Essays in the Spatial Economic Analysis of Social Interactions

Andreas Diemer

Department of Geography and Environment

London School of Economics

A thesis submitted to the Department of Geography and Environment
of the London School of Economics for the Degree of Doctor of Philosophy

London, July 2020

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 49,000 words, excluding all appendices and the bibliography.

Acknowledgements

Despite the many hours dedicated to researching social interactions, perhaps the strongest evidence I have found for their importance is the conviction, at the end of this journey, that I could not have completed my PhD without the support of the many people around me.

I am indebted to Michael Storper and Simona Iammarino for their outstanding supervision. Michael's inspiring curiosity and enthusiasm, and Simona's meticulous support and reassuring presence, carried me through this experience. Their guidance was invaluable. I am also grateful to Olmo Silva, for his precious advice, and for the time he found to share his knowledge and insights with me. A debt of thanks also goes to the other members of the faculty whose comments during departmental seminars or meetings enriched my work: Gabriel Ahlfeldt, Felipe Carozzi, Steve Gibbons, Vernon Henderson, Christian Hilber, Neil Lee, Henry Overman, and Lindsay Relihan, among many others. A special mention to Riccardo Crescenzi, for reviewing early drafts of Chapters 2 and 3, and to Andrés Rodríguez-Pose, along with my supervisors, for extending the valuable opportunity to collaborate on a project for the European Commission.

For Chapter 3, I am particularly grateful to Tanner Regan for long and fruitful exchanges. Thank you also to Francesco Lissoni, Adam Jaffe, and Jorge Pérez Pérez for discussing my work. In addition, I greatly benefited from comments and conversations with participants at the EPFL Virtual Innovation Seminar, the 5^{th} Geography of Innovation Conference in Stavanger, the RSA North America Conference in Montréal, the 2^{th} Regional and Urban Economics Workshop in Bogota, the 9^{th} European Meeting of the Urban Economics Association in Amsterdam, the 6^{th} Workshop in Economics of Innovation, Complexity and Knowledge in Turin, and the 2^{th} ICUBERD Conference in Pécs.

Chapter 4 benefited from helpful comments and suggestions received at the 7^{th} Central European Conference in Regional Science in Sopron, the 24^{th} Annual Workshop on the Economic Science with Heterogeneous Interacting Agents in London, and the RSA Annual Conference in Santiago de Compostela.

Financial support from the UK Economic and Social Research Council is gratefully acknowledged. I would also like to thank Michael Bailey at Facebook for providing access to the Social Connectedness data.

Many friendships sustained me along the way, with both those in my cohort, who shared the pleasures and pains of the PhD experience from the very beginning, and those from other cohorts who preceded or followed me in this journey. Special thanks to Tanner, for the numerous and inspiring cups of coffee we shared, Matt, for being there pretty much every step of the way, Lindsey and Louise, for always lifting me up (sometimes literally) with a quality climbing session. Thank you to the Bankside community, and especially Paroj, for helping me feel at home in London. Thank you also to the old friends, in particular Federico, Fabio, and Michele, for motivating me and for having my back. I am most grateful to my family, for their support and guidance from the very start, and for believing I could do this.

Finally, and above all, my gratitude goes to my wife, Tehminah, who encouraged me to pursue a PhD in the first place. I dedicate this work to you. You make it all worthwhile.

Abstract

This thesis examines the role of social interactions in economic geography from several different angles. It draws on and dialogues with literatures in related fields such as spatial and urban economics, regional science, economic sociology, and innovation economics, to explore how the geographical and social spaces are interlinked. The thesis comprises an introduction and three essays, all focused on the United States. The first essay considers the notion of social capital from a territorial perspective and investigates the role of manufacturing decline in its accumulation. It documents a positive relationship between the two, but also highlights significant challenges in the stability and interpretation of this result. The essay thus questions how well the notion of social capital lends itself to measurement and empirical analysis. The second essay uses a direct and broad measure of the social connectedness of regions to examine its role in transferring knowledge across the entire US geography. It uncovers a small yet significant and robust effect of social connection on knowledge flows as proxied by patent citations. The effect matters above and beyond the pre-existing geography of production and the professional networks of inventors. The third and final essay uses US social connectedness data to investigate how plausibly exogenous surges in the local demand for jobs in the oil and gas industry during the 'fracking boom' can affect the economy of spatially distant but socially proximate places. Findings support a role for social interaction in the diffusion of local economic shocks. This effect is likely explained by the relocation of transient workers within the industry, providing new aggregate evidence in support of the literature on job information networks. The overriding contribution of this thesis is to underscore with new empirical evidence the importance of social interactions in the spatial distribution of economic activity, not just locally but also over large scale geographies.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis comprises three empirical essays that examine the role of social interactions in economic geography from several different angles. It draws on and contributes to literatures in spatial and urban economics, regional science, economics of innovation, and economic sociology. The common theme to all three essays is the attention paid to the spatial nature of social interactions and their effects. In the first essay, I consider the notion of social capital, which the discipline of economics borrows from sociology to describe how interaction gives access to resources embedded in interpersonal networks. I review the geography of social capital, and investigate the role of manufacturing decline in its accumulation. In the second and third essays, I consider the notion of social connectedness, which describes linkages across places formed from interactions between their residents. I rely on a novel and direct measure of social connectedness based on the universe of friendship ties revealed on Facebook, a popular social media platform. With this measure, I document how interactions between places contribute to the diffusion of knowledge and economic shocks even over long distances. Below, I introduce the notion of social interactions and how it relates to economic geography, understood as a discipline that applies spatial thinking to the study of the economy. Section 1.3 summarises the three empirical chapters of this thesis. Section 1.4 outlines its contribution.

1.1 Conceptual and Empirical Considerations

Economics was once described as the most 'undersocialised' social science (Granovetter, 1985). This critique took aim at the conceptualisation of individuals in neoclassical theory as rational agents whose decisions are unaffected by their social environment. Over the years, however, economics has incorporated the notion of social interactions in the study of resource allocation, both in theory and in empirical applications. Institutional economics emphasised the importance of formal and informal institutions in shaping human interaction by setting incentives and defining the 'rules of the game in society' (North, 1990). At the macro level, endogenous growth theory and new economic geography acknowledge the importance of externalities and spatial linkages in determining economic development (Romer, 1986, 1990; Lucas, 1988; Krugman, 1991b; Krugman and Venables, 1995). In microeconomics, research strands in labour, urban, regional, and social economics consider the role of peers and neighbours in determining outcomes of individuals and places (Topa and Zenou, 2015).

But what are social interactions, and how are they relevant to understanding economic behaviour? The concept itself is quite intuitive. Social interactions are instances where the behaviour of an individual relates to that of someone else, or generally depends on contextual factors. There are many examples of interactions in daily life, so much so that we often engage in them "without knowing anything about it" (Ioannides, 2013, p. 1). Learning from classmates or colleagues, picking up a new hobby, getting involved in community work, or deciding what show to watch next on TV are all instances where behaviours are interdependent across individuals. Moreover, interactions take many forms. In markets, prices provide an indirect mechanism revealing information on all individual interdependences. This is the traditional domain of economics. However, many interactions are not mediated by markets and represent externalities. The behaviour of an agent depends on that of another without this relationship perforce being reflected in prices. The incorporation of non-market interactions into rigorous economic thinking has been one of the main developments in economics over the past three decades. In his seminal article *Economic Analysis of Social Interactions*, Manski (2000) discusses the economic notion of interactions and outlines empirical challenges for their identification. Throughout my thesis, I refer to this article for its conceptual clarity. It is thus useful to review some of its key points.

1.1.1 Mechanics of Interaction

Manski isolates three channels through which the actions of one agent might affect those of another. These are preferences, expectations, and constraints. Interaction through preferences occurs when agents' ordering of alternative courses of actions depends on the choice of other agents because they affect utility associated with each action. For example, deciding to attend a concert may depend on friends agreeing to also join the event. At times, the repeated nature of such choices links with the emergence of social norms, which determine what behaviours society deems appropriate and what it instead sanctions. Such is the practice of forming queues in shops, or the convention of reciprocating the purchase of a round of drinks at a British pub. Preference interaction may also stem from empathy, as a result of the utility drawn from observing the good state of others.[1]

Interaction may also occur through expectations, when an agent forms a belief about the pay-off to some choice based on the observation of group behaviour. For example, the presence of broken windows across a neighbourhood may suggest that crime goes undetected, in turn encouraging further crimes.[2] In financial markets, investors may believe that other agents reveal private information on a certain firm if they observe its stocks price increase. This can result in herd behaviour and irrational exuberance. Another consequence of expectation interaction can be statistical discrimination. According to this theory, employers form beliefs on the unobserved performance of an individual based on the perceived historical performance associated with the group that individual belongs to. This may in itself be due to discrimination, leading to a self-reinforcing process that can contribute to explaining the persistence of racial or gender inequalities in labour markets.

Finally, interaction can take place via the constraints implied by other agents' choices. Congestion is a common example. When many drivers decide to use the same road network at the same time, the average travel speed decreases for all cars as a result. Likewise, access to a restaurant is limited by its seating capacity when an establishment is very popular. Moreover, underperforming students may impinge on the achievements of other pupils by diverting the attention of the teacher. In sum, social interactions can affect individual behaviour through different channels and, I should stress, not always in positive ways. Manski points out that this distinction is more than a theoretical exercise, as different mechanisms may require different policy responses. An intervention that provides information to agents, for instance, can be effective in addressing distortions arising from expectation interactions and observational learning. By contrast, it is unlikely to influence agents' behaviour when they are interacting based on preferences or constraints.

1.1.2 Identification Challenges

Concerning empirics, Manski distinguishes between three observationally equivalent explanations for the correlation of behaviour within groups of agents: endogenous interactions, contextual or exogenous interactions, and correlated effects. Endogenous interaction effects arise when the behaviour of an individual is a function of the behaviour of other

[1] This is already recognised by Adam Smith in his *Theory of Moral Sentiments* (1869, p. 2): "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it."

[2] This hypothesis is known in sociology and criminology as the 'broken window theory'.

individuals in the reference group. Contextual effects are present when the behaviour of an agent varies with the exogenous characteristics of the reference group. By contrast, correlated effects refer to the exposure of individuals in the group to similar institutional environments, or the similarity of individual characteristics in the group (for instance, due to sorting). To clarify, it is useful to consider an example from peer effects in education. The correlation of student grades across classmates could be because a pupil's achievement motivates another to work harder (an endogenous effect), or it could be that the class composition facilitates learning (a contextual effect, for instance, related to gender balance). However, the quality of the class' teacher or the fact that students in the class share similar socio-economic backgrounds could also explain the observed correlation in grades (both correlated effects).

The econometrician's task is to separately identify each of these explanations, or at least to isolate the effect of social interactions proper (whether endogenous or contextual) from the presence of correlated effects and sorting. In the latter cases, the correlation in group behaviours is spurious. In addition, correct identification of endogenous effects has important implications for policy, as only the latter give rise to social multipliers. These exist when exposing an individual to an intervention has a knock-on effect on other individuals because of the reciprocal changes in behaviour induced by the treatment. This magnifies the outcomes of a policy when it targets the group.[3] However, in the absence of experimental data, or of information on the structure of all interactions, it is difficult to isolate endogenous and contextual effects due to what is known as the 'reflection problem' (Manski, 1993). Akin to simultaneity and reverse causality in traditional econometrics, the reflection problem refers to the difficulty of distinguishing between the influence of average group behaviour on the individual from that of the individual's on the group average (which arises mechanically, due to participation to the group). This is further complicated by the fact that an agent's behaviour is likely a function of his or her characteristics. Therefore, separately identifying endogenous effects from contextual ones is an almost insurmountable challenge. At the very least, however, the applied researcher should avoid misapprehending correlated effects as evidence for the relevance of social interactions. Considerable attention is given to this empirical issue throughout the dissertation.

[3] Assuming the endogenous effects are sufficiently strong and non-linear across targeted and non-targeted units so that the benefits of concentrating the treatment onto one group (e.g., a community or a region) are greater than the costs imposed on the other.

1.2 Towards a Geography of Social Interactions

Setting aside the conceptual definition of social interactions, an overarching theme common to all empirical work on this subject is the importance of space in mediating interactions between economic agents. In abstract terms, there is a 'social space' that bounds the scope of the relevant reference group within which interactions takes place. In practice, the true reference group is typically unobserved. The applied researcher thus relies on various empirical definitions depending on the subject at hand. In education, as discussed above, it is usually classroom peers. In labour economics, it could be co-workers or residential neighbours. For example, colleagues may pass on information about job opportunities. Exposure to better neighbourhood environments at a young age may increase long term economic outcomes of those who grow up there. In urban and regional economics, the reference group could be an entire city or industrial district. Due to agglomeration economies, co-location of workers and firms in the same cluster may result in higher productivity. The high-density of urban environments favours social interactions, making cities particularly susceptible to these types of non-market exchanges.

Abstracting from these examples, reference groups are often defined based on geographical proximity, assuming that social interactions need face-to-face contact and that the cost of interacting increases with physical distance (Storper and Venables, 2004). Because of this interdependence with distance, economic geography and spatial economics have been at the forefront of applied research on social interactions, notably for local interdependencies. But individuals may also interact over longer distances, for instance through participation to professional networks, or thanks to advances in information and communication technologies. Thus, there is a sense in which the scope of interaction cannot be defined a priori purely based on geographical considerations. This is acknowledged in theoretical work, yet limited data availability has until recently constrained the empirical study of social interactions beyond the local level. My thesis contributes to this literature by marshalling novel measures that allow studying interactions even across large distances. In so doing, I aim to empirically investigate the sociological foundations of economic geography and spatial economics more in general, by offering new evidence to bear on the economic relevance of social interactions at different spatial scales. What follows gives an overview of this dissertation, providing summaries of each essay.

1.3 Overview of the Thesis

This thesis is composed of three independent essays, all focusing on the United States and carrying out applied econometric work. Each essay investigates a different aspect of social interactions. Chapter 2 looks at social capital and its production, Chapter 3 considers the role of interactions in knowledge flows, and Chapter 4 focuses on the diffusion of local economic shocks over social networks. The common denominator to each essay is the interest in the spatial dimension of interactions, and the attention paid to the econometric identification of the effects studied. Moreover, each chapter develops ideas and considers limitations highlighted in the previous ones. Considerable learning took place during the writing of this project, and I present chapters in the chronological order they were developed, hoping that the reader can get a sense of this progress.

1.3.1 First Essay

The first essay studies the production of social capital across local labour markets in the US, with a focus on manufacturing decline as a possible determinant. Broadly understood, social capital can be defined as access to resources embedded in community relations (Durlauf and Fafchamps, 2005; Dasgupta, 2005). The origins of this concept are in sociology and political science, but over the past three decades it has attracted the attention of economists interested in determining how non-market interactions can mitigate problems of imperfect competition, externalities, incomplete or asymmetric information, and agent coordination. Similarly, economic geographers and regional scientists have borrowed this notion to characterise certain intangible features of local communities such as informal institutions, emphasising the spatial and historically-determined dimension of this concept.

I explore the concept of social capital as the aggregate manifestation of social interactions taking place within a defined geographical area. My aim is twofold. First, I conduct a descriptive investigation into the claim by Putnam (1995, 2000) that social capital was declining across US communities, paying attention to the territorial dimension of changes in a newly constructed index for local labour markets in the US. Second, I test the hypothesis that observed trends can be related to the progressive deindustrialisation of the US economy by exploiting plausibly exogenous shocks to local manufacturing activity stemming from changes in foreign productivity and trade policy. Indeed, the persistent decline of manufacturing employment observed since at least the 1970s raised concerns about the effect these changes may have on American society.

I attest a general decrease in stocks of social capital before the turn of the century, consistent with previous research published at that time. However, I also find that social capital subsequently grew stronger across many local communities, countering the historic decline. Moreover, I find that local shocks to the manufacturing sector, captured by rising import penetration from China, are associated with an increase, rather than a decline, in social capital over 1990-2007. This is in line with a theory on the social insurance effect of social interactions proposed by Becker (1974). However, I discourage the reader from interpreting this result as something that is necessarily socially desirable. Moreover, I also highlight important limitations to my findings, suggesting that further research is needed to provide conclusive evidence on this topic. Overall, the essay questions how well the traditional notion of social capital lends itself to measurement and interpretation.

1.3.2 Second Essay

The second essay considers the role of social interactions in knowledge flows. I focus on the transmission of technical and scientific knowledge by relying on patent citations as a proxy for these flows. Citations provide a powerful measure of economically relevant knowledge exchange, otherwise difficult to observe in different settings. Moreover, they speak to the process of innovation and technological change, which is a key determinant of long run economic growth. Building on the results of the previous chapter, the work in Chapter 3 introduces a new measure for social interactions that directly captures ties across the entire US geography as revealed by the universe of online friendships on Facebook, a popular social media platform. Aggregated to the level of counties, this metric informs about the social connectedness of different places with one another due to interactions between their residents.

My aim is to empirically examine the role of social connectedness in the diffusion of knowledge among agents located across distant geographies. This line of enquiry relates to an old question in economics about the role of localised knowledge spillovers in promoting the agglomeration of people and industries in space (Marshall, 1890). The essay outlines micro level channels by which learning might occur, and specifically how social connectedness might provide non-agglomerative mechanisms for the transmission of knowledge unrestricted to physically proximate agents.

The identification strategy relies on matching inventor citations with citations from patent examiners (a feature of the patenting process), whose own social geography is orthogonal to the inventor's on the same patent. By exploiting examiner-added citations as a control group for knowledge flows, this work identifies a small but significant and robust effect of social connectedness between places on their propensity to cite one another. This is independent of geographical distance or professional linkages between inventors and takes into account the endogenous location of relevant knowledge due to the pre-existing geography of production potentially related to Marshallian forces other than learning (i.e., matching, or sharing). I also show that the relevance of informal social ties has increased over time, although this may relate to changes in measurement accuracy. Furthermore, effects appears to be stronger for entrepreneurs (firms patenting for the first time), for patents that are common domain in a geographical sense, and for knowledge more distant in the technological space.

1.3.3 Third Essay

The third essay also relies on the social connectedness measure discussed in the previous chapter, but considers a different type of econometric setting. I study how localised economic shocks can propagate across a country through interactions of people over social networks independent of the physical distance that separates them. In particular, I look at shocks associated with the ‘fracking revolution’ in the US, taking place since the early 2000s. I describe how these plausibly exogenous surges in the local demand for jobs in the oil and gas industry can affect the economy of spatially distant but socially proximate places. While the majority of extant literature has focused on spatial spillovers of localised shocks to nearby areas, the role of networks in this process is relatively understudied. As outlined in the introduction to this thesis, it is possible that social interactions between places geographically far apart play a role irrespective of the physical distance between them.

In line with existing evidence, I find that the largest effects of localised shocks are felt in geographically proximate areas. However, social networks do play a role. On average, a million dollar per capita increase in oil and gas extraction in the top 25 most strongly socially connected counties raises per capita wages by about 2,000 dollars for workers reporting their income in counties located as far as 1,000 km away from the drilling site. This novel result is consistent with accounts of the fracking industry that discuss the importance of out-of-state hires and transient workers. It also provides new aggregate evidence in support of the literature on job information networks. I also document that new oil and gas production has positive inward effects on wages and employment in socially connected counties mostly in mining and transportation industries, and to some extent in services, with some downward pressure on manufacturing jobs. Although there are some caveats in the causal interpretation of these cross-industry results, the findings are consistent with previous accounts in the literature, underscoring the relevance of social connectedness as a metric to study spatial diffusion.

1.4 Common Themes, Contribution, and Limitations

Each project included herein offers its own specific contributions to relevant literatures. I discuss these more in detail within each chapter, where I also clarify how my research addresses gaps in existing applied research. At the same time, it is possible to distil some general learning points from my essays, which speak to the contribution of this dissertation as a whole.

Two interrelated themes stand out: measurement of social interactions, and integration of different types of linkages between places in geographical research. An important takeaway from the project in Chapter 2 is that reliance on more direct and if possible disaggregated metrics of social interactions is advisable in future work. In particular, pure network-based measures may offer more fruitful ground for applied analyses. Their interpretation is more intuitive, and their empirical study can rely on the tools developed by literatures in spatial econometrics and social network analysis. Chapter 3 acknowledges this point and explores the use of a new measure of social interactions based on the universe of online friendships in the US, as revealed by connections on Facebook. In so doing, the chapter heeds the opportunities offered by new sources of data based on large scale records. This allows me to study the role of social interactions beyond the local level, for the entire geography of a country. It also demonstrates how these data can be integrated with existing measurement strategies to uncover new dimensions of linkages between people and places. Chapter 4 considers more closely the relationship between geographical and social spaces, by studying the diffusion of localised economic shocks over social networks. The evidence presented in this chapter is of relevance to policy makers interested in local economic development. If being socially connected to thriving places can benefit local economies above and beyond immediately contiguous areas, then this research sheds light onto the importance of considering a new dimension of access to opportunity, namely one that takes into account the interaction of people across distant geographies. Further, this analysis reveals the potential of spillovers of place-based interventions beyond contiguous areas, in a way that depends on the geography of social interactions.

There are also some limitations to this work. These too are discussed more in detail within each chapter. Relevant to mention at this stage is the necessity to trade breadth of geographical coverage with scale at which social interactions are measured. The inherent challenge I faced in this thesis was to strike a balance between studying relationships between individuals at the micro level and capturing the geographical profile of these relationships across an entire country, independent of physical distance. Consistently with the focus of this dissertation, I privileged the latter, which required a degree of aggregation in the measure used for social interactions. This entailed a loss of precision in empirical measurement, shifting the focus of analysis from people to places. It also somewhat curtailed opportunities for econometric identification of the empirical relationships of interest. Future applied work could rely on more disaggregated data, perhaps measured directly at the micro level, leveraging sources that provide information on both geographical location and social ties of individuals. With appropriate linkages and due consideration for preserving the anonymity of all records, private sector data, population registers, and selected surveys increasingly offer these opportunities. Accessing and combining these data is still quite difficult, but there is growing recognition of the value it provides. On a more conceptual level, I also wish to highlight the limitations inherent to studying social interactions from a purely quantitative perspective, as I do in this thesis. I do not mean to suggest that applied economic analysis can offer definitive answers. There is a broader debate in social sciences regarding the subjective motivations and deeper purposes of behaviour in society, and the epistemological aptness of quantitative research to answer such questions. Looking ahead, research on social interactions will greatly benefit from dialogues with other disciplines that consider the social determinants of human behaviour, such as sociology or social psychology.

In conclusion, the overriding contribution of this thesis is to underscore with new empirical evidence the importance of social interactions in the spatial distribution of economic activity. My research suggests that this is not just true locally but also over large scale geographies. This bears on policy-relevant debates in economic geography and related disciplines about why and how space matters in economic development. I hope that this thesis will offer opportunities to reflect on the use of new data sources to address old questions in economic geography and spatial economics.

Chapter 2

Manufacturing Decline and Social Capital Production

2.1 Introduction

Almost two centuries ago, French political scientist Alexis de Tocqueville, in diplomatic visit to the US, wrote in length about the rich community life he found on the other side of the Atlantic, describing Americans as a nation of joiners. In his essays, Tocqueville highlighted the virtues of associational activity and civic life for the economic success and political stability of the American democracy (de Tocqueville, 1835). Today, these ideas continue to be the subject of lively discussion in social sciences, and the term 'social capital' was coined to broadly refer to the associational forces that underlie the formation of communities and societies. The effects of social capital on various measures of economic development, as well as the determinants of social capital formation, continue to be topics of great interest among scholars. In recent times, however, several studies have described a declining trend in American social capital (Putnam, 1995, 2000; Costa and Kahn, 2001). Parallel to this fall, the US has also witnessed a steady loss of manufacturing employment, at least since the 1970s (Fort et al., 2018). This decline is not spread evenly across the country and, as will be shown, different areas have been affected in different periods of time. This also entails that the socio-economic consequences of this decline may display distinct spatial patterns, aligned with the geographic distribution of economic shocks.

What happens to local communities when manufacturing disappears? The aim of this research is to examine changes in social capital across the US territory, with a focus on testing whether the decline of manufacturing may contribute to explaining the observed dynamics. Social scientists have studied how the industrial transformation of an economy affects communities and societies since at least the work of Polanyi (1945), who argued that the rapid economic changes experienced by eighteenth century England during the Industrial Revolution, if not regulated, would disrupt societal cohesion. Jahoda et al. (1971) discuss the 'weary community' of the industrial village of Marienthal, Austria, whose once thriving social life disappeared following the closure of its large flax-fibre factory, the town's main employer. The authors describe a general state of resignation, despair, and apathy in Marienthal. More recently, the outcome of the 2016 US presidential elections has sparked new debates focusing on the role of deindustrialisation in shaping the social fabric of American communities. Some commentators have argued that the election of Donald Trump to US President is a result of a 'revolt of the Rust Belt' (McQuarrie, 2017), highlighting the role played by economic transformation and its impact on left behind Americans across the country in securing Trump's victory.

To investigate the relationship between deindustrialisation and community life, I construct an index for social capital that aligns with the definition and measurement proposed by Putnam (2000), providing descriptive evidence for its changes over time. I find that social capital decreased before the turn of the century, but subsequently increased to levels higher than in 1990, which questions claims about its steady decline. Moreover, this paper follows the empirical approach of Autor et al. (2013a), exploiting variation in exposure to trade shocks across industries and local labour markets due to surges in import penetration from China. This variation is argued to isolate plausibly exogenous changes in local demand for manufacturing jobs, whose effect on social capital in local communities is of interest in this analysis. I document a positive relationship between trade shocks and social capital accumulation over 1990-2007. However, evaluating this association separately for sub-periods and for individual variables composing the index casts a doubt on the robustness of my findings. Overall, I argue that trade shocks to local manufacturing are likely to have increased, rather than decreased, social capital, but I do not believe I can provide conclusive evidence on this question.

The remainder of this paper is structured as follows. Section 2.2 takes stock of what is known on social capital. It offers a definition of this concept, discusses its determinants, and makes hypothesis about the specific role that manufacturing decline plays in this process. Section 2.3 presents the empirical methods, discussing measurement and the econometric strategy. Section 2.4 discusses empirical results, presenting descriptive as well as econometric findings. Section 2.5 concludes, outlining limitations and avenues for future work.

2.2 Conceptual Framework and Related Literature

2.2.1 Definitional Issues

Social capital generally refers to access to resources embedded in community relations. The origins of this concept are in sociology and political science, thanks to the pioneering works of Bourdieu (1986), Coleman (1988), Portes (1998), and Putnam et al. (1993), among others. Over the past three decades, this notion has increasingly attracted the attention of economists, interested in determining how non-market interaction can mitigate problems of imperfect competition, externalities, incomplete or asymmetric information, and agent coordination (Manski, 2000; Durlauf and Fafchamps, 2005). This strand of research addresses a dissatisfaction with the 'undersocialised' foundations of neoclassical economic theory (Granovetter, 1985). Similarly, economic geographers and regional scientists have borrowed this notion to characterise certain intangible features of local communities such as informal institutions, emphasising the spatial and historically-determined dimension of this concept (Storper, 2005; Rodríguez-Pose and Storper, 2006). But what exactly is meant by social capital? Despite the large number of studies on this subject, its definition remains elusive and fraught with disciplinary divides. Social capital is often described with broad brush strokes in terms of generalised trust, social norms, reputation, community governance, altruism, pro-social behaviour, and participation in civic life, to name a few examples. Researchers concur that this theoretical vagueness in turn impinges on empirical work (Durlauf, 2002; Manski, 2000). Reliance on disparate definitions can lead to mismeasurement or even erroneous identification, and affects comparability of findings across studies. In light of this ambiguity, and without aiming to provide a comprehensive discussion of social capital as a concept, it is useful to review some key definitional issues in order to clarify the purpose, scope, and limitations of the empirical work I carry out. Three interrelated issues stand out: the level of analysis, the tension between functional and causal definitions, and the community structure.

With respect to level of analysis, social capital can be defined as an attribute of individuals, or as pertaining to an entire community (or geographical area). Traditionally, social capital has been defined in aggregate terms, as being embedded in networks of individuals, or as the community realisation of a set of beliefs affecting informal institutions (e.g., trust, social norms, or culture more in general). Research in sociology and political science has tended to rely on this type of definition, including for instance the work of Putnam et al. (1993). Economists too, have often considered social capital as a community outcome (Zak and Knack, 2001; Knack and Keefer, 1997; Bowles and Gintis, 2002; Keefer and Knack, 2008).

However, Glaeser et al. (2002) point out that aggregate-level definitions make it difficult to understand the economic incentives and motivations underlying the accumulation of such capital, as communities are not decision makers. They thus advocate for a definition of social capital centred around the individual. Moreover, even when conceptualised as a network-level variable, it is useful to think about how micro-level interactions affect the way individuals form preferences, develop expectations, and face constraints (Manski, 2000; Durlauf and Fafchamps, 2005). At the same time, the authors note that the aggregate realisation of social capital in a group can differ from the simple sum of all its members' due to the presence of externalities and multiplier effects (Glaeser et al., 2002).

A second and related concern is that of confusion between functional and 'causal' definitions, as pointed out by Durlauf (2002, p. 460): "When social capital is defined as a set of norms or values that facilitate co-operation and efficiency, this is a functional notion. In contrast, when one argues that the co-operative behaviour of others leads to expectations under which co-operation is individually rational, this is a causal notion." In the former case, social capital is characterised in terms of the (typically favourable) function it performs. As a result, social capital so defined is almost invariably observed in conjunction with the very outcomes it is claimed to foster, resulting in a 'warm glow' interpretation of this concept as a necessarily favourable feature of economies. By contrast, social capital can also lead to negative outcomes. In particular, it can distort incentives, misallocate resources, and have inequitable distributional consequences to the extent that it does not reach evenly across all actors in an economy (Durlauf and Fafchamps, 2005; Dasgupta, 2005). To circumvent this issue, which passes an a priori judgement on the nature of the concept, Lin (1999), Burt (2000) and Dasgupta (2005) recommend to define social capital simply as interpersonal networks and their structure. This approach has a long tradition in analytical and economic sociology, at least since the work of Granovetter (1973, 1983, 1985, 2005). In economics, it is associated with a burgeoning literature on peer and neighbourhood effects (Topa and Zenou, 2015), and economic networks (Jackson et al., 2017).

Finally, and also related to the above, community structure also matters for a definition of social capital. Echoing the work of Granovetter (1973, 1983) on strong and weak ties,[1] and that of structural holes by Burt (1992), Putnam (2000) distinguishes between bonding and

[1] Granovetter (1973, 1983) defines 'weak ties' as the outer ring of acquaintances of an individual that are more likely to bridge her with resources and opportunities not otherwise available within the inner circle of close friends and family ('strong ties'). Building on this notion, he defends the 'strength of weak ties' and concludes that "The more local bridges in a community and the greater their degree, the more cohesive the community and the more capable of acting in concert" (Granovetter, 1973, p. 1376).

bridging social capital. The bonding dimension of this concept emphasises group belonging and homogeneity, with dense and often overlapping social ties among participants. It is often associated with insider-outsider issues whereby the benefits of bonding accrue to a set of people defined in terms of common ethnicity, background, geographical provenance, class, race, or other socio-economic features, with negative implications in terms of equity (Durlauf and Fafchamps, 2005; Dasgupta, 2005). Conversely, bridging occurs between individuals or groups not sharing similar characteristics, needs or interests. It relates to how well connected people are *across* (and despite their belonging to) different social groupings, with emphasis on the construction of wide-reaching social networks. Storper (2005) views bonding and bridging as operational counterparts of a deeper analytical distinction in traditional sociology between community and society, 'Gemeinschaft' and 'Gesellschaft' (Tönnies and Loomis, 1957; Durkheim, 1893). He stresses that there is an underlying tension between the two: intense bonding tends to come at the expense of bridging. In practice, places with high rates of civic participation and trust also tend to be socially more homogeneous. This tension poses an important empirical challenge to the extent that bonding and bridging are features of a community very difficult to observe using traditional social capital measures.[2] This difference is not merely conceptual. Empirical research shows that the benefits of social capital very often occur because of its bridging dimension, rather than the bonding type. In stark contrast to Putnam et al. (1993)'s findings on the positive role of social capital on the performance of Italian regions, Banfield (1958) discusses how strong family ties and nepotism in the southern Italian town of Chiaromonte ('Montegrano') curtailed its development prospects (see also Percoco, 2015, for recent evidence on this point).

This paper discusses Putnam (2000)'s claim that social capital has been falling steadily in American society, evaluating whether the decline of manufacturing also observed across the US contributes to explaining this trend. As such, I define social capital in a way that is consistent with what discussed in Putnam's work, and seek to measure it in a comparable fashion.[3] According to Putnam (1995, p. 664), social capital refers to "[...] features of social life-networks, norms, and trust that enable participants to act together more effectively to pursue shared objectives". In his later book, he describes it as: "[...] connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them [...] closely related to what some have called 'civic virtue' [...] most

[2] Jackson (2020) suggests that reliance on network definitions of social capital allows to make progress on this front, as different notions of social capital can be related to network topology and centrality measures. However, this requires detailed sociometric data at individual level.

[3] I refer to Section 2.3.1 for a discussion of measurement.

powerful when embedded in a dense network of reciprocal social relations" (Putnam, 2000, p.19). In keeping with these descriptions, I therefore adopt an aggregate definition of social capital, which emphasises participation in community life and civic-mindedness. As such, it is susceptible to the 'warm glow' effect discussed with respect to functional definitions, and it cannot distinguish between the bonding and bridging dimensions of this concept. I will however refer to literature in economics discussing individual-level determinants, for insights with respect to social capital accumulation. Moreover, adopting a functional definition does not affect empirical work to the extent that I study social capital as an outcome, rather than focusing on its effects.

2.2.2 Accumulation and Depletion of Social Capital

What determines changes in the stock of social capital over time? While the impact of social capital on economic outcomes has been studied extensively, research cannot yet fully explain the dynamics governing social capital production and depreciation processes, and their distribution in space. However, understanding the determinants of social capital accumulation and depletion is necessary in order to define hypotheses as to what role manufacturing decline might play in explaining the observed trends. I turn to the micro-level literature in economics for insights, noting that aggregate outcomes may differ due to externalities and multipliers. Many comparative static results, however, should still hold, and can help interpret empirical findings (Glaeser et al., 2002).

On the most basic level, participation to community organisations and civic engagement can be regarded as a decision on the allocation of time between work and leisure. Becker (1965) provides a formal discussion, emphasising the ambiguous effect of changes in wage rates, as they increase the opportunity cost of time but also raise overall income. The former reduces time dedicated to leisure as it becomes more expensive to forego working hours, while the latter increases it (assuming leisure is a normal good). It is difficult to establish which effect prevails on balance. This is ultimately an empirical question. However, this way socialisation is treated purely as consumption, rather than an investment. Becker (1965) makes some way forward by discussing the notion of 'productive consumption', which refers to commodities, such as business lunches, which contribute to work as well as to leisure. For productive consumption, the author notes, foregone earnings are relatively less important the more this type of leisure also contributes to earnings. Consequently, other things equal, income effects might prevail for such goods, inducing an increase in time dedicated to socialisation and community participation (assuming these do indeed represent productive consumption). Azzi and Ehrenberg (1975) make a similar case with respect to the decision to participate in religious activities, noting how it increases in age since individuals aim to maximise afterlife consumption, but the rate of increase in religious participation is lower the greater the increase in wage rates over the course of a lifetime.

A more direct treatment of social interaction is offered in Becker (1974), perhaps the first in economics to focus explicitly on this notion. The point of departure in his analysis is to let individual utility depend not only on own consumption, but also on the characteristics of other relevant peers. These can be influenced via some kind of contribution, which can be thought of as social interaction. To this end, Becker introduces the notion of 'social income', defined as the sum of own income plus the monetary value of relevant characteristics of others in one's social environment (including, for instance, their income) when no contributions are made. This social income can be spent on own consumption and on efforts to alter the characteristics of the social environment, where equilibrium expenditure is given by the typical marginal conditions from consumer theory. To illustrate, an agent might care to achieve a certain standing at the workplace, which depends on the opinions of colleagues. These can be influenced by making efforts to act amicably with peers. Perhaps more straightforward is the example of a family. In a couple, one's utility depends in part also on the welfare of his or her partner or spouse. In this case, assuming there is no cost in transferring resources across family members, the social income is effectively given by the sum of the couple's incomes, and contributions to the social environment can be regarded as monetary transfers to support the partner's consumption. This example can be extended more generally to the case where someone's utility depends on the welfare of anyone in a group of people the individual cares about (e.g., neighbours, or a community), providing a rationale for charitable giving or community engagement. Importantly, Becker notes that such preferences result in a kind of social insurance whereby members of a group increase their contributions to others when these are affected by some unexpected disaster. If social capital captures efforts by a community to provide support to its members, then a negative economic shock might result in its accumulation, rather than depletion. Kaplan (2012) formalises this idea with respect to the ability of households to insure their children against adverse labour market shocks by providing an option to move back into the parental home.

More recently, Glaeser et al. (2002) develop a model of investments in social capital which builds on the intuition of theories for physical and human capital. Their comparative statics suggest that social capital accumulation depends on life cycle considerations, geographic mobility, proximity, occupational category, education, and home ownership. An agent is more inclined to invest in social capital in the earlier stages of her life, as she will reap the returns on investment over a longer time. Similarly, expected geographical mobility deters accumulation of social capital as this tends to depreciate with distance. For the same reason, lower levels of expected mobility imply that home ownership raises the gains from social capital developed in the neighbourhood due to the local nature of positive externalities that are generated.[4] Occupational category matters too, because agents whose job offers a higher return on social skills will find it profitable to invest in social capital. Finally, the authors suggest that investments in social capital are complements to those in education. Intuitively, this could be driven by the fact that schooling offers opportunities for socialisation and learning of social skills, but it could also have to do with the fact that education is a proxy for the relative status of individuals, or simply that individuals with higher levels of education benefit more from social connections than those who are less educated.

Contributing to theory about aggregate determinants of social capital, Alesina and La Ferrara (2000) argue that group level heterogeneity such as income inequality, as well as ethnic and racial fractionalisation, reduce group participation and thus erode social capital. Brueckner and Largey (2008) consider the role of population density in socialisation, to formalise Putnam (1995, 2000)'s claim that sprawl reduces social capital by lowering the chance and raising the cost of meetings. If individuals do not consider the loss of interaction when choosing the size of their lots, a negative externality arises and the equilibrium density of cities will be too low.[5]

2.2.3 The Role of Manufacturing Decline

In light of the theories reviewed above, why and how might deindustrialisation affect social capital? Firstly, if social capital investments are indeed related to occupational returns to sociability (Glaeser et al., 2002), then a reorientation of the economy towards jobs that better remunerate interaction (e.g., in services) may in and of itself lead to an increase of social capital. In addition, three other possible mechanisms are considered: the loss of earnings, population mobility, and inequality.

Traditionally, manufacturing activity has been a source of well paying jobs that sustained the American middle class throughout the past decades. The manufacturing sector has

[4] DiPasquale and Glaeser (1999) specifically model this relationship, suggesting that homeowners are 'better citizens' due to the incentives they face to invest in neighbourhood relations. This claim is later discussed and tested by Hilber (2010), who evaluate the role of constrained housing supply in limiting dilution of locally accumulated social capital due to new entrants.

[5] In their attempt to validate the model's prediction with empirical evidence, however, the authors find social interaction to be decreasing in population density, rather than increasing.

tended to pay wage premia due to a greater number of hours worked, as well as due to higher hourly compensation, especially to relatively unskilled workers (Katz et al., 1989; Helper et al., 2012). In recent years, however, technological change and trade have caused a significant contraction in US manufacturing employment, and a shift of the US economy towards services (Autor et al., 2013a; Autor and Dorn, 2013), with varying exposure of local labour markets to these forces (Autor et al., 2013b). In particular, several studies document how trade-related shocks to US manufacturing induced by rising import penetration from China caused unemployment, falling wages, lower workforce participation, and surges in social security transfers (Autor et al., 2013a, 2014).[6] Accordingly, manufacturing decline might influence social capital accumulation by reducing earnings via changes in income and unemployment (or non-employment). If community engagement is a normal good with productive consumption features, then we might expect social capital to fall as manufacturing declines. Similarly, income loss may require households to take on extra work, reducing the time available for socialisation and social capital production. On the other hand, if Becker (1974) was right regarding social insurance effects, then lower earnings may increase the dependency on one's local community for various services in the form of reciprocal exchange of favours and support. Households may struggle to buy these on markets. For instance, a low income household may ask a good willing neighbour to look after their children while being at work, rather than paying a babysitter. In this context, social sanctions have greater bite and individuals may be less inclined to engage in anti-social behaviour. Community ties may thus grow stronger as individuals tap more into their social support system. It is also theoretically possible that lower wage rates reduce the opportunity cost of time spent working, inducing individuals to interact more with each other.

In addition, shocks to the manufacturing sector have a distinct geographic profile (Autor et al., 2013b), which matters as social capital might be affected through population mobility or due to rising inequalities across the US. In the former case, it is possible that workers partly adjust to changes in their economic environment by relocating to other local labour markets unaffected by the shock. Population churn might then cause social capital to fall due to community ties being broken (in case of outflows, see Glaeser et al., 2002), or because newcomers dilute existing systems of support (in case of inflows, see Hilber, 2010).[7]

[6] Additionally, manufacturing decline is found to explain almost half of the increase in US non-employment between 2000 and 2011, although this increase was partly masked by a corresponding decrease in the service sector due to local housing booms before 2007 (Charles et al., 2013, 2016).

[7] Previous work also shows that American workers respond differently to localised shocks depending on their income: whilst workers in the high-wage group move to other regions, those in the mid- and lower-end of the wage distribution tend to stay put (Bound and Holzer, 2000; Autor et al., 2014). If it is the former

In practice, however, market adjustments through the channel of labour mobility tend to be limited, at least in the short and medium terms (Blanchard et al., 1992; Autor et al., 2013a).[8] With respect to inequality, instead, social capital might decrease if manufacturing decline raises within-labour market divides, as discussed in Alesina and La Ferrara (2000). However, growing resentment among those negatively affected by unequal earnings may constitute a bond of its own. Depending on the magnitude of income inequality and distribution of earnings across groups, social capital may thus even grow stronger amongst the loosing side, for instance if deindustrialisation raises inequality *across* local labour markets. This is especially true if effects are geographically concentrated.

Having discussed these potential channels separately, however, it is important to note that they are likely to interact with each other in many ways that are hard to define a priori. The empirical analysis will thus focus on a reduced-form relationship that looks at overall effects without claiming to separately identify each mechanism. There is also ambiguity with respect to the qualitative direction of overall effects, which is ultimately an empirical question.

2.2.4 Related Applied Literature and Intended Contribution

This paper contributes to three different strands of literature. First, it discusses the claim that social capital has been declining across the US (at least up until the turn of the century). In a seminal contribution to this literature, Putnam (1995, 2000) investigates the dynamics of social capital formation and documents a persistent fall in American social capital over the course of the second postwar period. Putnam draws on a wide set of empirical measures including political participation, clubs, community associations, religious organisations, professional bodies, informal groups, trust, and altruism. For each, he shows that the predominant trend is one of falling engagement in social activities and increasing isolation of the American people. Amongst the hypothesised causes of this dynamic, Putnam enumerates increasing financial pressures and long working hours, urban sprawl, electronic entertainment and television, the changing structure of American families, female participation to the workforce, and demographic composition as the great 'civic

who are more active in the production of social capital, their mobility will negatively affect left-behind communities. Moreover, housing markets adjustments related to inelastic downward supply faced with negative shocks exacerbate these effects as they encourage the 'wrong' type of migration to declining areas, attracting individuals with low levels of human and social capitals, to the extent that the two are complements (Glaeser and Gyourko, 2005).

[8] This is consistent with a general declining trend in geographic mobility of workers documented for the US (Molloy et al., 2011, 2014, 2016). Molloy et al. (2016) also advances the hypothesis that this decline might be endogenously connected to lower levels of general social trust amongst Americans, further complicating the picture.

generation' of the pre-war period is replaced by their disengaged children and grandchildren. Putnam (2000, p. 274) also acknowledges that with no doubt "[...] global economic transformations are having an important impact on community life across America", yet he does not further explore this channel. Several papers have evaluated Putnam's claim, with mixed results. Paxton (1999) examines various indicators of social capital in the US over two decades and finds a moderate decrease in a general index measure, as well as falling levels of individual trust. Yet trust in institutions and participation to associations does not seem to have declined. Paxton also emphasises that to fully understand changes in American communities, it is important to study the dynamics of *dispersion* in social capital, not just its stocks, which is where geography can offer a valuable contribution. Costa and Kahn (2001) study changes in social capital since the 1950s, focusing on the residential dimension of this concept, that is to say, the community of family and friends formed outside the workplace within the private sphere of one's life. They distinguish between social capital produced within home through meetings with friends and family, and one generated outside home thanks to organisational membership or volunteering activity. Their findings suggest that, once education is controlled for, women's growing participation to the workforce predicts changes in social capital produced at home, while income inequality and community heterogeneity more in general explain the losses of social capital formed outside the home environment, in line with Alesina and La Ferrara (2000).

Second, my analysis dialogues with research on the determinants of social capital formation in general, and on the role of deindustrialisation in this process in particular. Rupasingha et al. (2006) offer a comprehensive analysis of social capital production in the US, in what is perhaps the paper most closely related to my work. The authors construct a county level index of social capital for various vintages. Their methodology provides the starting point for some of the empirical analysis in this paper. Moreover, their analysis represents one of the first nationwide studies at this geographical scale, shedding light on the spatial distribution of social capital *stock* in America, as well as on its local determinants. Their findings broadly align with the theoretical predictions outlined in Section 2.2.2, albeit with some exceptions. The authors confirm that formal education, age, and community homogeneity are strongly associated with social capital production in American counties. They also find a positive effect for community attachment and, contrary to Putnam, for female labour market participation. In their econometric model, the authors also consider the share of manufacturing workers in each county, with mixed results. However, their specification is cross-sectional or uses random effects to allow inclusion of time-invariant variables. Hence, it does not capture within-county changes in manufacturing employment that are key to understand the impact of deindustrialisation on social capital, nor does it consider changes over time in the outcome. My aim is to provide new and more credibly identified evidence specifically on the role of manufacturing decline, and to consider changes in social capital rather than its cross-section in levels. To the best of my knowledge, no other study looked at the relationship between manufacturing and social capital in a developed economy. Miguel et al. (2006) consider the case of Indonesia. The authors rely on survey data from 1985 and 1997, a period of rapid growth in the country, to examine how industrialisation affects density of voluntary community associations, informal social networks, and trust or cooperation. Their findings suggest that districts where manufacturing expands display positive changes in various measures of social interaction, while industrialisation in nearby districts negatively correlates with associational activity and mutual cooperation in the district itself. The key hypothesised mechanisms underlying these relationships are income-growth, higher inequality, and internal migration to neighbouring booming areas. In particular, it seems likely that the mass relocation of young Indonesians spurred by new employment opportunities in neighbouring areas may explain the disruptive effect on the social networks in origin communities. It is difficult to directly relate these findings to the American case, however.

Third, this paper also extents the literature on the societal effects of manufacturing decline by considering a novel non-economic outcome and its geography across US communities. Several studies document the adverse consequences linked to deindustrialisation. I already discussed Autor et al. (2013a) and Autor et al. (2014), who emphasise economic outcomes, notably wages, employment, and public-transfers. Autor and Dorn (2013) consider the distributional implications of a shift towards service jobs associated with skill-biased technological change and the loss of middle income job opportunities in the tradable sector. Autor et al. (2016) examine a non-economic outcome, documenting how rising import competition from abroad increased political polarisation across exposed US congressional districts and counties. More recently, Autor et al. (2019) investigate the impact of changes in manufacturing employment on family formation, the primordial social unit, by looking at how labour market shocks in this sector decrease the value of young men on the marriage market. They find that localised shocks to the manufacturing sector, which disproportionally affect men, lead to a surge in male mortality due to abuse of alcohol and drugs, shrinking employment and wages, and ultimately a fall in marriages among young adults accompanied by lower fertility rates and higher single-parent households. My aim is to contribute to this strand of research with new evidence on the strength of community participation.

Finally, a word of caution. The above sections outline the impact of manufacturing decline on social capital. However, it is also possible that social capital is a cause, rather than a result, of changes in the economic trajectory of places. There is ample research documenting the relevance of social capital for economic outcomes.[9] Putnam et al. (1993) provide one of the earliest empirical studies linking institutions and social capital to the functioning of an economy, by comparing the performance of regions in the north and south of Italy. Knack and Keefer (1997), Zak and Knack (2001) and Beugelsdijk et al. (2004) consider the social capital-growth nexus on a macro-level, while Iyer et al. (2005) and Tabellini (2010) examine its role in the local economic development of US and European regions. More in general, social capital is associated with several other economic or economically relevant outcomes such as improved educational attainments (Coleman, 1988), reduced crime (Sampson et al., 1997; Lederman et al., 2002; Akçomak and ter Weel, 2012), financial development (Guiso et al., 2004), entrepreneurship (Percoco, 2012a, 2015), and innovation (Akçomak and ter Weel, 2009; Crescenzi et al., 2013a,b), to cite a few. To address the possibility of reverse causation, along with other endogeneity issues, this analysis will rely on an instrumental variable approach (further details are discussed in Section 2.3.3 below).

