Three Empirical Studies on the Spatial Analysis of Social Interactions in Elections

by

Andreas Mastrosavvas

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

CARDIFF UNIVERSITY

December 2023

DECLARATION

This thesis has been submitted to Cardiff University for the degree of Doctor of Philosophy (PhD) in Social Sciences. It has been carried out at the School of Social Sciences (Social Data Science Lab) and the School of Geography and Planning of the University and funded by the UK Economic and Social Research Council (ESRC) via grant number 2272893. I certify that the thesis is the result of my own independent work, except where otherwise stated, and that any views expressed herein are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party or submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award. I have given consent for this thesis to be made available in the University's Open Access repository and, where approved, to be available in the University's library for inter-library loan and for the title and abstract to be made available to outside organisations, subject to the expiry of any University-approved bar on access. The thesis is approximately 45,000 words in length, excluding references.

This page has been intentionally left blank.

ABSTRACT

This thesis examines the relationship between the geographic structure of social networks and electoral outcomes. It draws on spatial analysis, econometric methods, and data on the density of social ties between the populations of different places to make empirical contributions to a body of related literatures in the social sciences. The thesis is comprised of an introduction, three independent empirical chapters, and a conclusion. The common theme uniting all three chapters is the observation that non-local, and even geographically distant social ties can play an important role in shaping local voting behaviour. In the first chapter, I investigate the effect of social proximity to trade-related employment shocks on vote choice in the 2016 UK EU membership referendum. Looking at referendum results in International Territorial Level 3 (ITL3) regions and individual-level voting data from the British Election Study, I examine whether regional social proximity to shocks in different local labour markets affected support for the Leave option. Instrumental variable estimates suggest that social spillover effects on voting behaviour are comparable to those of within-region exposure and travel an average distance of 74 to 102 kilometres. In the second chapter, I examine social spillovers on choice of voting method from local rollouts of all-mail voting policy in the United States. Combining individual-level administrative records on the entire electorate of North Carolina and data on the social ties between zip code tabulation areas (ZCTAs) across states, I look at how the share of social ties in policy-switching jurisdictions affected mail voting in the 2020 US presidential election. Difference-in-differences estimates are indicative of positive spillover effects, primarily originating from populous Western jurisdictions located over 2,000 kilometres away. In the third chapter, I explore the geography of partisan homophily in the 2020 US presidential election. I use the local Moran index to identify clusters of politically homophilous ZCTAs and examine their characteristics. The findings suggest that while spatial patterns of partisan homophily broadly track residential segregation along the urban-rural continuum, there are notable partisan differences in the relative density and geographic distance of social ties: homophilous Democratic-leaning areas are likely to have denser and more distant ties elsewhere than homophilous Republican-leaning areas. Overall, the thesis contributes new evidence on the role of social interactions in elections and its links to public policy.

This page has been intentionally left blank.

ACKNOWLEDGEMENTS

This thesis has benefitted from the support of the vibrant research community that I am fortunate to have been a part of during my doctoral studies. I am first and foremost grateful for the dedicated supervision of Scott Orford and Matthew Williams, and for the additional support of Luke Sloan and Pete Burnap. I am also thankful for the collegiality and feedback of participants in the annual conferences and colloquia of the British and Irish section of the Regional Science Association International (RSAI-BIS), the Royal Geographical Society with the Institute of British Geographers (RGS-IBG), and the Wales Institute of Social and Economic Research and Data (WISERD). I would also like to thank the Economic and Social Research Council (ESRC) and the Welsh Graduate School for the Social Sciences (WGSSS) for their financial support, and additionally the RSAI-BIS and the Learned Society of Wales (LSW) for their recognition of my work. My thanks further go to the fellow members of my doctoral cohort for creating an ever-present environment where academic matters could be dissected and escaped in equal measure, both during the most challenging phases of the COVID-19 pandemic and beyond. Finally, and most importantly, my deepest gratitude goes to my family, whose unwavering support made all this possible.

This page has been intentionally left blank.

CONTENTS

This page has been intentionally left blank.

LIST OF FIGURES

This page has been intentionally left blank.

LIST OF TABLES

This page has been intentionally left blank.

INTRODUCTION

This thesis is concerned with the social interactions between the residents of different places and their role in elections. It contains three empirical chapters, each addressing a question on this topic using spatial analysis, econometric methods, and a novel dataset on the geographic structure of social networks. In so doing, each chapter contributes to a body of closely related literatures in the social sciences, including economics, human geography, political science, and sociology. Of central relevance to all three contributions is the concept of neighbourhood effects, which has featured in the empirical study of elections across disciplines. Namely, the evidence presented in this thesis shows that new sources of relational data afford new ways of defining neighbourhoods, which can considerably improve our understanding of voting behaviour in the 21st century. I thus begin this introduction with a discussion of neighbourhood effects and related concepts in Section I.1. Equipped with this conceptual framework, I then expand on the rationale for the thesis and provide an overview of chapters in Section I.2.

I.1 Neighbourhoods, Space, and Dependence

It has been a few decades since Tobler introduced his 'first law of geography', declaring that 'everything is related to everything else, but near things are more related than distant things' (1970, 236). Other than just resembling the law of universal gravitation, its Newtonian counterpart, it has arguably enjoyed comparable status in the social sciences. In the subsections that follow, I briefly discuss the concepts attached to this influential statement and their connection to the study of what has come to be known as 'neighbourhood effects'.

I.1.1 Spatial Dependence

Tobler's first law is linked to the concept of spatial dependence, which can be loosely understood as 'a functional relationship between what happens at one point in space and what happens elsewhere' (Anselin, 1988, 11). Formally, we may write:

$$y_i = f(y_j) \qquad \text{for } i \neq j \tag{I.1}$$

Here, y_i and y_j represent observations of a random variable at locations i and j.[1] Correspondingly, the presence of spatial dependence is often formally expressed in terms of the non-zero covariance condition (Anselin and Bera, 1998):

$$Cov(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) \neq 0 \qquad \text{for } i \neq j \tag{I.2}$$

That is, higher values of y at one location are associated with either higher or lower values of y at a different location—a linear relationship that is respectively termed positive or negative spatial autocorrelation. By replacing y_j with another variable measured at j, x_j, one can also consider a spatial relationship between different variables, or spatial cross-correlation.

The link between Tobler's first law and spatial dependence is hence clear: the former expects the latter to be stronger as the geographic distance between any pair of considered locations decreases. In fact, yielding non-zero covariance as per (I.2) only becomes spatially meaningful when the configuration of the pairs ij has an interpretation in terms of relative proximity. Spatial dependence is thus often discussed with reference to 'neighbouring' locations (Anselin and Bera, 1998), where the neighbourhood of any given location is determined according to some measure of distance.

[1] Note that the term 'location' could imply any unit of analysis with a spatial representation. For instance, this could equally denote areal units or individuals embedded in them.

I.1.2 Spatial Interactions

Locations are connected by spatial interactions: flows of tangibles and intangibles, such as goods, services, people, and information (Fotheringham and O'Kelly, 1998). Formally, for n locations, flows between any pair of locations ij can be denoted by element T_{ij} in a $n \times n$, zero-diagonal matrix:

$$\mathbf{T} = \begin{bmatrix} 0 & T_{12} & \cdots & T_{1n} \\ T_{21} & 0 & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & 0 \end{bmatrix} \tag{I.3}$$

Depending on the application in question we may consider spatial interaction as directed or undirected. In the former case, the row sums O_i will represent origin outflows and the column sums D_j will represent destination outflows with $T_{ij} \neq T_{ji}$. In contrast, in the undirected case $\mathbf{T}$ is symmetric.

In a way then, when invoked in social science applications, Tobler's first law is as much a statement about geographic distance as it is about spatial interactions. The latter encapsulate the very channels via which social phenomena in one location may affect those in another. To invoke the law is thus to expect that the level of spatial interactions will be a negative function of geographic distance, and that this is a sound proxy for the relative differences across the pairwise flows of interest T_{ij}.

I.1.3 Defining Neighbourhoods

The neighbourhood of each location considered in an application can be represented using a spatial weights matrix. That is, for n locations, neighbourhood relations between any two locations i and j are typically captured by a $n \times n$, positive, zero-diagonal, row-normalised matrix $\mathbf{W}$ with elements w_{ij}:[2]

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix} \tag{I.4}$$

In line with Tobler's first law, the weights are usually set as a negative function of geographic distance, with weights that are equal to zero designating location-pairs as non-neighbours.

[2] Here, I employ the conventional structure in spatial econometrics. As discussed by Gibbons et al. (2015), different matrix structures are possible, such that, for instance, $\mathbf{W}$ is not row-normalised.

This might be the case, for instance, when the pair in question represents non-contiguous areal units, coordinate points further than a given Euclidean or geodesic distance from each other, or residents in different districts. Sometimes, it may be warranted to conceive of space as 'more than geography' (Beck et al., 2006). This may be especially so when geographic distance is likely to be an unreliable proxy for different spatial interactions of interest in line with the 'uncertain geographic context problem' (Kwan, 2012). Notice the correspondence between (I.3) and (I.4). One may well derive the weight w_{ij} on the basis of the level of some flow T_{ij}.[3] While this could retain the interpretation of i and j as units with a representation in geographic space, it would alter that of their neighbourhoods: being in the same neighbourhood would now be solely determined by the spatial interaction of interest, irrespective of how close the units are to each other in a geographic sense.

I.1.4 Neighbourhood Effects in Social and Geographic Space

One of the core tenets of the social sciences is that social interactions among individuals play an important role in shaping a range of outcomes and behaviours. Social interactions can be defined as 'direct interdependences in preferences, constraints, and beliefs of individuals, which impose a social structure on individual decisions.' (Durlauf and Ioannides, 2010, 452). Such interdependencies can of course be brought about by a variety of channels ranging from the exchange of resources to the transmission of information. Indeed, in line with the conceptual framework of Manski (2000), a social interaction takes place whenever an individual develops a preference, forms an expectation, or faces a constraint as a result of the outcomes or behaviour of another.[4] As discussed by Akerlof (1997), social interactions can be thought to occur in a 'social space' whereby interaction is more likely among socially proximate individuals, such as friends and relatives. The resemblance to Tobler's first law is not coincidental; in fact, the methodological literatures on dependence in both kinds of space are intrinsically linked.[5] This becomes intuitive in light of the discussion in the previous subsection: whether a neighbourhood is defined on the basis of geographic or social space, their analytical treatment can be very similar. It is for this reason that the term 'neighbourhood effects' is often equally used to describe dependence in both geographic and social space, just as the term 'neighbourhood' is used to describe relations between locations projected on either of the two.[6]

[3] Though note that non-geographic weight matrices will have distinct implications for causal inference when deployed in a spatial econometric analysis (e.g. see LeSage, 2014).

[4] Here, Manski (2000) positions social interactions in a utility maximisation framework.

[5] A detailed discussion of the correspondence between spatial econometrics and the econometrics of social interactions is offered by Gibbons et al. (2015).

[6] Other terms are also employed in particular domains and applications, such as 'peer effects', 'network effects', and 'spillover effects'. In line with Durlauf (2004), Ioannides (2013), and others, I consider these and other related terms as subsets of the term 'neighbourhood effects'.

Whatever their conceptualisation, the identification of neighbourhood effects is notoriously elusive (e.g. see Gibbons et al., 2015; Manski, 2000). Researchers in the social sciences are usually interested in one or more of the parameters in the equation below:[7]

$$y = \delta \mathbf{W} y + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \lambda \mathbf{W} u + \epsilon \tag{I.5}$$

Here, for n locations and k variables, y and u are respectively $n \times 1$ vectors of outcomes and unobservables, $\mathbf{X}$ is a $n \times k$ matrix of covariates, $\mathbf{W}$ is a $n \times n$ weights matrix defining neighbourhood structure, β and θ are $k \times 1$ vectors of parameters, while δ and λ are scalar parameters. Notice that the above implies two kinds of neighbourhood effect: that of neighbourhood outcomes on own outcomes captured by δ, and that of neighbourhood characteristics on own outcomes captured by θ—what Manski (1993) respectively refers to as 'endogenous' and 'contextual' interactions. In the absence of restrictions, this gives rise to Manski's 'reflection problem', whereby perfect collinearity between neighbourhood outcomes and characteristics prevents separately estimating both kinds of effect. Notice further that the presence of common unobserved characteristics, u, in locations within neighbourhoods may give rise to 'correlated effects' that threaten identification.[8] This may equally arise when there is endogenous sorting into neighbourhoods, or when the neighbourhoods are exogenous and outcomes are correlated with unobserved characteristics.

A growing literature investigates the kinds of restrictions that might be imposed in addressing the reflection problem and correlated effects (e.g. see Bramoullé et al., 2020 and Gibbons et al., 2015). An interesting case arises when one is interested in the neighbourhood effects of a particular treatment, as is often the case in applied settings. As shown by Dieye et al. (2014) and Bramoullé et al. (2020), given that the treatment affects outcomes both directly and indirectly through neighbourhood outcomes and characteristics, and is randomly assigned and uncorrelated with neighbourhood structure, it is in principle possible to separately estimate the three effects. In practice, control over randomisation is rare in many domains in which neighbourhood effects may be of interest. Here, some progress can be made by exploiting quasi-random variation in exposure to the treatment in line with the experimentalist paradigm that is adhered to in many modern observational studies (e.g. Angrist and Pischke, 2010).

[7] This is akin to the general nesting model (GNS) as this is conventionally expressed in the spatial econometric literature (e.g. Elhorst, 2014), which is preferred for simplicity. Different notation is typically employed in the econometrics of social interactions, while additional lagged terms may also appended for different units of analysis or time periods (e.g. see Gibbons et al., 2015).

[8] Using the example of students nested in schools as in Manski (1993), endogenous interactions arise when student achievement varies with the average achievement of others in the school, while contextual interactions arise when achievement responds to the socio-economic composition of the school. In contrast, correlated effects are exemplified by similar achievement due to shared teachers.

In such cases, Gibbons and Overman (2012) advance the argument for targeting the estimation of two rather than three parameters: one relating to the direct effect of exposure to the treatment and a composite parameter capturing both kinds of neighbourhood effect. Namely, these can be estimated using an ordinary least squares (OLS) regression of the outcome of interest on direct and neighbourhood exposure to the treatment—what is often referred to as the spatial lag of X (SLX) model in spatial econometrics—provided that variation in both of the latter is exogenous.[9] A key benefit of this approach is its consistency with several well-understood quasi-experimental estimators, which aid the task of isolating exogenous variation (Halleck Vega and Elhorst, 2015). As such, while this essentially amounts to evading the reflection problem and thus restricts one's ability to pick apart the channels via which neighbourhood effects are being transmitted, it offers a useful empirical framework for separating these from correlated effects, which is often substantively relevant in itself.

I.2 Overview of the Thesis

I.2.1 Background and Motivation

Few research domains have seen as sustained an interest in neighbourhood effects as that shown in the study of elections. Following the early studies of Tingsten (1937) and Key (1949), a long line of work has documented relationships between voters' political preferences and the aggregate characteristics of people in their residential environments.[10] Since the work of Cox (1969), who—much like Tobler (1970)—observed that there is 'a strong inverse relationship between distance and the formation of acquaintanceship' (92), it has been common to interpret such relationships as evidence of social interactions. The famous study of voters in the city of South Bend, Indiana by Huckfeldt and Sprague (1995) was perhaps most influential in this regard. Having collated one of the most thorough relational datasets of the time, the authors showed that voters' political preferences varied with such factors as the frequency and length of conversation with various peers in the local community. This further solidified a then developing and now widely held view in the empirical study of elections. That is, a voter's area of residence is to be seen as overlayed by a dense social network from which they receive politically relevant information—whether this flows through direct conversations among peers or more indirect means, such as observation (Books and Prysby, 1991). In other words, the geographic distance between voters' residences is, among other things, a proxy for the probability of social interaction between them.

[9] With reference to (I.5), this would yield estimates of β and $\pi = (\theta + \beta\delta) \div (1 - \delta)$, respectively representing the treatment and composite neighbourhood effects (Gibbons and Overman, 2012).

[10] While this literature is too vast to be contained in any one account, good reviews are offered by Ethington and McDaniel (2007), Gimpel and Reeves (2022), and Johnston and Pattie (2014).

Interestingly, not too long after the seminal study of Huckfeldt and Sprague (1995), Baybeck and Huckfeldt (2002a,b) would carry out a further detailed survey of voters in the cities of St Louis and Indianapolis, which received relatively limited attention. In it, the authors had a closer look at the geographic structure of voters' social networks: a somewhat peripheral topic in the South Bend study. Much like Cox (1969) had observed, they found that the probability of two voters being acquainted to each other declined steeply as the geographic distance between their residences increased. However, there was also substantial variation in the geographic distribution of individual voters' acquaintances. The authors further demonstrated that the geographically dispersed social networks arising from this variation were 'more likely to connect individuals that reside in socially and politically divergent settings' (Baybeck and Huckfeldt, 2002b, 217). Interpreting these findings in light of the conceptual framework of Granovetter (1973, 1983)—whereby 'social ties' are interpersonal relationships of varying strengths and 'weak ties' are social ties of relatively low strength that are likely to bridge disparate cliques of individuals—they also hypothesised that weak ties were likely to facilitate the spatial diffusion of politically relevant information along such networks. In other words, the information flows that feed into a voter's political preferences may well originate from a different residential environment than their own, and the extent in which this might be the case is likely to depend on the geographic structure of the localised social network in which the voter is embedded.

At the time of its publication, the study of Baybeck and Huckfeldt (2002a,b) was among just a handful of others measuring the geographic dispersion of social ties across different communities (e.g. see also Fischer, 1982; Wellman, 1979, 1996). However, interest in the topic grew with the uptake of the Internet and new communications technologies, partly owing to popular narratives on the diminishing role of geographic distance in constraining social interactions (e.g. Cairncross, 2001; Friedman, 2005). Recently, this culminated into a set of prominent studies by Bailey et al. (2018, 2020,2020) which provided one of the most comprehensive records of the global structure of social networks to date. Using data on the entire population of active users of Facebook—the popular social networking service—the authors developed a 'Social Connectedness Index' that measures the pairwise density of social ties between the user populations resident in different, granular spatial units.

The Social Connectedness Index is of particular relevance to Western countries where a commanding majority of adults were estimated to be users of the service at the time of its construction, such as the United States. Interestingly, in a similar way to Cox (1969), Bailey et al. (2018) showed that a substantial amount of the variation in the aggregate social ties between county-pairs is explained by geographic distance. Nevertheless, the authors also found strong evidence of variation in the geographic dispersion of social ties, with the population of a county having, on average, over 20 per cent of its aggregate social ties in counties that are located over 500 miles away.

In line with the rationale advanced by Baybeck and Huckfeldt (2002a,b), this could mean that social influences on voting behaviour in contemporary elections may also originate from substantially more distant localities than those typically considered. Indeed, there is an emerging body of evidence that is consistent with such long-distance 'social spillover effects' in a number of other domains, such as public health (Charoenwong et al., 2020; Holtz et al., 2020; Zhao et al., 2021), consumer behaviour (Makridis, 2022), and public finance (Wilson, 2022). Further reinforcing the relevance of this prospect in elections is the growing evidence on the political effects of social media and the Internet (Geraci et al., 2022; Zhuravskaya et al., 2020).[11] However, to this date, little empirical attention has been dedicated to the relationship between the geographic structure of social networks and electoral outcomes. This thesis constitutes one of the earliest attempts at addressing this gap in the literature.

I.2.2 Research Aims and Questions

The overall aim of this thesis is to examine how the aggregate social ties between the resident populations of different places relate to voting behaviour. In line with the conceptual discussion in Section I.1, this runs parallel to the methodological aim of exploring how aggregate social media data can be employed within a spatial econometric framework in tackling policy-relevant research questions in the empirical study of elections. These aims are pursued in the form of three empirical chapters, with each focusing on a particular research question relating to recent electoral events in the United Kingdom and the United States, and utilising the Social Connectedness Index by Bailey et al. (2018, 2020,2020) to flexibly define the geographic space over which social interactions are likely to occur. The research questions are listed below, indexed according to the corresponding chapter:

- RQ1 (Chapter One): How was support for the Leave option in the 2016 UK EU membership referendum affected by voters' social interactions with those exposed to import competition in other local labour markets?

- RQ2 (Chapter Two): How was the demand for mail ballots in no-excuse absentee voting jurisdictions in the 2020 US presidential election affected by voters' social interactions with those in jurisdictions switching to all-mail voting?

- RQ3 (Chapter Three): How did partisan homophily among the populations of different localities vary in geographic space in the 2020 US presidential election?

[11]Note that a number of studies of online social networks suggest that geographic distance can constrain both the volume and strength of social ties (e.g. see Gimpel and Reeves, 2022). However, in line with Granovetter (1973, 1983), insofar as new communications technologies facilitate information flows along weak ties, this can still have important implications for the geographic range of social influence.

As such, each chapter can be seen as making a substantive contribution to each of three distinct bodies of related literatures in the empirical study of elections. It is thus in principle possible for each chapter to be read independently from others and this introduction. Though as the chapters are closely tied by a common thematic and methodological core, the concepts and data discussed in this introduction will resonate across chapters. I provide an overview of each chapter in the subsections that follow.

I.2.3 Overview of Chapter One

The first chapter looks at the 2016 EU membership referendum in the United Kingdom, asking whether local exposure to trade-related employment shocks had social spillover effects on support for the Leave option in other localities. Regional exposure to Chinese import competition has been identified as an important driver of Brexit (Ballard-Rosa et al., 2021; Colantone and Stanig, 2018a; Steiner and Harms, 2021), which has contributed to growing interest in 'left-behind places' and place-based policy (Martin et al., 2021). Following on from the literature on the economic effects of trade (e.g. Acemoglu et al., 2016; Autor et al., 2013), research on its political effects has assumed that these are primarily contained within exposed local labour markets. I argue that while this might be the case for economic spillovers operating via industry linkages, aggregate demand, and reallocation, it is less likely to be so for social spillovers operating via the flow of information along social networks. In order to test this, I construct two measures of exposure to import competition for International Territorial Level 3 (ITL3) regions in England and Wales: one of exposure within the region and one of exposure in socially connected regions outside of the local labour market within which it is situated. Adopting an instrumental variable approach akin to that of Autor et al. (2013), and using both official election results and individual-level data from the British Election Study, I show that the two measures have comparable positive effects on voting for Leave in the referendum. Importantly, in a series of further checks, I find no evidence to suggest that the spillovers from the exposure of socially connected regions are accounted for by interregional economic links or restricted to voters facing economic difficulty. Overall, this chapter suggests that information flows between local labour markets are an important channel via which the political effects of economic shocks propagate, contributing new evidence on the political dimensions of 'left-behind places' and place-based policy.

I.2.4 Overview of Chapter Two

The second chapter considers the socio-spatial spillovers of electoral reforms on choice of voting method in the 2020 US presidential election. Since the COVID-19 pandemic, election officials across several jurisdictions have been faced with increases in the demand for voting by mail (Fortier and Stewart, 2021). Gauging this demand in advance is important for election administration as it has implications the acquisition and allocation of the relevant resources.

While previous work suggests that the local rollout of all-mail voting policies is one of the most relevant factors in the take-up of voting by mail (Herrnson and Stewart, 2023), little is known about its potential spillovers in non-local jurisdictions where mail ballots are widely available on request. I address this gap by leveraging a detailed administrative dataset on the entirety of registered voters in the state of North Carolina. Exploiting county-level variation in the rollout of all-mail voting induced by the COVID-19 pandemic, for each zip code tabulation area (ZCTA) in the state, I calculate the share of aggregate social ties in counties that switched to the policy between the 2016 and 2020 US presidential elections. Importantly, I show that all counties that switched to all-mail voting between elections are located in states lying several hundreds of kilometres away. Employing a difference-in-differences approach, I find that voters residing in ZCTAs with higher shares of aggregate social ties in policy-switching counties were substantially more likely to vote by mail in 2020. I also show that these socio-spatial spillovers were especially relevant for older voters, registered Democrats and unaffiliated voters, and those resident in metropolitan counties. In all, the evidence suggests that local rollouts of all-mail voting are likely to drive demand for mail ballots in non-local, socially connected jurisdictions, showcasing the utility of social media data in anticipating the social spillovers of electoral reforms.

I.2.5 Overview of Chapter Three

The third chapter explores the geography of partisan homophily in the 2020 US presidential election. In so doing, it takes a more descriptive approach compared to the previous two chapters, focusing on the characteristics of the 'neighbourhoods' defined by the geographic structure of social networks rather than the estimation of their causal effects. Limited inter-partisan contact is usually put forward as a contributing factor to the rise of political polarisation in the United States in recent years (Boxell et al., 2020; Enos, 2017; Finkel et al., 2020). Though the presence of the former is typically inferred from studies of partisan segregation in geographic space, which reveal a stark urban-rural divide in voting behaviour (e.g. Darmofal and Strickler, 2019; Brown and Enos, 2021). Adopting a similar measurement approach to such studies, but using data on aggregate social ties rather than geographic distance, I employ the local Moran index to identify clusters of ZCTAs that are segregated by partisanship in social space. In a series of multinomial logit specifications, I also explore the probability of membership in each cluster for ZCTAs in different regions and types of settlement, and with differences in the relative density and geographic distance of aggregate social ties in other areas. The empirical findings are broadly consistent with the expectations arising in studies of partisan segregation: as of the 2020 US presidential election, most areas were more likely to be socially connected to others with a similar composition of vote shares, with more urban areas being more likely to be both homophilous and Democratic-leaning.

However, I also show that the areas with the most relatively dense and geographically distant social ties are also most likely to fall in the latter category, indicating substantial spatial variation in the geographic range of co-partisan contact. These findings suggest that the political preferences of voters in urban, Democratic-leaning areas are likely to be more exposed to social spillovers from remote residential environments.

This page has been intentionally left blank.

CHAPTER ONE

Social Networks and Brexit: Evidence from a Trade Shock

Regional exposure to Chinese import competition has often been linked to support for the Leave option in the 2016 UK EU membership referendum. Looking at 143 harmonised International Territorial Level 3 regions covering England and Wales, and using data on the density of online social ties between them, I show that regional support for leaving the EU was also associated with exposure in socially connected regions. I first delineate 18 commuting zones based on interregional flows over three Census years. For each region, I then construct a measure of own exposure to Chinese import competition and a measure of exposure in a set of social neighbours located outside its commuting zone. Exploiting variation within commuting zones, and using an instrumental variable approach, I find that the two measures have comparable positive effects on the regional share of the Leave vote. In a series of checks, I do not find evidence that the effect of social neighbours' exposure is driven by an economic channel or a relationship between import competition and social ties. I also corroborate the regional results using survey data on vote choice. I interpret these findings as indicative of social spillovers between local labour markets: information flows from social neighbours are a likely channel behind the estimated spillover effects on voting outcomes.

1.1 Introduction

The result of the 2016 European Union membership referendum is often said to betray a ‘geography of discontent’ in the United Kingdom (McCann and Ortega-Argilés, 2021). Motivating this diagnosis is the link between local economic decline and voting to leave the EU (e.g., Becker et al., 2017; Colantone and Stanig, 2018a; Harris and Charlton, 2016)—the option that was formally opposed by the incumbent government before the referendum was held. Looking through the areas where the Leave option prevailed, one can quickly find ‘places that don’t matter’ (Rodríguez-Pose, 2018): former industrial hubs that have been adversely affected by economic globalisation. This observation has revitalised the debate on the political significance of spatial inequality and its policy implications for both the UK and the EU (e.g., Iammarino et al., 2018; Martin et al., 2021; McCann, 2020; Rajan, 2020; Sandbu, 2022). I contribute to this debate by showing that local support for leaving the EU is also likely to have responded to spillovers from exposure to global trade in other local labour markets.

I focus on the rise of China as a major exporter and its effects on the share of the Leave vote in 143 harmonised ITL3 regions covering the whole of England and Wales.[1] This contribution is thus closely related to existing work on the political effects of regional exposure to the ‘Chinese import shock’ in Great Britain (Ballard-Rosa et al., 2021; Colantone and Stanig, 2018a; Steiner and Harms, 2021). However, rather than only measuring within-region exposure, I also use the Social Connectedness Index (SCI) by Bailey et al. (2018, 2020) to construct a weighted measure of exposure in a set of five socially connected regions located outside each region’s local labour market. Comparing regions within local labour markets, I find that an increase by one standard deviation in either measure corresponds to an increase in the regional share of the Leave vote by roughly 3 percentage points. Adopting the shift-share instrumental variable approach of Autor et al. (2013), I retrieve supply-driven components of both measures using Chinese imports into the United States and obtain new estimates, which remain very close to the baseline. In augmented specifications, I find that spillover effects decay in the second set of five nearest social neighbours. Still, spillovers from the first set seem to travel a fair geographic distance: on average, a region’s five nearest social neighbours are between 74 and 102 kilometres away.

[1] ITL3 refers to the third level of disaggregation of the International Territorial Level classification. Following Brexit, ITL3 replaced the same level of the EU-wide Nomenclature for the Use of Territorial Statistics (NUTS3) in the publication of UK official statistics. The initial version of ITL3, which is the one employed in this chapter, mirrors the 2021 version of NUTS3. As I discuss in Appendix 1.A.1, creating a harmonised ITL3 classification—whereby a small number of ITL3 units are aggregated—allows me to uniformly observe all relevant variables.

Following the economic literature on the effects of import competition in the US, I define local labour markets as commuting zones derived using the hierarchical clustering approach of Tolbert and Sizer (1996). Specifically, I assign each harmonised ITL3 region to one of 18 commuting zones based on the mean commuting dissimilarity between region-pairs over three Census years. These aid me in addressing the problem of 'correlated effects' (Manski, 1993) under the identifying assumption that regional exposure to Chinese import competition is conditionally exogenous and does not affect social ties (see Bramoullé et al., 2020). As regions within the same local labour market are likely to be exposed to common unobserved shocks, commuting zones help me to account for such heterogeneity. In addition, as the Chinese import shock precedes the year in which the SCI is measured, restricting each region's social neighbours to be located in other commuting zones ensures that the social ties considered are unaffected by past mobility responses within local labour markets.

In principle, social ties between a region and its social neighbours could still be related the import shock via other channels, such as migration between local labour markets or social sorting. Measurement error would arise if the social weights within a region's set of social neighbours were affected by their exposure, while spillover effects would be overstated if the latter was positively associated with the probability of being in the set. In supplementary specifications, I fail to find evidence consistent with either scenario. Comparing social neighbours within regions, the effect of import-shock exposure on the density of social ties with the focal region is weak and indistinguishable from zero. Similarly, the probability of a region being in another's set of ten nearest social neighbours is weakly and negatively associated with its exposure. As such, despite the possibility of a small downward bias in the estimation of spillover effects, the empirical strategy is likely to satisfy the conditions for their detection.

I argue that there are two probable channels behind the estimated effects of social neighbours' import-shock exposure on the regional share of the Leave vote: economic spillovers operating via indirect effects on linked industries, aggregate demand effects, and reallocation effects (see Acemoglu et al., 2016), and social spillovers operating via information flows over social networks. Consistent with the latter explanation, I find that estimates are robust to the inclusion of regional controls for the start-of-period employment share in manufacturing, relative growth in GVA, and net-in migration from social neighbours. This is not a surprising result given that a large component of the economic spillovers of regional exposure to the shock will be contained within local labour markets (e.g., see Autor et al. 2016; Dorn and Levell 2021). A further analysis using individual-level data from Wave 9 of the British Election Study mirrors the baseline regional results. Importantly, I show that spillover effects are not restricted to voters facing economic difficulty.

In all, the empirical results suggest that social spillovers between local labour markets in England and Wales are likely to have been an important channel through which import competition with China affected voting in the 2016 EU membership referendum. This evidence offers a new perspective on the potential role of spatially heterogeneous economic shocks in national voting outcomes.

1.2 Background

An extensive literature documents that the growing participation of China in global trade since the early 1990s has had adverse effects on labour market outcomes in many Western regions that used to rely on manufacturing employment (e.g. see Autor et al., 2016; Dorn and Levell, 2021). Aided by the country's status as a major low-wage producer, Chinese imports displaced production in Western import-competing industries, resulting in substantial employment and wage losses in highly exposed local labour markets. Shadowing its economic effects, the political effects of the shock have also been observed in several countries: higher local exposure has been tied to increased support for Eurosceptic candidates in national elections within EU member states (e.g., Barone and Kreuter, 2021; Colantone and Stanig, 2018c; Dippel et al., 2021; Malgouyres, 2017), for Donald Trump in US presidential elections (Autor et al., 2020), and for the Leave option in the 2016 UK EU membership referendum (Colantone and Stanig, 2018a).[2]

A number of studies on the relationship between regional exposure to import competition and voting behaviour point to the presence of neighbourhood effects. For instance, Colantone and Stanig (2018a) measure exposure in Great Britain at the level of NUTS3 regions and find that its positive effect on individual-level support for Brexit persists when accounting for employment status and occupation. Similarly, Steiner and Harms (2021) measure exposure both at the level of NUTS3 regions and occupation classes, finding that the negative effect of regional exposure on individual-level support for EU membership persists when holding occupational exposure constant. Looking at individuals in the UK and and the EU, Colantone and Stanig (2018c) and Hays et al. (2019) also note that exposure at the level of NUTS2 regions corresponds to higher support for anti-establishment parties irrespective of employment status. As the authors argue, these findings suggest that labour market outcomes are not the only channel via which regional exposure to import competition affects voting.

[2] Guriev and Pappaioannou (2021) and Rodrik (2021) provide comprehensive reviews of the literature on the political effects of import competition in various countries.

Electoral geographers and political scientists have often argued that one of the reasons that one might observe neighbourhood effects is the high density of social ties among geographically proximate individuals (Cox, 1969; Johnston and Pattie, 2000, 2011, 2014; Miller, 1977). If a group of individuals is closely bound by interpersonal relationships that stimulate the flow of information, then the exposure of some to an economic shock may easily feed into the political preferences of others. Indeed, a nascent empirical literature is consistent with this prospect. Examining the policy preferences of Spanish voters, Liu et al. (2020) find that when almost half of one's friends and acquaintances have experienced economic hardship during the Great Recession, support for staying in the Euro currency is about 10 percentage points lower compared to when no peers are affected. Similarly, looking at the behaviour of Danish voters, Alt et al. (2022) find that a percentage point increase in the share of 2nd degree peers that have recently become unemployed increases the probability of voting for a left-wing party by 3.7 percentage points.[3]

While there is strong evidence to suggest that social ties decay with geographic distance (Bailey et al., 2018, 2020), treating geographic neighbourhoods and the localised social networks they contain as 'social islands' may conceal substantial spillover effects on voting outcomes. As Johnston and Pattie anticipate, 'such [localised] social networks are extremely unlikely to be isolated – many members will have links to either or both of other, nonlocal networks (based on workplaces or family/kin, for example) and separate networks in adjacent neighbourhoods: such external links are continual sources of new information to the importing networks, providing stimuli to which they respond, in some cases altering their attitudes and behaviour as a consequence' (2014, 3). That is, the behaviour of residents in any given area are likely to be shaped by conditions in other areas with which they have social ties.

To date, studies on the effect of the Chinese import shock on Brexit have assumed that different regions are independent of each other. This is in line with the expectation that the adverse economic effects of local exposure to the shock will be largely contained within local labour markets, which is a common relaxing assumption in the economic literature. Yet, even if economic spillovers between local labour markets are indeed too modest to bear any substantial political effects, the same may not hold for social spillovers involving the flow of information. This chapter therefore asks whether import competition with China is likely to have shaped voting in the referendum via this unexplored channel.

---

[3] I refer to n^{th} degree peers as those separated by an unweighted shortest path of length n along dyadic ties. That is, 1st degree peers are tied with each other, 2nd degree peers are not tied but have at least one common 1st degree peer, and so on.

1.3 Conceptual Framework

In this section, I introduce a simple conceptual framework that informs the empirical strategy of this chapter. This rests on the intuition that there are two main channels through which exposure to import competition in one region can influence voting outcomes in its social neighbours: (1) 'economic spillovers' operating via indirect effects on linked industries, aggregate demand effects, and reallocation and mobility effects, and (2) 'social spillovers' operating via information flows over social networks. Insofar as more economically integrated regions are likely to have more dense social ties, the two channels will not be readily distinguishable from each other. In addition, given that social ties as measured by the SCI are observed at a later time than the Chinese import shock, the former may be affected by the latter, posing challenges in the identification of spillover effects. As such, a conceptual discussion of the kinds of processes that fall under each channel is informative with respect to both the interpretation of empirical results and the ways in which these issues may be addressed. As I ground this discussion on relevant theories of voting behaviour, it is important to stress that the empirical analysis is not directly informative about the specific individual-level mechanisms anticipated by these theories.

1.3.1 Economic Voting and Economic Spillovers

Studies on the political effects of the Chinese import shock often subscribe to the economic voting hypothesis (Kinder and Kiewiet, 1981). Namely, they anticipate that voting behaviour is influenced by both personal economic circumstances (egotropic voting) and the state of the local economy (sociotropic voting).[4] The former case is exemplified by a voter experiencing a loss of employment, income, or wealth, and then becoming disaffected with incumbents or particular policies. However, even when one suffers no personal economic losses as a result of an import shock, their political preferences and voting behaviour may still sociotropically respond to its salient effects in the area in which they reside, such as deteriorating local public services (Feler and Senses, 2017), closures of local businesses and venues (Bolet, 2021), crime (Che et al., 2018; Dix-Carneiro et al., 2018) and anti-social behaviours such as substance abuse (Pierce and Schott, 2020).[5]

[4]While Kinder and Kiewiet (1981) originally defined sociotropic voting as relating to the state of the national economy, subsequent work recognises that voters often respond to local economic conditions (Kiewiet and Lewis-Beck, 2011). Local economic conditions have also been found to inform voters' evaluations of the national economy (Ansolabehere et al., 2014; Hansford and Gomez, 2015).

[5]As Kiewiet and Lewis-Beck (2011) suggest, sociotropic voting is often misrepresented as necessarily altruistic. However, insofar as local economic conditions shape one's own economic expectations, sociotropic voting can also be motivated by self-interest.

While egotropic and sociotropic voting may often be explicitly motivated by economic concerns, this need not be the case. For instance, Colantone and Stanig (2018b) show that local exposure to the Chinese import shock can also activate concerns about the cultural threat posed by immigration—a common issue in Eurosceptic campaigning across EU member states. Similarly, observing the positive effect of pub closures on voting for the Eurosceptic UKIP party in the UK, Bolet (2021) argues that the socio-cultural degradation induced by local exposure to an economic shock can also hurt residents' sense of place-based identity. As such, in the interest of simplicity, I refer to any behaviour that responds to personal economic circumstances as egotropic, and to any behaviour that responds to changes in local economic conditions as sociotropic—irrespective of whether such behaviours are primarily motivated by economic or other concerns.

Regional exposure to import competition can induce egotropic or sociotropic voting responses in other regions via economic spillovers. Egotropic voting responses will occur in other regions insofar as personal economic circumstances within them are impacted by interdependencies with the exposed region. Likewise, sociotropic voting responses may also occur if interdependencies with the exposed region affect the overall state of other regional economies. What do such interdependencies look like? Acemoglu et al. (2016) develop a framework for decomposing the employment impact of Chinese import competition within a local labour market, which is also informative about the kinds of economic spillovers that are possible. Namely, the local employment impact will be comprised of (a) direct effects on exposed industries, (b) indirect effects on linked industries, (c) reallocation effects, (d) and aggregate demand effects.

The direct effects of exposure to Chinese import competition within a local labour market are the job and wage losses of workers in local industries whose products directly compete with Chinese imports. Though firms operating in these industries are also likely to be in supplier or buyer relationships with other local and non-local firms operating in linked industries. Their direct exposure to the shock may thus also beget indirect effects on workers in those other firms. These may be contractionary 'upstream' effects whereby suppliers are themselves hit by suppressed demand from their buyers, or 'downstream' effects whereby buyers either face supply chain issues or benefit from the availability of cheaper alternatives. While an extensive theoretical and empirical literature on agglomeration economies establishes that firms in linked industries tend to be co-located (e.g., Glaeser 2010), some may not be.[6] As such, direct and indirect effects on local industries within one local labour market may also bear indirect effects on linked industries in others, in turn affecting voting behaviour.

