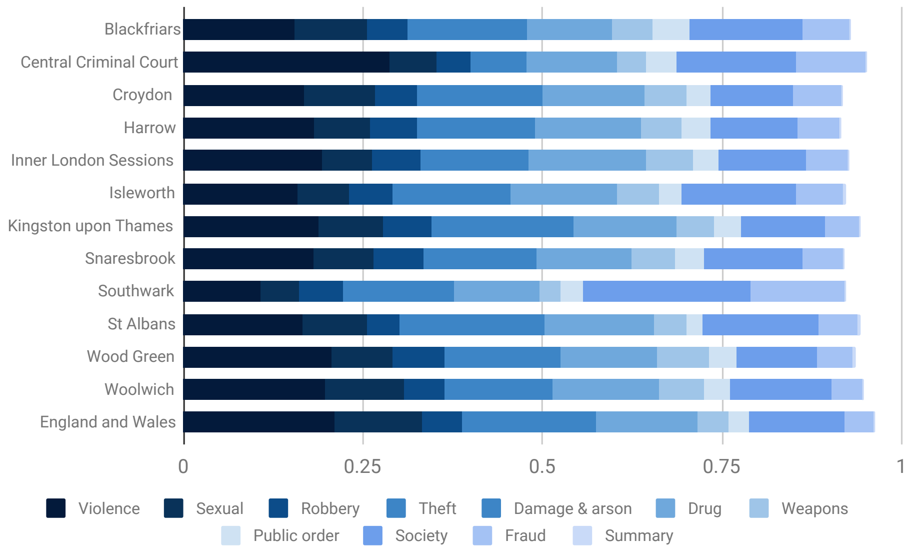


The performance measure is the percentage of allowed sentence appeals (overturned sentences) on the number of effective trials.

Figure 2.3: Case-mix in UK criminal courts



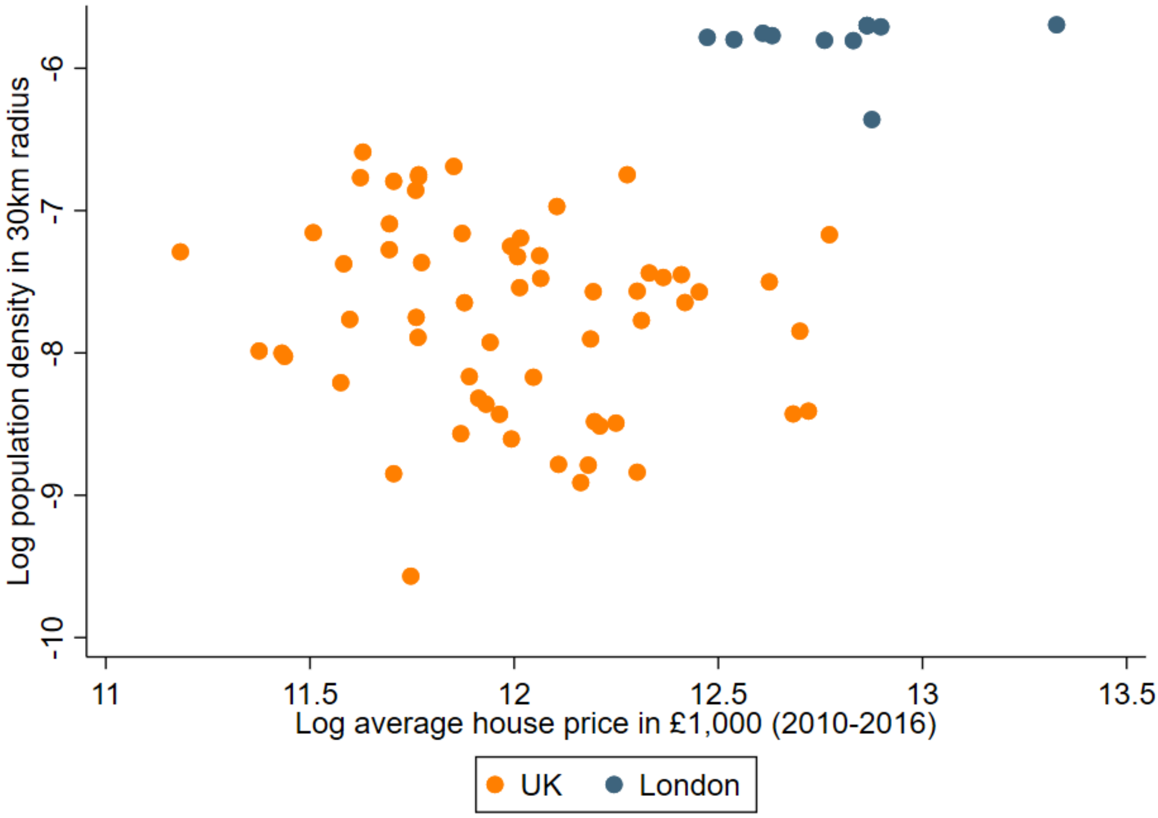
(a) Type of cases in London courts vs the rest of England and Wales



(b) London courts case-mix

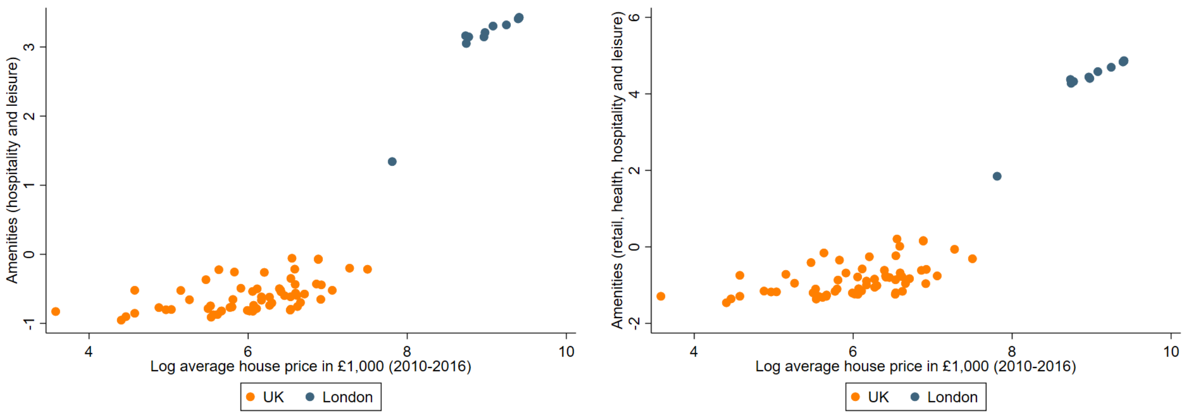
Figure 2.3a shows the average amount of cases by offence dealt with in London courts compared to the rest of England and Wales. On the y-axis: percentage of cases received. Each offence category reports the result from a two-sided t-test for the two means being equal with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Figure 2.3b shows instead the case composition in London courts. On the x-axis: percentage of cases received. The London area is defined as a circle of 35 km radius around Charing Cross. The offence categories are: Violence against the person, Sexual, Robbery, Theft, Criminal damage and arson, Drug, Possession of Weapons, Public order, Miscellaneous against society, Fraud, Summary.

Figure 2.4: Amenities vs house prices



Each data point is a Crown Court. On the vertical axis: population density in a 30 km radius circle around the court, in logarithm. On the horizontal axis: the logarithm of the average house price, in a 15 km radius, in £1000 from 2010 to 2016. The London area is defined as a circle of 35 km radius around Charing Cross.

Figure 2.5: Amenities vs house prices

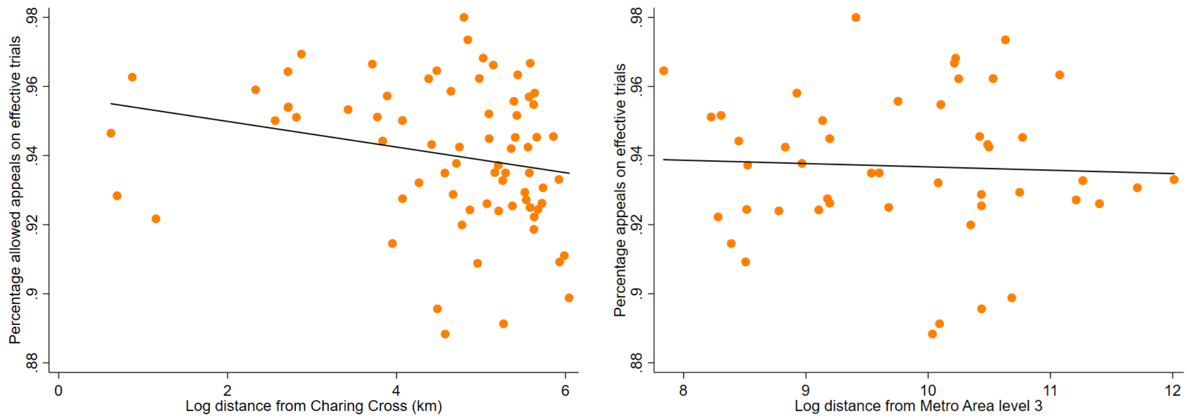


(a) Hospitality and leisure

(b) Health, hospitality, leisure and retail

Each data point is a Crown Court. On the vertical axis: measure of urban amenities equal to the first principal component of the hospitality and leisure business density (left), health, hospitality, leisure and retail business density (right). On the horizontal axis: the logarithm of the average house price, in a 15 km radius, in £1000 from 2010 to 2016. The London area is defined as a circle of 35 km radius around Charing Cross.

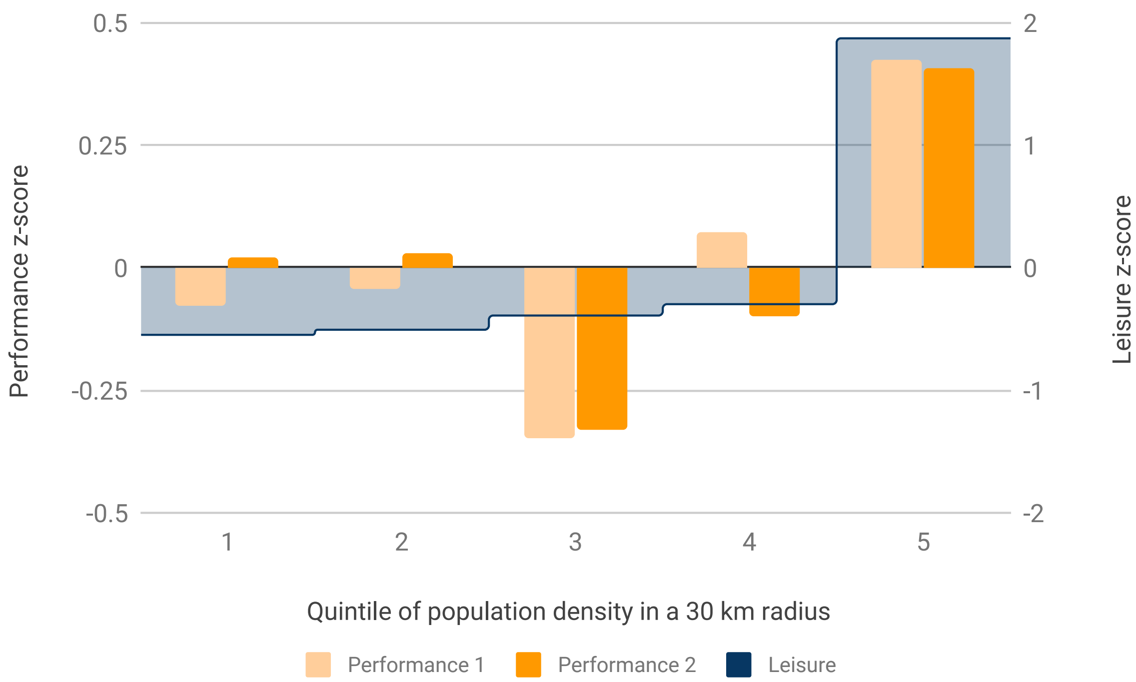
Figure 2.6: Performance and distance from metropolitan areas



(a) Distance from Charing Cross (b) Distance from Metro Area level 3

Each data point is a Crown Court. On the vertical axis: performance measured using the percentage of allowed sentence appeals on effective trials. On the horizontal axis: on the left the logarithm of the distance between the court and Charing Cross, on the right the logarithm of the distance between the court and the closest level 3 metropolitan area, defined according to OECD (2016).

Figure 2.7: Performance by density



Each bar represents the average performance z-score for the Crown Courts grouped by quintiles of the distribution of population density in a 30 km radius. The coloured area represents the z-score for a measure of amenities using the same grouping. Performance 1 is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials. Performance 2 equal to 1 minus the ratio between the number of sentence appeals and effective trials. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court.

2.9 Tables

Table 2.1: Circuit judges' salaries

Year	Standard salary	Old Bailey
2012	£128,296	£138,548
2013	£129,579	£139,933
2014	£130,875	£141,332
2015	£132,184	£142,745
2016	£133,506	£144,172

Source: Ministry of Justice (2015, 2016), Judicial Salaries. Each salary is effective from April, 1st of the relative year. The column to the right reports the salary of circuit judges at the Central Criminal Court in London (the Old Bailey) who receive a higher wage, band 6.1, compared to band 5 for all the other circuit judges.

Table 2.2: Judiciary performance, real wage and amenities

	Dependent variable: Allowed sentence appeals					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Log real wage	-0.0128***		-0.0114*	-0.0223***		-0.0194***
	(0.0040)		(0.0060)	(0.0061)		(0.0065)
Leisure		0.00364***	0.000667		0.00688**	0.00398
		(0.0013)	(0.0020)		(0.0029)	(0.0030)
Panel B						
Log real wage			-0.0101**			-0.0218***
			(0.0046)			(0.0057)
Log density 30 km		0.00550**	0.00276		0.00746*	0.00701**
		(0.0024)	(0.0027)		(0.0039)	(0.0032)
Case-mix				✓	✓	✓
Case-load				✓	✓	✓
Observations	73	73	73	69	69	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials. Log real wage is the logarithm of the court judges' real wage. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court. Log density 30 km is the logarithm of population density in a 30 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.3: Judiciary performance, real wage and amenities, all appeals

	Dependent variable: All sentence appeals					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Log real wage	-0.0212***		-0.0245***	-0.0392***		-0.0367***
	(0.0063)		(0.0086)	(0.0082)		(0.0088)
Leisure		0.00480**	-0.00158		0.00894**	0.00346
		(0.0022)	(0.0029)		(0.0041)	(0.0039)
Panel B						
Log real wage			-0.0214***			-0.0387***
			(0.0067)			(0.0080)
Log density 30 km		0.00563	-0.000192		0.00675	0.00595
		(0.0036)	(0.0038)		(0.0060)	(0.0048)
Case-mix				✓	✓	✓
Case-load				✓	✓	✓
Observations	73	73	73	69	69	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of sentence appeals and effective trials. Log real wage is the logarithm of the court judges' real wage. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court. Log density 30 km is the logarithm of population density in a 30 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.4: Placebo test

	Dependent variable: Allowed conviction appeals					
	(1)	(2)	(3)	(4)	(5)	(6)
Log real wage	0.000746	-0.00161				
	(0.0010)	(0.0016)				
Leisure			-0.000564*	-0.000317		
			(0.00032)	(0.00056)		
Log density 30 km					-0.00107**	-0.00142**
					(0.00048)	(0.00067)
Case-mix				✓	✓	✓
Case-load				✓	✓	✓
Observations	73	69	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed conviction appeals and effective trials. Log real wage is the logarithm of the court judges' real wage. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court. Log density 30 km is the logarithm of population density in a 30 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.5: Judiciary performance and metropolitan areas

	Dependent variable: percentage of appeals			
	(1)	(2)	(3)	(4)
London	0.0131***	0.0212**		
	(0.0047)	(0.0096)		
Metro area			0.00566	0.00950
			(0.0062)	(0.0071)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	62	58

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials. London is a dummy=1 if the court is located in a 35 km radius around Charing Cross. Metro area is a dummy=1 if the court is located in a level 3 metropolitan area according to the OECD (2016) definition. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester. Models 3-4 also exclude all courts in the London metropolitan area.

Table 2.6: Reduced form IV, judiciary performance and distance from Charing Cross

	Dependent variable:			
	Allowed appeals		All appeals	
	(1)	(2)	(3)	(4)
Log distance from Charing Cross	-0.00371**	-0.00727**	-0.00565*	-0.0126***
	(0.0017)	(0.0031)	(0.0029)	(0.0043)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials (1-2) or equal to 1 minus the ratio between the number of sentence appeals and effective trials (3-4). Log distance from Charing Cross is the logarithm of the distance in km between the court and Charing Cross. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.7: 2SLS Judiciary performance and real wage

	Dependent variable:			
	Allowed appeals		All appeals	
	(1)	(2)	(3)	(4)
Log real wage	-0.0108**	-0.0204***	-0.0164**	-0.0352***
	(0.0044)	(0.0076)	(0.0075)	(0.0099)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. 2SLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials (1-2) or equal to 1 minus the ratio between the number of sentence appeals and effective trials (3-4). Log real wage, the endogenous regressor, is the logarithm of the court judges' real wage. The excluded instrument is the logarithm of the distance in km between the court and Charing Cross. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester. The robust first-stage F-statistics at a 5% confidence level is 137.994 in columns 1, 3 and 54.211 in columns 2, 4.