2.3 Data and Empirical Methods

This paper considers the effects of manufacturing decline on social capital formation across the US. The falling share of manufacturing employment has several causes, notably changes in production technologies and increased import pressure from more competitive producers abroad (Fort et al., 2018). I follow Autor et al. (2013a) and focus on the latter, measured as the change in import penetration from China. As discussed in further detail below, this approach has the advantage of providing a plausibly exogenous measure of shocks to local labour demand in manufacturing. For the outcome variable, I borrow the methodology of Rupasingha et al. (2006) to construct an index for social capital at the level of local labour markets, defined consistently across all years considered in this analysis. I examine changes taking place over nearly two decades between 1990 and 2007. The choice of period is constrained on one end by availability of trade data, and on the other by the occurrence of the economic and financial crisis, which is likely to have affected the manufacturing sector and social capital independently from the relationship under study. Since this analysis relies on spatially aggregated information, it is potentially subject to bias due to the Modifiable Areal Unit Problem (Openshaw and Taylor, 1979; Briant et al., 2010). When moving from

[9] For conciseness, I mainly consider studies that define social capital on aggregate for countries or local communities, ignoring the extensive literature on peer effects and social networks.

individual unit observations to more aggregate data, the size and shape of the chosen level of geographical units can substantially influence the spatial statistics and model parameters estimated by the researcher. With this in mind, I rely on commuting zones to capture local labour markets. Commuting zones are groupings of counties developed by Tolbert and Sizer (1996) using hierarchical clustering methods as a spatial measure of labour markets that is not constrained by minimum population thresholds and maximises within-group commuting ties. They have the advantage of representing boundaries based on economic geography rather than administrative criteria. This is important in the present analysis because industry shocks are otherwise captured incorrectly, as they are likely to be spatially correlated among neighbouring counties. The 1990 classification used here defines 741 CZs covering all territories in the United States. Due to incomplete data availability, only 722 CZs in 48 continental states are used. Below, I discuss more in detail the measurement of key variables of interest, and the econometric strategy.

2.3.1 Social Capital Index

The outcome measure is a composite index created following the methodology of Rupasingha et al. (2006), henceforth the social capital index (SK). This metric captures some key empirical dimensions of social capital identified in extant literature by considering four variables: associational density and volunteering activity (Putnam et al., 1993; Fukuyama, 2001), voter turnout (Alesina and La Ferrara, 2000), rates of response to decennial censuses (Knack, 2002), and religious participation (Putnam, 2000). Associational density is the number of establishments per million population for community organisations such as civic and social associations, sport, recreation and bowling centres, religious, political, and professional organisations, and membership clubs. Religious participation is the count of active members in congregations of various faiths per thousand population. The social capital index, constructed for 1990, 2000 and 2007, is created by extracting the first component from a principal component analysis (PCA) that uses the four variables listed above. Before performing the PCA, each variable is standardised using the distribution across all the years considered, to avoid excessive loading of those with higher variance while at the same time ensuring comparability across years. Appendix Tables 2.B.1 and 2.B.2 give information on the Pearson's correlation coefficients for all variables used in the index, as well as details on principal components such as their loadings, correlations with each index variable, and the proportion of explained variance. The retained first component explains nearly half of the total variance across the variables over 1990-2007. To prevent possible outliers from driving empirical results, the top and bottom one percent values of the index are trimmed and replaced with those at the first and 99th percentiles. Each year's index is also normalised across all available vintages using the min-max method so as to allow capturing changes over time and to ease interpretation as the resulting indicator is bounded between zero and one (Nardo et al., 2008).

Data on the number of associations in each CZ are obtained from the County Business Patterns database (CBP), which gives the number of establishments by six-digits North American Industrial Classification System (NAICS) or, before 1997, by four-digits Standard Industrial Classification (SIC) codes. Relevant codes, chosen in line with Rupasingha et al. (2006), are listed in Table 2.B.3 in Appendix. Consistency in the coding of industries over the years was ensured. Turnout to presidential elections and rates of response to Decennial Census letters are obtained from Rupasingha et al. (2006), integrated with data from David Leip's Atlas of US Presidential Elections for year 2000 (Leip, 2016). Religious participation is measured using data on adherence to religious congregations of all faiths, collected by the Association of Statisticians of American Religious Bodies (ASARB) and distributed by the Association of Religion Data Archives (ARDA). These data give the total number of adherents to congregations affiliated with 149 religious bodies of Christian as well as other faiths for each county in the US.[10] Not all data was available consistently for the vintages considered in this paper. I use the average turnout to 1988 and 1992 elections for 1990, and that to 2004 and 2008 elections for 2007. The 2010 Census mail response and religious adherence rates were used for 2007. Aggregation of county data on voter turnout and responses to Census letters to CZ level was done using population-weighted averages. Associational density and rates of religious participation were obtained by adding up counts by CZ and normalising by total CZ population.

This analysis also considers the variables composing the index separately, to investigate trends in social capital formation more accurately. In so doing, this paper follows Knack and Keefer (1997) and also distinguishes within associational density between Putnam- and Olson-type organisations (Olson, 1982), that is to say, organisations that promote trust and cooperation (e.g., civic groups) versus those established with a rent-seeking purpose (e.g., business associations).

[10] Adherence was defined as "[...] all members, including full members, their children and the estimated number of other participants who are not considered members; for example, the 'baptized', 'those not confirmed', 'those not eligible for communion', 'those regularly attending services', and the like."

2.3.2 Local Shocks to Manufacturing

Local labour market shocks to the manufacturing sector are measured as the change in exposure of each commuting zone to import penetration from China. I obtain these data from Autor et al. (2013a). What follows describes the methodology behind this measure. For each CZ i and industry j, the changing exposure to Chinese imports at time t is given by:

$$\Delta IP_{i,t}^{US} = \sum_{j} \frac{L_{ij,t}}{L_{j,t}} \times \frac{\Delta M_{j,t}^{US}}{L_{i,t}} \tag{2.1}$$

This measure mimics a traditional shift-share variable by assigning to local labour markets the change in overall imports $M_{j,t}^{US}$ from China to the US in industry j in proportion to their share of national employment in that same industry (Bartik, 1991; Blanchard et al., 1992). Variation in exposure for each CZ is thus obtained from the local structure of employment, whether manufacturing or non-manufacturing, and from the local industrial composition within manufacturing activities at the start of each period.[11] I consider differences over 1990-2000 and 2000-2007. During this time, the US experienced a more than tenfold increase in imports from China, a sizeable change compared to imports from other low-income countries, and much larger than the corresponding increases in US exports (Autor et al., 2013a). This reflects the fact that over this period Chinese manufacturing saw a dramatic rise in competitiveness, owing to the country's progressive transition to market-economy combined with growing openness to trade and accession to the World Trade Organisation (WTO) in 2001. This sharp increase represents the supply-driven shock to US manufacturing exploited in this analysis. Because the distribution of shocks is right skewed, similarly to what done with social capital I recode the top one percent of values with those at the 99$^{\text{th}}$ percentile to prevent outliers from influencing the results.

Data on imports is from the UN Comtrade Database, available consistently for many high-income countries at six-digits Harmonised System (HS) product-level from 1991 onwards.[12] Information on local industrial employment composition is obtained from the 1990 and 2000 CBP databases, which along with establishment counts also tabulate employment, firm size class, and payroll for US counties and industries.[13] All county figures are aggregated to

[11] Following Autor et al. (2013a), controlling for the overall share of manufacturing jobs at the start of each period allows to narrow the source of variation to local specialisation patterns *within* manufacturing.

[12] The matching of these data to US industry codes is described in the Online Appendix to Autor et al. (2013a), available at: `https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121`.

[13] The aforementioned Online Appendix to Autor et al. (2013a) also discusses the methodology used to impute missing employment figures for small establishments.

CZ level. Import penetration is expressed in terms of thousand dollars per worker at 2007 prices.

2.3.3 Model Specification and Identification Strategy

Econometrically, the relationship between local exposure to import penetration from China and social capital can be described using a model of this form:

$$\Delta SK_{i,t} = \beta \Delta IP^{US}_{i,t} + \mathbf{X}'_{i,t}\gamma + \theta_t + \epsilon_{s(i),t} \tag{2.2}$$

Where $\Delta SK_{i,t}$ is the change in the social capital index (or one of its components) in CZ i over the decade starting at time t, $\Delta IP^{US}_{i,t}$ is the measure of import penetration from China, and $\mathbf{X}_{i,t}$ is a vector of controls for start of period socio-economic characteristics of the CZ, which might influence social capital accumulation independently from the shocks to manufacturing. I consider among others population density and rates of homeownership, college education, or elderly (ages 65 or older) in the population. All control variables are measured using data from the USA Integrated Public Use Microdata Series (IPUMS) by Ruggles et al. (2020), or US Decennial Censuses. In addition, the model includes a period dummy θ_t indicating whether differences are taken over 1990-2000 or 2000-2007.[14] To account for the potential spatial correlation of errors, the residual term $\epsilon_{s(i),t}$ is always clustered by state. Variation to this specification that include fixed effects for nine Census divisions or 48 states are also considered, to absorb any trend in social capital at these levels. When only one decade-equivalent change is considered, the model in Equation (2.2) is essentially a two-periods fixed effects model, as it is estimated in first differences. When examining long changes over 1990-2007, instead, the resulting stacked first differences model mimics a three period fixed effects regression with looser assumptions on the serial correlation of the error term, as pointed out by Autor et al. (2013a), who rely on a similar specification.

The main coefficient of interest in this analysis is β. However, there are several possible sources of bias. First, the increase in Chinese imports may be related to unobserved domestic shocks that also independently affect social capital. For instance, rising import penetration may be related to changes in US product demand and consumer preferences, sluggish productivity growth in US industry, or technology shocks specific to the US and high-income countries that favour less labour-intensive industries by sheltering them from

[14] All changes for the 2000-2007 period are expressed in ten-year equivalents, obtained by multiplying differences by a factor of 10/7, which allows like-for-like comparison of effects over the two periods.

international competition (e.g., automation).[15] Second, and importantly, it is also possible that it is changes in local social capital driving changes in local manufacturing performance, rather than the other way around, which is in turn reflected in growth of imports from China. This is essentially a story about reverse causality, which cannot be easily dismissed especially because the measure of exposure to trade shocks is apportioned to each CZ using the structure of manufacturing jobs at the beginning of the same period over which changes to social capital are defined. Stronger social cohesion may promote industrialisation due to trust, norms of reciprocity or enhanced entrepreneurship. Conversely, it is also possible that social networks foster rent-seeking behaviour leading to manufacturing losses due to conflictual bargaining among workers and firms.[16]

It is very difficult to fully discount these alternative stories. However, I rely on an instrumental variable strategy to make some way forward in the causal identification of the effect of shocks to manufacturing on social capital. The strategy, which parallels that in Autor et al. (2013a), attempts to isolate the supply-driven variation in trade shocks from China by considering growth in imports per worker from the same origin to other high-income countries over the same period, rather than to the US.[17] In addition, these imports are attributed to CZs using the structure of local industrial employment in the decade prior to the period considered. This mitigates issues of reverse causality assuming there is no serial correlation in unobserved CZ shocks which simultaneously determine local industry structure and social capital (Faggio and Overman, 2014). More formally, the instrument is obtained as:[18]

$$\Delta IP_{i,t}^{IV} = \sum_j \frac{L_{ij,t-1}}{L_{j,t-1}} \times \frac{\Delta M_{j,t}^{IV}}{L_{i,t-1}} \tag{2.3}$$

The model in Equation (2.2) is then estimated with Two Stage Least Squares (2SLS) using $\Delta IP_{i,t}^{IV}$ in the first stage to predict $\Delta IP_{i,t}^{US}$, conditional on the same set of controls and fixed effects as in the second stage. The identifying assumption is that the variation common to imports by US and other high income countries reflects changes that are specific to China, notably its high productivity growth and rapid erosion of trade barriers. The exclusion restriction then requires that supply-driven import pressure from China does not determine changes in social capital in any other way but through its effect on the local manufacturing

[15] Related to this point, Gagliardi et al. (2015) discuss evidence on the job implications of offshoring.

[16] The literature has not reached a firm conclusions yet on this question (see Rodríguez-Pose and Storper, 2006, for a comprehensive discussion), but either of these effects represents a threat to identification.

[17] These are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland, for which comparable trade data was available.

[18] Consistently with what done with US trade shocks, I recode the top one percent of values with those at the 99$^{\text{th}}$ percentile.

sector. Following Miguel et al. (2006), this paper does not attempt to separately identify the specific channels that link shocks to manufacturing and social capital, as these are likely to interact in ways that are difficult to predict. Separate regressions will however investigate the relationship between some of these mechanisms and social capital to offer some validation of what discussed conceptually in Section 2.2.3.

2.4 Results and Discussion

2.4.1 Descriptive Analysis

To provide context to the econometric analysis that follows, I begin the discussion of my findings by highlighting some descriptive facts about the key variables of interest. Table 2.B.4 gives descriptive statistics for levels and first differences (decade-equivalent changes) of all main variables used in this analysis.

Figure 2.1: Kernel density for distribution of social capital across years

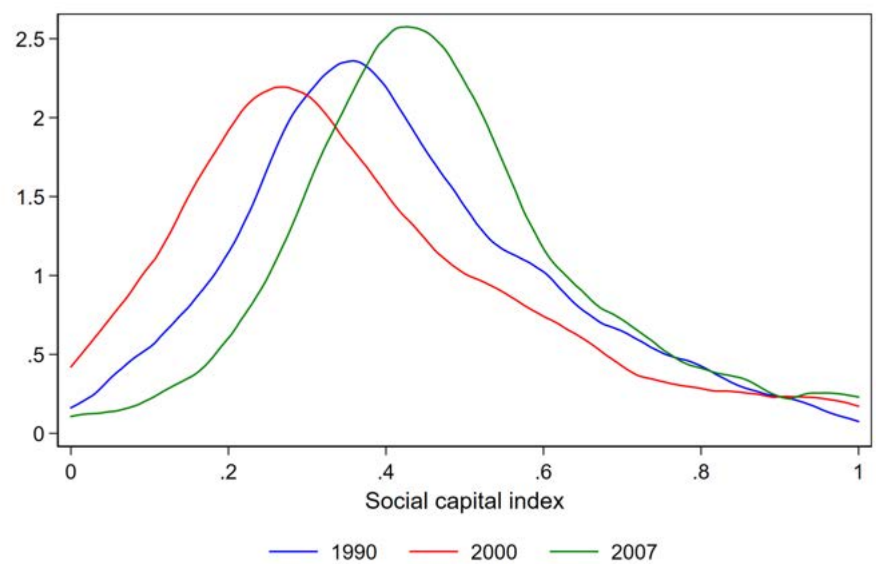
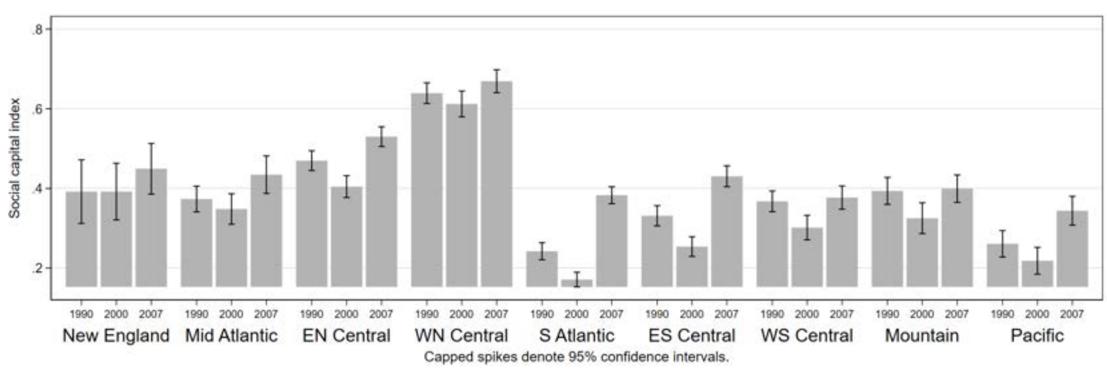


Figure 2.1 shows the distribution of the social capital index in 1990, 2000, and 2007. As suggested by Putnam (2000), it does appear that community strength fell over the decade leading up to the time the Harvard Professor published his research. According to my index, however, social capital actually increased between 2000 and 2007, inverting the declining trend described in the literature. On net, the stock of social capital seems to be higher in 2007 than it was even in 1990, denoting a high degree of accumulation since the turn of the century. T-tests for differences in means confirm these descriptive statements. Social

capital decreased on average by about half an index point between 1990 and 2000, and increased by over twice this amount during the subsequent seven years. Both differences are statistically significant at the highest conventional levels of confidence. I can also reject the hypothesis that the long term change between 1990 and 2007 is zero, confirming a net increase until before the economic and financial crisis.

The claim that social capital is in steady decline, therefore, is not fully supported by the data, at least not since after Putnam published his research. However, as emphasised by Paxton (1999), a more fruitful analysis should focus on the *dispersion* of this measure. For instance, aggregate trends might conceal geographic variation in accumulation over time. Accordingly, the histogram in Figure 2.2 shows average stocks of social capital for each Census division and year, along with 95% confidence intervals. In terms of levels, the graph shows that social capital is highest in the West North Central division, and in the Midwest more in general. By contrast, it is relatively scarce in the South Atlantic and Pacific areas. Figure 2.A.1 in Appendix maps social capital stocks across all commuting zones and years, providing a more detailed representation of the geographical distribution of this variable. Polygon shades reflect quartiles of the distribution of the index across all years. In terms of dynamics, Figure 2.2 also reveals that, with few exceptions (e.g., New England), the pattern of decline and subsequent rise in social capital is confirmed across most Census divisions, albeit with varying intensity. The South Atlantic division displays the highest volatility over the years.

Figure 2.2: Average stock of social capital by division and year



It can be instructive to investigate which of the index components is driving these findings. To this end, the maps in Appendix Figure 2.A.2 show the geographic distribution of each variable used to measure social capital. Similarly, the histograms in Appendix Figures 2.A.3 to 2.A.8 give average stocks for each variable by Census division and year, along with 95% confidence intervals. Without commenting these graphs in detail, I highlight that each subcomponent displays its own geography and dynamic, which somewhat departs from that in the aggregate index, although generally consistent. Interestingly, it appears that associational density, especially of the Putnam-type, is driving the growth trend (although Olson-type organisations tended to decline between 2000 and 2007). By contrast, rates of adherence to religious congregations fell across nearly all Census divisions since the 1990s. Voter turnout and Census mail response rate displayed somewhat more heterogeneous dynamics.

Since this analysis focuses on first differences, I provide maps for decade-equivalent changes in the main variables of interest. The geographic profile of social capital accumulation or depletion is visualised in Figure 2.3, while that for US import penetration from China is in Figure 2.4. In addition, Appendix Figures 2.A.9 to 2.A.11 also map changes in the instrumental variable for trade exposure, shares of manufacturing employment, and average household incomes. In all maps, polygons in blue denote quartiles of positive differences, while those in red are quartiles of negative differences. I always consider the distribution of changes in both decades, so that dynamics can be visually compared across periods of analysis. The maps highlight the decline and rise of social capital over the two decades I consider, and the near monotonic increase in exposure of local markets to the Chinese trade shock. Almost all communities across the US face this shock, which is particularly pronounced after the turn of the century when China joined the WTO. Areas where social capital declined the most during 1990-2000 include the Mountain region in the West, parts of the South (e.g., Texas, Alabama, and Virginia), and some areas in the Midwest (especially Montana and Indiana). During the subsequent decade, the strongest increases were in the East South Central and South Atlantic divisions. In the same period, this area was also highly exposed to trade shocks affecting its manufacturing industry, reflecting the concentration of industrial activity east of the Great Plains. Shocks were somewhat more contained during the preceding decade, and again concentrated in the eastern parts of the country. Visually comparing the two maps, there does not seem to be a clear association between local shocks to manufacturing and social capital accumulation. If anything, it would appear that communities grew stronger in areas that were more exposed to import penetration from China. Determining the nature of this relationship is the central task for the econometric analysis I propose in the next Section.

Before turning to the econometric analysis, however, it is useful to examine the association of the index of social capital with variables other than trade shocks that could also contribute to determine its production (or depletion). The purpose is twofold. First, this exercise helps validate the social capital measure I rely on in this analysis by naively look-

Figure 2.3: Decade-equivalent change in social capital

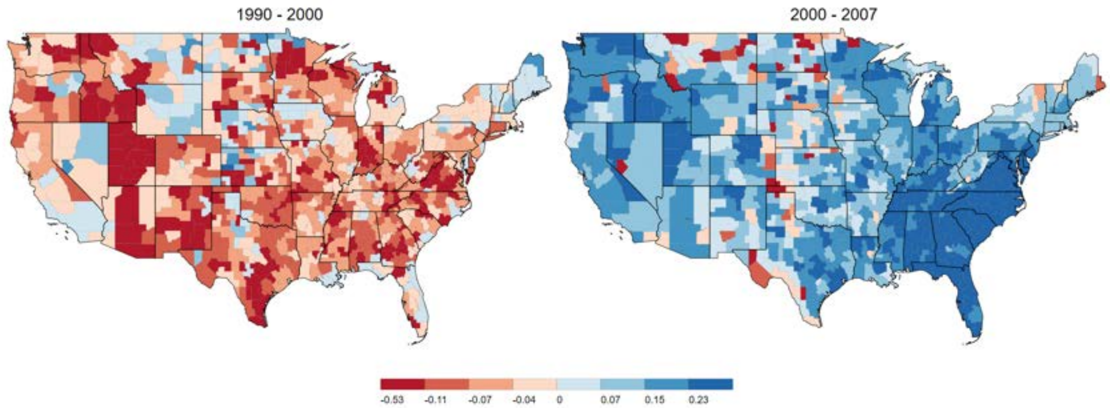
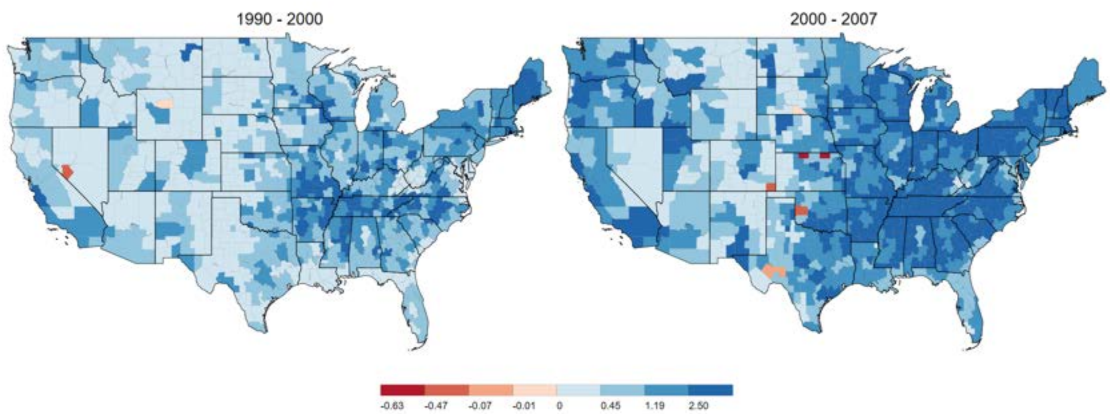


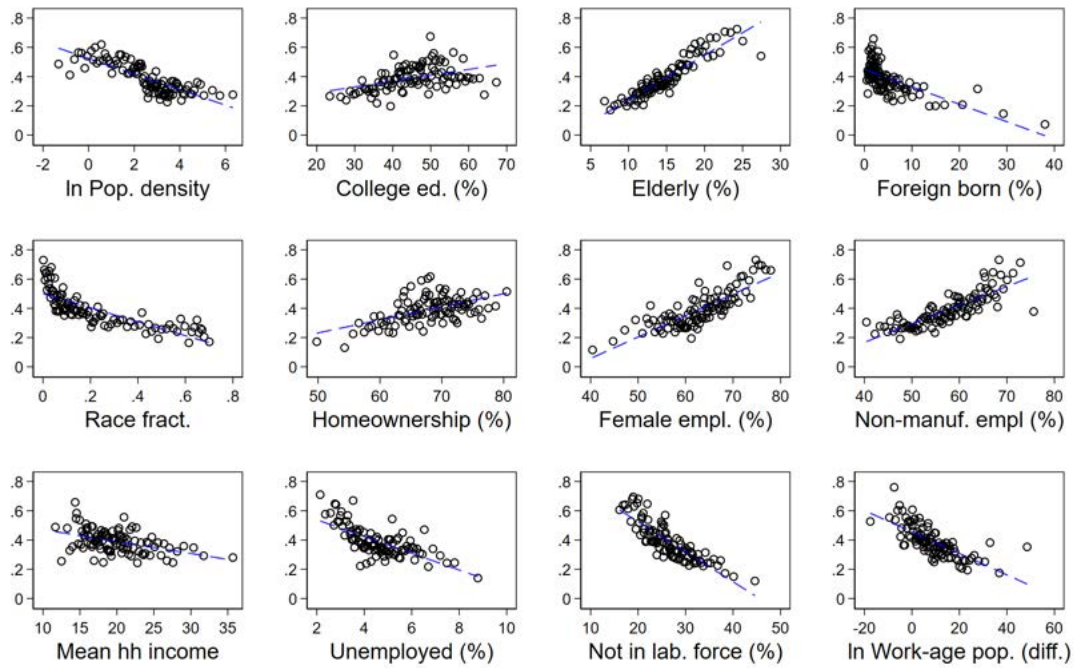
Figure 2.4: Decade-equivalent change in US per capita import penetration from China



ing at whether the correlations I observe can be reconciled with what is normally discussed in the literature. Second, looking at correlations with other variables can be informative with respect to what control variables should be considered for inclusion in the empirical models. Figure 2.5 gives binned scatterplots for a set of covariates I deem relevant to this analysis based on the theory I discussed in Section 2.2.2. I also show a linear fit to the data, obtained by regressing levels of social capital onto each covariate, along with a constant. Linear fit lines in blue denote significance of the bivariate correlation at the 95% level, using heteroscedasticity-robust standard errors. I do not include any control, nor do I attempt to address possible sources of bias in any other way. The correlations I present are simple statistical associations that should be interpreted with a degree of scepticism.

Social capital is found to be significantly correlated with all the variables I consider. Confirming the analysis in Brueckner and Largey (2008), it appears to decrease in population

Figure 2.5: Binned scatterplots for selected covariates of social capital



Binned scatter plots (100 bins). Linear fit lines in blue denote significance of the slope coefficient at the 95% level.

density, although imposing a linear fit might conceal a non-linear relationship. Perhaps this is due to the greater anonymity and impersonality of large cities, or a more favourable configuration of lower density places to social interaction. Social capital also increases with college education, consistently with the complementarities highlighted in the literature (Coleman, 1988; Glaeser et al., 2002). Similarly, it is greater in communities with a larger share of elderly population (Putnam, 1995, 2000) and homeowners (DiPasquale and Glaeser, 1999; Hilber, 2010), and lower where more people are of foreign origin or race fragmentation is higher (Alesina and La Ferrara, 2000).[19] Contrary to what suggested by Putnam (2000), there is a positive correlation between the rate of employment of working-age women and social capital. This however is consistent with the findings in Rupasingha et al. (2006). The transition of advanced economies towards service jobs might also explain changes in social capital. To test this, I consider the share of working-age population employed in activities other than manufacturing, finding a positive association. This highlights the need to control for initial shares of manufacturing employment in my empirical model, to narrow down the source of variation to the composition of jobs *within*

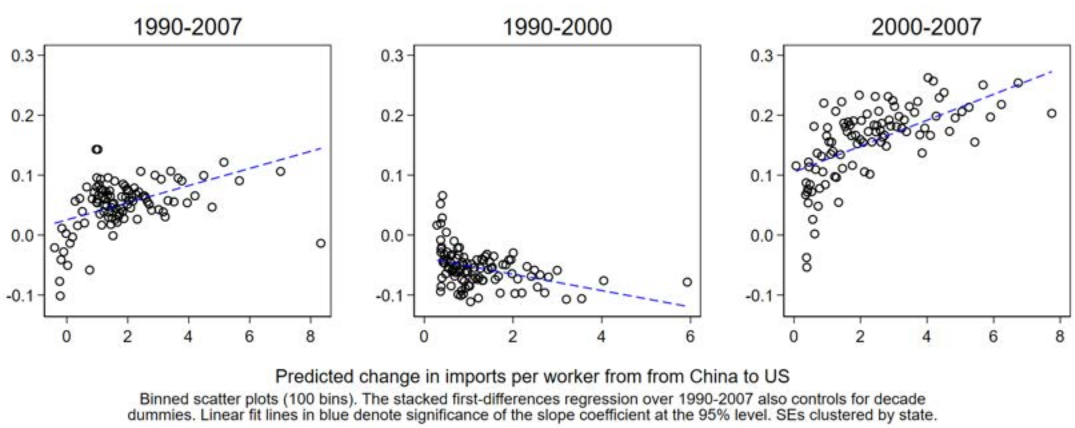
[19]The latter measure is computed as one minus the normalised Herfindahl-Hirschman Index (HHI) for the shares of population in the following groups: Whites, Blacks, Native Americans, and Asians or Pacific Islanders. This measure is consistent with the approach by Alesina et al. (2003).

manufacturing. Finally, I also consider four variables that represent possible mechanisms through which shocks to the manufacturing sector might affect social capital: average household income, unemployment, labour force participation, and changes to the working-age population (that is to say, net migration of workers). For all these I observe a negative relationship: higher income, unemployment or non-employment, and inward migration are associated with lower levels of social capital. It is difficult to interpret these correlation without additional information, however.[20] Next, I turn to the study of the relationship between exposure to trade shocks and social capital accumulation.

2.4.2 Econometric Analysis

I begin by reporting in Figure 2.6 a visual intuition of the analysis I perform. The binned scatterplots show the bivariate relationship between decade-equivalent changes in social capital and predicted changes in per worker import penetration from China. I consider the stacked first differences model in the first panel, and also show the relationship for the two sub-periods separately. As before, linear fit lines in blue denote significance of the slope coefficient at the 95% level, in this case using standard errors clustered by state.[21]

Figure 2.6: Binned scatterplot for changes in social capital and trade shocks



In all instances the relationship is statistically significant. However, social capital accumulation appears to display a heterogeneous response to shocks to the manufacturing sector. While there is a strong positive relationship overall during 1990-2007, only changes over 2000-2007 align with this finding. During 1990-2000, reflecting the decline observed in

[20] Appendix Figures 2.A.12 to 2.A.15 provide binned scatterplots for each variable composing the social capital index, with comparable results.

[21] Equivalent visualisations for the observed US trade shocks (endogenous variable) and trade shocks using imports from other high income countries (instrumental variable) are also available in Appendix Figures 2.A.16 and 2.A.17, showing the simple OLS and reduced-form relationships.

the raw data, social capital is inversely related to trade shocks. These relationships can be quantified. Table 2.1 reports baseline OLS regressions for the endogenous explanatory variable and its instrument (the reduced-form of the 2SLS with no controls). Columns (1) and (4) give stacked first differences models, which also control for decade dummies, while the other columns give estimates for each decadal change. The social capital index is rescaled to the 0-100 interval for ease of interpretation. In column (1), the positive and statistically significant coefficient of 1.14 suggests that a one thousand dollar increase in import penetration per worker raises social capital by just over one full index point. Faced with a negative shock to their local manufacturing activity, therefore, it appears that communities tend to grow stronger the more intensely they are exposed to the shock. This is especially true during the 2000-2007 period. In the simple OLS regression in column (2), it does not seem that trade shocks have any significant effect on social capital during 1990-2000. By contrast, the reduced form regressions in columns (4) to (6) highlight the heterogeneity of this relationship. In column (5), it would seem that a one thousand dollar increase in exposure to import penetration from China (as captured by imports to other high income economies) decreases social capital by about 1.2 index points. This effect is not only statistically significant, but also similar in magnitude to that over 1990-2007 in column (4), with the opposite sign. The inconsistency of these findings is somewhat puzzling and suggests that there is some other unobserved mediating factor determining the nature of this relationship over time.

Table 2.1: Baseline OLS and reduced-form regressions for social capital

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	1.144	-0.304	1.706			
	(0.263)^a	(0.230)	(0.334)^a			
Δ IP China to other				1.227	-1.172	1.858
				(0.283)^a	(0.445)^b	(0.404)^a
Adj. R^2	0.5039	0.0016	0.0863	0.5007	0.0227	0.0824
R^2	0.5046	0.0030	0.0875	0.5014	0.0240	0.0837
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. The social capital index is rescaled to the 0-100 interval.

As a natural next step, I show in Table 2.2 results from 2SLS regressions obtained from fitting the model in Equation 2.2 with predicted values for the potentially endogenous trade shocks. For every regression, I report the coefficient on the instrumental variable from the first stage, as well as the Kleibergen-Paap F-statistic to test for weak instruments.

Reassuringly, this metric is well above the conventional levels considered for the test across all specifications, suggesting that the instrument is relevant.

The first five columns give estimates for the stacked first differences model over 1990-2007. Column (1) shows the plain bivariate relationship between supply-induced shocks to local manufacturing and social capital production, controlling only for decade dummies and a constant. The positive and statistically significant coefficient confirms the findings reported for OLS models, with a similar (perhaps slightly greater) magnitude. Column (2) augments the model by controlling for levels of social capital at the beginning of each period, and initial shares of manufacturing jobs as measured by the CBP data. The former addresses the possibility of general trends in the production of social capital, while the latter allows to restrict the source of variation in the trade shocks variable to changes in local industry composition within manufacturing, rather than the share of local manufacturing jobs more in general. In so doing, it implicitly also captures any possible explanatory role of other non-manufacturing employment, and generalised changes to the structure of local economies (e.g., increased specialisation in services). The negative coefficient on initial levels of social capital suggests that local labour markets with a one index point greater level of community strength face a differential change of 0.14 points. That is to say, greater initial levels mitigate social capital accumulation by about 10-20% of the effect of trade shocks (depending on the specification). Initial shares of manufacturing employment have a qualitatively similar effect, although smaller in magnitude and not always significant. In column (3) I introduce a number of controls to address the possibility that social capital accumulation changes across places with higher initial levels of population density, college education, elderly or foreign born individuals, homeowners, or that have higher race fragmentation or female shares of employment. All these variables were found to significantly predict social capital in Section 2.4.1. Most coefficients are significant, with the exception of homeownership. Moreover, other things equal, it appears that population density is positively correlated with social capital accumulation, and so is race fractionalisation, although only weakly so. I do not attempt to interpret any of these coefficients as they are likely biased. Even after controlling for these factors, however, a unitary change in import penetration from China has a positive and significant effect on social capital accumulation. The magnitude, although slightly reduced, is comparable to that in previous models. In columns (4) and (5), I introduce dummies for Census divisions or states to absorb any trends specific to these geographical units.[22] Still, the coefficient on trade shocks is virtu-

[22] Models with state FEs drop two CZs due to overlap in the definition of the two geographical units. For Providence, Rhode Island, and Bridgeport, Connecticut, the CZ corresponds to the state itself.

ally unchanged. Because the inter-quartile range of decade-equivalent changes in exposure to import penetration from China over 1990-2007 is of approximately two thousand dollars per worker, the estimate in column (4) suggests that a CZ at the 75$^{\text{th}}$ percentile of exposure faced a differential positive increase in social capital of about two index points compared to another at the 25$^{\text{th}}$ percentile. This is about 15% of the standard deviation in social capital accumulation over 1990-2007.

Table 2.2: Main 2SLS regressions for social capital

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	1.468	1.336	1.059	1.029	1.101	-1.740	-1.131	-0.137	-0.191
	(0.333)^a	(0.381)^a	(0.383)^a	(0.377)^a	(0.344)^a	(0.759)^b	(0.703)	(0.423)	(0.378)
Social capital (level)		-0.143	-0.208	-0.205	-0.217	-0.0826	-0.0256	-0.313	-0.303
		(0.0265)^a	(0.0232)^a	(0.0271)^a	(0.0334)^a	(0.0463)^c	(0.0438)	(0.0414)^a	(0.0434)^a
Manuf. empl. CBP (%)		-0.0552	-0.0937	-0.103	-0.101	0.0270	0.0684	0.118	0.0535
		(0.0470)	(0.0475)^c	(0.0471)^b	(0.0572)^c	(0.0657)	(0.0644)	(0.0629)^c	(0.0572)
ln Density			1.085	0.770	1.367	-0.470	-0.371	1.814	2.253
			(0.295)^a	(0.363)^b	(0.441)^a	(0.480)	(0.455)	(0.572)^a	(0.687)^a
College (%)			0.114	0.167	0.242	-0.0693	0.0232	0.211	0.157
			(0.0592)^c	(0.0598)^a	(0.0731)^a	(0.110)	(0.0864)	(0.0778)^a	(0.0780)^b
Elderly (%)			0.404	0.318	0.409	0.473	0.346	0.212	0.301
			(0.117)^a	(0.136)^b	(0.166)^b	(0.124)^a	(0.128)^a	(0.229)	(0.268)
Homeownership (%)			0.0525	0.0286	0.0109	-0.191	-0.107	0.0764	-0.0576
			(0.0778)	(0.0657)	(0.0876)	(0.119)	(0.113)	(0.110)	(0.122)
Foreign born (%)			-0.236	-0.198	-0.234	0.0144	0.101	-0.270	-0.345
			(0.0505)^a	(0.0614)^a	(0.0745)^a	(0.102)	(0.0840)	(0.0900)^a	(0.0812)^a
Race frac.			3.109	0.848	-2.122	-1.079	-0.420	3.495	-0.787
			(1.781)^c	(2.146)	(2.198)	(2.928)	(2.901)	(4.014)	(3.507)
Women empl. (%)			0.151	0.0754	-0.177	0.177	0.0164	0.488	0.495
			(0.0721)^b	(0.113)	(0.134)	(0.116)	(0.100)	(0.153)^a	(0.125)^a
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6726^a	0.6623^a	0.6454^a	0.6290^a	0.7167^a	0.6885^a	0.5449^a	0.5364^a
Kleibergen-Paap F	248.21	128.49	113.69	107.89	100.45	18.51	16.13	65.70	59.28
Adj. R^2	0.0349	0.1101	0.1683	0.0979	0.1162	-0.0358	-0.0293	0.3023	0.3021
R^2	0.0369	0.1131	0.1752	0.1104	0.1235	-0.0085	-0.0136	0.3206	0.3128
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. The social capital index is rescaled to the 0-100 interval.

In columns (6) to (9) I consider changes over the 1990-2000 and 2000-2007 periods separately, with dummies for Census divisions or for states. Somewhat puzzlingly, the coefficient in column (6) for 1990-2000 is negative and significant in the model with Census divisions FEs, in contrast to results documented with the stacked first differences model. This coefficient is not very robust, as introducing state FEs in column (7) makes it insignificant. However, it casts a doubt regarding the stability of the statistical relationship I documented for 1990-2007. This is all the more the case as point estimates for the first differences model over 2000-2007 are also negative, although insignificant. While I do uncover a positive and statistically significant relationship between shocks to local manufacturing and social capital over the roughly two decades since 1990, I therefore encourage to interpret this result carefully. Next, to better understand what drives my finding, I examine changes on each variable composing the index separately.

Figure 2.7: Coefficients plot for components of the social capital index using 2SLS

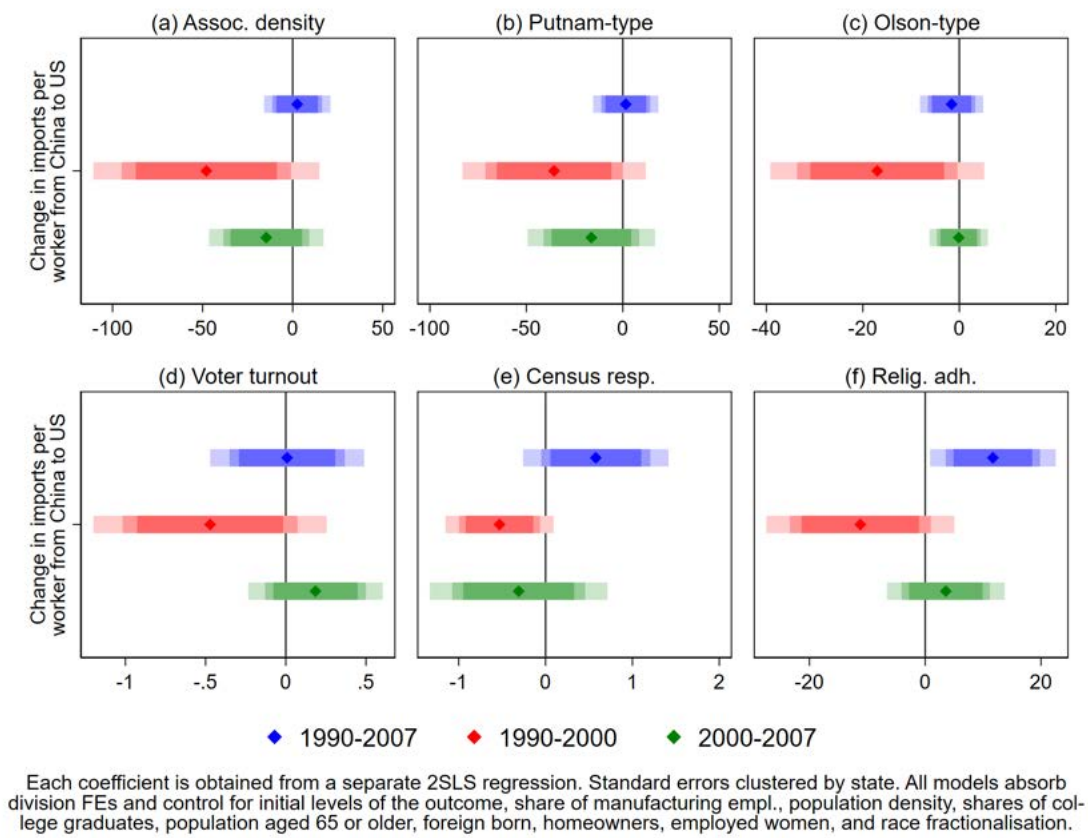


Figure 2.7 reports coefficients on trade shocks obtained from separate 2SLS regressions for different outcomes, using stacked first differences over 1990-2007 (in blue) or first differences over 1990-2000 and 2000-2007 (in red and green, respectively). I control for initial levels of the outcome (rather than the social capital index), and the full set of covariates presented in Table 2.2. All models also absorb Census division FEs.[23] Considering results for the 1990-2007 regressions, it appears that the positive coefficients I uncover for social capital are driven largely by an increase in rates of response to decennial Censuses, and by greater participation to religious organisations. By contrast, all index variables appear to be

[23] In Appendix, I also report a similar coefficients plot obtained by running simple bivariate OLS regressions for trade shocks and their instrument (Figure 2.A.18). I also include regression tables for all OLS and 2SLS models (Tables 2.B.5 to 2.B.16).

negatively associated with shocks to manufacturing during the decade leading up to the turn of the century.

Finally, before concluding, I qualitatively discuss which channels might be associated with the strengthening of local communities faced with shocks to their local manufacturing base over 1990-2007. Table 2.3 gives OLS (odd columns) and 2SLS (even columns) estimates for stacked first differences models. Instrumental variable regressions use the change in import penetration to non-US high income countries to predict the endogenous regressor of interest. I do not claim that the instrument is valid for each channel. Rather, I report these alternative estimates as an attempt to isolate the same identifying variation used for all previous results.

Table 2.3: Channels for social capital change, OLS and 2SLS regressions (1990-2007)

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS
Δ Manuf. empl. (%)	-0.181	-1.982								
	(0.233)	(0.786)^b								
Δ Unemployed (%)			1.636	3.510						
			(0.493)^a	(1.194)^a						
Δ Not in lab. force (%)					-0.166	2.086				
					(0.182)	(1.003)^b				
Δ Mean hh income							-0.916	-2.072		
							(0.327)^a	(0.796)^b		
Δ ln Working-age pop.									-0.242	2.989
									(0.0480)^a	(3.416)
Fixed effects	Div.	Div.	Div.	Div.	Div.	Div.	Div.	Div.	Div.	Div.
First stage coeff.		-0.3350^a		0.1892^a		0.3184^a		-0.3206^a		0.2221^a
Kleibergen-Paap F		64.08		30.99		40.62		40.57		0.94
Adj. R^2	0.5722	-0.0164	0.5880	0.0677	0.5724	-0.1745	0.5823	0.0651	0.5883	-6.2475
R^2	0.5778	-0.0023	0.5934	0.0806	0.5780	-0.1583	0.5878	0.0780	0.5937	-6.1471
N	1,444	1,444	1,444	1,444	1,444	1,444	1,444	1,444	1,444	1,444

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. The social capital index is rescaled to the 0-100 interval. All models include a constant and a time dummy, and control for initial levels of social capital, along with the full set of covariates in Table 2.2.

Columns (1) and (2) consider the change in local manufacturing jobs as a share of working-age population. While the OLS is inconclusive, the 2SLS estimate suggests that a one percentage point decrease in manufacturing employment is associated with a near two index points differential increase in social capital. Consistently, columns (3) and (4) imply that a one percentage point increase in unemployment raises social capital by roughly the same amount (or even more, according to the 2SLS estimate). An increase in the share of working-age population not in the labour force has a similar effect, although only statistically significant for the IV regression (columns 5 and 6). Mean household wage income is negatively associated with social capital. Every one thousand dollar decrease in average household wage per working-age adult differentially increases social capital by one, up to two, index points. This finding gives credit to Becker (1974)'s theory of social insurance, whereby a community whose members are hit by a negative shock comes together by providing compensating transfers of resources to those most negatively affected. Finally, columns (9) and (10) estimate the effect of log changes to the working-age population as a proxy for net labour mobility. The OLS coefficient suggests that a one percentage point increase in working-age population reduces social capital by roughly a quarter of an index point. This finding is consistent with what discussed in Hilber (2010) with respect to the dilution of local community goods as new members move in. The 2SLS estimate in column (10) is insignificant, yet it should be noted that the instrument is weak. All in all, the regressions in Table 2.3 align with the intuition that US communities grew stronger confronted with a trade-originating negative shock to their local manufacturing sector.

2.5 Conclusions

The persistent decline of manufacturing employment observed in the US since at least the 1970s raises concerns about the effect these changes may have on American society. In his investigation of social capital in the US, Putnam (1995, 2000) offers some evidence for a parallel long term decline in social capital. This paper investigated the possibility that these two trends are connected, with particular attention to manufacturing shocks originating from the plausibly exogenous surge in import competition from China during the 1990-2007 period. I proxy social capital using a measurement strategy known to the literature (Rupasingha et al., 2006), suitably re-interpreted to fit the scope of this work which focuses on changes and considers local labour markets as the relevant unit of analysis.

The paper makes several contributions. First, it provides new descriptive evidence regarding the claim that American social capital is declining, somewhat questioning its validity at least since the turn of the century. While I attest a decrease in stocks before year 2000, I also find that social capital subsequently grew to levels higher than in 1990. Second, I provide new evidence on the drivers of social capital accumulation, suggesting that a negative economic shock affecting a community may induce stronger bonds among its members, perhaps confirming the social insurance mechanisms discussed by Becker (1974), or reflecting the greater pay-off to socialisation derived from working in other sectors such as services (Glaeser et al., 2002). Separately regressing different hypothesised channels through which manufacturing shocks might affect social capital does confirm some of the effects expected theoretically, although none of these can be interpreted causally. Third, and related to the previous point, I document a novel, non-economic, consequence of manufacturing decline in general, and of trade shocks in particular. This ties together with a literature that shows how higher import penetration from China is associated with polarisation in voting behaviour, substance abuse, and lower rates of marriage (Autor et al., 2016, 2019). I uncover a positive relationship between trade shocks and social capital, but whether or not this is a socially desirable outcome I cannot say. I am personally sceptical of 'warm glow' interpretations. What I measure could just as well be a process of inward retreat by local communities facing hardship, united by anger and resentment at the distributional consequences of globalisation. It is here, perhaps, that the seeds for the 'revolt of the Rust Belt' were sown (McQuarrie, 2017).