[6] Lavoratori and Castellani (2021) illustrate the agglomeration of UK manufacturing firms.

Operating alongside direct and indirect effects is reallocation—the expansion of industries that seize the freed labour and capital. Notably, this includes displaced workers moving on to different firms. Again, the evidence on labour mobility responses to import competition suggests that this is unlikely to involve substantial flows between local labour markets (e.g., Autor et al., 2016; Dorn and Levell, 2021). Still, mobility may be more likely between certain regions, such as those with strong pre-existing social ties (e.g., see Munshi, 2020); in such cases, reallocation between local labour markets could plausibly be strong enough to influence voting outcomes.

Finally, aggregate demand effects relate to changes in consumption and investment. Such is the case when reduced spending on local goods and services results in additional employment losses in non-tradable sectors. Here too, as workers typically buy goods and services from firms that are in proximity to their residences and workplaces, a large component of the aggregate demand effects of a given local shock is likely to be contained within the exposed local labour market (e.g., Mian et al., 2013; Mian and Sufi, 2014). Though there are ways in which spending may be routinely carried out elsewhere, such as travel (e.g., Dunn and Gholizadeh, 2020). As such, insofar as a local shock affects spending in other local labour markets, spillover effects on voting outcomes may be expected to follow.

1.3.2 Social Influence and Social Spillovers

The economic voting hypothesis anticipates how aggregate economic conditions in the geographic neighbourhood in which voters are embedded can affect their behaviour. Though it is less informative about the ways in which this can be shaped by the economic circumstances of 1$^{\text{st}}$ or n^{th} degree peers in their social networks who may well reside elsewhere. Theories on the role of social influence in voting behaviour are attentive to two main mechanisms: information diffusion (e.g., Huckfeldt and Sprague, 1995) and conformity (e.g., Berelson et al., 1986). While all social influence may require the flow of information in some manner, the latter mechanism is more restrictive in that it presupposes that a particular preference is already held by a 1$^{\text{st}}$ degree peer—or indeed a group of 1$^{\text{st}}$ degree peers—before it gets passed on to another through persuasion or learned norms. In contrast, simple information diffusion includes cases where preferences respond to information received from others with no regard for their preferences.

The study of Alt et al. (2022) provides an intuitive setting in thinking about how economic circumstances within one's social network may influence voting through either mechanism. As the authors link unemployment shocks in Danish voters' 2$^{\text{nd}}$ degree peers with various political preferences, there are two elementary scenarios. Strict information diffusion would involve the focal peer merely becoming aware of the job loss of a 2$^{\text{nd}}$ degree peer and, as a result, adjusting their political preferences.

On the other hand, a strict conformity scenario would require that a 2nd degree peer adjusts their preferences after becoming unemployed, then compels a 1st degree peer to also adjust their preferences in a similar direction, who in turn compels the focal peer to do the same.[7] Of course, it is possible that the two mechanisms are mixed. For instance, it could be that the preferences of the 1st degree peer simply respond to the information that the 2nd degree peer has become unemployed and then the focal peer is persuaded to conform. It is also not difficult to conceive of more complex cases where the unemployment shock that triggers the information diffusion or conformity pressure that ultimately affects the focal peer's preferences originates from a peer of even higher degree. For simplicity, I thus refer to any combination of information diffusion and conformity pressures involving peers of any degree as information flows. Importantly, I make no distinction with respect to the mode over which information flows are transmitted. This may equally involve peers meeting in-person, visiting each other's neighbourhoods, telephone contact, text messaging, or interactions over the Internet and social media.

Regional exposure to import competition may affect voting outcomes elsewhere via social spillovers. That is, as the shock affects economic conditions and political preferences in one region, information flows relating to these conditions and preferences can affect voting behaviour in others. A burgeoning empirical literature shows that social influence is characterised by complex contagion (e.g, Christakis and Fowler, 2007; Sprague and House, 2017; Törnberg, 2018). That is, the probability of an individual changing their behaviour is conditional on the number of peers from which they receive a signal, the frequency in which they receive it, as well as the strength of their relationship with peers. Indeed, Alt et al. (2022) and Liu et al. (2020) respectively associate voter preferences to the shares of 2nd and 1st degree peers experiencing economic adversity, rather than considering a binary treatment. Social spillovers between regions are thus likely to be determined both by the extent of their exposure to import competition as well as the density of 1st degree social ties between them.

1.4 Data

1.4.1 Harmonised Regions

The prerequisite data are not all readily available for the same spatial units. As detailed in Appendix 1.A.1, I address this by drawing on the Code History Database produced by the Office for National Statistics (ONS) and correspondence tables produced by Eurostat to create harmonised regions. First, I create a harmonised local authority district (LAD) classification which is comprised of 366 regions that cover England, Scotland, and Wales.

[7] Alt et al. (2022) suggest that a strict conformity mechanism is unlikely in their study setting.

I then also create a harmonised ITL3 classification comprised of 143 regions covering England and Wales. The former classification ensures that the majority of variables, which are available for different versions and variants of LAD, can be uniformly observed over time, while the former ensures that these can be aggregated to the same level as data on social ties, which are available for the NUTS3 2016 classification—a predecessor of ITL3. However, while LAD in England and Wales are nested within ITL3 regions, the same does not hold for Scotland, meaning that it is necessarily excluded from main empirical specifications. Still, as discussed later in this section, Scotland can still be considered in correcting measures of regional exposure to import competition in England and Wales for measurement error, hence its inclusion in the harmonised LAD classification. Notably, Northern Ireland is wholly excluded from the analysis due to data limitations.

1.4.2 Mobility

I obtain data on interregional commuting flows measured at the time of each UK Census via the public 'Flow Data' portal operated by the UK Data Service. Importantly, these flows also provide information on the size of the resident workforce, as the data include the number of workers commuting within their region of residence. I also obtain data on interregional migration from ONS publications on annual internal migration estimates.

1.4.3 Social Ties

The SCI by Bailey et al. (2018, 2020) serves as the main measure of interregional social ties. The public release of the SCI used in this chapter was accessed via the Humanitarian Data Exchange of the UN Office for the Coordination of Humanitarian Affairs and is based on a snapshot of all active Facebook users (i.e. those who have logged in during the previous 30 days) as of August 2020. For NUTS3 2016 region-pairs uu', the SCI is defined as follows:[8]

$$SCI_{uu'} = SCI_{u'u} = \frac{FB\_Connections_{uu'}}{FB\_Users_u \times FB\_Users_{u'}} \tag{1.1}$$

That is, the SCI is the total number of Facebook friendships in both regions over the product of active users in each region. It can therefore be thought of as a measure of the relative probability of Facebook friendship between an active user that resides in region u and an active user that resides in u', or the density of Facebook friendships between the active user populations of each region.

[8]Before release, the SCI is scaled so that the global index has a maximum value of 1,000,000,000 and a minimum of 1. For region-pairs in England and Wales, the maximum is 2,871,114 and the minimum is 1,620, excluding each region's connectedness with itself.

I convert the SCI to the harmonised ITL3 region classification using the approach of Bailey et al. (2021).[9] For the vast majority of units, there is 1-to-1 correspondence between the latter and the NUTS3 2016 classification in which the SCI is available. Other than recoding, the conversion thus effectively involves aggregating SCI values relating to the four NUTS3 child regions referenced on Appendix Table 1.A.3 into values relating to the two harmonised ITL3 parent regions also shown on the same table. Formally, I recalculate the SCI as follows:

$$SCI_{rr'} = SCI_{r'r} = \sum_{u} \sum_{u'} PopShare_u \times PopShare_{u'} \times SCI_{uu'} \tag{1.2}$$

Here, u and u' are NUTS3 regions respectively mapping into harmonised ITL3 regions r and r', and $PopShare_u$ is the population of u as a share of the population of r based on ONS mid-2020 estimates. Following the conversion, I also set SCI values where $u = u'$ to zero, effectively discarding each region's ties with itself.

As the SCI is based on social media friendships, it is arguably not a perfect representation of the true density of social ties between regional populations. However, as the UK Office for Communications (2020) estimates that approximately 72 per cent of UK adults were social media users in 2020, and that 88 per cent of those had a Facebook account (i.e. over 63 per cent of all UK adults), the SCI likely constitutes a sound proxy for both online and offline social ties among regions.

1.4.4 Regional Exposure to Import Competition

I obtain data on regional employment by industry from the 1991 ONS Annual Employment Survey (AES) via the public 'Nomis' portal operated by the ONS. Industry breakdowns are accessed at the 3-digit level of the 1992 Standard Industrial Classification (SIC), which is equivalent to the first revision of the Statistical Classification of Economic Activities in the European Community (NACE Rev 1). Notably, regional figures produced by the survey relate to persons employed within the region as opposed to the residents of that region. Annual data on the value of goods imported from China into the UK are drawn from the UN International Trade Statistics Database (COMTRADE) for product categories of the third revision of the Standard Industrial Trade Classification (SITC Rev.3). Product breakdowns are accessed at the lowest level that is available in both 1991 and 2007, which mark the period over which import growth is observed. As imports are reported in US dollars and current prices, I also use annual currency conversion factors available on COMTRADE along with the ONS Consumer Price Index (CPI) to express them in sterling and in 2015 prices.

[9] In the absence of data on the regional populations of Facebook users, Bailey et al. (2018) suggest that scaling the SCI using total regional populations should bear similar results.

Associating imports in different product categories to the corresponding import-competing manufacturing industries is necessary for measuring regional exposure to import competition. In doing so, I use a crosswalk published by the World Integrated Trade Solution (WITS) service of the World Bank, which allows me to match the SITC Rev.3 product categories to 3-digit NACE Rev.1 industries.[10] Following Autor et al. (2013), I then measure regional exposure to Chinese import competition as follows:

$$ImpShock_r = \sum_d \frac{Employment_{rd(1991)}}{Employment_{r(1991)}} \times \frac{\Delta Imp_{d(1991,2007)(CN,UK)}}{Employment_{d(1991)}} \tag{1.3}$$

Here, r indexes the 143 harmonised ITL3 regions and d indexes 95 manufacturing industries at the 3-digit NACE Rev.1 level, and $ImpShock_r$ is expressed in pounds per worker. Exposure to import competition within a national industry will be high if the growth in competing Chinese imports between 1991 and 2007 was also high relative to the number of industry workers across all regions.[11] As such, regional exposure will be high if a large share of regional employment in 1991 was in highly exposed manufacturing industries. Consistent with evidence on the economic effects of the Chinese import shock (e.g., Autor et al., 2016; Dorn and Levell, 2021), this measure reflects the intuition that regions specialising in import-competing manufacturing industries were also the ones most likely to experience adverse labour market outcomes due to the growth of Chinese imports. Using 1991 as the base year in the import-shock exposure measure is in line with the onset of Chinese import growth in developed economies including the UK. Equally, using 2007 as the year up to which import growth is measured is preferred due to preceding the Great Recession and its confounding effects on trade. In the UK, Colantone and Stanig (2018a) show that the share of imports that came from China increased from just over 1 per cent to nearly 9 per cent over this period.

Insofar as a given region had a high number of residents that worked elsewhere in 1991, the regional import-shock exposure measure may misrepresent exposure among residents in that region. In addressing this concern, I follow Malgouyres (2017) by measuring place-of-residence—or ‘residential’—exposure by using data on commuting flows between regions.

[10] For a small number of cases (54 out of 3214 SITC Rev.3), there are 1-to-n matches between SITC Rev.3 and NACE Rev.1. In such cases, I follow Malgouyres (2017) and apportion the value of imports in each SITC product category into NACE industries based on the share of employment in each industry in 1991. I also discard any import data under the ‘special transactions’ SITC product category (9310) as these cannot be mapped to any given industry.

[11] National employment totals include England, Scotland, and Wales.

First, I calculate exposure in each Scottish harmonised LAD region h, $ImpShock_h$. I then calculate the residential measure for harmonised ITL3 regions in England and Wales as:

$$ResImpShock_r = \sum_{r^*} \frac{Commuters_{rr^*(1991)}}{Workers_{r(1991)}} \times ImpShock_{r^*} \tag{1.4}$$

Here, r indexes harmonised ITL3 regions, whereas r^* indexes harmonised ITL3 regions r' and Scottish harmonised LAD regions h. That is, for each harmonised ITL3 region r, the residential import-shock exposure measure is the weighted sum of the place-of-work measure in each harmonised ITL3 region and Scottish harmonised LAD r^* multiplied by the share of commuting flows from r to r^* relative to the resident labour force of r. Recall that harmonised ITL3 regions only cover England and Wales due to correspondence issues with Scottish harmonised LAD. Still, as import-shock exposure can be obtained for Scottish harmonised LAD, its consideration in the residential exposure of regions in England and Wales accounts for the potential exposure of commuters from these regions to Scotland.

Appendix 1.A.2 presents a detailed comparison of place-of-work and residential exposure to import competition, showing that for the vast majority of regions the two measures are very similar. As I employ the residential rather than the place-of-work measure in all empirical specifications, any subsequent references to import-shock exposure thus allude to the former unless explicitly stated.

1.4.5 Voting Outcomes

Regional voting outcomes are obtained from the UK Electoral Commission, which publishes the official referendum results alongside information on turnout and electoral size. Data on individual voting outcomes are drawn from Wave 9 of the British Electoral Study, which was held between the 24th of June and the 4th of July 2016. As the wave immediately follows the date of the UK EU membership referendum, individual voting outcomes represent self-reported vote choice.

1.5 Empirical Strategy

The main challenge in identifying spillover effects on regional voting outcomes from the exposure of socially connected regions to Chinese import competition is the well-known problem of 'correlated effects' (Manski, 1993). That is, estimates may be biased if there are unobserved similarities between social neighbours. Not least, similar regions may be more socially connected due to homophily between their populations (McPherson et al., 2001).

As shown by Dieye et al. (2014), given a randomised treatment that does not affect social ties, unbiased estimates of spillover effects can be obtained even in the presence of endogenous network processes. That is, if the treatment and its social lag are exogenous, their effects can be separated from correlated effects (Bramoullé et al., 2020). While spatial variation in exposure to Chinese import competition might be endogenous, various approaches may be employed in isolating an as-good-as random component of this variation. Then, the key condition to be satisfied is that social ties between social neighbours are unaffected by exposure to import competition. In this section, I describe the adopted empirical strategy and how this works toward satisfying the discussed conditions for identification.

1.5.1 Delineating Commuting Zones

If a given local labour market spans multiple regions, the latter are likely to be similarly affected by unobservables that correlate both with their exposure to import competition and their voting outcomes. These may range from common shocks affecting the productivity of the local workforce to similarities in culture and political representation. Insofar as such similarities are also shared with social neighbours, their exposure may too be endogenous. Also, given that social ties as measured by the SCI are observed after the Chinese import shock, the former may be shaped by mobility responses within local labour markets. I address these threats to the identification by focusing on variation within commuting zones, and only considering social neighbours that are located in different commuting zones. Specifically, I delineate 18 commuting zones using the approach of Tolbert and Sizer (1996) with harmonised ITL3 regions as building blocks. Geographic clusters derived in this manner are often considered as equivalent to local labour markets in the US literature on import competition (e.g. Autor et al., 2013, 2020).[12]

In UK official statistics, the most conceptually similar classification to commuting zones is that of Travel-to-Work Areas (TTWA) (Coombes and Bond, 2008). Though commuting zones are still preferred in this chapter for at least three reasons. First, there is poor correspondence between ITL3 regions and TTWA. While the clustering approach for TTWA could still be in principle applied on harmonised ITL3 regions, the Tolbert-Sizer approach remains preferable as it is more conservative with respect to the level of commuting flows that is tolerated between the delineated clusters.[13] Importantly, the Tolbert-Sizer approach also provides a more straightforward measure for evaluating commuting dissimilarity between region-pairs, which further aids with precluding substantial commuting flows between social neighbours.

[12] The European literature on the effects of import competition often equates local labour markets to NUTS3 regions (e.g., De Lyon and Pessoa, 2021; Dippel et al., 2021). Though NUTS does not preclude substantial interregional commuting flows as it is based on administrative geographies.

[13] For TTWAs, at least 70 per cent of those living in the area will also work there and 70 per cent of those working in the area will also live there. Commuting zones will, on average, group region-pairs where gross commuters are over 2 per cent of the workforce of the smaller region.

A difference between the approach used in this chapter and that of Tolbert and Sizer (1996) is that the latter only uses data for 1991 to measure the commuting dissimilarity between region pairs. In contrast, I take the mean commuting dissimilarity across three UK Census years: 1991, 2001, and 2011. Given that harmonised ITL3 regions are comparable to US counties—which form the building blocks used by Tolbert and Sizer (1996)—it may be argued that commuting zones delineated solely on the basis of 1991 data are sufficiently over time, as is usually the case in the US literature.[14] However, as social ties are observed in 2020, taking the mean dissimilarity across years works towards clustering any region-pairs which may have developed stronger social ties via long-term commuting adjustments to past import-shock exposure. Formally, for every harmonised ITL3 region-pair rr' and Census year t, I calculate the following proportional flow measure:

$$P_{rr't} = P_{r'rt} = \frac{Commuters_{rr't} + Commuters_{r'rt}}{min(Workers_{rt}, Workers_{r't})} \tag{1.5}$$

The measure thus represents the ratio of gross commuting flows between regions r and r' in Census year t over the minimum of the two regional populations of resident workers in that Census year. Subsequently, I obtain $P_{rr'}$, which is the mean commuting similarity between two regions across the three Census years. The commuting dissimilarity between two regions is then defined as:

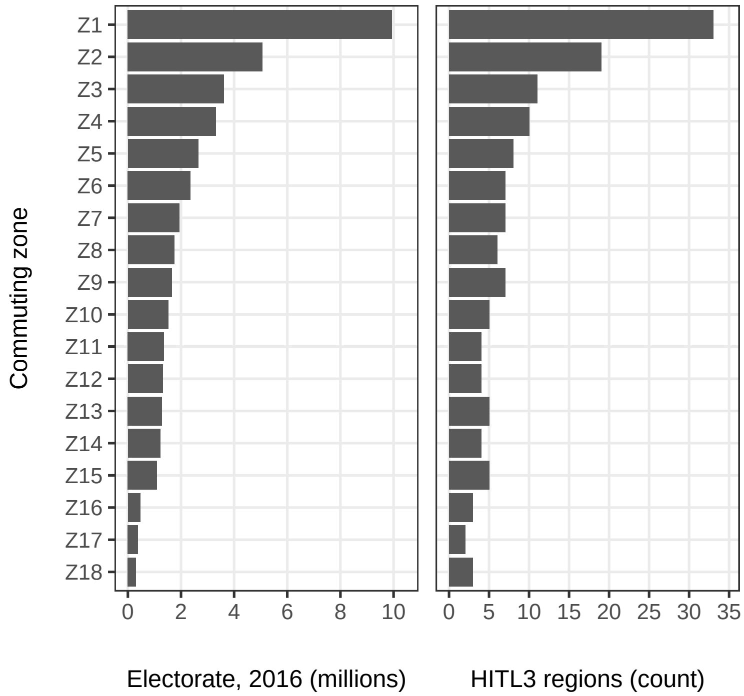
$$D_{rr'} = D_{r'r} = 1 - P_{rr'} \tag{1.6}$$

Just like Tolbert and Sizer (1996), I use average agglomerative hierarchical clustering to produce a dendrogram of harmonised ITL3 region clusters based on $D_{rr'}$ and select those above the average between-cluster dissimilarity cut-off of 0.98. Figure 1.1 lists the resulting 18 commuting zones by electorate size and the number of regions spanned, and Figure 1.2 maps them.[15] Notably, only three commuting zones overlap with higher levels of the ITL classification: zone Z7 with the North East of England ITL1 region, zone Z17 with the Cumbria ITL2 region, and zone Z13 with the Kent ITL2 region. Overall, it is shown that local labour markets cross administrative geographies and the England-Wales border.

[14]The population of a harmonised ITL3 region is, on average, almost twice that of a US county (417,620 against 207,544 in 2020) but is much less heterogeneous (standard deviations of 205,300 and 1,269,953).

[15]The clustering algorithm first produces 19 zones with the Isle of Wight forming its own zone. To allow for within-zone variation, the island is grouped with the most proximate zone (Z8).

Figure 1.1: Commuting Zones, by Electorate and Number of Regions

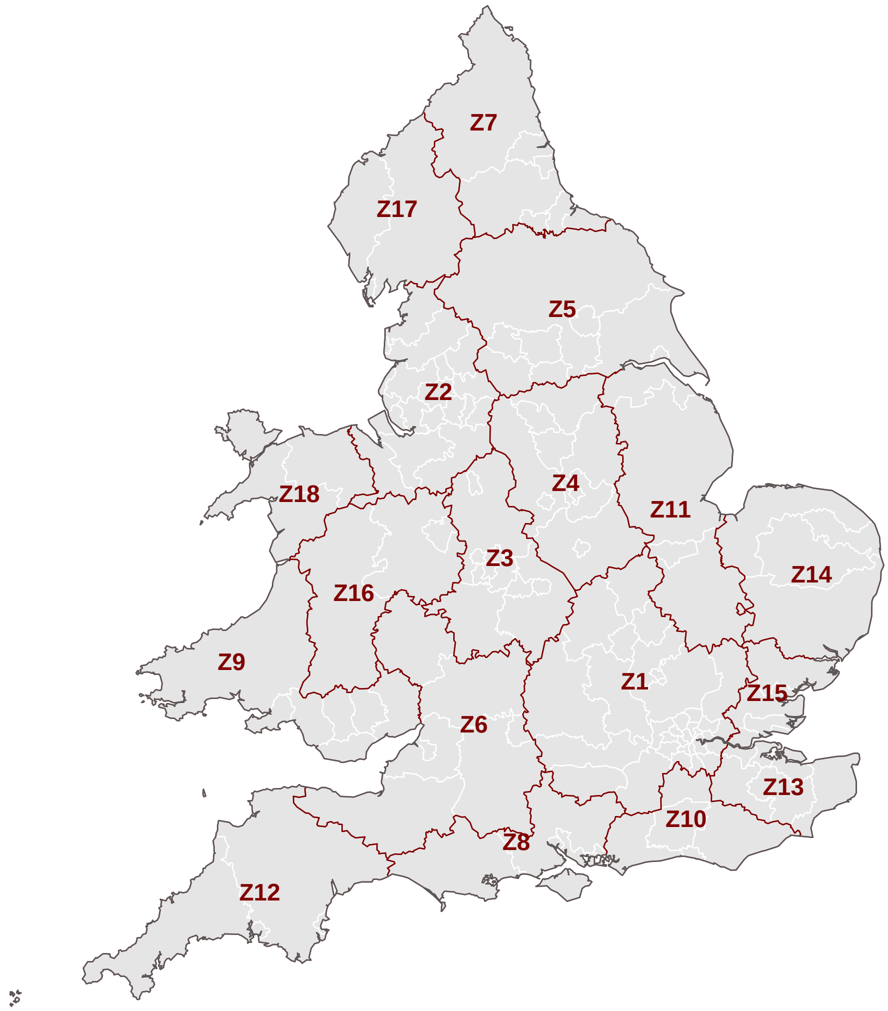


1.5.2 Creating Social Neighbour Matrices

I create a modified index, $\mathrm{SCI}_{(\text{-c})}$, which is equal to the SCI except where region-pairs belong to the same commuting zone or have a commuting dissimilarity $D_{rr'}$ of less than the between-zone average 0.98, in which case it is equal to zero. That is, I discard connections between regions in the same commuting zone and regions that share substantial pairwise commuting flows. Figure 1.3 maps the five nearest neighbours as defined by the SCI and $\mathrm{SCI}_{(\text{-c})}$ for the regions with the maximum, median, and minimum exposure to Chinese import competition.

For all three regions, the five nearest neighbours defined by the SCI are in very close geographic proximity. In contrast, the neighbours defined by $\mathrm{SCI}_{(\text{-c})}$ are more distant and in some cases lie well beyond the boundaries of the region's commuting zone. Appendix 1.A.3 offers a detailed discussion of the differences between the sets of social neighbours defined by each measure and their relationship to geographic distance. Notably, it is shown that a given region's respective first, fifth, and tenth social neighbour as defined by $\mathrm{SCI}_{(\text{-c})}$ is, on average, 74, 102, and 123 kilometres away. It is also shown that to discard region-pairs with commuting ties is to discard the most socially connected region-pairs. Yet as the remaining variation in the density of social ties is unrelated to the effects of the Chinese import shock on commuting patterns, it is better suited for the identification of spillover effects from social neighbours' exposure to the shock.

Figure 1.2: Commuting Zones and Harmonised ITL3 Regions



Notes: Commuting zones are drawn in red, and harmonised ITL3 regions in white. The zones are based on the hierarchical clustering approach of Tolbert and Sizer (1996) using a mean commuting dissimilarity measure computed across three Census years (1991, 2001, 2011). Zone labels are numbered according to 2016 electorate size, in descending order.

I create social neighbour matrices $\mathbf{W}_s$ by populating the elements (r, r') with zero except where $\text{SCI}_{rr'(\text{-c})}$ is in the sth set of five highest values in row i, in which case they contain that value. That is, for each region-row, non-zero cells in $\mathbf{W}_1$ identify the five most social connected regions that are not in its commuting zone or have substantial pairwise commuting flows with it — what I hereby simply refer to as its five nearest social neighbours. By storing the SCI value associated with each identified social neighbour, the matrix also captures the relative strength of their connection to the focal region. Similarly, $\mathbf{W}_2$ captures these relations with respect to the sixth to tenth nearest social neighbours. Partitioning matrices in sets allows me to test whether spillovers decay with distance in social space.

Figure 1.3: Five Nearest Neighbours, by Region and Measure

Notes: The focal region is shaded in red and its neighbours in black. SCI refers to the Social Connectedness Index (Bailey et al., 2018, 2020) after discarding the connectedness of each region with itself. $SCI_{(-c)}$ refers to the latter measure discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows.

1.5.3 Social Neighbours' Exposure to Import Competition

Alongside measures of within-region exposure to Chinese import competition I also calculate measures of social neighbours' exposure as follows:

$$\hat{w}_{s,r} \times ResImpShock = \sum_{k \in K_{s,r}} \frac{SCI_{rk} \times ResImpShock_k}{\sum_{k \in K_{s,r}} SCI_{rk}} \tag{1.7}$$

Here, $\hat{w}_{s,r}$ is a row vector from the row-normalised social neighbour matrix $\hat{\mathbf{W}}_s$ identifying the sth set of five nearest social neighbours of harmonised ITL3 region r and their relative weights in terms of the density of their social ties to r, $ResImpShock$ is the column vector of import-shock exposure within each harmonised ITL3 region, and k indexes social neighbours in the set $K_{s,r}$. That is, exposure in a set of five social neighbours is the SCI-weighted sum of their exposure over the sum of SCI values in the set. This measure can thus be interpreted as the expected exposure of the 1st degree peers in s of a resident in r.

Figure 1.4 maps regional exposure to Chinese import competition in England and Wales along with the five nearest social neighbours' exposure and the regional share of the Leave vote in the EU membership referendum. In terms of within-region exposure, the map broadly resembles those of previous UK studies (Ballard-Rosa et al., 2021; Colantone and Stanig, 2018a; Steiner and Harms, 2021). The most highly exposed regions are seen in South Wales, the Midlands, and the North of England, which is in line with the historical reliance of these regions on employment in import-competing manufacturing industries. In contrast, the least exposed regions are largely concentrated around London and the South East. However, different patterns emerge when looking at social neighbours' exposure.[16] For instance, Harrow and Hillingdon, North Yorkshire, and Herefordshire, are respectively in the first, third, and fourth deciles for within-region exposure, moving up to the sixth, eighth, and tenth deciles for social neighbours' exposure.

1.5.4 Baseline Specifications

For harmonised ITL3 regions r within commuting zones c, I estimate the following baseline regional specification using ordinary least squares (OLS) regression:

$$Leave_{rc} = \alpha_c + \beta ResImpShock_r + \gamma \hat{w}_{1,r} \times ResImpShock + \epsilon_{rc} \tag{1.8}$$

[16] Assuming linearity, the five nearest social neighbours' exposure explains 18 per cent of the variation in within-region exposure. For variation within commuting zones this reduces to 9 per cent. This suggests limited homophily in terms of exposure to Chinese import competition.

Here, $Leave_{rc}$ is the share of valid votes cast in region r in support of the Leave option in the 2016 UK EU membership referendum, α_c represents commuting zone dummies, $ResImpShock_r$ is the within-region exposure of r to Chinese import competition, and $\hat{w}_{1,r} \times ResImpShock$ is exposure in the first set of five nearest social neighbours of r. The coefficient of interest is therefore γ, which, following the standardisation of all independent variables to have zero mean and unit variance, is interpreted as the percentage point change in the regional share of the Leave vote for a one standard deviation increase in its five nearest social neighbours' exposure to Chinese import competition. For each coefficient, I report standard errors corrected for heteroscedasticity and autocorrelation within 31 lower commuting zones.[17] In an augmented specification, I also append exposure in the second set of five nearest social neighbours, $\hat{w}_{2,r} \times ResImpShock$, to (1.8).

For respondents to Wave 9 of the British Election Study i residing in harmonised ITL3 regions r within commuting zones c, I also estimate the following baseline individual-level specification using probit regression:

$$Leave_{irc} = \alpha_c + \eta ResImpShock_r + \theta \hat{w}_{1,r} \times ResImpShock + Z_i + \epsilon_{irc} \tag{1.9}$$

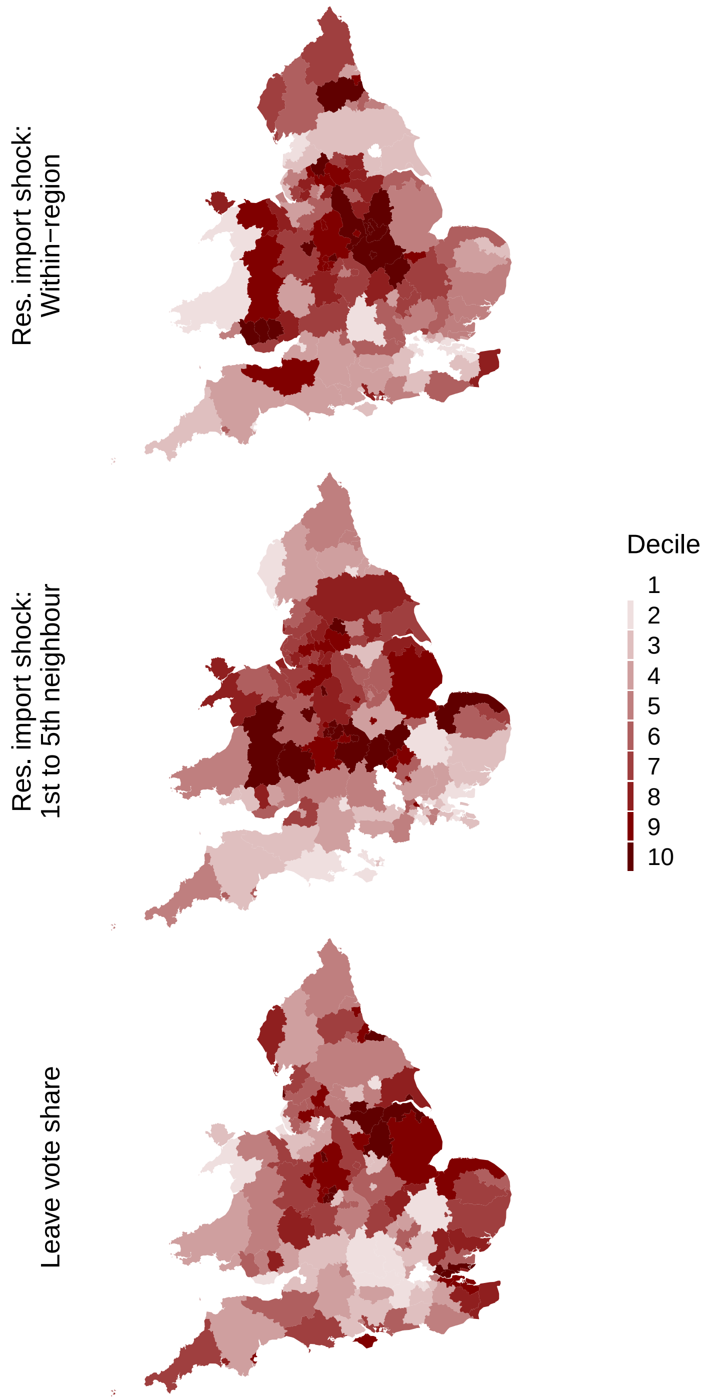
Here, $Leave_{irc}$ is a binary variable that is equal to one when an individual has voted in favour of leaving the EU, while $ResImpShock_r$ and $\hat{w}_{1,r} \times ResImpShock$ remain the same as in the baseline regional specification. Also included is the vector Z_i, containing controls for age and gender, and five education dummies.[18] I also calculate and report the average marginal effect θ' which can be interpreted as the average change in the probability of voting to leave the EU for a standard deviation increase in social neighbours' import-shock exposure. As with the regional specifications, I report heteroscedasticity-robust standard errors clustered at the level of lower commuting zones and estimate an augmented specification including import-shock exposure in the sixth to tenth social neighbour.

Across baseline specifications, the key assumption for the identification of the effects of interest is that the density of social ties between each pair of social neighbours is not affected by their respective exposure to import competition and that, conditional on any covariates, variation in social neighbours' exposure to import competition within commuting zones is as good as random.

[17] The 31 lower commuting zones are nested within the 18 commuting zones and are derived in the same way except for setting the between-cluster commuting dissimilarity cut-off to 0.95. That is, lower commuting zones will, on average, group region-pairs where gross commuters are more than 5 per cent, as opposed to 2 per cent, of the workforce of the smaller region.

[18] The five education dummies indicate whether the individual's highest qualification is at the postgraduate, undergraduate, A-level, GCSE A-C, or GCSE D-G level.

Figure 1.4: Import-Shock Exposure and Voting in Harmonised ITL3 Regions



Notes: The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set.

1.5.5 Instrumental Variable Estimation

Focusing only on variation within commuting zones in all baseline specifications accounts for potentially endogenous social, economic, and political factors varying at the level of local labour markets. Still, omitted variable bias from regional heterogeneity within local labour markets remains a concern. I address this using an instrumental variable approach akin to that of Autor et al. (2013). Namely, I attempt to isolate the variation in the import-shock exposure measures that relates to supply conditions in China by instrumenting Chinese imports into the UK with Chinese imports into the US. This approach rests on the idea that supply-side factors in China, such as productivity growth, will have similarly affected its capacity to export across developed trading partners. As such, insofar as there are no common demand shocks in a given pair of partners, import growth in the one can be used to retrieve a supply-driven component of import growth in the other that is plausibly orthogonal to regional characteristics.

I create instruments for within-region exposure, $ResImpShock_{r(\text{US})}$, and social neighbours' exposure, $\hat{w}_{s,r} \times ResImpShock_{(\text{US})}$, by first substituting growth in Chinese imports into the UK competing within industry d, $\Delta Imp_{d(1991,2007)(\text{CN,UK})}$, with growth in Chinese imports into the US competing within the same industry, $\Delta Imp_{d(1991,2007)(\text{CN,US})}$, in (1.5) and then recalculating the respective measures in (1.6) and (1.7).[19] Using these instruments, I re-estimate the baseline regional and individual-level specifications using two-stage least squares (2SLS) and IV probit regressions. The identifying assumption for these estimates is that, conditional on any covariates, variation in social neighbours' supply-driven exposure to import competition within commuting zones is as good as random, and that the density of social ties between each pair of social neighbours is not affect by their respective supply-driven exposure.

1.5.6 Relationship between Import Competition and Social Ties

While the social ties considered are plausibly unrelated to past mobility within local labour markets, in principle, mobility between local labour markets remains a threat to identification to the extent that this both responds to import competition and affects interregional social ties. As discussed in the conceptual framework, this is an unlikely prospect; indeed, Autor et al. (2016) argue that mobility frictions between local labour markets are, to a large extent, what makes spatially uneven exposure to import competition a persistent economic problem. Similarly, using individual-level administrative data on UK workers, De Lyon and Pessoa (2021) do not find evidence that workers exposed to Chinese import competition are more likely to switch the NUTS3 region in which they work.

[19] Colantone and Stanig (2018c) provide some evidence consistent with the expectation that Chinese import growth in the US is orthogonal to demand and technology shocks affecting voting behaviour in European economies including the UK.

Import competition may still affect social ties via other channels, such as social sorting. This could be the case, for instance, if individuals whose preferences have responded to exposure to the import shock in their local area have sought to establish ties with distant others holding similar preferences over the Internet. I thus perform a series of checks to gauge whether such or other unidentified processes are likely to be affecting the estimated spillover effects. The relevant specifications and analysis are presented in Appendix 1.A.4. Overall, other than the possibility of limited downward bias arising from a very weak observed relationship between import-shock exposure and selection in social neighbour sets, I fail to find evidence of a clear threat to identification.

1.6 Results

Table 1.1 presents the regional results. As all specifications include commuting zone dummies, estimates are interpretable as comparisons between different regions in the same commuting zone. The first column refers to the most parsimonious OLS specification including exposure to Chinese import competition within the region and its five nearest social neighbours. The third column appends exposure in the second set of five nearest social neighbours. The second and fourth columns present the respective 2SLS estimates based on the instrumental variable approach discussed in the previous section. For 2SLS estimates, I also report the Sanderson and Windmeijer (2016) conditional F-statistic for weak instruments. In all cases, the F-statistics remain above conventional levels indicating that each instrument provides an independent source of exogenous variation.

In the first column of Table 1.1 it is shown that exposure to Chinese import competition within a region and exposure in its five nearest social neighbours are associated with similar positive responses in voting in favour of leaving the EU. Comparing regions within the same commuting zone, a one standard deviation increase in either measure is associated with a roughly 2.8 percentage point increase in the regional share of the Leave vote.[20] The 2SLS estimates in the second column suggest that there is no clear omitted variable bias as the coefficients for both within-region and social neighbours' exposure remain close to OLS estimates. Though spillovers seem to decay beyond the first few social neighbours: as shown in the third column, the coefficient of exposure in the second set of five social neighbours is less than half than that of the first set and is statistically indistinguishable from zero. The respective 2SLS estimate on the fourth column is almost four times smaller, further suggesting that trade-related labour market conditions in the sixth to tenth social neighbour did not have substantial spillover effects on referendum results.

[20] As shown on Appendix Table 1.A.6, similar estimates are obtained from weighted-least-squares (WLS) regressions weighted by the regional number of votes cast in the referendum.

Table 1.1: Regional-Level Results

	Leave vote share			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	2.809^a	2.992^a	2.543^a	2.949^a
	(0.807)	(0.857)	(0.866)	(0.946)
1st to 5th neighbour	2.867^b	3.031^b	2.633^b	2.989^b
	(1.142)	(1.298)	(1.241)	(1.428)
6th to 10th neighbour			1.256	0.251
			(0.93)	(1.115)
Estimation method	OLS	2SLS	OLS	2SLS
Commuting zone dummies	✓	✓	✓	✓
Observations	143	143	143	143
Within-R^2	0.18	0.18	0.19	0.18
Sanderson-Windmeijer F:				
Within-region		346.9		376.1
1st to 5th neighbour		420.3		568.7
6th to 10th neighbour				552.3

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. $^a p < 0.01$, $^b p < 0.05$, $^c p < 0.1$

As proposed by Mummolo and Peterson (2018), I also report the mean range within commuting zones and the standard deviation of the within-distribution for the standardised import-shock exposure measures. Given that only variation within commuting zones is considered, these alternative theoretical shifts further aid with gauging the qualitative significance of the results. I express these as shares of the standard deviation of the full distribution of the respective import-shock exposure measure.