2.A Data

The judiciary variables were obtained from a number of sources. The information on the number of cases and trials derives from the Criminal Court quarterly statistics 2010-2016 (Ministry of Justice, 2017a). The information on appeals and the number of judges in each court is not published by the Ministry of Justice and was obtained in the response to two successive Freedom of Information requests (Ministry of Justice, 2018b,c). The information on different types of offences for individual Crown Courts is obtained from the Crown Courts receipts, disposals, and outstanding cases by offence group 2014-16 (Ministry of Justice, 2018a).

Performance: Performance measures are computed as:

$$Performance_c = 1 - \frac{Appeals_c}{Trials_c}$$

$Appeals_c$ can be the number of allowed sentence appeals, the number of sentence appeals or the number of allowed conviction appeals (in the variable used for the placebo test) against rulings from Crown Court c. $Trials_c$ is the number of effective trials in Crown Court c. Effective trials are trials where the jury is sworn in and a verdict is reached. I exclude Sheffield Crown Court from the sample because it is an outlier: performance measured using allowed sentence appeals is 7 standard deviations lower than the mean, performance measured using all appeals is 5.6 standard deviations lower than the mean. The results are robust to including Sheffield Crown Court in the sample.

Case-mix and case-load: I compute the case-mix starting from the percentage of received cases for each offence group compared to the total of received case per court. The categories are: violence against the person, sexual offences, robbery, theft offences, criminal damage and arson, drug offences, possession of weapons, public order offences, miscellaneous crimes against society, fraud offences, summary, and unknown. For each category I compute the average percentage for the years 2014-16. My measure of case-mix is the three first principals components of all the individual case categories' percentage prevalence in the court. Case-load instead is the ratio between the number of cases received by a Crown Court and the number of judges sitting in said court. There is no case-load data on the following Crown Courts: Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Wage: I compute the real salary as the ratio between the average circuit judges' nominal salary from the Ministry of Justice (2015, 2016) Judicial Salaries reported in Table 2.1 and the the average house price index from the HM Land Registry (2018) in a 15 km radius around the location of the Crown Court. The average house price index is calculated using weights based on the overlap between the circle around the court and the territory of the local authority. The 15 km radius of the circle is chosen approximately equal to the average commuting distance in the UK in 2015 (Le Vine, Polak, and Humphrey, 2017).

Amenities: I compute different indicators of the level of amenities in the city where Crown Court c is located using the following formula:

$$amenities_c = \frac{\sum_i^N \alpha_i \times \frac{intersect_{ic}}{area_{ic}}}{area_c}$$

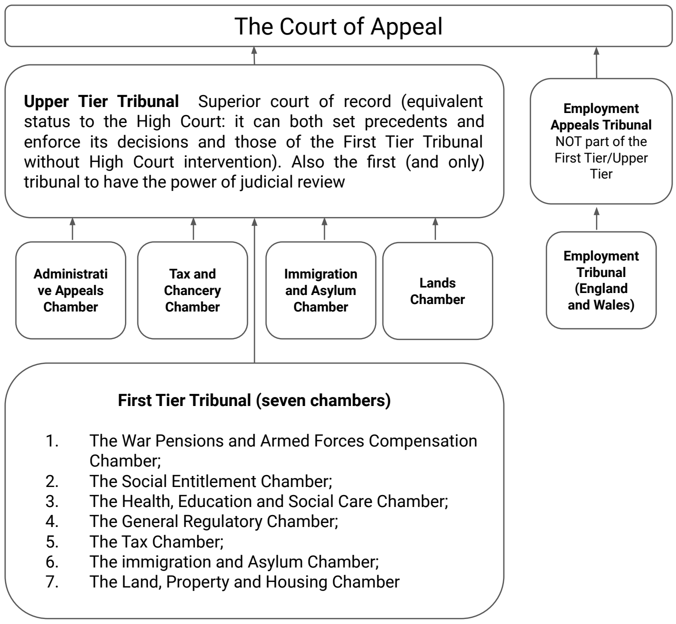
The circle around the Crown Court c is either of 30 or 50 km in radius. When the level of amenities is measured using population density, α_{ic} is the population living in output area i and $area_{ic}$ the area of the output area m^2. $intersect_{ic}$ is the intersection between the two. See Figure 2.B4 for a graphical representation. When the level of amenities is measured using business density, α_{ic} is the number of businesses in local authority i and $area_{ic}$ the area of the local authority in m^2. $intersect_{ic}$ is the intersection between the two. I compute the density of the following business categories from ONS (2017): G - Wholesale and retail trade; repair of motor vehicles and motorcycles, I - Accommodation and food service activities, Q - Human health and social work activities, R - Arts, entertainment and recreation. The two amenity variables I construct are the first principal components of respectively: categories I and R in the first variable (*Leisure*), all four categories in the second (*Amenities*).

Table 2.A1: Main variables, summary statistics

Variable	Mean	SD	Min	Max	N Obs
Performance (allowed)	0.940	0.020	0.888	0.980	73
Performance (all)	0.913	0.029	0.837	0.969	73
Real wage	0.630	0.288	0.218	1.351	73
Leisure (30 km)	0.008	1.421	-0.953	3.429	73
Density (30 km)	0.001	0.001	0.000	0.003	73
Amenities (30 km)	0.010	2.008	-1.460	4.870	73

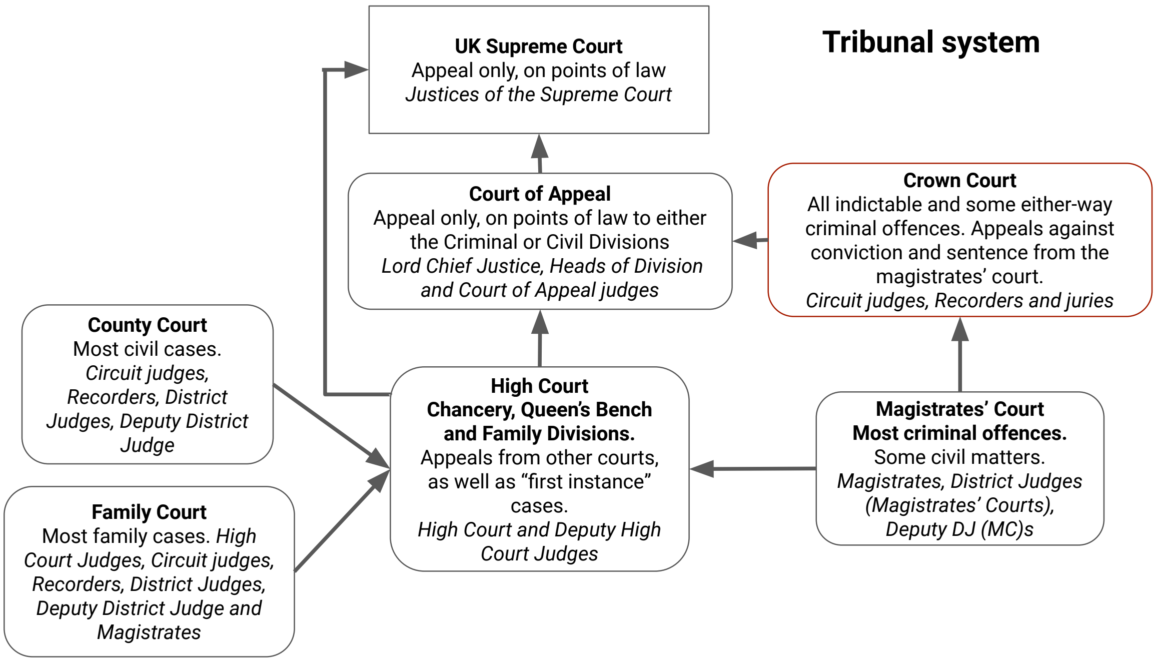
2.B Additional Figures

Figure 2.B1: Tribunals structure chart



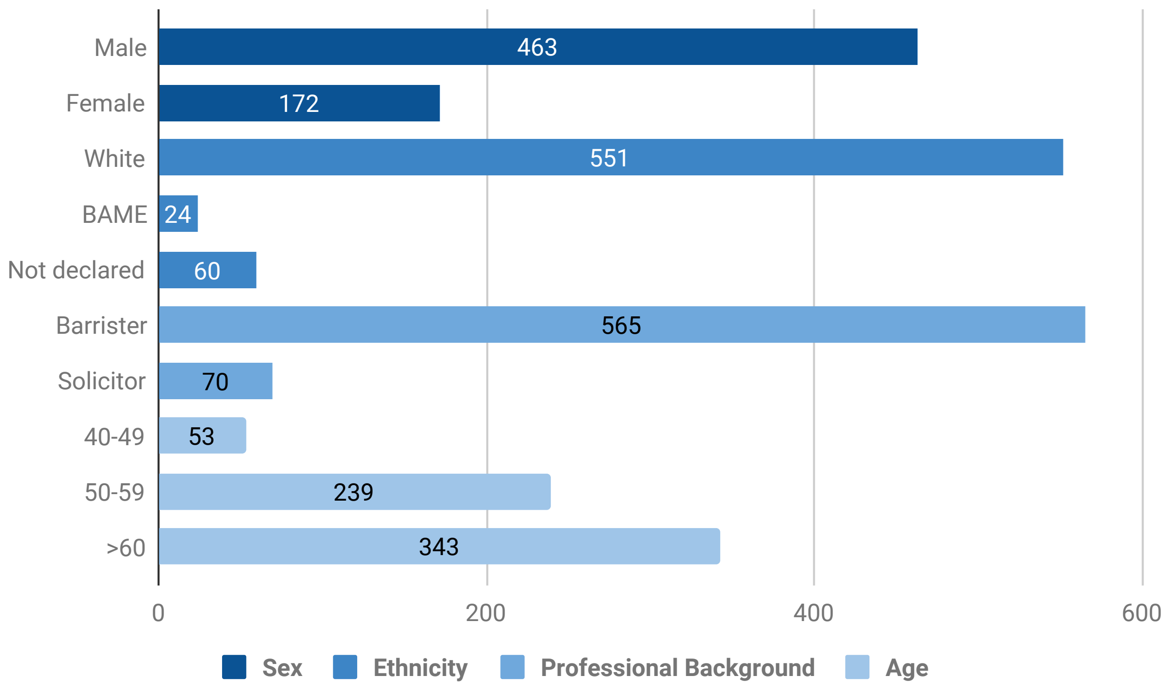
Source: The structure of the Tribunals Service in The Judicial System of England and Wales. A visitor's guide (Lady Justice Arden, 2016).

Figure 2.B2: The courts of England and Wales



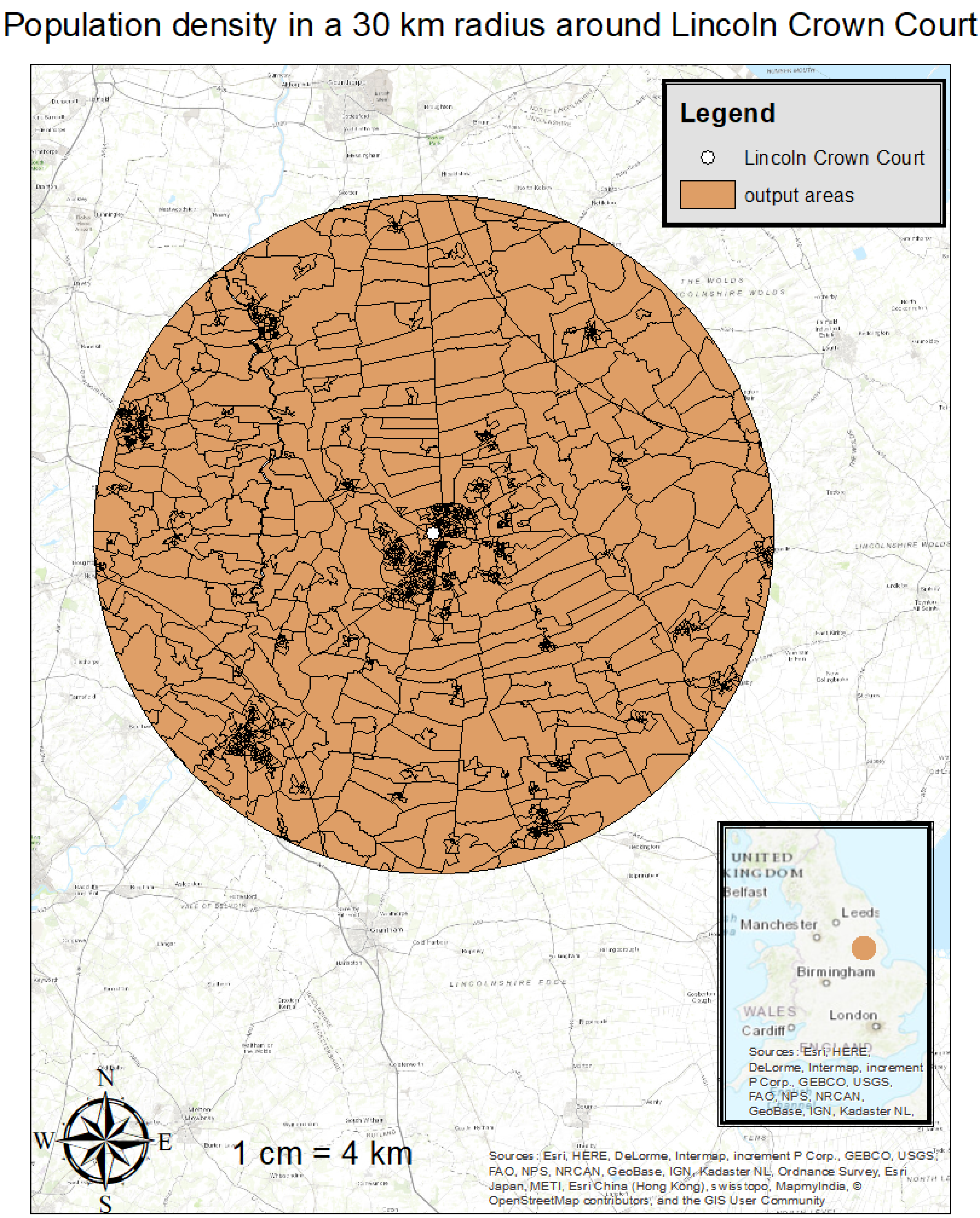
Source: The Structure of the Courts in The Judicial System of England and Wales. A visitor's guide (Lady Justice Arden, 2016).

Figure 2.B3: Circuit judges diversity statistics



Source: Judicial Diversity Statistics 2017, Table 1.1: Primary appointment of Judges in Courts in England and Wales, by gender, ethnicity, professional background and age, as of 1 April 2017 (Judiciary, 2017).

Figure 2.B4: Output areas in a 30 km radius around Lincoln Crown Court



2.C Additional Tables

Table 2.C1: Days in court by judge type

	Days sat in Court									
	Crown Court					County Court				
Type of judge	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016
Lords Justices	-	-	-	-	-	29	2	39	2	0
High Court judges	3,838	3086	2,935	3,075	2,976	152	161	215	94	161
Deputy High Court judges	206	8	24	2	93	267	271	179	80	13
circuit judges	80,621	79,621	77,211	80,745	81,454	34,714	34,262	33,446	33,667	33,583
Deputy circuit judges	1,229	582	432	665	519	463	512	276	339	625
Recorders	20,173	19,919	23,894	29,006	25,246	5,618	5,797	5,386	5,608	5,752
District judges	-	-	-	-	-	83,513	80,484	78,970	77,754	76,056
Deputy district judges	-	-	-	-	-	24,757	26,632	26,709	27,273	30,765
Unknown	481	380	556	473	506	0	0	0	0	0
Other	-	-	-	-	-	490	654	21,330	29,043	28,909
Total	106,548	103,596	105,052	113,966	110,794	150,002	148,774	166,548	173,858	175,864

Source: Royal Courts of Justice tables, Civil Justice statistics quarterly: January to March 2018 (Ministry of Justice, 2018d). Unknown judge means that the judge type was not recorded.