There are also significant limitations to this analysis that I wish to highlight. My results should not be overstated. I uncover a positive relationship between plausibly exogenous shocks to local manufacturing and social capital accumulation. However, this finding only holds for the stacked first differences specification over 1990-2007, and is not confirmed when looking at decadal changes separately before and after the turn of the century. Moreover, the variables underlying the social capital index display heterogeneous responses to these shocks, which in addition also depend on the time-frame being considered. All this casts doubts as to the stability and robustness of the statistical relationship I described. Ultimately, there does not seem to be sufficient evidence to conclude that the relationship between manufacturing decline and social capital can be described unequivocally. It would appear that communities react in different ways to trade shocks depending on a complex set of factors, often unobserved, and involving place-specific dynamics that interact with one another in a way that is hard to model quantitatively once and for all.[24]

The ambiguity of this paper's findings underscores several challenges that were encountered during the empirical work, which offer an opportunity for reflection, learning, and avenues for future research. Measurement is most prominent among these. I attempted to capture changes in stocks of social capital over time by relying on an index that aligns with the conceptual definition offered by Putnam (1995, 2000). However, this is just one of many possible ways to proxy for the strength of local communities. As emphasised, I cannot adequately distinguish between the bonding and bridging dimensions of this concept, nor does my measure easily lend itself to intuitive interpretations. In addition, there is the possibility that the nature of what is being measured changes over time too, so that as community participation takes new shapes and forms, the alleged decline of social capital is just a reflection of mismeasurement. Finally, and related to the previous point, much of the

[24]For instance, it might be that there is heterogeneity depending on the types of industries that face trade shocks, with some sectors naturally leading to new industries (e.g., semi-conductors in San Jose or Silicon Valley) and others leaving no prospects once they disappeared (e.g., automotive in Detroit). Testing this hypothesis, however, is beyond the scope of this analysis.

trends in social capital described by Putnam (1995, 2000) might simply have played out by the turn of the century, so that my definition and measurement strategy for this concept is anachronistic. Another takeaway from this paper, therefore, is that reliance on more disaggregated, direct, measures of social capital is advisable in future work. This aligns with what already discussed by Durlauf and Fafchamps (2005), who express scepticism about aggregate and index-based metrics for this concept. In particular, pure network-based interpretations of social capital, as advocated by Lin (1999), Burt (2000), and Dasgupta (2005), may offer more fruitful ground for empirical analyses, in line with literatures on peer and neighbourhood effects. Additional lines of enquiry on this topic could consider new ways to capture the territorial dimension of social capital and the heterogeneity of this concept, for instance by relying on large-scale survey micro-data, or by leveraging new sources of information such as mobile phone records or social networking platforms.[25]

[25] Arribas-Bel and Tranos (2018), for instance, review the opportunities offered by big data in geographical analysis. Poorthuis (2018) considers their application to determine the scope of urban neighbourhoods, while Tranos (2016) discusses the role of internet and social media in the emergence of 'digital social capital'. Blind et al. (2018) review methodological advances enabled by the use of finely geo-referenced data.

2.A Additional Figures

Figure 2.A.1: Stocks of social capital across US commuting zones

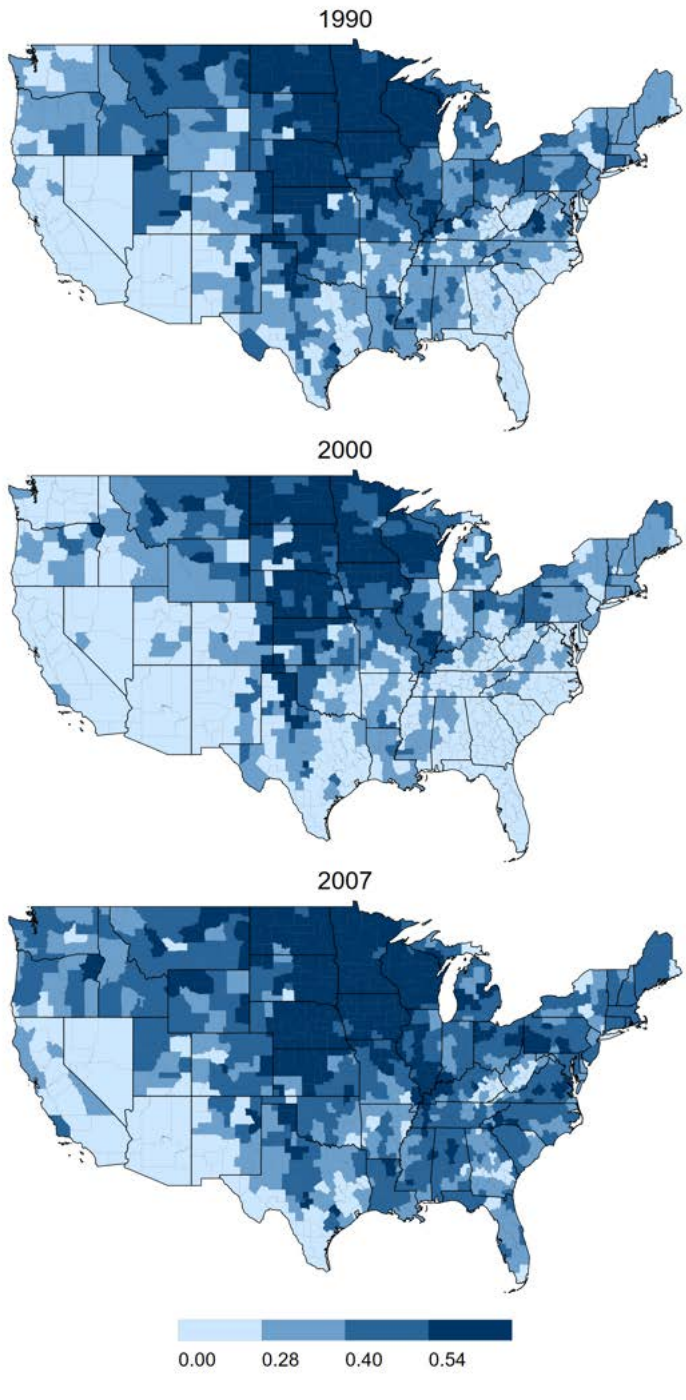


Figure 2.A.2: Components of social capital index across US commuting zones

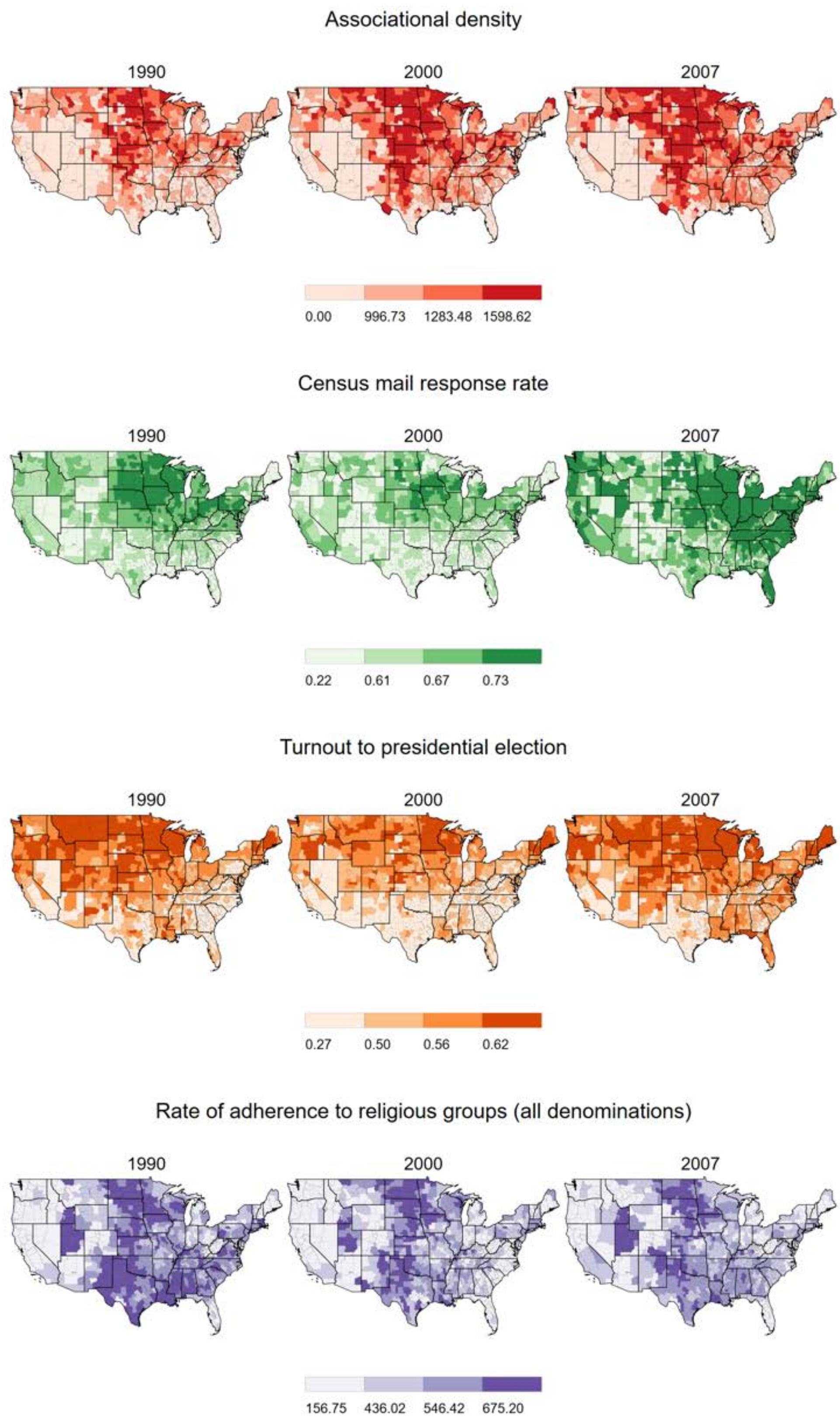


Figure 2.A.3: Average associational density by division and year

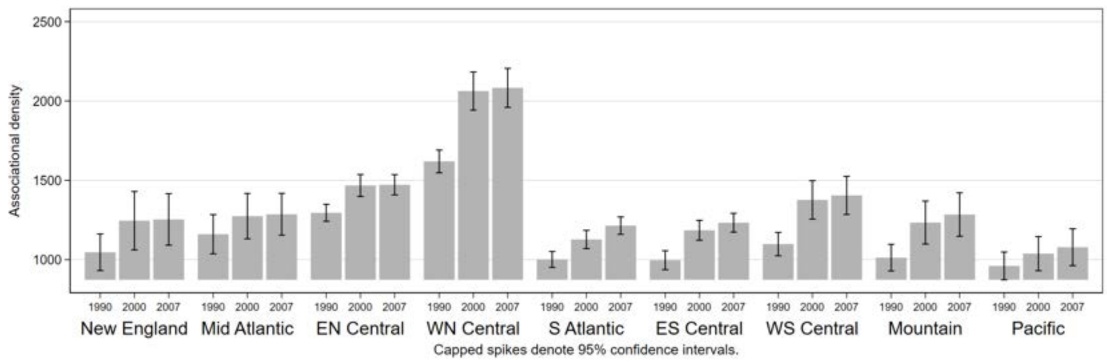


Figure 2.A.4: Average density of Putnam-type assoc. by division and year

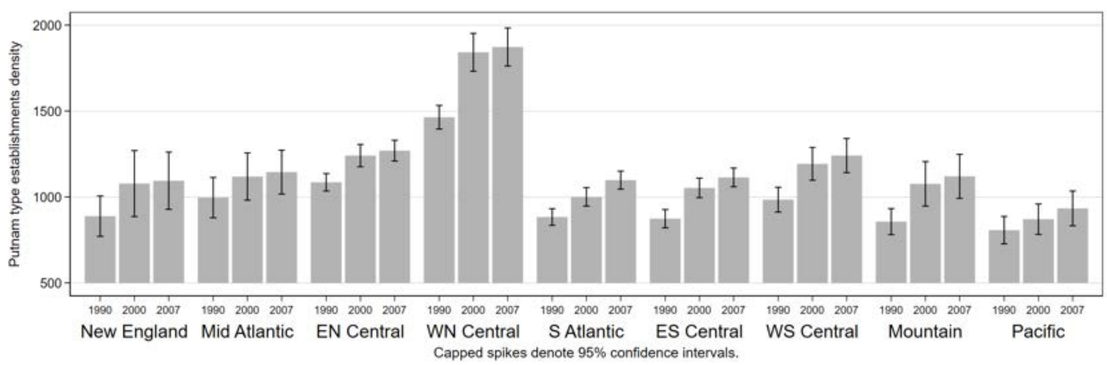


Figure 2.A.5: Average density of Olson-type assoc. by division and year

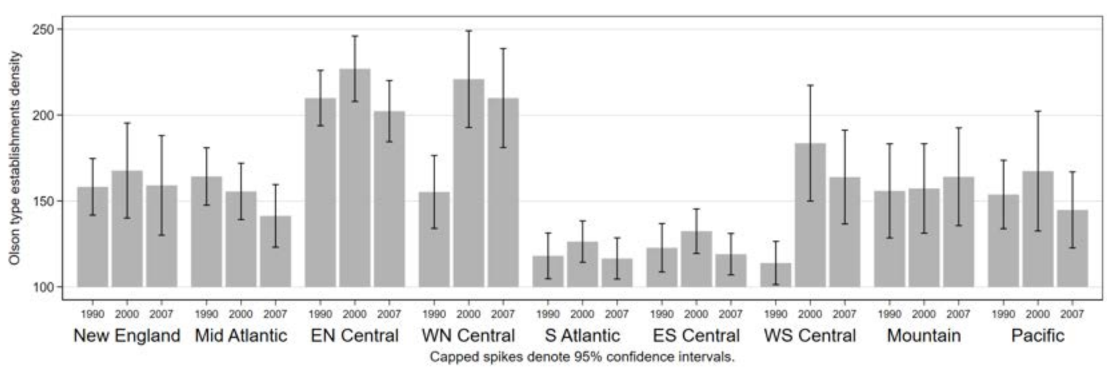


Figure 2.A.6: Average voter turnout by division and year

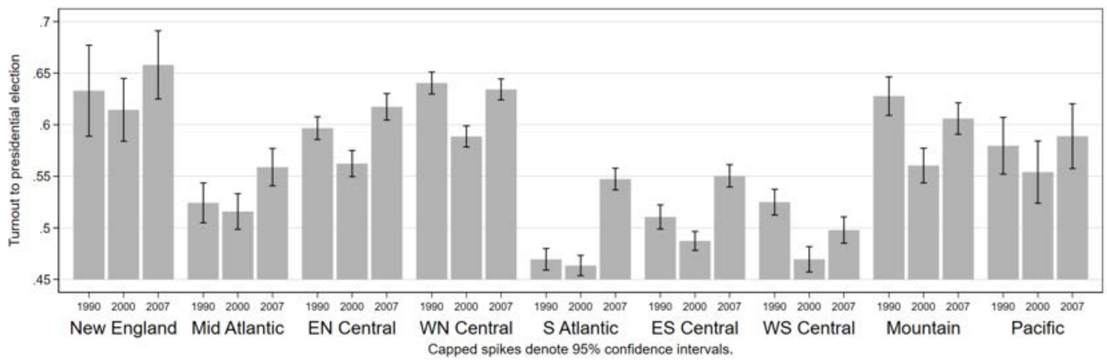


Figure 2.A.7: Average Census mail response rate by division and year

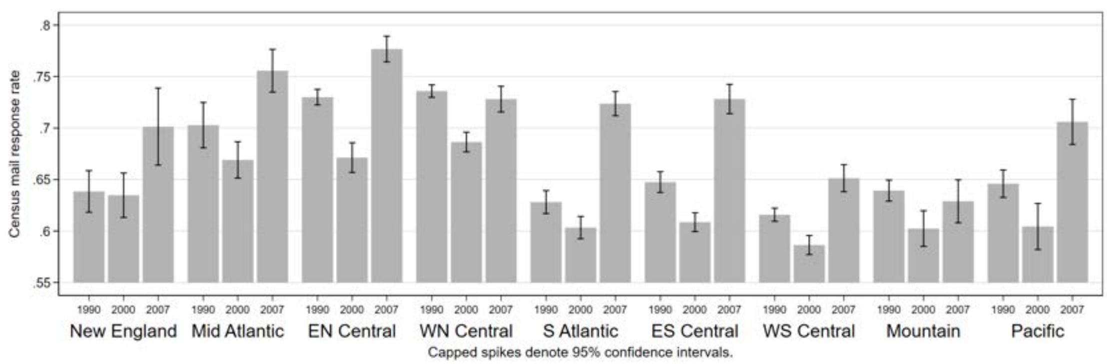


Figure 2.A.8: Average rate of adherence to congregations by division and year

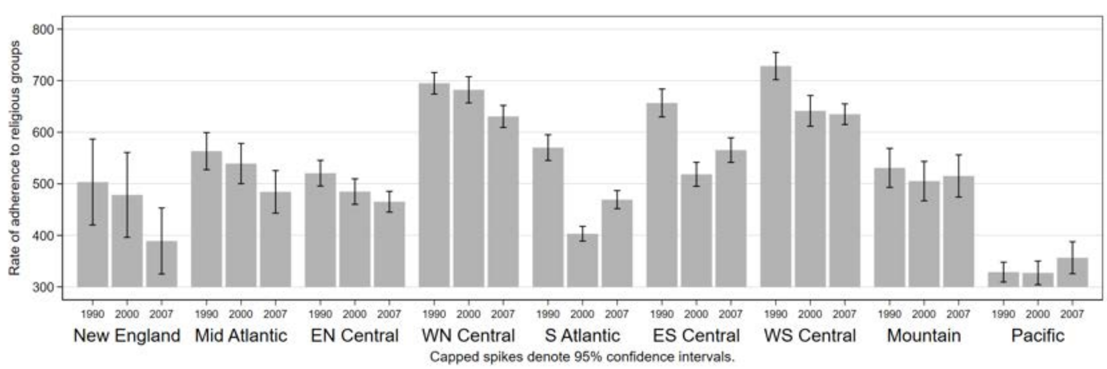


Figure 2.A.9: Decade-equivalent change in import penetration from China (IV)

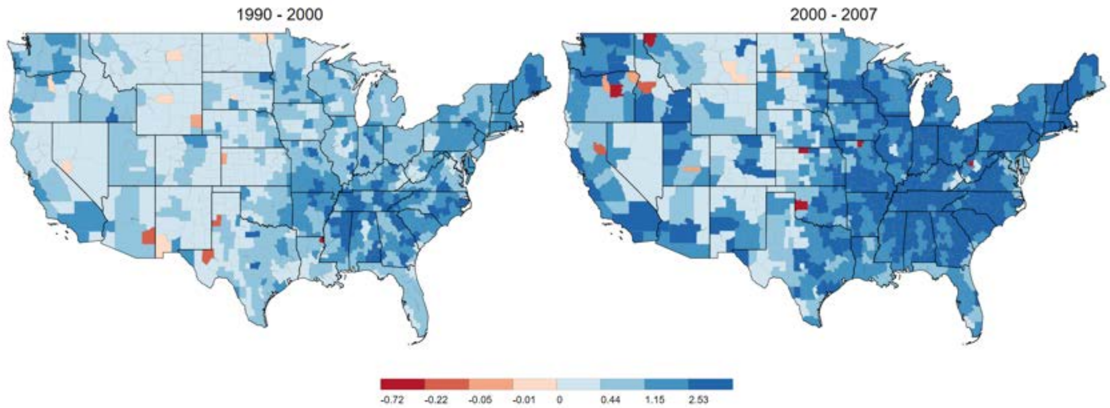


Figure 2.A.10: Decade-equivalent change in empl. share of working-age pop. in manuf.

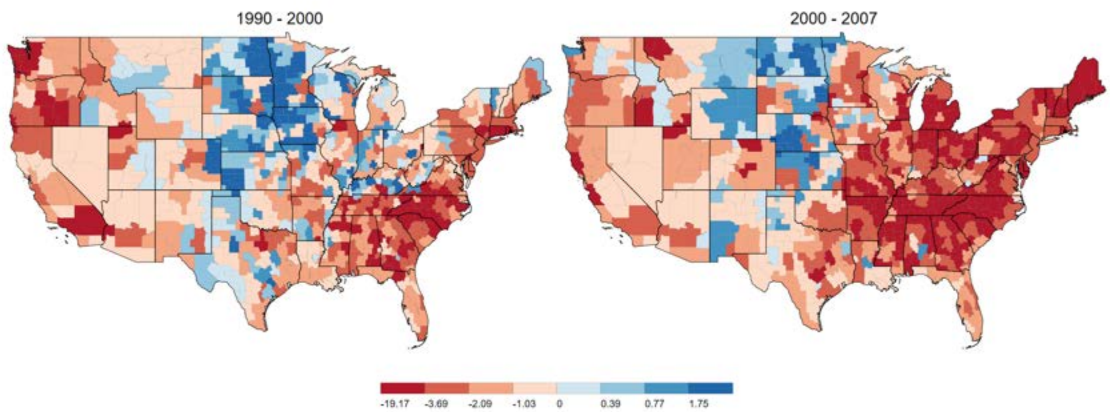


Figure 2.A.11: Decade-equivalent change in average household wage income per adult

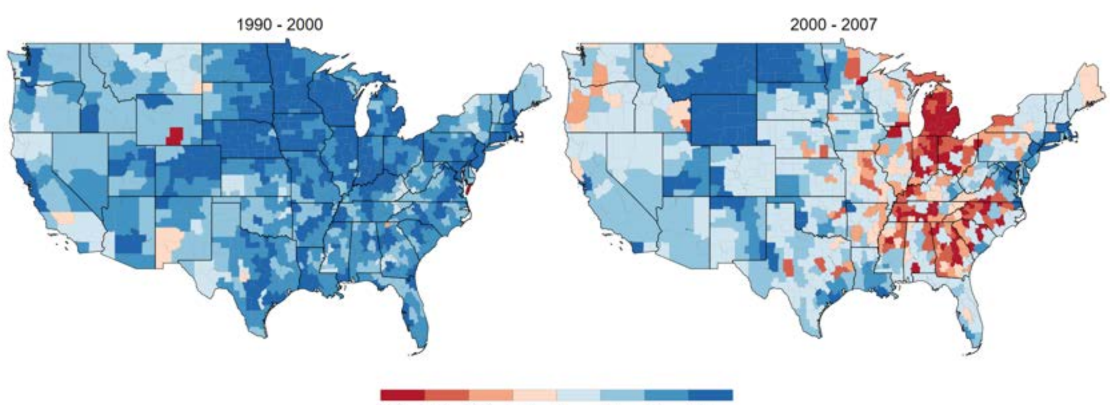
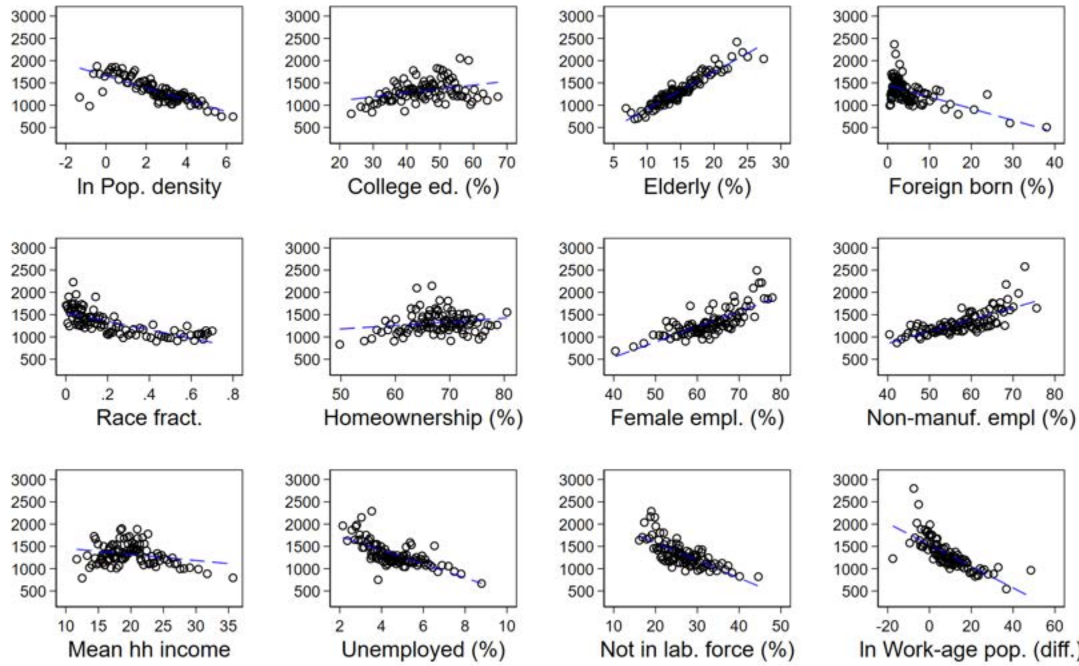
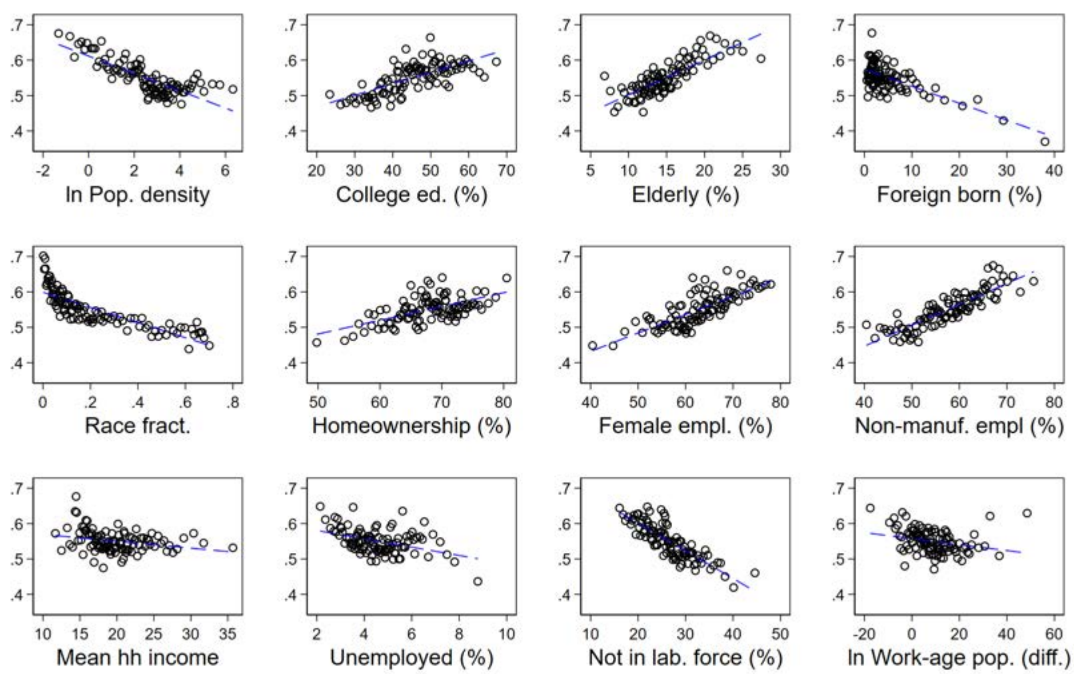


Figure 2.A.12: Binned scatterplots for selected covariates of associational density



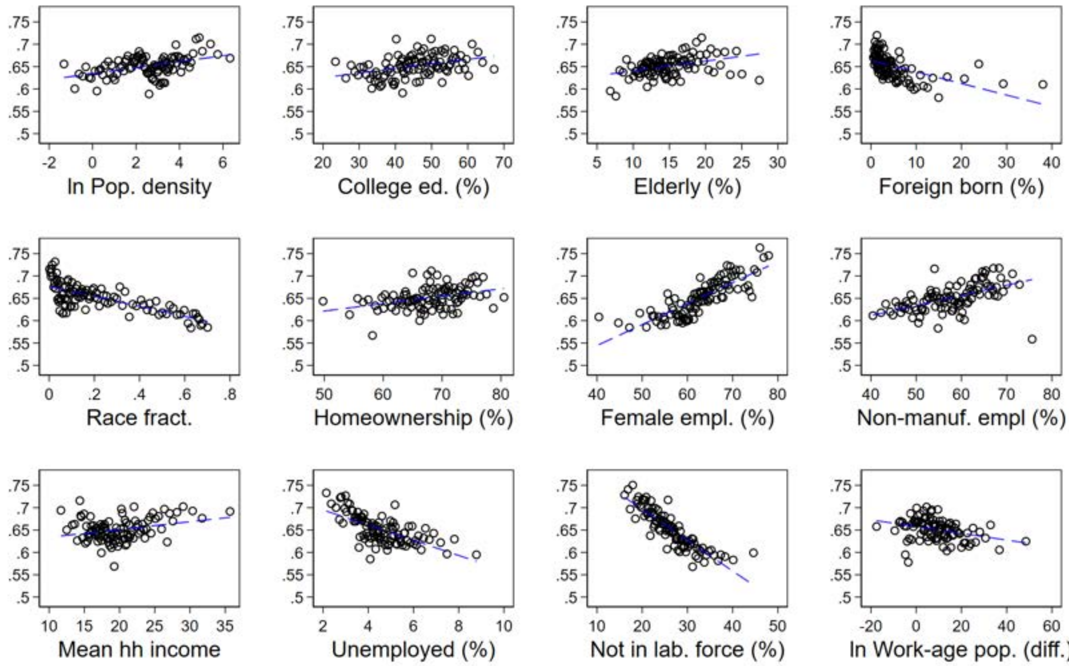
Binned scatter plots (100 bins). Linear fit lines in blue denote significance of the slope coefficient at the 95% level.

Figure 2.A.13: Binned scatterplots for selected covariates of voter turnout



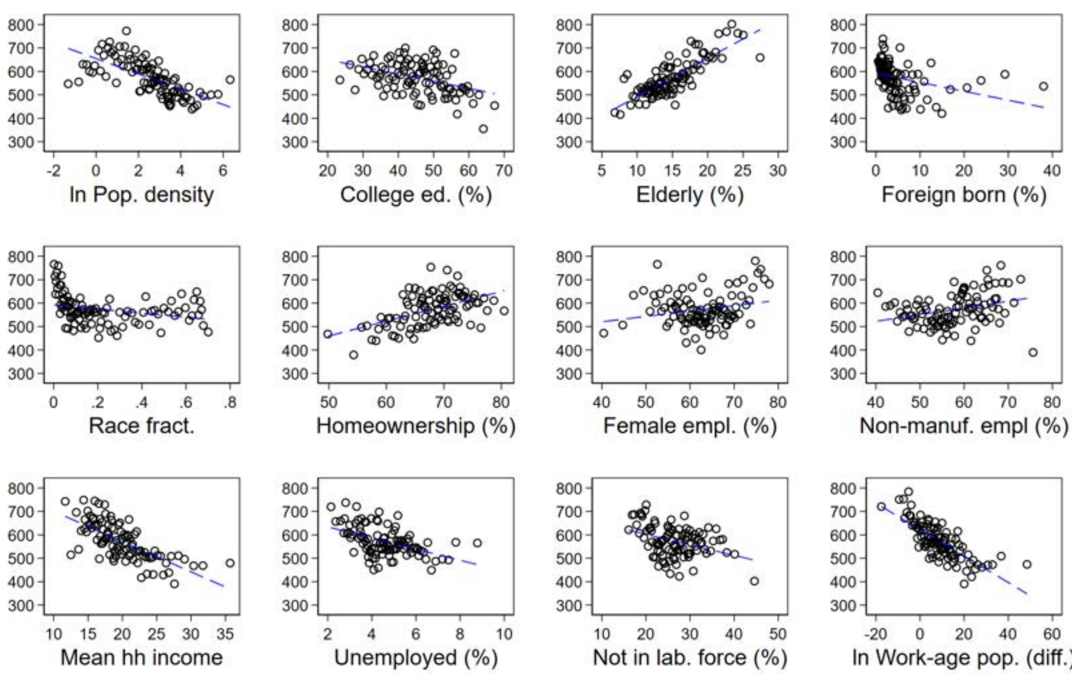
Binned scatter plots (100 bins). Linear fit lines in blue denote significance of the slope coefficient at the 95% level.

Figure 2.A.14: Binned scatterplots for selected covariates of Census mail response rate



Binned scatter plots (100 bins). Linear fit lines in blue denote significance of the slope coefficient at the 95% level.

Figure 2.A.15: Binned scatterplots for selected covariates of religious participation



Binned scatter plots (100 bins). Linear fit lines in blue denote significance of the slope coefficient at the 95% level.

Figure 2.A.16: Binned scatterplot for changes in social capital and US trade shocks

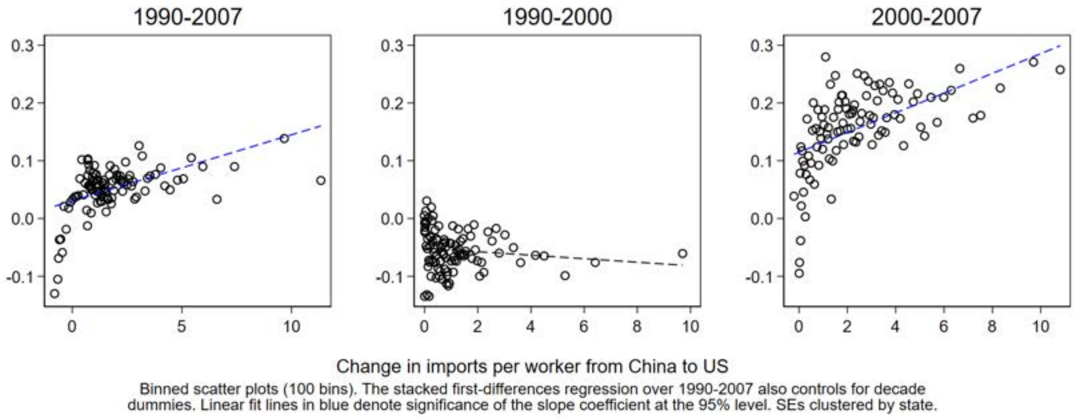


Figure 2.A.17: Binned scatterplot for changes in social capital and trade shocks (IV)

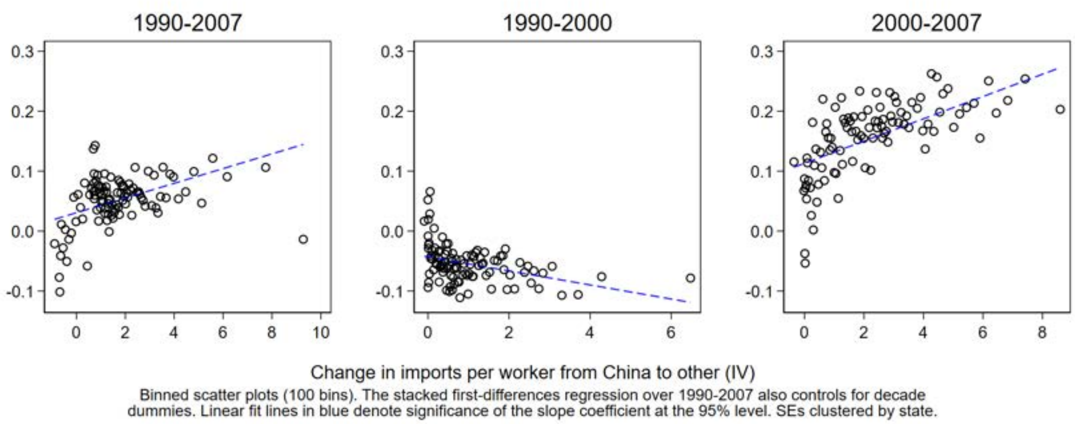
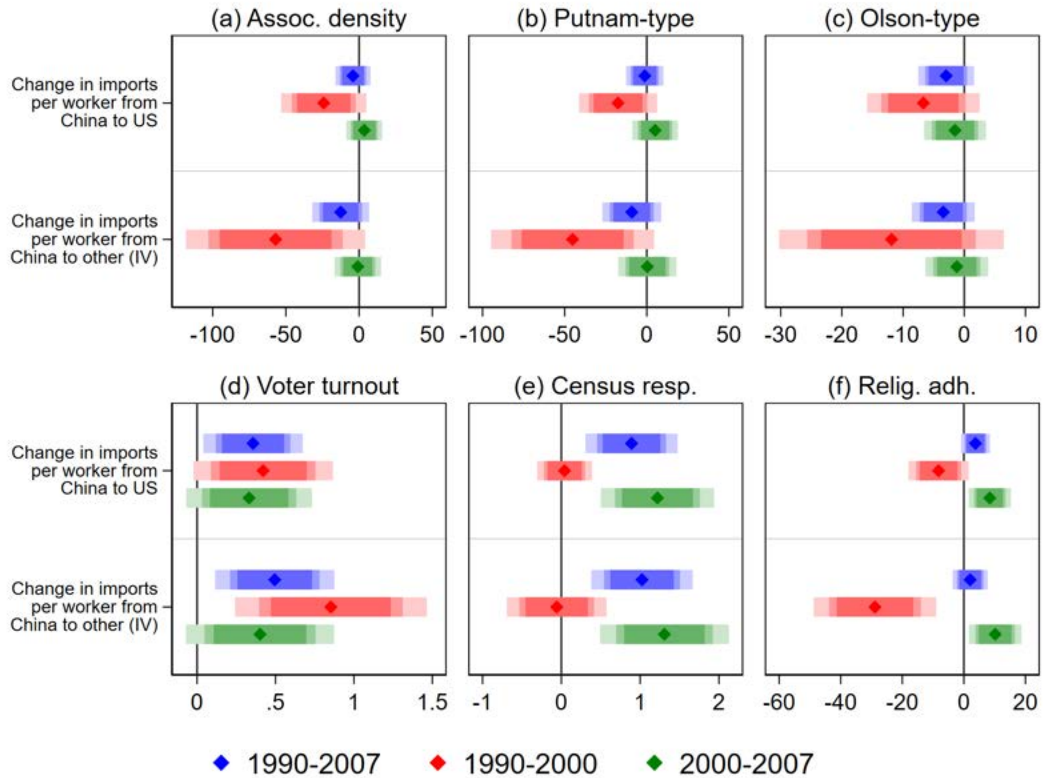


Figure 2.A.18: Coefficients plot for components of the social capital index using OLS



Each coefficient is obtained from a separate OLS regression. Standard errors clustered by state.

2.B Additional Tables

Table 2.B.1: Correlation matrix of social capital index variables

	Assoc. den.	Census resp.	Voter turnout	Rel. part.
Assoc. den.	1			
Census resp.	0.138^a	1		
Voter turnout	0.420^a	0.305^a	1	
Rel. part.	0.390^a	0.0809^a	0.119^a	1

Sig. levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All variables are standardised.

Table 2.B.2: Principal component analysis for social capital index

	PC1		PC2		PC3		PC4	
	Loading	Correl.	Loading	Correl.	Loading	Correl.	Loading	Correl.
Assoc. den.	0.600	0.796	-0.279	-0.282	0.302	0.262	-0.686	-0.468
Census resp.	0.379	0.502	0.652	0.659	0.624	0.541	-0.208	-0.142
Voter turnout	0.553	0.734	0.343	0.347	0.505	0.438	0.567	0.386
Rel. part.	0.436	0.578	-0.617	-0.624	0.515	0.447	0.406	0.277
Proportion	0.439		0.256		0.188		0.116	
Cumulative	0.439		0.696		0.884		1	

Table 2.B.3: Industry codes for associational density

Association	Type	SIC (87)	NAICS (97, 02)
Bowling centres	Putnam	7930	713950
Civic and social associations	Putnam	8640	813410
Physical fitness facilities	Putnam	7991	713940
Public golf courses	Putnam	7992	713910
Religious organisations	Putnam	8660	813110
Sports clubs, managers and promoters	Putnam	7941	711211
Business associations	Olson	8610	813910
Labour organisations	Olson	8630	813930
Political organisations	Olson	8650	813940
Professional organisations	Olson	8620	813920

Table 2.B.4: Summary statistics in levels and decade-equivalent changes

	1990							2000							Total						
	Mean	Std. Dev.	Min.	P25	P50	P75	Max.	Mean	Std. Dev.	Min.	P25	P50	P75	Max.	Mean	Std. Dev.	Min.	P25	P50	P75	Max.
Social capital	41.78	19.63	0	29	39	54	100	36.30	22.06	0	21	32	48	100	39.04	21.05	0	24	36	52	100
Assoc. density	1199.26	438.99	0	906	1134	1431	3264	1444.38	665.34	0	1046	1326	1692	6173	1321.82	576.63	0	962	1235	1554	6173
P-type density	1052.35	418.05	0	781	980	1265	3264	1266.55	602.77	0	897	1161	1480	5761	1159.45	529.47	0	832	1077	1365	5761
O-type density	146.91	102.52	0	87.35	133.57	188.22	1326	177.82	137.93	0	106.55	154.62	210.73	1753	162.37	122.46	0	96.80	143.66	200.93	1753
Census response	67.06	6.62	46	62	67	72	82	63.16	7.34	22	59	64	68	80	65.11	7.26	22	60	65	71	82
Voter turnout	56.95	9.32	30	50	57	63	87	53.08	8.33	27	47	53	59	78	55.02	9.04	27	49	55	61	87
Relig. part.	604.54	171.92	218	485	604	726	1000	540.49	178.51	157	404	515	645	1000	572.51	178.09	157	436	558	699	1000
Imports per worker (US)	0.36	0.64	0	0	0	0	9	1.12	1.90	0	0	1	1	41	0.74	1.47	0	0	0	1	41
Imports per worker (IV)	0.45	0.56	0	0	0	1	4	1.02	1.17	0	0	1	1	15	0.74	0.96	0	0	0	1	15
Pop. density (km2)	33.41	83.80	0	5	14	31	1631	37.30	91.99	0	5	16	35	1771	35.36	87.98	0	5	15	33	1771
College (%)	42.19	8.57	20	36	43	48	70	48.32	8.55	26	42	49	54	71	45.26	9.09	20	39	45	52	71
Elderly (%)	14.87	3.95	6	12	14	17	31	14.88	3.82	6	12	14	17	29	14.87	3.88	6	12	14	17	31
Homeownership (%)	69.89	5.09	42	67	70	73	82	65.67	5.37	41	63	66	69	81	67.78	5.64	41	64	68	72	82
Foreign born (%)	3.91	4.97	0	1	2	5	40	6.02	6.47	1	2	4	7	49	4.96	5.86	0	2	3	6	49
Race frac.	0.21	0.20	0	0	0	0	1	0.25	0.20	0	0	0	0	1	0.23	0.20	0	0	0	0	1
Women empl. (%)	61.14	6.63	33	57	62	66	77	64.33	7.11	41	60	64	69	80	62.74	7.06	33	58	63	68	80
Manuf. empl. CBP (%)	22.13	12.85	1	12	21	30	62	19.02	11.06	0	11	18	27	55	20.58	12.09	0	11	20	29	62
Mean hh income	18.09	3.63	11	16	17	20	34	21.71	4.16	13	19	21	24	40	19.90	4.30	11	17	19	22	40
Unemployed (%)	4.85	1.29	2	4	5	6	11	4.32	1.11	2	4	4	5	9	4.59	1.23	2	4	4	5	11
Not in lab. force (%)	26.73	5.20	16	23	26	30	48	26.39	5.55	15	23	26	30	49	26.56	5.38	15	23	26	30	49

	1990-2000							2000-2007							1990-2007						
	Mean	Std. Dev.	Min.	P25	P50	P75	Max.	Mean	Std. Dev.	Min.	P25	P50	P75	Max.	Mean	Std. Dev.	Min.	P25	P50	P75	Max.
Δ Social capital	-5.47	8.03	-46	-10	-5	-1	28	15.79	13.30	-53	8	16	24	47	5.16	15.29	-53	-6	2	17	47
Δ Assoc. density	245.12	354.45	-459	50	169	330	4054	52.79	298.90	-2064	-38	64	162	2374	148.95	341.56	-2064	0	107	238	4054
Δ P-type density	214.20	299.50	-459	49	151	294	3642	70.55	294.34	-2064	-12	77	165	2968	142.38	305.40	-2064	19	111	229	3642
Δ O-type density	30.91	111.32	-348	-14.05	9.97	49.54	1629	-17.75	106.01	-1163	-42.74	-13.00	0.00	839	6.58	111.35	-1163	-28.86	0.00	28.68	1629
Δ Census response	-3.91	5.55	-48	-6	-3	-0	9	10.87	9.22	-26	6	11	17	54	3.48	10.61	-48	-4	1	11	54
Δ Voter turnout	-3.87	3.98	-22	-6	-4	-1	9	7.26	4.67	-13	4	7	10	33	1.69	7.05	-22	-4	1	7	33
Δ Relig. part.	-64.04	102.65	-543	-113	-53	-5	288	-2.17	128.56	-620	-76	4	81	440	-33.10	120.33	-620	-95	-33	39	440
Δ Imports per worker (US)	1.14	1.44	-0	0	1	1	11	2.53	2.31	-1	1	2	3	11	1.84	2.04	-1	0	1	2	11
Δ Imports per worker (IV)	1.00	1.06	-0	0	1	1	9	2.42	2.07	-1	1	2	3	9	1.71	1.79	-1	0	1	3	9
Δ Mean hh income	3.62	1.50	-3	3	4	4	10	1.08	2.25	-6	-0	1	2	9	2.35	2.29	-6	1	2	4	10
Δ Unemployed (%)	-0.53	0.86	-3	-1	-0	0	2	0.34	1.61	-5	-1	0	1	5	-0.10	1.36	-5	-1	-0	1	5
Δ Not in lab. force (%)	-0.34	2.54	-12	-2	-0	1	8	-1.51	2.76	-15	-3	-1	0	7	-0.93	2.71	-15	-3	-1	1	8
Δ ln Working-age pop.	10.46	11.37	-26	3	10	16	65	7.14	9.05	-24	0	6	12	43	8.80	10.41	-26	2	8	14	65

Table 2.B.5: Baseline OLS regressions for associational density

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	-4.175	-24.16	3.574			
	(4.461)	(10.88)^b	(4.544)			
Δ IP China to other				-12.56	-57.06	-0.871
				(7.234)^c	(22.80)^b	(5.922)
Adj. R^2	0.0786	0.0082	-0.0006	0.0817	0.0279	-0.0014
R^2	0.0799	0.0096	0.0008	0.0830	0.0292	0.0000
N	1,444	722	722	1,444	722	722

Robust standard errors are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.6: Main 2SLS regressions for associational density

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	-15.02	-8.183	5.339	2.499	1.953	-47.93	-39.42	-14.67	-12.97
	(8.811)^c	(7.021)	(6.282)	(6.873)	(7.524)	(23.36)^b	(23.64)	(11.85)	(12.24)
Assoc. density (level)		0.0361	-0.0877	-0.123	-0.194	0.151	0.0900	-0.159	-0.196
		(0.0385)	(0.0418)^b	(0.0439)^a	(0.0425)^a	(0.0662)^b	(0.0730)	(0.0587)^a	(0.0580)^a
Manuf. empl. CBP (%)		-0.828	-0.211	0.424	0.131	3.323	2.916	3.650	2.387
		(1.349)	(1.203)	(1.462)	(1.880)	(2.607)	(2.763)	(1.883)^c	(2.107)
ln Density			-46.13	-68.01	-66.22	-99.31	-105.8	-35.22	-41.11
			(9.857)^a	(15.31)^a	(17.64)^a	(21.19)^a	(17.24)^a	(11.72)^a	(14.60)^a
College (%)			0.568	4.515	5.226	0.863	-1.509	5.033	6.244
			(2.170)	(2.249)^c	(1.795)^a	(3.334)	(2.326)	(2.625)^c	(2.690)^b
Elderly (%)			22.76	21.39	26.90	29.02	30.79	7.356	11.77
			(6.488)^a	(6.809)^a	(6.352)^a	(6.718)^a	(6.093)^a	(7.531)	(8.511)
Homeownership (%)			0.0827	-0.640	-0.349	-6.125	-8.338	-1.360	-0.0819
			(2.629)	(2.542)	(2.753)	(3.635)^c	(3.173)^b	(3.426)	(4.240)
Foreign born (%)			0.743	2.799	4.465	3.405	4.446	3.113	3.936
			(1.885)	(2.214)	(2.251)^c	(3.033)	(2.457)^c	(2.083)	(2.156)^c
Race frac.			106.4	66.43	11.50	305.5	234.9	-119.3	-147.9
			(51.52)^b	(53.67)	(63.79)	(85.54)^a	(87.05)^a	(64.03)^c	(78.13)^c
Women empl. (%)			2.070	-1.389	-2.137	-4.147	0.103	5.655	6.840
			(2.374)	(2.568)	(2.937)	(3.436)	(3.397)	(3.198)^c	(3.075)^b
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6674^a	0.6601^a	0.6441^a	0.6282^a	0.7200^a	0.6909^a	0.5436^a	0.5343^a
Kleibergen-Paap F	248.21	128.06	113.56	108.59	100.78	18.41	16.11	66.36	59.26
Adj. R^2	-0.0055	0.0016	0.0869	0.0791	0.1010	0.2679	0.2624	0.0401	0.0628
R^2	-0.0034	0.0050	0.0944	0.0919	0.1085	0.2872	0.2736	0.0653	0.0771
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.7: Baseline OLS regressions for Putnam-type associational density

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	-1.219	-17.48	5.086			
	(4.239)	(8.864)^c	(5.174)			
Δ IP China to other				-9.121	-45.17	0.347
				(6.679)	(18.50)^b	(6.603)
Adj. R^2	0.0541	0.0057	0.0002	0.0565	0.0243	-0.0014
R^2	0.0554	0.0070	0.0016	0.0578	0.0256	0.0000
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.8: Main 2SLS regressions for Putnam-type associational density

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	-10.91	-7.405	3.774	1.526	0.714	-35.61	-30.58	-16.30	-14.83
	(8.094)	(6.576)	(5.697)	(6.298)	(6.678)	(17.70)^c	(17.24)^c	(12.30)	(12.80)
P-type density (level)		0.0235	-0.103	-0.137	-0.211	0.124	0.0722	-0.208	-0.253
		(0.0382)	(0.0458)^b	(0.0467)^a	(0.0455)^a	(0.0522)^b	(0.0525)	(0.0708)^a	(0.0762)^a
Manuf. empl. CBP (%)		-0.423	0.0255	0.608	0.531	1.969	2.244	4.601	3.305
		(1.097)	(0.985)	(1.216)	(1.507)	(2.034)	(2.046)	(2.249)^b	(2.551)
ln Density			-37.78	-55.72	-56.40	-70.26	-76.59	-40.49	-49.52
			(8.905)^a	(13.58)^a	(15.69)^a	(18.81)^a	(15.85)^a	(14.36)^a	(17.21)^a
College (%)			0.452	3.626	4.367	0.300	-0.796	4.970	6.042
			(2.118)	(2.145)^c	(1.716)^b	(3.253)	(2.008)	(2.767)^c	(2.799)^b
Elderly (%)			19.99	19.10	24.60	22.38	23.84	11.54	16.74
			(5.761)^a	(5.886)^a	(5.515)^a	(6.596)^a	(6.304)^a	(7.055)	(8.052)^b
Homeownership (%)			0.224	-0.459	-0.357	-5.707	-6.045	-0.587	0.0101
			(2.311)	(2.236)	(2.435)	(3.578)	(3.163)^c	(3.083)	(3.806)
Foreign born (%)			-0.414	1.416	3.561	0.112	3.246	2.637	3.715
			(1.483)	(1.818)	(2.096)^c	(2.471)	(2.078)	(1.920)	(2.107)^c
Race frac.			86.19	59.29	13.18	243.1	191.5	-79.85	-95.13
			(41.48)^b	(44.50)	(53.30)	(76.43)^a	(77.43)^b	(54.38)	(68.49)
Women empl. (%)			1.622	-1.761	-2.390	-5.103	-1.723	5.410	7.514
			(2.224)	(2.647)	(2.993)	(3.239)	(2.679)	(3.343)	(3.746)^c
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6669^a	0.6603^a	0.6446^a	0.6285^a	0.7193^a	0.6906^a	0.5444^a	0.5347^a
Kleibergen-Paap F	248.21	128.16	113.85	108.55	100.62	18.54	16.26	66.46	59.34
Adj. R^2	-0.0060	-0.0030	0.0712	0.0681	0.0948	0.2219	0.2164	0.0557	0.0829
R^2	-0.0039	0.0005	0.0789	0.0810	0.1023	0.2424	0.2284	0.0805	0.0969
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.9: Baseline OLS regressions for Olson-type associational density

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	-2.956	-6.681	-1.512			
	(1.699)^c	(3.425)^c	(1.880)			
Δ IP China to other				-3.438	-11.89	-1.219
				(1.914)^c	(6.841)^c	(1.911)
Adj. R^2	0.0491	0.0061	-0.0003	0.0491	0.0115	-0.0008
R^2	0.0504	0.0074	0.0011	0.0504	0.0129	0.0006
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.10: Main 2SLS regressions for Olson-type associational density

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	-4.113	-6.650	-1.606	-1.606	-0.626	-17.01	-12.78	-0.122	0.288
	(2.317)^c	(2.734)^b	(2.383)	(2.440)	(2.907)	(8.266)^b	(9.841)	(2.253)	(2.476)
O-type density (level)		-0.253	-0.276	-0.289	-0.311	-0.142	-0.139	-0.305	-0.315
		(0.0956)^b	(0.112)^b	(0.121)^b	(0.131)^b	(0.119)	(0.127)	(0.101)^a	(0.110)^a
Manuf. empl. CBP (%)		-0.221	-0.216	-0.225	-0.431	1.115	0.547	-0.952	-0.928
		(0.387)	(0.424)	(0.443)	(0.593)	(0.792)	(0.935)	(0.793)	(0.976)
ln Density			-8.366	-14.29	-12.02	-27.14	-28.20	-3.426	-0.566
			(2.330)^a	(3.901)^a	(5.438)^b	(6.403)^a	(9.814)^a	(4.005)	(6.316)
College (%)			-0.0133	0.971	0.870	0.892	-0.241	-0.119	-0.167
			(0.426)	(0.475)^b	(0.642)	(0.603)	(0.936)	(0.887)	(1.202)
Elderly (%)			3.695	2.998	3.211	6.494	6.817	-0.340	-0.421
			(2.210)	(2.311)	(2.491)	(3.141)^b	(3.044)^b	(2.039)	(2.343)
Homeownership (%)			-0.550	-0.505	-0.318	-0.679	-2.318	-1.516	-1.061
			(0.852)	(0.771)	(0.924)	(1.647)	(2.031)	(0.632)^b	(0.896)
Foreign born (%)			0.468	0.894	0.671	1.678	0.399	0.291	0.371
			(0.546)	(0.657)	(0.577)	(1.021)	(0.572)	(0.685)	(0.836)
Race frac.			-0.148	-8.199	-11.59	23.46	19.29	-43.11	-54.66
			(15.43)	(15.80)	(18.90)	(18.72)	(17.56)	(23.13)^c	(31.34)^c
Women empl. (%)			1.261	1.057	1.030	1.432	2.266	2.124	1.646
			(0.729)^c	(0.779)	(1.133)	(0.937)	(1.461)	(1.297)	(1.846)
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6679^a	0.6605^a	0.6454^a	0.6294^a	0.7173^a	0.6843^a	0.5447^a	0.5362^a
Kleibergen-Paap F	248.21	129.90	113.07	108.62	100.46	18.52	16.42	65.57	59.00
Adj. R^2	0.0002	0.0742	0.1157	0.1100	0.1178	0.1230	0.1249	0.1255	0.1357
R^2	0.0023	0.0774	0.1230	0.1224	0.1251	0.1461	0.1382	0.1485	0.1489
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. Associational density is expressed in terms of organisations per million population.