The mean range within commuting zones is 1.97 and 1.52 for standardised exposure to Chinese import competition within a region and its five nearest social neighbours respectively, while the standard deviations of the corresponding within-distributions are respectively 0.78 and 0.63. Consequently, with reference to the estimates in the first column of Table 1.1, the Leave vote share of the most exposed region in a given commuting zone will, on average, be larger than that of the least exposed region by $1.97 \times 2.81 = 5.57$ percentage points.

Similarly, the Leave vote share of the region with the most exposed five nearest social neighbours within a commuting zone will, on average, be larger than that of the region with the least exposed neighbours by $1.52 \times 2.867 = 4.36$ percentage points. When one considers the 'typical' differences between two regions from the same commuting zone, the Leave vote share of the region with higher exposure to import competition is expected to be larger by $0.78 \times 2.81 = 2.19$ percentage points, while the Leave vote share of the region with more exposed social neighbours is expected to be larger by $0.63 \times 2.867 = 1.8$ percentage points.

Consequently, irrespective of the counterfactual shift considered, the effects of social neighbours' exposure on voting outcomes are comparable to that of within-region exposure. Considering that the referendum result was determined by a 4 percentage point margin, spillover effects are also clearly of substantive significance.

Are the estimated spillover effects on regional voting outcomes attributable to economic spillovers between social neighbours? In the first three columns of Table 1.2, I successively augment the baseline regional specification from the first column of Table 1.1 with a series of standardised regional economic controls. In line with the discussed conceptual framework, these controls attempt to absorb effects on regional voting outcomes arising from indirect effects on linked industries, aggregate demand effects, and reallocation and mobility effects.

In the first column, I control for the start-of-period employment share in manufacturing using data from the 1991 UK Census. According to the ONS, manufacturing goods accounted for nearly 70 per cent of intermediate consumption in the UK manufacturing sector around that period (1995). This control thus accounts for the fact that indirect effects on linked industries are likely to be largely contained within the sector. In the second column, I also append the change in regional gross value added (GVA) between 1997 and 2016 relative to the median using ONS data. This control, which is also considered by Colantone and Stanig (2018a), captures unequal growth in income generated across regional industries and thus proxies for aggregate demand effects. Lastly, in the third column, I also include the net-in migration rate from social neighbours between 2002 and 2016, which proxies for reallocation and mobility effects.

The first three columns of Table 1.2 show that while the inclusion of regional economic controls largely blocks the political effects of within-region exposure to import competition, spillover effects from the exposure of the five nearest social neighbours retain significance and nearly two-thirds of their size relative to the baseline specification. This suggests that the estimated spillover effects are unlikely to be primarily driven by economic spillovers between social neighbours. In the fourth column, I also control for demographic characteristics of the regional population, finding that the baseline estimate for spillover effects remains effectively unchanged.

What kinds of social spillovers might lie behind the estimated spillover effects on regional voting outcomes? Answering this question with any degree of confidence is prohibitive given the absence on data on the nature of information flows between socially connected regions. However, a broad distinction might be made between information transmitted via geographic processes, such as travel, and via processes that are not constrained by geographic distance, such as telecommunication.

Table 1.2: Regional-Level Robustness

	Leave vote share			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	1.086	0.807	1.083	1.866^{a}
	(1.432)	(1.48)	(1.438)	(0.677)
1st to 5th neighbour	1.995^{b}	1.810^{b}	2.103^{b}	2.639^{a}
	(0.908)	(0.782)	(0.955)	(0.903)
Manufacturing emp. share, 1991	2.766	2.247	2.02	
	(1.936)	(1.785)	(1.765)	
Δ Relative income, 1997-2015		-2.505^{b}	-2.16^{a}	
		(0.917)	(0.726)	
Net in-migration rate, 2002-2015			-1.884^{b}	
			(0.849)	
Pop. share aged over 65, 1991				-0.003
				(0.759)
Pop. share foreign-born, 1991				-6.21^{a}
				(0.452)
Estimation method	OLS	OLS	OLS	OLS
Commuting zone FE	✓	✓	✓	✓
Observations	143	143	143	143
Within-R^2	0.21	0.29	0.33	0.5

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. All independent variables are standardised to have zero mean and unit variance. The k^{th} neighbour is the k^{th} most socially connected region as measured by the SCI$_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. $^{a}p < 0.01$, ^{b} $p < 0.05$, ^{c} $p < 0.1$

In Appendix Table 1.A.7, I present estimates from alternative specifications looking at spillover effects on regional voting outcomes from the exposure of geographic neighbours to import competition. Like social neighbours, I define geographic neighbours as the closest regions located outside the focal region's commuting zone and that do not share substantial pairwise commuting flows with it. Across specifications, estimates are statistically indistinguishable from zero. The effects of spillovers from the five nearest geographic neighbours are far weaker than those from the five nearest social neighbours and of the opposite sign.

Interestingly, OLS estimates for spillovers from the second set of five nearest geographic neighbours are closer to those from the first set of five nearest social neighbours. In line with the discussion in Appendix 1.A.3, this is likely due to overlap between the two sets. In all, these results suggest that processes driving the estimated effects of social neighbours' exposure to import competition on regional voting outcomes are not constrained by geographic distance.

Table 1.3 presents estimates from individual-level specifications using data from Wave 9 of the British Election Study. The first and third columns present baseline probit estimates respectively excluding and including exposure to import competition in the second set of social neighbours, while the second and fourth columns present the corresponding IV probit estimates. To aid interpretation, I also report the average marginal effects for each coefficient in Appendix Table 1.A.8. The results reflect those of the regional analysis. Import-shock exposure in the tenth to sixth social neighbour of the region in which an individual resides does not seem to have an effect on vote choice. Conversely, comparing individuals within commuting zones, a one standard deviation increase in the five nearest social neighbours' exposure is associated with a 1.53 percentage point increase in the probability of having voted to leave the EU. The respective 2SLS estimate is higher at 2.13 per cent, suggesting potential downward bias in the probit estimates.

In Table 1.4, I explore whether spillover effects from the five nearest social neighbours' exposure to import competition on vote choice are restricted to particular categories of voters. In doing so, I supplement the probit model in the first column of 1.3 with dummies which are also interacted with social neighbours' exposure. Namely, I include dummies for individuals that report a moderate to high risk of poverty over the next 12 months, the unemployed, those in full-time education, those who followed political news on the Internet for over 1 hour in the previous 7 days, and those with both parents born in the UK.

Consistent with the regional robustness checks, spillover effects do not appear to be restricted to those facing economic difficulty. The first two columns of Table 1.4 show that effects on support for Brexit among voters who are at low risk of poverty and voters in employment remain close to the baseline, while the respective interaction terms are not significant.

Similarly, I find no evidence that spillover effects are restricted to those that follow political news online, suggesting that these are not necessarily mediated by an online channel. Interestingly, while within-region exposure to import competition appears to be associated with support for leaving the EU among those with foreign-born parents, exposure in socially connected regions does not seem to affect vote choice in this category. One potential explanation for this finding is that recent international migrants may have more geographically concentrated social networks in destination countries (e.g. Schelling, 1971).

Table 1.3: Individual-Level Results

	Voted Leave			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	0.0506^a	0.0414^b	0.0549^a	0.0476^b
	(0.0173)	(0.0168)	(0.0195)	(0.0201)
1^{st} to 5^{th} neighbour	0.042^b	0.0583^a	0.0445^b	0.0609^a
	(0.0172)	(0.0165)	(0.0181)	(0.0178)
6^{th} to 10^{th} neighbour			-0.0167	-0.0265
			(0.0177)	(0.0205)
Estimation method	Probit	IV Probit	Probit	IV Probit
Commuting zone FE	✓	✓	✓	✓
Demographic controls	✓	✓	✓	✓
Observations	22,231	22,231	22,231	22,231
Regions	143	143	143	143
Sanderson-Windmeijer F:				
Within-region		352		355.3
1^{st} to 5^{th} neighbour		418.1		591.8
6^{th} to 10^{th} neighbour				603.5

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. Demographic controls include age, gender, and five education dummies. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. $^a p < 0.01$, $^b p < 0.05$, $^c p < 0.1$

Table 1.4: Individual-Level Results with Interactions

	Voted Leave				
	(1)	(2)	(3)	(4)	(5)
Import-shock exposure:					
Within-region	0.0466^a	0.0503^a	0.0496^a	0.0444^b	0.046^a
	(0.0168)	(0.0176)	(0.017)	(0.0173)	(0.0175)
1st to 5th neighbour	0.0432^b	0.0408^b	0.047^a	0.0393^b	0.0054
	(0.0189)	(0.0172)	(0.0182)	(0.0189)	(0.044)
(× At risk of poverty)	-0.0044				
	(0.0274)				
(× Unemployed)		0.0186			
		(0.0678)			
(× Student)			-0.0747		
			(0.107)		
(× Follows news online)				-0.009	
				(0.0201)	
(× UK-born parents)					0.0289
					(0.0431)
At risk of poverty	-0.0218				
	(0.0267)				
Unemployed		0.248^a			
		(0.0607)			
Student			-0.635^a		
			(0.0908)		
Follows news online				-0.008	
				(0.023)	
UK-born parents					0.185^a
					(0.0602)
Estimation method	Probit	Probit	Probit	Probit	Probit
Commuting zone FE	✓	✓	✓	✓	✓
Demographic controls	✓	✓	✓	✓	✓
Observations	20,852	22,231	22,231	21,605	16,658
Regions	143	143	143	143	143

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. Demographic controls include age, gender, and five education dummies. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. $^a p < 0.01$, $^b p < 0.05$, $^c p < 0.1$

1.7 Conclusion

In this chapter, I have shown that local trade-related economic shocks within regions in England and Wales are likely to have had spillovers that boosted support for Brexit in socially connected places lying well beyond their boundaries. The estimated effects on voting behaviour are comparable to those of local exposure and close to the margin that ultimately determined the outcome of the 2016 UK EU membership referendum. Importantly, the results suggest that these spillovers are more likely to be driven by information flows over social networks, rather than industry linkages, spending, and mobility between regions. This evidence offers a new perspective on the role of import competition and 'left-behind places' in voting to the leave EU in the UK. Whereas previous studies have established that the vote benefited from economic and cultural insecurity responding to local economic decline, this chapter suggests that these concerns are likely to have spread more widely and may have even affected the preferences of remote individuals facing relatively little economic difficulty. The results point to a promising avenue of research on the role of social spillovers on voting outcomes from spatially uneven exposure to economic shocks in other national and electoral contexts.

1.A Appendix

1.A.1 Delineating Harmonised Regions

The SCI by Bailey et al. (2018, 2020), which is the employed measure of interregional social ties, is reported for NUTS3 regions as of the 2016 version of the classification. Mobility, employment, and voting data are available for local authority districts (LAD), with the exception of commuting flows from the 2011 UK Census, which are in some instances reported for Census Merged LAD (CMLAD). I address the resulting correspondence issues by drawing on the Office for National Statistics (ONS) Code History Database and the Eurostat NUTS correspondence tables to produce crosswalks to harmonised spatial units for which each variable can be uniformly observed over all periods in the dataset. The latter span from 1991, which is the base year of the import-shock exposure measures, to 2020, when the SCI is measured. The earliest LAD version in the data is 2009.

I first create a harmonised LAD classification, which is comprised of 364 LAD as of 2020 and two CMLAD as of 2011, covering England, Scotland, and Wales. As discussed in the main text, Northern Ireland is not considered due to data limitations, while Scotland is excluded from the main empirical specifications due to poor correspondence between Scottish LAD and NUTS3. Though its inclusion in the harmonised LAD classification aids with correcting regional import-shock exposure in England and Wales for measurement error. As shown on Table 1.A.1, the two CMLAD appearing in the dataset are aggregates of two respective pairs of LAD, meaning that any variable that is observed for the latter can be easily aggregated to the former. Nevertheless, as the dataset period spans multiple years, boundary changes cause 1-to-*n* relations between earlier LAD versions in the dataset and the 2020 version.

Table 1.A.1: LAD20 to CMLAD11 Correspondence

LAD20 Code	LAD20 Name	CMLAD11 Code	CMLAD11 Name
E06000052	Cornwall	E41000052	Cornwall, Isles of Scilly
E06000053	Isles of Scilly	E41000052	Cornwall, Isles of Scilly
E09000001	City of London	E41000324	Westminster, City of London
E09000033	Westminster	E41000324	Westminster, City of London

Notes: Data is drawn from the ONS Code History Database. Only includes CMLAD11 that appear in the dataset.

Table 1.A.2 presents the ten LAD that are involved in boundary changes as well as indexing the corresponding statutory instruments associated with each change. Statutory instrument documentation is accessible on Legislation.gov.uk.

Table 1.A.2: LAD Boundary Changes, 2009 to 2020

LAD Code	LAD Name	LAD Code (previous)	LAD Name (previous)	Statutory Instrument
S12000045	E. Dunbartonshire	S12000009	E. Dunbartonshire	353/2010
S12000045	E. Dunbartonshire	S12000043	Glasgow City	353/2010
S12000046	Glasgow City	S12000009	E. Dunbartonshire	353/2010
S12000046	Glasgow City	S12000043	Glasgow City	353/2010
S12000047	Fife	S12000015	Fife	430/2017
S12000047	Fife	S12000024	Perth & Kinross	430/2017
S12000048	Perth & Kinross	S12000015	Fife	430/2017
S12000048	Perth & Kinross	S12000024	Perth & Kinross	430/2017
E08000037	Gateshead	E08000020	Gateshead	595/2013
E06000057	Northumberland	E08000020	Gateshead	595/2013
E07000242	E. Hertfordshire	E07000097	E. Hertfordshire	596/2013
E07000243	Stevenage	E07000097	E. Hertfordshire	596/2013
W06000024	Merthyr Tydfil	W06000007	Powys	889/2009
W06000023	Powys	W06000007	Powys	889/2009

Notes: All codes on the first column except S12000046 were live in December 2020. Data is drawn from the ONS Code History Database. Excludes data for Northern Ireland.

As detailed in the statutory instrument documentation, the boundary changes only involve a few hundred square metres of land at a time, which largely cover non-residential areas. This is also reflected in the fact that there are no new LAD names arising from the changes. I thus create a crosswalk that ignores these boundary changes by only matching each of the ten successor LAD codes with the predecessor code corresponding to the same LAD name. This lets me match units from all LAD versions with a harmonised equivalent or parent.

I next create a harmonised ITL3 classification covering England and Wales, which corresponds with both harmonised LAD and NUTS3 as of 2016. As shown on Table 1.A.3, there are only two instances of mismatching between ITL3 and the latter two classifications: one is due to a NUTS3 region boundary change between the 2016 and 2021 versions, and the other is due to a 1-to-n match between a harmonised LAD and two ITL3 regions. Both instances are resolved by merging the four ITL3 regions into two aggregates. The resulting classification is thus comprised of 141 ITL3 regions and two ITL3 region aggregates.

Table 1.A.3: ITL321 to Harmonised LAD/NUTS316 Correspondence

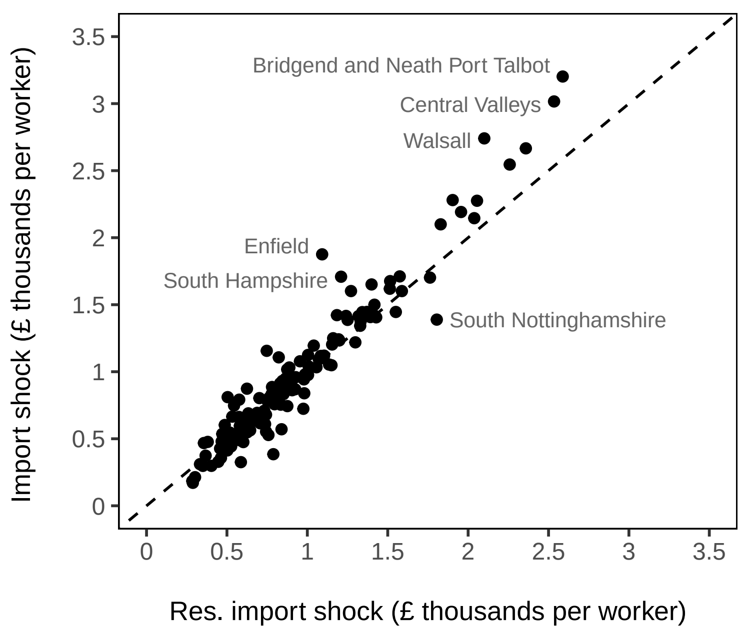
ITL321 Code	ITL231 Name	NUTS316/HLAD Code	NUTS316/HLAD Name
TLK24	Bournemouth, Christchurch & Poole	UKK22	Dorset CC
TLK25	Dorset	UKK22	Dorset CC
TLI31	Camden, City of London	E41000324	Westminster, City of London
TLI32	Westminster	E41000324	Westminster, City of London

Notes: Data is drawn from the ONS Code History Database and the Eurostat NUTS history tables. Excludes data for Scotland and Northern Ireland.

1.A.2 Comparing Regional Import-Shock Exposure Measures

Figure 1.A.1 compares the place-of-work and residential import-shock exposure measures. It is shown that the two measures are very similar for the vast majority of harmonised ITL3 regions. Median residential exposure is lower than its place-of-work counterpart by a modest 1.55 percentage points. Some of the more pronounced differences between the two measures are driven by the smoothing effect of the residential adjustment on extreme values of place-of-work exposure. For instance, Bridgend and Neath Port Talbot—the most affected region by both measures—is shown to move closer to the exposure of other regions as all residents commuting outside the region in 1991 were, as a matter of course, employed in less exposed regions as per the place-of-work measure. There are also cases where switching from place-of-work to residential exposure is associated with a substantial shift in the respective region's position in the distribution. For instance, Enfield moves from being the 11th most exposed region to the 43rd place, betraying its substantial commuting outflows to other regions in Greater London.

Figure 1.A.1: Comparison of Import-Shock Measures



Notes: Dots represent harmonised ITL3 regions. The two measures are equal for regions on the dashed line. Pounds are in 2015 prices.

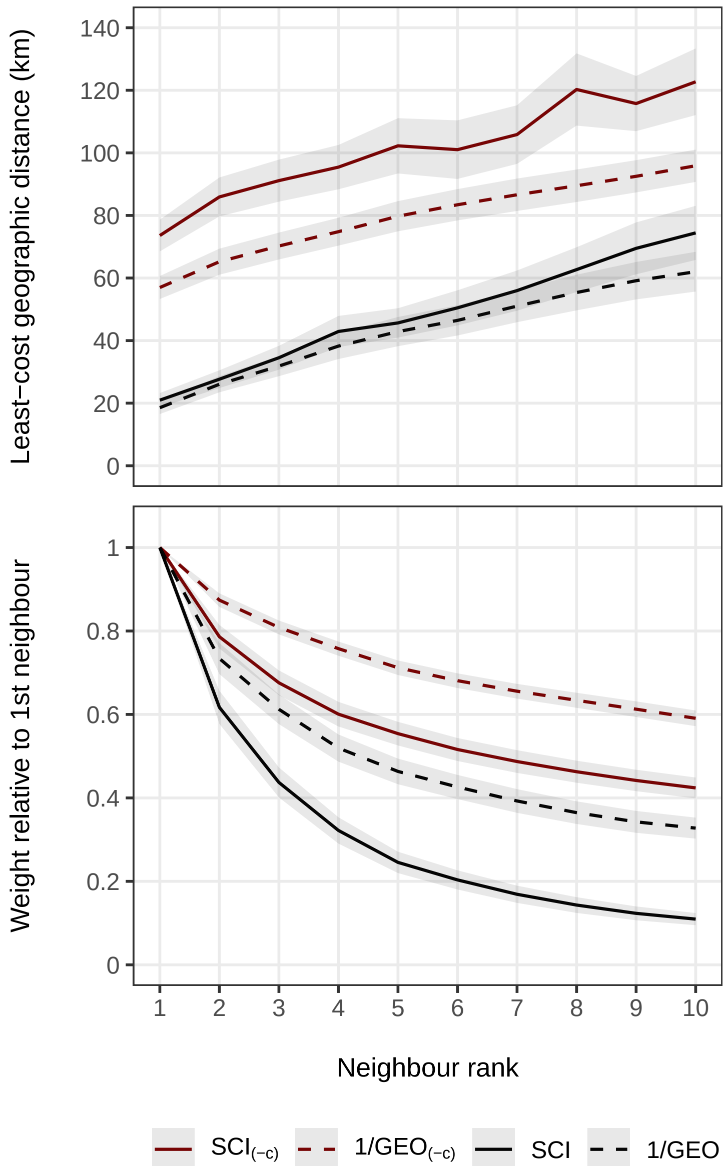
1.A.3 Comparing Sets of Social and Geographic Neighbours

The top panel of Figure 1.A.2 plots the mean least-cost geographic distance of the k^{th} nearest neighbour from the focal region as defined by the SCI and $\text{SCI}_{(\text{-c})}$. I also plot the mean distance by neighbour rank as defined by inverse geographic distance before and after discarding region-pairs within the same commuting zone and substantial pairwise commuting flows. Geographic distance is computed between the population-weighted centroids of harmonised ITL3 regions, which are derived using 1991 Census enumeration district centroids and populations as building blocks. The least-cost path algorithm is run over a 650 × 550 geographic transition matrix covering England and Wales, allowing for queen moves among cells on land while prohibiting movement along water surfaces. In the underlying raster layer, land buffers are placed around the isles of Wight and Anglesey to connect them to the British mainland.

It is shown that for all measures, neighbour rank increases with geographic distance. A given region's ten nearest neighbours as defined by the SCI are, on average, about as far away as its respective ten nearest geographic neighbours. However, looking at the 'commuting-discounted' measures, social neighbours are, on average, further away than their geographic counterparts. The first, fifth, and tenth social neighbour is on average 74, 102, and 123 kilometres away from the focal region, while the respective geographic neighbours are 57, 80, and 96 kilometres away. This suggests that there is less overlap between social and geographic neighbours in the 'commuting-discounted' measures, when compared to the raw measures. Indeed, as shown on Figure 1.A.3, the SCI and the raw geographic distance measure identify the same first nearest neighbour in 65 per cent of cases, up from 37 and 27 per cent for the second and third neighbour. In contrast, the respective overlap between $\text{SCI}_{(\text{-c})}$ and the 'commuting-discounted' geographic distance measure is lower at 41, 17, and 10 per cent.

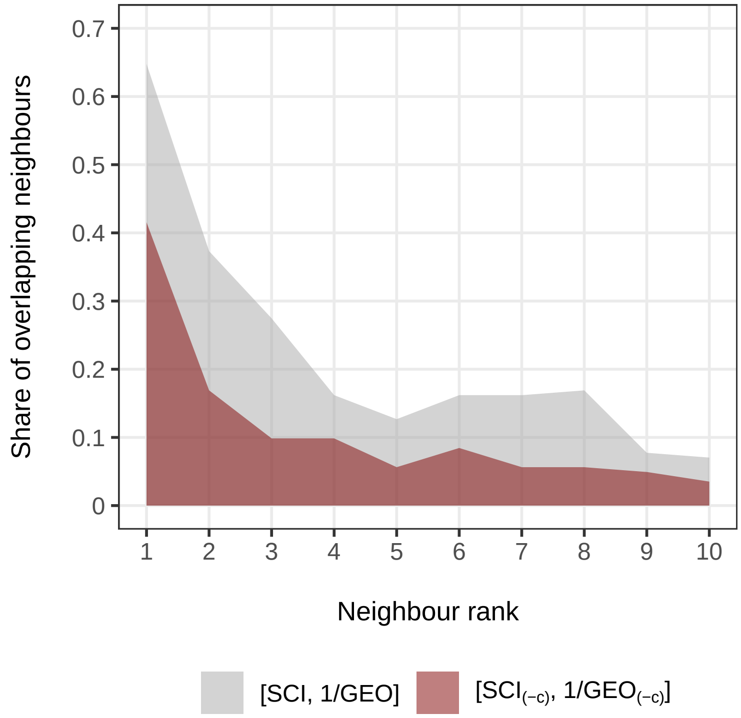
Switching from the SCI to $\text{SCI}_{(\text{-c})}$ does not only alter social neighbour ranks, but also the relative density of each neighbour's social ties with the focal region. As shown in the bottom panel of Figure 1.A.2, within each region, the mean SCI value of the k^{th} neighbour relative to that of the first neighbour as defined by the SCI declines very sharply. In contrast, the mean relative SCI value of the k^{th} neighbour as defined by the $\text{SCI}_{(\text{-c})}$ declines more smoothly, meaning there is less variation with respect to the density of their social ties with the focal region. A similar pattern is observed when switching from the raw to the 'commuting-discounted' geographic distance measure. However, in both cases, the relative density of social ties with the focal region is more heterogeneous, suggesting that proximity in social space decays more rapidly than in geographic space.

Figure 1.A.2: Mean Distance from and Weight of k^{th} Neighbour, by Measure



Notes: Shaded areas denote 95% confidence intervals. SCI refers to the Social Connectedness Index (Bailey et al., 2018, 2020) after discarding the connectedness of each region with itself. $SCI_{(-c)}$ refers to the latter measure after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Similarly, 1/GEO and $1/GEO_{(-c)}$ refer to respective measures based on inverse least-cost geographic distance.

Figure 1.A.3: Overlapping Neighbours, by Pairs of Measures



Notes: SCI refers to the Social Connectedness Index (Bailey et al., 2018, 2020) after discarding the connectedness of each region with itself. $SCI_{(-c)}$ refers to the latter measure discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Similarly, 1/GEO and $1/GEO_{(-c)}$ are respective measures based on inverse least-cost geographic distance.

1.A.4 Import Competition and Social Ties

For the ten nearest social neighbours k of harmonised regions ITL3 r, I estimate the following log-linear specification using both OLS and 2SLS regressions:

$$log(SCI_{kr}) = \alpha_r + \psi ResImpShock_k + \epsilon_{kr} \tag{1.10}$$

Here, $log(SCI_{kr})$ is the logged SCI between k and r, α_r are regional dummies, and $ResImpShock_k$ is within-region import-shock exposure in k. I standardise $ResImpShock_k$ to have zero mean and unit variance and report heteroscedasticity-robust standard errors corrected for autocorrelation within regions r and neighbours k. The coefficient ψ can thus be interpreted as the percentage point difference in social connectedness with region r that is expected between social neighbours whose import-shock exposure differs by one standard deviation. As such, a strong relationship would imply that the SCI-weighted measures of social neighbours' import-shock exposure as defined in (1.7) may suffer from measurement error, as the relative social weights of neighbours within sets could be codetermined with their exposure.

Another potential source of bias is that import-shock exposure may be associated with the probability of a region being in another's set of ten nearest social neighbours. Upward bias would arise in the estimation of spillover effects if more exposed regions were more likely to enter the set, while downward bias would arise in the opposite case. In gauging this prospect, I estimate the following linear probability specification using both OLS and 2SLS regressions:

$$I(r' \in \{K_{1,r}, K_{2,r}\})_{r'r} = \alpha_r + \xi ResImpShock_{r'} + \epsilon_{r'r} \tag{1.11}$$

Here, $r'r$ indexes all harmonised ITL3 region-pairs and $I(r' \in \{K_{1,r}, K_{2,r}\})_{r'r}$ is a binary variable that is equal to one when region r' is one of the ten nearest social neighbours of region r and zero otherwise. The coefficient ξ is thus interpreted as the relative probability of being one of the ten nearest social neighbours of region r that is expected when comparing regions whose import-shock exposure differs by one standard deviation. I again report heteroscedasticity-robust standard errors corrected for autocorrelation within r and r'.

Table 1.A.4 presents estimates from the discussed specifications. As shown in the first two columns, comparing regions in the set of ten nearest social neighbours of a focal region, a one standard deviation increase in exposure to import competition is associated with a 1.84 percentage point increase in social connectedness with the focal region. The effect is weak and statistically indistinguishable from zero, while the explained variation in social connectedness within sets stands at 0.2 percentage points.

As such, the estimated spillover effects on regional voting outcomes are unlikely to be driven by a relationship between the relative social weights of neighbours in the sets considered and their exposure to import competition. Notably, as seen in Appendix Table 1.A.5, the regional results discussed in the main text are robust to assigning equal weights to social neighbours in each set.

In the third and fourth column of Table 1.A.4 it is also shown that, across regions, a one standard deviation increase in exposure to import competition is associated with a decrease in the probability of being in another's set of ten nearest social neighbours by roughly 1.1 per cent and significant at the 90 per cent confidence level. While this may imply some downward bias in the estimation of spillover effects from the systematic selection of less exposed regions into social neighbour sets, the relationship is too weak to suggest a material threat to identification.

Table 1.A.4: Social Ties and Import-Shock Exposure

	log(SCI)		In $K_{1,r}$ or $K_{2,r}$	
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Neighbour	0.0184	0.0177	-0.0104^{c}	-0.0114^{c}
	(0.0195)	(0.0225)	(0.005)	(0.005)
Estimation method	OLS	2SLS	OLS	2SLS
Region FE	✓	✓	✓	✓
Neighbours per region	10	10	142	142
Observations	1,430	1,430	20,306	20,306

Notes: Robust standard errors in parentheses are clustered at the level of harmonised ITL3 regions and neighbours. The neighbour's import-shock exposure is standardised to have zero mean and unit variance. Log-linear model results in columns (1) and (2) only consider the ten nearest neighbours defined by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Linear probability model results in columns (3) and (4) consider all region-pairs. $K_{1,r}$ or $K_{2,r}$ respectively denote the first and second set of five nearest social neighbours of region r defined by the $\text{SCI}_{(\text{-c})}$.

1.A.5 Additional Tables

Table 1.A.5: Regional-Level Results, Equal Neighbour Weights

	Leave vote share			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	2.974^a	3.194^a	2.691^a	3.143^a
	(0.852)	(0.921)	(0.937)	(1.03)
1^{st} to 5^{th} neighbour	2.782^a	2.735^b	2.487^b	2.675^b
	(0.976)	(1.108)	(1.077)	(1.256)
6^{th} to 10^{th} neighbour			1.264	0.286
			(0.953)	(1.108)
Estimation method	OLS	2SLS	OLS	2SLS
Commuting zone FE	✓	✓	✓	✓
Observations	143	143	143	143
Within-R^2	0.18	0.19	0.19	0.19
Sanderson-Windmeijer F:				
Within-region		286.4		336
1^{st} to 5^{th} neighbour		834.8		978.4
6^{th} to 10^{th} neighbour				600.3

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the mean exposure in the set. $^a p < 0.01$, $^b p < 0.05$

Table 1.A.6: Regional-Level Results, Weighted by Votes Cast

	Leave vote share			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	3.048^a	3.197^a	2.74^a	3.187^a
	(0.877)	(0.957)	(0.975)	(1.043)
1^{st} to 5^{th} neighbour	2.544^c	2.77^c	2.373^c	2.766^c
	(1.279)	(1.409)	(1.360)	(1.487)
6^{th} to 10^{th} neighbour			1.148	0.043
			(1.02)	(1.195)
Estimation method	WLS	2SWLS	WLS	2SWLS
Commuting zone FE	✓	✓	✓	✓
Observations	143	143	143	143
Within-R^2	0.17	0.17	0.18	0.17
Sanderson-Windmeijer F:				
Within-region		336.4		343.4
1^{st} to 5^{th} neighbour		364.3		531.9
6^{th} to 10^{th} neighbour				543.4

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. All regressions are weighted by the regional number of votes cast in the referendum. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. ${}^a p < 0.01$, ${}^b p < 0.05$, ${}^c p < 0.1$

Table 1.A.7: Regional-Level Results, Geographic Neighbours

	Leave vote share			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	3.516^a	3.978^a	3.403^a	3.914^a
	(0.962)	(1.151)	(0.925)	(1.107)
1st to 5th neighbour	-0.0319	-0.802	-0.497	-1.054
	(1.149)	(1.224)	(0.916)	(1.096)
6th to 10th neighbour			1.749	0.971
			(1.429)	(1.462)
Estimation method	OLS	2SLS	OLS	2SLS
Commuting zone FE	✓	✓	✓	✓
Observations	143	143	143	143
Within-R^2	0.13	0.12	0.14	0.14
Sanderson-Windmeijer F:				
Within-region		325.5		348.1
1st to 5th neighbour		212		193.2
6th to 10th neighbour				365.7

Notes: Robust standard errors in parentheses are clustered at the level of 31 lower commuting zones. Import-shock exposure measures are standardised to have zero mean and unit variance. The k^{th} neighbour is the k^{th} most geographically proximate region as measured by $1/\text{GCD}_{(\text{-c})}$, which is inverse geographic distance discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the distance-weighted sum of their exposure over the sum of distances in the set. $^a p < 0.01$, $^b p < 0.05$, $^c p < 0.1$

Table 1.A.8: Individual-Level Results, Average Marginal Effects

	Voted Leave			
	(1)	(2)	(3)	(4)
Import-shock exposure:				
Within-region	0.0185[a]	0.0151[b]	0.02[a]	0.0174[b]
	(0.0063)	(0.0061)	(0.007)	(0.007)
1st to 5th neighbour	0.0153[b]	0.0213[a]	0.0162[b]	0.0222[a]
	(0.0063)	(0.006)	(0.0066)	(0.0065)
6th to 10th neighbour			-0.0061	-0.0097
			(0.0065)	(0.0075)
Estimation method	Probit	IV Probit	Probit	IV Probit
Commuting zone FE	✓	✓	✓	✓
Demographic controls	✓	✓	✓	✓
Observations	22,231	22,231	22,231	22,231
Regions	143	143	143	143

Notes: Robust delta-method standard errors in parentheses. Coefficients represent average marginal effects retrieved from the respective non-linear specifications. Import-shock exposure measures are standardised to have zero mean and unit variance. Demographic controls include age, gender, and five education dummies. The k^{th} neighbour is the k^{th} most socially connected region as measured by the $\text{SCI}_{(\text{-c})}$, which is equal to the SCI (Bailey et al., 2018, 2020) after discarding region-pairs in the same commuting zone or with substantial pairwise commuting flows. Import-shock exposure in a set of five neighbours is the SCI-weighted sum of their exposure over the SCI value sum in the set. $^a p < 0.01$,$^b\ p < 0.05$,$^c\ p < 0.1$

This page has been intentionally left blank.

CHAPTER TWO

Socio-Spatial Spillovers of All-Mail Voting: Evidence from North Carolina

All-mail voting has been introduced in several jurisdictions in the United States since the COVID-19 pandemic, with the resulting need for processing larger volumes of mail ballots posing challenges to election administration. Leveraging data on more than 4.5 million eligible voters in the state of North Carolina and on online social ties between zip code tabulation areas (ZCTA) across the country, I show that local rollouts of all-mail voting also have spillover effects on choice of voting method in distant areas. Using a difference-in-differences research design, I find that an increase in the ZCTA-level share of social ties in counties that switched to all-mail voting between the 2016 and 2020 presidential elections by one standard deviation (1.77 percentage points) corresponds to an increase in the probability of casting a mail ballot by roughly 3.8 percentage points. These socio-spatial spillovers are stronger for older voters, non-Republicans, and those residing in metropolitan counties. The findings suggest that considering the aggregate social ties of a local jurisdiction in other areas may aid election officials in planning for changes in the demand for mail ballots.

2.1 Introduction

All-mail voting is an electoral policy whereby all registered voters automatically receive a ballot by mail. While relatively rare prior to the COVID-19 pandemic, recent years have seen a marked uptake in its implementation in general elections in the United States. Recent work has primarily focused on the effects of all-mail voting on turnout (Amlani and Collitt, 2022; McGhee et al., 2022) and partisan gains (Barber and Holbein, 2020; Thompson et al., 2020; Yoder et al., 2021), while there has also been wider policy interest in how local jurisdictions may best prepare for administering all-mail elections.[1] However, little is known about the potential spillovers of all-mail voting on voting behaviour in other jurisdictions. Voter preferences are known to diffuse along social networks (e.g. Bond et al. 2012) while mail ballots are readily available on demand in many states. In the presence of socio-spatial spillovers from localised rollouts of all-mail voting, jurisdictions in such states could be faced with unexpected increases in demand for mail ballots.

In this chapter, I investigate socio-spatial spillovers on choice of voting method in North Carolina from local rollouts of all-mail voting in other states. For zip code tabulation areas (ZCTAs) in North Carolina, I first measure the share of aggregate social ties located in counties that rolled out the policy between the 2016 and 2020 presidential elections. In doing so, I rely on the Social Connectedness Index of Bailey et al. (2020), which is based on the near-universe of Facebook friendships, and the electoral policy dataset of Amlani and Collitt (2022). I then exploit detailed administrative records on the state electorate to examine the relationship between the ZCTA-level share of social ties switching to all-mail voting and the individual probability of voting by mail.

Using a difference-in-differences research design, I find that an increase in the share of social ties switching to all-mail voting by one standard deviation (1.77 percentage points) corresponds to an increase in the probability of voting by mail by roughly 3.8 percentage points. The estimate rises to 4.6 percentage points when restricting the panel to individuals that voted in both elections, in line with a small negative effect on turnout. In supplementary specifications, I show that the spillover effects are chiefly driven by policy changes in Western counties lying over 2,000 kilometres away. Across specifications, I account for local heterogeneity in exposure to the COVID-19 pandemic by including county-year fixed effects.

[1] In the context of US elections, the term 'jurisdiction' refers to the administrative unit that is responsible for election administration in a given geographic area. According to the National Conference of State Legislatures (2022a; 2022b), elections are most commonly administered at the county level while electoral reforms are typically implemented at the state level. In some cases, statewide legislation affords individual jurisdictions some discretion over electoral rules.

In line with the framework of Callaway et al. (2021), I interpret the baseline estimates as weighted averages of $ACRT(d|d)$: the average causal response associated with a marginal change in shares among units that experienced share d. This presupposes both common treated potential outcomes at marginally different treatment levels and common untreated potential outcomes. I provide suggestive evidence of the plausibility of the latter assumption by carrying out placebo tests using data on voting behaviour in the 2018 midterms. In evaluating the former assumption, I obtain new estimates by only comparing outcome evolution among individuals of the same political affiliation and age group who are also resident in the same county. This accounts for the possibility of selection bias along these salient dimensions. Reassuringly, the new estimates remain close to the baseline.

In a series of triple-difference specifications, I find evidence of heterogeneous spillover effects for different groups of voters. Democrats, unaffiliated voters, older voters, and metropolitan residents were all more likely to adopt the voting method in response to the share of social ties switching to all-mail voting. I relate these findings to previous work anticipating that age, partisanship, and urbanisation affect the convenience and social costs associated with the decision to vote by mail. In further checks, I fail to find evidence of heterogeneous spillover effects for voters experiencing polling place changes and voters in ZCTAs that are more socially connected to jurisdictions with higher exposure to the COVID-19 pandemic.

Finally, I present estimates from alternative binary difference-in-differences specifications comparing individuals in ZCTAs with below-mean shares of social ties switching to all-mail voting to successive groups of individuals in ZCTAs with different ranges of above-mean shares. Under a standard parallel trends assumption, which I support with further placebo tests, I interpret these estimates as lower bounds of $ATT(d|d)$: the average treatment effect among units that experienced shares in the range d. I find that socio-spatial spillovers are substantial at more extreme values. For regular voters in ZCTAs that saw shares between two and three standard deviations above the mean (7.34 and 9.11 per cent), the probability of voting by mail rose by almost 12 percentage points.

The empirical findings of this chapter suggest that local rollouts of all-mail voting trigger positive socio-spatial spillovers on the demand for mail ballots in other jurisdictions, which may have important implications for election administration. Other than contributing to the growing literature on the effects of all-mail voting on political outcomes, this evidence also relates to recent work highlighting the role of social networks in spatial spillovers of localised policies (Holtz et al., 2020; Zhao et al., 2021). That is, it also documents that social ties between distant areas can facilitate substantial spillovers over long distances: a neglected prospect in previous work.