Table 2.C2: Performance, real wage and amenities, other measures

	Dependent variable: Allowed appeals			
	(1)	(2)	(3)	(4)
Panel A				
Log real wage 30 km	-0.0146***	-0.0238***		
	(0.0042)	(0.0065)		
Panel B				
Leisure 50 km	0.00423***	0.00802***		
	(0.0014)	(0.0025)		
Log density 50 km			0.00713**	0.00857**
			(0.0027)	(0.0036)
Panel C				
Amenities	0.00256**	0.00487**		
	(0.00097)	(0.0021)		
Amenities 50 km			0.00287***	0.00520***
			(0.00098)	(0.0018)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials. Log real wage 30 km is the logarithm of the average court judges' real wage in a 30 km circle around the court. Leisure 50 km is the first principal component score of hospitality and leisure business densities in a 50 km radius circle around the Crown Court. Log density 50 km is the logarithm of population density in a 50 km radius circle around the Crown Court. Amenities and Amenities 50 km are the first principal component score of health, hospitality, leisure and retail business densities in respectively a 30 and 50 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.C3: Performance, real wage and amenities, other measures, all appeals

	Dependent variable: All appeals			
	(1)	(2)	(3)	(4)
Panel A				
Log real wage 30 km	-0.0251***	-0.0433***		
	(0.0065)	(0.0088)		
Panel B				
Leisure 50 km	0.00587***	0.0114***		
	(0.0021)	(0.0035)		
Log density 50 km			0.00808*	0.00958*
			(0.0041)	(0.0055)
Panel C				
Amenities	0.00330**	0.00607**		
	(0.0015)	(0.0029)		
Amenities 50 km			0.00387**	0.00711***
			(0.0015)	(0.0025)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of sentence appeals and effective trials. Log real wage 30 km is the logarithm of the average court judges' real wage in a 30 km circle around the court. Leisure 50 km is the first principal component score of hospitality and leisure business densities in a 50 km radius circle around the Crown Court. Log density 50 km is the logarithm of population density in a 50 km radius circle around the Crown Court. Amenities and Amenities 50 km are the first principal component score of health, hospitality, leisure and retail business densities in respectively a 30 and 50 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.C4: Placebo test, robustness

	Dependent variable: Allowed conviction appeals			
	(1)	(2)	(3)	(4)
Leisure 50 km	-0.000461	0.0000550		
	(0.00033)	(0.00057)		
Log density 50 km			-0.00121**	-0.000971
			(0.00059)	(0.00080)
Case-mix		✓		✓
Case-load		✓		✓
Observations	73	69	73	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed conviction appeals and effective trials. Leisure 50 km is the first principal component score of hospitality and leisure business densities in a 50 km radius circle around the Crown Court. Log density 50 km is the logarithm of population density in a 50 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.C5: Performance, real wage and amenities, robustness

	Dependent variable: Allowed sentence appeals					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Log real wage	-0.0283***	-0.0262***				
	(0.0072)	(0.0080)				
Leisure			0.00792***	0.00785***		
			(0.0029)	(0.0027)		
Log density 30 km					0.00814**	0.0107***
					(0.0040)	(0.0033)
Case-mix & load	✓	✓	✓	✓	✓	✓
Length trial	✓	✓	✓	✓	✓	✓
Litigious		✓		✓		✓
Observations	69	69	69	69	69	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of allowed sentence appeals and effective trials. Log real wage is the logarithm of the court judges' real wage. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court. Log density 30 km is the logarithm of population density in a 30 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Length trial is the logarithm of the average time in the court from when the case is listed to when it is completed. Litigious is the ration between the cases that go to jury trial and the overall amount of cases. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Table 2.C6: Performance, real wage and amenities, robustness, all appeals

	Dependent variable: All sentence appeals					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Log real wage	-0.0485***	-0.0423***				
	(0.0097)	(0.011)				
Leisure			0.0103**	0.0102***		
			(0.0041)	(0.0037)		
Log density 30 km					0.00763	0.0120**
					(0.0062)	(0.0051)
Case-mix & load	✓	✓	✓	✓	✓	✓
Length trial	✓	✓	✓	✓	✓	✓
Litigious		✓		✓		✓
Observations	69	69	69	69	69	69

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. OLS where the dependent variable is equal to 1 minus the ratio between the number of sentence appeals and effective trials. Log real wage is the logarithm of the court judges' real wage. Leisure is the first principal component score of hospitality and leisure business densities in a 30 km radius circle around the Crown Court. Log density 30 km is the logarithm of population density in a 30 km radius circle around the Crown Court. Case-mix is the three first principal components for twelve measures of prevalence of a category of offence in the court. Case-load is the total number of cases received (including those that never go to trial) over the number of judges sitting in the court. Length trial is the logarithm of the average time in the court from when the case is listed to when it is completed. Litigious is the ration between the cases that go to jury trial and the overall amount of cases. Sheffield Crown Court is excluded from all regressions. Models where case-load is used as a control also exclude Great Grimsby, Newport, Salisbury and Weymouth & Dorchester.

Chapter 3

Innovation as a determinant of religiousness

3.1 Introduction

Since Azzi and Ehrenberg (1975), a number of researchers have focused on the economic determinants of religion. A strand of literature has been devoted to settling the debate about the long-standing *secularisation hypothesis* that predicts the death of religion in an industrialised society (Bruce, 1992).[1] The most widely accepted explanations are that on one hand industrialisation and economic development bring a sense of existential security, making people less vulnerable and in need of religion (Inglehart and Norris, 2004),[2] whilst on the other they provide an increasing number of attractive alternatives that increase the opportunity cost of adhering to religious norms (Carvalho and Koyama, 2011). At the same time we know that starting with the industrial revolution, technological progress has had a profound impact on the way people live, work and interact with each other. Should we then expect scientific advances *per se* to have contributed to the diminishing importance of religion?

This paper provides empirical evidence on the negative relationship between innova-

[1]This is a consistent idea in European philosophy and sociology, tracing back to the works of Comte, Spencer, Weber, Marx. It seems to exhibit striking evidence in the data when looking at the low religious participation rates in Western Europe (Inglehart and Norris, 2004).

[2]Early works in the religious economic literature placed special focus on the social function of religion denominations (Iannaccone, 1992, 1994; Montgomery, 1996).

tion and religiousness[3] which was originally uncovered by Bénabou, Ticchi, and Vindigni (2015a). I test the relationship at both country and US state-level and show that it is robust to using different specifications as well as introducing different measures of religiousness and innovation. Similar to Barro and McCleary (2003a), who try to identify the causal effect of economic growth on religion in the presence of reverse causality, I employ instrumental variable analysis to try to pin down the effect of innovation on religiousness. I use the strength of patent protection law and, more generally, property rights protection to instrument patents per capita, the measure of innovation, in a cross-country analysis. Conversely, in the US state-level analysis I build an instrumental variable based on the spatial distribution of universities across US states.

I first build a simplified version of the theoretical framework in Bénabou, Ticchi, and Vindigni (2015a) with the intent of providing evidence on the effect of technological progress (innovation) on religiousness. The idea behind the model is that, although the repeated arrival and diffusion of new scientific knowledge has been welcomed by some governments due to its productivity and competitiveness enhancing abilities, others have deployed resources to impede their diffusion and thus absorb some of the challenges that it posed to religion as an institution.[4] The model predicts that faster technological progress will be accepted despite its damaging effect on the religious doctrine, as it leads to higher future output. I then proceed to test this finding empirically.

I measure religiousness using different variables originating from country-level averages of individual responses to questions included in 6 waves of the World Values Survey Association (2015) ad 4 waves of the European Value Study (2011) from 1985 to 2010. In

[3]While in the literature the terms *religiousness* and *religiosity* are often used interchangeably, encapsulating all dimensions of religion, religiosity can also be a narrower concept referring to one's dedication to religion (Gallagher and Tierney, 2013). In this paper the distinction between the two terms is not trivial. When using the term *religiousness* I am referring to all religious indicators in a broad sense. Instead, the term religiosity will specifically refer to one of the religious indicators that I will introduce later, reporting the fraction of people in a country that consider themselves religious.

[4]Historical evidence is replete with examples of recurrent tensions between organised religion and science, which have continued to remain relevant nowadays in a number of countries where governments have tried to impede the spread of knowledge. One of such cases is the increasingly vocal support for *creationism* as a valid alternative to Darwin's theory of evolution in the United States (Berkman and Plutzer, 2011). The same phenomenon has been witnessed in other countries, including Israel and Turkey (Burton, 2011). A case at the opposite ends of the spectrum is that of post-revolutionary Iran, where clerics and political leaders have actively promoted science and technology, fuelled by the desire to promote the country's international status. In fact, in 2002, Iran's Supreme Leader, Ayatollah Khamenei, issued a *stem cell fatwa* - a fatwa is religious opinion about whether an action is permissible - declaring the consistency of experimentation on human embryos with Shia Islam tradition, resulting in Iran being the first Muslim country to produce, culture and freeze them (Miremadi et al., 2012).

particular, I focus on four indicators, *religiosity*, *importance of religion*, *beliefs in God* and *weekly church attendance* that are widely used in the empirical literature on religion and economics.[5] Following Bénabou, Ticchi, and Vindigni (2015a), I measure innovation using World Intellectual Property Organization (2015) country-level data on patent filings. I also introduce a different indicator of innovation based on country-level production of scientific articles and technical reports.[6]

I show the existence of a strong negative relationship between innovation and religiousness across countries that remains unaffected when using the different measures of either the dependent or independent variables as well as controlling for all the other potential determinants of religiousness discussed in the literature. The findings provide evidence that countries with faster technological progress exhibit lower levels of religiousness. However, it is not possible to interpret the results as evidence of a causal relationship between innovation and religion, because of reverse causality. In an attempt to provide convincing evidence of an effect of innovation on religion, I implement instrumental variable analysis where innovation is instrumented by the level of patent protection in the country and by an index of property rights protection. The results confirm the existence of the negative relationship, even if the point estimates must be interpreted with caution.

I also perform the same analysis at US-state level, using measures of religiousness deriving from survey responses to equivalent questions from the Pew Reseach Center (2016) Religious Landscape Study. I find evidence of the same significant negative relationship between innovation, measured using patents per capita, and the different variables capturing religiousness. In this case, to address the endogeneity issue, I build an instrumental variable based on the spatial distribution of higher education institutions (universities and colleges) across the country, based on the idea that knowledge clusters will produce more innovation. The results, once again, display a negative relationship between innovation and religiousness that is robust to controlling for several state-level determinants of religion.

[5]Barro and McCleary (2003a,b); Guiso, Sapienza, and Zingales (2003); Huber (2005); Bénabou, Ticchi, and Vindigni (2015b) among others.

[6]From World Bank (2015) Development Indicators.

Related literature: This paper closely relates to Bénabou, Ticchi, and Vindigni (2015a) who unveil the existence of a strong negative relationship between religion and innovation measured using patents per capita and proceed to model a theoretical framework in which the two variables are endogenously determined by the interaction between the public, the government and religious organisations. In a later work, Bénabou, Ticchi, and Vindigni (2015b) provide empirical evidence in favour of a negative effect of religion on innovation showing how, at the individual level, *openness to innovation* is negatively related to religiousness. This paper differs by taking a diametrically opposite approach, seeking empirical evidence of determinants of religiousness in cross-country data. I use instrumental variables to attempt pinning down the effect on innovation on religiousness in the presence of reverse causality.

This paper is also related to the empirical literature that following Barro and McCleary (2003a) has tried to identify determinants of religion.[7] In particular, this paper contributes to the long-standing secularisation debate by proposing a different mechanism for the effect of economic development on religion, rather than increased existential security (Inglehart and Norris, 2004) or higher opportunity cost (Carvalho and Koyama, 2011). I argue instead that technological progress directly erodes religious beliefs and reduces the importance of religion.

This paper also fits within the literature that following the seminal paper by Azzi and Ehrenberg (1975), started applying economic tools to the study of religion. While most of the earlier works focused on modelling religious institutions,[8] more recently a vast empirical literature studying the relation between religions and several socio-economic indicators has flourished by either focusing on the effect of religious attitude in general (Barro and McCleary, 2003b; Gruber, 2005; Hungerman, 2005; Scheve and Stasavage, 2006) or on a specific religious denomination (Blum and Dudley, 2001; Galor and Moav, 2006; Becker and Woessmann, 2009; Cantoni, 2015; Woodberry, 2012). This paper focuses on the relationship between religion and innovation, rather than economic growth (Barro and McCleary, 2003b), economic attitude (Guiso et al., 2003), education (Glaeser and Sacerdote, 2008) or insurance (Scheve and Stasavage, 2006).

[7] See also Huber (2005); Mocan and Pogorelova (2017); Franck and Iannaccone (2014).

[8] For example, Azzi and Ehrenberg (1975); Stark and Bainbridge (1985); Sullivan (1985); Iannaccone (1992, 1994); Montgomery (1996); Finke and Stark (2005).

The discussion proceeds as follows: Section 3.2 presents the theoretical framework, Section 3.3 describes the data used in the analysis, Section 3.4 outlines the approach to the empirical analysis, Section 3.5 presents the results, Section 3.6 concludes.

3.2 Theoretical framework

In this section I present a simplified version of the model developed by Bénabou et al. (2015a), yet differentiated by ignoring the government and religious sectors and by making the cost of blocking innovation dependent on its speed. As such, the model is tailored to uniquely focus on how faster innovation negatively affects religiousness.

3.2.1 Representative agent

Let's assume a representative agent living over 2 periods with preferences

$$U_1 = \mathbb{E}_1[c_1 + \nu c_2 + r_2 g_2] \tag{3.1}$$

with (c_1, c_2) denoting the agent's consumption of the *secular good*[9] in the two periods and g_2 their consumption of the *religious good*. The agent's level of religiousness r_2, that can be thought as the intensity of religious beliefs, determines the utility they derive from the religious good. The agent maximises their utility and is uncertain about the future level of religiousness r_2 and technological progress, defined as a_2 that, without loss of generality, corresponds to the agent's income in period 2.[10]

3.2.2 Technological progress

In the first period some exogenous scientific discovery is made. Unless impeded, its spread will increase the level of technology $a_2 = (1+\gamma)a_1$ and reduce religiousness: $r_2 = (1-\delta)r_1$.[11] For simplicity I assume the process to be deterministic, so that equation

[9]In Bénabou et al. (2015a) where a continuum of agents is assumed, (c_1, c_2) are post-tax-and-transfer consumption. In a representative agent framework introducing taxes and transfers would be redundant.

[10]The model I propose circumvents the issue of endogenous preference. Another approach is to use a *imperfect empathy* framework, where the agent evaluates future utility based on their current utility function. See Bisin and Verdier (2011) for a detailed discussion.

[11]The model can be easily generalised to account for the fact that only some discoveries are detrimental to religion and factor in the probability of this type of scientific progress compared to that not

3.1 modifies to:

$$U_1 = c_1 + \nu c_2 + r_2 g_2$$
$$s.t. \quad c_1 + \varphi(\gamma) = a_1$$
$$c_2 + g_2 = a_2$$

Technological progress can be impeded (in period 1) by sustaining an up-front cost $\varphi(\gamma)$ that is increasing in the size of the innovation: the more attractive the idea, the harder it is to prevent it from spreading fast. Alternatively, to reconcile this intent with the representative agent framework, the more powerful the idea, the easier it is for the agent to embrace it. We also need to assume that $\varphi() \leq a_1$ or the agent would be unable to sustain the investment needed to block technological progress.

3.2.3 The agent's problem

In period two the agent chooses how to allocate their income over consumption of secular and religious goods. Given linear preferences, it is straightforward to see that they will consume only the good that they value more. Therefore, the allocation depends on the comparison between the marginal value of the secular good ν and the level of religiousness r_2. The optimal level of religious good in period two is:[12]

$$g(b_2, \nu) = \begin{cases} 0 & if \quad r_2 < \nu \\ a_2 & if \quad r_2 \geq \nu \end{cases} \tag{3.2}$$

In the first period, the only decision taken by the agent is whether to impede the innovation. Let's consider the two sub-cases:

Case $\boldsymbol{r_1 < \nu}$ When $r_1 < \nu$, the second period utility does not depend on religiousness, since r_2 can be at most equal to r_1, the value of the secular good is higher and the agent does not consume any amount of religious good.[13] Therefore, any investment made to impede innovation only reduces first period utility and it is wasteful. The agent optimally chooses:

contradicting the religious doctrine.

[12]Following Bénabou et al. (2015a), when $r_2 = \nu$ the indifference is broken in favour of the religious good without loss of generality.

[13]The model can be generalised to account for events that actually increase the level of religiousness.

$$(c_1, c_2) = (a_1, a_2) \quad g_2 = 0 \quad \chi_1 = 0 \tag{3.3}$$

where χ_1 represents the decision to block or not (taking value 1 for blocking, 0 for not blocking).

Case $r_1 \geq \nu$ When $r_1 \geq \nu$, second period utility is now increasing in the level of religiousness. There is a trade-off between higher income (given the assumption of income being equal to the level of technology a_2) and the value of the religious good. Stopping innovation does not decrease the latter but keeps income constant. On the other hand, enabling innovation increases second period income but reduces religiosity. We can define the agent's value function as a function of the blocking decision: $\chi_1 = \{0, 1\}$:

$$V(\chi_1, a_1, r_1) = a_1 - \chi_1\varphi(a_1) + \max\{\nu, r_1(1 - \delta(1 - \chi_1))\} \times a_1(1 + \gamma(1 - \chi_1)) \tag{3.4}$$

to split for simplicity in the value of blocking $V(\chi_1 = 1) \equiv V(B)$:

$$V(B) = a_1 - \varphi(\gamma) + r_1 a_1 \tag{3.5}$$

and the value of not blocking $V(\chi_1 = 0) \equiv V(NB)$:

$$V(NB) = a_1 + \max\{\nu, r_1(1 - \delta)\} \times a_1(1 + \gamma). \tag{3.6}$$

To characterise the solution it is important to distinguish two cases:

Case $(1 - \delta)r_1 < \nu$ In the first case, technological progress reduces religiousness up to the point where the secular good provides higher marginal utility than the religious good. In this case, assuming a reasonable range of parameters, it is optimal to block innovation when γ is not too large or r_1 is high. Therefore, when blocking occurs the agent's optimal choice modifies to:

$$(c_1, c_2) = (a_1 - \varphi(\gamma), 0) \quad g_2 = a_1 \quad \chi_1 = 1 \tag{3.7}$$

Otherwise, with no blocking, since $(1-\delta)r_1 < \nu$ the solution is the same as in 3.3.