Table 2.B.11: Baseline OLS regressions for voter turnout

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	0.357	0.421	0.332			
	(0.118)^a	(0.166)^b	(0.149)^b			
Δ IP China to other				0.495	0.853	0.401
				(0.142)^a	(0.228)^a	(0.176)^b
Adj. R^2	0.6313	0.0218	0.0256	0.6352	0.0505	0.0304
R^2	0.6318	0.0231	0.0269	0.6357	0.0518	0.0317
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Voter turnout is expressed in terms of percentage points.

Table 2.B.12: Main 2SLS regressions for voter turnout

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	0.592	0.266	0.100	0.00847	0.0266	-0.472	-0.329	0.185	0.127
	(0.172)^a	(0.216)	(0.187)	(0.179)	(0.169)	(0.271)^c	(0.238)	(0.156)	(0.161)
Voter turnout (level)		-0.149	-0.162	-0.161	-0.231	-0.213	-0.241	-0.113	-0.148
		(0.0497)^a	(0.0406)^a	(0.0378)^a	(0.0340)^a	(0.0424)^a	(0.0381)^a	(0.0501)^b	(0.0445)^a
Manuf. empl. CBP (%)		0.0145	-0.0172	-0.0168	-0.00830	0.0190	0.0239	-0.0559	-0.0402
		(0.0276)	(0.0244)	(0.0201)	(0.0190)	(0.0236)	(0.0185)	(0.0381)	(0.0306)
ln Density			0.550	0.290	-0.0413	0.0806	-0.283	0.497	0.233
			(0.196)^a	(0.196)	(0.147)	(0.251)	(0.194)	(0.290)^c	(0.200)
College (%)			0.0258	0.0221	0.146	0.0192	0.105	-0.0208	0.0706
			(0.0415)	(0.0411)	(0.0369)^a	(0.0418)	(0.0326)^a	(0.0533)	(0.0408)^c
Elderly (%)			0.0139	0.0211	0.0752	0.0166	0.0676	0.0312	0.0377
			(0.0558)	(0.0482)	(0.0548)	(0.0682)	(0.0666)	(0.0750)	(0.0663)
Homeownership (%)			0.0304	-0.000240	0.0131	0.0261	0.000583	-0.0546	-0.0706
			(0.0510)	(0.0444)	(0.0383)	(0.0640)	(0.0471)	(0.0485)	(0.0406)^c
Foreign born (%)			-0.119	-0.0916	-0.0600	-0.0737	-0.0433	-0.1000	-0.0604
			(0.0304)^a	(0.0399)^b	(0.0254)^b	(0.0729)	(0.0691)	(0.0382)^b	(0.0224)^a
Race frac.			4.289	3.603	2.756	-0.488	0.758	7.446	4.713
			(1.013)^a	(0.859)^a	(0.705)^a	(1.015)	(0.890)	(1.587)^a	(1.391)^a
Women empl. (%)			0.113	0.0874	0.00534	0.116	0.0340	0.108	0.0838
			(0.0351)^a	(0.0452)^c	(0.0388)	(0.0474)^b	(0.0312)	(0.0613)^c	(0.0548)
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6723^a	0.6609^a	0.6458^a	0.6300^a	0.7140^a	0.6782^a	0.5454^a	0.5365^a
Kleibergen-Paap F	248.21	129.28	113.39	108.05	100.02	18.78	16.54	65.51	59.31
Adj. R^2	0.0121	0.1205	0.2586	0.1314	0.1631	0.1097	0.1598	0.2085	0.2284
R^2	0.0141	0.1236	0.2648	0.1434	0.1701	0.1331	0.1727	0.2293	0.2402
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. Voter turnout is expressed in terms of percentage points.

Table 2.B.13: Baseline OLS regressions for Census response rate

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	0.891	0.0408	1.220			
	$(0.218)^a$	(0.130)	$(0.268)^a$			
Δ IP China to other				1.024	-0.0588	1.309
				$(0.240)^a$	(0.236)	$(0.305)^a$
Adj. R^2	0.5108	-0.0013	0.0919	0.5100	-0.0013	0.0851
R^2	0.5114	0.0001	0.0932	0.5107	0.0001	0.0864
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Census response rate is expressed in terms of percentage points.

Table 2.B.14: Main 2SLS regressions for Census response rate

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	1.225	1.242	0.542	0.578	0.718	-0.530	-0.269	-0.310	-0.191
	$(0.289)^a$	$(0.299)^a$	$(0.298)^c$	$(0.312)^c$	$(0.307)^b$	$(0.232)^b$	(0.227)	(0.381)	(0.365)
Census response (level)		-0.283	-0.474	-0.490	-0.508	-0.400	-0.403	-0.579	-0.504
		$(0.0443)^a$	$(0.0679)^a$	$(0.0760)^a$	$(0.0771)^a$	$(0.0663)^a$	$(0.0650)^a$	$(0.112)^a$	$(0.0801)^a$
Manuf. empl. CBP (%)		0.00951	-0.0228	-0.0457	-0.0425	0.00412	0.0234	0.108	0.0750
		(0.0298)	(0.0351)	(0.0327)	(0.0338)	(0.0223)	(0.0221)	$(0.0618)^c$	(0.0566)
ln Density			2.310	2.340	3.108	1.815	2.184	3.072	3.763
			$(0.224)^a$	$(0.305)^a$	$(0.337)^a$	$(0.315)^a$	$(0.424)^a$	$(0.402)^a$	$(0.412)^a$
College (%)			-0.0179	-0.0240	-0.0554	-0.0158	-0.0111	-0.0950	-0.168
			(0.0468)	(0.0571)	(0.0692)	(0.0510)	(0.0686)	(0.0727)	$(0.0735)^b$
Elderly (%)			-0.101	-0.141	-0.112	-0.139	-0.193	-0.137	-0.0545
			(0.0988)	(0.0888)	(0.0883)	(0.107)	(0.117)	(0.120)	(0.0941)
Homeownership (%)			-0.0486	-0.0418	-0.0613	-0.143	-0.0386	0.0278	-0.0509
			(0.0594)	(0.0553)	(0.0742)	(0.107)	(0.0890)	(0.0597)	(0.0724)
Foreign born (%)			-0.140	-0.153	-0.268	-0.0732	-0.0518	-0.152	-0.343
			$(0.0376)^a$	$(0.0457)^a$	$(0.0471)^a$	(0.0637)	(0.0529)	$(0.0722)^b$	$(0.0573)^a$
Race frac.			-4.669	-5.910	-6.289	-3.117	-3.485	-8.126	-7.965
			$(1.700)^a$	$(1.771)^a$	$(1.763)^a$	(1.982)	$(1.849)^c$	$(3.128)^b$	$(2.705)^a$
Women empl. (%)			0.171	0.201	0.0582	0.177	0.163	0.417	0.246
			$(0.0712)^b$	$(0.0890)^b$	(0.0830)	$(0.0571)^a$	$(0.0717)^b$	$(0.164)^b$	$(0.105)^b$
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6680^a	0.6617^a	0.6455^a	0.6281^a	0.7231^a	0.6904^a	0.5451^a	0.5357^a
Kleibergen-Paap F	248.21	130.05	114.46	108.99	100.35	18.90	16.47	65.86	59.04
Adj. R^2	0.0414	0.1064	0.2848	0.1995	0.2144	0.2539	0.2887	0.2270	0.2200
R^2	0.0434	0.1095	0.2908	0.2106	0.2209	0.2736	0.2996	0.2474	0.2319
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. Stacked first differences models control for a time dummy. Census response rate is expressed in terms of percentage points.

Table 2.B.15: Baseline OLS regressions for adherence rate to congregations

	(1) 1990-2007	(2) 1990-2000	(3) 2000-2007	(4) 1990-2007	(5) 1990-2000	(6) 2000-2007
Δ IP China to US	3.768	-8.205	8.410			
	$(1.750)^b$	$(3.644)^b$	$(2.553)^a$			
Δ IP China to other				2.047	-28.86	10.17
				(2.118)	$(7.403)^a$	$(3.179)^a$
Adj. R^2	0.0685	0.0118	0.0214	0.0656	0.0879	0.0255
R^2	0.0698	0.0132	0.0228	0.0669	0.0891	0.0268
N	1,444	722	722	1,444	722	722

Robust standard errors in parentheses are clustered by state. Significance levels: ${}^a p < 0.01$; ${}^b p < 0.05$; ${}^c p < 0.1$. All models include a constant. Stacked first differences models control for a time dummy. Adherence rate to religious groups is expressed in terms of adherents per thousand population.

Table 2.B.16: Main 2SLS regressions for adherence rate to congregations

	1990-2007					1990-2000		2000-2007	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ IP China to US	2.449	8.447	10.22	11.69	10.22	-11.20	-9.688	3.555	1.023
	(2.514)	$(3.633)^b$	$(3.790)^a$	$(4.047)^a$	$(3.951)^b$	$(6.046)^c$	(6.123)	(3.791)	(4.196)
Relig. part. (level)		-0.250	-0.256	-0.313	-0.393	-0.206	-0.155	-0.240	-0.320
		$(0.0287)^a$	$(0.0337)^a$	$(0.0429)^a$	$(0.0517)^a$	$(0.0365)^a$	$(0.0466)^a$	$(0.0643)^a$	$(0.0523)^a$
Manuf. empl. CBP (%)		-1.514	-1.079	-0.961	-1.034	-0.152	0.455	0.693	-0.128
		$(0.502)^a$	$(0.564)^c$	$(0.556)^c$	$(0.563)^c$	(0.437)	(0.353)	(0.828)	(0.816)
ln Density			-3.562	2.933	8.592	-9.632	-10.81	16.42	19.43
			(2.976)	(3.589)	$(4.462)^c$	$(4.133)^b$	$(5.656)^c$	$(6.281)^b$	$(9.013)^b$
College (%)			1.655	1.602	0.876	-0.370	0.158	2.526	0.654
			$(0.885)^c$	$(0.904)^c$	(0.851)	(0.869)	(0.853)	(1.632)	(1.109)
Elderly (%)			1.346	1.793	3.330	5.457	4.488	-4.150	-2.114
			(1.281)	(1.374)	$(1.525)^b$	$(0.889)^a$	$(0.937)^a$	$(2.365)^c$	(2.028)
Homeownership (%)			0.905	1.008	0.715	0.127	0.612	0.916	-0.00240
			(0.989)	(0.881)	(0.933)	(0.891)	(1.044)	(1.286)	(0.964)
Foreign born (%)			0.109	0.304	0.490	0.412	0.969	-0.658	-0.810
			(0.866)	(1.025)	(1.558)	(1.509)	(1.597)	(1.022)	(1.255)
Race frac.			-1.216	11.99	-22.64	-70.28	-77.37	67.88	25.73
			(18.65)	(19.17)	(25.23)	$(32.59)^b$	$(33.92)^b$	$(35.26)^c$	(24.90)
Women empl. (%)			-1.731	-2.368	-2.710	0.872	0.356	-1.993	1.181
			$(0.923)^c$	$(1.109)^b$	$(1.332)^b$	(1.043)	(1.294)	(2.048)	(1.471)
Fixed effects	No	No	No	Div.	State	Div.	State	Div.	State
First stage coeff.	0.8360^a	0.6655^a	0.6609^a	0.6449^a	0.6283^a	0.7158^a	0.6870^a	0.5461^a	0.5381^a
Kleibergen-Paap F	248.21	127.94	113.93	108.11	100.50	18.85	16.70	65.51	59.02
Adj. R^2	0.0013	0.1642	0.1699	0.1730	0.2099	0.1758	0.1117	0.2747	0.3008
R^2	0.0034	0.1671	0.1768	0.1844	0.2165	0.1975	0.1252	0.2938	0.3115
N	1,444	1,444	1,444	1,444	1,444	722	720	722	720

Robust standard errors in parentheses are clustered by state. Significance levels: ${}^a p < 0.01$; ${}^b p < 0.05$; ${}^c p < 0.1$. Stacked first differences models control for a time dummy. Adherence rate to religious groups is expressed in terms of adherents per thousand population.

Chapter 3

Social Connectedness and Knowledge Flows[1]

3.1 Introduction

Do inventors learn from the informal context that surrounds them? This paper empirically examines the role of social connectedness in the diffusion of knowledge among agents located across distant geographies. Social connectedness is conceptualised as the overall informal social environment of an agent, measured by the aggregate ties connecting the agent's neighbourhood to other neighbourhoods, net of her formal, professional, networks. The research question we address, therefore, is whether stronger informal social ties to other places can foster knowledge exchange with these places, above and beyond what would be explained by professional channels or by simple geographic proximity. While the paper is conceptually interested in the general case of knowledge flows, the empirical analysis focuses on patent citations. Citations provide a powerful measure of economically relevant knowledge exchange, otherwise difficult to observe in different settings. Moreover, they speak to the process of innovation and technological change, which is a key determinant of long run economic growth (Romer, 1986, 1990; Lucas, 1988; Aghion and Howitt, 1992).

This research relates to an old question in economics that considers the role of localised knowledge spillovers in promoting the agglomeration of people and industries in space (Marshall, 1890). As individuals come together and interact, they learn from each other and become more productive (Glaeser, 1999). With respect to innovation, the sharing and recombination of existing ideas supports the creation of more ideas (Carlino et al., 2007).

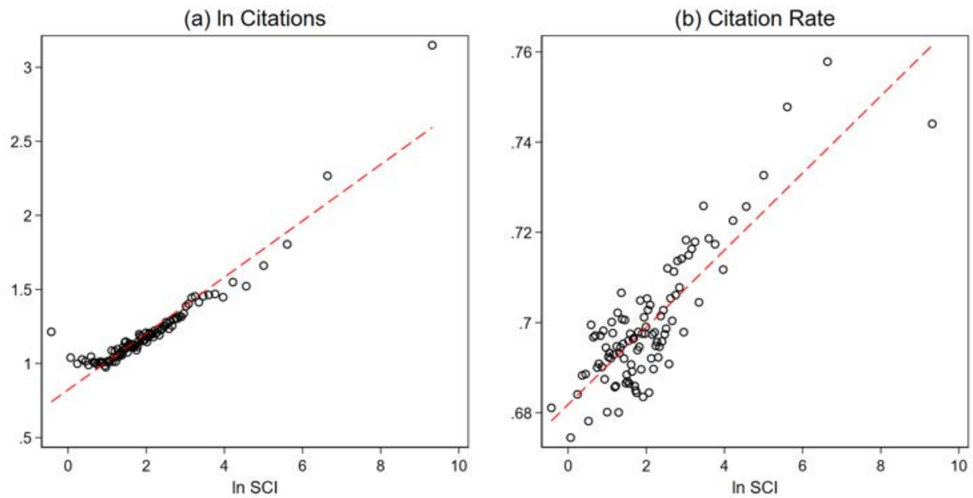
[1] This chapter is partially based on materials that are part of a bigger research agenda developed with Tanner Regan. His contributions to fruitful discussions are fully acknowledged, and this chapter is submitted as single-authored with his agreement.

A large body of empirical research has attempted to validate the notion of knowledge spillovers, frequently using patent data and patent citations to measure innovation and knowledge transfers. In keeping with the notion of agglomeration, these studies typically focus on the geographical dimension of spillovers (Jaffe et al., 1993; Audretsch and Feldman, 1996; Thompson and Fox-Kean, 2005; Murata et al., 2013). Yet numerous papers emphasise that the mechanisms underlying the spread of knowledge, whether intentional or unintentional, rely on interaction of people over networks (Bala and Goyal, 2000; Powell and Grodal, 2005; Henderson, 2007). Knowledge exchanges are only localised to the extent that physical proximity shapes the quantity and quality of social connections (Breschi and Lissoni, 2001; Storper and Venables, 2004). The increasing availability of data on networks and interaction has thus spurred a growing empirical literature that evaluates the role of social ties in the exchange of knowledge. Some papers consider inventor networks constructed from data on co-patenting (Agrawal et al., 2006; Breschi and Lissoni, 2009), others examine social proximity measures inferred from belonging to similar ethnic groups (Agrawal et al., 2008; Kerr, 2008). Our efforts focus on the role of informal connections across places, above and beyond what would be explained by professional relationships (co-inventor networks). We exploit a new broad and direct measure of social connectedness across counties in the US based on aggregate counts of the universe of online friendships on Facebook, a popular social media platform (Bailey et al., 2018b). The binned scatterplots in Figure 3.1 show the correlations of the log of social connectedness for county pairs with (a) the log of aggregate citation counts between the two counties, and (b) the propensity of inventors in these counties to cite one another. The strong and positive associations between these variables is the motivating fact underlying this analysis.[2]

Bailey et al. (2018b)'s comprehensive data on social connectedness allows us for the first time to study knowledge flows over informal networks on a large scale without relying on indirect proxies. Our empirical analysis jointly estimates the roles of geographic and social proximity in knowledge diffusion, explicitly focusing on informal ties rather than professional inventor networks, which are controlled for separately. We identify a small but significant and robust effect of informal ties between places on their propensity to cite one another. This is independent of geographical distance or professional linkages between inventors and takes into account the endogenous location of relevant knowledge due to the pre-existing geography of production. Interestingly, the effect of physical proximity (across regions, not within) is statistically insignificant once controlling for social connectedness

[2] Figures 3.A.1 and 3.A.2 in Appendix provide detailed maps that visualise the correlation between connectedness and citation networks for selected US counties.

Figure 3.1: Binned scatterplots for US citations by patents issued in 2016-2019



Citations are over 2016-2019. Citation rate is the number of applicant-added citations over all citations. Both graphs control for the log of citing- times cited-county population.

and professional inventor networks, suggesting that the latter two are more powerful predictors of knowledge exchange between places, certainly important to consider in addition to plain geographical distance. According to our preferred estimate, two counties at the 75$^{\text{th}}$ percentile of social connectedness are on average 1.2 percentage points more likely to cite one another than a pair of counties at the 25$^{\text{th}}$ percentile. While this result may appear somewhat abstract, a look at the data reveals that two otherwise neighbouring counties may occasionally display such a difference in connectedness strength with the same third county. The relationship we document is thus economically meaningful as it can potentially be achieved by simply crossing a border. We also show that the relevance of informal social ties has increased over time over the past two decades. Further, effects appears to be stronger for entrepreneurs (assignees observed patenting for the first time), for patents that are common domain in a geographical sense, and for knowledge more distant in the technological space.

The rest of the paper is organised as follows. Section 3.2 frames the problem conceptually. Section 3.3 discusses the data and the empirical methods. Section 3.4 presents the results of the analysis. Section 3.5 concludes highlighting limitations and future work.

3.2 Conceptual Framework and Related Literature

By critically reviewing the extant literature, what follows conceptualises the importance of social connectedness in knowledge exchange and how it relates to physical geography.

It also outlines key questions to be examined empirically and highlights the intended contribution of this paper to the relevant scholarship.

3.2.1 Geography and Social Learning

The work of Jaffe et al. (1993) pioneered research on the geography of knowledge flows and spillovers by documenting the propensity of patent citations to be localised in space. Thereafter, localisation of knowledge flows has been the subject of extensive investigation by literatures in urban and regional economics (Carlino and Kerr, 2015), economics of innovation (Audretsch and Feldman, 2004; Jaffe and de Rassenfosse, 2017), and regional science (Autant-Bernard et al., 2007b). Local knowledge exchange, or learning, is in fact one of the key drivers of urban agglomeration externalities (Duranton and Puga, 2004) and typically increases with density (Carlino et al., 2007). Despite the large body of theoretical and empirical studies documenting the geographical boundedness of learning, however, the micro-channels underlying this relationship are relatively understudied. What determines learning and why might learning be confined in space? An emerging consensus points to the key role of interaction over social networks. Among others, this is emphasised by Henderson (2007) with respect to the urban economics literature, and by Saxenian (1996), Breschi and Lissoni (2001), Feldman (2002), and Powell and Grodal (2005) with respect to innovation.[3] As people socialise, interpersonal communication sustains the exchange of new ideas and knowledge. Learning from interaction with others, henceforth also referred to as *social learning*, could be productivity enhancing compared to relying on autonomous efforts and search on books and archives (Lucas and Moll, 2014; Akcigit et al., 2018). Moreover, because knowledge can be 'sticky' due to its often tacit and embedded nature (Polanyi, 1966), successful transmission may sometimes require face-to-face communication (Storper and Venables, 2004).

Social interaction however is itself costly, as time and money go into the development and cultivation of social ties. Bala and Goyal (2000) present a model where network formation depends on each agent's trade-off of the costs of establishing and maintaining ties and the expected benefits obtained from these ties. This line of reasoning suggests that the exchange of knowledge via social interaction is localised insofar as physical distance determines the quantity and quality of socialisation. Social learning, it would seem, is more likely between agents located close to each other due for instance to lower commuting times and costs or increased frequency of meetings (Kim et al., 2020). Kerr and Kominers (2014) relate the role of agglomeration forces and interaction costs between firms to the emergence

[3] Goyal (2011) provides a more general discussion by reviewing theoretical models for social learning.

of clusters of different sizes and shapes across different industries. Catalini et al. (2020) show that travel costs play an important role in shaping collaborations between scientists, exploiting the introduction of low-cost airlines for identification. Agrawal et al. (2017) show that this is true even over small geographical areas. The paper documents a causal link between road networks and innovation, showing that a greater stock of interstate highways crossing a metropolitan area (MSA) is associated with a higher propensity to innovate and a greater probability of a localised citation conditioning on within-MSA distance between inventors. According to the authors, by improving local connectivity, roads allow inventors to access knowledge that, albeit local, is relatively more distant to them. Percoco (2016) provides similar evidence for Italy with respect to firm location and urban development. Other contributions in this area of research emphasise the role of geographic proximity, along with institutional and cognitive similarities, in determining collaboration patterns (Crescenzi et al., 2016, 2017). Autant-Bernard et al. (2007a) confirm the presence of network effects in R&D cooperation, which can compensate for physical distance. Moreover, the co-location of agents increases the probability of unintentional meetings that lead to the exchange of relevant knowledge. Cities, as a result, offer optimal environments for learning due to high density living (Glaeser, 1999; De la Roca and Puga, 2017; Davis and Dingel, 2019). These mechanisms lie at the heart of the reduced-form relationship between physical distance and knowledge flows widely documented in the literature. Breschi and Lissoni (2009), in particular, show that the localisation of patent citations is largely determined by the fact that citing inventors draw on a social network that is localised due to the limited geographical reach of their inter-firm mobility.

3.2.2 Heavy and Light Knowledge Conveyors in Innovation

There are several specific ways in which social learning may take place. Individuals could learn from friends, family, or colleagues. They could also learn in more indirect ways through friends of friends (e.g., via referral) and their broader social environment. Some types of knowledge might also diffuse because of reputation and status effects in the network, or because they are specific to certain cultural or ethnic groups (Agrawal et al., 2008). More generally, networks could facilitate exchange of knowledge in intended ways, due to higher trust among connected members and the belief that the exchange will be reciprocated in the future (Helsley and Strange, 2004). It could also be that knowledge is transferred unintentionally, for instance during a conversation in passing where one retains latent knowledge not purposefully obtained with some specific aim in mind (Breschi and Lissoni, 2001; Breschi et al., 2005). Finally, it seems plausible that different types of knowledge rely more intensely on different types of channels. Hollywood gossip, for instance, is more likely to travel over informal friendship ties than insights on regulatory developments in the pharmaceutical industry would be. A systematic survey and discussion of possible transfer mechanisms over social networks for different types of knowledge falls beyond the scope of this paper. Instead, the conceptual discussion that follows focuses on the kind of scientific and technical knowledge found in patents (ideas or inventions). The reasons for this are twofold. First, patents embody knowledge that is economically relevant and that determines, at least by some approximation, the rate of innovation, productivity, and growth of an economy. A growing literature in macroeconomics discusses endogenous growth models that are micro-founded onto the notion that social interaction spurs knowledge diffusion and innovation (Comin et al., 2012; Lucas and Moll, 2014; Akcigit et al., 2018; Buera and Oberfield, 2020). There is therefore an economic interest in studying this particular kind of knowledge. Second, patent citations 'leave a trail', allowing to track flows of knowledge which are otherwise notoriously difficult to observe. Patent citations are thus the empirical proxy for learning used in this analysis.

How might social interaction affect the flow of ideas and technological knowledge, as captured by patent citations? Three distinct mechanisms come to mind, which we generally refer to as mobility, meetings, and exposure. Among these, we distinguish between heavy and light knowledge conveyors in the process of social learning. Heavy knowledge conveyors are associated with interaction of inventors with colleagues in professional networks, or with geographic mobility of inventors themselves. Light conveyors, by contrast, are related to interaction in informal networks, and refer to less structured channels such as chance meetings, referrals, perceptions, and salience of market opportunities. Table 3.1 gives an overview.

Table 3.1: Overview of possible mechanisms for the effects of social learning

	Mobility	Meetings	Exposure
Heavy	Endogenous inventors' location	Endogenous inventors' networks	N/A
Light	N/A	Chance meetings and referrals	Salience of market opportunities

The distinction between heavy and light conveyors is important because, in our empirical framework, we are mainly interested in isolating the effect of the latter, which we argue operates through informal networks. We interpret informal networks in a broad sense, as the social environment in which inventors work, net of their professional ties (see Section 3.3.1 below for further details). In what follows, we refer to this concept simply as 'social connectedness'. The broad measure of social connectedness we adopt, however, is potentially driving both heavy and light conveyors of knowledge. For instance, social networks are known to correlate with labour mobility (Buechel et al., 2019). In line with the findings of Breschi and Lissoni (2009), mobile inventors carry ideas with them as they move across firms and places.[4] Social connectedness might matter, then, to the extent that it favours inventor mobility and influences their choice of location. Similarly, it is also possible that social connectedness determines professional networks, typically defined empirically as networks of co-inventors. According to this perspective, technical collaboration networks can be endogenous to one's informal social network. In other words, the likelihood that inventor A starts a collaboration with inventor B increases in the number of (and decreases in the length of) paths connecting A to B. Powell and Grodal (2005) refer to such ties as 'emergent networks', that is, unintentional networks that develop on the grounds of ongoing relationships of a different nature (friendship ties, common ethnicity, co-location or reoccurring meetings). Whilst not directly focusing on informal networks, Crescenzi et al. (2016) and Crescenzi et al. (2017) do show that social proximity in co-invention networks influences the probability of forming a collaboration in the future.[5] Professional networks are of paramount importance in innovation, as patents embody technical, often discipline-specific, ideas that require prior knowledge to be absorbed (Cohen and Levinthal, 1990).

The channels outlined so far point to relatively well-specified ways in which social connectedness can affect citation probability through heavy knowledge conveyors. However, interpreted this way, any observed impact of informal networks would effectively be nothing more but a reduced-form empirical correlation of limited interest if one can readily observe inventor collaboration networks or inventor mobility. In fact, the correlation should disappear once controlling for these variables (a task we take up in our empirical model). Is there, at least conceptually, a residual role for social connectedness to influence the flow of technical knowledge through lighter channels? One possibility is that social connectedness

[4] Lissoni (2018) provides a recent discussion with respect to international mobility and migration.

[5] In his account of the development of the modern radio industry, Maclaurin (1950, p.98) offers an interesting historical insight into the life of inventor Guglielmo Marconi that aligns with the endogenous collaboration network hypothesis as well as with that of inventor mobility: "Marconi was also greatly aided by his family connections. His mother was of the Irish aristocracy and moved in the 'best circles' in England. The family concluded that Guglielmo would have a better chance to commercialize his inventions there than in Italy. A visit was arranged in 1896, and the young inventor (then twenty-two) was introduced to government officials and capitalists who might be interested in the radio field. Among these officials was William Preece, engineer-in-chief of the British Post Office. He took a keen interest in Marconi, and planned a demonstration for the post-office engineers. Marconi, who had been steadily improving the workmanship of every part of his equipment, showed that messages could be sent up to eight miles. This success and the interest displayed by Preece led to the formation of the British Marconi company in 1897."

increases the rate of chance meetings and referrals. Research is increasingly considering the importance of serendipitous encounters in directing inventive activity and knowledge exchange. Catalini (2018) studies the exogenous reallocation of university researchers due to building renovation. Atkin et al. (2019) rely on cell-phone data to uncover the effect of unplanned meetings between workers from different firms on the propensity of these firms to cite each other's patents. Roche (2019) shows that chance interactions promoted by better connecting local road networks foster serendipitous knowledge exchanges within neighbourhoods, which can explain differences in their innovative performance. With respect to access to external sources of knowledge, intuitively, the probability of chance meetings occurring between individuals from different places increases in the number of ties connecting these places. In practice, this could happen through visits to distant friends, or even digitally via interaction on social media and online communication platforms.

Another tentative light channel linking social connectedness to patent citations emerges from a survey of inventors carried out by Jaffe et al. (2002), which aimed to shed light onto the black-box process of idea exchange in technical and scientific fields. The authors find that, asked about what factors had a significant influence on the development of their inventions, almost 60% of respondents cited the 'awareness of a commercial opportunity' while another 20% mentioned 'word of mouth or personal interaction'. Notably, 'joint work with others' was only mentioned by less than 15% of respondents. Moreover, 'word of mouth' and 'viewed a presentation or demonstration' also accounted for more than 30% of responses when asking about how citing inventors learned about the previous patent, compared to only about 5% of inventors who cited 'direct communication with the inventor'.[6] Taken together, these qualitative findings suggest that there might be something related to *salience* of ideas and identification of market opportunities in the process of scientific learning (labelled herein as the 'exposure' channel). This channel is not necessarily technical in nature nor is its scope confined to professional connections. More concretely, exposure-induced learning could be driven by preferences on the demand side (determining market opportunities for ideas both for consumers or downstream firms) or through awareness of supply side technological opportunities via knowledge of different but related products, solutions or applications prevailing in the (possibly geographically distant but) socially connected market. This intuition is taken up in the work of Breschi and Lenzi (2016), who, although focusing on professional ties between inventors, emphas-

[6]These figures are particularly high considering that the survey could not distinguish between citations made by the applicant from those included by the patent examiner during the review process. The frequency of inventors who answered '[learned] during patent application process' and 'never before now' (about 60% in total) is consistent with the average incidence of examiner-added citations (about 60% of all patent citations, according to the data used in this paper).

ise the importance of allowing for social connections between different cities as a means to "[...] enriching and renewing a city's knowledge base by facilitating access to fresh external knowledge" (p. 66). More recently, Akcigit et al. (2018) also emphasise the pernicious effects of restricted access to external knowledge, which can limit innovation productivity. This is due to the 'proximity paradox', whereby the absence of inflow of new ideas from interaction across clusters results in too much specialisation, cognitive lock-in, and lower idea quality (Miguélez and Moreno, 2015). The conceptual argument made in this paper is similar. The emphasis however is on latent knowledge embedded in informal connections, or 'knowledge in the air' as originally conceptualised by Marshall (1890). According to the proposition of this paper, ideas are not only channelled through one specific network connection but rather permeate the broader informal social context in which innovation occurs. Accordingly, our empirical analysis will attempt to isolate the effect of social connectedness on knowledge flows via light conveyors, as opposed to its influence through heavy channels such as inventor networks and mobility.

3.2.3 Contribution to Related Literature

This work dialogues with three main strands of literature, partly reviewed in the previous sections. First, it speaks to research on urban economics and agglomeration economies by outlining micro-level channels by which learning might occur, and specifically how social connectedness might provide non-agglomerative mechanisms for transmission of knowledge unrestricted to physically proximate agents. It also emphasises how this type of knowledge tends be technologically more distant, which offers a new perspective on debates about specialised and diversified industrial clusters, opening to new research questions about complementarity and substitution between internal and external sources of knowledge.

Second, this analysis contributes to the innovation literature pioneered by Jaffe et al. (1993), which relies on patent citations to capture the geographic localisation of knowledge exchange and spillover of ideas. Many analyses so far have implicitly used physical proximity as a proxy for social interaction. This might work well for *local* social interaction. It is nonetheless important to highlight the conceptual distinction that proximity is not a knowledge sharing mechanism in and of itself, but rather an approximate measure of social interaction when interaction is unobserved. Crucially, the need to proxy interaction this way in empirical work imposes a specific spatial boundary to the conceptual definition of social learning, which may or may not occur locally in a spatial sense. In fact, imposing an a priori spatial boundary to social learning would seem excessively restrictive considering the tremendous progress in ICT and the fall in travel costs observed in the past few decades. This is acknowledge in the literature, but the use of data on social connectedness in this paper allows to study informal social interactions at an unprecedented spatial scale. Whilst this reasoning by no means attempts to deny the importance of proximity in the exchange of knowledge, it does argue that there is a merit in conceptually separating the geographical from the social space. This work thus examines the conditions under which social learning might occur independently of geographical constraints, particularly beyond the local level.

Finally, this paper contributes to the growing scholarship on the role of social networks in the innovation process, by explicitly looking at informal social connections defined as the broader social environment to which inventors are exposed in their daily work. The extant literature emphasises the importance of professional ties over informal ones in the innovation process. There is ample scholarship documenting the importance of such ties among scientists and inventors. Sorenson et al. (2006) study the role of social proximity in inventor collaboration networks and find that being socially close does indeed facilitate the transfer of moderately complex knowledge, even when controlling for geographic and organisational proximity. Agrawal et al. (2008) interact measures of spatial and social proximity to investigate knowledge transfers. They find that belonging to the same ethnic community is a substitute for geographic proximity among inventors. Co-invention networks are also at the core of the analysis by Breschi and Lissoni (2009), who emphasise the importance of geographically-constrained inventor mobility in the localisation of knowledge flows. This statement is consistent with Agrawal et al. (2006), who argue that flows of knowledge between the origin and destination of mobile inventors are 50% higher than they would be in the absence of such links. It also aligns with Belenzon and Schankerman (2012), who discuss how state-border effects on citation of university patents and scientific publications are stronger for universities that rely more heavily on workforce educated in-state, and located in states where interstate scientific labour mobility is low. Maggioni et al. (2007) also show that collaboration in research networks and co-patenting represent a relational channel prevailing over geographical distance in explaining spillovers. Similarly, D'Este et al. (2013) find that firms' embeddedness in networks of agents working in technological related fields allows them to overcome the constraints of physical space in determining university-industry collaboration patterns. More recently, Zacchia (2019) shows that interaction of scientists on inter-firm networks drives knowledge spillovers, as captured by greater productivity and innovation outcomes from changes in R&D spending in connected firms.

As the literature on professional inventor networks is already rich, we are less interested in documenting additional evidence for this channel. Rather, we focus on light conveyors associated with informal networks. We argue that this is important for several reasons. First, when it comes to localisation, it is already established in the literature that physical distance is an inadequate proxy for social interaction (a finding that the analysis in this paper partly corroborates). Second, limiting the analysis to professional networks draws an incomplete picture due to the likely discipline-biased nature of such ties, which tend to convey specialised knowledge. By contrast, it is possible that informal types of connections lead to an entirely different type of knowledge exchange, due precisely to their more diverse composition. This distinction reminds conceptually of Granovetter (1973, 1983)'s 'strength of weak ties' hypothesis. According to this notion, it is more distant relationships (acquaintances, and friends of friends) that convey the most novel and valuable type of knowledge. Related research discusses the importance of local social capital, especially of the 'bridging' type, in the innovation process (Crescenzi et al., 2013a,b). However, studying social capital variables or population density allows to approximate this informal type of interaction only locally, whereas there is little empirical evidence that explicitly investigates the effects of informal ties beyond the urban or regional level, especially when it comes to the exchange of technical and scientific knowledge. Breschi and Lissoni (2009) show that the relevance of social networks in patent citations (knowledge transfer) decays in the number of ties needed to connect two inventors in the network. In other words, it appears that social proximity in professional networks of scientists (co-invention networks) matters above and beyond geographic proximity, but is limited to close professional ties. The authors further interpret the finding on distance-decay in social networks as suggestive evidence against the "[...] conventional wisdom that assigns great importance to more informal, non-market related knowledge exchanges such as those originating from kinship, friendship and social gatherings" (p. 466). Their analysis, however, does not directly measure informal and non-professional ties. Hence, there is value in presenting more direct evidence on this particular claim, especially since research linking informal networks to innovation dynamics is scant (Powell and Grodal, 2005).

A notable exception is the work of Bailey et al. (2018b), who rely on the same data used in this paper to explore empirical correlations of social connectedness with a broad set of outcomes, including patent citations. Their analysis relies on the case-control matching strategy by Jaffe et al. (1993), finding that connectedness positively correlates with innovative activity and knowledge flows. This paper aims to complement their work in several ways. It focuses exclusively on the learning outcome, carefully conceptualising the underlying relationship notably with respect to informal ties and light conveyors. It also improves the estimation framework by attempting to identify the causal effect of connectedness on knowledge flows using examiner added citations as a control group, and by controlling for inventor mobility, their professional networks, and other confounding channels. Finally, we document several important dimensions of heterogeneity in the effects of social connectedness that are consistent with the conceptual discussion.

3.3 Data and Empirical Methods

3.3.1 Variable Definition and Measurement

This paper relies on two main sources of data. Social connectedness is measured using information on friendship links on Facebook, a popular social media platform. Knowledge flows are proxied using patent citation data. Additionally, the analysis also uses data from the 2010 US Decennial Census and the Internal Revenue Service (IRS). What follows gives details on how the key variables of interest are defined.

Informal Social Networks: the Social Connectedness Index

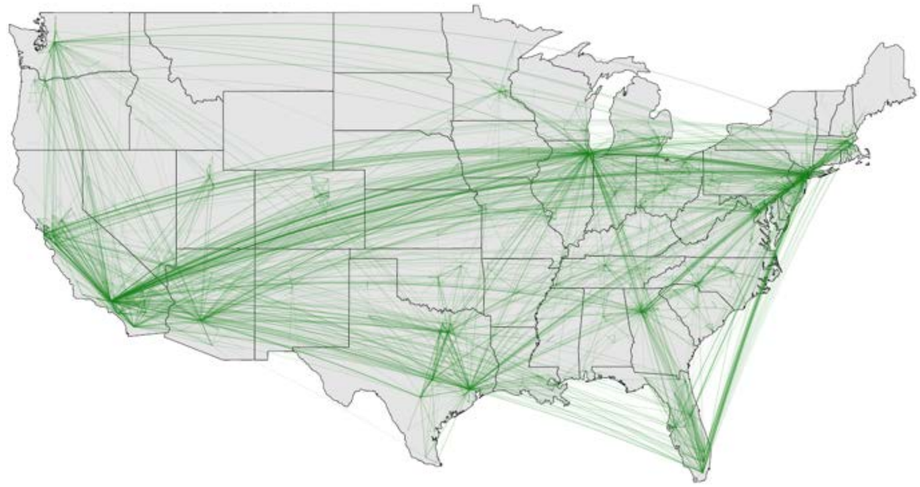
The proposed measure of informal social networks, or social connectedness, relies on an index developed by Bailey et al. (2018b): the Social Connectedness Index (SCI). This index essentially captures the social graph for the universe of *active* US Facebook users as of April 2016, aggregated up to the level of counties.[7] Users are deemed active if they interacted with Facebook in the 30 days prior to the April 2016 snapshot. Geographic location is assigned using the IP address from which users login most frequently. For all users i and j and for each pair of counties c and k, the index is constructed as:

$$SCI_{ck} = \sum_{i \neq j} \sum_{j} \mathbb{1}_{ij}, \text{ for } i \in c \text{ and } j \in k$$

Where $\mathbb{1}_{ij}$ is an indicator variable that takes the value of 1 if two users are friends with each other, and 0 otherwise. Due to confidentiality concerns, Facebook only releases a re-scaled version of these data. The index thus ranges between 0 and 1,000,000, the highest observed value, which is assigned to connections of Los Angeles County to itself (i.e., friendships within the county). The result is a weighted social graph consisting of 3,136 nodes and 9,462,485 edges. Figure 3.1 visualises the top one percent of edges in the data, assigning darker colours and thicker lines to stronger connections. The concentration of social ties between counties hosting the largest cities in the US is evident.

[7] In principle it would be more accurate to refer to Facebook *accounts* rather than *users*. However, the same expression as in Bailey et al. (2018b) is used here for consistency.

Figure 3.1: Top one percent of social connections across US counties



Unfortunately, Facebook does not release covariates for these data. However, it is possible to gauge some descriptive facts from secondary sources. At the time the data were extracted, there were over 220 million active monthly Facebook users in the United States and Canada.[8] A Pew Research Center study published in that same year estimates that about 70% of US adults (aged 18 or more) used the social media platform (Greenwood et al., 2016). Women, younger individuals (aged 50 or less), college educated and relatively poorer adults were slightly overrepresented, albeit by small margins. Most Facebook friendships are with people with whom users have ongoing interaction in real life. According to Hampton et al. (2011), ties between Facebook users tend to occur among high school or college peers (31%), immediate or extended family members (20%), co-workers (10%), and neighbours or acquaintances (9%). The remaining ties are with friends of friends, or 'dormant relationships', that may become useful to users in the future. However, only 3% of Facebook friendships are with someone the user has never met in person. Moreover, several studies have shown that Facebook ties are good predictors of real life friendships and friendship strength (Gilbert and Karahalios, 2009; Jones et al., 2013). All this suggest that there is strong potential in these data to be used to study social relationships on a large-scale (Bailey et al., 2018b).

Nevertheless, there are limitations in the use of the SCI to capture real-life ties. The geography of connectedness might be measured imprecisely to the extent that Facebook users do not represent the average American. Because friends are typically added on Facebook rather than deleted, it is also possible that this measure overestimates real-life

[8] Information obtained from Facebook's 2016 quarterly results report, retrieved at: `https://s21.q4cdn.com/399680738/files/doc_presentations/FB-Q4'16-Earnings-Slides.pdf`

interaction between people and places, a concern only partly mitigated by the fact that Facebook imposes a limit of 5,000 friendships on personal accounts. However, erroneous measurement is unlikely to bias estimates unless there are reasons to believe that this error is correlated with the outcome of interest.

Another important concern is that using the SCI involves a loss of precision in the measurement of relevant informal social networks, insofar as they are imputed to each inventor on the basis of their neighbourhood, rather than their actual social ties. We conceptualise informal networks as those broad-based relationships individuals entertain in their personal life beyond work. These include family and friends, but also extend to relationships beyond this inner circle of connections. There are two main ways to define such informal networks empirically. One way is to directly look at each agent's ties (social ties proper, or *interpersonal* networks), restricting these to non-professional relationships (professional ties in this application are inventors' co-patenting networks). Another way is to think of informal connections more generally as the social environment characterising the neighbourhood in which an individual lives or works (*neighbourhood* networks). We adopt the latter definition, which emphasises the value of weak ties (Granovetter, 1973, 1983). The composition of this broader social environment is the aggregate result of choices made by many individuals over many time periods, and therefore represents a potentially richer and more diverse source of knowledge and ideas than strong ties such as family and close friends. Ultimately, neighbourhood ties are simply interpersonal ties aggregated for all agents residing in a given spatial unit. This distinction however matters for at least two reasons. First, even though interpersonal and neighbourhood networks are likely to overlap (most people have friends that live geographically close), some agents in neighbourhood networks may never appear in interpersonal networks (even when considering high-degrees of separation), or enter at a social distance so high that interpersonal networks seem unlikely to matter more than the fact that the same contact can be established due to exposure to the same overall social environment. Second, neighbourhood networks can be considered to be time-invariant over a sufficiently large area and a sufficiently small period of time, due precisely to their aggregate and historically-determined nature. This mitigates endogeneity concerns in the definition of this variable. Moreover, by looking at the overall social environment in which inventors operate this measurement of informal social networks is faithful to Marshall (1890)'s original conceptualisation of spillovers as arising from knowledge 'as it were in the air'.