2.2 Background

The 2020 US presidential election saw an unprecedented 43 per cent of voters cast a mail ballot, up from 21 per cent in the 2016 election (Fabina and Scherer, 2022). As a shock to both the demand for and the supply of absentee voting options,[2] the COVID-19 pandemic was central to this development. On the demand side, preferences are likely to have responded to the perceived risk of contracting the virus while voting at the polling place on the day of the election. While there was a marked partisan divide, with Republicans being almost half as likely to vote by mail as Democrats (Stewart, 2020) and the latter being more likely to do so in response to projections about the pandemic (Kousser et al., 2021; Lockhart et al., 2020), voters of both political leanings were more likely to vote by mail when running an increased health risk from infection or experiencing a higher incidence of deaths from COVID-19 in their local area (Atkeson et al., 2022; Scheller, 2021; Stewart, 2021). Studies of elections in countries with more restrictive absentee voting options also find that local COVID-19 cases and mortality had a negative effect on turnout (Constantino et al., 2021; Fernandez-Navia et al., 2021; Noury et al., 2021; Picchio and Santolini, 2022), further suggesting that local health outcomes shaped voters' risk perceptions in the pandemic.

On the supply side, for voters in many jurisdictions, it was easier to vote by mail in the 2020 presidential election relative to the 2016 election (Amlani and Collitt, 2022; Herrnson et al., 2022). According to Amlani and Collitt (2022), some 748 counties in ten states switched to no-excuse absentee voting—eliminating the requirement for an excuse to apply for a mail ballot—with 747 counties in twenty states automatically sending a mail ballot application to every registered voter. A further 168 counties in seven states—California, Hawaii, Nevada, New Jersey, Montana, Utah, and Vermont—and the District of Columbia switched to all-mail voting. Preliminary evidence suggests that, conditional on individual characteristics as well as personal exposure to COVID-19 or in one's social network, voters in jurisdictions that switched to all-mail voting in the 2020 US presidential election were much more likely to vote by mail when compared to votersin jurisdictions switching to other policies (Herrnson and Stewart, 2023).

[2]The term 'absentee voting' is used to refer to any alternative to voting in person on the day of the election, including voting by mail and voting in person before the day of the election. I use the terms 'vote by mail' and 'cast a mail ballot' interchangeably to refer to the act of receiving a ballot by mail—a 'mail ballot'—which is then cast at a post box, ballot drop box, or any other drop site. Also note that while 'all-mail voting' involves the dispatch of a mail ballot to every registered voter, this does not typically preclude voters from voting in person at a polling place (National Conference of State Legislatures, 2022b)

Relative to voters in jurisdictions that offered no-excuse absentee voting both before and after the onset of the pandemic, the probability of voting by mail was higher by 22 percentage points in jurisdictions switching to all-mail voting and lower by 22 percentage points in jurisdictions switching to no-excuse absentee voting, with mailing applications bearing no apparent effect. Interestingly, looking at the staggered rollout of all-mail voting in three out of the five states within which the policy was adopted in the years preceding the pandemic, Thompson et al. (2020) detect similar positive effects of on voting by mail. Indeed, Herrnson and Stewart (2023) conclude that while policy changes played a more limited role than pandemic-related concerns in voter turnout in the 2020 presidential election, they are likely to have been much more important in voters' choice of voting method. The former finding is also echoed in more systematic studies on the effects of various policy changes on turnout in the election (Amlani and Collitt, 2022; McGhee et al., 2022), which find modest positive effects on turnout from switches to all-mail voting and mixed results for other policies.[3]

The unprecedented demand for voting by mail weighed heavily on the administration of the 2020 presidential election, with election officials reportedly facing varied challenges in recruiting, allocating, and training poll workers, as well as in sending, receiving, storing, and validating ballots (Fortier and Stewart, 2021; U.S. Election Assistance Commission, 2020a). As the processing of mail ballots differs from that of ballots that are completed in person at a polling place, managing expected increases in voting by mail requires advance preparation well before election day, particularly in securing the necessary staff, equipment, and infrastructure (U.S. Election Assistance Commission, 2010). Failure to do so can contribute to processing delays and faults, leaving a weakened public trust in election integrity (Levy et al., 2021; Stewart, 2011)—an issue that featured heavily in the polarised debate around the 2020 presidential election. Consequently, anticipating the size of demand for voting by mail is an important objective for election officials in local jurisdictions tasked with administering elections.

While there is much variation in the extent and ways in which different jurisdictions estimate expected demand for different voting methods, election experts increasingly highlight the importance of carefully considering the impact of planned policy changes (Hodgson et al., 2021; McGhee et al., 2021; Orey, 2021; U.S. Election Assistance Commission, 2020b). Given growing evidence from the wider implementation of all-mail voting in the aftermath of the pandemic, more information is becoming available to jurisdictions that consider adopting this or similar policies.

[3] A related strand of work examines the partisan effects of all-mail voting in various states (e.g. Barber and Holbein 2020; Thompson et al. 2020; Yoder et al. 2021), finding no evidence that the policy advantages any particular party.

However, to this date, much less is known about how all-mail voting might affect choice of voting method outside of the jurisdictions in which the policy is being rolled out. The prospect of such spatial dependence became especially salient during the pandemic, with an emerging literature suggesting that localised policy changes, such as the introduction and lifting of mobility restrictions, had spillovers on behaviour well beyond the boundaries of the areas in which they were implemented (Holtz et al., 2020; Zhao et al., 2021). Crucially, the social ties between individuals in different, often distant regions emerged as an important channel through which the spillovers propagated. As such, to the extent that local rollouts of all-mail voting trigger similar socio-spatial spillovers, other jurisdictions may be faced with unexpected increases in demand for mail ballots, with potentially important implications for election administration. This is a pressing question in the context of elections following the pandemic, given that many jurisdictions that first rolled out all-mail voting in recent years are set to continue the automatic dispatch of mail ballots to voters in future elections.[4]

This chapter offers the first empirical examination of socio-spatial spillovers on voting by mail from the implementation of all-mail voting in other jurisdictions. In doing so, it adopts a difference-in-differences research design focusing on voting behaviour in the 2016 and 2020 presidential elections in the state of North Carolina. As all-mail voting was rolled out in several jurisdictions between the two elections, the study period offers rare variation in voters' social ties in geographically disparate areas adopting the policy. Further, North Carolina is a natural setting for the empirical objectives of this chapter for at least three reasons. First, the North Carolina State Board of Elections publishes individual-level administrative records on the state electorate, providing a rarely detailed source of behavioural data on choice of voting method over time. Second, the southeastern state offered no-excuse absentee voting in both the 2016 and 2020 presidential elections, which makes it a policy-relevant setting: no-excuse absentee voting jurisdictions are more likely to see unexpected increases in demand for voting by mail than both excuse-required jurisdictions, where the voting method is not widely available, and all-mail voting jurisdictions, where it is usually the norm.[5] Third, North Carolina is often regarded as a battleground state, meaning that the electorate as a whole does not exhibit a strong partisan bias: an attractive property given the prospect of heterogenous spillover effects across partisan groups.

[4]According to the National Conference of State Legislatures (2022b), California, Hawaii, Nevada, Utah, and Vermont have made recent changes in relevant legislation permanent.

[5]North Carolina is also one of just ten no-excuse absentee voting states that did not switch to the automatic dispatch of mail ballot applications in the 2020 presidential election (Amlani and Collitt, 2022). Instead, the state opted for light reforms such as limiting the requirement for two witness signatures to one and extending the grace period for the return of mail ballots.

2.3 Conceptual Framework

Rooted in the rational choice tradition, seminal theories on how voters decide whether to vote consider this to be a function of three broad types of costs. While choice of voting method has received little theoretical attention, a modest empirical literature suggests that this is governed by similar considerations. Given imperfect knowledge about available policies, Downs (1957) first anticipated that voting is influenced by information costs: the time and resources that would have to be expended in making an informed decision. Indeed, experimental evidence shows that voters who receive informational messages about absentee voting options are more likely to vote by mail (Herrnson et al., 2019), and that presenting voters with guidelines on how to vote by mail increases electoral participation (Shineman, 2018). Later work increasingly emphasised that voters are also faced with social costs that encourage conformity with the behaviour of their peer or reference group (e.g. Fiorina, 1976; Riker and Ordeshook, 1968). Consistent with this expectation, experimental studies show that voters who receive promotional messages from co-partisans are more likely to switch to absentee voting (Hassell, 2017), as are those who receive messages emphasising its popularity (Smith and Sylvester, 2013) and its relevance to voters with similar characteristics (Hanmer et al., 2015). Lastly, convenience costs are incurred by the mundane tasks that voters have to undergo in casting their ballot, such as registration as well as finding and getting to their polling place (Wolfinger and Rosenstone, 1980). For instance, it has been shown that voters who are exposed to polling place changes, which plausibly increase the effort required in finding one's polling place, are more likely to switch to voting by mail (e.g. Amos et al. 2017; Clinton et al. 2021; Brady and McNulty 2011). As such, the decision to vote by mail is likely to be driven by voters' awareness of the process of casting a mail ballot, deem this to be more convenient than other options, and feel that their peers would endorse this decision.

I hypothesise that the experience of all-mail voting in a voter's social network can reduce the social, information, and convenience costs associated with voting by mail. First, and most evidently, social costs are likely to decrease due to conformity pressures from the increased use of mail ballots by peers in jurisdictions switching to all-mail voting. Second, information flows from peers in jurisdictions switching to all-mail voting can increase both awareness of voting by mail as a potential option as well as familiarity with the procedural steps involved, thereby decreasing information costs. As anticipated by Downs (1957) and a rich literature on the role of information diffusion in vote choice (e.g. Huckfeldt and Sprague, 1991; Katz and Lazarsfeld, 1955; Santoro and Beck, 2017), other voters are an important source of politically relevant information shaping voter preferences. Reduced convenience costs are a third prospect; insofar as voters in all-mail voting jurisdictions become more knowledgable about casting mail ballots, they may help their peers to do so too.

In the context of this chapter, reductions in social costs via conformity pressures and in information costs via increased awareness around voting by mail are arguably the most likely mechanisms. Voters in North Carolina do not automatically receive a mail ballot, unlike voters in jurisdictions implementing all-mail voting. Consequently, information costs relating to the process of applying for mail a ballot are likely to be unaffected from policy changes in these jurisdictions. Similarly, to the extent that rules around the completion and return of mail ballots differ between states, reductions in information costs relating to the procedural aspects of voting by mail are likely to be more limited. For instance, North Carolina is one of the few states where mail ballot completion requires witness signatures (National Conference of State Legislatures, 2022b). Given such differences, the reduction of convenience costs becomes an even less likely prospect.

The conceptual discussion thus far has only considered first-order effects. That is, it has only described how interacting with one or more peers in jurisdictions switching to all-mail voting might affect one's choice of voting method in North Carolina. Though it is possible that this gives rise to higher-order effects whereby voters who have no—or not many—peers in such jurisdictions are influenced by the behaviour of those who do. The reduction of convenience costs and information costs relating to the practicalities of casting mail ballot may play a greater role here, given that higher-order effects can arise from interactions between peers who both vote in North Carolina. In fact, a long-standing literature in electoral geography anticipates that a large part of politically relevant social interaction takes place among individuals in close geographic proximity, resulting in contextual effects whereby aggregate outcomes shape individual voting behaviour within a locality (e.g. Agnew, 1996; Cox, 1969; Johnston, 1983; Johnston and Pattie, 2000, 2014; Wheeler and Stutz, 1971). However, as Johnston and Pattie (2014) highlight, these localised social networks do not exist in a vacuum: while most social ties will plausibly be local, individuals are also likely to have non-local social ties which aid the spatial diffusion of preferences.[6] As such, the aggregate social ties of a locality in jurisdictions switching to all-mail voting can be thought as a kind of contextual influence on choice of voting method in that locality, which can be exerted both on voters who personally have peers in the jurisdictions in question as well as on as those who do not. I collectively refer to the first- and second-order effects arising from this process as socio-spatial spillovers.

[6]This description is echoed in the sociological literature on the effects of 'weak ties' (Granovetter, 1973, 1983): social ties that bridge disparate cliques of peers.

2.4 Data and Measurement

2.4.1 Policy Changes

I use the policy dataset of Amlani and Collitt (2022) to identify counties that switched to all-mail voting between the 2016 and 2020 presidential elections. The dataset covers the District of Columbia and all states except Alaska. I thus append data on Alaskan boroughs based on Herrnson and Stewart (2023).[7] As shown in Figure 2.1, 168 counties and the District of Columbia switched to all-mail voting, with 115 counties doing so as part of statewide rollouts in California, Hawaii, Nevada, New Jersey, and Vermont, and the remaining counties as part of local rollouts in most of Montana and parts of Utah.[8] It is further shown that large geographic distances separate North Carolina from most of these areas. Looking at distances among population centres, the District of Columbia lies closest to North Carolina at roughly 300 kilometres away, followed by New Jersey and Vermont at a respective 600 and 800 kilometres away.[9] All other states lie in the West at well over 2,000 kilometres away.

2.4.2 Social Ties

I use the Social Connectedness Index (SCI) by Bailey et al. (2020) to measure the social ties between zip code tabulation areas (ZCTAs) in North Carolina and counties that switched to all-mail voting between the 2016 and 2020 presidential elections.[10] I access the public release of the SCI via the Humanitarian Data Exchange of the UN Office for the Coordination of Humanitarian Affairs, which is, at the time of writing, based on a snapshot of all active Facebook users (i.e. those who have logged in during the previous 30 days) as of October 2021. The SCI is defined for ZCTA-pairs zz' as follows:

$$SCI_{zz'} = SCI_{z'z} = \phi \frac{FB\_Connections_{zz'}}{FB\_Users_z \times FB\_Users_{z'}} \tag{2.1}$$

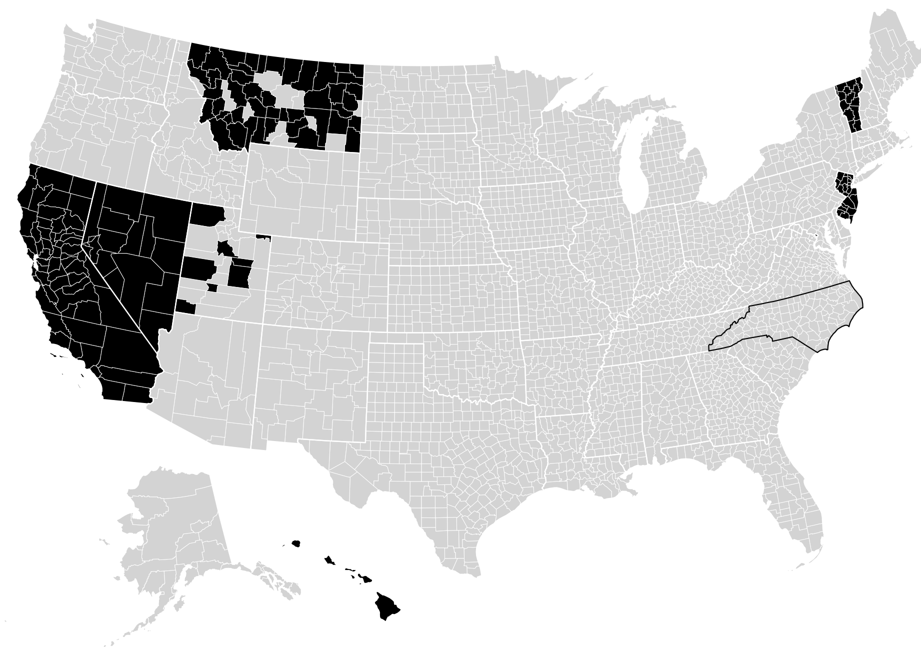
[7] Counties do not feature in the administrative geography of Alaska, Louisiana, and the District of Columbia. However, I follow the U.S. Census Bureau in considering Alaskan boroughs, Louisianan parishes, and the District of Columbia as county equivalents.

[8] Contrary to all other cases, Hawaii and Utah carried out the relevant statewide electoral reforms prior to the onset of the pandemic (National Conference of State Legislatures, 2022b).

[9] Geographic distance is calculated as the great-circle distance between 2020 population-weighted centroids as reported by the U.S. Census Bureau.

[10] As of 2020, there were approximately 32,000 ZCTAs in the United States, 808 of which were in North Carolina. The spatial units are produced and updated decennially by the U.S. Census Bureau using intersections of zip codes and census blocks as building blocks. Each ZCTA is coded according to the zip code in which the majority of addresses in the area are located.

Figure 2.1: US Counties Switching to All-Mail Voting between 2016 and 2020



Notes: Counties switching to all-mail voting are shaded in black. Thicker white lines delineate state borders. The state border of North Carolina is drawn in black. Alaska and Hawaii have been reprojected and rescaled for illustrative purposes. The map is based on Amlani and Collitt (2022). Data on Alaska are based on Herrnson and Stewart (2023).

That is, the SCI is the total number of Facebook friendships in both ZCTAs over the product of active users in each ZCTA. It can therefore be thought of as a measure of the relative probability of Facebook friendship between an active user that resides in z and an active user that resides in z'. Notice that the public release of the index is also multiplied by a scaling factor ϕ for privacy purposes.

A growing body of evidence suggests that Facebook friendships are a sound proxy for real-world social ties (e.g. Bailey et al. 2018; 2018; 2021; 2022, Kuchler et al. 2022, 2021). In principle, given that the date on which the SCI is measured succeeds the date of the 2020 presidential election by almost a year, it is possible that it does not provide an accurate representation of social ties at the time of the election. This concern is partly assuaged by Bailey et al. (2021), who find that the county-level version of the index does equally well in predicting spatial economic interactions today as it does in predicting the same interactions decades ago, suggesting that social ties between counties are fairly stable over time. While the version of the index used in this chapter is observed for smaller spatial units than counties, the prospect of measurement error is plausibly minimised by the relatively short temporal lag with respect to the 2020 presidential election.[11]

[11]There are roughly eight ZCTAs for every county in North Carolina.

Before joining data on social ties with data on county-level policies, I follow Bailey et al. (2021) and aggregate the SCI to the level of ZCTA-county pairs as follows:

$$SCI_{zc} = SCI_{zc} = \sum_{z' \in c} PopShare_{z'} \times SCI_{zz'} \tag{2.2}$$

Here, z' denotes ZCTAs mapping into counties c, and $PopShare_{z'}$ is the population of z' as a share of the population of c.[12] Notice that this calculation implicitly assumes that Facebook usage rates across ZCTAs within counties are roughly equal. While this cannot be tested in the absence of publicly available data on the spatial distribution of Facebook users, previous work using proprietary data suggests that similar values of the index are obtained when Facebook friendships are scaled by total population (e.g. Bailey et al. 2018, 2021). In a similar manner to Kuchler et al. (2022), I then calculate the share of social ties of each ZCTA in North Carolina z that are resident in a county c that switched to all-mail voting between the 2016 and 2020 presidential elections as follows:

$$ShareAMV_z = \sum_{c} \frac{SCI_{zc} \times Pop_c \times dAMV_c}{\sum_c SCI_{zc} \times Pop_c} \tag{2.3}$$

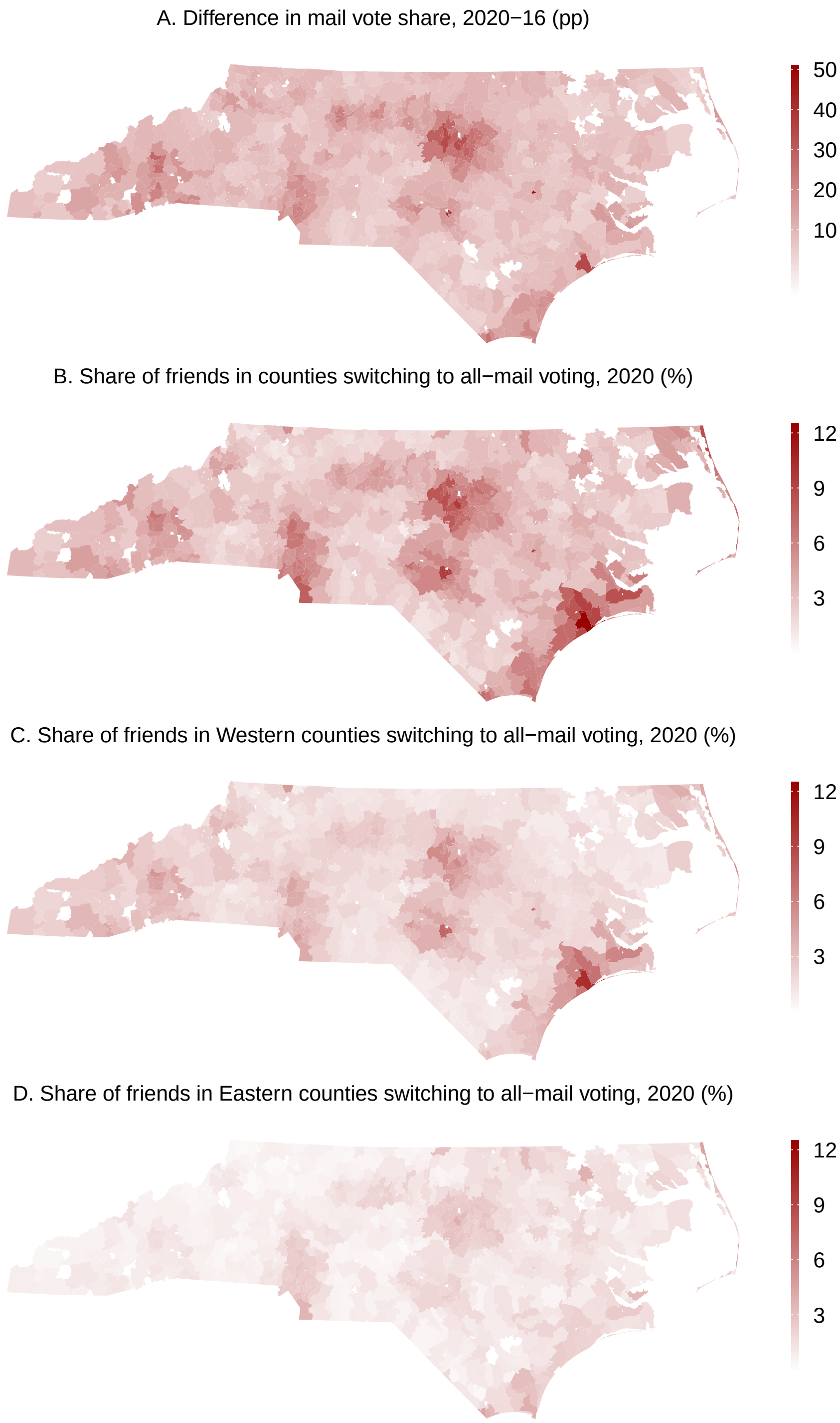
Here, $dAMV_c$ is a dummy variable that is equal to one when county c switched to all-mail voting between elections and Pop_c is the population of county c. There are two main benefits in using this measure over the mean SCI value of switching counties within ZCTAs. First, as the SCI is a relative measure, *ShareAMV* has a more intuitive interpretation. Second, as social influence is characterised by 'complex contagion' (e.g. Christakis and Fowler, 2007; Sprague and House, 2017), it also accounts for the fact that the probability of an individual changing their behaviour is likely to increase with the number of peers that are exerting the influence in question.

Panel B in Figure 2.2 maps the ZCTA-level share of friends in counties switching to all-mail voting between the 2016 and 2020 presidential elections for ZCTAs in North Carolina. A total of 701 ZCTAs are represented—fewer than the state total of 808. That is because the SCI is suppressed for areas with a very small number of Facebook users for privacy purposes.[13] Notably, there is substantial spatial variation with higher shares observed in and around metropolitan areas such as Asheville, Charlotte, Durham, Jacksonville, and Raleigh.

[12] I obtain data on the population share of each ZCTA within each county as of the 2020 Census using the 'Geocorr' application of the Missouri Census Data Centre.

[13] In practice, the suppressed SCI cells often correspond to non-residential areas.

Figure 2.2: Mail Ballots and Social Ties to Policy-Switching Counties, by ZCTA



Notes: The mail vote share is the share of voters that voted by mail, only considering voters who voted in both presidential elections and did not move ZCTAs between elections. Western counties are located in California, Hawaii, Nevada, Montana, and Utah and Eastern counties are located in New Jersey, Vermont, and the District of Columbia.

Panel A in Figure 2.2 maps the difference in the ZCTA-level share of voters that voted by mail between the 2016 and 2020 presidential elections.[14] Here too, it is shown that while most areas saw large increases in the demand for mail ballots, the largest were seen around metropolitan areas. In Panels C and D of the same figure, I also map additional measures $ShareAMV_{(W)}$ and $ShareAMV_{(E)}$, respectively denoting the share of friends in Western and Eastern counties that switched to all-mail voting. These are respectively calculated by multiplying the denominator of the fraction in equation (2.3) with a dummy variable identifying counties in California, Hawaii, Montana, Nevada, and Utah or counties in New Jersey and Vermont, and the District of Columbia. It is shown that most ZCTAs in North Carolina have a higher share of social ties in the substantially more distant Western counties—an observation that may at first seem counterintuitive. However, this is largely due to the fact that the latter group has a much larger population, not least because it includes California: the most populous state. Based on Census estimates, New Jersey, Vermont, and the District of Columbia had a combined population of just over 10 million in 2021: almost four times smaller than the population of California.

As discussed in the background section, pandemic-related concerns were a significant factor in voter turnout and choice of voting method in 2020. As such, to the extent that voters with more social ties in areas switching to all-mail voting are influenced by the local incidence of COVID-19 in these areas, the latter may constitute an important time-varying confounder in the relationship between the former and choice of voting method. In order to investigate this further, I draw on the COVID-19 Dashboard by the Centre for Systems Science and Engineering at Johns Hopkins University to collect data on the cumulative COVID-related deaths per 10,000 people from the onset of the pandemic to a month before the 2020 presidential election. For each ZCTA in North Carolina z, I then calculate the SCI-weighted average of that death rate, $DeathsPer10k_c$, in counties switching to all-mail voting c as follows:

$$SocProxDeathsAMV_z = \sum_c \frac{SCI_{zc} \times DeathsPer10k_c \times dAMV_c}{\sum_c SCI_{zc}} \tag{2.4}$$

The above measure can be interpreted as the social proximity of a ZCTA in North Carolina to COVID-19 deaths in counties switching to all-mail voting.

[14] The mail vote share in each election is calculated as the share of voters that voted by mail, only considering a sub-panel of voters that voted in both elections and did not move ZCTAs between elections. A total of 3,183,515 voters and 699 ZCTAs are represented. The data used is discussed in detail in subsection 2.4.3.

2.4.3 Voting

Using publicly available voter registration files published by the North Carolina Board of Elections, I obtain data on the entire population of individuals that were eligible to vote in both the 2016 and 2020 presidential election. Other than individual-level data on demographics and registration with a political party, the files include information on addresses, which allows me to match individuals to ZCTAs. I achieve this in three steps. First, I use the U.S. Census Bureau Geocoder to geolocate the addresses, which is successful in 93 per cent of cases.[15] I then match the geolocated addresses to ZCTAs by performing point-in-polygon spatial joins using ZCTA boundary data from the U.S. Census Bureau. Finally, for addresses that could not be geolocated, I consider the address zip code as equivalent to the ZCTA in which the individual resides. Recall that a ZCTA assumes the code of the most common zip code among its residents; the latter approach will thus give rise to more mismatches in cases where ZCTA residents are evenly distributed across multiple zip codes. Though given that the addresses matched in this way are a very low proportion of total addresses, the prospect of measurement error from potential mismatches is plausibly of little concern.

I further draw on the voter history files also published by the North Carolina Board of Elections, which provide data on past participation and choice of voting method in all elections for a moving window of ten years. Upon joining the registration and history files, and discarding observations in ZCTAs for which data on social ties is unavailable in either 2016 or 2020, I observe a panel of 4,755,385 individuals that were eligible to vote in both elections and a sub-panel of 3,831,827 that voted in both elections.[16]

As discussed in the conceptual framework section, there is evidence that changes in polling place locations can dissuade voters from in-person election day voting due to increased convenience costs. While the evidence suggests that voters reacting to these changes are most likely to abstain from voting all together, significant positive effects have also been observed on voting by mail. Local polling place changes may thus be a potentially important time-varying confounder in my analysis insofar as they are correlated with the share of social ties in counties switching to all-mail voting. To examine this further, I also identify the voters that were affected by such changes in the 2020 presidential election using the precinct-level polling place files published by the North Carolina Board of Elections.

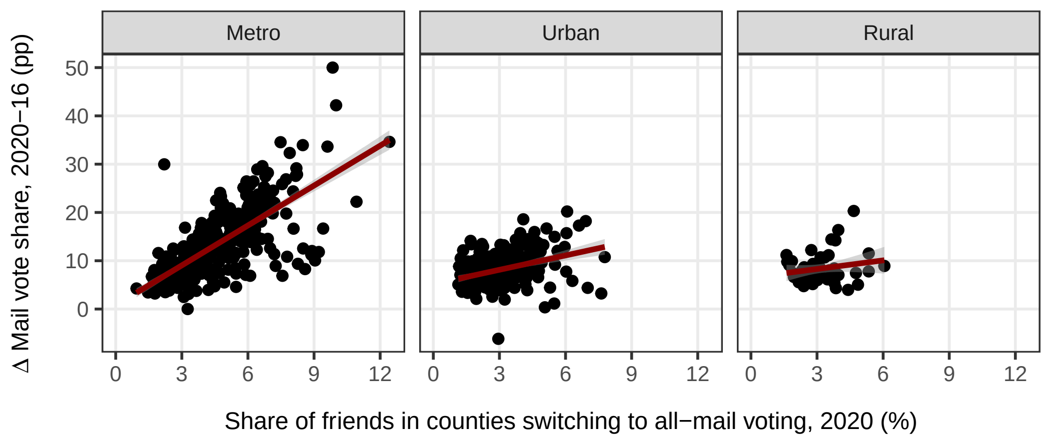
[15] I use the 'censusxy' R package by Prenner et al. (2022) to access the application programming interface of the U.S. Census Bureau Geocoder.

[16] A small number of individuals for whom there are duplicate, conflicting records in the voter history files are also excluded from the main panel.

2.4.4 Urbanisation and Broadband

As per the previous discussion of Figure 2.2, more densely populated areas appear to have more social ties to counties that switched to all-mail voting between the 2016 and 2020 presidential elections as well as having seen greater increases in demand for mail ballots between elections. To examine the role of urbanisation in this relationship, I use the county-level rural-urban continuum classification of the U.S. Department of Agriculture to characterise each voter as resident in a metropolitan, urban, or rural county.[17] An initial naive inspection of Figure 2.3 suggests a positive relationship between ZCTA-level differences in mail vote shares between elections and shares of social ties in counties switching to all-mail voting which weakens in a gradient-like fashion as urbanisation decreases.

Figure 2.3: Mail Ballots and Social Ties to Policy-Switching Counties, by Urbanisation



Notes: The dots represent zip code tabulation areas (ZCTA) as of 2020. The mail vote share is the share of voters that voted by mail, only considering voters who voted in both presidential elections and did not move ZCTAs between elections. Urbanisation categories are derived from the county-level rural-urban continuum codes by the U.S. Department of Agriculture.

A sizeable literature further documents the role of modern communication technologies, such as the Internet and social media, in influencing voting behaviour (Zhuravskaya et al., 2020). Given the fair geographic distance of most counties that switched to all-mail voting between the 2016 and 2020 presidential elections relative to North Carolina, such technologies have a potentially important role in facilitating social interactions between residents in the state and these counties. To investigate this, I use the 2019 version of the Census-tract-level Broadband Availability and Quality Index produced by the North Carolina Department of Information Technology.

[17] This is preferred over the Census Bureau's dichotomous rural-urban classification as previous work suggests that the influence of urbanisation on political outcomes is best understood as a gradient moving outward from metropolitan cores (e.g. Scala and Johnson 2017).

2.5 Empirical Strategy

The empirical objective of this chapter is to examine the causal relationship between the rollout of all-mail voting in various jurisdictions between the 2016 and 2020 US presidential elections and voting by mail in North Carolina. To this end, I begin with the following baseline linear probability difference-in-differences specification:

$$VBM_{izct} = \beta ShareAMV_z \times d2020_t + \eta_u + \theta_{ct} + \epsilon_{izct} \tag{2.5}$$

Here, i indexes individuals residing in ZCTAs z and counties c in North Carolina, and t indexes election years. Correspondingly, VBM_{izct} is a dummy variable indicating whether i voted by mail in t, $ShareAMV_z$ is the share of the aggregate out-of-z social ties of the residents of z that are located in the out-of-state counties that switched to all-mail voting between elections, and $d2020_t$ is a dummy variable identifying observations in the 2020 presidential election. The term η_u denotes any out of a pair of unit fixed effect terms η_z, and η_{iz}, respectively denoting ZCTA and individual-ZCTA fixed effects. Each of these terms serves to absorb time-invariant heterogeneity at the respective unit level, while η_{iz} further effectively discards individuals who moved ZCTAs between elections. Similarly, the county-year fixed effects η_{ct} absorb time-varying factors at the county level. Notably, the latter are likely to include the county-level incidence of COVID-19 cases and deaths and other local factors relating to the COVID-19 pandemic. Finally, ϵ_{izct} is a heteroscedasticity-robust error term clustered at the county-year level. The coefficient of interest β thus summarises the relationship between the ZCTA-level share of social ties switching to all-mail voting and the individual-level probability of voting by mail.

The interpretation of β crucially depends on the set of assumptions invoked (Callaway et al., 2021). Under the standard parallel trends assumption that is often invoked in difference-in-differences specifications with a continuous treatment, groups that experienced different levels of the treatment follow the same path of untreated potential outcomes. In this case, β will be a weighted average of $ACRT(d|d)$—the average causal response among those that experienced treatment level d to a marginal change in the treatment—plus a 'selection bias' term. Intuitively, the latter term captures differences in the effect of experiencing the same treatment level between groups that actually experienced marginally different levels. To identify $ACRT^*$,[18] which is the target parameter for the estimand $\hat{\beta}$, one must thus assume that there is no such bias.

[18]By $ACRT^*$ I denote the weighted average over all $ACRT(d|d)$.

To gauge whether the parallel trends assumption is plausible in the context of specification (2.5), I carry out a placebo test. Namely, I run a similar specification that also considers voting behaviour in the 2018 midterm election, which precedes the implementation of all-mail voting in the jurisdictions in question. Namely, I append a term interacting the treatment with a midterm dummy $d2018_t$ and effectively pretend that the treatment took place between the 2016 election and the 2018 midterm—earlier than it truly did. A near-zero coefficient on that interaction term would thus indicate common pre-trends across groups with different treatment levels, which is suggestive of a common path of untreated potential outcomes in the post-treatment period.

The choice of a placebo date that is between the original two time periods is unconventional as this typically precedes both periods. This is first motivated by the fact that data on the preceding 2012 presidential election is not readily available, given that the North Carolina Board of Elections publishes files for a 10-year moving window. While data on the 2014 midterm is available, the 2018 midterm is preferred for two reasons. As the latter recorded the highest turnout in a midterm in several decades (Misra, 2019), using this election partly assuages concerns over the lower salience of midterms relative to presidential elections, which may normally make comparisons between the two problematic. Paired with the closer temporal proximity of the 2018 midterm to the 2020 presidential election, this also aids me in retaining most of the panel of voters that were eligible to vote in both 2016 and 2020.

There is currently no consensus on how the kind selection bias described by Callaway et al. (2021) might be gauged. The authors propose that the specific context of the application in question can indicate whether assuming the absence of selection bias is plausible. In the context of this chapter, age, urbanisation, exposure to the pandemic, and partisanship are arguably the most salient potential sources of treatment effect heterogeneity. Older individuals as well as those in less urbanised areas are likely to incur higher convenience costs from in-person election day voting, also rendering them more likely to adjust their choice of voting method in response to the adoption of all-mail voting in their social network. Individuals in areas with a higher incidence of COVID-19 cases may too be more inclined to vote by mail given the heightened health risk associated with voting in person. Conversely, given that the incumbent Republican president was vocally opposed to the voting method in the weeks preceding the 2020 presidential election (Clinton et al., 2022), it is also possible that Republican voters would incur higher social costs from voting by mail relative to Democrats. These prospects become concerning for the identification of average causal responses to the ZCTA-level share of social ties switching to all-mail voting given that higher shares are spatially clustered in more urbanised areas (Figures 2.1 and 2.2), which also tend to consist of younger, more densely distributed, and more Democratic populations (Rodden, 2010).

That is, while higher shares may indeed have stronger causal effects on the probability of voting by mail, it might be difficult to gauge the extent in which differences between causal effects in higher and lower-share ZCTAs are due to experiencing a different treatment level or treatment effect heterogeneity.

The inclusion of county-year fixed effects in specification (2.5) arguably assuages some of the above concerns as it essentially restricts the analysis to comparisons of trends between ZCTAs in the same county. Other than being more homogenous in terms of treatment levels, spatially proximate ZCTAs are also likely to be much more similar in terms of urbanisation and exposure to the pandemic, thus minimising the prospect of selection bias along these dimensions. However, as there may still be substantial variation in age and partisanship at the sub-county level, I also run augmented specifications interacting the county-year fixed effects with party and age fixed effects,[19] effectively restricting the analysis to comparisons of the trends of co-partisans in the same age and county.

I further run a set of triple-differences specifications to explicitly examine treatment effect heterogeneity across different dimensions and separately estimate socio-spatial spillovers from the West and East.[20] In the former case, this involves regressing VBM_{izct} on $ShareAMV_z$ and the grouping variable at hand, the two-way interactions among themselves and the 2020 election dummy, the three-way interaction of all three variables, and the fixed effects. As is the case in the baseline specification, note that some of these terms will be effectively absorbed by the fixed effects. Correspondingly, the latter case involves regressing VBM_{izct} on the fixed effects, the two-way interactions of the geographically split treatments $ShareAMV_{z(W)}$ and $ShareAMV_{z(E)}$ with the 2020 election dummy, as well as the three-way interaction of all three variables.

Even in the absence of selection bias, depending on the distribution of the treatment, the estimand $\hat{\beta}$ will place more weight on particular parameters when averaging the average causal responses at different treatment levels $ACRT(d|d)$ (Callaway et al., 2021). This could mean that $\hat{\beta}$ is less representative of average causal responses at the tail ends of the distribution. To investigate this, I also run a set binary difference-in-differences specifications of the following form:

$$VBM_{izct} = \gamma 1[ShareAMV_z \in (l, h]] \times d2020_t + \eta_u + \theta_{ct} + \epsilon_{izct} \tag{2.6}$$

[19] Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties.

[20] In a difference-in-differences design with two treatments, two time periods, and no treated units in the first period, adding the interaction of the two treatments can eliminate contamination bias (de Chaisemartin and D'Haultfoeuille, 2022).

Here, l and h define the interval of treatment levels that determine a given treatment group, while observations where $ShareAMV_z$ is equal to or below the mean serve as the comparison group. In constructing the treated groups, I successively consider three sets of observations where $ShareAMV_z$ is respectively higher than the mean and up to one standard deviation above the mean, higher than one standard deviation above the mean and up to two standard deviations above the mean, and higher than two standard deviations above the mean up to three standard deviations above the mean. Notably, under a standard parallel trends assumption, the coefficient γ on each of the three resulting specifications can be interpreted as the average treatment effect on the treated $ATT(d|d)$: the causal effect of experiencing a share of social ties switching to all-mail voting d in the respective interval among those that experienced shares in that interval.[21] While not necessary for the identification of $ATT(d|d)$, an interpretation of the difference between any two estimands as an average causal response further rests on the previously discussed assumption that there is no selection bias between the treatment levels in question.

In summary, the binary specifications allow for the identification of ATT-type parameters, which can be policy-relevant in their own right, while also affording comparisons of causal effects between different intervals of shares, which can provide a sense of how average causal responses may vary locally. When interpreting the estimand $\hat{\gamma}$ as the causal effect on voting by mail associated with belonging to a given treatment group it is important to remember that, in the absence of never-treated units, the comparison group is inevitably consisted of units treated at low levels of the treatment. Consequently, insofar as effects are also present at these lower levels, it would be more suitable to interpret the estimand as a lower bound of the effect in question.