Case $(1-\delta)r_1 \geq \nu$ In this case, even after technological progress occurs, the marginal utility from the religious good is higher than the one derived from the secular good. The agent's choice will depend on the speed of technological progress, the level of technology and religiousness.[14] In the case of blocking the solution is the same as in 3.7.

3.2.4 Religiousness and technological progress

Before proceeding with comparative statics, it is worth abstracting from the representative agent case in order to create a link with the empirical analysis. From now onward, I refer to r as the level of religiousness in one country.[15] a is the level of technology in the same country, and ν the level of secular values in the country. Also, I drop time indexation and use notations r and a for r_1 and a_1. A simple comparison between the two cases $r < \nu$ and $r \geq \nu$ is enough to show that faster technological progress is associated with less impediments to innovation that negatively affect religiousness. Focusing on the blocking decision in the case $(1-\delta)r < \nu$:

$$\text{block if} \quad r \geq \frac{\varphi(\gamma)}{a} + \nu(1+\gamma) \tag{3.8}$$

it is clear how, for given r and ν faster technological progress γ would increment the right-hand side of the equation, thus reducing blocking. As scientific discoveries reduce religiousness, technological progress and innovation should negatively influence religious faith. Moreover, my model also allows considering other possible mechanisms that will affect the endogenous determination of beliefs and technology. For example, we may consider other variables that directly affect the cost of blocking such openness to trade. If $\varphi(\gamma)$ modifies to $\varphi(\gamma, \tau)$ which is increasing in τ, the amount of trade. An increase in τ will positively affect $\varphi(a, \tau)$ and the right-hand side of the equation, reducing innovation blocking. As a conclusion, if the implications of the model are correct, the data should show a negative relationship between religiousness and technological progress.

[14]See Appendix 3.A for derivation.

[15]We may think about it as the average number of religious people in the country or the average intensity of the country population's religious beliefs.

3.3 Data

The data used in this paper is comprised of two parts. The first is cross-country data, while the second is on the 48 contiguous US states. The measures used in the two parts of the empirical analysis are similar but have different sources, discussed below.

3.3.1 Country-level data

The main data-set used in this paper consists of repeated cross-sectional data on 105 countries merging religiousness indicators obtained from country averages of individual responses to surveys on the topic of religion, with a series of economic measures as well as variables capturing the main religious affiliations and the interplay between secular and religious power. The data spans over six years (1985, 1990, 1995, 2000, 2005 and 2010)[16] corresponding to the the six available waves of the World Value Survey and the four waves of the European Value Study, henceforth WVS and EVS, for a total of 327 observations. The WVS is a collection of national level surveys conducted in different countries all over the world in different waves covering topics such as development, democratisation, gender equality, social capital, subjective well-being, and, obviously, religion. The EVS administers a very similar questionnaire to people living in Europe. The two studies are perfectly analogous for most questions.[17] If a country was surveyed in both studies for the same wave I keep the observation from the EVS.

Religion intensity: I use four different indicators of the intensity of religiousness from responses to the WVS/EVS questionnaires: *religiosity*, *beliefs in God*, *weekly church attendance* and *importance of religion*. Religiosity is measured as the share of people in one country that to the question: *"Independently of whether you attend religious services or not, would you say you are: 1) A religious person; 2) Not a religious person; 3) An atheist"* answered 1). Similarly, Beliefs in God is the share of those responding *"Yes"* to

[16] Not all countries are surveyed in the same year, I attribute a specific year from the range to indicate a joint WVS-EVS wave. The WVS waves are: 1981-1984, 1989-1993, 1994-1998, 1999-2004, 2005-2009 and 2010-2014. For the EVS: 1981-1984, 1990-1993, 1999-2001 and 2008-2010. Table 3.1 provides an overview of the number of countries in each wave. See European Value Study (2011); World Values Survey Association (2015) for further details.

[17] This is not surprising as WVS builds on the EVS. On both websites it is possible to find instructions on how to merge the two data-sets.

"Do you believe in God?"; Weekly Church Attendance equals the share of responses *"1) More than once a week"* or *"2) Once a week"* to the question: *"Apart from weddings and funerals, about how often do you attend religious services these days?"*[18] Finally, Importance of Religion is defined on a scale from 1 *"Less important"* to 4 *"Very important"*.[19] I will mainly focus on *religiosity* and *importance of religion* for the following reasons. First, the role and importance of God varies dramatically between monotheistic and polytheistic religions. On the other hand, church attendance despite being widely used in the literature (Barro and McCleary, 2003b; Scheve and Stasavage, 2006; Aleksynska and Chiswick, 2013) is a shallower measure of religiosity (Inglehart and Norris, 2004) as it may depend on network effects (Gruber, 2005; Huber, 2005) and other *material* or *social* motivations supplementary to the spiritual ones (Iannaccone and Berman, 2006; Levy and Razin, 2012).

Innovation: The main explanatory variable is innovation. To measure it, I use patent applications as a proxy following Bénabou et al. (2015b), using data from the World Intellectual Property Organization (2015) on the total patent application filed in a country by residents. I also introduce an alternative measure of innovation based on the logarithm of the number of scientific and technical journal articles per capita.[20] Finally, I introduce two variable that I use to instrument patents per capita: the index of patent protection developed by Park (2008) and the 2010 International Property Rights Index developed by Property Rights Alliance (Strokova, 2010).

Other controls: I introduce a series of country variables to control for other factors that can affect the level of religiousness of the population. In particular, I record the predominant religion, past levels of religious adherence, government regulation of religion, religious plurality, the presence of a communist regime and a series of socio-economic control variables including GDP per capita, immigration, population, trade, education and

[18] Other potential answers from the questionnaire were: *"3) Once a month; 4) Only on special holy days; 5) Once a year; 6) Less often; 7) Never, practically never"*. Formulations of the options 3) to 5) slightly differ in the EVS. However, this does not affect the consistency of the variable across the two surveys given how it is defined.

[19] The original variable is in inverted scale and has been re-coded.

[20] The data on the number scientific and technical journal articles and countries' population is derived from the World Bank (2015).

urbanisation. I use data on religious affiliation shares from the World Religion Dataset (Maoz and Henderson, 2013) to compute dummies for religious predominance, assigning the variable a value equal to 1 if more than 40% of the population adheres to a specific religion.[21] I borrow a series of variables that describe the regulation of religion in the country from Barro and McCleary (2003b).[22] In particular, I use dummy variables indicating: the presence of a state religion in 1970;[23] the existence of regulation of religion in the country; the presence of a Communist regime in 1970; and, for the same year, an indicator of religious plurality.[24] From the World Bank Development Indicators database (World Bank, 2015) for each country I collect GDP per capita, the share of immigrants as a fraction of the total population, the share of urban population, total population, trade as a fraction of GDP, FDI as a fraction of GDP, the expenditure on research and development as a fraction of GDP and the number of researchers for 100 million people in the country.[25] I also use the total capital stock and total factor productivity from the Penn World Tables (Feenstra et al., 2013). Finally, to measure the level of human capital I include educational attainment as the average years of secondary and tertiary schooling in the country from Barro and Lee (2013).

3.3.2 US state-level data

The US state-level religiousness variables are derived from the Religious Landscape Study by Pew Reseach Center (2016) and refer to 2014.[26] The three religiousness indicators I collect are: importance of religion, beliefs in God and weekly church attendance. These variables are practically identical to those constructed from responses to the EVS and

[21]In doing so, I consider all smaller Protestant denominations as Protestant.

[22]Religious Adherence Data from Barro (2003).

[23]Barro (2003) constructs such indicators for three different years: 1900, 1970 and 2000. I choose 1970 to be able to capture the long-lasting effect of communist regimes on the religious landscape of a county (Minarik, 2014) and, at the same time, use an observation preceding the waves of the EVS/WVS surveys.

[24]Namely, the Herfindahl index of religious concentration, a proxy for the competitiveness of the religious market that, according to the *religious market theory* should be a strong predictor of religious participation (Finke and Stark, 1988; Finke et al., 1996; Iannaccone et al., 1997). This measure and the literature based on evidence deriving from it was criticised by Chaves and Gorski (2001) and Voas et al. (2002) who explain how by construction it tends to be correlated with measures of religiosity by a mechanical relationship. This does not pose any issues in this paper as religiousness variables and the plurality index come from different sources and have not been computed from the same raw data.

[25]All variables are recorded for all years available between 1980 and 2014.

[26]The data consists in responses to a telephone survey administered to a representative sample of around 35000 adults living in the US (Pew Reseach Center, 2016).

WVS. They are built using the same strategy as that outlined in 3.3.1 from responses to analogous questions. The measure of innovation I use is once again patents per capita. The state-level data on patent applications per 1000 people is published by the United States Patent and Trademark Office (USTPO, 2015). I control for other state-level variables that could affect the level of religiousness in the state. I use US Census Bureau data on income per capita, educational attainment, state population and the urbanisation rate.[27] Then, I also control for the size of the state and the average percentage of Republicans in the two houses composing the state legislature from the Book of States (CSG, 2018).

3.4 Empirical strategy

I use the following reduced form model to test for the influence of innovation on religion:

$$reli_{it} = \beta_0 + \beta_1 I_{it}^{(4)} + \boldsymbol{\beta_2}\mathbf{X_{it}} + \boldsymbol{\beta_3}\mathbf{R_{it}} + \boldsymbol{\beta_4}\mathbf{W_i} + \gamma_\mathbf{t} + \epsilon_\mathbf{it} \quad (3.9)$$

where $reli_{it}$, the dependent variable, is a measure of religiousness and innovation. The main explanatory variable is defined as:

$$I_{it}^{(4)} = Log\left(\frac{1}{4}\sum_{\tau=t-4}^{t-1} P_{i\tau}\right) \quad (3.10)$$

which is the logarithm of the average of the lagged values over the last four years of P_{it}, patents per capita. There are two main advantages of using this measure compared to $LogP_{it}$. The first is that the average smooths out potential idiosyncratic yearly variations. The second is related to the sample size. Indeed, if there is a gap in the time series of patents per capita in a country, I am still able to include it in the analysis as long as the lagged values are available. $\mathbf{X_{it}}$ is a vector of socio-economic country-level control variables including GDP per capita[28], immigration as a fraction of total population, urbanisation,

[27]Income, educational attainment and population are from the American Community Survey 2012-2016 estimates (U.S. Census Bureau, Geography Program, 2016). Educational attainment refers to the population above 25 and is measured as the fraction of people with a graduate degree. The urbanisation rate is derived from the 2010 census data.

[28]I compute the log of the average of 4 lagged values of GDP per capita, using the same formula from Equation 3.10 to control for yearly fluctuations in GDP. I do not use the same approach for the other variables in $\mathbf{X_{it}}$ as they do not fluctuate as much as GDP per capita over time and there is no gain in terms of sample size.

education, trade and foreign direct investment both measured as a fraction of GDP. Such variables have been found in the literature to be determinants of religiousness. Higher income should reduce religiousness according to the classic theory of secularisation as higher standard of living and higher existential security make people in less need of religion (Inglehart and Norris, 2004). Urbanisation also is a negative predictor for two reasons: first, the social aspect of religion may be more salient in rural areas; second, the urban environment offers a larger menu of alternative activities (McCleary and Barro, 2006). Educational attainment is also found to be associated with lower religiousness in empirical studies (Hungerman, 2014; Mocan and Pogorelova, 2017). Immigration, instead, should have the opposite effect as immigrants tend to be more religious than the natives as religion is a way of affirming their ethnic and cultural diversity (Alesina et al., 2003; Aleksynska and Chiswick, 2013; Opfinger, 2014). I also include trade and FDI because more openness could also negatively affect religiousness. $\mathbf{R_{it}}$ is a vector of dummies indicating if one of four of the main denominations Protestantism, Islam-ism, Catholicism and Orthodox Christianity, is the predominant religion.[29] $\mathbf{W_i}$ is a vector of time-invariant characteristics such as the presence of a communist regime, a state religion, regulation of religion and the level of religious plurality.[30] I control for communist regimes as they promoted State atheism and tried to corner religious organisations (Minarik, 2014). Conversely, state religions are associated with higher religiousness (Barro and McCleary, 2003a). The effect of plurality is instead ambiguous: while secularisation predicts that the availability of more religious denominations will lead people to cast doubts on their faith (Martin, 1978), more competition could actually force religious organisation to perform better and attract more followers. Finally, I include a time fixed-effect, γ_t, to control for possible time trends of the dependent variable.

I start by testing the model in Equation 3.9 as pooled OLS with standard errors clustered at country level. However, a simple Breusch and Pagan (1980) test shows that country fixed effects have non-zero variance estimates which means they will be biased. Also, using panel data analysis does not seem appropriate given the characteristics of

[29] I exclude Buddhism and Hinduism as in the richest specifications only 4 countries (India, Japan, Thailand and Vietnam) have predominance of one of such religions.

[30] All such measured are derived from Barro and McCleary (2003a) who record them in 1900, 1970 and 2000 with the exception of the index of regulation of religion. I use the 1970 observations as they are closer but precede the dependent variables records.

this panel. It is, indeed, unbalanced, with the sample changing at each time period, and over-representation of developed countries always present since the earlier EVS/WVS.[31] Also, religious outcome variables slowly change over time compared to the explanatory variable of interest, meaning that it has little explanatory power over a short time span. The solution I employ is to transform the data-set in a cross-sectional panel by only using the latest observation for each country in the sample. Figure 3.1 provides an overview of the countries included in the sample and, for each country, what survey wave is used. Equation 3.9 then modifies to:

$$reli_i = \beta_0 + \beta_1 I_i^{(10)} + \boldsymbol{\beta_2}\mathbf{X_i} + \boldsymbol{\beta_3}\mathbf{R_i} + \boldsymbol{\beta_4}\mathbf{W_i} + \boldsymbol{\beta_5}\boldsymbol{\omega_i} + \epsilon_i \tag{3.11}$$

where the main explanatory variable measuring innovation, $I_i^{(10)}$, is computed as in Equation 3.10 but using a 10-year average, which betters fits the slow pace of the belief erosion mechanism.[32] The vectors X_i, R_i and W_i are defined as above.[33] $\boldsymbol{\omega_i}$ is a vector of dummies controlling for the wave of the survey that includes the country i's observation.

Identifying the causal effect of innovation on religions is a complex task. The theoretical model shows that innovation and religiousness are endogenously determined. Also, both variables have evolved over a long time-frame (centuries for religion) and their evolution may depend on other factors unaccounted in the model from Equation 3.11. For this reason, I employ instrumental variable analysis in an attempt to more convincingly pin down the direction of causality from innovation to the measures of religiousness. In choosing the instrument, I focus on patent protection, a variable that has an effect on the number of patents per capita but has no clear direct influence on the level of religion, thus arguably exogenous. I estimate the model:

$$\begin{aligned} I_i^{(10)} &= \alpha_0 + \alpha_1 PatProt_i + \boldsymbol{\alpha_2}\mathbf{X_i} + \boldsymbol{\alpha_3}\mathbf{R_i} + \boldsymbol{\alpha_4}\mathbf{W_i} + \boldsymbol{\alpha_5}\boldsymbol{\omega_i} + \xi_i \\ reli_i &= \beta_0 + \beta_1 \hat{I}_i^{(10)} + \boldsymbol{\beta_2}\mathbf{X_i} + \boldsymbol{\beta_3}\mathbf{R_i} + \boldsymbol{\beta_4}\mathbf{W_i} + \boldsymbol{\beta_5}\boldsymbol{\omega_i} + \epsilon_i \end{aligned} \tag{3.12}$$

[31] See Table 3.1 in the Appendix for further detail.

[32] In the pooled OLS I chose a 4-year average because there is a 5-year gap between successive waves of the EVS/WVS. Using a longer time span would mean an overlap between different observations of the independent variable, innovation.

[33] The only difference is that the lagged average of GDP per capita is now computed over 10 years, similar to innovation.

where $PatProt_i$ is the excluded instrument and all the other right-hand side variables are defined as above. A legitimate concern about the instrument chosen is related to the direction of causality between patent protection and innovation. While individuals and firms are more encouraged to innovate in countries where the product of their work can be better protected and, consequently, monetised, it is also likely that some countries enforce stricter patent protection because of the high volume of innovation produced by domestic firms. If the endogenous variable *causes* the instrument, the latter will not be valid. In this perspective, a potential alternative class of instruments is given by *institutional characteristics*. However, institutions often affect religiousness through other mechanisms than innovation, thus violating the exclusion restriction.[34] Moreover, patent protection could still be preferred due to being the more *relevant* instrument. I employ a *best of two worlds* approach by introducing another instrument that is correlated with patent protection but is related to a broader spectrum of institutional features, without obvious problems of endogeneity. I use Property Rights Alliances's International Property Rights Index (Strokova, 2010). The IPRI is a weighted average of several indicators that determine the level of property rights protection in the country. Patents protection is only a partial component of the index, which instead is also composed of measures of legal protection of property rights as well as the level of protection of tangible goods and all other categories of intangibles.[35] As such it is correlated with patent protection but arguably not caused by innovation.

3.4.1 US state-level analysis

I also investigate the existence of the same relationship across US states, using the measures of religiousness from the Pew 2014 Religious Landscape Study. For consistency with the country-level analysis, I measure innovation using the logarithm of the 10-year average (2005 to 2014) of state-level patents per capita from the USTPO (2015). I estimate the

[34] This is the case for a history of communism Minarik (2014).