Importantly, the assumption that the SCI captures informal connections relies on the ability to separately account for interpersonal professional connections, which extant literature finds to significantly influence the exchange of technical knowledge. It is otherwise possible that that social connectedness simply picks up a very noisy estimate of professional ties among inventors. The empirical measurement of such connections is discussed jointly with the patent citation data below.

Knowledge Flows and Professional Networks: USPTO PatentsView

Contrary to the claim that ‘Knowledge flows [...] leave no paper trail’ (Krugman, 1991a, p. 53), Jaffe et al. (1993) argue that in fact they sometimes do, for instance in the form of patent citations. Following this intuition, this analysis relies on patent data released by the United States Patent and Trademark Office (USPTO) to measure knowledge transfers. In particular, the USPTO’s PatentsView platform offers access to large structured data on over 40 years of patents and patent citations from 1976 until today. From 2001 onwards, these data also include valuable information on who made the citation, the patent’s applicant or its examiner. As discussed in Section 3.3.2, this information is at the core of the proposed identification strategy. There are well known limitations to the use of patent citations as a measure of knowledge flows (Pavitt, 1985; Griliches, 1998; Bessen, 2008). Firstly, patenting is selective, meaning that not all ideas or inventions are observed. In order to be patented, an invention needs to be novel, non trivial and commercially viable. Very often, these criteria make it easier to patent inventions in manufacturing-related activities rather than services, and there is bias within manufacturing industries too. It entails that patents typically represent outcomes of applied, rather than basic research. There is also a strategic component to patenting. Obtaining and maintaining a patent is costly, so that it is likely that only the most valuable ideas are filed for intellectual protection. Similarly, some firms may prefer to maintain their invention entirely secret. Finally, patents necessarily represent a form of knowledge that is relatively structured and that can be codified. This means that the more tacit kinds of knowledge are not captured by this measure. Arguably, however, tacit knowledge is also the kind for which social ties, interpersonal communication and face-to-face contact matter the most.[9]

With these caveats in mind, what follows describes the construction of the estimating sample. We begin by taking the population of citations sent by patents issued in the 2002-2019 period. Each citing and cited patent is mapped onto US counties using the mode of the location of listed inventors residing in the US, breaking ties randomly.[10] We prefer the use of inventor location, rather than the assignee’s. Lychagin et al. (2016) show that

[9] Provided there is also sufficient absorptive capacity, especially relevant in the case of technical knowledge.
[10] In earlier results, not reported herein, the mapping was also carried out using the location of the first inventor for whom this information was available, with no substantial change in findings.

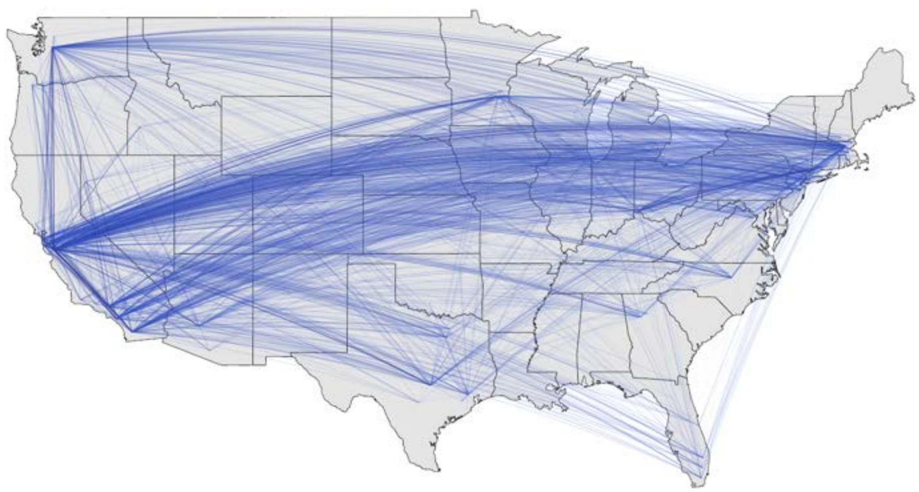
the geographic location of a firm's researchers better explains cross-firm spillovers than that firm's establishment location. For each citation, we retain its source, whether it was added by the applicant or by the examiner. We then merge in all available patent and county level information, such as issue and application years, technology fields, links over inventors' networks, bilateral geographical distances, social connectedness, and a set of controls based on the 2010 US Census.[11] Technological fields are based on the International Patent Classification (IPC), which provides a hierarchical system of codes.[12] We consider IPC classes (3-digit) and subclasses (4-digit), henceforth IPC3 and IPC4 classes respectively. Moreover, based on this classification, the World Intellectual Property Organisation (WIPO) provides a list of fields that have the advantage of being largely mutually exclusive, with adequate level of differentiation, and appropriate within-field homogeneity (Schmoch, 2008). While a single patent could be associated with multiple IPC classes or subclasses, in the vast majority of cases there is only one WIPO field.[13] A complete list is available in Appendix Table 3.B.1. The first listed IPC3 and IPC4 classes are also retained, for robustness checks in the empirical analysis. Some sampling restrictions are applied. First, only national flows of knowledge are considered. Citing and cited patents with no inventors residing in the US at the time the patent was issued are thus dropped. Moreover, citations originating from or received by patents located outside continental US states are also dropped. The sample is then restricted to citing patents whose elapsed time between application and issue date was below the 95$^{\text{th}}$ percentile in the distribution because of concerns of unobserved heterogeneity in the top 5% group. Similarly, we drop cited patents whose elapsed time to citation (their 'age' at the time of citing, using differences in application dates) was above the 95$^{\text{th}}$ percentile in the distribution. Finally, we restrict our attention to citations originating from patents issued after 2016, as this is the date when the social graph of Facebook was extracted.

The resulting estimating sample consists of 489,230 citing patents and 11,349,396 citations, of which on average about 60% were added by the applicant. Appendix Tables 3.B.2 and 3.B.4 offer descriptive details for citing and cited patents. A large number of patents had all citations made by the applicant (29%) or all citations made by the examiner (22%). This is in line with previous findings (Thompson, 2006; Alcácer and Gittelman, 2006; Alcácer et al., 2009). Still, one might worry that these extreme value patents could bias the analysis. Thus, robustness checks will show that results are unchanged even when

[11] I refer to Appendix Table 3.B.6 for a list of all variables.
[12] Detailed information on this system is available at this link: https://www.wipo.int/classifications/ipc/en/
[13] In the few exceptions, the field most frequently associated with the listed IPC classes is retained.

these patents are dropped from the sample. Another concern relates to the fact that inventors might cite other patents whose assignee is the same.[14] These citations do not strictly speaking represent knowledge spillovers since they occur within the boundaries of the same organisation, and are less interesting for the case of knowledge flows. Self-citations so defined represent about 10% of citations in the sample. They are not used in the analysis. Figure 3.2 maps the top one percent of knowledge flows in the data (aggregate bilateral citation counts for county pairs, irrespective of their direction), assigning darker colours and thicker lines to larger flows. There is a striking overlap between the flows represented in this map, and the social connections in Figure 3.1. An alternative way of visualising this relationship is proposed in Appendix Figures 3.A.1 and 3.A.2.

Figure 3.2: Top one percent of knowledge flows (citations) across US counties



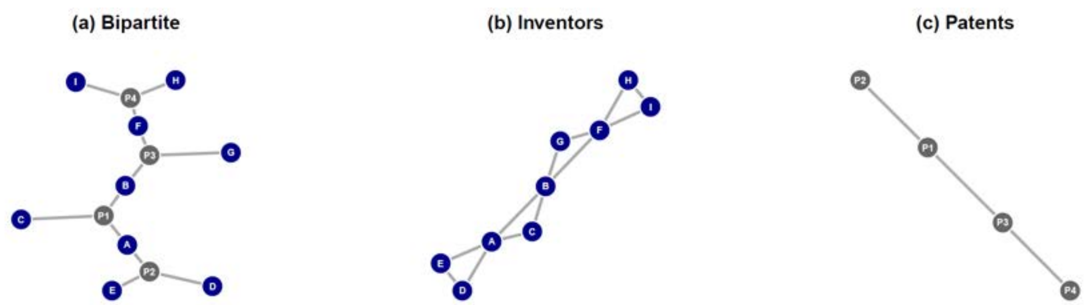
Finally, the professional network of inventors is measured in line with existing empirical literature as a co-inventor, or co-patenting, network. To obtain the network, this methodology crucially relies on inventor name disambiguation. Luckily, PatentsView data feature disambiguated inventor identifiers obtained through a discriminative hierarchical coreference algorithm proposed by Nicholas Monath and Andrew McCallum from University of Massachusetts Amherst.[15] We rely entirely on these data and do not attempt to disambiguate inventor names in alternative ways. Using the unique identifiers for *all* listed inventors (not just those located in the US), we construct a dummy for professional networks indicating whether citation patent pairs had a common inventor (self-citation), whether they shared a co-inventor (first-degree connection), and whether they shared the co-inventor of

[14] An assignee is the legal person to whom ownership of the patent is granted, typically a firm, a university, or another organisation. PatentsView provides assignee-disambiguation. For details, please refer to this webpage: `http://www.patentsview.org/community/methods-and-sources`.

[15] Details on the discriminative hierarchical coreference algorithm are available at this webpage: `http://www.patentsview.org/data/presentations/UMassInventorDisambiguation.pdf`.

a co-inventor (second-degree connection). Figure 3.3 illustrates this network. We begin with a bipartite representation of the data in panel (a), where each inventor (blue nodes) is linked to patents (grey nodes). This graph can be converted to a one-mode projection for inventors (panel b), showing co-patenting relationships. In this example, A and B are connected by a first degree tie due to the common authorship of patent P1. F is the co-inventor of B, a co-inventor of A, meaning he or she shares a second degree connection with A. Finally, projecting the graph in (a) by patents allows to track whether a citation falls within the inventors' network. As shown in panel (c), patent P1 is connected to P2 and P3 due to a common inventor ('degree zero'). Patent P1 is also connected to P4 via a co-inventor, F, while patent P2 is linked to P4 via a second degree connection due to F being the co-inventor of B who is co-inventor of A.

Figure 3.3: Illustration of a professional network based on co-patenting



Just under 20% of citations in the data are linked by a professional connection. Unfortunately, it was not computationally feasible to build higher order network links. Reassuringly, however, Breschi and Lissoni (2009) document that the effect of inventor networks on patent citations drops sharply in the degree of social distance.

3.3.2 Empirical Strategy

Researchers studying the geographic localisation of knowledge exchange face the challenge of controlling for the pre-existing geography of production, that is, the propensity of industries to cluster in space. Firms or workers might exchange knowledge locally within a given industry simply as a result of their co-location due to mechanisms other than learning. That is to say, inventors might disproportionately cite nearby knowledge not because of some spatial friction that limits their access to knowledge produced farther away, but simply because the most relevant knowledge tends to be created locally anyway (and the inventor is located in that cluster precisely for that reason). This would not be a problem if the concentration of relevant activities were entirely driven by spatial frictions, but the literature shows that there are other reasons for the emergence of industrial clusters. Indeed, agglomeration may also arise in the presence of economic externalities due to matching and sharing benefits, such as thicker labour markets or input-output relationships (Duranton and Puga, 2004). In this setting, the correlation between knowledge flows and proximity would be spurious. It is therefore important to empirically isolate learning as a distinct channel other than matching and sharing.

This empirical concern applies analogously to analyses that focus on the social space, rather than the physical one (i.e., social connections). Firstly, because homophily in social relationships typically entails that similar people like each other, thus making it rather likely that the social network measure also reflects the geographic concentration of industries (an example of what Manski, 2000, termed 'correlated effects'). For instance, software engineers are likely to be friends with each other, but also tend to work in the same industries, which cluster around Silicon Valley. At the same time, in Silicon Valley workers might share knowledge independently of these friendship links. Secondly, and more trivially, the clustering of industrial activity matters because in this particular analysis social connectedness is imputed to inventors on the basis of their geographic location. Another possible biasing factor relates to common unobserved environmental factors in the respective locations of each agent, which simultaneously affect their propensity to interact and the possibility of observing a flow of knowledge without the need that interaction is associated to knowledge exchange. This could be the case, for instance, if two large university colleges facilitate interaction between graduate students, thus creating social ties, whilst at the same time being host to important research centres that use knowledge produced by the other university. Yet this knowledge could be sourced autonomously in complete absence of social learning, despite the existence of social links between students. What follows describes the proposed empirical strategy to address these shortcomings.

Examiner Citations as Control Group

To identify the effect of social connectedness on knowledge flows using patent citation data, this paper relies on a strategy devised by Thompson (2006). This strategy exploits information available on patents from 2001 onwards about the source of each citation: whether it was the patent's applicant, or if the citation was included by the examiner during the review process.[16] Examiner-added citations are then used as a control group

[16]Examiners are specialised administrative officers whose job is to deliberate whether or not a patent can be granted. The patent examination process is described in detail in Cockburn et al. (2002).

for knowledge flows. In practice, a variable C_{ij} is coded to denote whether a citation of patent j by patent i can be interpreted as a flow of knowledge:

$$C_{ij} = \begin{cases} 1, & \text{if } j \text{ is cited by the applicant} \\ 0, & \text{if } j \text{ is cited by the examiner} \end{cases}$$

Levels of social connectedness between the counties $c(i)$ and $c(j)$ where patents i and j where created are compared for $C_{ij} = 1$ against those for $C_{ij} = 0$, controlling for other possible confounding factors, notably physical geography. This is achieved by means of a Linear Probability Model (LPM) that estimates how physical and social distances influence the likelihood that the citation of patent j by patent i is made by patent i's applicant, as opposed to its examiner.[17] Econometrically, this relationship can be represented as follows:

$$\begin{aligned} C_{ij} = \beta \ln SCI_{c(i)c(j)} + \delta \ln DIS_{c(i)c(j)} + \eta NET_{ij} + X'_{c(i)c(j)}\gamma \\ + \psi_{c(i)} + \psi_{c(j)} + \theta_i + \theta_j + \mu_{g(i)g(j)} + \pi_i + \pi_j + \epsilon_{z(i)z(j)} \end{aligned} \tag{3.1}$$

Where $\ln SCI_{c(i)c(j)}$ is the natural log of the Social Connectedness Index between counties $c(i)$ and $c(j)$, $\ln DIS_{c(i)c(j)}$ is the natural log of physical distance (great circle, in thousand kilometres), NET_{ij} is the professional networks dummy, and $X_{c(i)c(j)}$ is a vector of bilateral controls defined at county level. Note that controlling for professional networks allows to interpret β as the effect of informal social connections in the inventor's neighbourhood. Additionally, all specifications also include citing and cited counties fixed effects (FEs) $\psi_{c(i)}$ and $\psi_{c(j)}$, citing and cited patents cohort fixed effects θ_i and θ_j (using issue years[18]), and a technology-pair fixed effect $\mu_{g(i)g(j)}$ for both patents (using WIPO fields). We also explore the use of citing and cited patents fixed effects π_i and π_j. Finally, $\epsilon_{z(i)z(j)}$ is an error term double-clustered by citing and cited commuting zones $z(i)$ and $z(j)$,[19] allowing for two-way cluster-robust standard errors following the method of Cameron et al. (2011). This adjustment is required when one clustering dimension is not nested within the other.

[17] A linear probability model is preferred over the probit or logit options due to the use of high dimensional fixed effects, which would make probit and logit estimation computationally very demanding. Moreover, this allows for a more straightforward interpretation of coefficients as marginal effects.

[18] Application year cohort fixed effects were also tested in robustness checks, with no change in findings.

[19] Commuting zones (CZs) are groupings of counties developed by Tolbert and Sizer (1996) using hierarchical clustering methods as a spatial measure of labour markets that is not constrained by minimum population thresholds and that maximises within-group commuting ties. They have the advantage of representing boundaries based on economic geography rather than administrative criteria. There are more than 700 CZs in the US.

Main Sources of Bias

This paper is especially interested in correctly estimating β. However, the estimating equation will give biased estimates if $\mathbb{E}[\epsilon_{z(i)z(j)}| \ln SCI_{c(i)c(j)}] \neq 0$. With this in mind, there are several possible sources of bias. An important one, already cited, is the fact that industries operating in similar technological fields tend to agglomerate in space for reasons not limited to knowledge flows. Examiner citations, combined with technology-pair fixed effects, are precisely meant to control for this by providing a set of 'control' citations that is orthogonal to the physical and social geographies of the applicant.

Using examiner citations as controls requires two main assumptions (Thompson, 2006). Firstly, this method posits that citations made by the applicant are on average more likely to represent genuine knowledge transfers than citations by the examiner. Examiner citations are assumed to rather reflect an administrative act required to complete the scope of prior art available for that patent. In other words, this strategy requires that examiner citations can be credibly interpreted as counterfactuals for inventor citations: knowledge that the inventor ought to have had, but did not, and that this knowledge was not in turn added by the examiner as a result of knowledge flows. In partial support for this claim, a survey of inventors confirms that applicant citations do represent a measure of knowledge transfers - although noisy - and that when inventors were unaware of citations made in their patent, this was typically due to the citation being added by the examiner (Jaffe et al., 2000). Note that this method does not require that *all* applicant citations reflect knowledge flows. Indeed some citations may have been added by the patent attorney (Jaffe and de Rassenfosse, 2017). However, as long as applicant citations are systematically *more likely* to reflect a knowledge flow than examiner ones, incorrectly attributed citations are simply noise. Any signal in the data can still be uncovered using a larger sample, as we indeed do.

A second identifying assumption is that examiners do not learn via their social connections or geographic location. In other words, the geographic and social locations of examiners must be orthogonal to the predominant knowledge base of the patent being examined, so that examiners cannot learn about prior art from the same localised knowledge flows that are specific to the particular set of technologies of the examined patent. Importantly, this same requirement must also hold for the social space: the position of examiners in the network of social relationships must be exogenous to the predominant technological class of the citing patent so that exposure to the same social networks as the inventors cannot be the reason why examiners cite the patent. These conditions address the well known observation that firms and workers in specialised industries co-locate (sorting), and that people with similar characteristics are more likely to interact socially (homophily). Both these conditions are likely to be met in our data. Cockburn et al. (2002) and Thompson (2006) point out that most examiners work from one office located in Alexandria, VA.[20] Moreover, within a given subject area, patents are assigned to examiners in the order by which applications are filed to the office, which introduces an extra dimension of randomness in case one worries about the place of origin of the examiner before relocating to Virginia.[21] As regards the social space, exogeneity in the physical location of examiners allows to draw the same conclusion for connectedness, to the extent that the latter is defined for geographical units, and that it reflects relationships between the full population of Facebook users, and not just inventors. This is another advantage of using the SCI. Figure 3.4 gives additional credit to this assumption. The kernel density plots show the distributions of geographical distance (a) and social connectedness (b) for applicant (in blue) and examiner citations (control, in red), along with a distribution for control citations whose origin was replaced with that of Alexandria, VA, where examiners are actually located (in green). Comparing these fictional distributions to those of applicant citations and observed examiner citations, it is evident that examiners tend to draw citations from the social network (and geographic location) of applicants rather than their own,[22] confirming the orthogonality requirement discussed above.

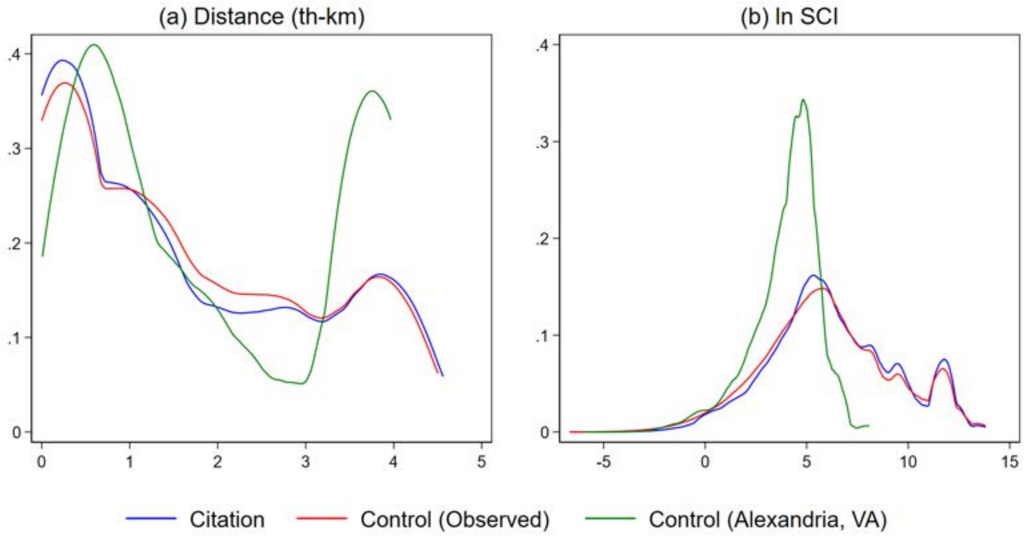
In addition, as discussed in Thompson (2006), estimates are potentially biased if examiner citations that do represent knowledge transfers are systematically more likely to cite patents farther away geographically or socially than knowledge flows measured by applicant citations are. This could occur if applicants strategically decide not to actively search for relevant patents that are produced by more remote counties (again, both geographically or socially), which would in turn be cited by examiners. These patents, however, if they do represent knowledge flows for examiners, would presumably take longer time to reach examiners through word of mouth due to their remoteness (unless examiners are systematically closer socially and geographically to these patents), so that one should observe that examiners cite on average older patents. Figure 3.5, however, shows that examiners tend to cite patents that have existed for a shorter period of time. This observation is

[20] A list of Patent Technology Centers, which provide examinations for patents in nine broadly defined technological fields, is available at: `https://www.uspto.gov/patent/contact-patents/patent-technology-centers-management`. The area codes associated to the telephone numbers of each Supervisory Patent Examiner (571) are indeed those of Alexandria, VA.

[21] Recent literature has documented the tendency of examiners to specialise by technological areas, see Righi and Simcoe (2019).

[22] This is likely due to the existence of communities of practice across different industries and technological fields (a similar argument holds for physical geography, due to industrial clustering).

Figure 3.4: PDFs of distance and ln SCI for citation and control flows



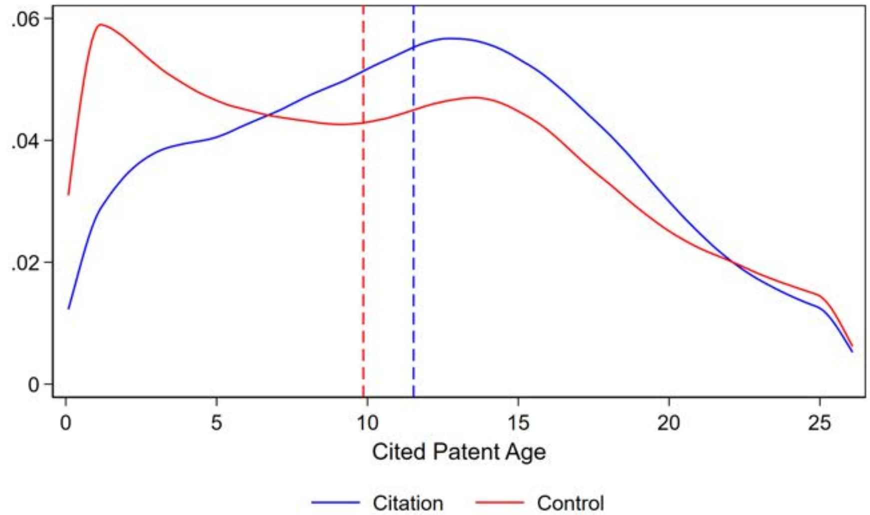
Kernel density plots, epanechnikov kernel with 0.3 bandwidth. Observed control citations are assigned to the county of the applicant. The Alexandria, VA control assigns citations to the examiners' county.

also consistent with the fact that examiners are not affected by social or geographic proximity when citing (or not as much affected as applicants). Conversely, it takes time for the same knowledge to reach applicants over these barriers, who as a result tend to cite older patents. It is of course possible that applicants omit younger patents on purpose, because these are more likely to be in competition with theirs to establish novelty. In principle, however, this should not happen. Rule 56 (37 CFR 1.56) of the *Manual of Patent Examining Procedure* imposes a 'Duty of Disclosure, Candor, and Good Faith', which requires applicants to inform the USPTO of any relevant prior art they are aware of that is material to patentability.[23] Moreover, any patent specific propensity to cite strategically can be absorbed with citing patent fixed effects.

Finally, note that combining technology pair dummies with dummies for citing and cited patent cohorts and patent-level FEs to some extent mimics the case-control matching method first implemented by Jaffe et al. (1993) and often used in this literature (including by Bailey et al., 2018b). In fact, by combining estimates of the within technology-pairs, cohorts, citing and cited patents effects with the examiner-control method, we believe that this analysis imposes stricter constraints on the data.

[23]The original text is available at: `https://www.uspto.gov/web/offices/pac/mpep/s2001.html`.

Figure 3.5: PDFs of cited patent age for citation and control knowledge flows



Kernel density plots, epanechnikov kernel with 0.5 bandwidth. Patent age is defined as the elapsed time between citing and cited patents application dates, plus 18 months. Dashed vertical lines are mean values.

Additional Identification Concerns

Citing patent fixed effects can also be useful if examiner citations are more likely to be added in patents that have lower geographical or social matching rates. In other words, they control for different propensities of citations to be added by the examiner, which may be correlated with the outcome at patent level. They also capture unobserved examiner characteristics and whether the patent has an institutional assignee or not. Thanks to the county-level fixed effects, moreover, any source of bias deriving from different propensities of counties to generate, patent, or cite ideas, as well as their initial stock of patents, is absorbed. In fact, county fixed effects solve any issue related to observed or unobserved characteristic specific to each county. In addition, the set of bilateral controls for differences in observable characteristics of counties $c(i)$ and $c(j)$ mitigates issues related to omitted variables specific to each county pair. For example, both counties might host important universities that also engage intensely in research activities. Students from county c who enrol to the university in county k increase social connectedness between the county pair (and the same is also true for students of county k pursuing a degree in county c). At the same time, the research activity of the two universities might increase the likelihood of citations of patents generated in the two counties independently of their social connectedness. Not controlling for such effects would lead to spurious results. Accordingly, a dummy coding the presence of a large, leading, research intensive university in both counties is used to capture this source of bias. The data is obtained from the 2018 Times Higher Education

Ranking of US universities, of which the top 50 institutions are retained. Another concern is that the SCI is simply a result of past migration patterns between county pairs. A variable capturing the log of gross migration flows between all county pairs is therefore also included in the analysis. This variable is constructed using rolling five-year cumulative counts of yearly county-to-county migration flows. It is assigned to each patent using county and application year information. The data on mobility come from the Statistics of Income Division (SOI) of the US Internal Revenue Service (IRS). They provide one of the most detailed sources of information at this level, based on address changes in the records of all individual income tax forms filed between 1990 and today.[24] Additional bilateral controls include absolute differences in: the share of adult population with a bachelor degree or higher, the share of children born in 1980-1984 who become inventors in the 2001-2014 period (by CZ where they grew up)[25], population density, median income, unemployment rates, and shares of White, Black, Asian and Hispanic Americans in each county.[26]

3.4 Results and Discussion

3.4.1 Main Regression Models

We begin by showing in Table 3.1 the results of baseline models regressing the main variables of interest separately one from the other. The models are estimated using Equation 3.1, selectively restricting the coefficients β, δ, η and γ to zero. Each coefficient is estimated in its raw form and with key fixed effects. For ease of reading, the outcome is expressed in percentage points. Columns (1), (3) and (5) give raw correlations between citations and social connectedness, geographical distance, and inventors' professional networks, respectively. Columns (2), (4) and (6) restrict the sample to citations across assignees and counties, and introduce the main set of fixed effects used in this analysis: dummies for citing and cited patent counties, cohorts (issue years), and pairs of WIPO technology fields. Restricting the sample is important for two reasons. First, we are interested in studying the impact of social connectedness *across*, rather than within the same region. Second, and most importantly, this allows to implicitly control for inventor mobility, which was one of the heavy knowledge conveyors discussed in Section 3.2.2. When the sample excludes within-county and within-assignee citations, inventor self-citations (accounted for by the professional network dummy) necessarily denote instances where the inventor changed em-

[24]Data retrieved at: `https://www.irs.gov/uac/soi-tax-stats-migration-data`.

[25]This variable is obtained from Bell et al. (2018), please consult the original paper for further details. The original data can be downloaded from: `https://opportunityinsights.org/data/`.

[26]Unless otherwise specified, all variables are defined at county-level and are constructed using data from the 2010 US Decennial Census.

ployer (at least for that particular patent), favouring one located in a different county. Because the professional network dummy also controls for endogenous inventor networks (another heavy channel), the SCI coefficient in this specification should only capture light conveyors of knowledge such as chance meetings, referrals, or salience of market opportunities (refer to Table 3.1 for an overview of all mechanisms). Moreover, with respect to fixed effects, note that other than the previously mentioned omitted variable concerns, county dummies also allow to account for the fact that larger county pairs naturally display higher social connectedness.[27]

Table 3.1: Baseline Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln SCI	0.242	0.450					0.266
	$(0.116)^b$	$(0.0560)^a$					$(0.109)^b$
ln Distance			-0.504	-0.367			-0.00243
			$(0.194)^a$	$(0.0626)^a$			(0.0998)
Prof. network					3.070	3.303	2.963
					$(0.772)^a$	$(0.562)^a$	$(0.556)^a$
Counties FEs		•		•		•	•
Years FEs		•		•		•	•
WIPO pairs FEs		•		•		•	•
Other county		•		•		•	•
Other assignee		•		•		•	•
Adj. R^2	0.0006	0.1139	0.0005	0.1138	0.0017	0.1143	0.1144
R^2	0.0006	0.1145	0.0005	0.1144	0.0017	0.1150	0.1151
N	11,349,396	8,851,560	11,335,849	8,839,486	11,349,396	8,851,560	8,839,486

Two-way cluster-robust standard errors for citing and cited CZ pairs (Cameron et al., 2011). Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. The outcome variable is expressed in terms of percentage points.

In Column (2), the positive and significant coefficient of 0.45 suggests that a one percent change in social connectedness leads to a 0.0045 percentage points increase in the probability of citation. Equivalently, it means that doubling the SCI yields a 0.31 percentage point increase in citation likelihood ($\beta \times \ln 2$). This is more than ten times smaller than the 4.37 percentage points estimated by Bailey et al. (2018b) for the same change using Jaffe et al. (1993)'s case-control matching method. Interestingly, Column (4) shows that physical distance displays a very similar effect, although with opposite sign. A county twice as far to where knowledge is produced is a quarter of a percentage point less likely to cite that knowledge, compared to another located only half that distance away. Column (6) shows the effect of professional networks. Being the co-inventor of a patent, having co-invented with an inventor of that patent, or sharing a co-inventor with an inventor of that patent

[27] Their inclusion equals to controlling for the natural logarithm of the product of each county's population, which combined with the logarithm of the SCI mimics a measure of logged relative probability of friendship (Bailey et al., 2018b) giving the number of existing connections over the number of total possible connections between two regions.

increases the probability of citation by just over 3 percentage points. These effects are all statistically significant at the highest conventional level. By contrast, Column (7) shows that when estimating all parameters simultaneously and controlling for the same variables mentioned above, the coefficient on distance becomes insignificant. Social connectedness, about 60% of the original magnitude, is only significant at the 5% level. Although slightly reduced, unsurprisingly the coefficient on professional networks also remains statistically significant. This specification represents the basis on which all other main models in this paper are estimated.

Table 3.2: Main Regressions

	(1)	(2)	(3)	(4)	(5)
ln SCI	0.266	0.435	0.129	0.0498	0.444
	$(0.109)^b$	$(0.112)^a$	$(0.0270)^a$	$(0.0279)^c$	$(0.110)^a$
ln Distance	-0.00243	0.0688	0.0229	-0.0136	0.0472
	(0.0998)	(0.0898)	(0.0186)	(0.0137)	(0.104)
Prof. network	2.963	2.934	0.517	0.274	3.056
	$(0.556)^a$	$(0.555)^a$	$(0.145)^a$	$(0.0746)^a$	$(0.541)^a$
WIPO pairs FEs	•	•	•	•	•
Bilat. controls		•	•	•	•
Within citing			•	•	
Within cited				•	
Interaction samp.					•
Adj. R^2	0.1144	0.1145	0.6044	0.6620	0.1219
R^2	0.1151	0.1151	0.6216	0.7070	0.1226
N	8,839,486	8,835,705	8,761,974	8,332,097	8,833,640

Two-way cluster-robust standard errors for citing and cited CZ pairs (Cameron et al., 2011). Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. The outcome variable is expressed in terms of percentage points. All specifications use citing and cited year and county fixed effects. The sample excludes citations within same assignee or same county. Bilateral controls: gross migration, top 50 college, differences in education, inventors, density, income, ethnicity. Interaction controls: main effects for own CZ or state, other state, elapsed time, assignee age, IPC4 technological distance.

The main results of the present analysis are reported in Table 3.2. For reference, the first column in this table copies the estimates of Column (7), Table 3.1. Column (2) introduces a vector of bilateral controls for all county pairs. In particular, the log of gross migration flows across counties over the previous five years addresses the concern that social connectedness does nothing more than to proxy population mobility between regions. Moreover, a dummy coding the presence of a major college in both citing and cited counties addresses the concern of spurious correlation due to the co-presence of students and researchers, with the former affecting friendship links and the latter generating citations, without any actual relationship between the two. Other controls include differences in education attainment, inventor and population densities, incomes, and ethnicities. Reassuringly, the coefficient on the SCI remains significant, even increasing in magnitude. Details on the marginal effects of each bilateral control are available in Appendix Table 3.B.7. Columns (3) and (4) introduce fixed effects for citing and cited patents. While the marginal effect of social connectedness is much smaller, it remains significant in both cases. At the same time, restricting our estimates to within-citing or within-citing and cited patents effects might be excessive. The identifying variation effectively would only come from the list of cited patents within each citing patent, when the cited patent is also cited by other patents, net of all other fixed effects. Accordingly, our preferred specification is that in Column (2). This suggests that two counties at the 75$^{\text{th}}$ percentile of the SCI are 1.2 percentage points more likely to cite one another than a pair of counties at the 25$^{\text{th}}$ percentile (see the Appendix, Table 3.B.6, for summary statistics). Finally, Column (5) reports the same specification in (2) but includes main effects for several dimensions of heterogeneity that we intend to explore using this estimating sample: different spatial boundaries (same county, same state, other state), cited patent age in years, maximum age of citing patent assignees, and technological distance (deciles of distance across IPC4 classes). This specification is included here for reference as it represents the baseline for all models that include heterogeneous SCI effects. This ensures that the intercept is the same across specifications even as the coefficient on connectedness is broken down by different variables, allowing like-for-like comparison (see Section 3.4.3 for further details). Despite this change, all coefficients are comparable in magnitude to those in Column (2).

3.4.2 Robustness Checks

Before investigating heterogeneous effects, what follows explores the robustness of estimates in Column (3), Table 3.2, to changes in model specifications and in the sample. Table 3.3 summarises the findings. Column (1) simply copies the estimates of the preferred specification (3) in Table 3.2, for reference. Column (2) shows that the estimates are robust to including fixed effects for application year cohorts, rather than issue year, for both citing and cited patents. Column (3) addresses the possibility of omitted variable bias due to assignment of location as the most frequently observed one among all inventors on the patent. Bias could arise if the other locations of co-inventors are also likely to be the most socially connected ones to the modal county of the patent. Knowledge flows from these counties would then be erroneously attributed to connectedness, while in reality they can be explained by (unobserved) co-location of one of the inventors. To address this, we restrict the estimating sample to citations made and received by patents with a single inventor. In such instances, location is necessarily assigned correctly with our method and there is no omitted variable bias of this kind. Doing so significantly reduces the size of the estimating sample, which falls to roughly 700,000 citations. Despite this very restrictive test, the coefficients on social connectedness and professional networks remain statistically significant. In fact, both increase somewhat in magnitude (especially the latter), suggesting that inventors who patent alone might disproportionately rely on informal and professional ties for access to knowledge (or perhaps this is simply due to more accurate measurement of location, further research may wish to explore this claim more in detail).

Table 3.3: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln SCI	0.435	0.351	0.505	0.305	0.288	0.400	0.385
	(0.112)^a	(0.126)^a	(0.216)^b	(0.0899)^a	(0.0878)^a	(0.165)^b	(0.0965)^a
ln Distance	0.0688	0.00651	-0.138	0.0411	0.0408	-0.137	0.0386
	(0.0898)	(0.0956)	(0.194)	(0.0717)	(0.0739)	(0.210)	(0.0935)
Prof. network	2.934	2.879	6.979	2.794	2.320	2.856	2.446
	(0.555)^a	(0.530)^a	(0.960)^a	(0.529)^a	(0.441)^a	(0.840)^a	(0.515)^a
Tech. pairs FEs	WIPO	WIPO	WIPO	IPC3	IPC4	WIPO	WIPO
Bilat. controls	•	•	•	•	•	•	•
Appl. year FEs		•					
Single-authored			•				
Non coastal						•	
Trimmed							•
Adj. R^2	0.1145	0.1133	0.2038	0.1241	0.1555	0.1315	0.0935
R^2	0.1151	0.1140	0.2103	0.1256	0.1623	0.1326	0.0944
N	8,835,705	8,835,704	717,586	9,091,697	9,066,472	5,047,385	5,815,090

Two-way cluster-robust standard errors for citing and cited CZ pairs (Cameron et al., 2011). Significance levels: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$. The outcome variable is expressed in terms of percentage points. All specifications use citing and cited year and county fixed effects. The sample excludes citations within same assignee or same county. Bilateral controls: gross migration, top 50 college, differences in education, inventors, density, income, ethnicity. The single-authored sample drops citations sent or received by patents with multiple authors. The non coastal sample drops citations originating or received in Census divisions bordering the Atlantic and Pacific coasts. The trimmed sample drops patents with citations added exclusively by the applicant or the examiner.

Columns (4) and (5) replace fixed effects for WIPO technology field pairs with fixed effects at IPC class (3-digit) and subclass (4-digit) levels. This entails moving from a set of just under 1,200 possible combinations to over 300,000 and 13 millions respectively, since there are more than 550 IPC classes and 3,700 subclasses. Despite this demanding change, the coefficient on connectedness is only sightly reduced and remains significant at the 99% level. Column (6) restricts the sample to non coastal areas only, dropping all citations originating from or destined to Census divisions not bordering the Atlantic and Pacific coasts. It addresses the concern that population and economic activity naturally cluster along the coasts, and so does innovation activity. As a result, more interaction is to be expected between coastal areas, as well as greater exchange of knowledge (more coast-to-coast citations), without the two being necessarily causally related to each other (essentially, an omitted variable bias due to an unobserved 'coast effect'). The size of the estimating sample is significantly reduced, but results are not affected by this restriction either. We infer that our findings are not limited to coastal locations but apply throughout the US territory. Finally, Column (7) trims the data by excluding patents whose citations were added exclusively by the applicant or by the examiner. As discussed, these represent a large group in our sample, and there is a concern that results are mainly driven by these patents. Reassuringly, this trimming does not alter findings.

Appendix Table 3.B.8 repeats the same exercise but includes fixed effects for citing patents across all models, despite concerns that this specification might be too restrictive. Once again, there does not seem to be any sizeable change in the coefficients compared to the original estimates, with the exception of single-authored patents, where the sample is likely too small to allow precise estimate of within-citing effects (indeed, the coefficient magnitude is stable, but standard errors are inflated).

3.4.3 Heterogeneous Effects

This section explores possible dimensions of heterogeneity in the marginal effects of social connectedness. In line with the literature and with the conceptual framework outlined in Section 3.2, we investigate three main drivers, described separately below. To empirically test for heterogeneous effects, we estimate models that take the following general form:

$$\begin{aligned} C_{ij} = \sum_h \beta_h \ln SCI_{c(i)c(j)} \times INT_h + \sum_h \delta_h \ln DIS_{c(i)c(j)} \times INT_h + \eta NET_{ij} \\ + X'_{c(i)c(j)}\gamma + \xi_{ij} + \psi_{c(i)} + \psi_{c(j)} + \theta_i + \theta_j + \mu_{g(i)g(j)} + \epsilon_{z(i)z(j)} \end{aligned} \tag{3.2}$$

Where all variables are defined as in Equation (3.1), with FEs for citing and cited counties, issue year cohorts, and WIPO technology pairs. In addition, the interaction term INT_h takes different values depending on the heterogeneous margin of interest:

$$INT_h = \begin{cases} GEO_{c(i)c(j)} & \text{for } h = 1 \\ AGE_{ij} & \text{for } h = 2 \\ ENT_i & \text{for } h = 3 \\ TDS_{g(i)g(j)} & \text{for } h = 4 \end{cases}$$

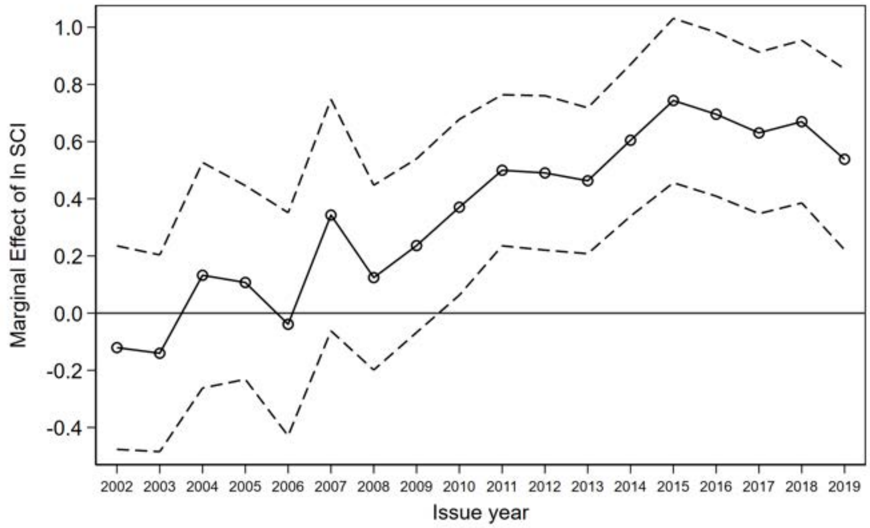
The vectors of coefficients of interest β_h and δ_h are selectively restricted to zero depending on the interaction that is being explored, setting h to 1, 2, 3 or 4. We consider heterogeneity over discrete geographical boundaries $GEO_{c(i)c(j)}$, cited patent age AGE_{ij}, citing assignee age ENT_i (elapsed time since first patent), and quintiles of technological distance $TDS_{g(i)g(j)}$. Note that all interaction terms are treated as categorical variables, so that β_h and δ_h are allowed to vary across all values of the interacted term. At the same time, we always include main effects for all interacted terms, captured by ξ_{ij}, to ensure the model has the same intercept no matter which dimension of heterogeneity we are investigating. The sample restrictions discussed in Section 3.4.1 are always applied: we drop within county and assignee citations. In the absence of any interaction term, thus, the baseline model reported in Table 3.2, Column (5), is estimated. An additional driver of heterogeneity we examine is the issue year cohort of the citing patent. In this particular case, however, we construct a new estimation sample dating back to patents issued in 2002, to consider a longer time-span. As the overall estimate of β is not directly comparable to that discussed so far anyway, the main effects term ξ_{ij} for all interaction variables is omitted (citing patent year cohort dummies are absorbed anyway). We also do not consider heterogeneity in the geographical distance coefficient δ here. The model specification is otherwise the same as in (3.2). We begin by discussing this last case of heterogeneity.

Time Trends

Sonn and Storper (2008) show that geographical proximity has become more important for knowledge production over time, despite advances in information and communication technologies. Using the case-control matching method by Jaffe et al. (1993), the authors reveal a greater likelihood of observing US citations to the same state or city in 1997 compared to 1975. The propensity to rely on local knowledge increases almost monotonically over this period. The underlying causes for this trend, the authors argue, have to do with greater reliance on tacit and non-codified knowledge on the technological frontier, faster product lifecycles requiring more rapid innovation rates, and more complex organisational strategies in knowledge production. More recently, Bloom et al. (2020) document a progressive decline in the productivity of research, defined as the ratio of total factor productivity (TFP) growth and the effective number of researchers. The authors thus conclude that "ideas are getting harder to find". Their result aligns with previous evidence by Jones (2009) on the changing nature of innovation, which he argues is becoming increasingly difficult and requires greater collaborative efforts. In keeping with these findings, we formulate the hypothesis that social connectedness may have also become more relevant over time, as a means to compensate for the increasingly demanding task of accessing relevant knowledge.

We test whether the effect of connectedness changes over time by allowing β to vary depending on the issue year cohort of the citing patent. To this end, we introduce a new

Figure 3.1: Marginal effects by citing patent issue year



The graph displays coefficients obtained from a regression where ln SCI is interacted with the patent's issue year, controlling for issue year main effects, ln Distance, inventor networks, and differences in county level observables. Dashed lines are 95% CIs.

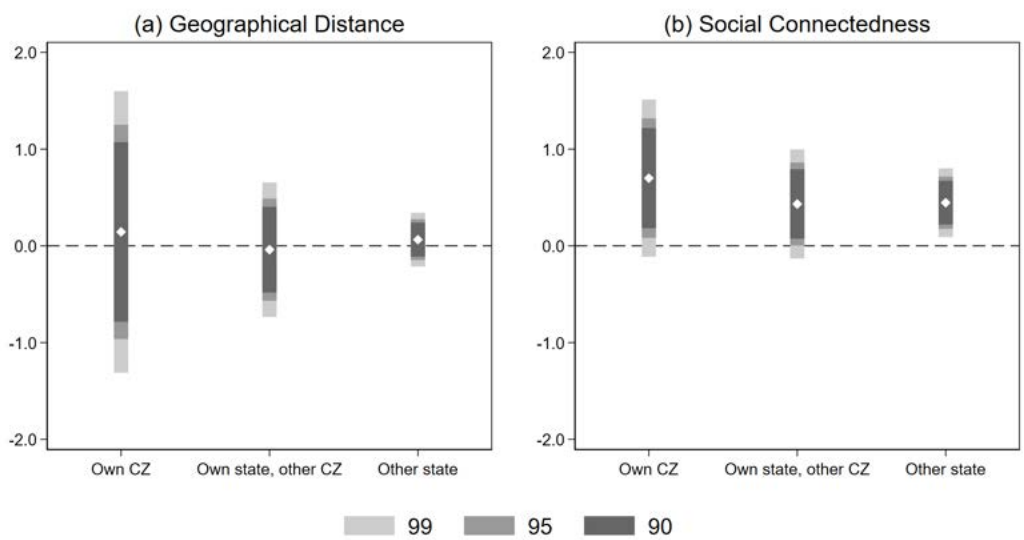
sample that includes all patents issued from 2002 onwards. Information on the source of citation, crucial for the identification strategy, was unavailable before this period. The sample construction follows the same method described in Section 3.3.1, with the exception that the size of the resulting list of citations is too large to work with, so a stratified random sample of 20% is drawn. Randomisation is performed at the level of citing patents to ensure that the drawn sample does not over-represent patents with many citations. The resulting estimating sample consists of 364,372 citing patents and 7,212,370 citations, of which about 60% on average are made by applicants. Appendix Tables 3.B.3 and 3.B.5 offer descriptive statistics for citing and cited patents.

Our results are summarised by the coefficient plot in Figure 3.1, which reports marginal effects of social connectedness over time. All coefficients were obtained from the same regression that interacts the log of the connectedness index with citing patent issue year dummies, controlling for issue year main effects, geographical distance, inventor networks, and differences in county-level observables. The results support our hypothesis. Not only are point estimates significantly higher in recent years compared to the early 2000s, marginal effects are not statistically different from zero at the 95% level before 2010. As a robustness check, an alternative regression is run where the application year of the citing patent is used, rather than the issue year. Results, reported in Appendix Figure 3.A.3, are unaffected. This finding also provides an additional reason for the decision to restrict this analysis to knowledge flows occurring in the 2016-2019 period. It should be noted, however, that the increasing magnitude of the effects could potentially also be related to the measure of social connectedness becoming more accurate over time, as it reflects a snapshot taken in 2016.

Spatial Boundaries

In this instance, we explore whether social connectedness becomes more important at greater distances. This would be consistent with the notion that connectedness allows to substitute for informal interaction that would otherwise occur locally due to geographic proximity (of two different counties, as we consider cross-county flows only - we do not test for substitution of co-location in the same county). For the same reason, we are also interested in comparing these results with what would happen if we only used physical distance as a proxy for this kind of interaction. As argued in Section 3.2.1, distance is likely to be inadequate in capturing this effect. To validate this, we would expect the coefficient on physical distance to be insignificant across discrete spatial boundaries capturing progressively larger areas.