2.6 Results

Table 2.1 presents baseline estimates for both the full panel of North Carolinian voters that were eligible to vote in both the 2016 and 2020 presidential elections as well as the sub-panel of those that voted in both elections. Appendix Table 2.A.1 also displays some key summary statistics for the full panel, showing that 15.6 per cent voted by mail in 2020, up from 3.22 per cent in 2016. Looking at the first two columns of Table 2.1, an increase in the ZCTA-level share of social ties that switched to all-mail voting between elections by one standard deviation (1.77 percentage points) is shown to correspond to an increase in the probability of voting by mail by roughly 3.8 percentage points.

[21] Notice that, as per the terminology of Callaway et al. (2021), $ATT(d|d)$ is a 'level effect' whereas $ACRT(d|d)$ is a 'slope effect'. That is, the former is informative with regards to untreated potential outcomes while the latter concerns treated potential outcomes under different treatment levels.

In the third column, a similar figure is obtained when leveraging comparisons among individuals in the same age group, political affiliation, and county, suggesting that baseline estimates are unlikely to suffer from selection bias along these dimensions.

Table 2.1: Difference-in-Differences Estimates, Continuous Treatment

	Voted by mail					
	Panel A: *Eligible in both elections*			**Panel B:** *Voted in both elections*		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of friends in counties switching to AMV × d_{2020}	0.0384^a (0.0038)	0.0385^a (0.0055)	0.0393^a (0.0038)	0.0459^a (0.038)	0.047^a (0.0056)	0.0468^a (0.0037)
Observations	9,510,770	7,712,362		7,663,654	6,367,022	
Individuals	4,755,385	3,856,181		3,83x1,827	3,183,511	
ZCTA	701	700		701	699	
Counties	100	100		100	100	
Fixed effects:						
ZCTA	✓			✓		
Indiv. × ZCTA		✓	✓		✓	✓
County × Year	✓	✓		✓	✓	
(×Age × Party)			✓			✓

Notes: Robust standard errors in parentheses are clustered at the county-year level in columns 1, 2, 4, and 5, and at the age-party-county-year level in columns 3 and 6. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties. The treatment is standardised to have zero mean and unit variance and is measured at the level of zip code tabulation areas (ZCTA) as of 2020. $^a p < 0.001$

In Appendix Table 2.A.2, I also report the results of the placebo tests corresponding to each specification in Table 2.1. Across columns, it is shown that the estimated effects of the ZCTA-level share of social ties that switched to all-mail voting on the probability of voting by mail in the 2018 midterm election are near zero. This implies that the outcomes of individuals that eventually experienced different shares followed similar pre-trends, which is suggestive evidence that the identification assumption of common untreated potential outcomes is plausible.

The last three columns of Table 2.1 show that a moderately stronger effect is observed when restricting the panel to voters that voted in both elections. As shown in Appendix Table 2.A.3, this is partly explained by a weak negative relationship between the share of social ties switching to all-mail voting and the probability of turning out to vote.

As there is evidence that local rollouts of all-mail voting between 2016 and 2020 had positive effects on turnout (Amlani and Collitt, 2022; McGhee et al., 2022), it seems unlikely that this is driven by negative spillovers from decreased turnout in jurisdictions implementing the policy. While uncovering the mechanisms underpinning this relationship is beyond the scope of this chapter, an alternative explanation that has found some empirical support is that the implementation of absentee voting policies increases the convenience of voting at the expense of the social pressure to vote—especially among casual voters (Burden et al., 2014; Thompson, 2004). As those opting to cast a mail ballot prior to the day of the election may become less engaged in political conversations, it is possible that their peers are in receipt of fewer prompts to turn out to vote.

Table 2.2 presents difference-in-differences estimates from specifications where the treatment is split into the respective shares of social ties that are located in Western and Eastern counties that switched to all-mail voting between the 2016 and 2020 presidential elections. It is shown that the identified effects on choice of voting method among North Carolinian voters seem to be mostly driven by policy changes in the more distant and populous Western counties. In specifications using the full panel and county-year fixed effects, an increase in the share of social ties in Western counties by one standard deviation (0.012 percentage points) corresponds to an increase in the probability of voting by mail by roughly 3 percentage points. In contrast, an increase in the share of social ties in Eastern counties by one standard deviation (0.0077 percentage points) corresponds to an increase in probability of roughly 1 percentage point. Socio-spatial spillovers from the West remain stronger when considering equal marginal changes: a 1 percentage point increase in Western shares corresponds to a 2.5 percentage point increase in probability, while a 1 percentage point increase in Eastern shares corresponds to a 1.2 percentage point increase. Notably, spillover effects from Eastern counties approach zero in the specifications with age-party-county-year fixed effects, which might be indicative of selection bias in the baseline estimates.

It is important to note that the results on Table 2.2 do not necessarily imply that social interactions with peers in Western counties are qualitatively more influential. Given that shares of social ties in Eastern counties are generally much lower, it could be that the weaker spillovers from these counties are due to smaller differences in the causal effects associated with marginally different shares at these lower levels. Though the results do suggest that spillover effects on voting by mail from the rollout of all-mail voting can and do travel long geographic distances.

Table 2.2: Difference-in-Differences Estimates, Split Continuous Treatment

	Voted by mail					
	Panel A: *Eligible in both elections*			**Panel B:** *Voted in both elections*		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of friends in Western counties switching to AMV $\times d_{2020}$	0.0302[a] (0.0044)	0.0286[a] (0.0058)	0.032[a] (0.0042)	0.0358[a] (0.0045)	0.0348[a] (0.0062)	0.0386[a] (0.0046)
Share of friends in Eastern counties switching to AMV $\times d_{2020}$	0.0086[b] (0.0032)	0.0098[c] (0.0044)	0.0033 (0.0033)	0.0127[a] (0.0033)	0.0139[b] (0.0048)	0.0051 (0.0037)
Observations	9,510,770	7,712,362		7,663,654	6,367,022	
Individuals	4,755,385	3,856,181		3,831,827	3,183,511	
ZCTA	701	700		701	699	
Counties	100	100		100	100	
Fixed effects:						
ZCTA	✓			✓		
$Indiv. \times ZCTA$		✓	✓		✓	✓
$County \times Year$	✓	✓		✓	✓	
$(\times Age \times Party)$			✓			✓

Notes: Robust standard errors in parentheses are clustered at the county-year level in columns 1, 2, 4, and 5, and at the age-party-county-year level in columns 3 and 6. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties. The treatment variables are standardised to have zero mean and unit variance and are measured at the level of zip code tabulation areas (ZCTA) as of 2020. Each column also includes the three-way interaction among the treatment variables and the 2020 dummy. Western counties are located in California, Hawaii, Nevada, Montana, and Utah and Eastern counties are located in New Jersey, Vermont, and the District of Columbia. $^a p < 0.001$; $^b p < 0.01$; $^c p < 0.05$

Table 2.3: Triple-Difference Estimates, Heterogeneous Effects

	Voted by mail - *eligible in both elections*					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of friends in counties switching to AMV × d_{2020}	0.0332^{a} (0.0063)	0.0494^{a} (0.0068)	0.041^{a} (0.0061)	0.0375^{a} (0.0054)	0.041^{a} (0.006)	0.0323^{a} (0.0043)
(× Aged 65 or over)	0.0218^{a} (0.0038)					
(× Democrat)		-0.005 (0.0032)				
(× Republican)		-0.0232^{a} (0.004)				
(× Urban county)			-0.0201^{b} (0.0073)			
(× Rural county)			-0.0307^{b} (0.01)			
(× Pol. place change)				0.0054^{b} (0.0019)		
(× Soc. prox. deaths)					-0.0024 (0.0033)	
(× Broadband index)						0.011^{a} (0.002)
Observations			7,712,362			7,161,450
Individuals			3,856,181			3,580,725
ZCTA			700			699
Counties			100			100
Indiv-ZCTA FE	✓	✓	✓	✓	✓	✓
County-Year FE	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors in parentheses are clustered at the level of county-years. Continuous independent variables are standardised to have zero mean and unit variance. The treatment is measured at the level of zip code tabulation areas (ZCTA) as of 2020. In the first four columns, the respective reference groups are those aged under 65, those who are unaffiliated or affiliated with another political party, those residing in metropolitan counties, and those residing in precincts where there were no polling place changes between elections. $^{a}p < 0.001$; $^{b}p < 0.01$

Table 2.3 presents estimates from a set of triple-difference specifications exploring heterogeneous effects for different groups of individuals. All specifications are based on the full panel of individuals who were eligible to vote in both the 2016 and 2020 presidential elections and include individual-ZCTA and county-year fixed effects, which makes them analogous to the baseline specification corresponding to column 2 of Table 2.1. As shown on column 1, individuals aged over 65 were more likely to switch to voting by mail in response to rollouts of all-mail voting in other jurisdictions. This echoes evidence on the relative popularity of the voting method among older voters (Barreto et al., 2006; Plescia et al., 2021), in line with expectation this demographic might be faced with higher convenience costs when voting in person. Column 2 further shows that socio-spatial spillovers from the rollout of all-mail voting are weaker for registered Republicans relative to both Democrats and unaffiliated voters. Given that the Republican Party leadership was publicly opposed to the voting method (Clinton et al., 2022), one potential explanation for this finding is that Republicans would need to overcome higher social costs in order to cast a mail ballot in 2020 presidential election. Column 3 also suggests that spillover effects decline in a gradient-like fashion as urbanisation decreases. This is consistent with the expectation that voters in the metropolitan periphery are likely to face higher travel costs when accessing polling places (Gimpel and Schuknecht, 2003), and that voters in less urbanised, more closely knit communities may be more incentivised to comply with the social norm of voting in person (Funk, 2010).

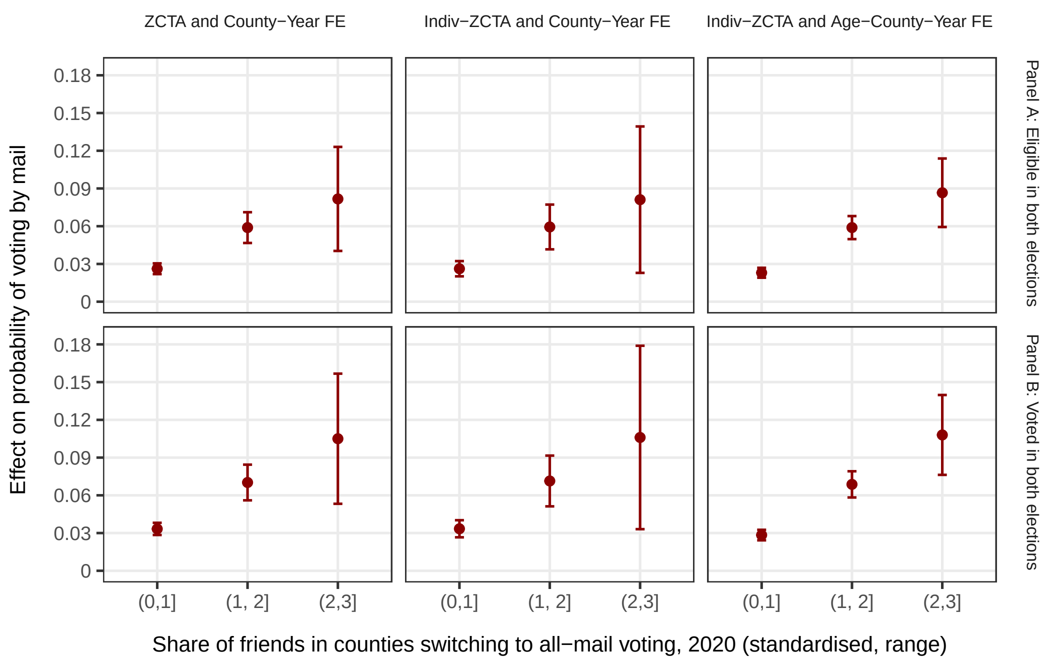
Columns 4 and 5 of Table 2.3 show that there is little to no spillover effect heterogeneity by exposure to polling place changes and social proximity to instances of pandemic-related deaths in jurisdictions switching to all-mail voting. This assuages concerns over the confounding due to these salient time-varying factors. Further, the last column shows that choice of voting method among individuals in Census tracts with better broadband availability was slightly more responsive to the share of social ties switching to all-mail voting.[22] While this could indicate that online social interactions are more conductive to socio-spatial spillovers in line with previous work on the political effects of the Internet and social media (e.g. Geraci et al. 2022; Zhuravskaya et al. 2020), it is difficult to disentangle the role of broadband availability from that of other correlated tract-level characteristics.

As discussed in the previous section, difference-in-difference estimates in specifications with a continuous treatment may misrepresent average causal responses on the tail ends of the distribution of the treatment. Figure 2.4 presents estimates from alternative binary specifications successively comparing individuals experiencing treatment levels equal to or below the mean to different treatment groups experiencing higher levels.

[22] Notice that this specification is based on a smaller panel due to data limitations.

Namely, the comparison group is consisted of individuals in ZCTAs where the aggregate share of social ties in counties switching to all-mail voting was below 3.8 per cent, while the respective ranges for the three treatment groups that are separately considered are over 3.8 and below 5.57 per cent (i.e. over the mean and up to one standard deviation above), over 5.57 and below 7.34 per cent (i.e. over one and up to two standard deviations above the mean), and over 7.34 and below 9.11 per cent (i.e. over two and up to three standard deviations above the mean). As shown in Appendix Figure 3.A.1, placebo tests indicate that all comparison and treatment groups followed similar trends in voting behaviour prior to the 2020 presidential election. Thus, the estimates can be interpreted as lower bounds of the causal effect on the probability of voting by mail for those that experienced a given range of shares of social ties switching to all-mail voting.

Figure 2.4: Difference-in-Differences Estimates, Binary Treatment



Notes: Each point corresponds to a separate specification where the comparison group is consisted of observations for which the standardised ZCTA-level share of friends in counties switching to all-mail voting is zero or lower (i.e. equal to or below the mean). The mean is 3.8% the standard deviation is 1.77%. The error bars denote 95% confidence intervals. Robust standard errors are clustered at the county-year level in the plots on the left and middle and at the age-party-county-year level in the plots on the right. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties.

It is shown that treatment effects become more substantial as shares increase. Looking at specifications using the full panel of individuals, the probability of voting by mail increased by roughly 6 percentage points in ZCTAs where between 5.57 and 7.34 per cent of all residents' social ties switched to all-mail voting between elections. When considering the sub-panel of those who voted in both the 2016 and 2020 presidential elections, the corresponding estimate rises to 7 percentage points. To put this into context, this suggests that a jurisdiction with 100,000 regular voters that experienced shares in the discussed range could have received at least 7,000 fewer mail ballots had none of its aggregate social ties switched to all-mail voting. The respective effect for experiencing shares in the range between 7.34 and 9.11 per cent is stronger by a factor of approximately 1.5—albeit less precisely estimated.

Notably, under the assumption of no selection bias between ranges of shares, comparisons between treatment effects at successive ranges suggest that causal responses decay as shares increase. For instance, in the specifications using the full panel and including ZCTA and county-year effects, the difference in treatment effects between the first and second ranges stands at 3.3 percentage points, while the difference between the second and third ranges stands at 2.3 percentage points. This implies that the results obtained using the original continuous treatment are more representative of causal responses to marginal changes closer to the mean.

2.7 Conclusion

In this chapter, I have shown that local rollouts of all-mail voting can affect demand for mail ballots in distant areas. Exploiting a detailed administrative dataset on North Carolinian voters and data on the near-universe of Facebook friendships among ZCTAs across the United States, I find that social interactions with jurisdictions switching to all-mail voting are likely to have contributed to the popularity of voting by mail in the 2020 presidential election. The analysis highlights the role of social network structure in the spatial diffusion of voter preferences, departing from conventional notions of proximity in electoral geography.

I further find that older voters, non-Republicans, and voters in metropolitan areas were most likely to switch to voting by mail in response socio-spatial spillovers from the implementation of all-mail voting in other jurisdictions. This is in line with the expectation that older voters and those in metropolitan suburbs face increased convenience costs in accessing polling places, and that elite political messaging increased the social costs of voting by mail faced by Republicans in the 2020 presidential election.

In all, these results contribute to a growing body of evidence on the effects of various electoral policies on choice of voting method, which has attracted interest following the COVID-19 pandemic. While previous work has confirmed the intuition that election officials in jurisdictions implementing all-mail voting are likely to be faced with unprecedented numbers of mail ballots, I show that demand is also likely to increase in socially connected no-excuse absentee voting jurisdictions. Given that the latter is a potentially unexpected prospect, information on aggregate social ties can be a relevant input in resource allocation decisions in such jurisdictions.

2.A Appendix

2.A.1 Additional Tables

Table 2.A.1: Summary Statistics, Panel Eligible to Vote in Both 2016 and 2020

	Mean	SD	Min	Max
Individual-level variables (N = 4,755,385)				
t = 2016				
Voted	0.891	0.311	0	1
Voted by mail	0.0322	0.177	0	1
t = 2020				
Voted	0.876	0.33	0	1
Voted by mail	0.156	0.363	0	1
ZCTA-level variables (N = 701)				
Share of friends in:				
Counties switching to AMV	0.038	0.0177	0.0095	0.124
Western counties switching to AMV	0.0243	0.012	0.0075	0.103
Eastern counties switching to AMV	0.0137	0.0077	0.002	0.0451

Table 2.A.2: Placebo Tests, Continuous Treatment

	Voted by mail					
	Panel A: *Eligible in all elections*			**Panel B:** *Voted in 2016 and 2020*		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of friends in counties switching to AMV × d_{2018}	-0.0055^{c} (0.0027)	-0.0056 (0.0033)	-0.005^{c} (0.0022)	-0.0045 (0.0026)	-0.0041 (0.0033)	-0.0037 (0.0022)
Share of friend in counties switching to AMV × d_{2020}	0.0386^{a} (0.0041)	0.0386^{a} (0.0051)	0.0395^{a} (0.0034)	0.046^{a} (0.041)	0.047^{a} (0.0052)	0.0469^{a} (0.0033)
Observations	13,656,237	11,233,104		11,275,353	9,430,338	
Individuals	4,552,079	3,744,368		3,758,451	3,183,511	
ZCTA	701	700		701	699	
Counties	100	100		100	100	
Fixed effects:						
ZCTA	✓			✓		
Indiv. × ZCTA		✓	✓		✓	✓
County × Year	✓	✓		✓	✓	
(Age × Party)			✓			✓

Notes: Robust standard errors in parentheses are clustered at the county-year level in columns 1, 2, 4, and 5, and at the age-party-county-year level in columns 3 and 6. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties. The treatment is standardised to have zero mean and unit variance and is measured at the level of zip code tabulation areas (ZCTA) as of 2020. Individuals for whom there are no observations in any given year are excluded. $^{a}p < 0.001$; $^{c}p < 0.05$

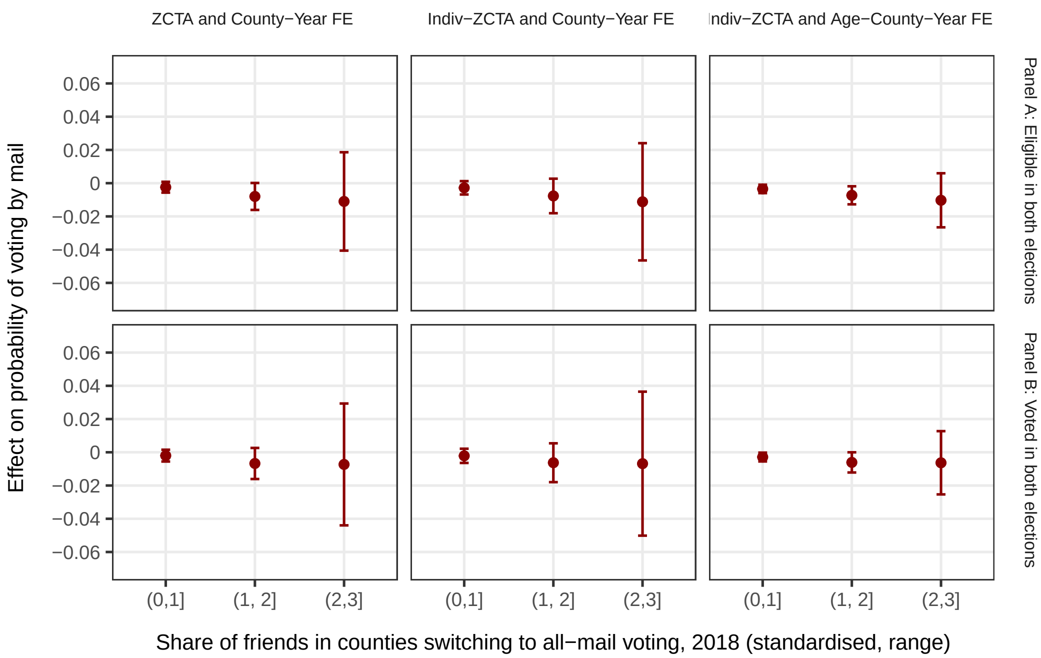
Table 2.A.3: Difference-in-Differences Estimates for Turnout, Continuous Treatment

	Voted - *eligible in both elections*		
	(1)	(2)	(3)
Share of friends in counties switching to AMV × d_{2020}	-0.0076[a] (0.0022)	-0.0089[c] (0.0037)	-0.0086[a] (0.0022)
Observations	9.510,770	7,712,376	
Individuals	4,755,385	3,856,181	
ZCTA	701	700	
Counties	100	100	
Fixed effects:			
$ZCTA$	✓		
$Indiv. \times ZCTA$		✓	✓
$County \times Year$	✓	✓	
$(\times Age \times Party)$			✓

Notes: Robust standard errors in parentheses are clustered at the county-year level in columns 1 and 2, and at the age-party-county-year level in column 3. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties. The treatment is standardised to have zero mean and unit variance and is measured at the level of zip code tabulation areas (ZCTA) as of 2020. $^{a}p < 0.001$; $^{b}p < 0.01$; $^{c}p < 0.05$

2.A.2 Additional Figures

Figure 2.A.1: Placebo Tests, Binary Treatment



Notes: Each point corresponds to a separate specification where the comparison group is consisted of observations for which the standardised ZCTA-level share of friends in counties switching to all-mail voting is zero or lower (i.e. equal to or below the mean). The mean is 3.8% the standard deviation is 1.77%. The error bars denote 95% confidence intervals. Estimates correspond to the interaction of the treatment with the 2018 midterm dummy. Robust standard errors are clustered at the county-year level in the plots on the left and middle and at the age-party-county-year level in the plots on the right. Age is considered as a dummy variable identifying those aged 65 or over. Party affiliation is considered in terms of three groups: Democrats, Republicans, and all others including the unaffiliated and those affiliated with minor parties.

This page has been intentionally left blank.

CHAPTER THREE

The Geography of Partisan Homophily in the 2020 US Presidential Election

Partisan segregation in the United States is often interpreted as evidence of limited social interaction among out-partisans, or partisan homophily. Though little is in fact known about the arrangement of voters in social space. In this chapter, I draw on data on the pairwise density of social ties between 22,537 zip code tabulation areas (ZCTA) and 2020 US presidential election results to explore the extent in which different places are socially connected to politically similar others. Using the local Moran index, I first identify clusters of ZCTAs where there is evidence of partisan homophily or heterophily. In a series of multinomial logit specifications, I then examine differences in the probability of each cluster in different settlement types and regions, and among areas with differences in the relative density and geographic distance of social ties elsewhere. In line with accounts of a segregated social space, I find that partisan homophily is the norm across areas, broadly tracking partisan segregation along the urban-rural continuum. However, the evidence is also suggestive of substantial spatial variation in the density and geographic range of co-partisan social ties: on average, homophilous Democratic-leaning areas, which are most likely to be in cities and suburbs, are much more likely than their Republican-leaning counterparts to be socially connected to distant areas.

3.1 Introduction

Residential segregation by partisanship, or partisan segregation, has been on the rise in the United States over the last few decades. Between the 1992 and 2020 presidential elections, the share of the electorate that lived in counties where the winning candidate gained more than 60 per cent of the two-party vote—often termed 'landslide' counties—gradually rose from roughly 30 per cent to 58 per cent (Bishop, 2020; Darmofal and Strickler, 2019).[1] While such trends are not unprecedented in American electoral history, the recent rise in partisan segregation is characterised by a distinctive urban-rural divide (Mettler and Brown, 2022): Democratic candidates have been gaining higher shares of the vote in more urbanised areas, with Republicans largely dominating elsewhere. Importantly, partisan segregation is also observed at fine spatial resolutions (e.g. Brown and Enos, 2021; Johnston et al., 2020; Kinsella et al., 2021, 2015; Myers, 2013; Rohla et al., 2018; Scala and Johnson, 2017); indeed, even within Democratic-leaning metropolitan counties, the electoral scale has been shown to tip in favour of Republican candidates as one moves away from metropolitan cores.

The relative importance of different factors in the recent rise of partisan segregation has been a subject of lively scholarly and public debate. The residential sorting thesis, which was popularised by Bishop and Cushing (2009), posits that Americans have been more likely to move to areas that are populated by co-partisans, either due to being drawn to like-minded others or socio-cultural factors that correlate with partisanship. However, more systematic work has shown that while partisanship and its correlates do factor into mobility decisions, they are usually outweighed by more universal criteria such as housing affordability, school quality, and the prevalence of crime (Gimpel and Hui, 2015; Martin and Webster, 2020; Mummolo and Nall, 2017). As these factors are expected to constrain mass sorting into politically homogeneous areas, attention has shifted to alternative explanations, such as partisan conversion—be it due to changing local, structural, or institutional conditions—and the replacement of older voters with the young and new. Indeed, using registration data on the near-universe of American voters, Brown et al. (2023) identify generational change and switches in party affiliation as the leading drivers of changes in county-level partisan homogeneity between 2008 and 2020 in Democratic-leaning and Republican-leaning areas respectively.[2]

[1] A steeper trend emerges in the share of counties that saw a landslide victory, which rose from roughly 26 to 77 per cent between 1992 and 2020 (Bishop, 2020; Darmofal and Strickler, 2019).

[2] For changes in partisan homogeneity in Democratic-leaning counties, Brown et al. (2023) also identify new adult entries in the electorate, which include immigrants, as an important driver. In line with the residential sorting thesis, the authors also find that mobility plays a role in counties of both leanings, albeit at a much lesser extent than other drivers.

Many have called attention to the potential implications of rising partisan segregation. One concern is that the concentration of Democratic voters in densely populated, single-seat districts can magnify the disconnect between the popular vote and the partisan composition of local, state, and national legislatures (Chen and Rodden, 2013; Hopkins, 2017; Rodden, 2019). This can then also give rise to discrepancies between implemented policies and policy preferences; for instance, Nall (2018) argues that partisan segregation within metropolitan areas led to the neglect of urban preferences for public transit in favour of suburban preferences for new roads. Finally, others fear that segregation promotes polarisation (e.g. Bishop and Cushing, 2009; Enos, 2017): to the extent that voters interact with others in their residential environment, the spatial clustering of co-partisans may restrict exposure to others with differing political identities and views. As experimental evidence suggests that intergroup contact often acts to constrain affective and ideological distance between disparate social groups (Mutz, 2002; Pettigrew and Tropp, 2006), this may have plausibly contributed to the rise in inter-partisan animosity (Finkel et al., 2020; Iyengar et al., 2019) and intra-partisan ideological homogeneity (Boxell et al., 2017; Gentzkow, 2016) recorded in recent decades, with adverse implications for the function of democratic institutions.[3,4]

The link between rising partisan segregation and issues around political representation is straightforward: as with the related issue of partisan gerrymandering (McGhee, 2020), these merely arise from changes in the arrangement of partisans across political boundaries. Though the claim that partisan segregation implies limited intergroup contact, which might in turn be contributing to political polarisation, is arguably more controversial. As Abrams and Fiorina (2012) suggest, this crucially rests on the assumption that, by and large, an American voter's residential environment—at whatever spatial scale this might be measured—is indeed a sound proxy for their social network. Questions of this sort have interested electoral geographers and urban sociologists for some time. Albeit often constrained by data limitations, relevant work has consistently shown that the spatial concentration of Americans' social ties can vary substantially depending on the kind of settlement in which they reside as well as their demographic characteristics, with higher earners residing in more urbanised areas being more likely to have more geographically distant social ties (e.g. Baybeck and Huckfeldt, 2002a,b; Fischer, 1982; Huckfeldt, 1983, 1982; Logan and Spitze, 1994).

[3]While most scholars agree that Americans have become more affectively polarised in recent decades, there is more disagreement on ideological polarisation. As Gentzkow (2016) demonstrates, both sides of the debate can be right depending on measurement: survey evidence does suggest that most Americans do not hold more extreme views than they used to, but it also shows that partisans are more likely to hold opposing viewpoints than in the past.

[4]As shown by Mason (2015) and others, inter-partisan animosity does not presuppose ideological distance between partisans. Moreover, partisan affect has been shown to act as a strong cue for non-political judgements and behaviours among Americans (Iyengar and Westwood, 2015).

Also emphasised in this literature is the role of telephony and the Internet in expanding the geographic distance over which Americans interact with kin and non-kin alike (e.g. Mok et al., 2007, 2010; Takhteyev et al., 2012; Wellman, 1979, 1996; Wellman and Potter, 1999; Wellman et al., 2003). Indeed, using data on social media friendships, Bailey et al. (2018) find that, on average, the share of social ties of a county's population living more than 200 miles away is 30 per cent. Taken together, these findings suggest that while partisan segregation may be indirectly informative about partisan homophily—the tendency of voters in any given area to be more socially connected to those who voted similarly elsewhere—it may not provide a comprehensive picture of its extent or the geographic ranges over which it is likely to occur.[5]

In this chapter, I explore partisan homophily across zip code tabulation areas (ZCTA) in the United States using data on voting behaviour in the 2020 presidential election and on the density of Facebook friendships between ZCTA-pairs shortly after the election. I operationalise the measurement of partisan homophily as autocorrelation of voting behaviour in social space using the local Moran index, which allows me to identify clusters of ZCTAs of different partisan leanings where there is strong evidence of partisan homophily or heterophily with respect to social ties in other areas. With ZCTAs as the level of measurement, partisan homophily (heterophily) describes cases where the population of a given area tends to have denser social ties with those of others with a similar (dissimilar) composition of partisan vote shares. I find that 70 per cent of the US population resides in relatively homophilous areas, with a further 7 and 23 per cent respectively residing in relative heterophilous areas and areas where there is insufficient evidence of homophily or heterophily.

Employing multinomial logistic regressions, I also examine the probability of membership in the various local Moran clusters for areas in different regions and settlement types, and with differences in the relative density and geographic distance of social ties in other areas. I show that Democratic-leaning cities and suburbs are generally both more likely to be homophilous and heterophilous when compared to rural, Republican-leaning areas, while cities and suburbs of the South are the most likely to be heterophilous. In additional specifications, I examine the probability of cluster membership for areas with differences in the relative density and geographic distance of social ties in other areas. Overall, these findings suggest that while partisan homophily broadly tracks partisan segregation along the urban-rural continuum, voters in urban Democratic-leaning areas are much more likely to interact with distant co-partisans than those in rural Republican-leaning areas.

[5] As defined by McPherson et al. (2001), homophily is 'the principle that a contact between similar people occurs at a higher rate than among dissimilar people'. As the authors discuss, a rich literature suggests that most communities are likely to exhibit homophily to some degree, whether this is defined on the basis of political orientation, age, religion, education, occupation or other social characteristics.

The empirical contribution of this chapter is threefold. First, it offers the first country-wide ecological study of partisan homophily in the United States using data on the social ties rather than the geographic distances between the populations of disparate areas. This addresses a long-standing quagmire in relevant literature whereby the observation of partisan residential segregation is typically equated with that of partisan homophily. As such, the focus of this chapter on actual social ties allows for a more accurate and nuanced assessment of the extent in which different places in the United States are politically homophilous. Second, this chapter also contributes to a growing literature examining the electoral geography of the United States at fine spatial resolutions. Namely, it complements a small number of studies of voting patterns at the below-county level (e.g. Johnston et al., 2020; Kinsella et al., 2021, 2015; Myers, 2013; Rohla et al., 2018), offering a deeper understanding of local political context in different types of settlement along the urban-rural continuum. Lastly, this chapter is linked to conceptual work on the uncertain geographic context problem in spatial analysis (Kwan, 2012; Fowler et al., 2020), shedding light on some of its implications for the geographic study of electoral outcomes in the United States. That is, given that social interactions with co-partisans are a potentially important contextual influence on political preferences (Huckfeldt and Sprague, 1995), the empirical findings are informative with respect to differences in the geographic distances over which this is likely to be exerted on voters in different kinds of localities.

3.2 Data and Methods

3.2.1 Social Connectedness Index

I observe the relative density of social ties between ZCTAs across the United States using the Social Connectedness Index (SCI) by Bailey et al. (2020).[6] This is based on all active users of Facebook—the popular online social networking service—as of October 2021, and is computed for ZCTA-pairs ij as follows:[7]

$$SCI_{ij} = SCI_{ji} = \phi \frac{FB\_Connections_{ij}}{FB\_Users_i \times FB\_Users_j} \tag{3.1}$$

The SCI is thus the total number of Facebook friendships between users in i and j over the product of users in each area, scaled by a factor ϕ. The latter is applied for privacy purposes and ensures that values of the index will range from 1 to 1,000,000.

[6]The SCI is publicly available and distributed via the Humanitarian Data Exchange of the United Nations Office for the Coordination of Humanitarian Affairs.

[7]A Facebook user is deemed active when they have used the service in the previous 30 days.

Consequently, higher values of the index denote a higher density of friendships between the respective populations of active Facebook users, by a factor that is equal to the ratio of the higher value over the lower value. Note that the SCI is available for 22,537 out of the total 33,642 ZCTAs spanning the United States as the index is suppressed for areas where there is a very low number of users. These areas are thus necessarily excluded from the analysis in this chapter. As shown on Appendix Figure 3.A.1, these are mainly concentrated in sparsely populated regions of the Midwest and the West, and account for under 2 per cent of the total population.

As discussed by Bailey et al. (2020), there is growing evidence that Facebook friendships closely resemble real-world social ties among Americans. Indeed, surveys suggest that almost 70 per cent of adults in the United States were Facebook users as of 2021 (Vogels and Anderson, 2021) and that the vast majority of users who are friends on Facebook have met in-person (Duggan et al., 2015). What is more, there is evidence to suggest that there is little variation in Facebook usage rates by partisanship, settlement type, and demographic characteristics (Vogels and Anderson, 2021; Vogels et al., 2021). The SCI is thus plausibly a sound proxy for the density of social ties between localities.

In expressing the neighbourhood relations between ZCTA-pairs, I first construct a social neighbour matrix $\mathbf{A}$ with elements a_{ij} as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } SCI_{ij} > 1 \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

This implies that I treat ZCTA-pairs with an SCI value of 1 as not being socially connected, or not being neighbours in social space, whereas I deem ZCTA-pairs with SCI values above 1 as social neighbours. It further implies that I discard the social ties of each area with itself to focus on its relations with others, resulting in a zero-diagonal matrix. Recall that as SCI values are scaled to range between 1 and 1,000,000,000, ZCTA-pairs with a value of 1 have the lowest pairwise density of social ties observed across the United States. While there is no available information on the actual density of social ties represented by a value of 1, the within-ZCTA distribution of SCI values is plausibly consistent with its interpretation as near-zero density. On average, an SCI value of 1 will characterise the social ties of an area with more than half of all others, with the next lower value being larger by a factor of 387, indicating a sharp decline to the global minimum. This empirical pattern echoes studies of personal social networks in large communities (e.g Fischer, 1982), suggesting that individuals tend to have a few first-degree social ties but are not directly tied to most others.

I combine information on the density of social ties and social neighbourhood relations by constructing a spatial weights matrix $\mathbf{W}$ with elements w_{ij} as follows:

$$w_{ij} = \begin{cases} SCI_{ij} & \text{if } a_{ij} = 1 \text{ and } SCI_{ij} \text{ is in } \min_i(\sum_j a_{ij}) \text{ highest in row} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

Intuitively, the above implies that I consider ZCTA-pairs that are social neighbours to be more proximate in social space when they have a higher density of social ties. It also implies that I only consider the social ties of each area with its 1,347 most connected social neighbours: the minimum number of neighbours across all areas. The latter approach has two advantages. First, by restricting all ZCTAs to have the same number of social neighbours, it improves comparability of social relations across areas. The choice of the minimum number of neighbours across all areas as the cut-off ensures that this is achieved while retaining the maximum amount of information on the social ties among them.[8] Second, it results in a sparse spatial weights matrix with a sparsity of 0.06: a desirable property in the computation of local Moran index, which is the preferred measure of partisan homophily discussed later in this section.

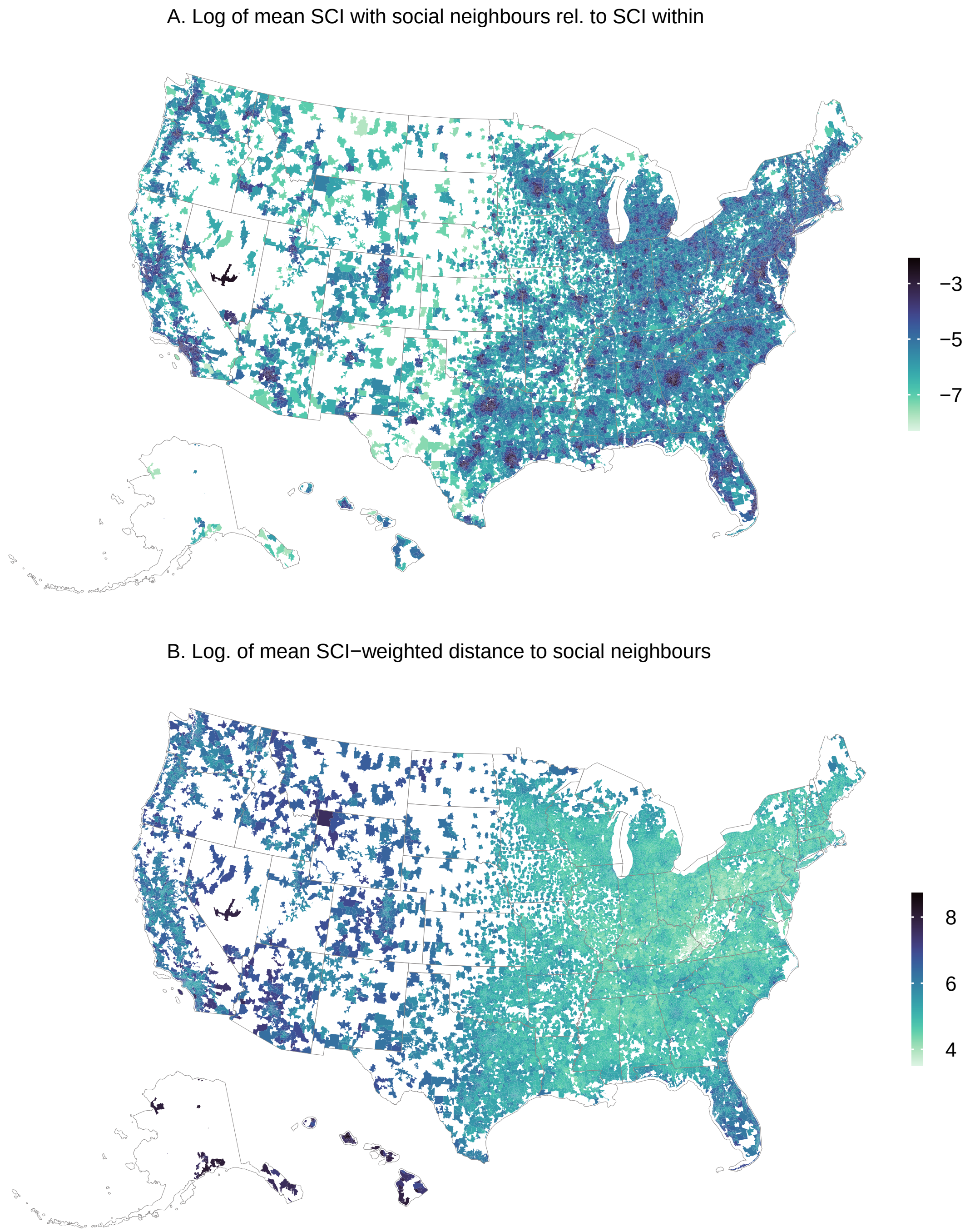
Panel A of Figure 3.1 maps the spatial variation in the logged mean of social connectedness to social neighbours relative to social connectedness within ZCTAs. It is shown that, on average, the social ties of ZCTAs which are closer to population centres are more likely to be located in their social neighbours. Panel B also maps variation in the logged SCI-weighted geographic distance to social neighbours. This further suggests that, on average, the social ties of more urbanised areas are located farther away, as are those of areas in remote and coastal regions. Taken together, these patterns appear consistent with the 'community liberated' argument in urban sociology (e.g. see Wellman, 1979), which contends that urbanites maintain more sparsely knit, spatially dispersed networks relative to the inhabitants of rural areas.