[35] The IPRI is composed of 3 major categories: Legal and Political Environment (LP), Physical Property Rights (PPR) and Intellectual Property Rights (IPR). Each of them is broken down in sub-components with an assigned score. The sub-components are: Judicial Independence, Political Stability, Rule of Law and Corruption for LP; Protection of Physical Property Rights, Registering Property and Access to Loans for PPR; Protection of Intellectual Property Rights, Patent Protection and Copyright Piracy for IPR (Strokova, 2010).

following model:

$$reli_s = \beta_0 + \beta_1 I_s^{(10)} + \boldsymbol{\beta_2}\mathbf{X_s} + \boldsymbol{\beta_3}\mathbf{r_s} + \epsilon_s \tag{3.13}$$

$I_s^{(10)}$ is the logarithm of the 10-year average of patents per capita. $\mathbf{X_s}$ is a vector of state-level variables including the logarithm of income per capita, the logarithm of state population, the logarithm of state area, the state-level average educational attainment, the proportion of people living in urban areas in the state and the average share of Republicans between the two legislative houses. $\mathbf{r_s}$ is a vector of region fixed-effects, including north-east, mid-west, south and west. As for the country-level strategy, I employ an instrumental variable to try to pin down the direction of causality from innovation to religion. In this case, I build a measure based on the spatial distribution of universities and colleges in each state.[36] While the number of education institutions may be endogenous, their location relative to each other is hardly dependent on the religiousness in the state. To capture the proximity of universities between each other I use the following strategy. First, for each university u_i I generate a polygon such that all points in the polygon are closer to university u_i rather than any other university. Formally, the polygon associated with university u_i is:

$$E_i = \{x \in X \mid d(x, u_i) < d(x, u_j) \; \forall \; j \neq i\} \tag{3.14}$$

where x is any point in the set X of all locations in continental US and $d(a, b)$ is the Euclidean distance between locations a and b. For each state, I then take the average of the areas of all such polygons and normalise it by the number of universities in the state. If universities are clustered, the polygons are going to be smaller on average. Therefore, the average polygon's area is going to be smaller as well. The normalisation is needed because when there are more universities in a state, the space will be split in more polygons that will mechanically be smaller. As the measure is also affected by state size, I will control for it accordingly. The idea behind such measure is that just like in the case of firms,[37] clusters of universities generate positive spillovers that lead to more

[36] I only focus on institutions with more than 1000 enrolled students. The data on the location of universities is retrieved from the US Department of Education.

[37] See (Carlino and Kerr, 2015).

and faster innovation.[38] The model I proceed to estimate is then:

$$\begin{aligned} I_s^{(10)} &= \alpha_0 + \alpha_1 V_s + \boldsymbol{\alpha_2}\mathbf{X_s} + \boldsymbol{\alpha_3}\boldsymbol{r_s} + \xi_s \\ reli_s &= \beta_0 + \beta_1 \hat{I}_s^{(10)} + \boldsymbol{\beta_2}\mathbf{X_s} + \boldsymbol{\beta_3}\boldsymbol{r_s} + \epsilon_s \end{aligned} \tag{3.15}$$

where V_s is the excluded instrument built as explained above and all other right-hand side variables are the same defined in model 3.14. Similarly to the cross-country analysis, the choice of the instrument deserves further discussion. A higher level of education is often associated to lower religiousness (Glaeser and Sacerdote, 2008; Hungerman, 2014; Mocan and Pogorelova, 2017) and some religious beliefs are in open contrast with standard academic curricula.[39] Thus, it may be possible that in less religious states there would be a lower demand for academic institutions and as a result less universities. This concern, however, is mitigated by the fact that most of the oldest colleges in the US were founded by religious congregations.[40] Another more pressing issue relates to the exclusion restriction. It is indeed likely that the proximity to universities affects religiousness via different mechanisms than through innovation. For example, similarly to innovation, established academic notions in contrast with religious teachings may directly erode beliefs and spillover to individuals not actually enrolled in higher education. Another potential mechanism is migration: if more religious individuals shun academia, states with more universities will see a influx of irreligious people. The instrument partially addresses these concerns by focusing on the concentration rather than the amount of academic institutions and it is normalised by the number of universities in the state. However, if there still existed a direct, albeit less obvious, effect of university clusters on state level of religiousness the exclusion restriction would be violated.

[38] The literature has recently uncovered the role of institutions and universities in particular as determinants of innovation (Carlino and Kerr, 2015), while the traditional discussion on natural advantages has focused on factors such as the presence of natural harbours or coal deposits. In particular, there is evidence that the presence of universities increases the number of patents in a given area (Andersson, Quigley, and Wilhelmsson, 2009).

[39] A case in point is the ongoing fight between evolution and creationism in the US (Berkman and Plutzer, 2010).

[40] The first institutions of higher learning in the US were Harvard, Yale (founded by the Puritans) and William and Mary (founded by the Anglicans). One of the objectives of the early colleges was advancing the study of religion and training a new class of ministers, despite not being seminaries (Geiger, 2016).

3.5 Results

In this section I will first discuss the results from estimating the analysis using country-level data, then proceed to look at the US state-level analysis.

3.5.1 Country-level analysis

The strong negative correlation between innovation and religiousness that was the original motivation for Bénabou et al. (2015a) is very clearly depicted in Figure 3.2 using all four indicators of religiousness: religiosity, importance of religion, beliefs in God and weekly church attendance. Table 3.2 shows a strong significant negative relationship between the lagged average of the innovation measures and two different indicators of religiousness: religiosity and importance of religion. The point estimates are affected by introducing controls for the state's regulation of religion, waves of the surveys as well as for all other potential determinants of religiousness. In the richest specifications, a 10% increase in patents per capita is associated with a reduction of approximately 0.36 percentage points in the amount of people qualifying themselves as religious and 0.012 standard deviation (σ) fall in the importance of religion.[41] The sign of the estimates is consistent, but the point estimates are smaller and the standard errors relatively larger in the richer specifications. The effects may appear small but are actually important when we account for the large variation in patents per capita across countries. Indeed, a σ increase in the explanatory variable (log average patents per capita) is associated with a 8 percentage points decrease in religiousness. This is just less 1/2 σ fall in religiousness.[42] Similarly, the same ($+\sigma$) change in the explanatory variable leads to a fall in the importance of religion equal to 1/3 σ. A better way to understand the significance of such magnitudes is to look at the within country evolution of religiousness. For example, a 8 percentage points fall in religiousness is equal to the mean decrease in religiousness from the first survey entry to the last one for all those countries that became less religious over time. In all regressions, the other controls display coefficients that, whenever statistically significant, have the expected sign according to the discussion from section 3.4. The neg-

[41] For completeness, I also provide the results of the pooled OLS regression from Equation 3.9 in Table 3.B1.

[42] These values refer to the sample used in the models presented in Table 3.2.

ative relationship between innovation and religiousness exists when also using different measures such as beliefs in God and weekly church attendance.[43] I perform a series of additional robustness checks to confirm the validity of the results. In particular, I cluster standard errors by geographical areas; I use years of secondary education from Barro and Lee (2013), *in lieu* of tertiary, to measure educational attainment; I control for the presence of state religion and plurality in the year 2000 rather than 1970; I change the measure of *predominant religion* to religious fraction with more than 50% of adherents rather than 40%. In all cases the results are not qualitatively affected.[44] I also replicate the same model using a different measure of innovation, scientific articles per capita.[45] The sign of the coefficient is negative in all specifications and also significant at a 5% level when importance of religion is the dependent variable. In the richest specifications, a 10% increase in scientific articles per capita is associated with a reduction of approximately 0.086 percentage points in the amount of people qualifying themselves as religious and a 0.014 standard deviations (σ) in the importance of religion.[46] When I use the patent protection index as an instrumental variable (Table 3.4 Panel A), the estimate coefficients are consistent with the OLS ones in models 1-2, 5-6 and larger in the richest specification.[47] The standard errors are instead uniformly larger, which is consistent with inference based on weak instruments. The sign of the estimates shows the expected negative relationship in all models. In Panel B I reproduce the same regressions using the second instrument, the International Property Rights Performance Index. The estimates are larger than in Panel A, while the standard errors are not, hence the coefficients are more precise even in presence of a weak instrument (the efficient first stage F-statistics is 4.951). Once again, the estimated coefficients are negative and consistent with the hypothesised relationship between innovation and religiousness.

[43] See Table 3.B2 for reference.

[44] See Tables 3.B3 and 3.B4 in Appendix 3.B.

[45] Similar to patents per capita, I use the logarithm of the 10 previous years' average, as in Equation 3.10.

[46] See table 3.B5 for the same regressions using beliefs in God and weekly church attendance as dependent variables.

[47] The composition of the sample has changed as patent protection data are not available for the whole set of countries. When we compare the coefficients to the equivalent OLS estimates using the same sample, the 2SLS point estimates are still larger but the magnitude of the increase is lower. See Table 3.B6 for the OLS regression using only the countries for whom patent protection data is available.

3.5.2 US state-level analysis

Figure 3.4 shows instead the existence of a similar relationship at US-state level between innovation measured using patents per capita and three different measures of religiousness: importance of religion, beliefs in God and weekly church attendance. The estimates from Table 3.5 confirm that higher innovation is negatively associated with lower religiousness even when introducing a rich set of controls that are potential determinants of religiousness. In the richer specification from column 2, a 10% increase in the amount of patents per 1000 people in the state is associated with approximately a 0.05 σ fall in the importance of religion. Similarly, columns 4 and 6 show that for a 10% increase in state-level patents per 1000 people there is a 0.34 percentage point drop in the amount of people who believe in God and an approximately 0.38 percentage point fall in the amount of people who attend a religious service weekly. Another way to better understand the magnitude of these effects is looking at a σ change in the explanatory variable of interest. That is associated to a fall in the importance of religion that is equal to 1/2 σ and 1/7 of the whole interquartile range. For beliefs in God it translates to a fall of approximately 1/2 σ and 1/5 of the interquartile range; for weekly church attendance the associated reduction is approximately 40% σ and 10% of the in interquartile range.

I then introduce the instrumental variable based on the location of higher education institutes in a state and, in particular, their proximity. The estimates still show the existence of a clear and statistically significant negative relationship between patents per capita and religiousness. The coefficients in columns 1, 3, 5 are consistent with the OLS estimates. The coefficients in the richer specifications, however, are much larger and need to be interpreted with caution because of the weakness of the instrument and the difficulty in determining the direction of the bias of the OLS estimate.[48]

[48] Table 3.B7 shows the reduced form IV regression. We can see that the instrumental variable based on the proximity of higher education institutions has a positive significant relationship with the dependent variable, regardless of the measure of religiousness used. The sign of the coefficients is as expected, given that larger areas on average imply that universities are further away from each other, hence there is less clustering.

3.6 Conclusion and directions for future research

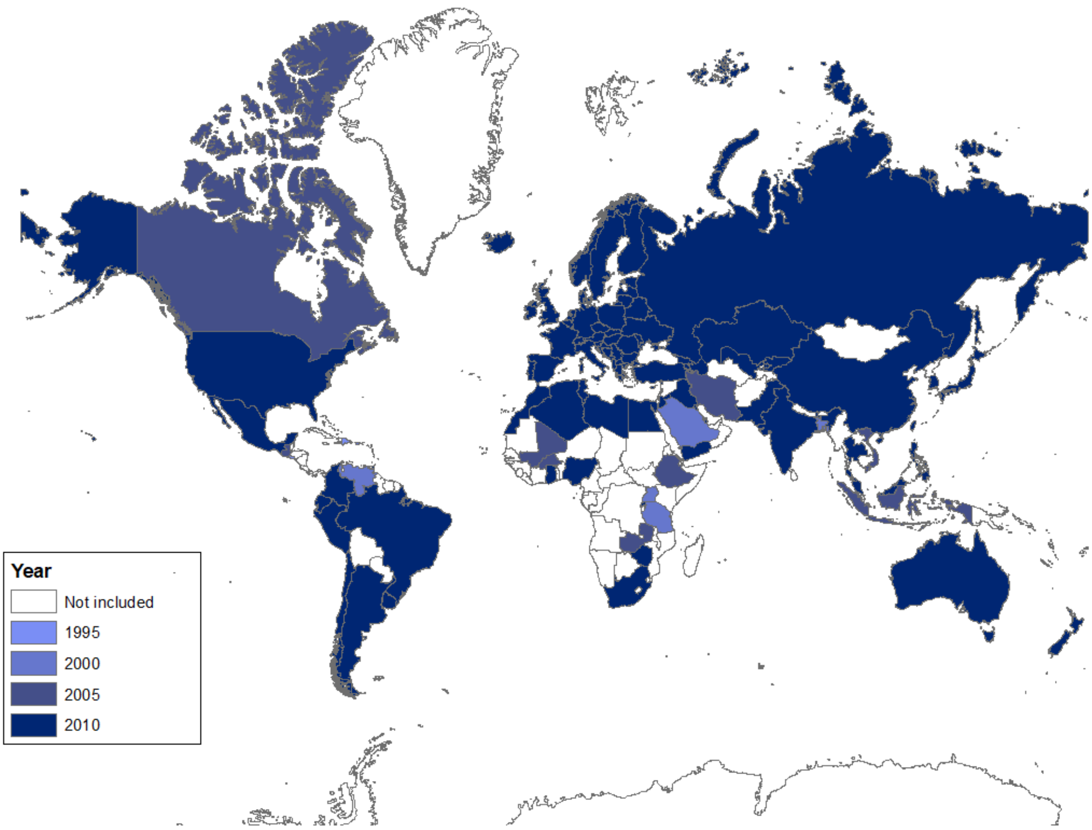
This paper empirically investigates the relationship between religion and innovation as originally uncovered by Bénabou, Ticchi, and Vindigni (2015a), yet using different measures of innovation and control variables at both cross-country and US state-level. I find clear evidence of the existence of a negative relationship between innovation and different measures of religiousness that is robust to different specifications and to controlling for other potential determinants of religiousness.

In a context where religiousness and innovation have co-evolved across countries over centuries and are both affected by other factors such as institutions, identification proves challenging. To address this problem, I build instrumental variables based on the level of patent and property rights protection in a country or the proximity of higher education institutions to each other. The instrumental variable analysis confirms the existence of a negative relationship between innovation and religion, although the estimates should be interpreted with caution due to the weakness of the instruments used.

Understanding the effect of innovation on the level of religiousness in a country contributes to the long-standing discussion on *secularisation* by providing a new mechanism that explains the drastic reduction in religiousness in the Western world. More generally, this paper provides empirical evidence on innovation as one of the determinants of religiousness. This has been understudied due to the lack of comprehensive data and sources of exogenous variation. As more countries are being covered by surveys such as the EVS and WVS, there will be an opportunity for future research to track how the changes in the level of religiousness across countries are influenced by different factors, such as innovation, and expand the analysis to a larger set of countries.

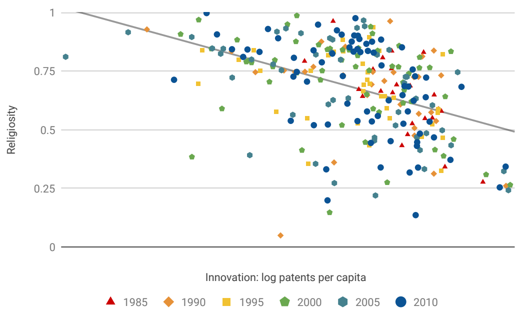
3.7 Figures

Figure 3.1: Countries in the EVS/WWS waves

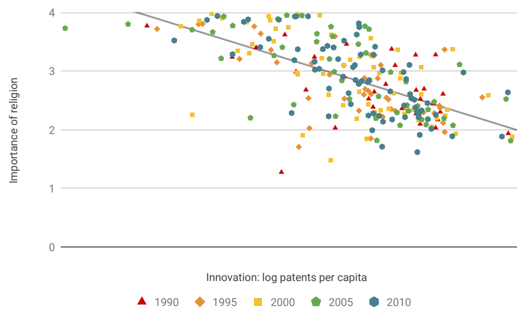


World countries included in the sample used in this paper. The colour indicates the latest wave in which each country is present (up to 2010).