Figure 3.2: Marginal effects by spatial boundaries



Each graph displays coefficients obtained from the same regression, where ln Distance and ln SCI are interacted with dummies for spatial boundaries. Main effects for spatial boundaries are also included. Dashed lines are 95% CIs.

Results reported graphically in Figure 3.2 only partly confirm our conjecture. The plot displays coefficients on physical distance (a) and connectedness (b) broken down by three discrete spatial boundaries: citations within CZ, within own state (but not own CZ), and across states. They are all obtained from the same regression, as in Equation (3.2). When controlling for social connectedness, distance does not appear to play any role in explaining knowledge flows no matter what boundaries are considered. In fact, its effects are estimated quite precisely at zero for citations across state boundaries. By contrast, point estimates on social connectedness are significant across all spatial boundaries. However, there is no evidence that the importance of connectedness increases as one considers progressively more (physically) distant interactions, although the estimates become increasingly precise. This confirms once more that social connectedness is a powerful predictor of knowledge flows, helping explain effects otherwise attributed to geographical distance. However, there is no evidence supporting the hypothesis that social and geographical proximity are strictly speaking substitutes. This contrasts with the findings by Agrawal et al. (2008), who study the interaction effect between geographical distance and co-ethnicity of inventors on citation likelihood.

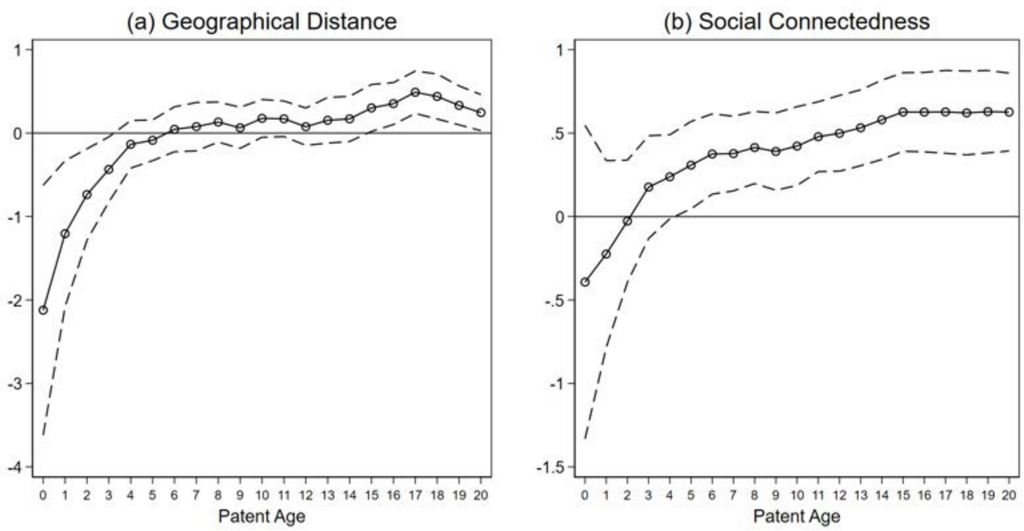
Patent Age

This section considers the role of elapsed time to citation in mediating the effect of social connectedness and geographical distance. Elapsed time to citation can be though of as the 'age' of patent j when it was cited by i at time t, measured as:

$$AGE_{ij} = t_i^{app} - t_j^{app} + 18\ months$$

Where t_i^{app} is the application date of citing patent i, and t_j^{app} is the application date of cited patent j. Since November 29, 2000, all applications received by the USPTO are published 18 months after being filed irrespective of whether or not they are granted. We thus consider this to be the relevant 'birth date' for cited patents. Patent age, initially expressed in months, is then discretised into years using a floor function that assigns the greatest integer less than or equal to the value in months divided by twelve. We conjecture that the impact of social connectedness might change over the interval defined by AGE_{ij}. The pattern of heterogeneity, however, is uncertain a priori. It is possible that social and geographic proximity matter most for the citation of young patents, when frictions in knowledge flows are highest. For geographic proximity, this effect is documented in Jaffe et al. (1993), where it is shown that localisation of citations decreases as the cited patent becomes 'older'. In the case of social connectedness, analogously, stronger informal ties might be especially relevant for the exchange of knowledge that is yet to become common domain. By contrast, it is also possible that once a patent does become common knowledge, its citation depends increasingly on the presence of some linkage, whether of geographical or social nature, which nudges inventors to tap into that pool of ideas as opposed to another. Older patents, for instance, might have been 'forgotten'. Making predictions about the direction of heterogeneity is further complicated by the fact that geographic proximity and social connectedness are not independent from each other, so that at different points in time the effect of one might influence that of the other. Ultimately, thus, this is an empirical question.

Figure 3.3: Marginal effects by cited patent age



Each graph displays coefficients obtained from the same regression, where ln Distance and ln SCI are interacted with cited patent age (app. - app. + 18 months). Main effects for patent age are also included. Dashed lines are 95% CIs.

Figure 3.3 graphically reports the marginal effects on geographical distance (a) and the SCI (b), allowing the coefficients to vary across the age of cited patents.[28] Vertical bars denote 95% confidence intervals. Controlling for the effect of social connectedness, geographic proximity matters most for the citation of young patents, confirming previous results by Jaffe et al. (1993). Greater distance between two counties decreases the probability of patents produced in one to cite those produced in the other during their first five years of circulation. The friction imposed by geographical distance is strongest for very young patents, then falls sharply and becomes largely insignificant. By contrast, controlling for the effect of geographical distance, social connectedness displays the opposite pattern. The marginal effect of the SCI is insignificant for cited patents aged five years or younger, but increases almost monotonically after that. The synchrony in the fading effect of physical geography as that of social connections becomes relevant is striking. It suggests that there is some degree of interaction between the two effects over time. It is difficult to interpret the graph unambiguously, however. It appears that as patents become common domain in a spatial sense, their likelihood of being cited depends increasingly on social connections.

[28]An unreported coefficient controls for the effect on all patents older than 20 years.

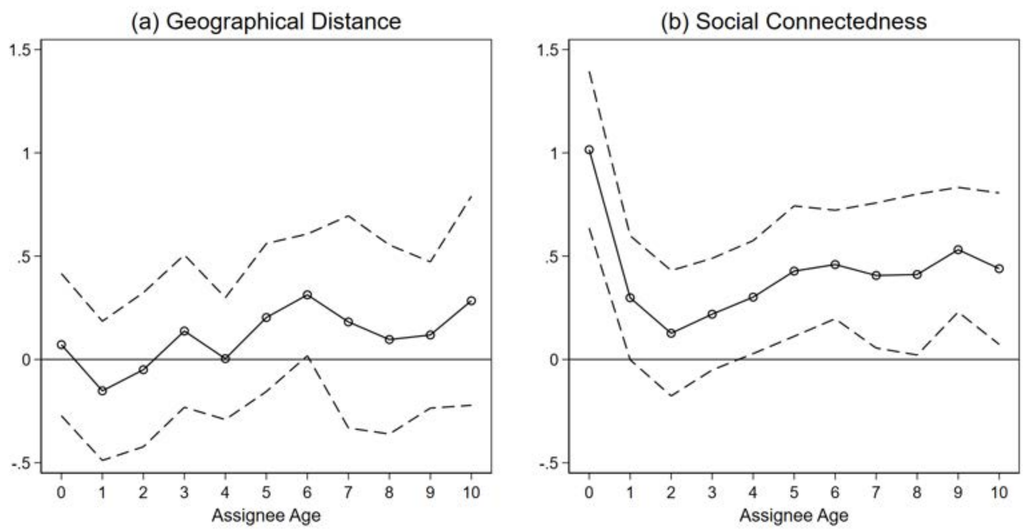
This could reflect a degree of bias in the sources of available knowledge inventors tap into, whereby they disproportionately rely on knowledge produced in places with stronger informal ties to their location. It could also show that social connectedness mitigates a propensity for older patents to become forgotten (without necessarily being obsolete).

Entrepreneurship and Garage Inventors

Does social connectedness matter differentially for organisations at different stages in their life? In particular, are entrepreneurs and garage inventors disproportionately reliant on their informal social environment as a source of ideas and innovation? In many organisations, inventors 'work for hire' with little flexibility in terms of process, and relatively strict guidance with respect to expected outputs. This is likely to be the case especially for more junior inventors in established teams, and generally in larger firms. For instance, Agrawal et al. (2010) find that inventors employed by large firms in company towns (places where innovation is concentrated in a single organisation) are more likely to draw on knowledge produced within the firm's institutional boundaries. The type of 'light' contributions channelled by social connectedness, such as salience of market opportunities, experimental ideas, or chance meetings, are perhaps of secondary relevance for this group. By contrast, they should matter most for more independent types of organisations, such as smaller and younger firms, entrepreneurs, and garage inventors (that is, inventors who work independently, on their own, often at the early stages of a new idea). Duranton and Puga (2001) introduce the concept of 'nursery cities' to highlight the role that access to diversified knowledge observed in large urban agglomerations plays in fostering innovation and entrepreneurship. Analogously, we test the hypothesis that social connectedness provides a similar source of advantage in the early stages of a firm's economic life. Consistent with this hypothesis, Percoco (2012a) shows that local social capital is positively associated with entrepreneurship in Italian cities, not least because of a possible effect on information transfers.

A variable ENT_i is created, which tracks the maximum age of all assignees listed for citing patent i. Assignee age is defined by exploiting disambiguated identifiers on organisations owning each patent. Organisations are assumed to have been established at the time they were issued their first patent. Subsequently, for any patent i, assignee age is the elapsed time between the issue year of the citing patent, and the issue year of the first observed patent for that same assignee. By construction, therefore, year zero is when none of the assignees owning the invention had previously patented. We think of them as entrepreneurs, or garage inventors. Equation (3.2) is then estimated for $h = 3$, allowing the coefficients on geographical distance and social connectedness to vary over assignee age. Results are reported in Figure 3.4.

Figure 3.4: Marginal effects by maximum age of citing assignee(s)



Each graph displays coefficients obtained from the same regression, where ln Distance and ln SCI are interacted with the maximum assignee age since first patent. Main effects for assignee age are also included. Dashed lines are 95% CIs.

While the marginal effect of geographical distance in (a) is mostly indistinguishable from zero across all values of assignee age, the effect of social connectedness in (b) is at least twice as large for garage inventors and start-up firms (year zero), than it is for older organisations. This difference is statistically significant compared to coefficient values for firms that are up to three years older. During this period, in fact, social connectedness does not matter for citation probability. From year four onwards, then, stronger informal ties matter again, although with reduced strength compared to garage inventors. This pattern is consistent with demographic studies of firms. Bartelsman et al. (2005) find that in the US firms enjoy a honeymoon phase in their first year of life, with the probability of exiting the market increasing significantly in the second year before settling at a constant rate. By year three, about 30% of newly established firms will have exited the market. Interpreting our results through this lens would suggest that while social connectedness is strongest for start-up firms, it is also lowest among firms that are more likely to fail. Perhaps, then, firms that survive this high risk phase and are still observed patenting as they age were somewhat advantaged by their greater social connectivity. This interpretation, however, is largely speculative and cannot be tested within the scope of this analysis. It is also possible, in fact even likely, that the proposed garage inventor measure correlates with the size of the patenting firm, with smaller firms (indeed potentially also young firms) disproportionately relying on external sources of knowledge.

Technological Distance

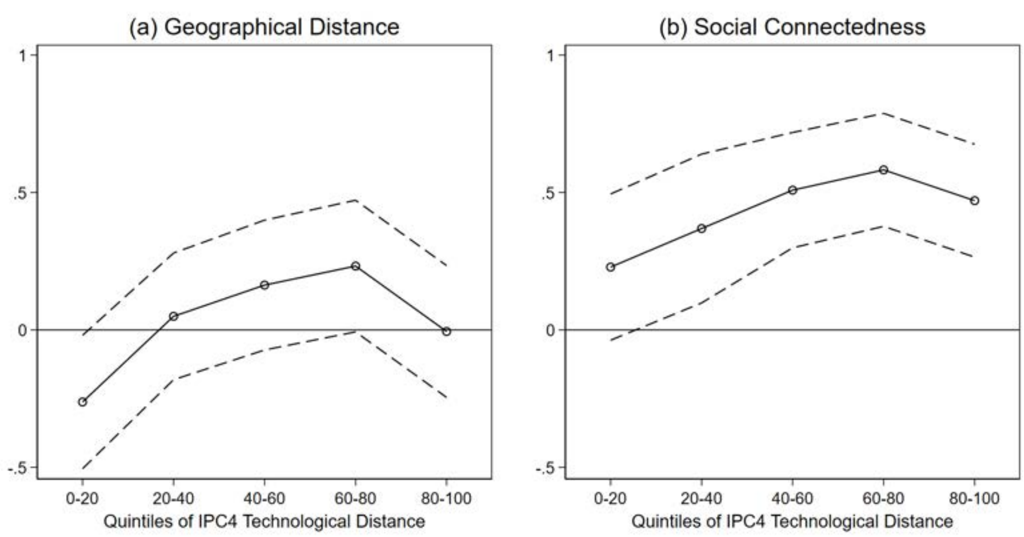
In this section, we explore the possibility that social connectedness matters differentially for the flow of ideas depending on the type of knowledge that is exchanged. It is well known that higher density leads to more innovation (Carlino et al., 2007). However, this relationship is non monotonic, since patenting rates are highest at medium levels of population density (Carlino et al., 2007; Henderson, 2007). Building on this finding, Berkes and Gaetani (2020) propose a model where informal interaction spurred by high density living sustains knowledge exchange across *distant* technologies. In other words, while overall innovation occurs in medium-sized specialised clusters, it would appear that 'unconventional innovation', as the authors call it, builds on informal interactions made possible by very dense urban agglomerations. Following this intuition, we investigate whether informal interaction fostered by stronger connectedness, rather than spatial proximity, can play a similar role in bridging gaps between different communities of inventors across the US. According to this hypothesis, social proximity would allow the diversity of knowledge bases typical of large urban agglomerations to exist beyond the constraints of geography. Feldman and Audretsch (1999) show that greater diversity in the industrial composition of a region is associated with higher rates of local innovation. One can think of informal social connectedness as a way to tap into a broader pool of knowledge. This hypothesis is consistent with research suggesting that a city with strong connections to other clusters benefits from the renewal and enrichment of its knowledge base by gaining access to new external ideas (Bathelt et al., 2004; Breschi and Lenzi, 2016; Akcigit et al., 2018), conditional on having sufficient absorptive capacity to do so (Miguélez and Moreno, 2015).

Technological distance is measured as the cosine dissimilarity in the reference set of each pair of technologies (Yan and Luo, 2017), using IPC technology classes (IPC3) or subclasses (IPC4), and the complete list of citations made by patents issued over the 2002-2019 period. Because each citing and cited patent can belong to multiple classes or subclasses, we consider a weighted average measure. We proceed as follows (we discuss classes only, the method is the same for subclasses). First, we inflate the citation list by assigning to each citing and cited patent all the classes they are associated with. We then assign a citation of each patent pair proportionally to the number of citing and cited classes for that pair. For instance, if citing patent i belongs to two classes and cited patent j belongs to four classes, each class pair is assigned one eighth of that citation. The resulting dataset is then collapsed summing up weighted citations by citing and cited classes. This is used to compute the cosine dissimilarity measure. In particular, for every pair of citing $g(i) = \mathcal{A}$ and cited $g(j) = \mathcal{B}$ classes, technological distance is measured as:

$$TDS_{g(i)g(j)} = 1 - \frac{\sum_k C_{\mathcal{A}k} C_{\mathcal{B}k}}{\sqrt{\sum_k C^2_{\mathcal{A}k}}\sqrt{\sum_k C^2_{\mathcal{B}k}}} \tag{3.3}$$

Where $C_{\mathcal{A}k}$ and $C_{\mathcal{B}k}$ denote the weighted number of citations sent from patents in technology class $\mathcal{A}$ and technology class $\mathcal{B}$ to patents in technology class k, with k indexing all available classes. Intuitively, the fraction in (3.3) gives the similarity in the two vectors representing the distribution of citations of each class to all classes (the cosine of their angle), which is bounded in the $[0, 1]$ interval. Subtracting this value from one thus gives a measure of dissimilarity, or distance, based on how different the knowledge bases of the two classes are. Finally, we assign a weighted average of this measure to each patent pair in the estimating sample, based on all the technology classes associated with the citing and cited patents. We also recode the variable in terms of quintiles over the distribution in 2016-2019 (we retain the same variable name for simplicity). Equation (3.2) is then estimated for $h = 4$.

Figure 3.5: Marginal effects by technological distance (IPC4)



The graph displays coefficients obtained from a regression where ln SCI is interacted with quintiles of technological distance, controlling for quintile main effects, ln Distance, inventor networks, and differences in county level observables. Dashed lines are 95% CIs.

Figure 3.5 graphically reports the marginal effects of geographical distance (a) and social connectedness (b), allowing the coefficients to vary across quintiles of technological distance between citing and cited patents (IPC4 level). Vertical bars denote 95% confidence intervals. Although somewhat noisy, these estimates appear to give some credit to our hypothesis with respect to social connectedness. The coefficients display a clear positive sloping trend, with the SCI being statistically insignificant for the bottom quintile of technological distance. Moreover, the point estimates on the most technologically distant groups of citations are about twice as large as that measured for the first quintile. By contrast, there does not seem to be any statistically significant relationship between geographical distance and citation irrespective of which quintile is considered. These results hold also if distance between classes (IPC3), rather than subclasses, is considered (Appendix Figure 3.A.4).

3.5 Conclusions

This paper explored the role of informal social interaction, defined in terms of social connectedness, in the transfer of knowledge as captured by patent citation data. Using an index of aggregate Facebook ties to measure social connectedness between places, it finds that social proximity does seem to matter, positively influencing the probability of observing a citation between two places. This is robust to controlling for physical distance, the pre-existing geography of production (e.g., clustering due to other Marshallian forces such as matching or sharing), and the existence of professional links between any inventor involved in creating the citing or the cited patent (up to two degrees of distance). Interestingly, these effects seem to explain away the statistical significance of physical proximity. This suggests that informal social connectedness, despite its likely correlation with geographical distance, offers perhaps a more accurate measure to study knowledge flows. By this we do not mean to say that being socially connected can replace the importance of being co-located. Our analysis did not directly test for substitution of co-location in the *same* county, nor was it conclusive with respect to substitution between social and geographical proximity *across* counties. Rather, we note that physical proximity and social connectedness appear to be two ways by which inventors can access existing knowledge. In practice, most inventors will rely on both, especially to the extent that physically proximate places are also likely to be strongly connected socially. We document that the age of the cited patent might play a role in explaining the relevance of geographical, as opposed to social proximity. In the early stages of knowledge creation, spatial frictions are strong and spatial proximity facilitates access to knowledge. However, as knowledge becomes common domain in a geographical sense, the informal social environment in which inventors operate is increasingly important in shaping knowledge flows across counties, irrespective of physical distance. We also show that social connectedness matters most for entrepreneurs and garage inventors, and that it contributes bridging gaps between technologically distant knowledge areas. Our key takeaway is that no inventor is an island, as knowledge creation is inherently a social process. This is not just true for interactions with colleagues in the profession, but also with respect to informal ties in the inventors' social environment.

In terms of magnitude, the effect of informal interaction is quite small. According to our preferred specification, doubling social connectedness increases citation likelihood by about a third of a percentage point. Social connectedness, however, can be economically meaningful. Two counties at the 75^{th} percentile of social connectedness are on average 1.2 percentage points more likely to cite one another than a pair of counties at the 25^{th} percentile. To be more concrete, consider the following example. The counties of Colleton and Dorchester in South Carolina neighbour each other geographically. The latter, however, has a connectedness strength to Santa Clara County in California (one of the top patenting counties in the US) at the 75^{th} percentile of the overall distribution for county-pairs, while the former is only at the 25^{th} percentile. Between 2016 and 2019, there were fourteen times as many applicant citations between Santa Clara and Dorchester, than between Santa Clara and Colleton.[29] This difference is striking considering that the two counties are contiguous and certainly within commuting distance from each other. Moving inventors from one to the other can potentially have implications for their exposure to ideas. While admittedly anecdotal, and granted that it is hard to imagine that there is actually a sharp discontinuity in connectedness at the county border, this example helps illustrate the local variation existing in this measure, and the tangible difference that social connectedness can make for knowledge flows. There are several other instances where this type of change can be achieved by moving relatively close in space. Appendix Figure 3.A.5 shows counties connected to Santa Clara, CA, with strength at least as strong as the upper quartile (in blue), or at least as weak as the lower quartile (in green). Evidently, green and blue counties are frequently located in close proximity.[30]

There are several limitations to the present work. The most important concern relates to measurement. What is the SCI capturing in practice? With the level of aggregation used in this analysis, we can only gauge an indirect picture. Ideally, one would observe the entire social graph of inventors, allowing to explicitly account for the nature and strength of connections, as well as more generally to study the topography of this graph. The

[29] In terms of propensity, the likelihood of observing a citation by an applicant, compared to all citations, is 20% greater between Santa Clara and Dorchester, than between Santa Clara and Colleton.

[30] More systematically, Appendix Figure 3.A.6 shows that of all county pairs strongly and weakly connected to the same third county, over 5% are within 400 kilometres of distance from each other, and over 20% are within a 1000 kilometres catchment area.

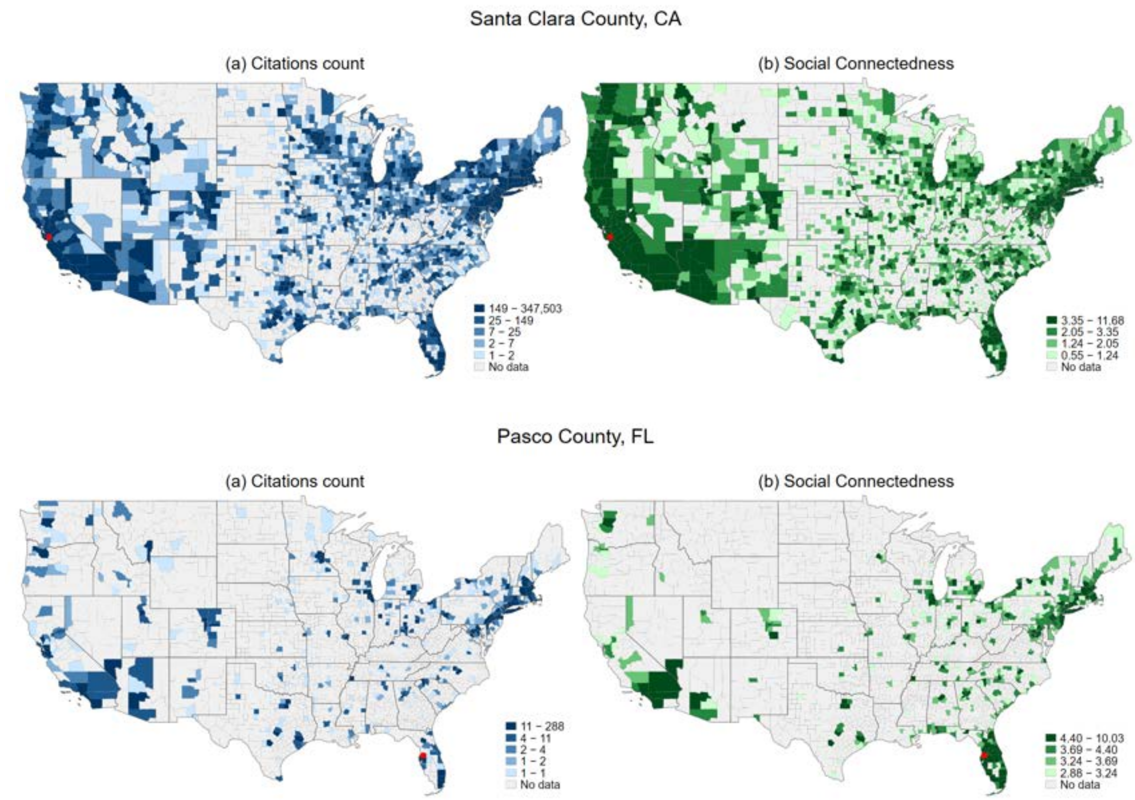
SCI, however, also has some advantages over analyses of this kind. To our knowledge, for instance, this index represents the most comprehensive measure of revealed social interaction available yet for the entire geography of the US. Moreover, failing to observe the full network of inventors, we align to previous work by measuring the professional network of inventors as proxied by co-patenting links. Future work could consider focusing on a subset of the data to construct higher-order connections, which could not be done in this paper due to computational constraints. Another problem relates to the possible endogeneity of the SCI measure. Omitted variable bias, for instance, could arise to the extent that people and economic activity tend to cluster around certain areas in response to natural comparative advantages and history. Our estimating framework has attempted to mitigate this concern, along with robustness checks that restricted the sample of citations to exchanges between non-coastal regions. Admittedly, however, this strategy is incomplete. The ideal experiment would randomly re-wire the social connectivity of all US citizens and measure the resulting effects on knowledge exchange. Finally, a reminder that all results depend on the identifying assumptions underlying the use of examiner citations as a control group. The literature is yet to form a clear view regarding the nature of these citations and possible biases they may cause (Alcácer and Gittelman, 2006; Alcácer et al., 2009; Righi and Simcoe, 2019). In the ideal picture, the examiners simply fill in all technological connections to a patent that the applicant was not aware of. In practice, however, citations are potentially also added by patent attorneys, and examiners might be limited by their own imperfect search process. As such, results should be interpreted as the relative effect of knowledge flows to the applicant, above and beyond any bias accruing to the examiner (rather than relative to an ideal omniscient actor). This, however, is likely to work against the detection of any effect. A comparison of our estimates to those of Bailey et al. (2018b), who use a case-control matching approach and estimate stronger effects, would indeed suggest that any distortion in our method biases results downward. The estimate we provide is thus conservative. We also express a word of caution in terms of the way knowledge flows are measured in this paper. We relied on patents due to the ease of tracking exchanges via citations and to the availability of structured data, but these data have well-known limitations (see Section 3.3.1). Future work could investigate other types of knowledge exchange that would be more likely to be channelled over informal ties.

There are also ways in which this work can be refined and expanded. One possibility is to investigate whether stronger social connectedness is significantly associated to weaker industrial agglomeration locally. Similarly, it would be interesting to study what types of clusters rely more on this resource. Could it be that large diversified urban agglomerations draw on this connectedness, or is it smaller, more specialised clusters that reap most benefits from stronger informal ties to actors elsewhere? Another important although more challenging question would be to distinguish SCI-mediated knowledge flows from pure spillovers. Indeed, observing that knowledge is more likely to flow from one place to another does not necessarily entail that it causes productivity-enhancing spillovers, or that the exchange took place outside market boundaries. In its simplest form, this analysis would investigate whether stronger social connectedness is associated to the production of higher-quality ideas holding inputs constant, where quality can be approximated using counts of downstream citations. This could be additionally integrated with the study of spillovers between specific industries, contributing to the understanding of how different 'trees of knowledge' emerge. Finally, another line of inquiry could take a closer look at the nature of populations and their social ties, exploring how and why people in different places are interconnected.

In conclusion, while this paper has attempted to set the ground for a sound investigation into the physical and social geographies of knowledge exchange, evidently a great amount of work still lies ahead.

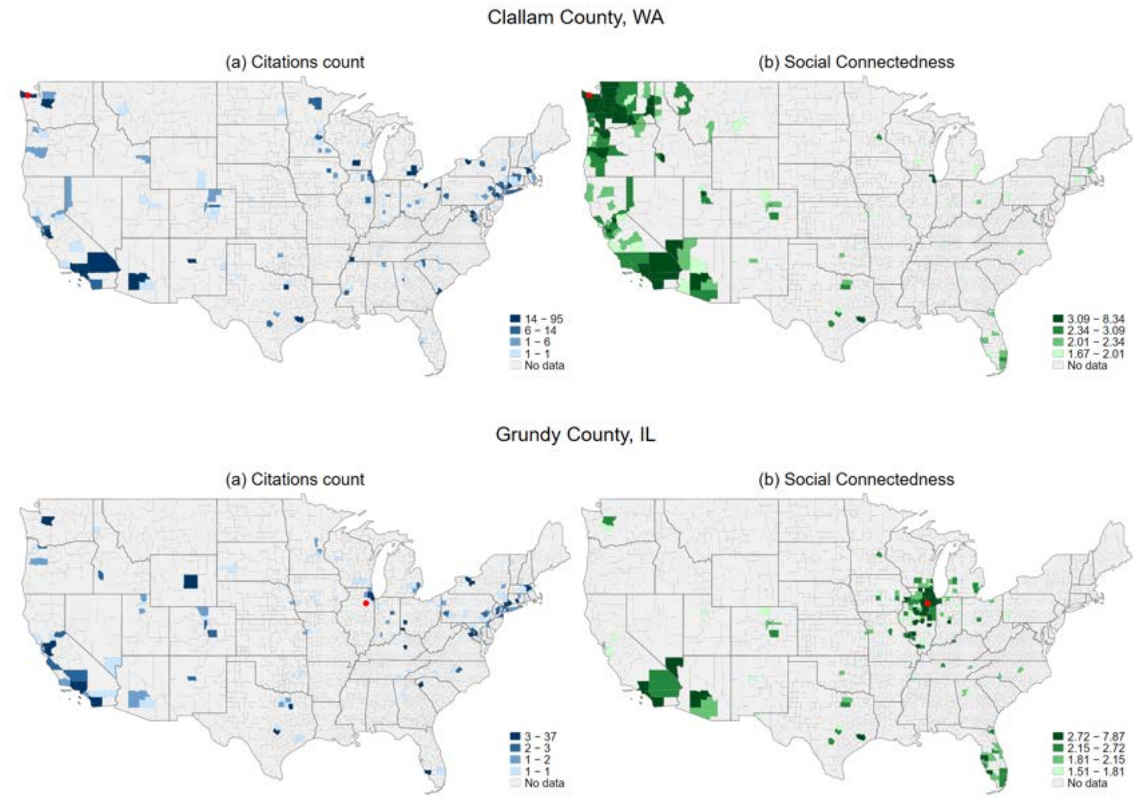
3.A Additional Figures

Figure 3.A.1: Network Maps of US Counties by Quartiles



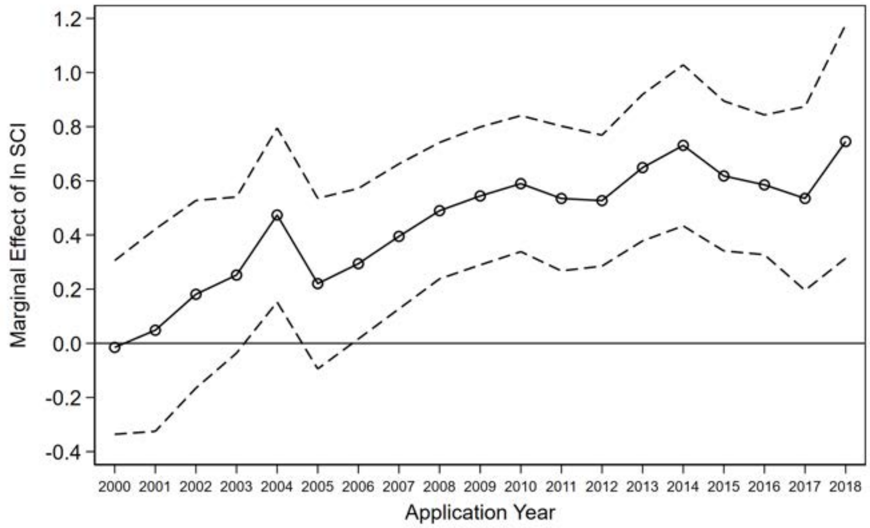
Notes: Panel (a) in each map shows, for a given citing county, all counties that receive citations by patents issued in the 2016-2019 period. Polygons are coloured proportional to quartiles of received citation counts. Panel (b) shows the log of social connectedness for counties most strongly connected to the citing one, limiting the sample to the same number of counties as those receiving at least one citation in panel (a). Polygons are coloured proportional to quartiles of connection strength. The similarity in panels (a) and (b) for each citing county suggests that there is a correlation between knowledge flows and social connectedness. Citing counties were selected to represent respectively the 99$^{\text{th}}$, 75$^{\text{th}}$, 50$^{\text{th}}$, and 25$^{\text{th}}$ percentiles in the distribution of sent citations, conditional on citing at least 100 different counties.

Figure 3.A.2: Network Maps of US Counties by Quartiles (continued)



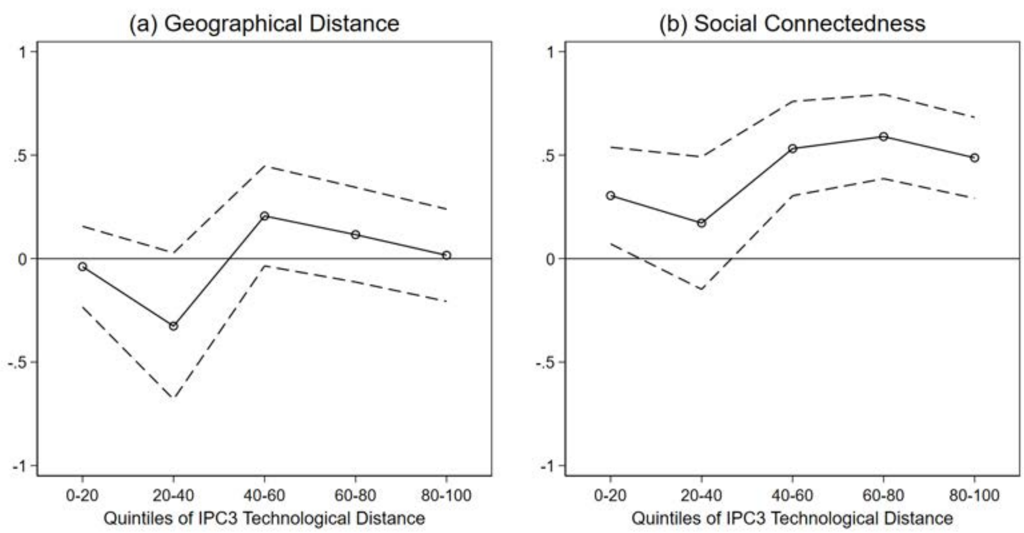
Notes: Panel (a) in each map shows, for a given citing county, all counties that receive citations by patents issued in the 2016-2019 period. Polygons are coloured proportional to quartiles of received citation counts. Panel (b) shows the log of social connectedness for counties most strongly connected to the citing one, limiting the sample to the same number of counties as those receiving at least one citation in panel (a). Polygons are coloured proportional to quartiles of connection strength. The similarity in panels (a) and (b) for each citing county suggests that there is a correlation between knowledge flows and social connectedness. Citing counties were selected to represent respectively the 99$^{\text{th}}$, 75$^{\text{th}}$, 50$^{\text{th}}$, and 25$^{\text{th}}$ percentiles in the distribution of sent citations, conditional on citing at least 100 different counties.

Figure 3.A.3: Marginal effects by citing patent application year



The graph displays coefficients obtained from a regression where ln SCI is interacted with the patent's application year, controlling for application year main effects, ln Distance, inventor networks, and differences in county level observables. Dashed lines are 95% CIs.

Figure 3.A.4: Marginal effects by technological distance (IPC3)



The graph displays coefficients obtained from a regression where ln SCI is interacted with quintiles of technological distance, controlling for quintile main effects, ln Distance, inventor networks, and differences in county level observables. Dashed lines are 95% CIs.

Figure 3.A.5: Strongly and weakly connected counties to Santa Clara, CA

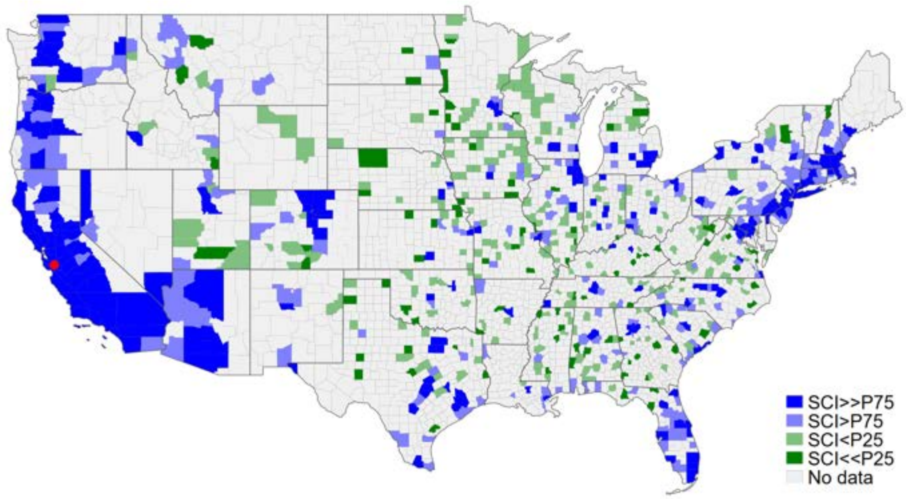
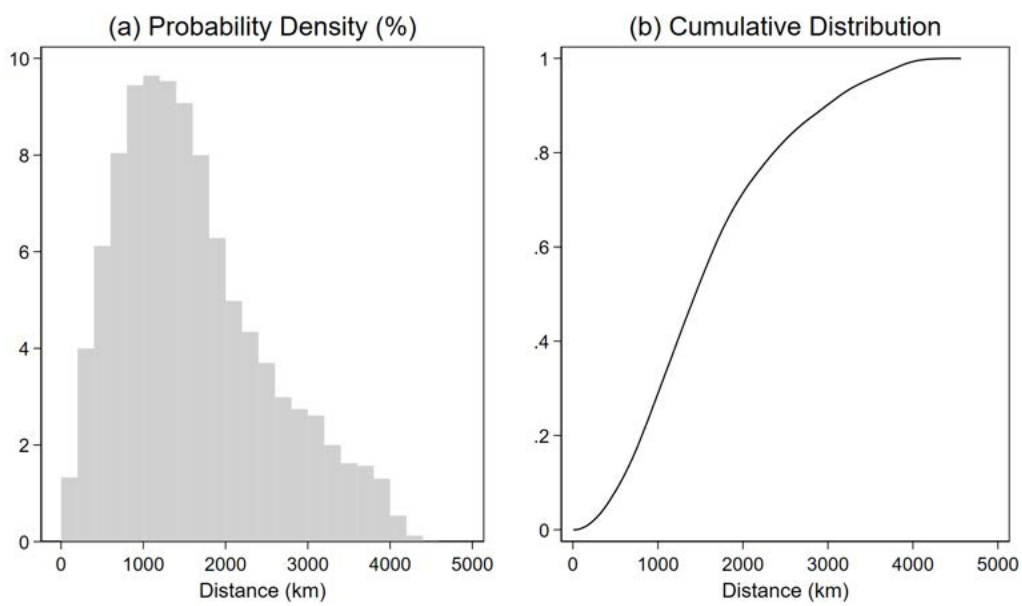


Figure 3.A.6: PDF and CDF of strongly and weakly connected county pairs



PDF and CDF for geographical distance between county pairs below the lower and above the upper quartiles of social connectednes strength with the same third county. In (a), bin width is 200 km.

3.B Additional Tables

Table 3.B.1: Complete list of WIPO technology fields

Code	Field Title
1	Electrical engineering: Electrical machinery, apparatus, energy
2	Electrical engineering: Audio-visual technology
3	Electrical engineering: Telecommunications
4	Electrical engineering: Digital communication
5	Electrical engineering: Basic communication processes
6	Electrical engineering: Computer technology
7	Electrical engineering: IT methods for management
8	Electrical engineering: Semiconductors
9	Instruments: Optics
10	Instruments: Measurement
11	Instruments: Analysis of biological materials
12	Instruments: Control
13	Instruments: Medical technology
14	Chemistry: Organic fine chemistry
15	Chemistry: Biotechnology
16	Chemistry: Pharmaceuticals
17	Chemistry: Macromolecular chemistry, polymers
18	Chemistry: Food chemistry
19	Chemistry: Basic materials chemistry
20	Chemistry: Materials, metallurgy
21	Chemistry: Surface technology, coating
22	Chemistry: Micro-structural and nano-technology
23	Chemistry: Chemical engineering
24	Chemistry: Environmental technology
25	Mechanical engineering: Handling
26	Mechanical engineering: Machine tools
27	Mechanical engineering: Engines, pumps, turbines
28	Mechanical engineering: Textile and paper machines
29	Mechanical engineering: Other special machines
30	Mechanical engineering: Thermal processes and apparatus
31	Mechanical engineering: Mechanical elements
32	Mechanical engineering: Transport
33	Other fields: Furniture, games
34	Other fields: Other consumer goods
35	Other fields: Civil engineering

Table 3.B.2: Summary statistics for citing patents

	Mean	Median	Std. Dev.	Min.	Max.
Issue year	2017.39	2017	1.07	2016	2019
Application year	2014.69	2015	1.77	2008	2019
Citations per patent	23.20	6	90.64	1	4154
Share of applicant citations	0.62	0.79	0.40	0	1
Cited WIPO	2.82	2	2.71	1	34
Cited IPC3 (first)	3.15	2	3.59	1	65
Cited IPC4 (first)	4.47	3	6.50	1	169
Assignee age (max)	19.11	15	15.06	0	43
Number of citing patents					489,230
Share of citing patens with only applicant citations					0.29
Share of citing patens with only examiner citations					0.22

Table 3.B.3: Summary statistics for citing patents, 20% random sample (2002-2019)

	Mean	Median	Std. Dev.	Min.	Max.
Issue year	2011.37	2012	5.11	2002	2019
Application year	2008.39	2009	5.34	1995	2019
Citations per patent	19.79	7	60.75	1	4204
Share of applicant citations	0.57	0.67	0.40	0	1
Cited WIPO	2.82	2	2.48	1	33
Cited IPC3 (first)	3.15	2	3.20	1	67
Cited IPC4 (first)	4.36	3	5.47	1	169
Assignee age (max)	17.42	15	13.38	0	43
Number of citing patents					364,372
Share of citing patens with only applicant citations					0.21
Share of citing patens with only examiner citations					0.24

Table 3.B.4: Summary statistics for cited patents

	Applicant		Examiner		Total	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
SCI	20,684.26	80,484.62	18,607.88	81,280.64	20,488.80	80,562.16
ln SCI	6.20	3.03	5.80	3.12	6.17	3.04
Distance (km)	1,573.88	1,426.67	1,630.35	1,376.18	1,579.19	1,422.09
ln Distance	6.06	2.63	6.27	2.48	6.08	2.61
Prof. network	0.19	0.39	0.12	0.33	0.19	0.39
Same inventor	0.07	0.26	0.07	0.26	0.07	0.26
Co-authored	0.07	0.25	0.03	0.17	0.07	0.25
Shared co-author	0.05	0.22	0.02	0.15	0.05	0.22
Same assignee	0.11	0.31	0.09	0.29	0.11	0.31
Issue year	2004.01	6.75	2004.49	7.34	2004.05	6.81
Application year	2001.08	6.27	2001.68	6.86	2001.13	6.33
Patent age (since app. +18m)	11.64	6.12	10.93	6.61	11.57	6.17
Patent age (since issue)	10.21	6.59	9.63	7.11	10.16	6.64
Same county	0.12	0.33	0.11	0.31	0.12	0.33
Same CZ	0.05	0.23	0.04	0.20	0.05	0.23
Other state	0.72	0.45	0.77	0.42	0.73	0.45
Number of applicant citations						10,179,877
Number of examiner citations						1,169,519
Total number of citations						11,349,396

Table 3.B.5: Summary statistics for cited patents, 20% random sample (2002-2019)

	Applicant		Examiner		Total	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
SCI	20,058.38	82,154.42	17,332.93	80,318.63	19,691.74	81,915.13
ln SCI	6.13	3.00	5.70	3.07	6.07	3.01
Distance (km)	1,607.31	1,419.45	1,643.28	1,371.51	1,612.15	1,413.15
ln Distance	6.12	2.60	6.33	2.41	6.15	2.58
Prof. network	0.16	0.37	0.10	0.31	0.16	0.36
Same inventor	0.06	0.25	0.06	0.24	0.06	0.25
Co-authored	0.06	0.23	0.03	0.16	0.05	0.23
Shared co-author	0.04	0.20	0.02	0.13	0.04	0.19
Same assignee	0.10	0.30	0.08	0.27	0.10	0.30
Issue year	2000.18	7.10	1999.50	7.49	2000.09	7.16
Application year	1997.57	6.63	1997.03	7.05	1997.50	6.69
Patent age (since app. +18m)	10.17	5.99	8.74	6.31	9.98	6.06
Patent age (since issue)	9.07	6.28	7.78	6.59	8.90	6.34
Same county	0.12	0.33	0.10	0.30	0.12	0.32
Same CZ	0.05	0.22	0.04	0.20	0.05	0.22
Other state	0.74	0.44	0.79	0.41	0.74	0.44
Number of applicant citations						6,090,796
Number of examiner citations						1,121,574
Total number of citations						7,212,370

Table 3.B.6: Overview of all variables used in the analysis

For county pairs:	Mean	Std. Dev.	Min.	25th Pct.	Median	75th Pct.	Max.
SCI	146.29	3,220.04	0.00	1.90	7.69	30.84	1,000,000.00
ln SCI	2.04	2.17	-6.67	0.64	2.04	3.43	13.82
Distance (km)	1,530.45	1,055.45	0.00	714.56	1,281.88	2,188.26	4,561.70
ln Distance	7.00	1.01	0.00	6.57	7.16	7.69	8.43
Gross mig. flow	216.88	2,870.38	0.00	0.00	0.00	0.00	325,606.00
ln Gross mig. flow	0.99	2.19	0.00	0.00	0.00	0.00	12.69
D Bachelor (%)	12.88	9.58	0.00	5.20	11.00	18.70	63.20
D Inventors (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.01
D Density	1,846.81	6,049.87	0.00	203.37	595.61	1,474.77	69,467.53
D Median income	7,937.54	6,442.72	0.00	2,836.00	6,334.00	11,598.00	47,098.00
D Unemployment (%)	2.60	2.09	0.00	1.00	2.10	3.70	24.70
D White (%)	21.14	16.64	0.00	7.58	17.26	31.31	94.65
D Black (%)	11.04	11.88	0.00	2.45	7.03	15.65	81.53
D Asian (%)	4.17	5.43	0.00	0.96	2.35	4.80	33.00
D Hispanic (%)	11.64	12.66	0.00	2.63	7.07	16.21	95.06
For patent pairs:	Mean	Std. Dev.	Min.	25th Pct.	Median	75th Pct.	Max.
Citation	0.91	0.29	0.00	1.00	1.00	1.00	1.00
Prof. network	0.19	0.39	0.00	0.00	0.00	0.00	1.00
Same inventor	0.07	0.26	0.00	0.00	0.00	0.00	1.00
Co-authored	0.07	0.25	0.00	0.00	0.00	0.00	1.00
Shared co-author	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Same assignee	0.11	0.31	0.00	0.00	0.00	0.00	1.00
Same county	0.12	0.33	0.00	0.00	0.00	0.00	1.00
Same CZ	0.05	0.23	0.00	0.00	0.00	0.00	1.00
Other state	0.73	0.45	0.00	0.00	1.00	1.00	1.00
Issue year	2004.06	6.80	1982.00	1999.00	2004.00	2010.00	2019.00
Application year	2001.14	6.33	1981.00	1997.00	2001.00	2006.00	2017.00
Patent age (since app. +18m)	11.57	6.17	0.00	7.00	12.00	16.00	26.00
Patent age (since issue)	10.16	6.64	0.00	4.00	10.00	15.00	27.00
Tech. distance (IPC3)	0.32	0.32	0.00	0.00	0.27	0.56	1.00
Tech. distance (IPC4)	0.40	0.32	0.00	0.03	0.39	0.65	1.00
For the estimation sample:	Mean	Std. Dev.	Min.	25th Pct.	Median	75th Pct.	Max.
ln SCI	5.36	2.41	-6.67	3.89	5.43	7.01	11.37
ln Distance	7.02	1.28	1.43	6.52	7.32	8.03	8.43
Prof. network	0.08	0.27	0.00	0.00	0.00	0.00	1.00

Table 3.B.7: Main regressions with details on bilateral controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln SCI	0.266	0.345	0.359	0.435	0.129	0.435	0.0498	0.0720	0.444
	(0.109)^{b}	(0.111)^{a}	(0.110)^{a}	(0.112)^{a}	(0.0270)^{a}	(0.113)^{a}	(0.0279)^{c}	(0.0246)^{a}	(0.110)^{a}
ln Distance	-0.00243	-0.00559	0.00836	0.0688	0.0229	0.0754	-0.0136	0.00438	0.0472
	(0.0998)	(0.100)	(0.100)	(0.0898)	(0.0186)	(0.0888)	(0.0137)	(0.0160)	(0.104)
Prof. network	2.963	2.962	2.958	2.934	0.517	2.828	0.274	0.388	3.056
	(0.556)^{a}	(0.557)^{a}	(0.556)^{a}	(0.555)^{a}	(0.145)^{a}	(0.541)^{a}	(0.0746)^{a}	(0.114)^{a}	(0.541)^{a}
ln Gross mig. flow		-0.0662	-0.0612	-0.0406	0.0175	-0.0351	0.0325	0.0234	-0.0373
		(0.0542)	(0.0539)	(0.0622)	(0.0165)	(0.0617)	(0.0167)^{c}	(0.0154)	(0.0592)
Top 50 colleges=1			-0.391	-0.507	-0.0470	-0.497	-0.106	-0.0511	-0.460
			(0.236)^{c}	(0.261)^{c}	(0.0373)	(0.253)^{c}	(0.0588)^{c}	(0.0370)	(0.268)^{c}
D Bachelor (%)				0.0285	0.00229	0.0292	-0.000114	0.00141	0.0270
				(0.0106)^{a}	(0.00208)	(0.0102)^{a}	(0.00189)	(0.00193)	(0.0105)^{b}
D Inventors (%)				161.3	-6.204	159.5	-24.35	-14.25	167.4
				(70.95)^{b}	(18.49)	(69.66)^{b}	(12.62)^{c}	(18.19)	(84.31)^{b}
D Density				0.0000191	-0.00000261	0.0000186	0.000000372	-0.00000263	0.0000254
				(0.0000126)	(0.00000683)	(0.0000120)	(0.00000396)	(0.00000499)	(0.0000119)^{b}
D Median income				-0.0000102	-0.00000347	-0.0000107	-0.000000439	-0.00000365	-0.0000101
				(0.0000133)	(0.00000314)	(0.0000129)	(0.00000321)	(0.00000297)	(0.0000135)
D Unemployment (%)				-0.0265	-0.0223	-0.0293	-0.00819	-0.0181	-0.0287
				(0.0403)	(0.00772)^{a}	(0.0392)	(0.00967)	(0.00703)^{b}	(0.0400)
D White (%)				-0.00854	0.000103	-0.00826	-0.0000926	-0.000173	-0.00957
				(0.00425)^{b}	(0.000932)	(0.00412)^{b}	(0.00104)	(0.000960)	(0.00397)^{b}
D Black (%)				0.0120	0.00409	0.0113	0.00325	0.00308	0.0125
				(0.00725)^{c}	(0.00149)^{a}	(0.00701)	(0.00175)^{c}	(0.00150)^{b}	(0.00710)^{c}
D Asian (%)				0.00873	0.00326	0.00860	-0.000414	0.00321	0.00498
				(0.0106)	(0.00268)	(0.0101)	(0.00118)	(0.00246)	(0.0114)
D Hispanic (%)				0.00108	-0.000393	0.000824	0.00138	-0.000560	0.00139
				(0.00597)	(0.00137)	(0.00591)	(0.00151)	(0.00133)	(0.00612)
WIPO pairs FEs	•	•	•	•	•	•	•	•	•
Bilat. controls				•	•	•	•	•	•
Within citing					•		•	•	
Within cited							•		
Interaction samp.									•
Adj. R^2	0.1144	0.1144	0.1144	0.1145	0.6044	0.1137	0.6620	0.5895	0.1219
R^2	0.1151	0.1151	0.1151	0.1151	0.6216	0.1143	0.7070	0.6066	0.1226
N	8,839,486	8,835,898	8,835,898	8,835,705	8,761,974	8,761,974	8,332,097	8,332,097	8,833,640

Two-way cluster-robust standard errors for citing and cited CZ pairs (Cameron et al., 2011). Significance levels: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$. The outcome variable is expressed in terms of percentage points. All specifications use citing and cited year and county fixed effects. The sample excludes citations within same assignee or same county. Interaction controls: main effects for own CZ or state, other state, elapsed time, assignee age, IPC4 technological distance. Column (6) estimates the same model as (4), restricting the sample to that in (5). Similarly, (8) estimates the model in (5) on the sample used in (7). These restrictions allow to compare coefficient changes due to changes in the specification, as opposed to changes in the sample. The reduced ln SCI coefficient in (5) can be largely attributed to the effect of citing patent dummies. By contrast, large part of the fall in the magnitude of ln SCI effects in (7) is due to a change in the sample, as opposed to the use of cited patent dummies.