3.2.2 Voting Data

Given that official election results are not published for ZCTAs, which is the level at which the data on social ties is available, I obtain 2020 US presidential election results at the precinct level. Namely, I use the dataset compiled by the Voting and Election Science Team (2020) of the University of Florida and Wichita State University, which includes information on precinct boundaries.

[8] Assuming that Facebook usage rates are similar across ZCTAs, on average, an estimated 74 per cent of the social ties of an area in all others lie in its 1,347 social neighbours.

Figure 3.1: Social Connectedness and Geographic Distance, by ZCTA



Notes: SCI refers to the Social Connectedness Index by Bailey et al. (2020). The social neighbours of a ZCTA are the 1,347 most socially connected ZCTAs to it as measured by the SCI.

While most precincts fall entirely within a single ZCTA, nearly half of them do not. In order to harmonise the data on voting and social ties, I aggregate the former to the ZCTA level in three steps. First, I retrieve the spatial intersections between precinct and ZCTA polygons.[9] I then compute the percentage of the population of each ZCTA that is in each intersection using the High Resolution Settlement Layer (HRSL): a detailed, publicly available, 30-metre population grid produced by the Facebook Connectivity Lab and the Centre for International Earth Science Information Network (CIESIN).[10] Finally, I use the derived population weights to apportion the precinct-level votes for each presidential candidate to each ZCTA.

In principle, to the extent that electoral support for any given presidential candidate is unevenly spatially distributed within any precinct spanning more than a single ZCTA, using population counts to apportion votes from such a precinct to intersecting ZCTAs can give rise to measurement error. However, other than the fact that fewer than half of precincts span multiple ZCTAs, there are two further points assuaging this concern. First, even within these precincts, the population overwhelmingly resides in a single ZCTA: on average, 78 per cent of the population of a precinct spanning ZCTAs falls within a single ZCTA. These precincts are also fairly homogeneous in terms of voting behaviour with an average of almost 70 per cent of the local vote being cast in favour of the leading candidate. Taken together, these points suggest that any measurement error arising from within-precinct variation in voting behaviour across ZCTA boundaries is unlikely to be substantial in the vast majority of cases.

I measure the partisan leaning of each area by calculating the Democratic two-party vote share in the 2020 US presidential election as follows:

$$r_i(D) = \frac{v_i(D)}{v_i(D) + v_i(R)} \tag{3.4}$$

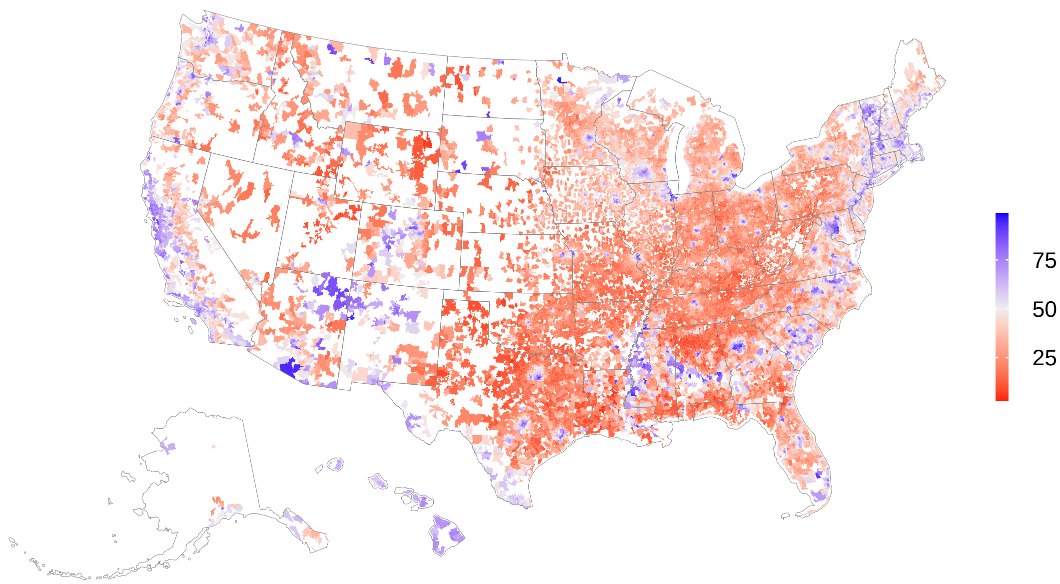
Here, $r_i(D)$ is the Democratic two-party vote share in ZCTA i, with $v_i(D)$ and $v_i(R)$ respectively corresponding to votes for the Democratic presidential candidate, Joe Biden, and the Republican presidential candidate, Donald Trump. As such, high values of $r_i(D)$ indicate a strong Democratic leaning, with low values indicating a strong Republican leaning. Figure 3.2 maps the measure, reflecting the well-documented pattern of strong Democratic support in population centres. Note that $r_i(D)$ can also be thought of as a measure of partisan homogeneity, with values closer to 0.5 indicating heterogeneity in voting behaviour.

[9] I use the 'tigris' R package by Walker and Rudis (2023) to obtain US Census Bureau shapefiles.

[10] The HRSL is distributed via the Humanitarian Data Exchange of the United Nations Office for the Coordination of Humanitarian Affairs.

While the two-party vote share is often preferred as a straightforward measure of both partisan leaning and partisan homogeneity, it is important to note that it disregards votes for presidential candidates that are not affiliated with any the two major political parties. In principle, this measure could then be misleading in cases where a high share of local votes are cast in favour of such 'independent' candidates. However, such cases are exceedingly rare in practice: on average, in the 2020 US presidential election, 1.7 per cent of votes within a ZCTA were cast in favour of independent candidates, with over 95 per cent of ZCTAs seeing independent vote shares of below 5 per cent.[11] Appendix Figure 3.A.2 maps independent vote shares across ZCTAs, showing that these are fairly even distributed in geographic space.

Figure 3.2: Democratic Two-Party Vote Share, 2020 US Pres. Election, by ZCTA



Notes: ZCTA-level vote shares are obtained by first apportioning precinct-level election results using the High Resolution Settlement Layer (HRSL).

3.2.3 Demographic Data and Settlement Classification

I use 5-year estimates from the American Community Survey (ACS) produced by the US Census Bureau to observe a series of demographic characteristics of ZCTA populations as of 2020.[12] I further obtain information on the type of settlement best describing each ZCTA by using the Locale Assignments File produced by the National Centre for Education Statistics (NCES) (Geverdt, 2019).

[11] I use the term 'independent' to refer to both unaffiliated presidential candidates and candidates affiliated with parties other than the Democratic and Republican parties.

[12] I use the 'tidycensus' R package by Walker and Herman (2023) to obtain demographic data from the US Census Bureau application programming interface (API).

The NCES framework offers a rarely detailed settlement classification at the ZCTA level. This is achieved by deriving classes based on both the extent in which ZCTA populations reside in different broad types of settlement as these are defined by official bodies—rural areas, urban clusters, urbanised areas, and principal cities—as well as their geographic distance from these types of settlement. A detailed description of the official definitions of settlements used in deriving the NCES classification is given by Geverdt (2018). Rural areas, urban clusters, and urbanised areas are defined by the US Census Bureau and are constructed from Census tracts and blocks. Urbanised areas are those containing 50,000 or more people, urban clusters are those containing more than 2,500 and less than 50,000 people, while rural areas are those outside urbanised areas and urban clusters.[13] Similarly, principal cities are defined by the US Office of Management and Budget and correspond to incorporated and Census-designated places within core based statistical areas. While there are several alternative qualifying criteria, these generally require that a place has a resident population of at least 10,000 people and a substantial number of workers from other places.

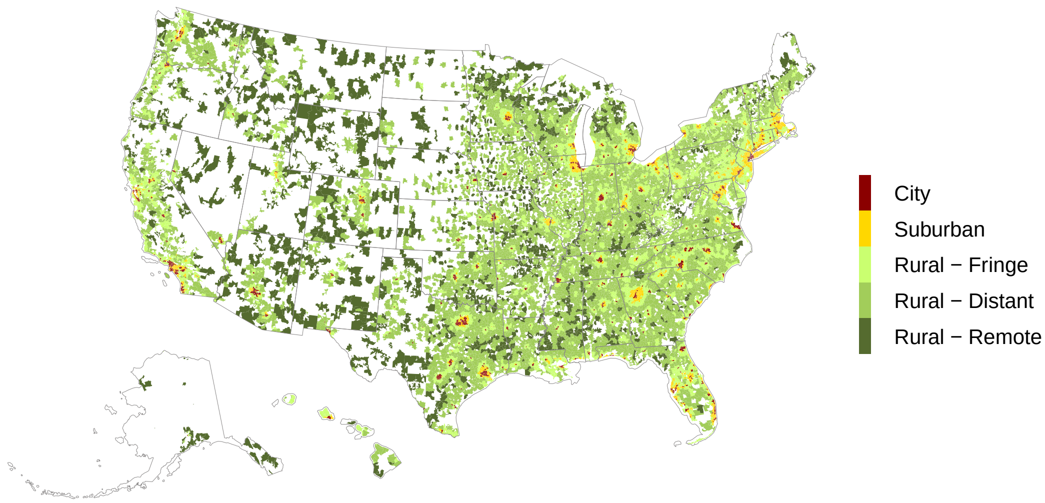
At the lowest level, the NCES classification is comprised of 14 classes. For brevity, I use an aggregated version of the classification comprised of the following 5 classes, which are also mapped on Figure 3.3:

- *City*: Territory inside an urbanised area and inside a principal city.
- *Suburb*: Territory outside a principal city and inside an urbanised area.
- *Rural – Fringe*: Census-defined rural territory that is less than or equal to 5 miles from an urbanised area, or rural territory that is less than or equal to 2.5 miles from an urban cluster, or territory inside an urban cluster.[14]
- *Rural – Distant*: Census-defined rural territory that is more than 5 miles but less than or equal to 25 miles from an urbanised area, or is more than 2.5 miles but less than or equal to 10 miles from an urban cluster.
- *Rural – Remote*: Census-defined rural territory that is more than 25 miles from an urbanised Area and also more than 10 miles from an urban cluster.

[13] In 2020, the US Census Bureau defined both urbanised areas and urban clusters as 'urban areas'. The NCES classification framework is thus based on definitions as of the 2010 Census.

[14] Territories inside urban clusters are classified as 'towns' by NCES. As very few ZCTAs fall under this class, I merge these with the related 'rural fringe' class. The latter is preferred over suburbs as urban clusters are generally much more distant relative to principal cities.

Figure 3.3: ZCTA Settlement Type Classification



Notes: Settlement types are derived from the NCES Locale Assignments File (Geverdt, 2019)

3.2.4 Identifying Politically Homophilous Clusters

In gauging whether there is evidence of partisan homophily or heterophily in each ZCTA, I use the local Moran index (Anselin, 1995): a popular local indicator of spatial association commonly employed in studies of partisan segregation (e.g. Darmofal and Strickler, 2019; Kinsella et al., 2021, 2015). However, unlike these studies, this chapter is primarily concerned with the arrangement of voters in social space (homophily) rather than geographic space (segregation). This mandates a different approach in operationalising the interactions between disparate locations. Namely, I compute the local Moran index for the Democratic two-party vote share in the 2020 US presidential election in each ZCTA as follows:[15]

$$I_i = \frac{z_i \sum_j \hat{w}_{ij} z_j}{\sum_i z_i^2} \tag{3.5}$$

Here, i and j index disparate ZCTAs, $\hat{w}_{ij}$ denotes elements of the row-standardised version of the spatial weights matrix $\mathbf{W}$ defined in (3.1), and z_i represents the standardised Democratic two-party vote share in i. What makes I_i interpretable as a local measure of partisan homophily rather than partisan segregation is the way $\mathbf{W}$ is constructed. That is, rather than being populated on the basis of geographic distance between ZCTA-pairs, its elements are based on the pairwise density of social ties.

[15] I use the 'spdep' R package by Bivand (2022) to calculate local Moran statistics.

In standardising vote shares, I employ the Empirical Bayes (EB) approach proposed by Assunção and Reis (1999) as follows:[16]

$$z_i = \frac{r_i(D) - \beta}{\sqrt{\alpha + (\beta \div T_i)}} \tag{3.6}$$

In the above, $r_i(D)$ is the Democratic two-party vote share in area i, T_i is the total number of two-party votes cast in i, with β and α respectively denoting the empirically estimated mean and variance of a prior Gamma distribution.[17] The EB approach reflects the intuition that, insofar as the number of votes cast vary considerably across ZCTAs, the Democratic two-party vote share may be a relatively less precise estimate of the underlying partisan leaning in ZCTAs were few votes were cast. As this would violate the assumption of constant variance across locations in the calculation of the local Moran index, the standardisation of shares as in (3.6) serves to account for such cases.

In effect, by computing the local Moran index as in (3.5), I conceptualise partisan homophily as positive autocorrelation in voting behaviour in social space. As is the case with operationalisations of the index as a measure of autocorrelation in geographic space, the main utility of this approach is that it facilitates significance testing against the null of no autocorrelation. As proposed by Anselin (1995), I do so by using a conditional permutation approach. Note that this departs from conventional notions of statistical significance in that pseudo p-values are obtained from simulated, empirical distributions of the index at each location.[18] Note further that the probability of false positives increases with the number of locations (de Castro and Singer, 2006), necessitating the downward adjustment of any pre-selected significance cut-off. I thus derive a critical p-value using the false discovery rate (FDR) criterion proposed by Benjamini and Hochberg (1995), setting the significance cut-off parameter to 0.05. With 22,537 locations, critical p-values derived in this way can be as low as $\frac{0.05}{22537} = 2.2 \times 10^{-6}$. In order to allow for the detection of pseudo p-values below this threshold, I construct the simulated local distributions of the local Moran index using 499,999 permutations.

After identifying locations where values of the local Moran index are significant, I assign them to politically homophilous and politically heterophilous clusters defined using the popular local Moran scatterplot approach discussed by Anselin (1995, 1996).

[16] I use the 'rgeoda' R package by Li and Anselin (2023) to calculate EB shares.

[17] More specifically, $\beta = \sum_i v_i(D) \div \sum_i T_i$ and $\alpha = [\sum_i T_i(r_i(D) - \beta)^2] \div T - \beta \div (T \div n)$ where $v_i(D)$ is the number of democratic votes in i and n is the number of observations.

[18] I describe values of the local Moran index as 'significant' when these correspond to pseudo p-values beyond a specified critical value along their local simulated distributions. As suggested by Efron and Hastie (2016), these values may be instead thought of as 'interesting'.

Locations with above and below-mean Democratic two-party vote shares $r_i(D)$ that respectively have social neighbours with above and below-mean SCI-weighted Democratic two-party vote shares $\sum_j \hat{w}_{ij} r_i(D)$ are assigned to two respective homophilous clusters—'High-High' and 'Low-Low'—and locations where the opposite relations hold are respectively assigned to two heterophilous clusters—'High-Low' and 'Low-High'. As such, the former group areas of similar partisan leaning where there is evidence of positive autocorrelation in voting behaviour in social space, and the latter represent areas where there is evidence of negative autocorrelation.

3.2.5 Examining the Determinants of Cluster Membership

To examine the ZCTA-level features associated with the various politically homophilous and heterophilous clusters, I employ a set of multinomial logit specifications. I model the probability that ZCTA i is assigned to a cluster c as:

$$P_{ic} = \frac{e^{x_i' \beta_c}}{\sum_{k=1}^{m} e^{x_{ic}' \beta_{kc}}} \qquad \text{for} \quad c = 1, \ldots, m \tag{3.7}$$

Here, x_i' is a vector of covariates and c indexes one of five categories: the High-High, Low-Low, High-Low, and Low-High Moran clusters as well as a 'not significant' category for cases where there is insufficient evidence of homophily or heterophily. Typically, the odds ratio e^{β_c} capturing the marginal impact of each variable on the relative probability of belonging to category c as opposed to some other category c' can be obtained by normalising the coefficient $\beta_{c'}$ to zero. However, this approach can lead to difficulties in interpreting effects across categories and can quickly become unwieldy in more complex specifications including interactions. As such, I instead report the fitted probabilities $\hat{P}_{ic}$ and their respective 95% confidence intervals using the approach of Fox and Andersen (2006),[19] which can be straightforwardly interpreted as the probability that i is assigned to c for particular values of the covariates.

3.3 Results

Figure 3.4 shows the spatial distribution of local Moran clusters across ZCTAs. The High-High and High-Low clusters represent areas that had a relatively higher Democratic two-party vote share in the 2020 US presidential election and were, respectively, relatively more and less likely to be socially connected to areas with similar shares.

[19] I use the 'effects' R package by Fox and Weisberg (2019) to obtain the fitted probabilities and their confidence intervals.

Similarly, the Low-Low and Low-High clusters are comprised of areas with a relatively more Republican leaning in voting behaviour that were respectively, and in relative terms, more politically homophilous and heterophilous. Also mapped are the areas for which there is relatively limited evidence of partisan homophily or heterophily in terms of aggregate social ties in other areas. Figure 3.5 further plots the share of all ZCTAs and the share of the total population that is occupied by each cluster and within each US region—the South, Midwest, Northeast, and West—along with their corresponding 95 per cent confidence intervals.[20]

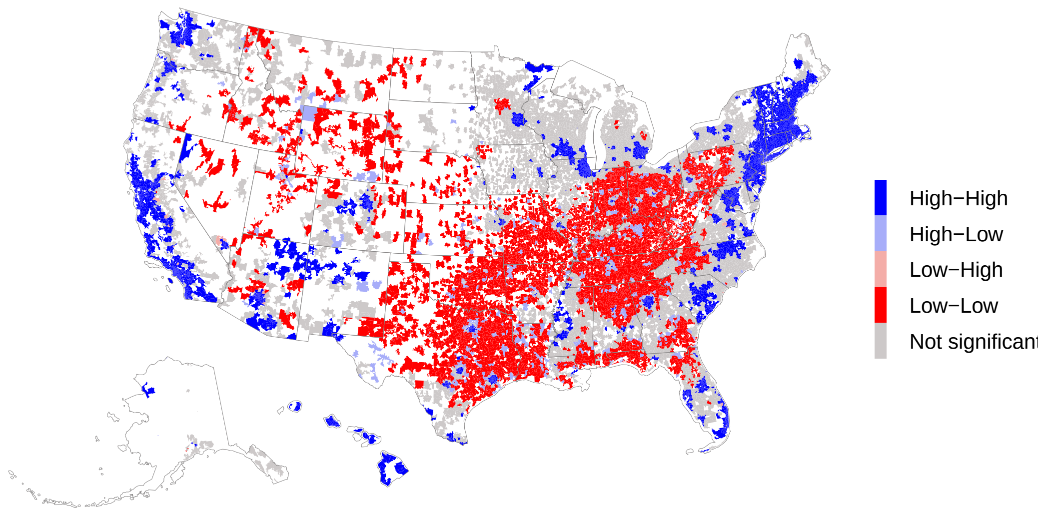
It is shown that there is evidence of either partisan homophily or heterophily in 67 per cent of all ZCTAs, representing a 77 share of the total population. Areas where the evidence is limited are more likely to be found in the Midwest than elsewhere, constituting 47 per cent of all areas in the region. This compares against a respective 30, 29, and 26 per cent in the West, the South, and the Northeast, suggesting a greater incidence of partisan homophily or heterophily across areas in these regions. Notably, there is substantial spatial variation within regions: for instance, one is less likely to come across evidently homophilous or heterophilous areas in settlements around upstate New York in the Northeast, along the coastal plains in the South, and away from the coast in the West than elsewhere in each respective region.

Despite not being evident in a substantial share of areas, partisan homophily appears to be the norm across ZCTAs in the United States: homophilous areas account for roughly 70 per cent of the total population and all ZCTAs, with heterophilous areas accounting for 4.5 per cent of ZCTAs and 7 per cent of the population. Of the ZCTAs identified as politically homophilous or heterophilous, 57 per cent are Democratic-leaning, representing roughly 83 per cent of the population among these areas. The higher population share of Democratic-leaning areas relative to their share of ZCTAs is in line with the well-documented spatial concentration of Democrats around population centres, and the greater geographic dispersion of Republicans in rural areas.

Partisan heterophily is almost exclusively observed among Democratic-leaning areas, with just 0.24 per cent of all ZCTAs identified as Republican-leaning and heterophilous, representing 0.3 per cent of the total population. Interestingly, the share of the population in heterophilous Democratic-leaning areas relative to that in homophilous areas of the same leaning matches the respective share of ZCTAs at roughly 12 per cent. This suggests that there are no major differences in the spatial concentration of voters in homophilous and heterophilous Democratic-leaning areas. Further, looking at the regional distribution of heterophilous Democratic-leaning areas, it is shown that these are most common in the South and the Midwest. These are also the regions within which Democratic-leaning areas are the least spatially dispersed.

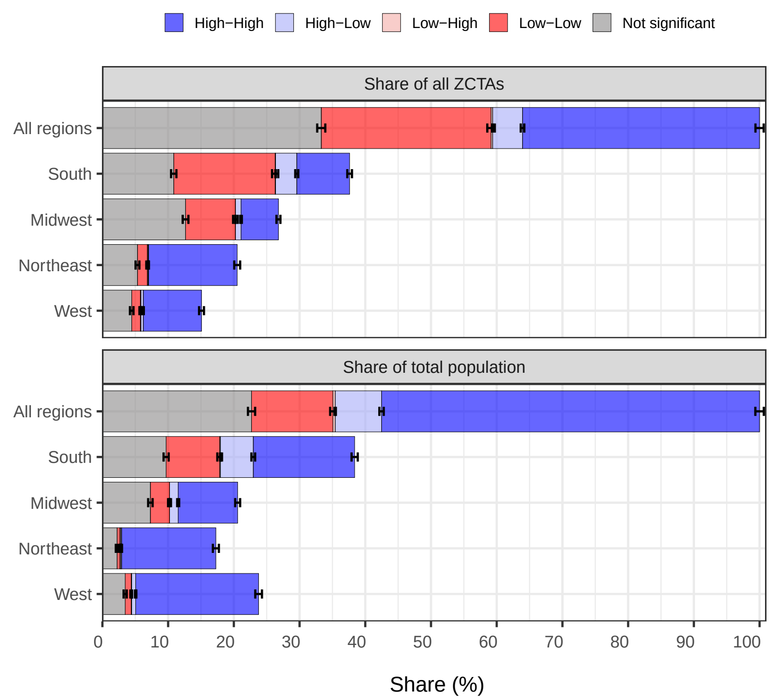
[20] Appendix Table 3.A.1 presents these results in table format.

Figure 3.4: Local Moran Clusters, by ZCTA



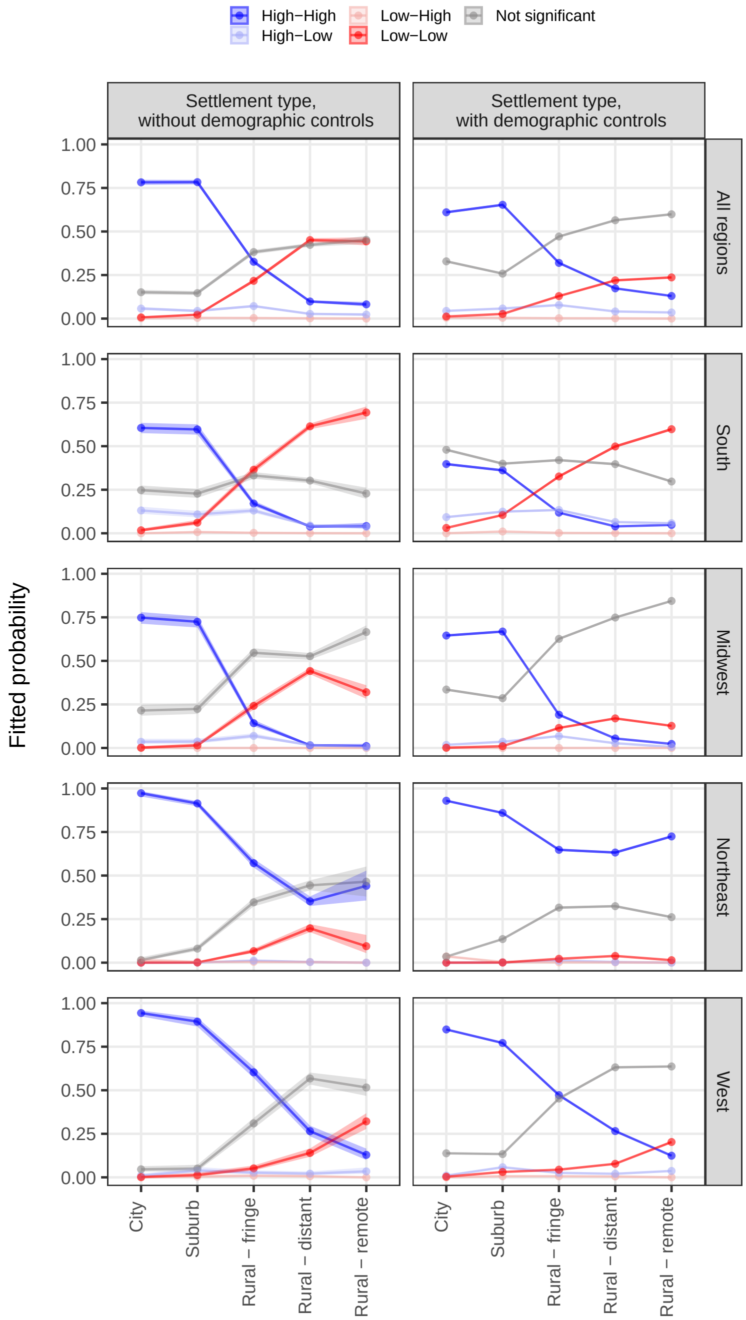
Notes: Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the 1,347 most connected areas as measured by the Social Connectedness Index (Bailey et al., 2020).

Figure 3.5: ZCTA and Population Shares, by Local Moran Cluster and Region



Notes: Black whiskers denote 95% confidence intervals. Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the 1,347 most connected areas as measured by the Social Connectedness Index (Bailey et al., 2020).

Figure 3.6: Fitted Probability of Local Moran Cluster, by Region and Settlement



Notes: Each column corresponds to a multinomial logit regression with cluster membership as the outcome. Points and shaded areas respectively correspond to fitted probabilities and 95% confidence intervals obtained using the approach of Fox and Andersen (2006). Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the 1,347 most connected areas as measured by the Social Connectedness Index (Bailey et al., 2020).

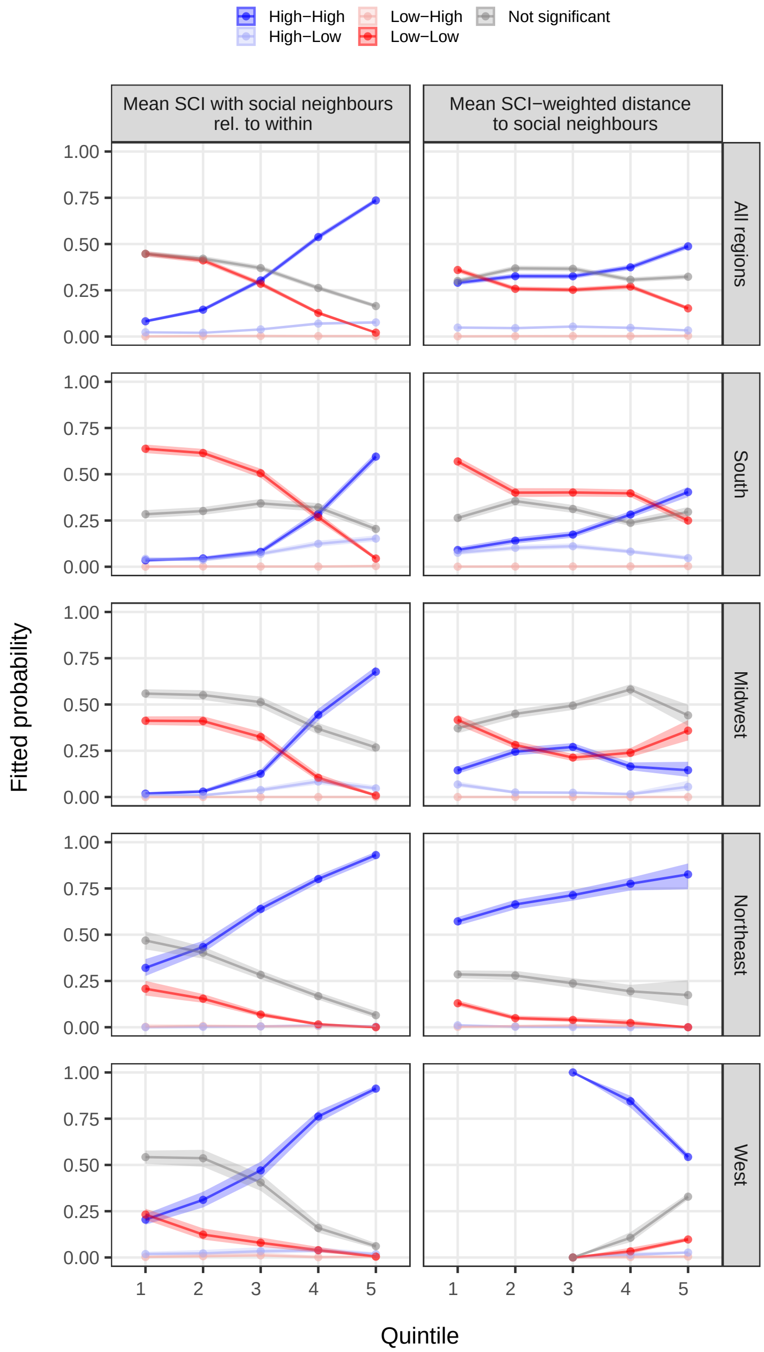
How does partisan homophily relate to urbanisation across regions? The first column of Figure 3.6 displays fitted probabilities from multinomial logit specifications regressing local Moran cluster membership on the interaction of US region and settlement type as per the NCES classification.[21,22] On average, more urbanised settlements have a higher probability of belonging to a homophilous Democratic-leaning cluster and a lower probability of belonging to a homophilous Republican-leaning cluster. With the exception of areas in the South, the relative probability of being in a politically homophilous as opposed to neither a homophilous nor a heterophilous cluster also tends to decrease with urbanisation. For instance, looking at the edges of the urban-rural continuum, a ZCTA in a city is 2.5 times more likely to be homophilous and Democratic-leaning than a remote rural area is likely to be homophilous and Republican-leaning. As such, when compared to rural areas across the continuum, cities and suburbs are generally more likely to be politically homophilous when it comes to their aggregate social ties in other areas. Cities, suburbs, and rural fringes in the South are also the most likely areas to be Democratic-leaning and heterophilous with a probability of slightly over 10 per cent: almost triple than the probability of an area of either partisan leaning being heterophilous elsewhere.

To what extent is the association between partisan homophily and urbanisation across regions explained by demographic characteristics? The second column of Figure 3.6 displays fitted probabilities from multinomial logit specifications mirroring those of the first column while additionally controlling for ZCTA-level median age, median household income, white population share, population share in owner-occupied housing, and the adult population share with a university degree. It is shown that, on average, a similar broad pattern holds: net of demographic characteristics, as urbanisation increases an area is more likely to be homophilous and Democratic-leaning and less likely to be homophilous and republic Republican leaning, with more urbanised settlements being more likely to be homophilous. However, compared to specifications without demographic controls, the probability of being in neither a politically homophilous nor a heterophilous cluster gains an average of 7 percentage points across regions and settlements, suggesting that partisan homophily is at least partly explained by spatial variation in the demographic composition of different settlement types. This is most salient in cities and suburbs of the South, where the gain in probability respectively stands at 23 and 22 percentage points. Rural areas of the Northeast are outliers in that the exact opposite holds; that is, when accounting for their demographic composition, the probability of these areas being politically homophilous increases by roughly 20 percentage points.

[21] I match each ZCTA to a parent county, state, and region using the 'Geocorr' application of the Missouri Census Data Centre, which provides correspondence tables based on the majority population share within each geography as of the 2020 Census.

[22] Appendix Table 3.A.2 presents these results in table format.

Figure 3.7: Fitted Probability of Local Moran Cluster, by Network Features



Notes: Each column corresponds to a multinomial logit regression with cluster membership as the outcome. Points and shaded areas respectively correspond to fitted probabilities and 95% confidence intervals obtained using the approach of Fox and Andersen (2006). Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the social neighbours of each area: the 1,347 most connected areas as measured by the Social Connectedness Index (SCI) (Bailey et al., 2020).

As discussed in the previous section, there is substantial variation in the relative density of social ties that the populations of different areas maintain elsewhere and the geographic distance of these social ties. Figure 3.7 presents fitted probabilities from multinomial logit specifications regressing local Moran cluster membership respectively on quintiles of the mean density of social ties between each ZCTA and its social neighbours over the density of social ties within the area, and on the mean geographic distance of the social neighbours of each ZCTA weighted by the density of social ties with them.[23] Looking at the first column, it is shown that, on average, areas with a higher relative density of social ties in other areas are substantially more likely to be in homophilous Democratic-leaning clusters. A politically homophilous area in the first quintile of relative density, in which the density of social ties within the median area is 500 times higher than the mean density of its ties in other areas, is over 5 times more likely to be Republican-leaning. In contrast, a politically homophilous area in the fifth quintile, in which the within-density of social ties of the median area is 41 times higher than the mean density of its ties elsewhere, is 35 times more likely to be Democratic-leaning. Notably, the disproportionate increase in the relative probability of homophilous Democratic-leaning clusters in higher quintiles is in line with a decrease in the probability of an area being identified as neither homophilous nor heterophilous. In other words, areas with a higher relative density of social ties in others are both more likely to be politically homophilous and Democratic-leaning.

Looking at the second column of Figure 3.7, it is shown that there is also an overall positive relationship between the probability of a political homophilous area being Democratic-leaning and the geographic distance of its social ties in other areas. Most notably, when moving from the fourth to the fifth quintile, in which the respective mean geographic distance of the social ties of the median area stands at 212 and 479 kilometres, the relative probability of a politically homophilous area being Democratic-leaning doubles to 3.2. Though there is also notable regional heterogeneity: while similar relationships hold in the South and the Northeast, in the Midwest the positive relationship turns negative after the third quintile, while being negative throughout in the West. That is, within these regions, when compared with areas with social ties lying at different geographic distances, the areas with the most geographically distant social ties are less likely to be homophilous and Democratic-leaning. These lower relative probabilities are primarily compensated by higher a relative probability of areas being homophilous and Republican-leaning in the Midwest, and a lower relative probability of being politically homophilous in the West. Finally, the areas with the highest relative density of social ties are more likely to be heterophilous and Democratic-leaning relative to others, albeit with a low absolute probability of 8 per cent. Though the same does not hold, on average, for areas with the most distant ties.

[23] Appendix Table 3.A.3 presents these results in table format.

3.4 Discussion and Conclusion

When examined in conjunction with the findings of studies on partisan segregation, the empirical results of this chapter suggest that the geography of partisan homophily in the United States broadly tracks that of the latter. Perhaps the most direct comparison can be made with the findings of Darmofal and Strickler (2019): whereas in this chapter partisan homophily is measured as autocorrelation in social space in voting behaviour in the 2020 US presidential election at the ZCTA level, the authors measure partisan segregation as autocorrelation in geographic space in voting behaviour in the 2016 US presidential election at the county level. [24] Broadly similar regional patterns emerge; both politically homophilous and segregated areas of a Republican leaning are most likely to be found in, around, and between Texas and Kentucky in the South, while their Democratic-leaning counterparts are most common along the Northeastern and Western coasts. A fairly close correspondence also emerges with the findings of Brown and Enos (2021), who measure individual-level exposure to Republicans in most American voters' residential environments using voter registration data as of June 2018.[25] Just as exposure to registered Republicans quickly lowers as one approaches urban cores, the probability of an area being Democratic-leaning in its voting behaviour in 2020 as well as having denser social ties in other areas that voted similarly becomes substantially higher. What is more, just as Democrats are, on average, less likely to be exposed to Republicans in their residential environment, politically homophilous areas are more likely to be Democratic-leaning rather than Republican-leaning.

Overall, the evidence presented in this chapter supports a long-held expectation regarding the electoral geography of the United States, which has nevertheless remained largely untested at scale. That is, by and large, segregation by partisanship in geographic space does go hand in hand with segregation by partisanship in social space. With the majority of ZCTAs across the United States identified as exhibiting a degree of partisan homophily with respect to their aggregate social ties in other areas, it thus seems likely that the prevalence of partisan homophily is a relevant factor in the rise of affective and ideological polarisation in the country in recent years. As is the case with spatial variation in voting behaviour (Scala and Johnson, 2017), the position of voters along the urban-rural continuum also appears to play an important role in spatial variation in partisan homophily that goes beyond differences in the demographic composition of local populations.

[24]For Darmofal and Strickler (2019), two counties are considered as geographic neighbours when they are contiguous.

[25]In calculating individual-level exposure to Republicans, Brown and Enos (2021) consider the thousand nearest neighbours in terms of residence.

Crucially, this chapter also highlights a relatively neglected prospect in previous work: while partisan homophily tends to coincide with partisan segregation, there is substantial variation in the geographic range over the latter might occur. In other words, one may have some confidence that a voter residing close to co-partisans is also likely to have co-partisan peers in their social network, though the extent in which these peers are also located nearby is likely to vary depending on the kind of locality in question. When compared to voters in areas with a strong Republican leaning, which are most likely to be rural, voters in areas with a strong Democratic leaning, which are most likely to be cities and suburbs, are much more likely to be socially connected to co-partisans in distant areas. Indeed, these areas are often located several hundreds of kilometres away. As discussed by Baybeck and Huckfeldt (2002a,b), while social ties with non-local co-partisans may not expose voters to politically divergent contexts, they may act as channels via which information about non-local socioeconomic conditions feeds into local political preferences. Such spillovers are thus likely to be particularly relevant in understanding the evolving political preferences of urban, strongly Democratic-leaning areas, particularly when seen in light of the political effects of exposure to heterogeneous urban environments (e.g. Enos, 2017) and the importance of co-partisan ties in preference formation (e.g. Huckfeldt and Sprague, 1995). Importantly, this may also have implications for the plausibility of previously proposed policy responses to political polarisation, such as the rollout of compulsory voting in blue states (Rodden, 2015). That is, to the extent that compulsory voting is both relatively uncontroversial in Democratic-leaning jurisdictions and creates incentives against polarising campaigns, its rollout by any given Democratic state legislature could in principle also facilitate the reduction polarisation in other states.