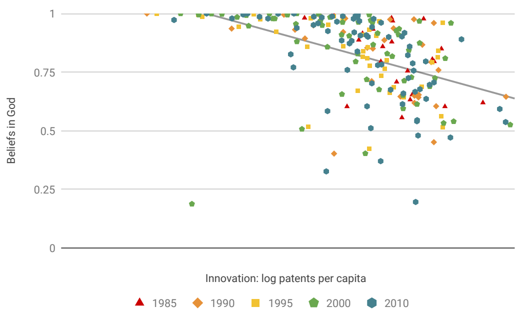
Figure 3.2: Different measures of religion and innovation



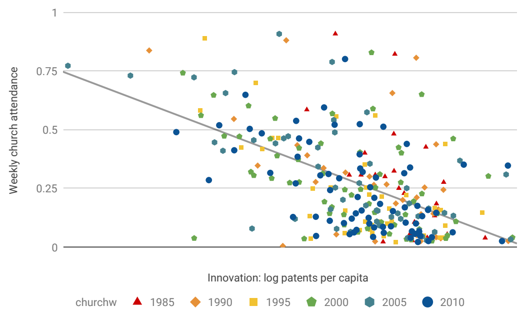
(a) Religiosity and innovation



(b) Importance of religion and innovation



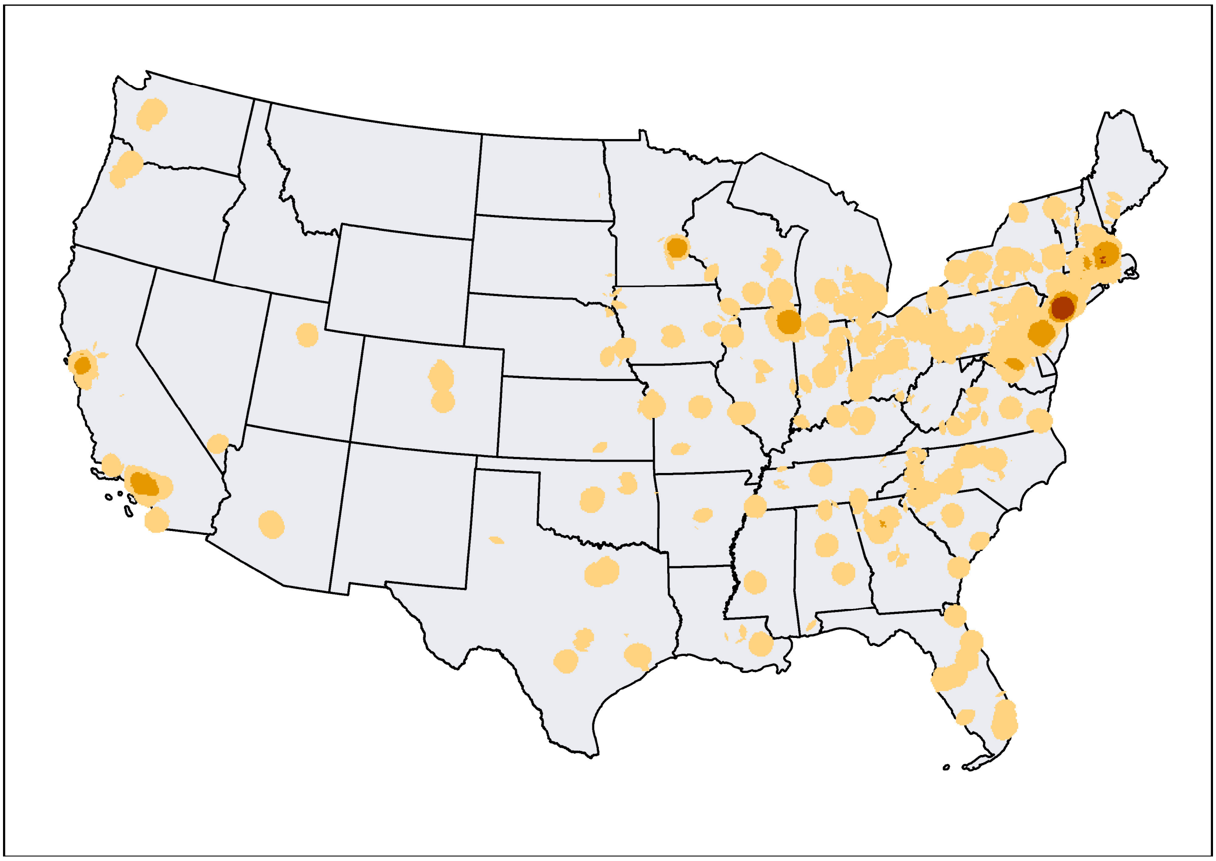
(c) Beliefs in God and innovation



(d) Weekly church attendance and innovation

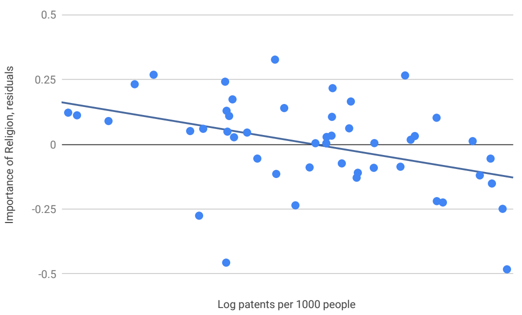
On the y-axis: one of the four religiousness measures, religiosity in 3.2a, importance of religion in 3.2b, beliefs in God in 3.2c, weekly church attendance in 3.2d on the logarithm of state income per capita. On the x-axis: the logarithm of patents per capita. The graph pools observations from different survey waves, indicated by different colours.

Figure 3.3: Density of higher education institutions in the US

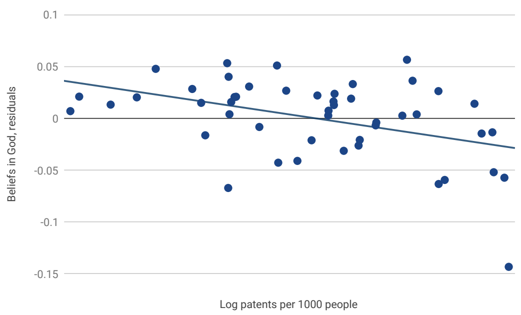


Density of US universities, computed as point density per km^2 in a 50-km neighbourhood around each institution. The values are classified in four categories using Jenks' natural breaks; a darker colour is associated with higher density.

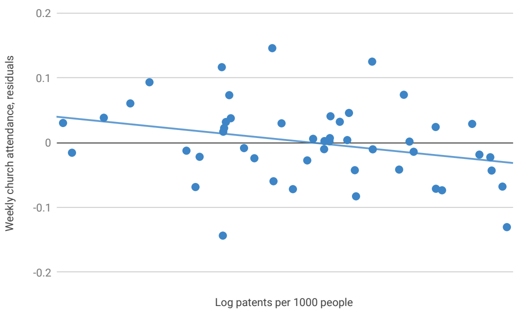
Figure 3.4: Different measures of religion and innovation



(a) Importance of religion and innovation, US states



(b) Beliefs in God and innovation, US states



(c) Weekly church attendance and innovation, US states

On the y-axis: the residuals from regressing one of the three religiousness measures, importance of religion in 3.4a, beliefs in God in 3.4b, weekly church attendance in 3.4c on the logarithm of state income per capita. On the x-axis: the logarithm of patents per 1000 people in the state.

3.8 Tables

Table 3.1: Observations per each EVS and WVS wave

Wave	**EVS**	**WVS**	**Total**
1985	15	8	23
1990	28	14	42
1995	-	51	51
2000	32	37	69
2005	-	56	56
2010	44	42	86
Total	**119**	**208**	**327**

The table reports the number of countries in each wave of the EVS and WVS surveys.

Table 3.2: Religion and innovation, country-level

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0427***	-0.0344***	-0.0380**	-0.209***	-0.129***	-0.0842***
	(0.0076)	(0.0096)	(0.015)	(0.022)	(0.026)	(0.030)
Log GDP per capita			0.00860			-0.0632
			(0.020)			(0.054)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.220	0.447	0.659	0.488	0.714	0.765
Observations	87	82	74	88	83	75

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-10$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-10$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.3: Religion and innovation using scientific articles, country-level

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Log articles per capita	-0.0429***	-0.0357***	-0.00901	-0.196***	-0.154***	-0.0963**
	(0.0070)	(0.0085)	(0.019)	(0.021)	(0.026)	(0.047)
Log GDP per capita			-0.0191			-0.0535
			(0.022)			(0.073)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.230	0.426	0.525	0.402	0.706	0.747
Observations	101	96	85	102	97	86

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log articles per capita is the logarithm of the average number of scientific articles and technical reports from years $t-10$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-10$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.4: 2SLS Religion and innovation

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Log patents per capita	-0.0418***	-0.0298*	-0.0560	-0.250***	-0.164***	-0.202
	(0.0095)	(0.017)	(0.053)	(0.031)	(0.045)	(0.14)
Observations	70	68	66	71	69	67
Panel B						
Log patents per capita	-0.0761***	-0.0488***	-0.0768*	-0.259***	-0.212***	-0.371**
	(0.013)	(0.013)	(0.047)	(0.030)	(0.037)	(0.15)
Observations	82	78	72	83	79	73
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. 2SLS regression where the dependent variable is: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-10$ to $t-1$. In Panel A he excluded instrument is the patent protection index from Park (2008). In Panel B the excluded instrument is the international Property Rights Index from Strokova (2010). *Controls* include the log of the average of GDP per capita from years $t-10$ to $t-1$, education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions. The robust first stage F-statistics in models 3 and 6 is 6.935 in Panel A, 4.951 in Panel B.

Table 3.5: Religion and innovation, US states

	Dependent variable:					
	Rel Importance		Belief in God		Church weekly	
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation	-0.115***	-0.125***	-0.0273***	-0.0358***	-0.0350***	-0.0399***
	(0.034)	(0.038)	(0.0058)	(0.0094)	(0.012)	(0.015)
All controls		✓		✓		✓
Area FE	✓	✓	✓	✓	✓	✓
R^2	0.747	0.824	0.697	0.771	0.573	0.681
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variable: state average of importance of religion ranked on a 1-4 scale in columns 1-2, average number of people in the state declaring to believe in God in columns 3-4, average number of people in the state declaring to attend church at least once a week in columns 5-6. Innovation is the logarithm of the average patents per 1000 people from 2005 to 2014. *All controls* include: the log of the average state income per capita, educational attainment as percentage of people older than 25 with a graduate degree, the state urbanisation rate, the logarithm of state population, the logarithm of state area, the average share of Republican legislators across the state lower and upper houses. *Area FE* are fixed effects for regions of the contiguous US: north-east, mid-west, south and west. In columns 2, 4 and 6, Nebraska is excluded as the data on the number of Republicans is missing since the state legislature is composed of a single non-partisan house.

Table 3.6: 2SLS Religion and innovation, US states

	Dependent variable:					
	Rel Importance		Belief in God		Church weekly	
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation	-0.0699**	-0.437*	-0.0290***	-0.117**	-0.0311*	-0.249*
	(0.029)	(0.25)	(0.0047)	(0.051)	(0.017)	(0.13)
All controls		✓		✓		✓
Area FE	✓	✓	✓	✓	✓	✓
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variable: state-average of importance of religion ranked on a 1-4 scale in columns 1-2, average number of people in the state declaring to believe in God in columns 3-4, average number of people in the state declaring to attend church at least once a week in columns 5-6. The excluded instrument is the average size of the polygon of all points that are closest to a specific university, weighted by the number of universities in the state. Innovation is the logarithm of the average patents per 1000 people from 2005 to 2014. *All controls* include: the log of the average state income per capita, educational attainment as percentage of people older than 25 with a graduate degree, the state urbanisation rate, the logarithm of state population, the logarithm of state area, the average share of Republican legislators across the state lower and upper houses. *Area FE* are fixed effects for regions of the contiguous US: north-east, mid-west, south and west. In columns 2, 4 and 6, Nebraska is excluded as the data on the number of Republicans is missing since the state legislature is composed of a single non-partisan house. The first-stage robust F-statistics is 9.45 in 1, 3, 5. It falls to 2.590 when adding other regressors in 2, 4, 6.

3.A Derivations

In Section 3.2 it was shown how when $r_1 \geq \nu$ there is a trade-off between higher income and the value of the religious good and the agent's value function can be written as:

$$\times a_1(1 + \gamma(1 - \chi_1)) \tag{3.16}$$

and it was split for simplicity in the value of blocking $V(B)$:

$$V(B) = a_1 - \varphi(\gamma) + r_1 a_1 \tag{3.17}$$

and the value of not blocking $V(NB)$:

$$V(NB) = a_1 + \max\{\nu, r_1(1 - \delta)\} \times a_1(1 + \gamma). \tag{3.18}$$

The agent will, therefore, opt for blocking when $V(B) \geq V(NB)$, that can be rewritten using Equations 3.17 and 3.18 as:

$$- \varphi(\gamma) + r_1 a_1 - \max\{\nu, r_1(1 - \delta)\} a_1(1 + \gamma) \geq 0 \tag{3.19}$$

To characterise the blocking decision we need again to distinguish the two cases:

Case $\boldsymbol{r_1 \geq \nu \geq (1 - \delta) r_1}$ Technological progress reduces religiousness up to the extent that the secular good provides higher marginal utility than the religious good. In this case, Equation 3.19 modifies to:

$$- \varphi(\gamma) + r_1 a_1 - \nu a_1(1 + \gamma) \geq 0, \tag{3.20}$$

yielding an optimal decision choice equal to:

$$\text{block if} \quad r_1 \geq \frac{\varphi(\gamma)}{a_1} + \nu(1 + \gamma). \tag{3.21}$$

If ν is very high, which means the secular good is very valuable, the inequality is never verified and blocking does not occur. We then need to assume a range of parameters such that the inequality may be verified (ν small enough) so that techonological progress will be the determinant of the blocking decision.

Case $\boldsymbol{(1 - \delta) r_1 \geq \nu}$ When after innovation the value of the religious good is still higher than the one derived from the secular good, Equation 3.19 modifies to:

$$- \varphi(\gamma) + r_1 a_1 - r_1(1 - \delta) a_1(1 + \gamma) \geq 0, \tag{3.22}$$

and the blocking decision becomes:

$$\text{block if} \quad r_1 \geq \frac{\varphi(\gamma)}{a_1} + r_1(1-\delta)(1+\gamma). \tag{3.23}$$

or

$$\text{block if} \quad r_1 \geq \frac{\varphi(\gamma)}{a_1(\delta - \gamma + \gamma\delta)} \quad \textit{with} \quad \gamma < \frac{\delta}{1-\delta} \tag{3.24}$$

assuming $\gamma < \frac{\delta}{1-\delta}$, there will be blocking with a high level of beliefs r_1, low level of technological progress γ or low level of technology a_1 and vice-versa.

However, in case $\gamma > \frac{\delta}{1-\delta}$ the inequality is reversed:

$$\text{block if} \quad r_1 < \frac{\varphi(a_1)}{a_1(\delta - \gamma + \gamma\delta)} \quad \textit{with} \quad \gamma > \frac{\delta}{1-\delta} \tag{3.25}$$

As the right-hand-side is negative, in the latter case it is always optimal not to block innovation. This is obvious as $\gamma > \frac{\delta}{1-\delta}$ implies that the gain from technological progress is higher than the loss from reduced religiousness. In a two-period model the agents do not account for the fact that they are getting closer to the threshold where $\nu \geq (1-\delta)r_1$.

3.B Additional Tables

Table 3.B1: Pooled sample

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0408***	-0.0285***	-0.0215	-0.192***	-0.116***	-0.0560*
	(0.0085)	(0.010)	(0.014)	(0.022)	(0.024)	(0.029)
Log GDP per capita			-0.00890			-0.105**
			(0.021)			(0.049)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.182	0.381	0.425	0.392	0.616	0.680
Observations	274	263	237	259	249	225

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at country level. Each observation represents one country-survey wave pair. Dependent variable: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-4$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-4$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B2: Other measures of religiousness and innovation

	Dependent variable:					
	Beliefs in God			Weekly Church Attendance		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0527***	-0.0303***	-0.0307**	-0.0604***	-0.0416***	-0.0153
	(0.0062)	(0.0092)	(0.015)	(0.0081)	(0.0090)	(0.011)
Log GDP per capita			-0.00430			-0.0307*
			(0.023)			(0.018)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.287	0.621	0.636	0.407	0.620	0.604
Observations	76	72	66	87	82	74

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring to believe in God in columns 1-3, average number of people in the country declaring to attend church at least once a week in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-10$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-10$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B3: Religion and innovation, robustness checks

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0380**	-0.0396**	-0.0383**	-0.0840***	-0.0648*	-0.0918***
	(0.015)	(0.018)	(0.015)	(0.030)	(0.038)	(0.030)
All controls	✓	✓	✓	✓	✓	✓
R^2	0.527	0.525	0.552	0.765	0.769	0.772
Observations	74	74	75	75	75	76

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at World Bank (2015) region level in parentheses in column 1 and 4. Robust standard errors in parentheses in columns 2-3 and 5-6. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-10$ to $t-1$. *All controls* include: the log of the average of GDP per capita from years $t-10$ to $t-1$ education as the average years of tertiary schooling in the country from Barro and Lee (2013) (secondary schooling in columns 2 and 5), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP, dummies for the regulation of religion in the state, the presence of a communist regime in 1970 (2000 in columns 3 and 6), the presence of a state religion in 1970 (2000 in columns 3 and 6), dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970 (2000 in columns 3 and 6), a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B4: Religion and innovation, robustness checks

	Dependent variable:					
	Beliefs in God			Weekly Church Attendance		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0307**	-0.0370**	-0.0354**	-0.0152	-0.00677	-0.0173
	(0.015)	(0.017)	(0.015)	(0.011)	(0.012)	(0.011)
All controls	✓	✓	✓	✓	✓	✓
R^2	0.636	0.642	0.632	0.604	0.615	0.587
Observations	66	66	67	74	74	75

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at World Bank (2015) region level in parentheses in column 1 and 4. Robust standard errors in parentheses in columns 2-3 and 5-6. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring to believe in God in columns 1-3, average number of people in the country declaring to attend church at least once a week in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t - 10$ to $t - 1$. *All controls* include: the log of the average of GDP per capita from years $t - 10$ to $t - 1$ education as the average years of tertiary schooling in the country from Barro and Lee (2013) (secondary schooling in columns 2 and 5), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP, dummies for the regulation of religion in the state, the presence of a communist regime in 1970 (2000 in columns 3 and 6), the presence of a state religion in 1970 (2000 in columns 3 and 6), dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970 (2000 in columns 3 and 6), a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B5: Other measures of religiousness and innovation using scientific articles