Table 3.B.8: Robustness checks with citing patent dummies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln SCI	0.129	0.100	0.113	0.119	0.111	0.137	0.189
	$(0.0270)^a$	$(0.0290)^a$	(0.0959)	$(0.0263)^a$	$(0.0260)^a$	$(0.0401)^a$	$(0.0371)^a$
ln Distance	0.0229	0.00713	-0.0218	0.0181	0.0129	0.0199	0.0241
	(0.0186)	(0.0192)	(0.0596)	(0.0196)	(0.0191)	(0.0413)	(0.0276)
Prof. network	0.517	0.505	0.810	0.520	0.517	0.380	0.767
	$(0.145)^a$	$(0.130)^a$	$(0.372)^b$	$(0.146)^a$	$(0.148)^a$	$(0.164)^b$	$(0.243)^a$
Tech. pairs FEs	WIPO	WIPO	WIPO	IPC3	IPC4	WIPO	WIPO
Bilat. controls	•	•	•	•	•	•	•
Whithin citing	•	•	•	•	•	•	•
Appl. year FEs		•					
Single-authored			•				
Non coastal						•	
Trimmed							•
Adj. R^2	0.6044	0.6038	0.6649	0.6046	0.6084	0.6154	0.4178
R^2	0.6216	0.6210	0.7014	0.6222	0.6283	0.6369	0.4402
N	8,761,974	8,761,973	692,110	9,015,773	8,989,740	4,958,565	5,801,815

Two-way cluster-robust standard errors for citing and cited CZ pairs (Cameron et al., 2011). Significance levels: ${}^a p < 0.01$; ${}^b p < 0.05$; ${}^c p < 0.1$. The outcome variable is expressed in terms of percentage points. All specifications use citing and cited year and county fixed effects. The sample excludes citations within same assignee or same county. Bilateral controls: gross migration, top 50 college, differences in education, inventors, density, income, ethnicity. The single-authored sample drops citations sent or received by patents with multiple authors. The non coastal sample drops citations originating or received in Census divisions bordering the Atlantic and Pacific coasts. The trimmed sample drops patents with citations added exclusively by the applicant or the examiner.

Chapter 4

Diffusion of Fracking Shocks over Social Networks

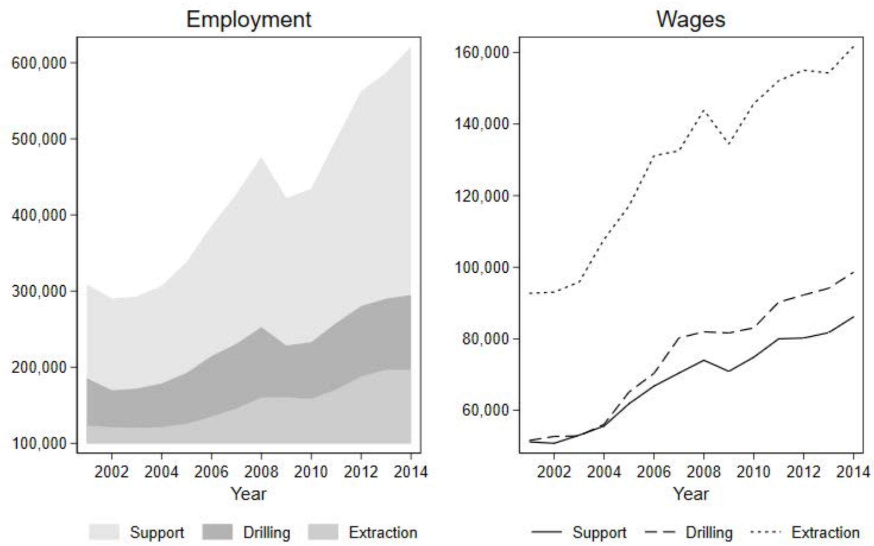
4.1 Introduction

The aim of this paper is to study how localised economic shocks can propagate across a country through interaction of people on social networks. In particular, the paper looks at shocks associated to the 'fracking revolution' in the US, taking place since the early 2000s. Fracking, or hydraulic fracturing, is a resource-extraction technology that uses highly-pressured liquid to obtain gas and oil from shale rock deposits. The presence of rich oil and gas deposits in shale formations across the US has been known for some time. It was however only around the turn of this century that a combination of technological innovation in extraction techniques and favourable market conditions allowed these reserves to be profitably exploited (DOE, 2009; Wang and Krupnick, 2015). Due to hydraulic fracturing and horizontal drilling, domestic production of oil and gas has been increasing steadily in the US. In 2017, crude oil production exceeded 1972 levels, and natural gas production reached a new record-high (EIA, 2018). As a result, drilling, extraction and support jobs in oil and gas operations nearly doubled between 2001 and 2014, with nominal wages growing by about 60% according to the US Bureau of Labor Statistics (Figure 4.1). Feyrer et al. (2017) emphasise that this activity is highly localised, making it a suitable case for the analysis proposed herein.

There is abundant literature on the regional economic effects of natural resources and the spillovers of local shocks to the economy in general, many of which focus on the US.[1]

[1] Outside the US, other recent evidence specific to oil and gas extraction is offered for instance by Caselli and Michaels (2013) for Brazil, by Borge et al. (2015) for Norway, by Percoco (2012b) for Italy, and by Gibbons et al. (2016) for the UK. The latter paper offers evidence on fracking, which is rare for Europe given widespread bans on this technology. All papers emphasise the importance of local institutional arrangements in determining economic effects, so the rest of this discussion focuses on the US.

Figure 4.1: Growth in employment and annual average pay in private sector firms



Source: US Bureau of Labor Statistics, Quarterly Census of Employment and Wages

Some scholars argue that the discovery of natural riches can harm local economies, in line with the resource curse literature (Sachs and Warner, 1995, 2001), for instance by crowding out employment from other sectors (Corden and Neary, 1982). Jacobsen and Parker (2016) show that the 1970s US oil boom caused harm to long term income and employment of local communities despite some short-term gains. Similarly, Black et al. (2005) study the boom and bust of the coal sector in four US states around the same period, finding small employment spillovers only into sectors producing locally traded goods. By contrast, several papers highlight the benefits that can accrue to regional economies. Michaels (2011) shows that in the very long run oil abundant counties in the southern US increased local employment density in mining as well as manufacturing, contributing to population growth, better infrastructure, and higher per capita income. Other studies confirm that resource extraction can benefit the manufacturing sector rather than harming it, contributing to local economic development (Fetzer, 2014; Weber, 2014; Allcott and Keniston, 2017). Some studies have also looked at non-monetary outcomes, such as marriage and fertility rates, and risky sexual behaviour. Shale gas extraction is associated with an increase in marital and non-marital birth rates due to the higher earnings potential of low-skilled men (Kearney and Wilson, 2018), as well with higher gonorrhea rates, with significant spatial spillovers from fracking sites (Cunningham et al., 2020). Bartik et al. (2019) provide a comprehensive discussion of the economic and welfare consequences of fracking for local communities, studying a wide range of outcomes including income, employment, housing and crime. The paper documents net average welfare gains from hydraulic fracturing across US shale plays, albeit with large heterogeneity between them. Recently, Feyrer et al. (2017) look at the dispersion of fracking-determined income shocks over space, time, and industries using novel data on yearly production of oil and gas from new wells in US counties between 2004 and 2014. The authors find that the effect of fracking on income and employment becomes larger as one considers the wider region around the county where production occurs, peaking at about 100 miles of distance. This effect is persistent over time and, while changes in mining wages disappear within two years, workers in other industries such as transport, manufacturing, and services, benefit from sustained growth in their earnings. Taken together, these results suggest that benefits from local shocks can propagate to the wider economy of a country.

While the majority of extant literature has focused on geographic spillovers of localised shocks to *proximate* areas, however, the role of networks in this process is relatively understudied. To address this gap, Amarasinghe et al. (2018) jointly investigate the role of geographic, transport and ethnic networks in the propagation of mining-related shocks across African administrative districts. Their findings highlight the importance of road and ethnic networks in the diffusion of economic shocks well beyond immediately contiguous areas. Other scholars have focussed on the macroeconomic relevance of networks in transmitting micro-level shocks, studying for instance input-output relationships between firms (Acemoglu et al., 2012; Carvalho, 2014). This paper is interested in studying the relevance of *social* networks, or better, the social connectedness of places arising from the interaction of people across the entire US geography. Bailey et al. (2018b) show that social connectedness correlates with many economic outcomes, including trade flows, mobility and innovation.[2] The micro-level literature on economic networks provides valuable insights into some of the mechanisms underlying these findings (Jackson et al., 2017). From the viewpoint of this paper, of particular relevance is the work of Calvó-Armengol and Jackson (2004, 2007), who develop a network model in which workers rely on their social relationships to obtain information about employment opportunities. The model predicts positive correlation of wages and employment status on networks. This intuition finds validation in subsequent empirical work on the labour market effects of information and referral networks (Bayer et al., 2008; Patacchini and Zenou, 2012; Beaman, 2012; Dustmann et al., 2016; Gee et al., 2017).[3]

[2]In a related paper, Bailey et al. (2018a) use micro-data on social connections to show that exposure to fluctuations in housing prices via one's network influences beliefs about attractiveness of property investments and ultimately housing market activity of this individual.
[3]See Topa (2011) and Topa and Zenou (2015) for recent reviews of this literature.

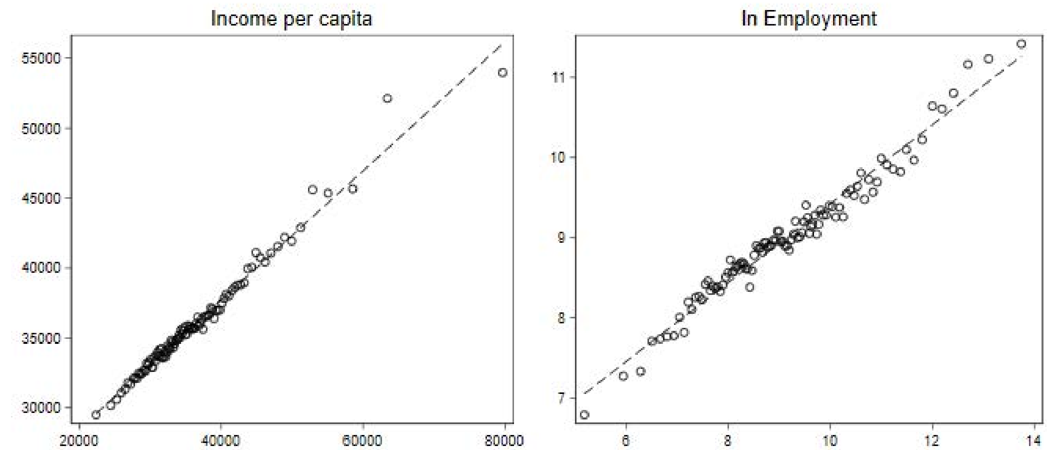
Empirically, this can also be seen in the aggregate data used for the analysis in this paper. The binned scatterplot in Figure 4.2 plots percentiles of income per capita and log employment in US counties against averages of the same measures taken over the top five percent most closely socially connected counties.[4] Evidently, there appears to be a strong positive autocorrelation of both income and employment in the network. Note that this is not in itself evidence of endogenous network effects. It is well possible that counties that are socially connected are similar in demographic composition due to sorting of people into places and networks, or that connected counties are exposed to the same economic shocks. It is also uncertain whether the correlation arises because of a change in outcomes of connected counties, or of local socio-economic conditions. Borrowing the terminology of Manski (1993), the relationship described in Figure 4.2 could be the result of correlated or contextual effects, the latter being especially difficult to distinguish from endogenous effects. As emphasised by Gibbons and Overman (2012) and Gibbons et al. (2015), it is possible to make some way forward in the identification of the desired effects if one can find exogenous instruments as a source of variation in the network variables. In this respect, fracking provides a suitable setting for the study of such effects insofar as resource extraction is a function of the exogenous pre-existing geology of shale formations. As such, the study of diffusion of fracking shocks can also be interpreted as the reduced-form analysis of endogenous network effects. This aligns with what recommended in Gibbons and Overman (2012), who suggest to rely on spatially lagged X models (SLX) in place of spatial autoregressive models (SAR), as the former are more amenable to be studied with an experimentalist paradigm in mind. More in general, while it may be difficult to separately identify endogenous and contextual effects, studying plausibly exogenous fracking shocks can help address concerns related to correlated effects.

Finally, this analysis also indirectly dialogues with a broader line of investigation concerning the effects of localised shocks to labour demand, geographic mobility, and the subsequent adjustments to equilibrium in labour markets (Blanchard et al., 1992; Bound and Holzer, 2000; Notowidigdo, 2011; Manning and Petrongolo, 2017; Amior and Manning, 2018; Ahlfeldt et al., 2020). Most of these studies discuss the limited role that mobility of low-skill workers play in the adjustment process. Conversely, by studying the fracking industry, this paper documents effects operating predominantly through the channel of low skilled employment.

Motivated by the stylised fact noted in Figure 4.2, and building on the existing literature

[4]Social connections are defined in terms of number of online friendships the counties share. More information on these data are available in Section 4.3.

Figure 4.2: Autocorrelation of county labour market outcomes in social networks



Binned scatterplot using percentiles of x-values. The y-axis measures averages over the top 5% strongest social connections of each county. Observations are over the 2004-2012 period, absorbing year-trends.

Sources: US BLS Quarterly Census of Employment and Wages; Facebook Social Connectedness Index

discussed above, this paper therefore aims to investigate how localised exogenous shocks from new oil and gas production diffuse via social connectedness across the entire US geography. To the best of my knowledge, no research has looked at this question yet. The paper thus aims to describe a new geography for the income and employment effects of resource booms, which has been largely overlooked in local economic development studies. In line with existing evidence (Feyrer et al., 2017), I find that the largest effects of localised shocks are felt in geographically proximate areas. However, social networks do play a role. On average, a million dollar per capita increase in oil and gas extraction in the top 25 most strongly socially connected counties raises per capita wages by about 2,000 dollars for workers reporting their income in counties located as far as 1,200 km away from the drilling site. This effect is likely to be explained by the relocation of itinerant workers within the industry, providing new aggregate evidence in support of the literature on job information networks. This finding is of relevance to policy makers interested in local economic development. If being socially connected to thriving places can benefit local economies above and beyond immediately contiguous areas, then this research sheds light onto the importance of considering a new dimension of access to opportunity, namely one that takes into account the interaction of people across distant geographies. Further, this analysis reveals the potential of spillovers of place-based interventions beyond contiguous areas, in a way that depends on the geography of social interactions.

Importantly, it is beyond the scope of this analysis to evaluate the overall welfare effects of hydraulic fracturing. While some places stand to gain in terms of wages or employment, there are well documented negative externalities associated with this extraction technology. Fracking has been associated to environmental damages (Howarth and Ingraffea, 2011) including deterioration of air quality (Colborn et al., 2014; Roy et al., 2014; Caulton et al., 2014) and contamination of water reserves due to by-products of the drilling process (Olmstead et al., 2013; Warner et al., 2013; Jackson et al., 2013; Vengosh et al., 2013; Fontenot et al., 2013). It was also found to increase crime rates, inequality and road traffic accidents (James and Smith, 2017; Graham et al., 2015), and to lower educational outcomes (Cascio and Narayan, 2015; Rickman et al., 2017). Shale gas extraction has even been linked to seismic activity (Koster and van Ommeren, 2015). In line with hedonic models, these externalities have been found to negatively affect house prices (Muehlenbachs et al., 2015; Gibbons et al., 2016). Moreover, the analysis in this paper is limited to short-term responses to the resource boom, thus overlooking potential adjustments following a bust in the medium- and long-term.

The remainder of the paper is structured as follows. Section 4.2 conceptualises the role of social networks in the transmission of economic shocks across local markets. Section 4.3 outlines the empirical strategy adopted by this paper and presents the econometric model. Section 4.4 discusses the main results. Section 4.5 concludes.

4.2 Conceptual Framework

This section discusses a conceptual framework useful to motivate the empirical analysis of the paper, clarifying how localised shocks can diffuse in space via networks. Consider an economy organised in multiple local labour markets (regions), producing two goods. One is traded (e.g., manufactures and energy), another is not (e.g., local services). There is a fixed number of homogeneous workers in the economy, each supplying inelastically one unit of labour. In this context, labour supply to local markets is fully determined by the workplace location choice of workers. As emphasised by Allcott and Keniston (2017), geographic spillovers from fracking are a consequence of general equilibrium effects in the economy. There are two main mechanisms through which a positive shock to the local energy producing sector can diffuse to other markets. One is via multipliers in the tradable and non-tradable sectors. Another is via labour mobility.[5] These two channels are interdependent, as there might be relocation of workers both across labour markets and sectors. Nonetheless, it is useful to consider them separately, for clarity. Moreover, this

[5]Another channel could operate via redistribution of tax revenues by the producer state to different counties, although this is arguably unrelated to the network structure of the economy.

analysis is especially interested in the channel operating through worker mobility, which will be given special attention.

4.2.1 Industry Multipliers

Fracking can be thought of as a positive increase in local labour demand in the oil and gas extraction industry within the tradable sector. This shock has a direct effect on employment in the affected industry, but is also likely to increase wages in other local industries, whether tradable or non-tradable, depending on the local elasticity of labour supply (as will be explained more in detail below). Moretti (2010) discusses the impact on other tradable and non-tradable sectors. As a result of higher local wages and employment, demand for non-traded local services also increases due to higher local incomes, benefiting industries such as construction, retail, restaurants, entertainment and personal care, among others. In their study of booming resource sectors on de-industrialisation, Corden and Neary (1982) term this the 'spending effect'. Some of this additional demand results in higher wages, some other leads to expansion of the non-tradable sector, with new jobs filled by workers moving into the local market from elsewhere in the economy. Benefits are thus shared between existing workers and new ones who relocate as a consequence of the shock. Mobility in this case is key to transmission of the shock. With fixed labour in the overall economy, supply in originating markets falls, which raises wages as firms compete for a reduced pool of workers (unless the production technology allows perfect substitution with capital). With respect to tradables, the effect is ambiguous. Due to higher overall wages in the local market experiencing the shock, firms face higher production costs. This affects their competitiveness as they cannot adjust output prices, which are fixed across all regions. Some production is likely to relocate to other regions, leading to a contraction of the local tradable sector, but potentially expanding it elsewhere in the economy. Corden and Neary (1982) refer to this as the 'resource movement effect'.[6] Conversely, some local and non-local tradable industries may stand to gain due to input-output linkages and demand for intermediate goods (Hirschman, 1958). The increased production of oil and gas may require specialised inputs related to drilling, storing, and refining, for instance. This may affect employment locally, if these industries tend to cluster geographically, but can also result in job creation elsewhere in the economy. These spillovers have to do with the geography of production networks, which is not the focus of this paper. However, any additional local employment effect also diffuses to other regions by selectively attracting new workers

[6] This, they argue, combined with the 'spending effect', explains the Dutch disease phenomenon, that is, the simultaneous expansion and contraction of industries in the tradable sector, where a booming resource extraction activity is associated with a weakening manufacturing base.

depending on social networks, as emphasised in the next section. In the specific case of fracking, Fetzer (2014) also highlights how trade costs and pipeline constraints in oil and gas lead to falling local energy prices. This 'energy effect' counteracts higher labour costs, potentially sustaining an expansion, rather than contraction of the tradable sector.

4.2.2 Selective Worker Mobility

The multiplier effects discussed so far allow for indirect diffusion of localised shocks across regions, but do not clarify how diffusion relates to interaction over social networks. To this end, it is useful to consider the mobility channel more closely. The key takeaway form the previous paragraphs is that a shock to the energy extraction sector can have a knock-on effect on other sectors, whether tradable or non-tradable. The relative impact on employment and wages is then mediated by the elasticity of local labour supply, which, in the proposed setting with fixed total workforce, amounts to the ability of workers to relocate or commute across local markets. This mobility, however, is selective, so that the propensity to take on work in a particular local market is higher for some region-pairs than for others. Increasingly, spatial general equilibrium models allow for constraints in worker mobility due to frictional spatial linkages (Amior and Manning, 2018; Ahlfeldt et al., 2020). There may be differences in preferences or constraints across workers in different local labour markets influencing the mobility outcome. Two key channels come to mind when thinking about social networks: preferences for location, and job search. The former can be traced back conceptually to the work of Sjaastad (1962), who discusses the non-monetary 'psychic costs' of leaving behind family and friends (or, symmetrically, the gains from re-joining them). Moretti (2011) was perhaps the first to acknowledge this in a formal model, by introducing idiosyncratic worker attachment to places, as individuals weight-off relative preferences for location-pairs. The second channel, job search, emphasises spatial frictions in access to information. Recent contributions in this area of research include Manning and Petrongolo (2017) and Schmutz and Sidibé (2019). Conceptually, this paper focuses on the information channel associated with job search.

Who gets to hear about job opportunities in distant markets? The news might not reach evenly across regions. The role of social networks, in this interpretation, is grounded on an intuitive argument: the greater the intensity of social interaction between two places, the higher the probability that information is channelled across these markets. It is also possible to relate this statement to micro-level foundations. According to the aforementioned model by Calvó-Armengol and Jackson (2007), individuals are more likely to receive information through their network about jobs paying higher wages than their current one ('better jobs') if a larger share of their social ties connects to agents with jobs paying higher wages ('satisfied agents'). Intuitively, the more agents in the network have better jobs, the more likely they are to have first-hand information on better jobs. At the same time, since workers compete for information on better jobs, if more agents in a network are satisfied with their current job, the likelihood that the information is passed on to someone else in the network increases, eventually reaching a dissatisfied agent who can take up the better job.[7] In short, by this argument localised shocks are more likely to diffuse between places that are more strongly connected with one another socially.

As emphasised in Monte et al. (2018), the choice of workplace location in response to a localised labour demand shock can result in either permanent relocation of workers across regions (effectively migration), or simple commuting. In fact, the authors point out that the effects of a shock are heterogeneous depending on the commuting openness of the affected area, as this influences local labour supply elasticity. I therefore consider both cases. With migration, spatial diffusion of shocks operates through general equilibrium effects mediated by labour and housing supply. This adjustment is best described with the local labour market model of Moretti (2011), where a demand shock in the destination region generates a real wage change at the origin due to falling housing demand.[8] The model makes several simplifying assumptions which, however, allow to highlight the critical role played by the local elasticity of labour supply in the transmission of shocks. Social networks play a role to the extent that the likelihood of relocation between region-pairs increases with the social connectedness of these regions. In addition, one could also imagine that migrant workers send remittances to social connections back in their origin region.

With commuting, diffusion operates directly via new jobs or higher nominal wages, as workers reside close enough to the fracking site to take on new jobs without changing their place of residence. Commuting is a particularly relevant case to consider in this analysis, for two reasons. First, sociological accounts of the oil and gas industry document that employees often do not live directly by the drilling site but rather in the surrounding areas, due to negative externalities linked to drilling, as well as limited provision of services and consumption amenities where extraction takes place (Christopherson and Rightor, 2012). Second, most jobs generated by fracking tend to be relatively short-lived, mainly occurring in relation to the set-up of the drilling site. As a result, employees are frequently out-of-town hires: transient workers active on several sites across vast regions, temporarily living

[7] I refer to the original paper for analytical derivations of these findings.

[8] In his framework, perfect substitution between capital and labour means that nominal wages do not adjust to the outflow of workers. Thus, gains accrue solely via real wages due to falling house prices. Introducing imperfect substitution in the production technology, however, would allow for gains in nominal wages too.

in purposely arranged caravan camps while maintaining their permanent residence in a different state (Jacquet, 2011; Christopherson and Rightor, 2012). Workers effectively act as if they were commuting over long distances for as long as they are needed to fulfil the job. They do not change their permanent address, but travel across the entire economy depending on availability of jobs in the industry. Under these conditions, the nominal wages gained by the workers leave the host community and are recorded in places potentially kilometres away from the drilling site. While some of these gains may be spent locally around the wells, most of the money is likely to be used elsewhere. Finally, commuting is also relevant from an empirical viewpoint. The geographical units of analysis in this paper are US counties, which do not represent self-contained labour markets. Conceptually discussing commuting thus allows to remain a priori agnostic regarding the definition of the catchment area of local labour markets.

4.2.3 Commuting with Social Connections

What follows formalises the intuition about selective mobility in the spirit of Ahlfeldt et al. (2015), focusing on commuting. As discussed above, sociological accounts of shale-gas workers suggest that this adjustment channel should prevail. A theory for the spatial diffusion of fracking shocks over social networks cannot abstract from what is known about industry practices. Analytically, the temporary long haul relocation of workers who do not change their original place of residence can indeed be thought of as analogous to commuting. In their quantitative spatial model, Ahlfeldt et al. (2015) provide a useful way to think structurally about the determinants of commuting flows in a gravity form. The commuting part of the model can be adapted to the context at hand by introducing a social connectedness term that counterweights the effect of geographical distance in determining commuting probabilities, where the act of commuting is interpreted in a broad way, to encompass the case of transient workers who do not change their place of residence.[9]

Consider an economy divided into $i = 1, ..., S$ discrete locations (regions). Each location offers a fixed amount of land, available for residential or commercial use. Land income is earned by absentee landlords and spent outside the economy. As before, workers are homogeneous and mobile, inelastically supplying one unit of labour. They choose residence i and workplace j pairs that maximise their utility. For simplicity, imagine there is now

[9]What follows provides a synthetic description of the model for illustrative purposes, which is also somewhat simplified. A comprehensive discussion of the model falls beyond the scope of this analysis which is by and large empirical. Please refer to the paper by Ahlfeldt et al. (2015) and companion supplementary materials for a detailed description and complete analytical derivation.

only one industry. Firms produce a single final good, traded at unit price. Indirect utility for worker o living in i and commuting to j is given by:

$$v_{ij,o} = \frac{z_{ij,o} B_i w_j Q_i^{\beta-1}}{d_{ij}} \tag{4.1}$$

Where B_i and Q_i are residential amenities and cost of land consumption, w_j are wages paid at the workplace, d_{ij} are commuting costs, and $z_{ij,o}$ is an idiosyncratic preference term specific to each worker that depends on residential and workplace location. The disutility from commuting is modelled as an iceberg cost $d_{ij} = e^{\kappa\tau_{ij} - \eta\sigma_{ij}} \in [1, \infty)$ which increases in geographical distance between place of work and residence, τ_{ij}, but decreases in the degree of social connectedness between the two, σ_{ij}, with strengths of κ and η respectively. When thinking about transient workers in the fracking industry, this cost can be interpreted as the overall decrease in utility from distance to home arising, for instance, due to less effective job search. The idiosyncratic preference term $z_{ij,o}$ captures heterogeneity in individual preferences for places of work and residence, and is drawn from an independent Fréchet distribution:

$$F(z_{ij,o}) = e^{-T_i E_j z_{ij,o}^{-\epsilon}}, \qquad T_i, E_j > 0, \epsilon > 1 \tag{4.2}$$

Where T_i is a scale parameter that determines the utility that the average worker derives from living in region i, E_j captures the average utility from working in region j, and ϵ is a shape parameter that describes the dispersion of idiosyncratic preferences across workers.

Because indirect utility increases monotonically in the idiosyncratic term $z_{ij,o}$, which follows a Fréchet distribution, indirect utility for any worker living in region i and working in j is also Fréchet distributed. In equilibrium, workers choose to live and work in a location pair ij such that their utility is maximised, taking into account commuting costs. Ahlfeldt et al. (2015) show that, as the maximum of Fréchet distributed variables also follows a Fréchet distribution, the probability that a worker commutes from i to j is given by:

$$\pi_{ij} = \frac{T_i E_j (d_{ij} Q_i^{1-\beta})^{-\epsilon} (B_i w_j)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s (d_{rs} Q_r^{1-\beta})^{-\epsilon} (B_r w_s)^{\epsilon}} \tag{4.3}$$

Other things equal, individuals prefer living in regions with higher amenities B_i (e.g., not living in close proximity to the wells), low cost of land Q_i, and higher average idiosyncratic utility T_i. Similarly, they privilege regions with higher wages w_j and average idiosyncratic utility E_j as a workplace. Moreover, by conditioning 4.3 on place of residence, it is possible to obtain the probability of commuting to j for a worker living in i, where all terms indexed with i are fixed:

$$\pi_{ij|i} = \frac{E_j(w_j/d_{ij})^{\epsilon}}{\sum_{s=1}^{S} E_s(w_s/d_{is})^{\epsilon}} \tag{4.4}$$

This highlights that workers are more likely to commute to regions where they can earn higher wages and draw higher average utility relative to those in all other workplace locations s. It also shows that the probability of working in j decreases in the bilateral resistance term d_{ij}, relative to that across all possible locations d_{is} (multilateral resistance). As a result, the income a worker living in i can expect to earn is given by the expression:

$$\mathbb{E}[w_j|i] = \sum_{j=1}^{S} \pi_{ij|i} w_j \tag{4.5}$$

Whereby the expected wage for an individual residing in i reflects the weighted average of wages that can be earned across all workplace locations j that can be accessed from the place of residence, with weights proportional to a measure of distance that takes into account commuting costs. Note that the expression in (4.4) implies a semi-log gravity commuting equation:

$$\ln \pi_{ij} = -\delta\tau_{ij} + \gamma\sigma_{ij} + \varsigma_j, \qquad \delta = \epsilon\kappa, \gamma = \epsilon\eta \tag{4.6}$$

Where the log probability of commuting between i and j decreases in geographical distance τ_{ij} with strength δ, and increases in social connectedness σ_{ij} with strength γ. Workplace characteristics are absorbed by the fixed-effect ς_j. This highlights the dependence of expected wages earned by living in i on the geographical distance and social connectedness with workplace location.

These last two equations are also helpful in that they provide a link between this conceptual discussion and the applied analysis of this paper. An empirical counterpart to (4.5) consistent with the relationship highlighted in (4.6) expresses wages observed in region i as a weighted average of wages in all other connected locations:

$$\Delta w_{i,t} = \gamma \times m(\Delta w, s)_{i,t} + \epsilon_{i,t} \tag{4.7}$$

Where $m(\Delta w, s)_{i,t}$ is a function determining 'spatial' averages, considering geographical or social distance. As we observe multiple realisation of wages over time in the data, Equation (4.7) is indexed with t for each year, and expressed in first differences to account for time-invariant unobservables. Moreover, acknowledging the above-mentioned challenges associated with estimation of a SAR model of this kind (Gibbons and Overman, 2012; Gibbons et al., 2015), spatially lagged Δw can be replaced with plausibly exogenous characteristics of each region that correlate with wages, such as fracking shocks. The resulting reduced-form SLX model provides the workhorse empirical specification used in this analysis. The next section discusses more in detail the empirical methods used in this paper to identify the role of social networks in the transmission of localised economic shocks in space. The methods were designed taking into account the conceptual intuitions developed in the above paragraphs.

4.3 Data and Empirical Methods

4.3.1 Variables Definition and Measurement

This paper relies on two main sources of data. To capture social networks, it uses a newly released measure of social connectedness that draws on information on the universe of online friendship links on Facebook, a popular social media site. On the other hand, the paper uses data from Feyrer et al. (2017) to measure fracking shocks and labour market outcomes.[10] What follows gives details on the original sources and definitions of all variables. The geographical units of analysis used throughout the paper are counties located in the contiguous US, observed yearly between 2004 and 2012.

Labour Market Outcomes

Labour market outcomes are measured for all US counties in the sample using information from two sources: the Quarterly Census of Employment and Wages (QCEW) by the Bureau of Labor Statistics (BLS), and the Adjusted Gross Income (AGI) statistics of the Internal Revenue Service (IRS). The former has the advantage of providing information disaggregated to the level of six NAICS industries.[11] The latter gives information on wages and salaries (of main interest in this paper), but also includes data on other sources of income such as rents, royalties, and other non-wage business revenues. Importantly for this analysis, the data are collected in different ways. The BLS data are reported by employers at their location, and therefore accurately describe economic activity where it takes place. The IRS data, on the other hand, are based on declarations filed by employees at their address of permanent residence, thus giving information on money earned (and likely spent)

[10] The data are available at this link: `https://www.aeaweb.org/articles?id=10.1257/aer.20151326`

[11] These are: natural resources and mining; transportation, trade and utilities; construction; manufacturing; education and health services; government (local, state and federal levels).

where people live. The two should be the same to the extent that people live and work in the same county, but can differ in case of commuting or temporary relocation across county borders. In other words, the income of a worker living in county i and working in county j will be allocated to i by IRS data and to county j by BLS data. The IRS outcomes are thus more likely to pick up any effect that might be observed on commuting and transient workers, who indeed appear to represent the bulk of earners in the industry.

Local Economic Shocks

To measure local economic shocks, Feyrer et al. (2017) compile a new dataset using information obtained from Enverus (formerly Drillinginfo), a company that systematically gathers data on the oil and gas industry. For each county, the authors isolate wells that began producing in any given year, and compute the total value of new production in that year as the quantity of oil and gas produced by those wells, times its market value (using EIA prices). All figures are then deflated to 2014 USD using the CPI and scaled by the one-year lagged value of county employment, to ensure the measure is comparable across differently sized counties. The resulting measure of local economic shocks from fracking is thus the per capita value of new oil and gas production in any given year, or more formally, for each county i in year t:[12]

$$\Delta X_{i,t} = \frac{\Delta Q_{i,t}^{oil} \times P_{i,t}^{oil} + \Delta Q_{i,t}^{gas} \times P_{i,t}^{gas}}{L_{i,t-1}} \tag{4.8}$$

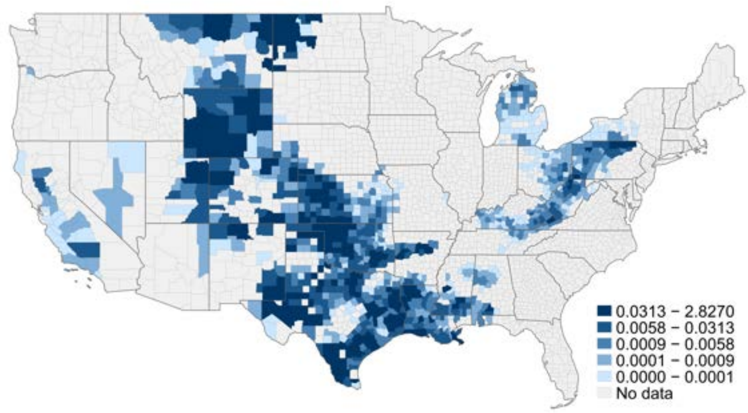
In line with Feyrer et al. (2017), the estimating dataset excludes the smallest two percent of counties in the sample, as these represent outliers especially when expressed in per capita terms. We refer to the original paper for any further detail on these data. Appendix Table 4.B.1 ranks the top 20 US states in terms of new per capita production over 2005-2012, along with average yearly changes in employment and wages using BLS data.[13] The five states experiencing the largest shocks were North Dakota, Wyoming, New Mexico, Oklahoma, and Texas. To give a more detailed overview of the spatial distribution of these shocks, Figure 4.1 maps quintiles of the total value of new production of oil and gas per capita over the 2005-2012 period.

Social Network Matrices and Socially Lagged Shocks

The proposed measure of social networks, or social connectedness, relies on an index developed by Bailey et al. (2018b): the Social Connectedness Index (SCI). This index essen-

[12] Per capita and per worker are used interchangeably in what follows.
[13] Appendix 4.A includes all additional figures, Appendix 4.B includes all additional tables.

Figure 4.1: Total value of new production per capita in 2005-2012 (in millions)



tially captures the social graph for the universe of *active* US Facebook users as of April 2016, aggregated up to the level of counties.[14] Users are deemed active if they interacted with Facebook in the 30 days prior to the April 2016 snapshot. Geographic location is assigned using the IP address from which users login most frequently. For all users m and n and for each pair of counties i and j, the index is constructed as:

$$SCI_{ij} = \sum_{m \neq n} \sum_{n} \mathbb{1}_{mn}, \text{ for } m \in i \text{ and } n \in j \tag{4.9}$$

Where $\mathbb{1}_{ij}$ is an indicator variable that takes the value of 1 if two users are friends with each other, and 0 otherwise. Due to confidentiality concerns, Facebook only releases a re-scaled version of these data. The index thus ranges between 0 and 1,000,000, the highest observed value, which is assigned to Los Angeles County to Los Angeles County connections. The result is a weighted social graph consisting of 3,136 nodes and 9,462,485 edges. Despite some limitations in terms of user representativeness, the SCI can be thought of as one of the most comprehensive measures of revealed social interaction available to date for the entire US geography. At the time the data were extracted, there were over 220 million active monthly Facebook users in the United States and Canada.[15] Moreover, concerns

[14] In principle it would be more accurate to refer to Facebook *accounts* rather than *users*. However, the same expression as in Bailey et al. (2018b) is used here for consistency.

[15] Information obtained from Facebook's 2016 quarterly results report, retrieved at: `https://s21.q4cdn.com/399680738/files/doc_presentations/FB-Q4'16-Earnings-Slides.pdf`. Unfortunately, Facebook would not release covariates for these data. However, it is possible to gauge some descriptive facts from secondary sources. A Pew Research Center study published in that same year estimates that about 70% of US adults (aged 18 or more) used the social media platform (Greenwood et al., 2016). Women, younger individuals (aged 50 or less), college educated and relatively poorer adults were slightly overrepresented, albeit by small margins. Most Facebook friendships are with people with whom users have ongoing interaction in real life. According to Hampton et al. (2011), ties between Facebook users tend to occur among high school or college peers (31%), immediate or extended family members (20%), co-workers (10%), and neighbours or acquaintances (9%). The remaining ties are with friends of friends, or 'dormant relationships', that may become useful to users in the future. However, only 3% of Facebook friendships

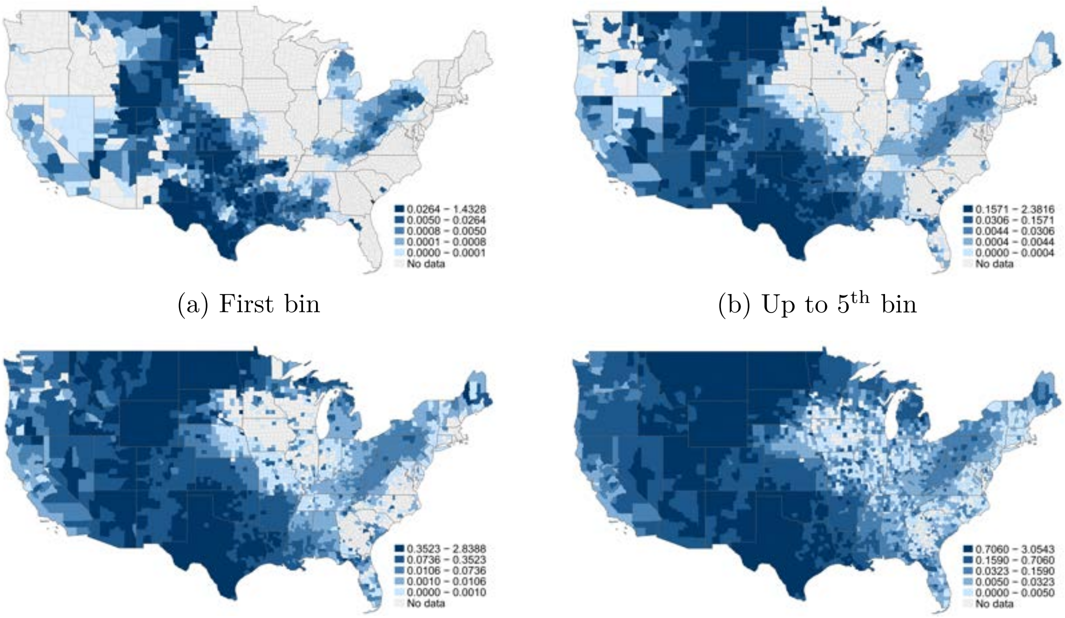
about possible bias introduced into the present analysis due to erroneous measurement should be minor unless there are reasons to believe that mismeasurement is systematic and correlated with the outcome of interest.

As will be discussed in Section 4.3.2, the empirical analysis considers the impact that shocks occurring in one place have on counties that are socially connected to this place, at varying degrees of (social) distance. To obtain matrices of social weights suitable for this analysis, I proceed as follows. First, the SCI is normalised by the product of each county's population for all pairs. This corrects for the fact that larger counties mechanically share more friendship links, giving a measure that is comparable across county-pairs. It can be thought of as the relative probability of friendship between any two counties, henceforth $RSCI$ (Bailey et al., 2018b). For simplicity, in what follows I still refer to this normalised measure as social connectedness, despite the transformation. Second, for every county the resulting distribution of connectedness to all other counties is discretised into 20 bins of five nearest *social* neighbours each. Connections ranking below the 100$^{\text{th}}$ neighbour in terms of strength are discarded, assuming there is a steep decay in network effects. This assumption can be directly tested in the data and appears to be valid, as will be shown. I thus obtain 20 matrices G_d (one for each bin), where each element g_{ij} takes the value of 1 if county i is socially connected to county j at distance-bin d, and 0 otherwise. A comparable set of matrices W_d based on the first 20 bins of five nearest neighbours in terms of *geographical* distance is also produced. The map in Figure 4.A.1 in Appendix shows these bins for the top five largest counties in terms of new oil and gas production over the 2005-2012 period. Counties are coloured in progressively lighter shades as the geographical or social distance of each bin increases (unit increases from 1 to 20).

Finally, for each county and year, I create lagged measures of fracking shocks $G_{d,i}\Delta X_t$ in the social space, computed as the total new production per capita occurring in each bin of five nearest social neighbours respectively. To this end, I sum up new production for each bin, and divide this by the total number of workers in the same bin during the previous year. A measure $W_{d,i}\Delta X_t$ of spatially lagged shocks is also obtained using the same method, based on geographical nearest neighbours. Figure 4.2 visualises the cumulative total value of socially lagged fracking shocks in US counties up to selected distance-bins. For each county, the choropleth maps show the total per capita value of new production taking place in its social neighbours over 2005-2012, with darker polygons corresponding to higher

are with someone the user has never met in person. Moreover, several studies have shown that Facebook ties are good predictors of real life friendships and friendship strength (Gilbert and Karahalios, 2009; Jones et al., 2013). All this suggest that there is strong potential in these data to be used to study social relationships on a large-scale (Bailey et al., 2018b).

Figure 4.2: Cumulative socially lagged new production in 2005-2012



(a) First bin

(b) Up to 5$^{\text{th}}$ bin

(c) Up to 10$^{\text{th}}$ bin

(d) Up to 20$^{\text{th}}$ bin

quintiles of the distribution. It is noteworthy that, by and large, socially weighted shocks are greatest in counties in close geographical proximity to where the extraction takes place, and display distinct decay patterns over space. This is due to positive correlation between social and spatial distances: individuals are more likely to become friends and interact with peers living close to them (Bailey et al., 2018b). Moreover, it is also interesting to notice that by the 20$^{\text{th}}$ bin, nearly every county in the contiguous US is exposed to fracking shocks through one of its social neighbours.

4.3.2 Identification Strategy

Baseline Specification

This paper is interested in estimating the inward effects on wages and employment in county i of new oil and gas production taking place in socially connected counties, conditional on energy production in i itself. Additionally, due to the spatial clustering of drilling sites, it is also important to consider inward effects from counties located in close geographic proximity to i, which could potentially bias results upwards if not accounted for (James and Smith, 2020). To this end, and consistent with what discussed in Section 4.2, the following empirical model in first differences is estimated using ordinary least squares (OLS):

$$\Delta Y_{i,t} = \beta \times \Delta X_{i,t} + \sum_{d=1}^{20} \gamma_d \times G_{d,i}\Delta X_t + \sum_{d=1}^{20} \delta_d \times W_{d,i}\Delta X_t + \theta_t + \epsilon_{i,t} \tag{4.10}$$

Where $\Delta Y_{i,t}$ denotes the change in income or employment per capita in county i in year t, $\Delta X_{i,t}$ is the value of new production per capita in the county itself, $G_{d,i}\Delta X_t$ is the total value of new production in the network of county i for twenty bins of five nearest social neighbours each, and $W_{d,i}\Delta X_t$ is a comparable measure computed over 20 concentric doughnuts of five nearest geographical neighbours each. Additionally, the model includes year dummies θ_t to account for general time trends. Robustness checks also include county fixed effects α_i, which renders parameters in Equation 4.10 comparable to difference-in-difference estimators. Finally, not explicitly mentioned in the model are also one-year lags of all new production variables (in the county itself, as well as in socially- and spatially-lagged counties), to account for possible dynamic effects of fracking shocks, whereby past production may continue to affect outcomes in subsequent years. This is in line with Feyrer et al. (2017). The error term $\epsilon_{i,t}$ is heteroscedasticity-robust and clustered by spatial bins (whether geographical or social, depending on the case at hand). Standard errors are adjusted using the approach discussed in Colella et al. (2019) to obtain cluster-robust inference in the presence of unobserved dependence of $\epsilon_{i,t}$ in the geographical or social space. This is implemented in Stata using the package `acreg`. The adjustment is akin to that in Conley (1999) but allows greater flexibility in the definition of the distance metric. I set the distance in terms of nearest neighbours bins, with a cut-off threshold at 10, the 50$^{\text{th}}$ neighbour, also allowing for a decay structure in the cross-sectional dependence using a linearly decreasing Bartlett kernel as distance increases (similar to Newey and West, 1987).