The findings also suggest that the 2020 US presidential election saw differences in the regional distribution of politically homophilous areas and spatial heterogeneity in the relationship between partisan homophily on the one hand and settlement type and network features on the other. A substantial share of areas representing as much as a fifth of the total population were identified as neither homophilous nor heterophilous, with the majority being distributed in the South and the Midwest. Also most common in these regions were the few heterophilous areas that were more likely to be socially connected to politically dissimilar others, primarily represented by Democratic-leaning areas along urban fringes with relatively geographically proximate social ties. The South was further identified as somewhat of a regional outlier in that, in juxtaposition to other regions, less urbanised areas were more likely to be identified as politically homophilous. As such, researchers interested in political preferences in particular regions and settlements should be cautious of this heterogeneity.

Examining the geography of partisan homophily on the basis voting behaviour in the 2020 US presidential election aids with the task of maximising the correspondence between the particular dates and geographic units for which partisanship and social ties are observed.

As spatial voting patterns are likely to change over time, it is important to note that developing a better understanding of how this geography evolves will require revisiting its measurement as more relevant data become available. While there are a number of barriers in their collation for all localities, voter registration data could also plausibly improve measurement, given that registered partisan affiliation may offer a more explicit indicator of partisan identity than voting behaviour.

Finally, recall that, in line with data availability, this chapter has examined partisan homophily in terms of the aggregate social ties between ZCTAs. Though partisan homophily is also likely to vary within these localities: a prospect that is arguably reinforced given the evidence on variation in partisan segregation at very fine spatial resolutions (Brown and Enos, 2021). Indeed, the empirical analysis has shown that while there is substantial spatial variation in the degree in which the density of social ties within a ZCTA exceeds the mean density of its social ties in others, the latter is, on average, dwarfed by the former. However, as argued in this chapter, this variation may still play an important role in the formation of political preferences insofar as it represents differences in the flow of information between distant areas. As anticipated by Granovetter (1973, 1983), influence and information often diffuse via 'weak ties' bridging otherwise disconnected groups of individuals.

3.A Appendix

3.A.1 Additional Figures

Figure 3.A.1: Availability of the Social Connectedness Index, by ZCTA

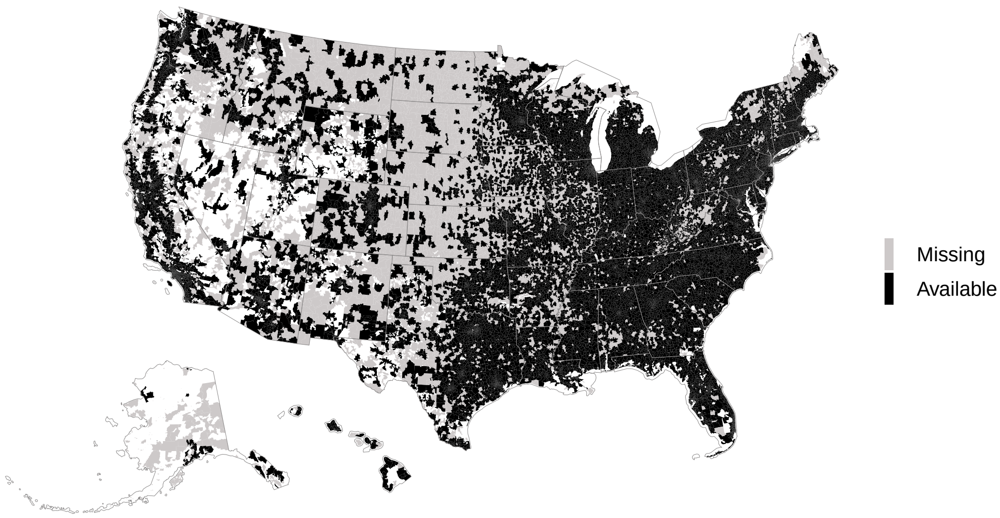
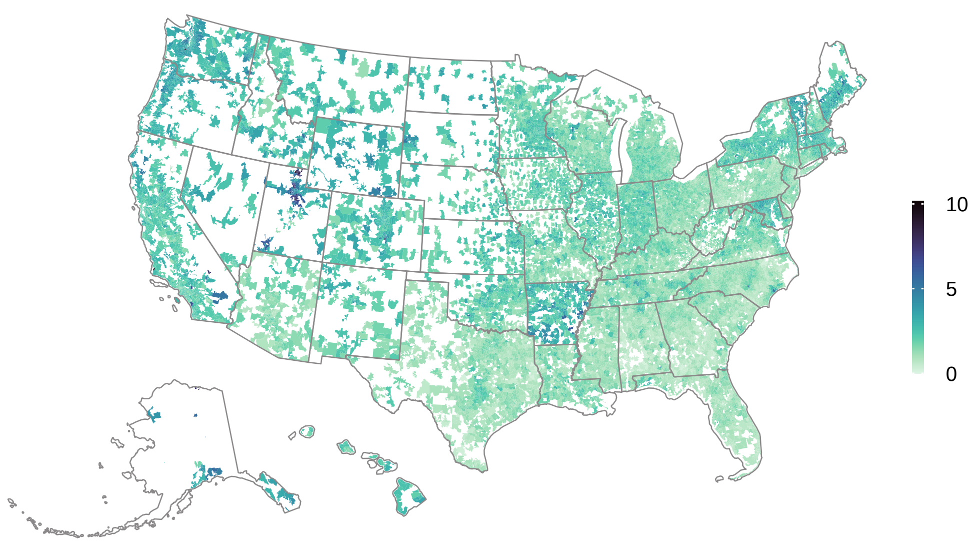


Figure 3.A.2: Independent Vote Share, 2020 Pres. Election, by ZCTA



Notes: The term ‘independent’ refers to candidates that are either unaffiliated or not affiliated with the Democratic or the Republican party.

3.A.2 Additional Tables

Table 3.A.1: ZCTA and Population Shares, by Local Moran Cluster and Region

A. *Share of all ZCTAs (%)*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions	36.07	4.55	0.24	25.84	33.3
	(0.32)	(0.14)	(0.03)	(0.29)	(0.31)
South	8.04	3.22	0.07	15.43	10.85
	(0.18)	(0.12)	(0.02)	(0.24)	(0.21)
Midwest	5.69	0.86	0	7.58	12.65
	(0.15)	(0.06)	(0)	(0.18)	(0.22)
Northeast	13.48	0.1	0.1	1.5	5.34
	(0.23)	(0.02)	(0.02)	(0.08)	(0.15)
West	8.86	0.37	0.07	1.32	4.46
	(0.19)	(0.04)	(0.02)	(0.08)	(0.14)

B. *Share of total population (%)*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions	57.52	7.06	0.33	12.4	22.69
	(0.33)	(0.17)	(0.04)	(0.22)	(0.28)
South	15.43	5.01	0.11	8.15	9.69
	(0.24)	(0.15)	(0.02)	(0.18)	(0.2)
Midwest	9.06	1.31	0	2.91	7.3
	(0.19)	(0.08)	(0)	(0.11)	(0.17)
Northeast	14.33	0.13	0.14	0.43	2.23
	(0.23)	(0.02)	(0.03)	(0.04)	(0.1)
West	18.71	0.6	0.08	0.91	3.46
	(0.26)	(0.05)	(0.02)	(0.06)	(0.12)

Notes: Standard errors are shown in brackets. Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the 1,347 most connected areas as measured by the Social Connectedness Index (Bailey et al., 2020).

Table 3.A.2: Fitted Probability of Local Moran Cluster, by Region and Settlement

A. *Without demographic controls*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions					
City	0.78	0.06	0	0.01	0.15
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Suburb	0.78	0.04	0	0.02	0.15
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - fringe	0.33	0.07	0	0.22	0.38
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - distant	0.1	0.03	0	0.45	0.42
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.08	0.02	0	0.44	0.45
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
South					
City	0.6	0.13	0	0.02	0.25
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Suburb	0.6	0.11	0.01	0.06	0.23
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - fringe	0.17	0.13	0	0.36	0.33
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - distant	0.04	0.04	0	0.61	0.3
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.04	0.04	0	0.69	0.23
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Midwest					
City	0.75	0.04	0	0	0.22
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Suburb	0.72	0.04	0	0.01	0.22
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - fringe	0.14	0.07	0	0.24	0.55
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)

(continued on next page)

Table 3.A.2: (continued)

	High-High	High-Low	Low-High	Low-Low	Not sig.
Midwest					
Rural - distant	0.02	0.02	0	0.44	0.53
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.01	0	0	0.32	0.67
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Northeast					
City	0.97	0	0.01	0	0.01
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Suburb	0.91	0	0	0	0.08
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - fringe	0.57	0.01	0	0.07	0.35
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - distant	0.35	0	0	0.2	0.44
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.44	0	0	0.09	0.46
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
West					
City	0.94	0.01	0	0	0.05
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Suburb	0.89	0.04	0	0.01	0.05
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - fringe	0.6	0.03	0.01	0.05	0.31
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - distant	0.27	0.02	0.01	0.14	0.57
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.13	0.03	0	0.32	0.52
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)

(continued on next page)

Table 3.A.2: (continued)

B. *With demographic controls*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions					
City	0.61	0.04	0.01	0.01	0.33
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Suburb	0.65	0.06	0	0.03	0.26
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - fringe	0.32	0.08	0	0.13	0.47
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - distant	0.17	0.04	0	0.22	0.56
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - remote	0.13	0	0.03	0.24	0.6
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
South					
City	0.4	0.09	0	0.03	0.48
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Suburb	0.36	0.12	0.01	0.1	0.4
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - fringe	0.12	0.13	0	0.33	0.42
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - distant	0.04	0.06	0	0.5	0.4
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - remote	0.05	0.06	0	0.6	0.3
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Midwest					
City	0.65	0.02	0	0	0.34
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Suburb	0.67	0.04	0	0.01	0.29
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - fringe	0.19	0.07	0	0.12	0.63
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

(continued on next page)

Table 3.A.2: (continued)

	High-High	High-Low	Low-High	Low-Low	Not sig.
Midwest					
Rural - distant	0.05	0.03	0	0.17	0.75
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - remote	0.02	0.01	0	0.13	0.84
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Northeast					
City	0.93	0	0.04	0	0.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Suburb	0.86	0	0	0	0.14
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - fringe	0.65	0.01	0	0.02	0.32
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Rural - distant	0.63	0	0	0.04	0.32
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - remote	0.72	0	0	0.01	0.26
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
West					
City	0.85	0.01	0	0	0.14
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Suburb	0.77	0.06	0.01	0.03	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - fringe	0.47	0.03	0.01	0.04	0.45
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - distant	0.27	0.02	0	0.08	0.63
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Rural - remote	0.12	0.04	0	0.2	0.64
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: Panels A and B each correspond to a multinomial logit regression with cluster membership as the outcome. Fitted probabilities and standard errors (in brackets) are obtained using the approach of Fox and Andersen (2006). Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the 1,347 most connected areas as measured by the Social Connectedness Index (Bailey et al., 2020).

Table 3.A.3: Fitted Probability of Local Moran Cluster, by Network Features

A. *Mean SCI with social neighbours rel. to within*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions					
1^{st} quintile	0.08	0.02	0	0.45	0.45
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
2^{nd} quinitile	0.14	0.02	0	0.41	0.42
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
3^{rd} quintile	0.3	0.04	0	0.29	0.37
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
4^{th} quintile	0.54	0.07	0	0.13	0.26
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
5^{th} quintile	0.74	0.08	0	0.02	0.16
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
South					
1^{st} quintile	0.03	0.04	0	0.64	0.28
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
2^{nd} quinitile	0.05	0.04	0	0.61	0.3
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
3^{rd} quintile	0.08	0.07	0	0.51	0.34
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
4^{th} quintile	0.29	0.12	0	0.27	0.32
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
5^{th} quintile	0.6	0.15	0	0.04	0.2
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
Midwest					
1^{st} quintile	0.02	0.01	0	0.41	0.56
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
2^{nd} quintile	0.03	0.01	0	0.41	0.55
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
3^{rd} quintile	0.13	0.04	0	0.32	0.51
	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)

(continued on next page)

Table 3.A.3: (continued)

	High-High	High-Low	Low-High	Low-Low	Not sig.
Midwest					
4^{th} quintile	0.44	0.08	0	0.1	0.37
	(0.02)	(0.01)	(0.00)	(0.01)	(0.02)
5^{th} quintile	0.68	0.05	0	0.01	0.27
	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)
Northeast					
1^{st} quintile	0.32	0	0	0.21	0.47
	(0.02)	(0.00)	(0.00)	(0.02)	(0.02)
2^{nd} quintile	0.43	0	0.01	0.15	0.4
	(0.02)	(0.00)	(0.00)	(0.01)	(0.02)
3^{rd} quintile	0.64	0	0.01	0.07	0.28
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
4^{th} quintile	0.8	0.01	0.01	0.02	0.17
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
5^{th} quintile	0.93	0	0	0	0.06
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
West					
1^{st} quintile	0.2	0.02	0	0.23	0.54
	(0.01)	(0.00)	(0.00)	(0.01)	(0.02)
2^{nd} quintile	0.31	0.02	0.01	0.12	0.54
	(0.02)	(0.01)	(0.00)	(0.01)	(0.02)
3^{rd} quintile	0.47	0.03	0.01	0.08	0.4
	(0.02)	(0.01)	(0.00)	(0.01)	(0.02)
4^{th} quintile	0.76	0.04	0	0.04	0.16
	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)
5^{th} quintile	0.91	0.02	0	0	0.06
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)

(continued on next page)

Table 3.A.3: (continued)

B. *Mean SCI-weighted distance to social neighbours*					
	High-High	High-Low	Low-High	Low-Low	Not sig.
All regions					
1st quintile	0.29	0.05	0	0.36	0.3
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
2nd quintile	0.33	0.05	0	0.26	0.37
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
3rd quintile	0.33	0.05	0	0.26	0.37
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
4th quintile	0.37	0.05	0	0.27	0.31
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
5th quintile	0.49	0.03	0	0.15	0.31
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
South					
1st quintile	0.09	0.08	0	0.57	0.26
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
2nd quintile	0.14	0.1	0	0.4	0.35
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
3rd quintile	0.17	0.11	0	0.4	0.31
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
4th quintile	0.28	0.08	0	0.4	0.24
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
5th quintile	0.4	0.05	0	0.25	0.3
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Midwest					
1st quintile	0.14	0.07	0	0.42	0.37
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
2nd quintile	0.25	0.02	0	0.28	0.45
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
3rd quintile	0.27	0.02	0	0.21	0.49
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)

(continued on next page)

Table 3.A.3: (continued)

	High-High	High-Low	Low-High	Low-Low	Not sig.
Midwest					
4th quintile	0.16	0.02	0	0.24	0.58
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
5th quintile	0.14	0.06	0	0.36	0.44
	(0.02)	(0.01)	(0.00)	(0.03)	(0.03)
Northeast					
1st quintile	0.57	0.01	0	0.13	0.29
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
2nd quintile	0.66	0	0.01	0.05	0.28
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
3rd quintile	0.71	0	0.01	0.04	0.24
	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
4th quintile	0.78	0	0.01	0.02	0.19
	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)
5th quintile	0.83	0	0	0	0.17
	(0.04)	(0.00)	(0.00)	(0.00)	(0.03)
West					
3rd quintile	1	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
4th quintile	0.84	0.01	0	0.03	0.11
	(0.00)	(0.00)	(0.84)	(0.01)	(0.01)
5th quintile	0.54	0.03	0	0.1	0.33
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)

Notes: Panels A and B each correspond to a multinomial logit regression with cluster membership as the outcome. Fitted probabilities and standard errors (in brackets) are obtained using the approach of Fox and Andersen (2006). Local Moran index values measure autocorrelation in EB-standardised (Assunção and Reis, 1999) Democratic two-party vote shares as of the 2020 US presidential election, in the social space spanned by the social neighbours of each area: the 1,347 most connected areas as measured by the Social Connectedness Index (SCI) (Bailey et al., 2020).

This page has been intentionally left blank.

CONCLUSION

The three empirical studies presented in this thesis have contributed new evidence on the relationship between the geographic structure of social networks and electoral outcomes, focusing on elections in the United Kingdom and the United States. Pursuant of this aim, each study has demonstrated the utility of using aggregate social media data in a spatial econometric framework in advancing research on the role of neighbourhood context in voting behaviour. Other than their joint contribution to this literature, each study has also addressed domain-specific questions with distinct implications for the research and policy areas in question. In this conclusion, I offer a final discussion of the main contributions of the thesis (Section II.1), along with overviews of their policy implications (Section II.2), their limitations (Section II.3), and promising directions for future research (Section II.4).

II.1 Main Contributions

The overarching contribution of this thesis has been to provide new evidence suggesting that the aggregate social ties between the populations of distant places constitute an important contextual influence on voting behaviour. This has remained a largely neglected prospect in the long-standing literature on neighbourhood effects in elections, which has typically placed more emphasis on the role of local geographic context (e.g. Ethington and McDaniel, 2007; Gimpel and Reeves, 2022; Johnston and Pattie, 2014). While early studies of individual voters within local communities have documented substantial variation in the geographic dispersion of social ties (e.g. Baybeck and Huckfeldt, 2002a,b), more systematic examinations of its relationship to electoral outcomes have been lacking, partly owing to data limitations. In this thesis, I have exploited newly available data on the population of social media users to begin addressing this gap in the form of three empirical studies. The studies presented in Chapter One and Chapter Two have investigated the causal social spillover effects of localised events on voting behaviour in the United Kingdom and the United States. Taken together, the findings of these studies suggest that the effects of localised shocks on both vote choice and choice of voting method often spread over longer geographic distances than those that are usually considered in the empirical study of elections, adding to the growing body of evidence on the presence of long-distance social spillovers in other domains (e.g. Charoenwong et al., 2020; Holtz et al., 2020; Zhao et al., 2021; Makridis, 2022; Wilson, 2022). The third chapter has further investigated partisan differences in the geographic dispersion of social ties in the United States, suggesting that some groups of voters may be more exposed to such spillovers than others.

Other than their shared contributions to the broader literature on neighbourhood effects in elections, each chapter in this thesis has also made unique substantive contributions to distinct strands of work. Chapter One contributes to the literature on the effects of Chinese import competition on support for the Leave option in the 2016 UK EU membership referendum (Ballard-Rosa et al., 2021; Colantone and Stanig, 2018a; Steiner and Harms, 2021). Whereas previous work has found that exposure to Chinese import competition within a voter's local labour market is likely to have had positive sociotropic effects on their support for the Leave option, the chapter examines the neglected possibility of social interactions between local labour markets.[1] Namely, it poses the following research question:

- RQ1: How was support for the Leave option in the 2016 UK EU membership referendum affected by voters' social interactions with those exposed to import competition in other local labour market

[1] Recall that in the context of Chapter One local labour markets are equivalent to commuting zones delineated using the approach of Tolbert and Sizer (1996).

Using data on the density of social ties among ITL3 regions in England and Wales, the empirical analysis in Chapter One showed that, when it comes to support for the Leave option, being resident in a region that is socially connected to others in different local labour markets that are highly exposed to Chinese import competition had a positive effect of similar magnitude to that of being resident in a region that is itself highly exposed. That is, the evidence suggests that voters' social interactions with those exposed to import competition in other local labour markets had a sizeable positive effect on support for the Leave option. Other than showing that the social interactions in question occurred over fairly long geographic distances, with the socially connected regions considered on average lying between 74 and 102 kilometres apart, the analysis also showed that these are unlikely to represent purely economic processes involving linked industries, aggregate demand effects, or reallocation of labour.[2]

The evidence presented in Chapter One has implications for the wider international literature on the political effects of trade (e.g. Autor et al., 2020; Barone and Kreuter, 2021; Colantone and Stanig, 2018c; Dippel et al., 2021; Malgouyres, 2017). Namely, it highlights that trade-related employment shocks, which disproportionately affect local labour markets with a historical reliance on exposed manufacturing industries, can have significant effects on electoral outcomes beyond the local labour markets in which they occur via social interactions. This represents a departure from a common assumption that is often implicitly carried over from the parallel literature on the economic effects of trade (e.g. Autor et al., 2013, 2016; Dorn and Levell, 2021), which anticipates that social interactions will be largely contained within local labour markets. Chapter One shows that while this might be defensible in the case of spillovers operating via economic channels, it is less likely to be so in the case of social spillovers operating via information flows among voters. Consequently, empirical specifications that do not accommodate social interactions between local labour markets may underestimate the political effects of trade. Chapter Three has shown that the identification of social spillover effects can be pursued as part of the instrumental variable approaches that are common in the literature, insofar as the interregional social ties considered are uncorrelated to instrumented measures of trade-shock exposure. Reassuringly, it is also shown that the latter is unlikely to be an overly restrictive condition due to expected frictions in the mobility of workers between local labour markets.

Chapter Two contributes to the literature on the effects of all-mail voting—the automatic dispatch of mail ballots—on voting by mail in the United States (Thompson et al., 2020; Herrnson and Stewart, 2023). While previous work has shown that all-mail voting is the most consequential reform for rises in the demand for mail ballots, its potential social spillovers in jurisdictions where mail ballots are available upon request, so-called 'no-excuse absentee voting' jurisdictions, have received next to no attention. Chapter Two addresses

[2] Recall that these types of economic spillovers anticipated in the framework of Acemoglu et al. (2016).

this gap in the literature by focusing on voting behaviour in the 2020 US presidential election, which saw an unprecedented number of jurisdictions roll out the reform in response to the public health risks imposed by the COVID-19 pandemic.[3] Specifically, the chapter poses the following research question:

- RQ2: How was the demand for mail ballots in no-excuse absentee voting jurisdictions in the 2020 US presidential election affected by voters' social interactions with those in jurisdictions switching to all-mail voting?

The empirical analysis in Chapter Two relies on administrative data on the entire electorate of the no-excuse absentee voting state of North Carolina and data on the density of social ties among ZCTAs across the United States. The evidence shows that residing in a ZCTA where a higher share of the aggregate social ties of the local population reside in jurisdictions that switched to all-mail voting between the 2016 and 2020 US presidential elections had a positive effect on the probability of voting by mail among North Carolinian voters. The identified effects are substantial: the estimated increase in probability is nearly 10 percentage points for voters in ZCTAs with the highest shares. In other words, social interactions with the residents of jurisdictions switching to all-mail voting are likely to have increased the demand for mail ballots in North Carolina in the 2020 US presidential election. Further, it is also shown that these social spillover effects originate from jurisdictions lying well away from state borders, and are most pronounced among voters who are older, resident in metropolitan counties, and Democratic or unaffiliated with a political party.

Aside from its contribution to the literature on the effects of all-mail voting, the findings of Chapter Two are also relevant to the wider policy evaluation literature on the effects of electoral reforms on political participation (Amlani and Collitt, 2022; McGhee et al., 2022) and partisan gains (Barber and Holbein, 2020; Thompson et al., 2020; Yoder et al., 2021) in the United States. As electoral reforms are usually rolled out at the county and state levels, the findings imply that, insofar as local rollouts of any given electoral reform affect local preferences, the social interactions of the local population with those of other jurisdictions could plausibly have an effect on such outcomes elsewhere. Consequently, failing to consider these socio-spatial spillovers, as is often the case in the literature, may limit researchers' ability to explain and anticipate relevant trends in voting behaviour. Moreover, in observational research designs leveraging comparisons between jurisdictions introducing new reforms and those that do not, the presence of unmodeled socio-spatial spillover effects between the localities in focus could also lead to biased estimates of the direct effects of reforms.

[3] Recall that in many jurisdictions the emergency electoral reforms introduced during the COVID-19 pandemic have since become permanent (National Conference of State Legislatures, 2022b).

As exposure to any given reform is a binary quantity, Chapter Two shows that one approach to the identification of these effects is the calculation of the share of aggregate social ties exposed and its use in a difference-in-differences specification subject to the usual conditions around untreated potential outcomes and selection bias.

Chapter Three contributes to the growing literature on partisan segregation in the United States (Brown and Enos, 2021; Darmofal and Strickler, 2019; Johnston et al., 2020; Kinsella et al., 2021, 2015). Previous work has often theorised that spatial patterns in partisan segregation are likely to be indicative of similar patterns in partisan homophily; however, this assumption has remained largely untested owing to data limitations. Chapter Three pairs data on the density of online social ties between ZCTAs across the United States with data on voting behaviour in the 2020 US presidential election to address this gap, posing the following research question:[4]

- RQ3: How did partisan homophily among the populations of different localities vary in geographic space in the 2020 US presidential election?

The empirical findings of Chapter Three confirm some widely held expectations arising in previous research. First, based on voting behaviour in the 2020 presidential election, most ZCTAs in the United States were politically homophilous: that is, more likely to be socially connected to politically similar rather than dissimilar others. Second, more urbanised ZCTAs were more likely to be Democratic-leaning and homophilous, while more rural ZCTAs were more likely to be Republican-leaning and homophilous as well as being less likely to be homophilous relative to more urbanised ZCTAs. Third, the broad regional patterns of partisan homophily follow those of partisan segregation, with most Republican-leaning and homophilous ZCTAs being spread across the rural Midwest and South, and most Democratic-leaning and homophilous ZCTAs being seen across cities and on the Western and Northeastern coasts. However, unlike previous research, the findings also reveal partisan differences in the average geographic distances over which co-partisan social ties are likely to be maintained. Overall, the ZCTAs with the highest relative density of social ties in other areas and the highest average geographic distances of social ties were substantially more likely to be Democratic-leaning and homophilous.

Methodologically, Chapter Three presents a novel application of the local Moran index (Anselin, 1995) on the measurement of spatial patterns in partisan homophily. In this effort, it also makes novel substantive contributions to the interface between the established literatures on partisan segregation and political polarisation in the United States (e.g. Hopkins, 2017; Mettler and Brown, 2022; Myers, 2013; Rohla et al., 2018; Scala and Johnson, 2017).

[4]The data on social ties come from the ZCTA-level version of the Social Connectedness Index by Bailey et al. (2020), which is the same dataset employed in Chapter Two.

Researchers have often alluded to widespread partisan segregation as a potential driver of political polarisation operating via constraints on inter-partisan contact. Chapter Three offers the first country-wide evidence suggesting that partisan segregation in geographic space does indeed tend to coincide with segregation in social space, reinforcing the prospect that the former is a relevant factor in the rise of political polarisation. At the same time, the finding that local Democratic-leaning populations are most likely to be connected over long geographic distances sheds new light on the role of geographic processes in the formation of partisan identities. Namely, it suggests that the political preferences of voters in Democratic-leaning areas are likely to be more exposed to non-local influences from different residential environments when compared to the preferences of voters in Republican-leaning areas.[5]

II.2 Policy Implications

Other than demonstrating that the geographic structure of social networks matters when it comes to electoral outcomes, the three chapters presented in this thesis have also discussed the diverse implications that this relationship can have for policy. First, recall how Chapter One discussed that the outcome of the 2016 UK EU membership referendum is often associated with a trend declining trust in democratic institutions in the United Kingdom (e.g. McCann and Ortega-Argilés, 2021). Given the higher probability of support for the Leave option in regions experiencing relative economic decline, or 'left-behind places', it has been argued that 'place-based policies' aimed at promoting growth by leveraging the competitive advantages of local economies are likely to also be effective in reducing political discontent (Martin et al., 2021; Rodríguez-Pose, 2018). The new evidence presented in Chapter One suggests that policy responses of this kind also bear the promise of positive spatial externalities in socially connected regions. That is, as information flows from regions exposed to import competition are likely to have aided the spatial diffusion of political discontent, local economic growth could similarly drive wider increases in public trust in institutions as well as easing the social tensions highlighted by the referendum (e.g. Williams et al., 2022).

Recall further that Chapter Two was partially motivated by the observation that gauging the local demand for voting by mail is an important task facing election officials across the United States, who are required to make advance arrangements in ensuring the timely and effective processing of mail ballots (U.S. Election Assistance Commission, 2010). Indeed, the empirical findings show that no-excuse absentee voting jurisdictions are likely to be faced with unexpected increases in this demand when higher shares of the aggregate social ties of the local population are resident in jurisdictions where all-mail voting is being rolled out.

[5] Interestingly, this is echoed in Chapter Two, which shows that socio-spatial spillover effects on choice of voting method from the rollout of all-mail voting are weaker for Republicans (Table 2.3).

This has straightforward implications for election administration: the use of information on the aggregate social ties of the local electorate, as this is captured by measures such as the Social Connectedness Index (Bailey et al., 2018, 2020), could lead to improvements in resource allocation decisions. In line with heterogeneity in the approaches employed in predicting demand for mail ballots across jurisdictions, this could range from simply reviewing the share of social ties experiencing reforms to incorporating this information in predictive models.

Finally, as per the discussion in Chapter Three, the finding that the geography of partisan homophily in the United States is characterised by relatively stronger social connections over long geographic distances among Democratic-leaning areas, may make some proposed to policy responses to rising political polarisation more attractive. For instance, Rodden (2015) proposes that the local introduction of compulsory voting in states controlled by Democratic legislatures—a policy that would likely face greater resistance in Republican-controlled states—would serve to reduce polarisation by removing the incentive for political candidates to campaign on extreme positions to mobilise their respective state electoral bases. Consequently, given the heightened prospect of interstate social interactions between the populations of Democratic-leaning areas, a possible moderate turn in local campaigning from the introduction of compulsory voting within a blue state could plausibly also be reflected in more moderate political preferences in other states.

II.3 Limitations

Each of the three empirical chapters within this thesis have discussed their specific limitations, particularly as these relate to issues around the geographic and temporal scope of inference. That is, while each chapter highlights relationships between the geographic structure of social networks and electoral outcomes at the time and place of the elections considered, additional evidence would be required before one could be confident that similar relationships will hold elsewhere.[6] Perhaps more importantly, some key limitations of this thesis are shared across chapters and largely stem from what has also constituted one of its strengths: the use of aggregate social media data.

All chapters have relied on the Social Connectedness Index by Bailey et al. (2018, 2020,2020) to observe the pairwise density of online social ties between localities, which is itself based on the universe of friendships between active users of Facebook.

[6]This of course also applies to the policy implications discussed in the previous section. Section II.3 offers some directions on the kind of evidence that may be collected in examining these further.

While each chapter has reviewed the strong and growing evidence that the index constitutes a reliable proxy for the geographic structure of all social ties in the United Kingdom and the United States, there is no publicly available information on variation in usage of the service in the periods and spatial resolutions considered. In the absence of this information, it is in principle possible that, for some pairs of localities, index values are less representative of the social ties of people who do not use Facebook.This could imply bias in the estimated effects on voting behaviour in the more causally oriented analyses in Chapter One and Chapter Two, given evidence on the relationship between Internet use and political participation (Geraci et al., 2022; Zhuravskaya et al., 2020). Reassuringly, the results are in both cases robust to the inclusion of available controls for Internet use, which assuages these concerns.

The use of aggregate social media also places limits on the interpretation of findings in line with broader critiques of the neighbourhood effects literature (e.g. King, 1996). When looking at relationships between individual-level voting behaviour and aggregate characteristics measured at the level of spatial units, picking apart the precise causal processes at play is often prohibitive. Indeed, while the first two chapters of this thesis have established relationships between voting behaviour and social context as captured by the non-local aggregate social ties of the population of the locality in which the voter is resident, it is difficult to single out individual mechanisms. Take for example the findings of Chapter One establishing that individual-level support for Brexit responds to regional social proximity to import competition in other local labour markets. While the evidence suggests that the spillovers in question operate via information flows, the available data prevent a detailed understanding of how the latter might be affecting voting behaviour at the individual level. In line with the conceptual discussion in the chapter, this could involve a wide range of possibilities, such as telephone conversations about the state of the economy, the sharing of online news stories on cultural issues, or physical observation of anti-social behaviour to name a few. Recall also that, in the absence of information on the social ties between individual voters, it is also not possible to gauge the extent in which the identified spillovers involve first-degree interactions with non-local peers in exposed local labour markets, or higher-degree interactions with local peers.[7]

Note further that the use of aggregate social media data is restrictive with respect to the interpretation of heterogeneous spillover effects. For instance, take the findings of Chapter Two looking at voting behaviour in North Carolina: it is estimated that a marginal increase in the ZCTA-level share of the non-local social ties that switched to all-mail voting was associated with a greater increase in the probability of voting by mail among unaffiliated voters when compared to registered Republicans. So what is behind this heterogeneity?

[7] Again, recall that first-degree peers are those with a direct interpersonal connection, while n-degree peers are those with at least one common peer of degree $n-1$.

Is could equally be that unaffiliated voters have more social ties to these jurisdictions or that they are more sensitive to the same increase in the share of social ties exposed to the electoral reform. Again, while the identification of heterogeneous effects may, for policy or practical purposes, be satisfactory in itself, it is difficult to say which of the two processes is at play without information on the whereabouts of individual voters' social ties.

As discussed in the introduction of this thesis, an additional factor limiting the interpretation of the evidence presented in Chapter One and Chapter Two stems from the adopted empirical approach. That is, both chapters have forgone the objective of separately estimating the effects of 'endogenous' and 'contextual' interactions (Manski, 1993), instead targeting a composite parameter which can be more straightforwardly estimated in a quasi-experimental setting (e.g. see Gibbons and Overman, 2012). While this has clear benefits for the identification of spillover effects, it naturally restricts one's capacity to separate some aggregate-level mechanisms. For instance, in the case of Chapter One, the empirical analysis does not reveal the relative extent in which exposure to import competition in socially connected regions affects local support for Brexit directly or indirectly via increases in support in these regions.

A final limitation relates specifically to Chapter Three. As the empirical analysis in the chapter is entirely based on spatial units, inferences about the behaviour of individual voters from its findings can be subject to the ecological fallacy. For instance, in a strict sense, the finding that Democratic-leaning ZCTAs are more likely to be socially connected to distant others relative to their Republican-leaning counterparts does not necessarily imply that Democrats on average have more distant co-partisan social ties than Republicans.[8] As such, when reviewing the evidence presented in the chapter, it is important to remember that partisanship refers to the composition of local electorates within ZCTAs rather than to individual voters.

II.4 Future Research

Notwithstanding its limitations, this thesis has shown that aggregate social media data have the potential to uncover novel aspects of spatial dependence in voting behaviour, especially when combined with established approaches to causal inference. As such, a lot stands to be gained from future research that continues to leverage this kind of data to assess the generalisability of the findings of this thesis in different national contexts and in the long-distance spillovers of different kinds of localised shocks.

[8] Though note that as it has been shown that it is fairly common for ZCTAs to be highly politically homogeneous, this may be of limited concern in the context of the 2020 US presidential election.

As seen in Chapter One, import competition has been shown to affect political preferences in places beyond the UK, such as the United States (Autor et al., 2020), Germany (Dippel et al., 2021), France (Malgouyres, 2017), and Italy (Barone and Kreuter, 2021). As such, examining whether localised import-shock exposure in these countries bears similar long-distance social spillovers on voting behaviour as those identified in Chapter One for the UK would be an important step in establishing their relevance to the wider literature on the political economy trade, which has constituted one of the most fruitful attempts at developing an understanding of the origins of rising political discontent in Western democracies in the 21st century. In line with parallel strands of work in this domain (e.g. see Guriev and Pappaioannou, 2021), it is also important to examine whether other relevant economic and non-economic shocks also beget these kinds long-distance spillovers on electoral outcomes associated with political discontent. Here, of particular interest are localised shocks relating to austerity (Fetzer, 2019), automation (e.g. Anelli et al., 2021; Frey et al., 2018), and immigration (e.g. Barone et al., 2016; Bratti et al., 2020; Dinas et al., 2019; Edo et al., 2019), given the strong evidence on their relevance to political preferences.

Further, in improving evidence on the policy implications of Chapter One, an equally promising strand of future work may evaluate the spatial externalities of place-based policies on political preferences, or outcomes relating to wider social cohesion as captured, for instance, by indicators developed using individual-level social media data (e.g. Williams et al., 2022). For instance, Crescenzi et al. (2020) find that local labour markets in the UK that saw tangible improvements as a result of EU structural funds were more likely to support remaining in the union in the referendum; naturally, establishing whether the localised effects of such economic interventions have long-distance spillovers on trust in institutions and social cohesion is of policy interest. Given the availability of the Social Connectedness Index for most Western countries, incorporating the modelling of long-distance spillovers for many of the discussed example cases is unlikely to be fundamentally different to the empirical approach of Chapter One, as identification of direct effects generally relies on similar quasi-experimental estimators.

Following on from Chapter Two, a natural direction for gauging the generalisability of findings is to extend the analysis of the socio-spatial spillovers of all-mail voting on choice of voting method to states beyond North Carolina. While one of the reasons that the latter was preferred as the geographic context for the analysis of Chapter Two was the public availability of voter registration data, barriers to access to voter files in several other no-excuse absentee voting states are generally low (National Conference of State Legislatures, 2023), making this an attractive possibility. Similar kinds of questions merit attention in the parallel strand of work on the effects of electoral reforms on political participation (Amlani and Collitt, 2022; McGhee et al., 2022) and partisan gains (Barber and Holbein, 2020; Thompson et al., 2020; Yoder et al., 2021).

Namely, it would be of policy interest to additionally examine the extent of socio-spatial spillovers on these outcomes from local rollouts of both all-mail voting and similar reforms, such as the mailing of absentee ballot applications. An interesting prospect also arises from the evidence on the effects of polling place changes on political participation. For instance, as previous work has raised concerns that polling place consolidation may have suppressive effects on turnout (e.g. Amos et al. 2017; Clinton et al. 2021; Brady and McNulty 2011), it would be of policy interest to examine the extent that social interactions with the residents of affected areas have wider negative spatial externalities on political participation. In contrast to US-specific electoral reforms, this question is also likely to be of relevance to different national contexts such as the UK, where polling place changes have also been shown to play an important role in voter turnout (e.g. Orford et al., 2011, 2008) and where the Social Connectedness Index is likely to be a similarly good proxy for the geographic structure of social networks. A possible limitation here is that the index is available at lower spatial resolutions outside the US, which may be limiting for popular research designs leveraging comparisons between proximate voters in changing polling precincts. Though as seen in Chapter One, whilst the spatial resolution of the index in countries like the UK is generally coarser, analyses focusing on administrative units around urban centres may overcome this limitation as these tend to be relative more granular.

Note that a further potential avenue for future work arising from both Chapter One and Chapter Two is the examination of the dynamic effects of the long-distance spillovers on voting behaviour over time, which would afford a more accurate of these processes than cross-sectional or two-period designs. While this is of course a highly appealing prospect, it is also subject to greater methodological challenges. For one, the Social Connectedness Index is renewed at infrequent intervals and undergoes substantial methodological changes which could hinder a longitudinal analysis spanning far beyond the year of any given iteration of the index. Further, given newly identified challenges in the estimation of dynamic effects in the econometric literature (e.g. Roth et al., 2023), the selection of appropriate estimators for each application remains a non-trivial exercise. In the continued absence of more detailed data, it would seem necessary to proceed with a narrow focus on a small number of successive elections whilst carefully evaluating assumptions around treatment effect heterogeneity.

Following on from the contributions of Chapter Three, a promising research direction would involve a closer examination of homophily between the populations of voters in distant urban, Democratic-leaning areas. While these areas have been identified as the most likely to be relatively strongly socially connected to distant and politically similar others, less is known about the possible demographic features underlying this relationship. Many urban environments across the US have been known to be residentially segregated by features like ethnicity and income (e.g. Jargowsky, 2020).

As these features have previously been associated to individual differences in the geographic dispersion of social ties (e.g. Fischer, 1982), it would be important to determine whether long-distance social ties between voters in Democratic-leaning areas are also connecting voters in homophilous socio-economic environments. To the extent that they do, this could have implications for the exposure of particular social groups of voters to socio-spatial spillovers. For instance, as recent evidence suggests that contact between social out-groups in segregated urban environments has significant effects on political preferences (Enos, 2017), this could mean that the preferences of more distally connected groups are more amenable to spatial diffusion. In line with the use of aggregate social media in the calculation of local Moran indices in Chapter Three, a promising methodological approach in developing a better understanding of the relationship between partisan and social homophily is the calculation and comparison of multivariate local Moran indices (Wolf et al., 2022) to reveal the relative importance of different demographic factors in driving the former. As with previous chapters, an examination of the geography of political homophily in other national contexts that also present distinct dynamics in urban segregation and desegregation, such as the UK (e.g. Harris, 2023; Lan et al., 2020), is also necessary to establish the generalisability of the findings of Chapter Three.