	Dependent variable:					
	Beliefs in God			Weekly Church Attendance		
	(1)	(2)	(3)	(4)	(5)	(6)
Log articles per capita	-0.0499***	-0.0314***	-0.0395**	-0.0643***	-0.0598***	-0.0278
	(0.0058)	(0.0097)	(0.019)	(0.0081)	(0.0093)	(0.022)
Log GDP per capita			0.0141			-0.0334
			(0.025)			(0.029)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.323	0.615	0.624	0.345	0.649	0.664
Observations	85	82	73	99	94	83

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. Dependent variable: average number of people in the country declaring to believe in God in columns 1-3, average number of people in the country declaring to attend church at least once a week in columns 4-6. Log articles per capita is the logarithm of the average number of scientific articles and technical reports from years $t-10$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-10$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B6: Religion and innovation, 66-country sample

	Dependent variable:					
	Religiosity			Importance of Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
Log patents per capita	-0.0443***	-0.0323***	-0.0471***	-0.209***	-0.134***	-0.102***
	(0.0076)	(0.010)	(0.017)	(0.024)	(0.026)	(0.038)
Controls			✓			✓
State religion		✓	✓		✓	✓
Wave FE		✓	✓		✓	✓
R^2	0.280	0.507	0.557	0.507	0.731	0.767
Observations	70	68	66	71	69	67

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Each observation represents one country in the latest available survey for that country. The sample includes only the 66 countries for whom patent protection data is available and are used for the instrumental variable analysis. Dependent variable: average number of people in the country declaring themselves as religious in columns 1-3, country-average of importance of religion ranked on a 1-4 scale in columns 4-6. Log patents per capita is the logarithm of the average patents per capita from years $t-10$ to $t-1$. Log GDP per capita is the log of the average of GDP per capita from years $t-10$ to $t-1$. The other controls are education as the average years of tertiary schooling in the country from Barro and Lee (2013), the urbanisation rate, the number of migrants as a fraction of population, trade and FDI as a fraction of GDP. *State religion* variables include dummies for the regulation of religion in the state, the presence of a communist regime in 1970, the presence of a state religion in 1970, dummies for the predominant religion (more than 40% of the population) being Catholicism, Islam, Protestantism or Orthodox Christianity and the Herfindahl index of religious concentration measured in 1970. *Wave* is a set of dummies for the wave of the survey the observation is taken from. The dummy for presence of a communist region, originally from Barro (2003), takes a value of 0.204 for Germany and 0.5 for Vietnam to account for the countries' divisions.

Table 3.B7: Reduced form IV Religion and innovation, US states

	Dependent variable:					
	Rel Importance		Belief in God		Church weekly	
	(1)	(2)	(3)	(4)	(5)	(6)
IV	6.573***	10.77*	2.729***	2.873*	2.929*	6.134***
	(2.38)	(5.94)	(0.82)	(1.43)	(1.54)	(2.14)
All controls		✓		✓		✓
R^2	0.657	0.783	0.589	0.689	0.505	0.659
Observations	48	47	48	47	48	47

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variable: state average of importance of religion ranked on a 1-4 scale in columns 1-2, average number of people in the state declaring to believe in God in columns 3-4, average number of people in the state declaring to attend church at least once a week in columns 5-6. The instrument variable on the right hand side is the average size of the polygon of all points that are closest to a specific university, weighted by the number of universities in the state. *All controls* include: the log of the average state income per capita, educational attainment as percentage of people older than 25 with a graduate degree, the state urbanisation rate, the logarithm of state population, the logarithm of state area, the average share of Republican legislators across the state lower and upper houses. *Area FE* are fixed effects for regions of the contiguous US: north-east, mid-west, south and west. In columns 2, 4 and 6, Nebraska is excluded as the data on the number of Republicans is missing since the state legislature is composed of a single non-partisan house.

Bibliography

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American economic review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth 1*, 385–472.

Aleksynska, M. and B. R. Chiswick (2013). The determinants of religiosity among immigrants and the native born in europe. *Review of Economics of the Household 11*(4), 563–598.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic growth 8*(2), 155–194.

Andersson, R., J. M. Quigley, and M. Wilhelmsson (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics 66*(1), 2–15.

Ash, E. and W. B. MacLeod (2016). The performance of elected officials: Evidence from state supreme courts. Technical report, National Bureau of Economic Research.

Azzi, C. and R. Ehrenberg (1975). Household allocation of time and church attendance. *The Journal of Political Economy*, 27–56.

Baldwin, J. and M. McConville (1978). Plea bargaining and plea negotiation in england. *Law & Soc'y Rev. 13*, 287.

Barro, R. and R. M. McCleary (2003a). International determinants of religiosity. Technical report, National Bureau of Economic Research.

Barro, R. J. (2003). Religion adherence data.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics 104*, 184–198.

Barro, R. J. and R. McCleary (2003b). Religion and economic growth. Technical report, National Bureau of Economic Research.

Bean, C. (2016). Independent review of uk economic statistics. *HM Treasury, Cabinet Office, The Rt Hon Matt Hancock MP and The Rt Hon George Osborne MP 11*.

Becker, S. O., S. Heblich, and D. M. Sturm (2018). The impact of public employment: evidence from Bonn.

Becker, S. O. and L. Woessmann (2009). Was weber wrong? a human capital theory of protestant economic history. *The Quarterly Journal of Economics 124*(2), 531–596.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy 122*(3), 507–553.

Behrens, K. and F. Robert-Nicoud (2015). Agglomeration theory with heterogeneous agents. In *Handbook of regional and urban economics*, Volume 5, pp. 171–245. Elsevier.

Bénabou, R., D. Ticchi, and A. Vindigni (2015a). Forbidden fruits: the political economy of science, religion, and growth. Technical report, National Bureau of Economic Research.

Bénabou, R., D. Ticchi, and A. Vindigni (2015b). Religion and innovation. *American Economic Review 105*(5), 346–51.

Berdejó, C. and N. Yuchtman (2013). Crime, punishment, and politics: an analysis of political cycles in criminal sentencing. *Review of Economics and Statistics 95*(3), 741–756.

Berkman, M. and E. Plutzer (2010). *Evolution, creationism, and the battle to control America's classrooms*. Cambridge University Press.

Berkman, M. B. and E. Plutzer (2011). Defeating creationism in the courtroom, but not in the classroom. *Science 331*(6016), 404–405.

Berry, S. and J. Waldfogel (2010). Product quality and market size. *The Journal of Industrial Economics 58*(1), 1–31.

Besley, T. (2004). Paying politicians: theory and evidence. *Journal of the European Economic Association 2*(2-3), 193–215.

Besley, T. and M. Ghatak (2010). Property rights and economic development. In *Handbook of development economics*, Volume 5, pp. 4525–4595. Elsevier.

Besley, T., J. G. Montalvo, and M. Reynal-Querol (2011). Do educated leaders matter? *The Economic Journal 121*(554), F205–227.

Besley, T. and A. Payne (2003). Judicial accountability and economic policy outcomes: Evidence from employment discrimination charges. Technical report, IFS Working Papers, Institute for Fiscal Studies (IFS).

Besley, T. and T. Persson (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review 99*(4), 1218–44.

Besley, T. and M. Reynal-Querol (2011). Do democracies select more educated leaders? *American political science review 105*(3), 552–566.

Bils, M. and P. J. Klenow (2001). Quantifying quality growth. *American Economic Review 91*(4), 1006–1030.

Bisin, A. and T. Verdier (2011). The economics of cultural transmission and socialization. In *Handbook of social economics*, Volume 1, pp. 339–416. Elsevier.

Blanes i Vidal, J. and C. Leaver (2011). Are tenured judges insulated from political pressure? *Journal of public economics 95*(7-8), 570–586.

Blanes i Vidal, J. and C. Leaver (2015). Bias in open peer-review: evidence from the english superior courts. *The Journal of Law, Economics, and Organization 31*(3), 431–471.

Blum, U. and L. Dudley (2001). Religion and economic growth: was weber right? *Journal of Evolutionary Economics 11*(2), 207–230.

Bomboy, S. (2014, January). Can the nation's capital really be moved to nebraska? `https://constitutioncenter.org/blog/can-the-nations-capital-really-be-moved-to-nebraska`.

Bowen, D. C. and Z. Greene (2014a). Legislative Professionalism Component Scores, V1.1.1. `https://doi.org/10.7910/DVN/27595`.

Bowen, D. C. and Z. Greene (2014b). Should we measure professionalism with an index?

a note on theory and practice in state legislative professionalism research. *State Politics & Policy Quarterly 14*(3), 277–296.

Bratton, K. A. and K. L. Haynie (1999). Agenda setting and legislative success in state legislatures: The effects of gender and race. *The Journal of Politics 61*(3), 658–679.

Breusch, T. S. and A. R. Pagan (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies 47*(1), 239–253.

Briant, A., P.-P. Combes, and M. Lafourcade (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics 67*(3), 287–302.

Bruce, S. (1992). *Religion and modernization: Sociologists and historians debate the secularization thesis.* Oxford University Press.

Burton, E. K. (2011). Evolution and creationism in middle eastern education: a new perspective. *Evolution: International Journal of Organic Evolution 65*(1), 301–304.

Campante, F. R. and Q.-A. Do (2010). A centered index of spatial concentration: Expected influence approach and application to population and capital cities.

Campante, F. R. and Q.-A. Do (2014). Isolated capital cities, accountability, and corruption: Evidence from us states. *American Economic Review 104*(8), 2456–81.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2019). Capital cities, conflict, and misgovernance. *American Economic Journal: Applied Economics 11*(3), 298–337.

Cantoni, D. (2015). The economic effects of the protestant reformation: testing the weber hypothesis in the german lands. *Journal of the European Economic Association 13*(4), 561–598.

Cappelli, P. and K. Chauvin (1991). An interplant test of the efficiency wage hypothesis. *The Quarterly Journal of Economics 106*(3), 769–787.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Carey, J. M., R. G. Niemi, and L. W. Powell (1998). The effects of term limits on state legislatures. *Legislative Studies Quarterly*, 271–300.

Carey, J. M., R. G. Niemi, and L. W. Powell (2009). *Term limits in state legislatures.* University of Michigan Press.

Carey, J. M., R. G. Niemi, L. W. Powell, and G. F. Moncrief (2006). The effects of term limits on state legislatures: a new survey of the 50 states. *Legislative Studies Quarterly 31*(1), 105–134.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. In *Handbook of regional and urban economics*, Volume 5, pp. 349–404. Elsevier.

Carlino, G. A. and A. Saiz (2019). Beautiful city: Leisure amenities and urban growth. *Journal of Regional Science 59*(3), 369–408.

Carvalho, J.-P. and M. Koyama (2011). *Development and religious polarization: The emergence of reform and ultra-Orthodox Judaism.* Department of Economics, University of Oxford.

Caselli, F. and M. Morelli (2004). Bad politicians. *Journal of Public Economics 88*(3-4), 759–782.

Chaves, M. and P. S. Gorski (2001). Religious pluralism and religious participation. *Annual review of sociology*, 261–281.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review 86*(1), 54–70.

Clark, W. A., Y. Huang, and S. Withers (2003). Does commuting distance matter?: Commuting tolerance and residential change. *Regional Science and Urban Economics 33*(2), 199–221.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, Volume 5, pp. 247–348. Elsevier.

Compton, J. and R. A. Pollak (2007). Why are power couples increasingly concentrated in large metropolitan areas? *Journal of Labor Economics 25*(3), 475–512.

Costa, D. L. and M. E. Kahn (2000). Power couples: changes in the locational choice of the college educated, 1940–1990. *The Quarterly Journal of Economics 115*(4), 1287–1315.

Courts Act (1971). `http://www.legislation.gov.uk/ukpga/1971/23/section/59`.

Cragg, J. G. and S. G. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9*(2), 222–240.

Cross, F. B. and J. F. Spriggs (2010). Citations in the us supreme court: An empirical study of their use and significance. *U. Ill. L. Rev.*, 489.

CSG (2018). *The Book of the States 2018.* Council of State Governments, Book Of the States. CSG.

Dal Bó, E., P. Dal Bó, and J. Snyder (2009). Political dynasties. *The Review of Economic Studies 76*(1), 115–142.

Dal Bó, E., F. Finan, O. Folke, T. Persson, and J. Rickne (2017). Who becomes a politician? *The Quarterly Journal of Economics 132*(4), 1877–1914.

Davis, D. R. and J. I. Dingel (2019). A spatial knowledge economy. *American Economic Review 109*(1), 153–70.

DuMond, J. M., B. T. Hirsch, and D. A. Macpherson (1999). Wage differentials across labor markets and workers: does cost of living matter? *Economic inquiry 37*(4), 577–598.

Eakins, K. R. (2006). Lawyers in the legislature: The case of ohio. *The Social Science Journal 43*(4), 717–725.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy 122*(3), 554–620.

Engstrom, E. J., J. R. Hammond, and J. T. Scott (2013). Capitol mobility: Madisonian representation and the location and relocation of capitals in the united states. *American Political Science Review 107*(2), 225–240.

European Value Study (2011). European values study 1981-2008, longitudinal data file. Data File Version 2.0.0, GESIS Data Archive, Cologne, ZA4804.

Faggio, G., T. Schluter, and P. vom Berge (2019). Interaction of public and private employment: Evidence from a german government move.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). Income distribution, product quality, and international trade. *Journal of political Economy 119*(4), 721–765.

Feenstra, R. C., R. Inklaar, and M. Timmer (2013). The next generation of the penn world table. Technical report, National Bureau of Economic Research.

Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly journal of economics 123*(2), 703–745.

Ferraz, C. and F. Finan (2009). Motivating politicians: The impacts of monetary incentives on quality and performance. Technical report, National Bureau of Economic Research.

Ferraz, C. and F. Finan (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review 101*(4), 1274–1311.

Finke, R., A. M. Guest, and R. Stark (1996). Mobilizing local religious markets: Religious pluralism in the empire state, 1855 to 1865. *American Sociological Review*, 203–218.

Finke, R. and R. Stark (1988). Religious economies and sacred canopies: Religious mobilization in american cities, 1906. *American Sociological Review*, 41–49.

Finke, R. and R. Stark (2005). *The churching of America, 1776-2005: Winners and losers in our religious economy.* Rutgers University Press.

Franck, R. and L. R. Iannaccone (2014). Religious decline in the 20th century west: Testing alternative explanations. *Public Choice 159*(3-4), 385–414.

Gagliarducci, S. and T. Nannicini (2013). Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association 11*(2), 369–398.

Gagliarducci, S., T. Nannicini, and P. Naticchioni (2010). Moonlighting politicians. *Journal of Public Economics 94*(9-10), 688–699.

Galasso, V. and T. Nannicini (2011). Competing on good politicians. *American political science review 105*(1), 79–99.

Gallagher, S. and W. Tierney (2013). *Religiousness/Religiosity*, pp. 1653–1654. New York, NY: Springer New York.

Galor, O. and O. Moav (2006). Das human-kapital: A theory of the demise of the class structure. *The Review of Economic Studies 73*(1), 85–117.

Galster, G., R. Hanson, M. R. Ratcliffe, H. Wolman, S. Coleman, and J. Freihage (2001). Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing policy debate 12*(4), 681–717.

Gardiner, B., R. Martin, P. Sunley, and P. Tyler (2013). Spatially unbalanced growth in the british economy. *Journal of Economic Geography 13*(6), 889–928.

Geiger, R. L. (2016). *The history of American higher education: Learning and culture from the founding to World War II*, Volume 80. Princeton University Press.

Gibson, B. and P. Cavadino (2008). *The Criminal Justice System: An Introduction.* Introductory Series. Waterside Press.

Glaeser, E. L. and J. D. Gottlieb (2006). Urban resurgence and the consumer city. *Urban studies 43*(8), 1275–1299.

Glaeser, E. L. and J. D. Gottlieb (2009, December). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature 47*(4), 983–1028.

Glaeser, E. L., J. Kolko, and A. Saiz (2001). Consumer city. *Journal of economic geography 1*(1), 27–50.

Glaeser, E. L. and B. I. Sacerdote (2008). Education and religion. *Journal of Human Capital 2*(2), 188–215.

Glaeser, E. L. and R. E. Saks (2006). Corruption in america. *Journal of public Economics 90*(6-7), 1053–1072.

Gordon, S. C. and G. A. Huber (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science 2*(2), 107–139.

Greaves, T., T. Stowell, M. Doel, D. Lowry, L. Farris, J. Liebert, and J. Marsh (2017). The case for (and against) moving parliament out of london. *The Guardian*. https://www.theguardian.com/politics/2017/jul/18/the-case-for-and-against-moving-parliament-out-of-london.

Gruber, J. H. (2005). Religious market structure, religious participation, and outcomes: Is religion good for you? *Advances in Economic Analysis & Policy 5*(1).

Guiso, L., P. Sapienza, and L. Zingales (2003). People's opium? religion and economic attitudes. *Journal of monetary economics 50*(1), 225–282.

Handbury, J. (2019). Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities. Technical report, National Bureau of Economic Research.

Handbury, J. and D. E. Weinstein (2015). Goods prices and availability in cities. *Review of Economic Studies 82*(1).