On a final and more general note, the findings of this thesis suggest that any future work on the role of local geographic context in electoral outcomes would be, somewhat counter-intuitively, well-served to frame this as encompassing a non-local social context defined by the aggregate social ties of the local population in other, potentially distant localities. Again, while this prospect has been at times theoretically entertained, it has largely eluded empirical attention in the neighbourhood effects literature. In this spirit, whereas this thesis commenced with a reiteration of Tobler's influential first law of geography, it seems fitting that it should end with the often forgotten second, which is perhaps a good reminder of the evidence presented herein: 'the phenomenon external to a geographic area of interest affects what goes on inside' (1999, 87).

BIBLIOGRAPHY

Abrams, S. J. and M. P. Fiorina (2012). "The Big Sort" That Wasn't: A Skeptical Reexamination. *PS: Political Science & Politics 45*(2), 203–210.

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016). Import Competition and the Great US Employment Sag of the 2000s. *Journal of Labor Economics 34*(S1), S141–S198.

Agnew, J. (1996). Mapping Politics: How Context Counts in Electoral Geography. *Political Geography 15*(2), 129–146.

Akerlof, G. A. (1997). Social Distance and Social Decisions. *Econometrica 65*(5), 1005–1027.

Alt, J. E., A. Jensen, H. Larreguy, D. D. Lassen, and J. Marshall (2022). Diffusing Political Concerns: How Unemployment Information Passed between Social Ties Influences Danish Voters. *The Journal of Politics 84*(1), 383–404.

Amlani, S. and S. Collitt (2022). The Impact of Vote-by-Mail Policy on Turnout and Vote Share in the 2020 Election. *Election Law Journal: Rules, Politics, and Policy 21*(2), 135–149.

Amos, B., D. A. Smith, and C. S. Clair (2017). Reprecincting and Voting Behavior. *Political Behavior 39*(1), 133–156.

Anelli, M., I. Colantone, and P. Stanig (2021). Individual vulnerability to industrial robot adoption increases support for the radical right. *Proceedings of the National Academy of Sciences 118*(47), e2111611118.

Angrist, J. D. and J.-S. Pischke (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht, NL: Kluwer Academic Publishers.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis 27*(2), 93–115.

Anselin, L. (1996). The Moran Scatterplot as an ESFA Tool to Assess Local Instability in Spatial Association. In M. Fischer, H. Scholten, and D. Unwin (Eds.), *Spatial Analytical*

Perspectives on GIS in Environmental and Socio-Economic Sciences. London, UK: Taylor & Francis.

Anselin, L. and A. Bera (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In A. Ullah and D. E. Giles (Eds.), *Handbook of Applied Economic Statistics*, pp. 237–289. New York, NY: Marcel Dekker.

Ansolabehere, S., M. Meredith, and E. Snowberg (2014). Mecro-Economic Voting: Local Information and Micro-Perceptions of the Macro-Economy. *Economics & Politics 26*(3), 380–410.

Assunção, R. M. and E. A. Reis (1999). A New Proposal to Adjust Moran's I for Population Density. *Statistics in Medicine 18*(16), 2147–2162.

Atkeson, L. R., W. L. Hansen, M. T. Oliver, C. D. Maestas, and E. C. Wiemer (2022). Should I Vote-by-Mail or In Person? The Impact of COVID-19 Risk Factors and Partisanship on Vote Mode Decisions in the 2020 Presidential Election. *PLOS One 17*(9), 1–21.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review 110*(10), 3139–83.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review 103*(6), 2121–68.

Autor, D. H., D. Dorn, and G. H. Hanson (2016). The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade. *Annual Review of Economics 8*(1), 205–240.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives 32*(3), 259–80.

Bailey, M., E. Dávila, T. Kuchler, and J. Stroebel (2018). House Price Beliefs And Mortgage Leverage Choice. *The Review of Economic Studies 86*(6), 2403–2452.

Bailey, M., P. Farrell, T. Kuchler, and J. Stroebel (2020). Social Connectedness in Urban Areas. *Journal of Urban Economics 118*, 103264.

Bailey, M., A. Gupta, S. Hillenbrand, T. Kuchler, R. Richmond, and J. Stroebel (2021). International Trade and Social Connectedness. *Journal of International Economics 129*, 103418.

Bailey, M., D. Johnston, T. Kuchler, D. Russel, B. State, and J. Stroebel (2020). The Determinants of Social Connectedness in Europe. In S. Aref, K. Bontcheva, M. Braghieri, F. Dignum, F. Giannotti, F. Grisolia, and D. Pedreschi (Eds.), *Social Informatics*, Cham, pp. 1–14. Springer International Publishing.

Bailey, M., D. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2022). Peer Effects in Product Adoption. *American Economic Journal: Applied Economics 14*(3), 488–526.

Ballard-Rosa, C., M. A. Malik, S. J. Rickard, and K. Scheve (2021). The Economic Origins of Authoritarian Values: Evidence From Local Trade Shocks in the United Kingdom. *Comparative Political Studies 54*(13), 2321–2353.

Barber, M. and J. B. Holbein (2020). The Participatory and Partisan Impacts of Mandatory Vote-By-Mail. *Science Advances 6*(35), eabc7685.

Barone, G., A. D'Ignazio, G. de Blasio, and P. Naticchioni (2016). Mr. rossi, mr. hu and politics. the role of immigration in shaping natives' voting behavior. *Journal of Public Economics 136*, 1–13.

Barone, G. and H. Kreuter (2021). Low-Wage Import Competition and Populist Backlash: The Case of Italy. *European Journal of Political Economy 67*, 101970.

Barreto, M. A., M. J. Streb, M. Marks, and F. Guerra (2006). Do Absentee Voters Differ from Polling Place Voters?: New Evidence from California. *Public Opinion Quarterly 70*(2), 224–234.

Baybeck, B. and R. Huckfeldt (2002a). Spatially Dispersed Ties Among Interdependent Citizens: Connecting Individuals and Aggregates. *Political Analysis 10*(3), 261–275.

Baybeck, B. and R. Huckfeldt (2002b). Urban Contexts, Spatially Dispersed Networks, and the Diffusion of Political Information. *Political Geography 21*(2), 195–220.

Beck, N., K. S. Gleditsch, and K. Beardsley (2006). Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy. *International Studies Quarterly 50*(1), 27–44.

Becker, S. O., T. Fetzer, and D. Novy (2017). Who Voted for Brexit? A Comprehensive District-Level Analysis. *Economic Policy 32*(92), 601–650.

Benjamini, Y. and Y. Hochberg (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 289–300.

Berelson, B. R., P. F. Lazarsfeld, and W. N. McPhee (1986). *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago, IL: University of Chicago Press.

Bishop, B. (2020). For Most Americans, the Local Presidential Vote Was a Landslide. *The Daily Yonder*.

Bishop, B. and R. G. Cushing (2009). *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart.* New York, NY: Houghton Mifflin Harcourt.

Bivand, R. (2022). R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data. *Geographical Analysis 54*(3), 488–518.

Bolet, D. (2021). Drinking Alone: Local Socio-Cultural Degradation and Radical Right Support—The Case of British Pub Closures. *Comparative Political Studies 54*(9), 1653–1692.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature 489*(7415), 295–298.

Books, J. W. and C. L. Prysby (1991). *Political Behavior and the Local Context.* Westport, CT: Praeger Publishers.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups. *Proceedings of the National Academy of Sciences 114*(40), 10612–10617.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2020). Cross-Country Trends in Affective Polarization. National Bureau of Economic Research Working Paper 26669.

Brady, H. E. and J. E. McNulty (2011). Turning Out to Vote: The Costs of Finding and Getting to the Polling Place. *The American Political Science Review 105*(1), 115–134.

Bramoullé, Y., H. Djebbari, and B. Fortin (2020). Peer Effects in Networks: A Survey. *Annual Review of Economics 12*(1), 603–629.

Bratti, M., C. Deiana, E. Havari, G. Mazzarella, and E. C. Meroni (2020). Geographical Proximity to Refugee Reception Centres and Voting. *Journal of Urban Economics 120*, 103290.

Brown, J. R., E. Cantoni, R. D. Enos, V. Pons, and E. Sartre (2023). The Increase in Partisan Segregation in the United States. Nottingham Interdisciplinary Centre for Economic and Political Research (NICEP) Discussion Paper 2023-09.

Brown, J. R. and R. D. Enos (2021). The Measurement of Partisan Sorting for 180 Million Voters. *Nature Human Behaviour 5*(8), 998–1008.

Burden, B. C., D. T. Canon, K. R. Mayer, and D. P. Moynihan (2014). Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform. *American Journal of Political Science 58*(1), 95–109.

Cairncross, F. (2001). *The Death of Distance: How the Communications Revolution Will Change Our Lives.* Boston, MA: Harvard Business School Press.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021). Difference-in-Differences with a Continuous Treatment. arXiv preprint 2107.02637.

Charoenwong, B., A. Kwan, and V. Pursiainen (2020). Social Connections with COVID-19–Affected Areas Increase Compliance with Mobility Restrictions. *Science Advances 6*(47), eabc3054.

Che, Y., X. Xu, and Y. Zhang (2018). Chinese Import Competition, Crime, and Government Transfers in US. *Journal of Comparative Economics 46*(2), 544–567.

Chen, J. and J. Rodden (2013). Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science 8*(3), 239–269.

Christakis, N. A. and J. H. Fowler (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine 357*(4), 370–379.

Clinton, J. D., N. Eubank, A. Fresh, and M. E. Shepherd (2021). Polling Place Changes and Political Participation: Evidence from North Carolina Presidential Elections, 2008–2016. *Political Science Research and Methods 9*(4), 800–817.

Clinton, J. D., J. Lapinski, S. Lentz, and S. Pettigrew (2022). Trumped by Trump? Public support for Mail Voting in Response to the COVID-19 Pandemic. *Election Law Journal: Rules, Politics, and Policy 21*(1), 19–33.

Colantone, I. and P. Stanig (2018a). Global Competition and Brexit. *American Political Science Review 112*(2), 201–218.

Colantone, I. and P. Stanig (2018b). The Economic Determinants of the'Cultural Backlash': Globalization and Attitudes in Western Europe. Baffi carefin centre research paper 2018-91.

Colantone, I. and P. Stanig (2018c). The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe. *American Journal of Political Science 62*(4), 936–953.

Constantino, S. M., A. D. Cooperman, and T. M. Q. Moreira (2021). Voting in a Global Pandemic: Assessing Dueling Influences of Covid-19 on Turnout. *Social Science Quarterly 102*(5), 2210–2235.

Coombes, M. and S. Bond (2008). Travel-to-Work Areas: the 2007 Review. Technical report, Office for National Statistics.

Cox, K. R. (1969). The Voting Decision in a Spatial Context. *Progress in Geography 1*, 81–117.

Crescenzi, R., M. Di Cataldo, and M. Giua (2020). It's not about the money. eu funds, local opportunities, and euroscepticism. *Regional Science and Urban Economics 84*, 103556.

Darmofal, D. and R. Strickler (2019). *Demography, Politics, and Partisan Polarization in the United States, 1828-2016*. Cham, CH: Springer Nature.

de Castro, M. C. and B. H. Singer (2006). Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association. *Geographical Analysis 38*(2), 180–208.

de Chaisemartin, C. and X. D'Haultfoeuille (2022). Two-way Fixed Effects and Differences-in-Differences Estimators with Several Treatments. National Bureau of Economic Research Working Paper 30564.

De Lyon, J. and J. P. Pessoa (2021). Worker and Firm Responses to Trade Shocks: The UK-China Case. *European Economic Review 133*, 103678.

Dieye, R., H. Djebbari, and F. Barrera-Osorio (2014). Accounting for Peer Effects in Treatment Response. Iza discussion paper 8340.

Dinas, E., K. Matakos, D. Xefteris, and D. Hangartner (2019). Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-Right Parties? *Political Analysis 27*(2), 244–254.

Dippel, C., R. Gold, S. Heblich, and R. Pinto (2021, 05). The Effect of Trade on Workers and Voters. *The Economic Journal 132*(641), 199–217.

Dix-Carneiro, R., R. R. Soares, and G. Ulyssea (2018). Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization. *American Economic Journal: Applied Economics 10*(4), 158–95.

Dorn, D. and P. Levell (2021). Trade and inequality in europe and the us.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy 65*(2), 135–150.

Duggan, M., N. B. Ellison, C. Lampe, A. Lenhart, and M. Madden (2015). Demographics of Key Social Networking Platforms. Pew Research Center.

Dunn, A. and M. Gholizadeh (2020). The Geography of Consumption and Local Economic Shocks: The Case of the Great Recession. BEA Working Paper 2020-09.

Durlauf, S. N. (2004). Neighborhood Effects. In J. V. Henderson and J.-F. Thisse (Eds.), *Cities and Geography*, Volume 4 of *Handbook of Regional and Urban Economics*, pp. 2173–2242. Amsterdam, NL: Elsevier.

Durlauf, S. N. and Y. M. Ioannides (2010). Social Interactions. *Annual Review of Economics 2*(1), 451–478.

Edo, A., Y. Giesing, J. Öztunc, and P. Poutvaara (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review 115*, 99–143.

Efron, B. and T. Hastie (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science.* Cambridge, UK: Cambridge University Press.

Elhorst, J. P. (2014). *Linear Spatial Dependence Models for Cross-Section Data.* New York, NY: Springer.

Enos, R. D. (2017). *The Space Between Us: Social Geography and Politics.* New York, NY: Cambridge University Press.

Ethington, P. J. and J. A. McDaniel (2007). Political Places and Institutional Spaces: The Intersection of Political Science and Political Geography. *Annual Review of Political Science 10*(1), 127–142.

Fabina, J. and Z. Scherer (2022). Voting and Registration in the Election of November 2020. U.S. Census Bureau.

Feler, L. and M. Z. Senses (2017). Trade Shocks and the Provision of Local Public Goods. *American Economic Journal: Economic Policy 9*(4), 101–43.

Fernandez-Navia, T., E. Polo-Muro, and D. Tercero-Lucas (2021). Too Afraid to Vote? The Effects of COVID-19 on Voting Behaviour. *European Journal of Political Economy 69*, 102012.

Fetzer, T. (2019). Did Austerity Cause Brexit? *American Economic Review 109*(11), 3849–86.

Finkel, E. J., C. A. Bail, M. Cikara, P. H. Ditto, S. Iyengar, S. Klar, L. Mason, M. C. McGrath, B. Nyhan, D. G. Rand, L. J. Skitka, J. A. Tucker, J. J. V. Bavel, C. S. Wang, and J. N. Druckman (2020). Political Sectarianism in America. *Science 370*(6516), 533–536.

Fiorina, M. P. (1976). The Voting Decision: Instrumental and Expressive Aspects. *The Journal of Politics 38*(2), 390–413.

Fischer, C. S. (1982). *To Dwell Among Friends: Personal Networks in Town and City.* Chicago, IL: University of Chicago Press.

Fortier, J. and C. Stewart, III (2021). Lessons Learned in the 2020 Election: Report by AEI and the MIT Election Lab to the U.S. Election Assistance Commission. American Enterprise Institute and MIT Election Data and Science Lab.

Fotheringham, S. A. and M. E. O'Kelly (1998). *Spatial Interaction Models: Formulations and Applications*. Dordecht, NL: Kluwer Academic.

Fowler, C. S., N. Frey, D. C. Folch, N. Nagle, and S. Spielman (2020). Who are the People in my Neighborhood?: The "Contextual Fallacy" of Measuring Individual Context with Census Geographies. *Geographical Analysis 52*(2), 155–168.

Fox, J. and R. Andersen (2006). Effect Displays for Multinomial and Proportional-Odds Logit Models. *Sociological Methodology 36*(1), 225–255.

Fox, J. and S. Weisberg (2019). *An R Companion to Applied Regression* (3rd ed.). Thousand Oaks CA: SAGE Publications Inc.

Frey, C. B., T. Berger, and C. Chen (2018). Political machinery: did robots swing the 2016 US presidential election? *Oxford Review of Economic Policy 34*(3), 418–442.

Friedman, T. (2005). *The World is Flat: A Brief History of the Twenty-First Century*. New York, NY: Farrar, Straus, and Giroux.

Funk, P. (2010). Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System. *Journal of the European Economic Association 8*(5), 1077–1103.

Gentzkow, M. (2016). Polarization in 2016. White Paper, Toulouse Network for Information Technology.

Geraci, A., M. Nardotto, T. Reggiani, and F. Sabatini (2022). Broadband Internet and Social Capital. *Journal of Public Economics 206*, 104578.

Geverdt, D. (2018). Education Demographic and Geographic Estimates Program (EDGE): Locale Boundaries File Documentation, 2017 (NCES 2018-115). Department of Education. Washington, DC: National Center for Education Statistics.

Geverdt, D. (2019). Education Demographic and Geographic Estimates Program (EDGE): ZIP Code Tabulation Area (ZCTA) Locale Assignments File Documentation (NCES2018-077). Department of Education. Washington, DC: National Center for Education Statistics.

Gibbons, S. and H. G. Overman (2012). Mostly Pointless Spatial Econometrics? *Journal of Regional Science 52*(2), 172–191.

Gibbons, S., H. G. Overman, and E. Patacchini (2015). Spatial Methods. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 115–168. Amsterdam, NL: Elsevier.

Gimpel, J. and J. Schuknecht (2003). Political Participation and the Accessibility of the Ballot Box. *Political Geography 22*(5), 471–488.

Gimpel, J. G. and I. S. Hui (2015). Seeking Politically Compatible Neighbors? The Role of Neighborhood Partisan Composition in Residential Sorting. *Political Geography 48*, 130–142.

Gimpel, J. G. and A. Reeves (2022). Electoral Geography, Political Behavior and Public Opinion. In T. J. Rudolph (Ed.), *Handbook on Politics and Public Opinion*. Northampton, MA: Edward Elgar Publishing.

Glaeser, E. L. (2010). *Agglomeration economics*. University of Chicago Press.

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 201–233.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology 78*(6), 1360–1380.

Guriev, S. and E. Pappaioannou (2021). The Political Economy of Populism. *Journal of Economic Literature 59*(3).

Halleck Vega, S. and J. P. Elhorst (2015). The SLX Model. *Journal of Regional Science 55*(3), 339–363.

Hanmer, M., P. S. Herrnson, and C. Smith (2015). The Impact of E-mail on the Use of New Convenience Voting Methods and Turnout by Overseas Voters: A Field Experiment to Address Their Challenges with New Technology. *Election Law Journal: Rules, Politics, and Policy 14*(2), 97–110.

Hansford, T. G. and B. T. Gomez (2015). Reevaluating the Sociotropic Economic Voting Hypothesis. *Electoral Studies 39*, 15–25.

Harris, R. (2023). A Tale of Four Cities: Neighbourhood Diversification and Residential Desegregation in and Around England's 'No Majority' Cities. *The Geographical Journal*.

Harris, R. and M. Charlton (2016). Voting Out of the European Union: Exploring the Geography of Leave. *Environmenft and Planning A: Economy and Space 48*(11), 2116–128.

Hassell, H. J. G. (2017). Teaching Voters New Tricks: The Effect of Partisan Absentee Vote-by-Mail Get-Out-the-Vote Efforts. *Research & Politics 4*(1), 2053168017694806.

Hays, J., J. Lim, and J.-J. Spoon (2019). The Path from Trade to Right-Wing Populism in Europe. *Electoral Studies 60*, 102038.

Herrnson, P. S., M. J. Hanmer, and H. Y. Koh (2019). Mobilization Around New Convenience Voting Methods: A Field Experiment to Encourage Voting by Mail with a Downloadable Ballot and Early Voting. *Political Behavior 41*(4), 871–895.

Herrnson, P. S., M. J. Hanmer, M. Weil, and R. Orey (2022). The Impact of the COVID-19 Pandemic on Election Administration, Voting Options, and Turnout in the 2020 US Election. *Publius: The Journal of Federalism 52*(3), 452–475.

Herrnson, P. S. and C. Stewart (2023). The Impact of COVID-19, Election Policies, and Partisanship on Voter Participation in the 2020 U.S. Election. *Election Law Journal: Rules, Politics, and Policy 22*(2), 129–144.

Hodgson, Q. E., J. Kavanagh, A. Garg, E. W. Chan, and C. Sovac (2021). Options for Ensuring Safe Elections. RAND Corporation.

Holtz, D., M. Zhao, S. G. Benzell, C. Y. Cao, M. A. Rahimian, J. Yang, J. Allen, A. Collis, A. Moehring, T. Sowrirajan, D. Ghosh, Y. Zhang, P. S. Dhillon, C. Nicolaides, D. Eckles, and S. Aral (2020). Interdependence and the Cost of Uncoordinated Responses to COVID-19. *Proceedings of the National Academy of Sciences 117*(33), 19837–19843.

Hopkins, D. (2017). *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics.* New York, NY: Cambridge University Press.

Huckfeldt, R. and J. Sprague (1991). Discussant Effects on Vote Choice: Intimacy, Structure, and Interdependence. *The Journal of Politics 53*(1), 122–158.

Huckfeldt, R. R. (1982). *Politics in Context: Assimilation and Conflict in Urban Neighborhoods.* New York, NY: Agathon Press.

Huckfeldt, R. R. (1983). Social Contexts, Social Networks, and Urban Neighborhoods: Environmental Constraints on Friendship Choice. *American Journal of Sociology 89*(3), 651–669.

Huckfeldt, R. R. and J. Sprague (1995). *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign.* New York, NY: Cambridge University Press.

Iammarino, S., A. Rodriguez-Pose, and M. Storper (2018). Regional Inequality in Europe: Evidence, Theory and Policy Implications. *Journal of Economic Geography 19*(2), 273–298.

Ioannides, Y. M. (2013). *From Neighborhoods to Nations: The Economics of Social Interactions.* Princeton, NJ: Princeton University Press.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science 22*(1), 129–146.

Iyengar, S. and S. J. Westwood (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science 59*(3), 690–707.

Jargowsky, P. (2020). *Racial and Economic Segregation in the US: Overlapping and Reinforcing Dimensions*, pp. 151–168. Northampton, MA: Edward Elgar Publishing.

Johnston, R. (1983). The Neighbourhood Effect Won't Go Away: Observations on the Electoral Geography of England in the Light of Dunleavy's Critique. *Geoforum 14*(2), 161–168.

Johnston, R., D. Manley, K. Jones, and R. Rohla (2020). The Geographical Polarization of the American Electorate: A Country of Increasing Electoral Landslides? *GeoJournal 85*(1), 187–204.

Johnston, R. and C. Pattie (2000). People Who Talk Together Vote Together: An Exploration of Contextual Effects in Great Britain. *Annals of the Association of American Geographers 90*(1), 41–66.

Johnston, R. and C. Pattie (2011). *Putting Voters in their Place: Geography and Elections in Great Britain.* New York, NY: Oxford University Press.

Johnston, R. and C. Pattie (2014). Social Networks, Geography and Neighbourhood Effects. In *The SAGE Handbook of Social Network Analysis*, London, UK. SAGE Publications Ltd.

Katz, E. and P. F. Lazarsfeld (1955). *Personal Influence: The Part Played by People in the Flow of Mass Communications.* Glencoe, IL: Free Press.

Key, V. (1949). *Southern Politics in State and Nation.* New York, NY: Alfred A. Knopf.

Kiewiet, D. R. and M. S. Lewis-Beck (2011). No Man is an Island: Self-Interest, the Public Interest, and Sociotropic Voting. *Critical Review 23*(3), 303–319.

Kinder, D. R. and D. R. Kiewiet (1981). Sociotropic Politics: the American Case. *British Journal of Political Science 11*(2), 129–161.

King, G. (1996). Why Context Should Not Count. *Political Geography 15*, 159–164.

Kinsella, C., C. McTague, and K. Raleigh (2021). Closely and Deeply Divided: Purple Counties in the 2016 Presidential Election. *Applied Geography 127*, 102386.

Kinsella, C., C. McTague, and K. N. Raleigh (2015). Unmasking Geographic Polarization and Clustering: A Micro-Scalar Analysis of Partisan Voting Behavior. *Applied Geography 62*, 404–419.

Kousser, T., S. Hill, M. Lockhart, J. L. Merolla, and M. Romero (2021). How Do Americans Want Elections to Be Run During the COVID-19 Crisis? *Research & Politics 8*(2).

Kuchler, T., Y. Li, L. Peng, J. Stroebel, and D. Zhou (2021). Social Proximity to Capital: Implications for Investors and Firms. *The Review of Financial Studies 35*(6), 2743–2789.

Kuchler, T., D. Russel, and J. Stroebel (2022). The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook. *Journal of Urban Economics 127*, 103314.

Kwan, M.-P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers 102*(5), 958–968.

Lan, T., J. Kandt, and P. Longley (2020). Geographic scales of residential segregation in english cities. *Urban Geography 41*(1), 103–123.

Lavoratori, K. and D. Castellani (2021). Too Close for Comfort? Microgeography of Agglomeration Economies in the United Kingdom. *Journal of Regional Science 61*(5), 1002–28.

LeSage, J. P. (2014). What Regional Scientists Need to Know about Spatial Econometrics. *Review of Regional Studies 44*(1), 13–32.

Levy, A., F. Sánchez, and A. Kolieboi (2021). Anxieties of Voting-by-Mail in 2020: Contending with the Past, the Pandemic, and Political Change. Center for Civic Design.

Li, X. and L. Anselin (2023). *rgeoda: R Library for Spatial Data Analysis.* R package version 0.0.1.

Liu, L., A. Kuo, and J. Fernandez-Albertos (2020). Economic Crisis, Social Networks, and Political Preferences. *Socio-Economic Review 20*(1), 199–231.

Lockhart, M., S. J. Hill, J. Merolla, M. Romero, and T. Kousser (2020). America's Electorate Is Increasingly Polarized along Partisan Lines about Voting by Mail during the COVID-19 Crisis. *Proceedings of the National Academy of Sciences 117*(40), 24640–24642.

Logan, J. R. and G. D. Spitze (1994). Family Neighbors. *American Journal of Sociology 100*(2), 453–476.

Makridis, C. A. (2022). The Social Transmission of Economic Sentiment on Consumption. *European Economic Review 148*, 104232.

Malgouyres, C. (2017). Trade Shocks and Far-Right Voting: Evidence from French Presidential Elections. Robert Schuman Centre for Advanced Studies Research Paper 2017-21.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies 60*(3), 531–542.

Manski, C. F. (2000). Economic Analysis of Social Interactions. *The Journal of Economic Perspectives 14*(3), 115–136.

Martin, G. J. and S. W. Webster (2020). Does Residential Sorting Explain Geographic Polarization? *Political Science Research and Methods 8*(2), 215–231.

Martin, R., B. Gardiner, A. Pike, P. Sunley, and P. Tyler (2021). *Levelling up Left Behind Places: The Scale and Nature of the Economic and Policy Challenge.* Abington, UK: Routledge.

Mason, L. (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science 59*(1), 128–145.

McCann, P. (2020). Perceptions of Regional Inequality and the Geography of Discontent: Insights from the UK. *Regional Studies 54*(2), 256–267.

McCann, P. and R. Ortega-Argilés (2021). The UK 'Geography of Discontent': Narratives, Brexit and Inter-Regional 'Levelling Up'. *Cambridge Journal of Regions, Economy and Society 14*(3), 545–564.

McGhee, E. (2020). Partisan Gerrymandering and Political Science. *Annual Review of Political Science 23*(1), 171–185.

McGhee, E., J. Paluch, and M. Romero (2021). Vote-by-Mail and Voter Turnout in the Pandemic Election. Public Policy Institute of California.

McGhee, E., J. Paluch, and M. Romero (2022). Vote-By-Mail Policy and the 2020 Presidential Election. *Research & Politics 9*(2), 20531680221089197.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology 27*(1), 415–444.

Mettler, S. and T. Brown (2022). The Growing Rural-Urban Political Divide and Democratic Vulnerability. *The ANNALS of the American Academy of Political and Social Science 699*(1), 130–142.

Mian, A., K. Rao, and A. Sufi (2013). Household Balance Sheets, Consumption, and the Economic Slump. *The Quarterly Journal of Economics 128*(4), 1687–1726.

Mian, A. and A. Sufi (2014). What Explains the 2007–2009 Drop in Employment? *Econometrica 82*(6), 2197–2223.

Miller, W. L. (1977). *Electoral Dynamics in Britain since 1918.* London, UK: MacMillan Press Ltd.

Misra, J. (2019). Behind the 2018 U.S. Midterm Election Turnout: Voter Turnout Rates Among All Voting Age and Major Racial and Ethnic Groups Were Higher Than in 2014. U.S. Census Bureau.

Mok, D., R. Basu, and B. Wellman (2007). Did Distance Matter Before the Internet?: Interpersonal Contact and Support in the 1970s. *Social Networks 29*(3), 430–461.

Mok, D., B. Wellman, and J. Carrasco (2010). Does Distance Matter in the Age of the Internet? *Urban Studies 47*(13), 2747–2783.

Mummolo, J. and C. Nall (2017). Why Partisans Do Not Sort: The Constraints on Political Segregation. *The Journal of Politics 79*(1), 45–59.

Mummolo, J. and E. Peterson (2018). Improving the Interpretation of Fixed Effects Regression Results. *Political Science Research and Methods 6*(4), 829–835.

Munshi, K. (2020). Social Networks and Migration. *Annual Review of Economics 12*(1), 503–524.

Mutz, D. C. (2002). Cross-Cutting Social Networks: Testing Democratic Theory in Practice. *The American Political Science Review 96*(1), 111–126.

Myers, A. S. (2013). Secular Geographical Polarization in the American South: The Case of Texas, 1996–2010. *Electoral Studies 32*(1), 48–62.

Nall, C. (2018). *The Road to Inequality: How the Federal Highway Program Polarized America and Undermined Cities.* New York, NY: Cambridge University Press.

National Conference of State Legislatures (2022a). Election Administration at State and Local Levels.

National Conference of State Legislatures (2022b). Voting Outside the Polling Place: Absentee, All-Mail and Other Voting at Home Options.

National Conference of State Legislatures (2023). Access to and Use of Voter Registration Lists.

Noury, A., A. François, O. Gergaud, and A. Garel (2021). How Does COVID-19 Affect Electoral Participation? Evidence from the French Municipal Elections. *PLOS One 16*(2).

Office for Communications (2020). Adults' Media Use and Attitudes Report 2020.

Office for National Statistics (1995). . United Kingdom Input-Output Analytical Tables, 1995.

Orey, R. (2021). How Data Can Solve the Elections Resource Allocation Problem. Bipartisan Policy Center.

Orford, S., C. Railings, M. Thrasher, and G. Borisyuk (2011). Changes in the Probability of Voter Turnout when Resiting Polling Stations: a Case Study in Brent, UK. *Environment and Planning C: Government and Policy 29*(1), 149–169.

Orford, S., C. Rallings, M. Thrasher, and G. Borisyuk (2008). Investigating Differences in Electoral Turnout: The Influence of Ward-Level Context on Participation in Local and Parliamentary Elections in Britain. *Environment and Planning A: Economy and Space 40*(5), 1250–1268.

Pettigrew, T. F. and L. R. Tropp (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology 90*(5), 751.

Picchio, M. and R. Santolini (2022). The COVID-19 Pandemic's Effects on Voter Turnout. *European Journal of Political Economy 73*, 102161.

Pierce, J. R. and P. K. Schott (2020). Trade Liberalization and Mortality: Evidence from US Counties. *American Economic Review: Insights 2*(1), 47–64.

Plescia, C., S. Sevi, and A. Blais (2021). Who Likes to Vote by Mail? *American Politics Research 49*(4), 381–385.

Prenner, C., B. Fox, and C. Kenny (2022). *censuxy: Access the U.S. Census Bureau's Geocoding API System.* R package version 1.1.1.

Rajan, R. (2020). *The Third Pillar: How Markets and the State Leave the Community Behind.* London, UK: William Collins.

Riker, W. H. and P. C. Ordeshook (1968). A Theory of the Calculus of Voting. *The American Political Science Review 62*(1), 25–42.

Rodden, J. (2010). The Geographic Distribution of Political Preferences. *Annual Review of Political Science 13*(1), 321–340.

Rodden, J. (2015). *Geography and Gridlock in the United States*, pp. 104–120. Cambridge University Press.

Rodden, J. (2019). *Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide.* New York, NY: Cambridge University Press.

Rodrik, D. (2021). Why Does Globalization Fuel Populism? Economics, Culture, and the Rise of Right-Wing Populism. *Annual Review of Economics 13*(1), 133–170.

Rodríguez-Pose, A. (2018). The Revenge of the Places that Don't Matter (and What to Do About It). *Cambridge Journal of Regions, Economy and Society 11*(1), 189–209.

Rohla, R., R. Johnston, K. Jones, and D. Manley (2018). Spatial Scale and the Geographical Polarization of the American Electorate. *Political Geography 65*, 117–122.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's Trending in Difference-in-Differences? Synthesis of the Recent Econometrics Literature. *Journal of Econometrics 235*(2), 2218–2244.

Sandbu, M. (2022). *The Economics of Belonging.* Princeton, NJ: Princeton University Press.

Sanderson, E. and F. Windmeijer (2016). A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables. *Journal of Econometrics 190*(2), 212–221.

Santoro, L. R. and P. A. Beck (2017). Social Networks and Vote Choice. In *The Oxford Handbook of Political Networks.* New York, NY: Oxford University Press.

Scala, D. J. and K. M. Johnson (2017). Political Polarization Along the Rural-Urban Continuum? The Geography of the Presidential Vote, 2000–2016. *The ANNALS of the American Academy of Political and Social Science 672*(1), 162–184.

Scheller, D. S. (2021). Pandemic Primary: The Interactive Effects of COVID-19 Prevalence and Age on Voter Turnout. *Journal of Elections, Public Opinion and Parties 31*, 180–190.

Schelling, T. C. (1971). Dynamic Models of Segregation. *The Journal of Mathematical Sociology 1*(2), 143–186.

Shineman, V. A. (2018). If You Mobilize Them, They Will Become Informed: Experimental Evidence that Information Acquisition Is Endogenous to Costs and Incentives to Participate. *British Journal of Political Science 48*(1), 189–211.

Smith, K. and D. E. Sylvester (2013). Is It the Message or the Person? Lessons from a Field Experiment About Who Converts to Permanent Vote by Mail. *Election Law Journal: Rules, Politics, and Policy 12*(3), 243–260.

Sprague, D. A. and T. House (2017). Evidence for Complex Contagion Models of Social Contagion from Observational Data. *PLOS One 12*(7).

Steiner, N. D. and P. Harms (2021). Trade Shocks and the Nationalist Backlash in Political Attitudes: Panel Data Evidence from Great Britain. *Journal of European Public Policy*, 1–20.

Stewart, III, C. (2011). Adding up the Costs and Benefits of Voting by Mail. *Election Law Journal 10*(3), 297–301.

Stewart, III, C. (2020). How We Voted in 2020: A Topical Look at the Survey of the Performance of American Elections. MIT Election Data and Science Lab.

Stewart, III, C. (2021). Voting amid COVID: The Response of Voters to Voting during the Pandemic. MIT Political Science Department Research Paper 2022-1.

Takhteyev, Y., A. Gruzd, and B. Wellman (2012). Geography of Twitter Networks. *Social Networks 34*(1), 73–81.

Thompson, D. F. (2004). Election Time: Normative Implications of Temporal Properties of the Electoral Process in the United States. *American Political Science Review 98*(1), 51–63.

Thompson, D. M., J. A. Wu, J. Yoder, and A. B. Hall (2020). Universal Vote-by-Mail Has No Impact on Partisan Turnout or Vote Share. *Proceedings of the National Academy of Sciences 117*(25), 14052–14056.

Tingsten, H. (1937). *Political Behavior: Studies in Electoral Statistics.* Totonwa, NJ: Bedminster (translation in 1963).

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography 46*, 234–240.

Tobler, W. R. (1999). Linear Pycnophylactic Reallocation Comment on a Paper by D. Martin. *International Journal of Geographical Information Science 13*(1), 85–90.

Tolbert, C. M. and M. Sizer (1996). US Commuting Zones and Labor Market Areas: A 1990 Update. Staff Paper AGES-9614, Economic Research Service, Rural Economy Division.

Törnberg, P. (2018). Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion. *PLOS One 13*(9).

U.S. Election Assistance Commission (2010). *Election Management Guidelines*, Chapter 7: Absentee Voting and Vote by Mail, pp. 49–58.

U.S. Election Assistance Commission (2020a). Lessons Learned from the 2020 Primary during COVID-19. Joint COVID Working Group of the Cybersecurity and Infrastructure Security Agency and the Elections Infrastructure Government and Sector Coordinating Councils.

U.S. Election Assistance Commission (2020b). Preparing for In-Person Voting during COVID-19: Resource Allocation. Joint COVID Working Group of the Cybersecurity and Infrastructure Security Agency and the Elections Infrastructure Government and Sector Coordinating Councils.

Vogels, E. A. and M. Anderson (2021). Social Media Use in 2021. Pew Research Center.

Vogels, E. A., B. Auxier, and M. Anderson (2021). Partisan Differences in Social Media Use Show Up for Some Platforms, but Not Facebook. Pew Research Center.

Voting and Election Science Team (2020). *2020 Precinct-Level Election Results.* Harvard Dataverse. Dataset version 40.

Walker, K. and M. Herman (2023). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames.* R package version 1.4.3.

Walker, K. and B. Rudis (2023). *tigris: Load Census TIGER/Line Shapefules.* R package version 2.0.1.

Wellman, B. (1979). The Community Question: The Intimate Networks of East Yorkers. *American Journal of Sociology 84*(5), 1201–1231.

Wellman, B. (1996). Are Personal Communities Local? A Dumptarian Reconsideration. *Social Networks 18*(4), 347–354.

Wellman, B. and S. Potter (1999). *Networks in the Global Village.* New York, NY: Routledge.

Wellman, B., A. Quan-Haase, J. Boase, W. Chen, K. Hampton, I. Díaz, and K. Miyata (2003). The Social Affordances of the Internet for Networked Individualism. *Journal of Computer-Mediated Communication 8*(3), JCMC834.

Wheeler, J. O. and F. P. Stutz (1971). Spatial Dimensions of Urban Social Travel. *Annals of the Association of American Geographers 61*(2), 371–386.

Williams, M. L., A. Sutherland, V. Roy-Chowdhury, T. Loke, A. Cullen, L. Sloan, P. Burnap, and P. Giannasi (2022). The Effect of the Brexit Vote on the Variation in Race and Religious Hate Crimes in England, Wales, Scotland and Northern Ireland. *The British Journal of Criminology 63*(4), 1003–1023.

Wilson, R. (2022). The Impact of Social Networks on EITC Claiming Behavior. *The Review of Economics and Statistics 104*(5), 929–945.

Wolf, L. J. et al. (2022). Confounded Local Inference: Extending Local Moran Statistics to Handle Confounding. *SocArXiv*.

Wolfinger, R. E. and S. J. Rosenstone (1980). *Who Votes?* New Haven, CT: Yale University Press.

Yoder, J., C. Handan-Nader, A. Myers, T. Nowacki, D. M. Thompson, J. A. Wu, C. Yorgason, and A. B. Hall (2021). How Did Absentee Voting Affect the 2020 U.S. Election? *Science Advances 7*(52), eabk1755.

Zhao, M., D. Holtz, and S. Aral (2021). Interdependent Program Evaluation: Geographic and Social Spillovers in COVID-19 Closures and Reopenings in the United States. *Science Advances 7*(31), eabe7733.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics 12*(1), 415–438.