Helpman, E. (1998). The size of regions. *Topics in public economics: Theoretical and applied analysis*, 33–54.

Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review 64*(4), 640–656.

Henderson, V. (2003). The urbanization process and economic growth: The so-what question. *Journal of Economic growth 8*(1), 47–71.

HM Land Registry (2018). Uk house price index. https://www.gov.uk/govhttp://landregistry.data.gov.uk/app/ukhpi/browse?from=2017-04-01&location=http%3A%2F%2Flandregistry.data.gov.uk%2Fid%2Fregion%2Funited-kingdom&to=2018-04-01ernment/publications/judicial-salaries-and-fees-2016-to-2017.

Hoffman, M. and E. Lyons (2018). A time to make laws and a time to fundraise? on the relation between salaries and time use for state politicians.

Home Office (2017). Home office funds new video enabled justice programme. =https://www.gov.uk/government/news/home-office-funds-new-video-enabled-justice-programme.

Huber, G. A. and S. C. Gordon (2004). Accountability and coercion: Is justice blind when it runs for office? *American Journal of Political Science*, 247–263.

Huber, J. (2005). Religious belief, religious participation, and social policy attitudes across countries. In *Annual Meetings of the Midwest Political Science Association, Chicago*.

Hungerman, D. M. (2005). Are church and state substitutes? evidence from the 1996 welfare reform. *Journal of Public Economics 89*(11), 2245–2267.

Hungerman, D. M. (2014). The effect of education on religion: Evidence from compulsory schooling laws. *Journal of Economic Behavior & Organization 104*, 52–63.

Iannaccone, L. R. (1992). Sacrifice and stigma: reducing free-riding in cults, communes, and other collectives. *Journal of political economy*, 271–291.

Iannaccone, L. R. (1994). Why strict churches are strong. *American Journal of Sociology*, 1180–1211.

Iannaccone, L. R. and E. Berman (2006). Religious extremism: The good, the bad, and the deadly. *Public Choice 128*(1-2), 109–129.

Iannaccone, L. R., R. Finke, and R. Stark (1997). Deregulating religion: the economics of church and state. *Economic Inquiry 35*(2), 350–364.

Inglehart, R. and P. Norris (2004). Sacred and secular: Religion and politics worldwide. *Cambridge/Mass.*

Jefferson, M. (1989). Why geography? the law of the primate city. *Geographical review 79*(2), 226–232.

Jenkins, S. (2017). Parliament needs to leave london and reconnect with the people. *The Guardian.* https://www.theguardian.com/commentisfree/2017/jul/17/brexit-britain-needs-mps-out-of-westminster-provinces-neglected.

Judicial Appointments Commission (2017). 052 circuit judge 2017 information page. Retrieved from https://www.judicialappointments.gov.uk/36 on 12/04/2017.

Judiciary (2017). Judicial diversity statistics. https://www.judiciary.uk/about-the-judiciary/who-are-the-judiciary/diversity/judicial-diversity-statistics-2017/.

Judiciary UK (2018a). Courts judiciary. https://www.judiciary.uk/about-the-judiciary/who-are-the-judiciary/judicial-roles/judges/.

Judiciary UK (2018b). Structure of the courts tribunal system. https://www.judiciary.uk/about-the-judiciary/the-justice-system/court-structure/.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics 133*(1), 97–126.

Kleven, H. J., C. Landais, and E. Saez (2013, August). Taxation and international migration of superstars: Evidence from the european football market. *American Economic Review 103*(5), 1892–1924.

Kleven, H. J., C. Landais, E. Saez, and E. Schultz (2014). Migration and wage effects of taxing top earners: evidence from the foreigners' tax scheme in denmark. *Quarterly Journal of Economics 129*(1), 333–378.

Kurtz, K. (2015, December). Who we elect: The demographics of state legislatures. *State Legislatures Magazine*. Retrieved from http://www.ncsl.org/research/about-state-legislatures/who-we-elect.aspx.

Kurtz, K. T. (2010). 75 years of institutional change in state legislatures. *Book of the States 2010 Edition.*

Kurtz, K. T., G. Moncrief, R. G. Niemi, and L. W. Powell (2006). Full-time, part-time, and real time: Explaining state legislators' perceptions of time on the job. *State Politics & Policy Quarterly 6*(3), 322–338.

Lady Justice Arden (2016). The judicial system of england and wales. a visitor's guide. https://www.judiciary.uk/wp-content/uploads/2016/05/international-visitors-guide-10a.pdf.

Le Vine, S., J. Polak, and A. Humphrey (2017). Commuting trends in england 1988-2015. *Department for Transport.*

Lee, S. (2010). Ability sorting and consumer city. *Journal of urban Economics 68*(1), 20–33.

Levy, G. and R. Razin (2012). Religious beliefs, religious participation, and cooperation.

American economic journal: microeconomics 4(3), 121–151.

Lovell, W. (1892). Queen eleanor's crosses. *Archaeological Journal 49*(1), 17–43.

Malleson, K. (2006). The new judicial appointments commission in england and wales: New wine in new bottles? *Appointing Judges in an Age of Judicial Power 39*, 40–42.

Maoz, Z. and E. A. Henderson (2013). The world religion dataset, 1945–2010: Logic, estimates, and trends. *International Interactions 39*(3), 265–291.

Martin, D. (1978). *A general theory of secularization*. Blackwell Oxford.

Maskin, E. and J. Tirole (2004). The politician and the judge: Accountability in government. *American Economic Review 94*(4), 1034–1054.

Matter, U. and A. Stutzer (2015). The role of lawyer-legislators in shaping the law: evidence from voting on tort reforms. *The Journal of Law and Economics 58*(2), 357–384.

Mattozzi, A. and A. Merlo (2008). Political careers or career politicians? *Journal of Public Economics 92*(3-4), 597–608.

McCleary, R. M. and R. J. Barro (2006). Religion and political economy in an international panel. *Journal for the Scientific study of religion 45*(2), 149–175.

Miller, M. C. (1995). *The high priests of American politics: the role of lawyers in American political institutions*. Univ. of Tennessee Press.

Milligan, K., E. Moretti, and P. Oreopoulos (2004). Does education improve citizenship? evidence from the united states and the united kingdom. *Journal of public Economics 88*(9-10), 1667–1695.

Minarik, P. (2014). Employment, wages, and religious revivals in postcommunist countries. *Journal for the Scientific Study of Religion 53*(2), 296–315.

Ministry of Justice (2015). Judicial salaries and fees: 2015 to 2016. `https://www.gov.uk/government/publications/judicial-salaries-and-fees-2015-to-2016`.

Ministry of Justice (2016). Judicial salaries and fees: 2016 to 2017. `https://www.gov.uk/government/publications/judicial-salaries-and-fees-2016-to-2017`.

Ministry of Justice (2017a). Criminal court statistics quarterly: January to march 2017. `https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-january-to-march-2017` (Accessed 02/07/2018).

Ministry of Justice (2017b). A guide to criminal court statistics. Available at `https://www.gov.uk/government/publications/a-guide-to-criminal-court-statistics`.

Ministry of Justice (2018a). Crown court receipts, disposals and outstanding cases by offence group. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/743438/cc-rdos-by-offence-group-q2-2018.csv/preview`.

Ministry of Justice (2018b). Freedom of information act (foia) request 180327013.

Ministry of Justice (2018c). Freedom of information act (foia) request 180413009.

Ministry of Justice (2018d). Royal courts of justice annual tables. Available at `https://www.gov.uk/government/statistics/civil-justice-statistics-quarterly-january-to-march-2018`.

Ministry of Justice, Data Access & Compliance Unit (2017a, July). Freedom of information act (foia) request 112705.

Ministry of Justice, Data Access & Compliance Unit (2017b, September). Freedom of information act (foia) request 113594.

Miremadi, T., G. H. Salekdeh, N. Aghdami, M. Gharanfoli, M. Vasei, A. Kouhkan, and H. Baharvand (2012). Stem cell research and therapy in the islamic republic of iran: pioneering in the islamic world. *Stem cells and development 22*(1), 51–57.

Mocan, N. and L. Pogorelova (2017). Compulsory schooling laws and formation of beliefs: Education, religion and superstition. *Journal of Economic Behavior & Organization 142*, 509–539.

Moncrief, G. F. and P. Squire (2017). *Why states matter: An introduction to state politics.* Rowman & Littlefield.

Montès, C. (2014). *American capitals: a historical geography.* University of Chicago Press.

Montgomery, J. D. (1996). Contemplations on the economic approach to religious behavior. *The American Economic Review*, 443–447.

Moretti, E. (2004). Human capital externalities in cities. In *Handbook of regional and urban economics*, Volume 4, pp. 2243–2291. Elsevier.

Moretti, E. and D. J. Wilson (2017, July). The effect of state taxes on the geographical location of top earners: Evidence from star scientists. *American Economic Review 107*(7), 1858–1903.

NCSL (2015a). The term-limited states. Retrieved from http://www.ncsl.org/research/about-state-legislatures/chart-of-term-limits-states.aspx.

NCSL (2015b). Who we elect. Retrieved from http://www.ncsl.org/press-room/ncsl-releases-in-depth-report-on-state-legislator-demographics.aspx.

NCSL (2017, June). Full- and part-time legislatures. Retrieved from http://www.ncsl.org/research/about-state-legislatures/full-and-part-time-legislatures.aspx.

NCSL (2018). Annual versus biennial legislative sessions. Retrieved from http://www.ncsl.org/research/about-state-legislatures/annual-versus-biennial-legislative-sessions.aspx.

NCSL (2019). National conference of state legislatures. Retrieved from http://www.ncsl.org/research.aspxx.

Neumark, D., J. Zhang, and S. Ciccarella (2008). The effects of wal-mart on local labor markets. *Journal of Urban Economics 63*(2), 405–430.

North, D. C. (1991). Institutions. *Journal of economic perspectives 5*(1), 97–112.

OECD (2016). Functional urban areas in oecd countries. Available at https://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm.

Office for National Statistics (2018). Uk business counts - local units by industry and employment size band. https://www.nomisweb.co.uk/datasets/idbrlu.

Olea, J. L. M. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

ONS (2012). Home people, population and community population and migration population estimates 2011 census 2011 census: Population and household estimates for small areas in england and wales, march 2011. https://www.ons.gov.uk/peoplepopulationandcommunity/

populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23.

ONS (2017). Regional gross value added (balanced) from 1998 to 2016. Available at https://www.ons.gov.uk/economy/grossvalueaddedgva/bulletins/regionalgrossvalueaddedbalanceduk/1998to2016#london-dominates-top-five-nuts3-regions.

Opfinger, M. (2014). Two sides of a medal: The changing relationship between religious diversity and religiosity. *Review of Social Economy 72*(4), 523–548.

Oster, E. (2012). Hiv and sexual behavior change: Why not africa? *Journal of health economics 31*(1), 35–49.

Ottaviano, G., T. Tabuchi, and J.-F. Thisse (2002). Agglomeration and trade revisited. *International Economic Review*, 409–435.

Pande, R. (2011). Can informed voters enforce better governance? experiments in low-income democracies. *Annu. Rev. Econ. 3*(1), 215–237.

Park, W. G. (2008). International patent protection: 1960–2005. *Research policy 37*(4), 761–766.

Pew Reseach Center (2016). Religious landscape study.

Powell, L. and K. Kurtz (2014). Measuring the power of state legislative leaders. In *APSA 2014 Annual Meeting Paper*.

Powell, L. W. (2012). *The influence of campaign contributions in state legislatures: the effects of institutions and politics*. University of Michigan Press.

Propper, C. and J. Van Reenen (2010, April). Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance. *Journal of Political Economy 118*(2), 222–273.

QS (2019). Qs world university rankings. Retrieved from https://www.topuniversities.com/university-rankings/world-university-rankings/2019.

Rappaport, J. (2008). Consumption amenities and city population density. *Regional Science and Urban Economics 38*(6), 533–552.

Repetto, L. (2018). Political budget cycles with informed voters: evidence from italy. *The Economic Journal 128*(616), 3320–3353.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy 90*(6), 1257–78.

Roberts, J., R. Hodgson, and P. Dolan (2011). "it's driving her mad": Gender differences in the effects of commuting on psychological health. *Journal of health economics 30*(5), 1064–1076.

Scheve, K. and D. Stasavage (2006). Religion and preferences for social insurance. *Quarterly Journal of Political Science 1*(3), 255–286.

Schultz, U. and G. Shaw (2003). *Women in the world's legal professions*. Bloomsbury Publishing.

Senior Courts Act (1981). https://www.legislation.gov.uk/ukpga/1981/54/contents.

Sexton, B. (2018). How to become a judge. https://www.allaboutlaw.co.uk/stage/becoming-a-lawyer/how-to-become-a-judge.

ShanghaiRanking Consultancy (2018). Academic ranking of world universities. Retrieved from http://www.shanghairanking.com/Academic-Ranking-of-World-Universities-2018-Press-Release.html.

Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica 73*(4), 1367–1382.

Smart, M. and D. M. Sturm (2013). Term limits and electoral accountability. *Journal of public economics 107*, 93–102.

Snyder Jr, J. M. and D. Strömberg (2010). Press coverage and political accountability. *Journal of political Economy 118*(2), 355–408.

Social Security Administration (2019). Popular names in 1961. https://www.ssa.gov/cgi-bin/popularnames.cgi.

Squire, P. (1992). Legislative professionalization and membership diversity in state legislatures. *Legislative Studies Quarterly*, 69–79.

Squire, P. (2000). Uncontested seats in state legislative elections. *Legislative Studies Quarterly*, 131–146.

Squire, P. (2012). *The evolution of American legislatures: Colonies, territories, and states, 1619-2009*. University of Michigan Press.

Squire, P. (2017). A squire index update. *State Politics & Policy Quarterly 17*(4), 361–371.

Squire, P. and G. Moncrief (2015). *State legislatures today: Politics under the domes*. Rowman & Littlefield.

Stark, R. and W. S. Bainbridge (1985). *The future of religion: Secularization, revival, and cult formation*. Univ of California Press.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80.

Strokova, V. (2010). International property rights index: 2010 report.

Strömberg, D. (2015). Media coverage and political accountability: Theory and evidence. In *Handbook of media Economics*, Volume 1, pp. 595–622. Elsevier.

Sullivan, D. H. (1985). Simultaneous determination of church contributions and church attendance. *Economic Inquiry 23*(2), 309–320.

Tabarrok, A. and E. Helland (1999). Court politics: The political economy of tort awards. *The Journal of Law and Economics 42*(1), 157–188.

Tabuchi, T. (2001). On interregional price differentials. *The Japanese Economic Review 52*(1), 104–115.

Tabuchi, T. and A. Yoshida (2000). Separating urban agglomeration economies in consumption and production. *Journal of Urban Economics 48*(1), 70–84.

Tan, W. (2019, May). Indonesia plans to move its capital from jakarta. here's why. https://www.cnbc.com/2019/05/02/indonesias-joko-widodo-decides-to-move-capital-from-jakarta.html.

The Telegraph (2015). Virtual courts could deliver real justice. Available at https://www.telegraph.co.uk/news/uknews/law-and-order/11414466/Virtual-courts-could-deliver-real-justice.html.

Times Higher Education (2019a). World university rankings. Retrieved from `https://www.timeshighereducation.com/world-university-rankings/2019/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats`.

Times Higher Education (2019b). World university rankings 2019 methodology. Retrieved from `https://www.timeshighereducation.com/sites/default/files/the_2019_world_university_rankings_methodology_pwc.pdf`.

Tocqueville, A. (2003). *Democracy in America: and Two essays on America.* Penguin UK.

UNESCO Institute for Statistics (2011). International standard classification of education.

U.S. Census Bureau, Geography Program (2016). American community survey 5-year estimates — geodatabase format. `https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.2016.html`.

U.S. Census Bureau, Population Division (2017, December). Annual estimates of the resident population: April 1, 2010 to july 1, 2017. `https://www.census.gov/data/datasets/2017/demo/popest/state-detail.html#par_textimage`.

U.S. Census Bureau, Population Division (2018, May). Annual estimates of the resident population: April 1, 2010 to july 1, 2017. Retrieved from `https://www.census.gov/data/datasets/2017/demo/popest/total-cities-and-towns.html#ds`.

USTPO (2015). Utility patents per state over time.

Van Ommeren, J. and M. Fosgerau (2009). Workers' marginal costs of commuting. *Journal of Urban Economics 65*(1), 38–47.

Voas, D., A. Crockett, and D. V. Olson (2002). Religious pluralism and participation: Why previous research is wrong. *American Sociological Review*, 212–230.

Vote Smart (2019). Vote smart - the voter's self defense system. Retrieved from `https://votesmart.org/`.

Woodberry, R. D. (2012). The missionary roots of liberal democracy. *American Political Science Review 106*(02), 244–274.

World Bank (2015). World bank development indicators. World bank open data.

World Intellectual Property Organization (2015). Wipo statistics database.

World Values Survey Association (2015). World values survey 1981-2014 longitudinal aggregate. Aggregate File v.20150418, JDSystems, Madrid, Spain.

Wright, G. C. (2007). Do term limits affect legislative roll call voting? representation, polarization, and participation. *State Politics & Policy Quarterly 7*(3), 256–280.

Young, C. and C. Varner (2011). Millionaire migration and state taxation of top incomes: Evidence from a natural experiment. *National Tax Journal 64*(2), 255.