The main parameters of interest are captured by the vector γ_d. Controlling for county i's own production and for production in i's 100 nearest geographical neighbours, γ_1 gives the inward effect on outcomes in i of production in the five counties i is most strongly socially connected to, net of inward effects from other socially connected counties up to the 100$^{\text{th}}$ social neighbour. Similarly, γ_2 estimates this effect for the next five most strongly socially connected counties, γ_3 considers the ones after that, and so forth. This set-up allows to study how far in the social network fracking shocks are felt. The effects are expected to be strongest among the nearest social neighbours (the socially closest counties), and decay rapidly as one moves out in the network. Note that geographically and socially neighbouring counties are likely to overlap due to the tendency to interact over close physical distances noted earlier. As a result, jointly estimating parameters on both social and geographical lags of fracking shocks is likely to yield biased results. In baseline specifications, therefore, the model in Equation 4.10 is estimated separately for geographical and social lags, respectively constraining either γ_d or δ_d to zero. Next, I address some further concerns with respect to this baseline specification and describe the proposed solutions.

Endogenous Network Formation

Social networks form endogenously as a result of several unobserved factors. Bias is introduced in the proposed estimating equation if these factors correlate with the outcome of interest. Two main concerns stand out. First, as already mentioned above, geographical and social neighbours are likely to overlap due to the fact that people are more likely to interact when they live close to each other. In other words, it is hard to separately estimate the effect of geographical and social proximity to the extent that the two are co-determined. Second, there could be reverse causality whereby fracking shocks determine the observed network by creating incentives for workers to relocate or commute between counties, rather than the other way around. The latter concern is particularly severe considered that the social connectedness data has no time dimension. The SCI gives a snapshot of networks connecting counties in 2016 only, which is posterior to the period under analysis. In practice, for large enough counties, it seems unlikely that the aggregate ties of all residents would be affected by the mobility of workers in one particular industry in any sensible way, unless local multipliers are strong enough to generate a sizeable migration and commuting response across other industries. In this case, the identifying assumption is that social connectedness represents a structural, slow-changing, feature of places determined over the long term and unaffected by the mobility of few workers over a relatively short time period. For smaller counties, however, this assumption is likely to fail. Reassuringly, however, as mentioned above, the smallest two percent of counties is dropped from the estimating sample, which further mitigates this concern. In addition, I address concerns about reverse causality and geographical distance by creating a new measure of social connectedness that partials-out bilateral migration and distance between counties. In particular, social connectedness between counties i and j, or better, the relative probability of friendship between the two, can be represented analytically by the following relationship:

$$RSCI_{ij} = f(d_{ij}, M_{ij}, \upsilon_{ij}) \tag{4.11}$$

Where d_{ij} denotes the geographical distance separating i and j (due to the cost of interacting over space), M_{ij} captures cumulative mobility between i and j, and υ_{ij} is a bilateral residual term for each place-pair combination. Assuming this relationship is log-linear, I estimate the following empirical model:

$$\ln RSCI_{ij} = \beta \times \ln d_{ij} + \gamma \times \ln(M_{ij} + 1) + \upsilon_{ij} \tag{4.12}$$

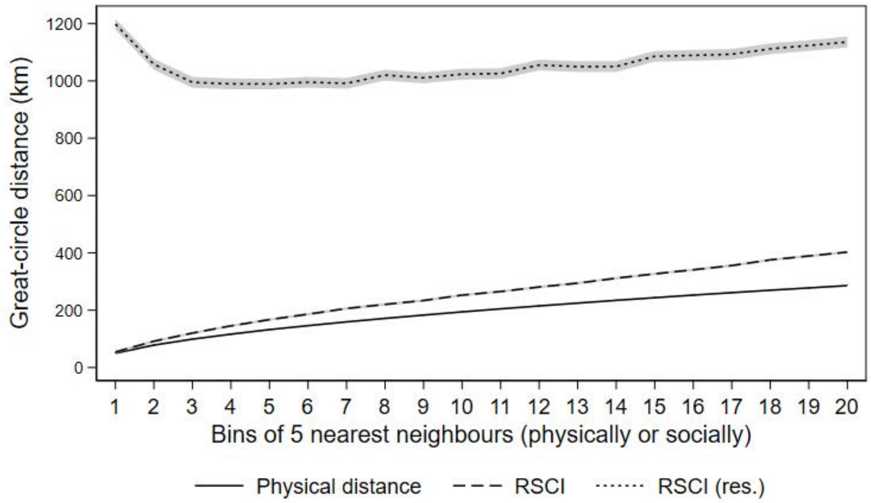
Where M_{ij} captures cumulative gross migration flows between all county pairs in the 2002-2016 period. This variable is constructed using counts of yearly county-to-county migration flows, obtained from the IRS Statistics of Income Division (SOI).[16] I predict the residuals $\hat{\upsilon}_{ij}$ and use those to create alternative matrices G_d^{res}, discretising the distribution of social connectedness captured by $\hat{\upsilon}_{ij}$ the same way outlined in Section 4.3.1. The third column in the set of maps in Appendix Figure 4.A.1 shows the resulting bins. I then compute alternative measures of socially lagged shocks based on these matrices. Appendix Figure 4.A.3 maps them. Note how there is no clearly emerging spatial pattern over the different distance bins. These measures are indeed based on a set of connectedness matrices that do not depend on physical distance or past mobility, thus further mitigating the endogeneity concerns expressed above, and allowing to jointly estimate γ_d and δ_d in one model.

Figure 4.3 shows the average value across all counties of the median geographical distance of each county from each bin, for the three set of matrices discussed above. By construction, average geographical distance increases monotonically as bins of neighbours farther away in space are considered. Interestingly, the same is true when bins formed using the plain measure of social connectedness ($RSCI$) are considered. Note that this does not need to be the case by construction, but is due to the aforementioned relationship between likelihood of interaction and physical distance. Despite this, it appears that neighbours in the social space are systematically farther away in a geographical sense than physical neighbours are. This is evidenced by the fact that the 99 percent confidence intervals drawn around mean values in each bin are non-overlapping. Finally, observe how average geographical distances for bins formed using the partialled-out measure of social connectedness are much greater than those in both other measures. On average, the median social neighbour in the first bin of each county is nearly 1,200 kilometres apart geographically from that county.

In terms of interpretation, the residual term $\hat{\upsilon}_{ij}$ can be thought of in a broad sense as anything, net of past migration, that supports interaction over physical distance. Examples

[16] These data provide one of the most detailed sources of information on migration at this level, based on address changes in the records of all individual income tax forms filed between 1990 and today. For more information, see: `https://www.irs.gov/uac/soi-tax-stats-migration-data`

Figure 4.3: Average across all counties of the median distance in each bin



Predicted values of the median distance between county i and all counties j in each bin, across all counties. Grey-shaded areas report 99 percent confidence intervals around the predicted mean. The t-statistic of the difference in means test between the first bin for physical distance and RSCI is 7.39.

include accessibility and transportation networks, business and professional collaborations, as well as knowledge networks and socio-cultural ties. Irrespective of endogeneity concerns, whilst there is an interest in studying the overall role of social connectedness in and of itself, it is especially relevant to policy makers to know whether the residual term $\hat{\upsilon}_{ij}$ matters above and beyond physical distance and past migration, since the former is at least partly amenable to policy intervention (e.g., via improvements in transport infrastructure).

Endogenous Production and Instrumental Variable

A final concern with the baseline model is that new production of oil and gas might be endogenous. As pointed out by Feyrer et al. (2017), production is a function of two factors. On the one hand, it requires the presence of oil and gas deposits, or plays, which are exogenously determined by geology. On the other, exploitation of available resources depends on the endogenous decision of mining companies to invest in extraction activities. Endogeneity is linked to the fact that firms might prioritise sparsely populated areas or high unemployment areas due to cost saving considerations. In the former case, the firm can save on land leases and royalties. In the latter, it can pay relatively lower nominal wages to local workers. Firms might also try to avoid regulatory responses from local policy makers in more populated areas. Moreover, the timing of extraction can depend on international fluctuations in oil and gas prices.

The use of time dummies addresses concerns related to changing prices for oil and gas, while estimating the baseline model in first differences mitigates issues related to prioritisation of certain counties over others, assuming the drivers of this decision are fixed. Explicitly introducing county fixed effects further addresses this issue. In addition, I follow Feyrer et al. (2017) and instrument production as a function of county and play-year fixed effects. The predicted per capita value of new production for every county and year is obtained in two steps. First, the following equation is estimated:

$$\ln(\Delta Q^{oil}_{i,t} \times P^{oil}_{i,t} + \Delta Q^{gas}_{i,t} \times P^{gas}_{i,t} + 1) = \alpha_i + \pi_{p,t} + \epsilon_{i,t} \tag{4.13}$$

Whereby the total value of new production is obtained as a combination of time variant, play-specific, technological shocks, and a county's time invariant characteristics (e.g., its area). Expressing the outcome in logs allows for non-linearities in this relationship. Second, I obtain predictions for total new production in every county and normalise this by lagged employment:

$$\Delta Z_{i,t} = \frac{\exp(\hat{\alpha}_i + \hat{\pi}_{p,t}) - 1}{L_{i,t}} \tag{4.14}$$

The validity of this instrument, which mimics a traditional shift-share measure, relies on the identifying assumption that a county's production is a sufficiently small share of the overall production of the play in each year, which can thus be considered exogenous. I rely on the same definitions of plays used by Feyrer et al. (2017), who in some instances combine small plays into larger groups in support of instrument validity. Figure 4.A.2 in Appendix maps these plays. Note that, for every bin defined by the matrices G_d, G^{res}_d and W_d, I construct equivalent measures by aggregating the predicted value of new production and employment within each bin d, then dividing the former by the latter.

Finally, I estimate the following empirical model using two stage least squares (2SLS), where the estimates of new production $\Delta Z_{i,t}$ (and equivalent lagged ones) are used as instruments in the first stage to predict observed new production:

$$\Delta Y_{i,t} = \beta \times \Delta \hat{X}_{i,t} + \sum_{d=1}^{20} \gamma_d \times G^{res}_{d,i} \Delta \hat{X}_t + \sum_{d=1}^{20} \delta_d \times W_{d,i} \Delta \hat{X}_t + \theta_t + \epsilon_{i,t} \tag{4.15}$$

Note that the set of matrices G^{res}_d is used, which allows to estimate the inward effects on i of new production in counties socially connected to i, while also controlling for shocks in geographically neighbouring places. This model, either estimated with OLS using observed new production or with 2SLS using the above described instruments, represents the preferred specification for most results presented in this paper. Because implementing the

Colella et al. (2019) standard error correction is computationally very demanding, I cluster residuals by commuting zones (Tolbert and Sizer, 1996) in 2SLS estimates. However, reduced form estimates for 2SLS regressions are also provided, where standard errors are again corrected for spatial clusters. Next, I discuss my findings. Table 4.B.2 in Appendix gives summary statistics for all the main variables used in the analysis.

4.4 Results and Discussion

This section summarises the key results of this paper. Due to the large number of coefficients, each with a similar interpretation, the findings are best reported graphically rather than with traditional regression tables. I thus present coefficient plots summarising the magnitude of the estimated effects γ_d (on the first vertical axis) for different bins of nearest neighbours (on the horizontal axis).[17] This allows to visualise how the average effect of fracking shocks in a county's social network changes as one considers progressively farther away neighbours. Grey areas denote 90, 95 and 99 percent confidence intervals, respectively in lighter shades. In the same diagram, I also overlay the average kilometre distance of neighbouring counties in each bin (measured on the second vertical axis and displayed in light gray). This allows to intuitively grasp how far away geographically fracking shocks disperse over social networks. I report findings that compare the strength of diffusion using the plain measure of social connectedness ($RSCI$) and the one obtained by partialling-out physical distance and migration.[18] Appendix 4.A provides similar coefficient plots capturing the effect of shocks occurring in neighbours in terms of geographical distance (δ_d) as well as in terms of an additional measure of social connectedness that considers nearest neighbours using the $RSCI$, but forcing neighbours to be at least 200 kilometres apart geographically. The latter is included for robustness. I also report 2SLS along with reduced form estimates. Tables for all underlying regressions, including 2SLS first stages, are reported in Appendix 4.B.

4.4.1 Effects on Wages

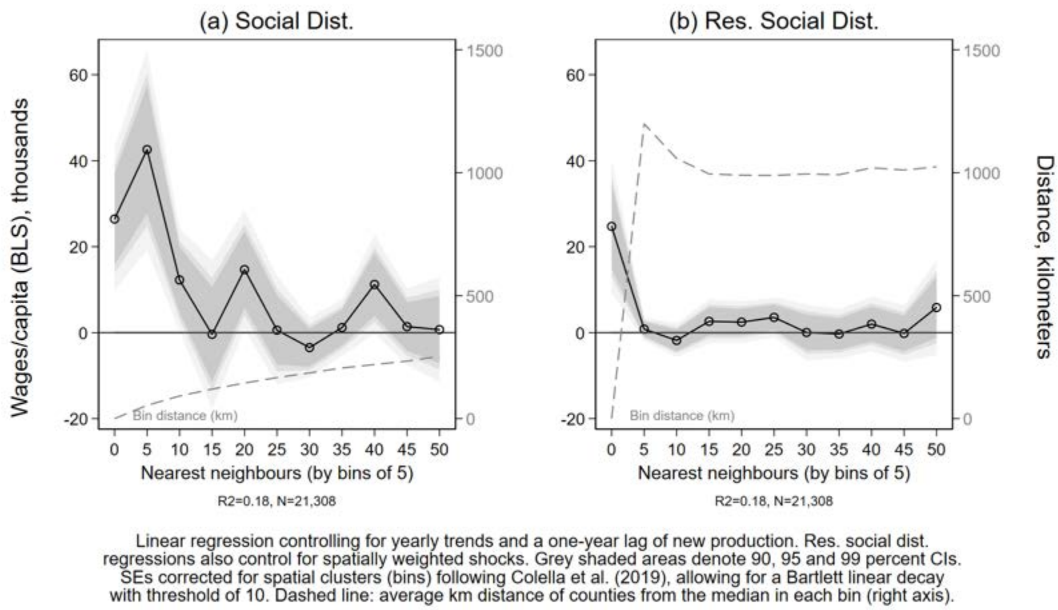
I begin by showing effects on wages. The conceptual framework predicts that in response to positive shocks to a social neighbour, wages should increase by a larger amount the more socially connected two counties are. In this specific application, a fracking-related shock

[17] The empirical estimates always include the full set of 20 distance-bins, although only the first 10 are reported (that is, up to the 50^{th} neighbour), since coefficients are mostly insignificant after that.

[18] Since the graphs are read left-to-right, the horizontal axis is more easily interpreted as capturing distance in social networks rather than proximity/connectedness. I therefore title each graph as 'Social Distance' and 'Residuals of Social Distance', respectively.

to a county with which many friendship connections are shared should matter more for local outcomes than one taking place in a more 'socially distant' county. Figure 4.1 reports OLS estimates of the effects for BLS wages, that is to say, wages reported by employers at their location. Changes in this outcome should reflect direct gains by workers in the industry itself (or employed in activities immediately tied to it, such as transportation), to the extent that they live close enough to the drilling site. They can also reflect gains made by workers elsewhere in different industries due to local multiplier effects and input-output relationships (e.g., higher wages gained in non-tradable services due to higher local demand). Table 4.B.3 in Appendix reports exact estimates for all coefficients.

Figure 4.1: Coefficients plot for wages (BLS) using OLS (main)

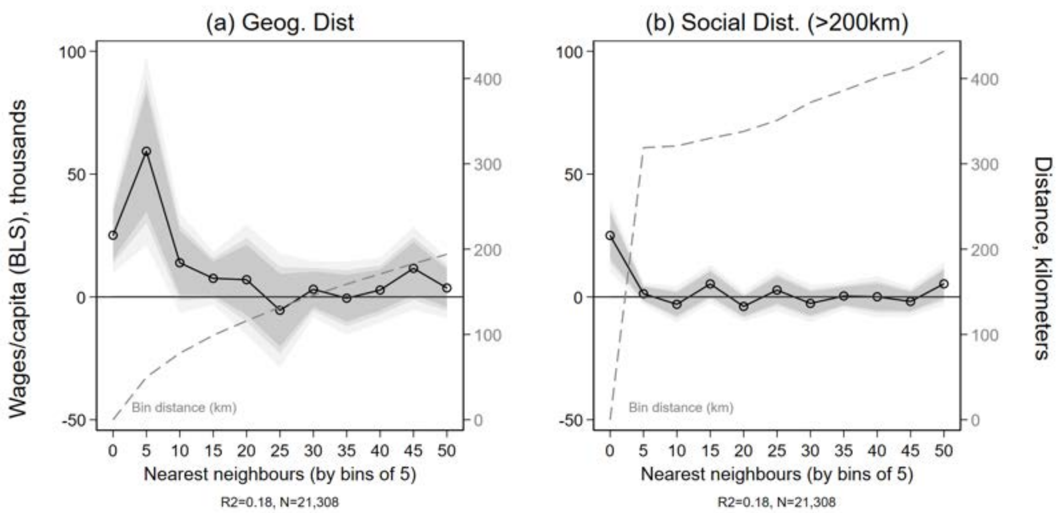


In both panels, a one million dollar increase in new production in oil and gas per capita is associated with an increase of wages per capita of about 25,000 dollars in the county itself.[19] Panel (a) further shows that, controlling for own production and inward effects from other social neighbours, a marginal increase in production taking place in the first five nearest social neighbours increases wages by as much as 42,000 dollars per capita, while new production in the next five neighbours raises wages in the socially connected county by just under 12,000 dollars per capita. Effects decay rapidly after that and converge towards zero. When the plain measure of social connectedness is used, therefore, it appears that shocks diffuse in space up to about 100 kilometres away. To what extent is this an effect

[19]This estimate is lower but comparable in magnitude to that of Feyrer et al. (2017), who give a point estimate of about 34,000 dollars.

specific to interaction via networks, as opposed to simple geographic proximity? Panel (b) in the same figure suggests that geography is by and large the main reason for this. None of the socially lagged shocks appear significant in this model. This can also be confirmed by looking directly at the impact of *spatially* lagged shocks, shown in panel (a) of Figure 4.2. The plot shows that effects are much larger for new production taking place in the first five nearest geographical neighbours, up to about 60,000 dollars per capita. Interestingly, however, effects decay much more rapidly after that for geographical distance than for social distance, and are only marginally significant at the 10 percent level. This would suggest that looking at social networks can provide a more accurate representation of economic interaction than simple geographical distance, especially as one considers relationships over progressively more (geographically) distant places.

Figure 4.2: Coefficients plot for wages (BLS) using OLS (additional)

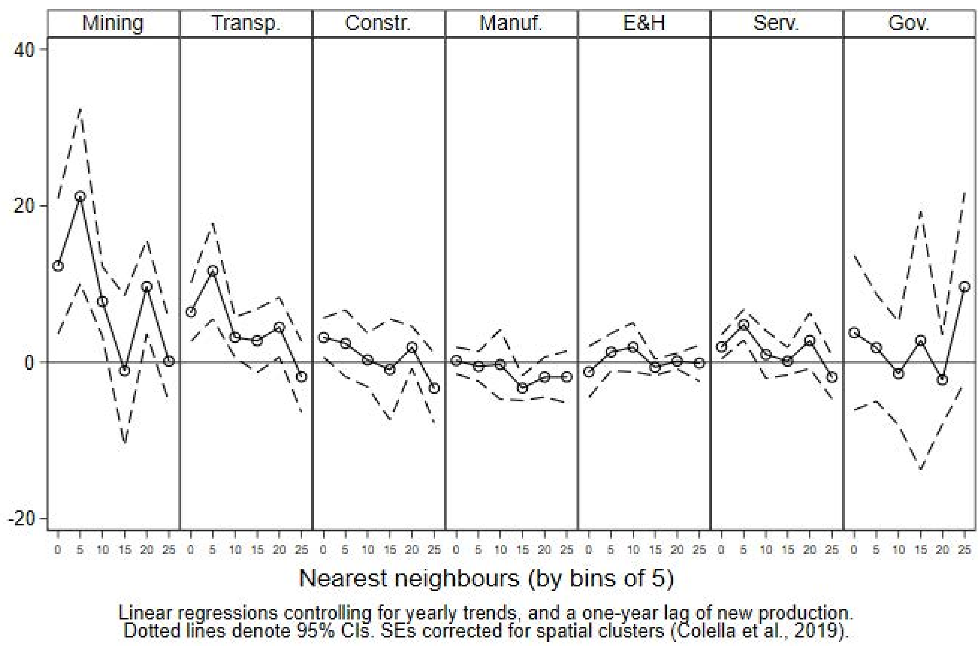


Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Introducing county fixed effects to the BLS wage regressions leaves the results virtually unchanged (Appendix Figure 4.A.4), except for a small effect of about 2,000 dollar per capita associated to new production in the five most socially connected counties (panel b, partialled-out measure). Finally, considering instrumental variable estimates also confirms these findings, although the point estimates are somewhat larger (Appendix Figure 4.A.6).

As mentioned, shocks diffusing from socially connected places can be felt beyond the mining and extraction industry itself due to input-output relationships and local multipliers. To gauge which industries are more likely to benefit from the effects described above, Figure 4.3 offers a sector breakdown of the OLS estimates obtained using the simple social connectedness measure.[20] The diagram shows that, in relative terms, the largest surge in wages is observed in mining activities and extraction activities, followed by transportation. In addition, it appears that services benefit somewhat from fracking shocks, although only in the most closely socially connected counties. This aligns to previous findings in the literature and with what discussed in Section 4.2.1. A corresponding set of results for geographical distance is available in Appendix (Figure 4.A.10).

Figure 4.3: Coefficients plot for wages (BLS) by industry using OLS (soc. dist)

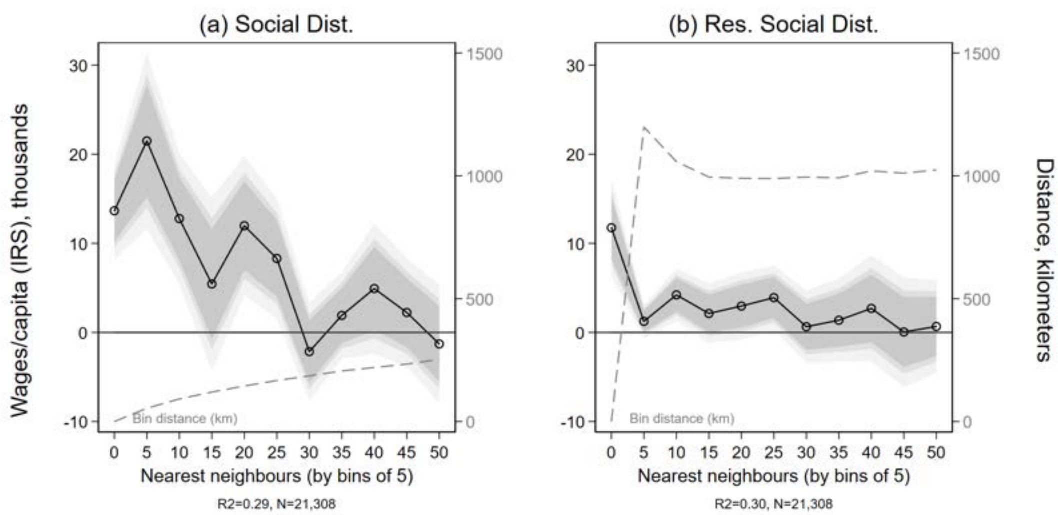


In sum, it seems that most of the effects of fracking shocks accrue to workers directly involved in extraction activities and, unsurprisingly, diffusion is therefore limited to areas immediately surrounding the drilling site (which also tend to be the most socially connected ones). This could simply be businesses registered or operating around the wells, with workers commuting daily. However, as discussed, a large portion of workers involved in extraction activities is often transient and from out-of-state. Does social connectedness play a role in the flows of transient workers? In particular, could it be that transient workers are disproportionately attracted to drilling sites if they live in places with stronger social ties

[20] I only report results based on the plain measure of social connectedness, rather than the partialled-out one, because there were no detectable effects on BLS wages in the latter. Feyrer et al. (2017) provide more details on effects across industries, space and time. An alternative way of studying this question could rely on the 'fields of influence' approach proposed by Sonis and Hewings (1992), which looks at perturbations in industry input-output relationships. However, an application of this method falls beyond the scope of this analysis.

to these sites? This would be consistent with the literature on job information networks. Directly testing this hypothesis is difficult. However, valuable indirect evidence can be obtained by looking at wages declared by workers at their place of permanent residence. To the extent that employees are transient and do not change their home address, this should reflect their county of origin. To this end, Figure 4.4 reports OLS estimates of the effects for IRS wages. Table 4.B.7 in Appendix reports exact estimates for all coefficients.

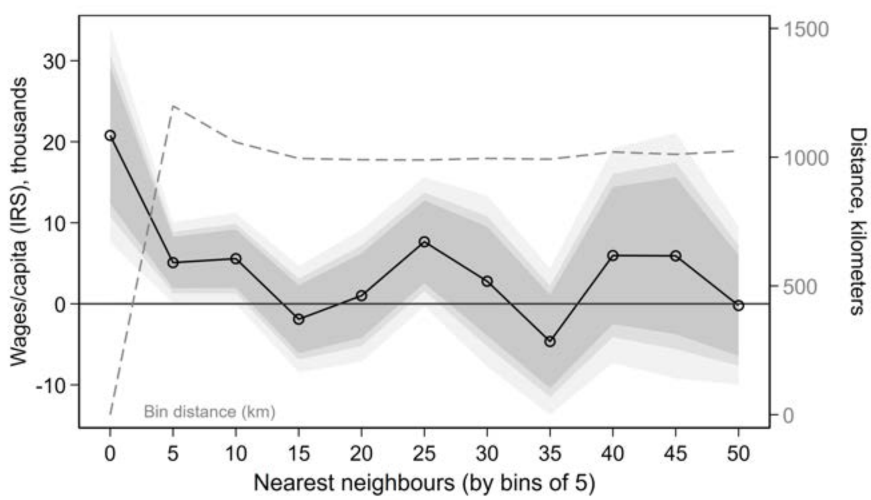
Figure 4.4: Coefficients plot for wages (IRS) using OLS



Linear regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Panel (a) largely confirms previous findings, although the relevance of social connectedness ($RSCI$) decays more slowly, with effects diffusing up to the 25$^{\text{th}}$ nearest social neighbour, or about 170 kilometres away. The key take-away from this diagram, however, is in panel (b), which uses the partialled-out measure of social connectedness to define neighbours. In this case, it appears that fracking shocks can result in small but significant wage increases up to the 25$^{\text{th}}$ most closely connected county, which corresponds to a pattern of spatial diffusion to regions over 1,000 kilometres away from where the initial shock was experienced. More accurately, a million dollar per capita increase in oil and gas extraction raises per capita wages by about 2,700 dollars per capita on average for workers reporting their incomes in counties located as far as 1,200 kilometres away from the drilling site, but strongly socially connected to it (up to the 25th nearest social neighbour, net of physical distance and migration). This finding is robust to controlling for the effects of new production in the county itself, and for inward effects from new production in the 100 counties surrounding it. The result is also confirmed when absorbing county fixed effects, and when using 2SLS estimators.[21] Figure 4.5 summarises results for IRS wages obtained using the proposed instrumental variable strategy on the partialled-out measure of social connectedness. Standard errors are adjusted to allow spatial correlation in the network measure by clustering over social bins. Results are slightly larger in this case, although decay is faster. A marginal increase in new production in connected counties is associated with an increase of wages of over 5,000 dollars per capita up to the 10th nearest social neighbour, once again controlling for incoming shocks from geographically proximate counties. These estimates are statistically significant at the 99 and 95 percent level for the first and second nearest neighbours bins respectively. Based on these estimates and the summary statistics reported in Table 4.B.2, the average combined effect on wages of a one standard deviation change in new production in the ten most strongly connected counties is of about 400 additional dollars per capita each year.

Figure 4.5: Coefficients plot for wages (IRS) using 2SLS (res. soc. dist.)



2SLS regression controlling for yearly trends, a one-year lag of new production, and spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis). R2=0.29, N=21,308.

According to the models presented in this analysis, geographical dispersion is almost one order of magnitude larger than that described by Feyrer et al. (2017), who place it at about 160 kilometres. In terms of interpretation, however, the evidence of dispersion documented herein differs. This is not money that is directly earned in far away places. Rather, I argue, it is information about new high paying jobs that travels over distance as a result of social networks, selectively attracting transient workers from regions across the country. The wage increases, thus, are earned by employees deployed on-site, but declaring their income in their place of origin. Whether and to what extent these accrued gains are transferred

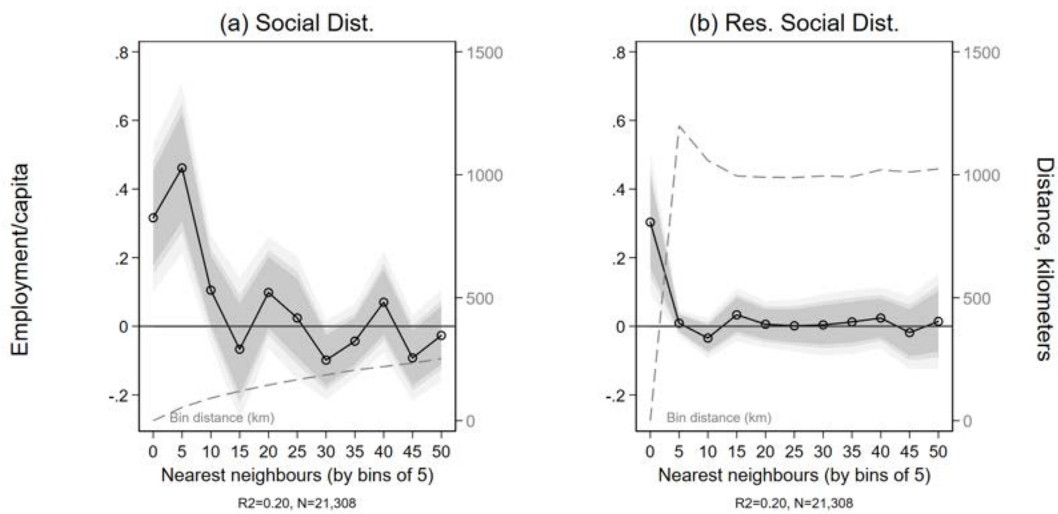
[21] See Appendix, Figures 4.A.11 and 4.A.13, and Tables 4.B.8 and 4.B.10, respectively.

back to their homes and injected into the local economies of distant places is hard to tell. However, the evidence from BLS wage regressions would not suggest that this takes place in any appreciable way, at least in the short run.[22]

4.4.2 Effects on Employment

What is the effect of new production of oil and gas in socially connected places on the employment of a county? Figure 4.6 provides baseline OLS estimates for this relationship (Table 4.B.11 in Appendix reports exact estimates for all coefficients).

Figure 4.6: Coefficients plot for employment using OLS



Linear regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

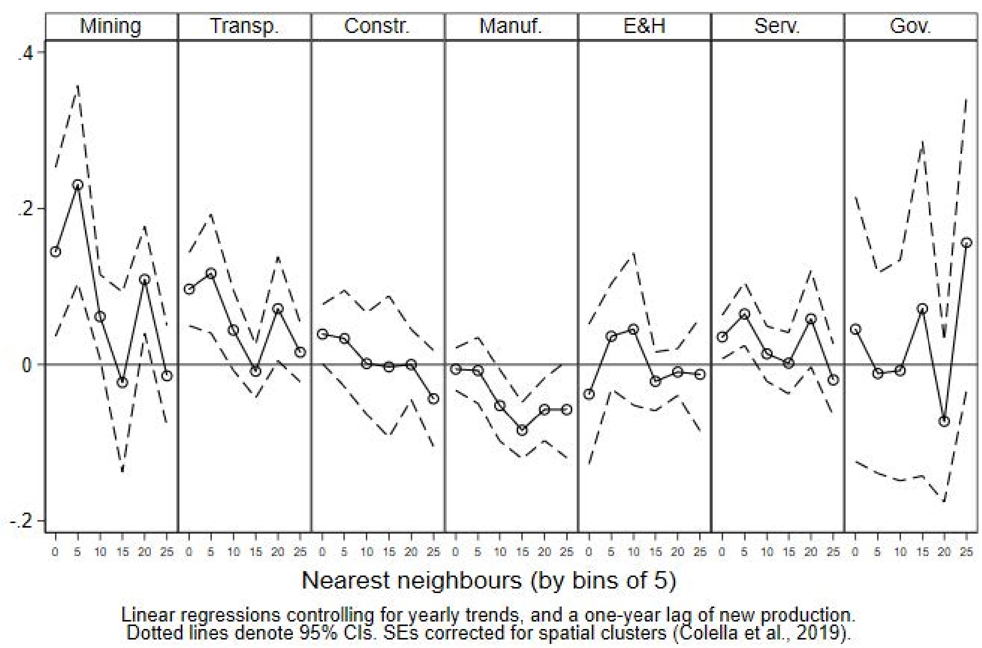
Each million dollar per capita of new production is associated with the creation of about a third of a new job for every existing job in the county itself. This estimate is comparable to that of Feyrer et al. (2017) at 0.4 using OLS. Conversely, fracking activity in the first five most strongly connected counties raises employment by just under half a new job for every existing one (panel a), corresponding to spatial dispersion of just over 50 kilometres. This effect, however, decays rapidly after that and is barely significant when the subsequent bin of social neighbours is considered (at about 100 kilometre distance on average). This pattern is consistent with the hypothesis of workers commuting over short distances to benefit from the jobs created by fracking. Indeed, once the role of space in determining

[22] Unfortunately, the IRS wage measure does not provide an industrial breakdown. This would have allowed to test whether surges in wages occur in extraction related sectors despite the geographical distance from the sites.

social networks is partialled-out, there are no effects of new production in socially connected counties on a county's employment (panel b). Results are largely confirmed when county fixed effects are introduced, as well as when 2SLS estimates are considered.[23]

It would thus appear that most of the dispersion of new job creation can be explained by geography rather than social networks. Once again this can be confirmed by looking directly at dispersion over nearest geographical neighbours (4.A.21), where effects are stronger (about 0.6 jobs) and at comparable average physical distances (about 50 kilometres on average in the first bin). Worthy of mention is that 2SLS estimates uncover some significant effects on employment of new production in the closest social neighbours even when distance and migration flows are partialled-out, suggesting dispersion in space up to 1,200 kilometres on average. These effects, however, are very small in magnitude (less than 0.1 of a new job for every existing one) and barely distinguishable from zero.

Figure 4.7: Coefficients plot for employment by industry using OLS (soc. dist.)



Similarly to what done with wages, we can look at a sector breakdown for employment creation. I report OLS estimates for the plain measure of social connectedness only, since there were no clearly discernible effects for the partialled-out measure of networks.[24] Consistent with the findings on wage-gains, Figure 4.7 shows that most job creation occurs directly in mining and transportation, with some new employment also being generated in

[23]See Appendix, Figures 4.A.18 and 4.A.20, and Tables 4.B.12 and 4.B.14, respectively.
[24]A corresponding set of results for geographical distance is available in Appendix (Figure 4.A.25).

services. Interestingly, it appears that new production of oil and gas in socially connected counties has small negative effects on manufacturing employment, especially over greater distances (third to fifth bin of social neighbours, corresponding geographically to about 80 to 130 kilometres). This gives some credit to the resource curse literature, suggesting that new job opportunities in fracking attract workers away from tradable goods production. Note that there were no clear effects on manufacturing wages, potentially due to rigidities in the sector.[25]

4.5 Conclusions

This paper has considered how plausibly exogenous shocks to local labour demand linked to hydraulic fracturing can diffuse in space over social networks. The empirical evidence supports qualitative predictions obtained from a simple conceptual framework and aligns with anecdotal findings from the sociological literature on fracking workers.

New production of oil and gas has positive inward effects on wages and employment in socially connected counties, mostly in mining and transportation industries, and to some extent in services, with some downward pressure on manufacturing jobs. Most of the diffusion over social networks is limited in space, but not all of it is simply a result of geographic proximity. This analysis also detected small effects of social networks irrespective of geographical considerations. In particular, I presented estimates obtained using a measure of social connectedness that partials-out any role of physical space in social interactions. These estimates suggest that a million dollar per capita increase in oil and gas extraction raises per capita wages by about 2,700 (OLS) and 5,000 (2SLS) dollars per capita for workers reporting their income in counties located as far as 1,200 kilometres away from the drilling site, but strongly socially connected to it. Evaluating 2SLS wage estimates using observed data on oil and gas production suggests that a county gains on average 400 dollars per capita each year from a one standard deviation increase in resource extraction in the top ten counties it interacts with socially.

This novel result is consistent with accounts of the fracking industry that discuss the importance of out-of-state hires and transient workers. It also provides new aggregate evidence in support of the literature on job information networks. Future work could examine this finding more closely using micro-data, helping understand the characteristics

[25] Due to the largely overlapping nature of dispersion over social and geographical neighbours, however, I refer the reader to the analysis by Feyrer et al. (2017) for a more detailed account of how fracking affects employment and wages in different industries in spatially contiguous areas and over time. The results presented herein are intended to briefly show that it is possible to obtain consistent results even when a measure based on social rather than physical distance is considered.

of itinerant workers and examining possible ‘push factors’ related to their employment patterns. It could also consider the dynamic dimension of cross-sectional shock dispersion, which in the case of hydraulic fracturing might be affected by a subsequent bust of the resource boom.

4.A Additional Figures

Figure 4.A.1: Bins of nearest neighbours for top producing counties in 2005-2012

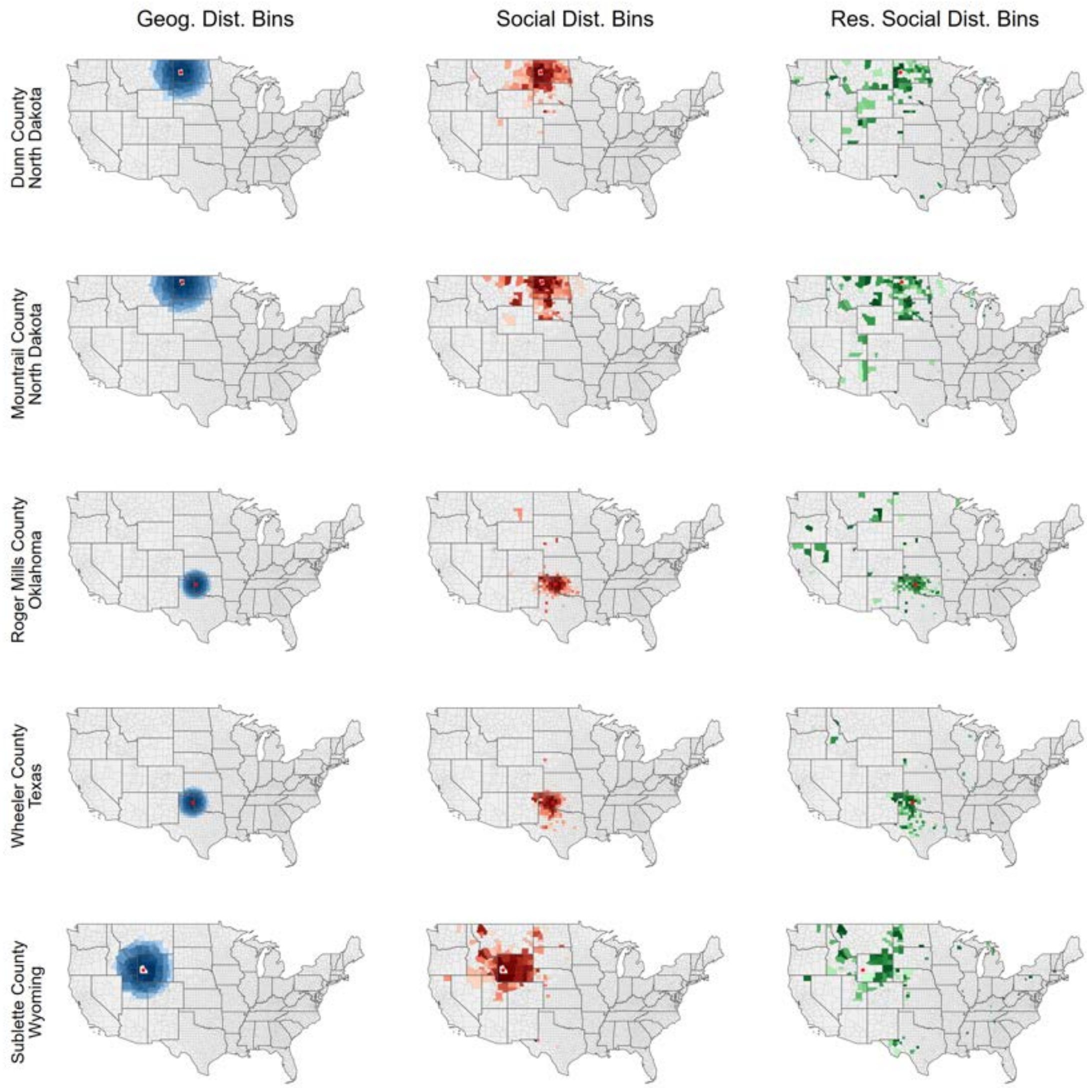
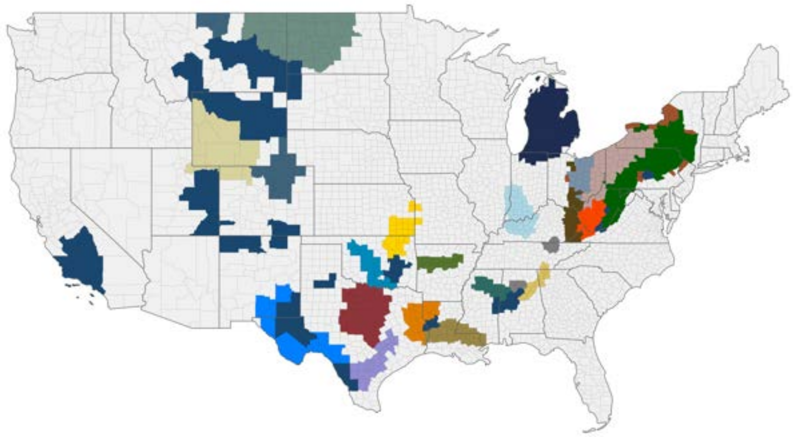
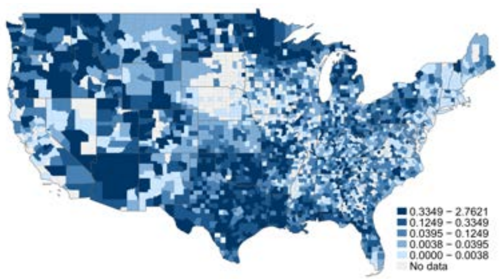


Figure 4.A.2: US Shale plays, 23 designations and one ‘other’ category

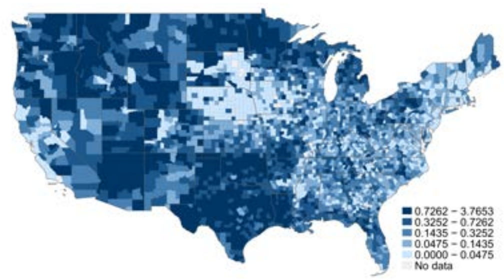


Source: EIA Shapefile “Major Tight Oil and Shale Gas Plays in Lower 48 States”

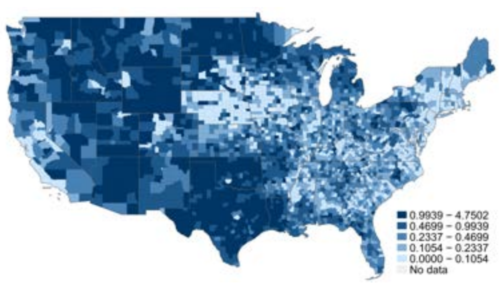
Figure 4.A.3: Cumulative socially lagged new production (partialled-out RSCI)



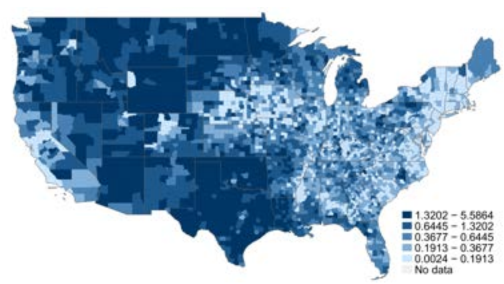
(a) First bin



(b) Up to 5$^{\text{th}}$ bin



(c) Up to 10$^{\text{th}}$ bin



(d) Up to 20$^{\text{th}}$ bin

4.A.1 Regression Coefficients Plots for Wages (BLS)

Figure 4.A.4: Coefficients plot for wages (BLS) using OLS with county FEs

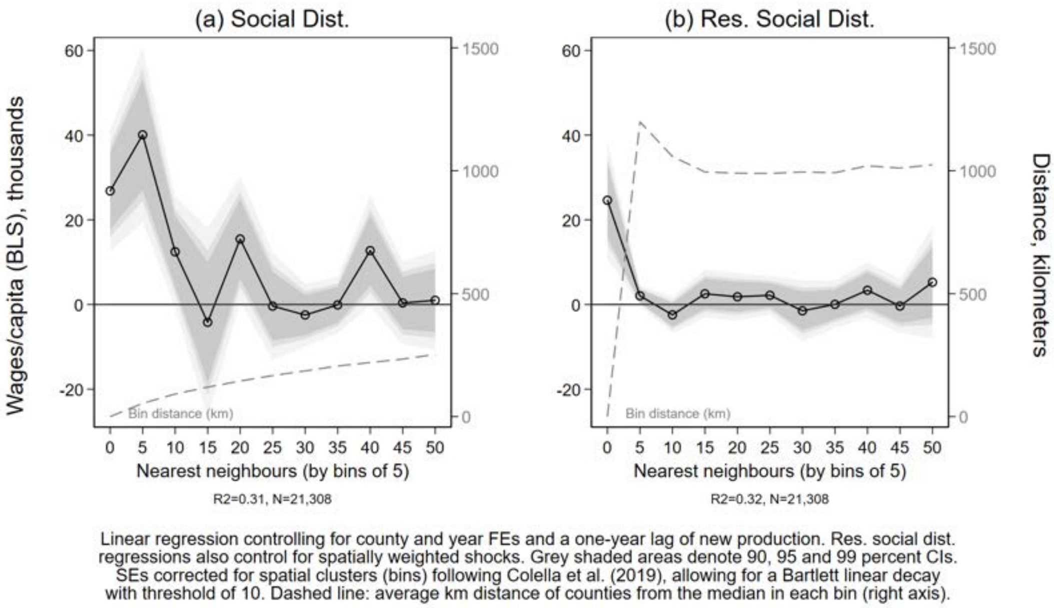


Figure 4.A.5: Coefficients plot for wages (BLS), reduced form of 2SLS

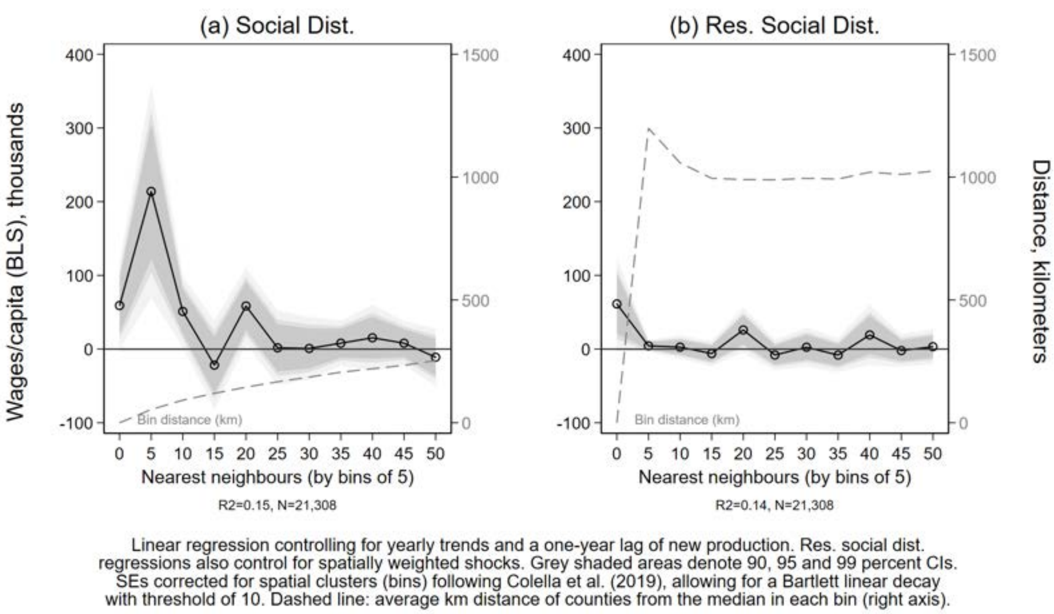


Figure 4.A.6: Coefficients plot for wages (BLS) using 2SLS

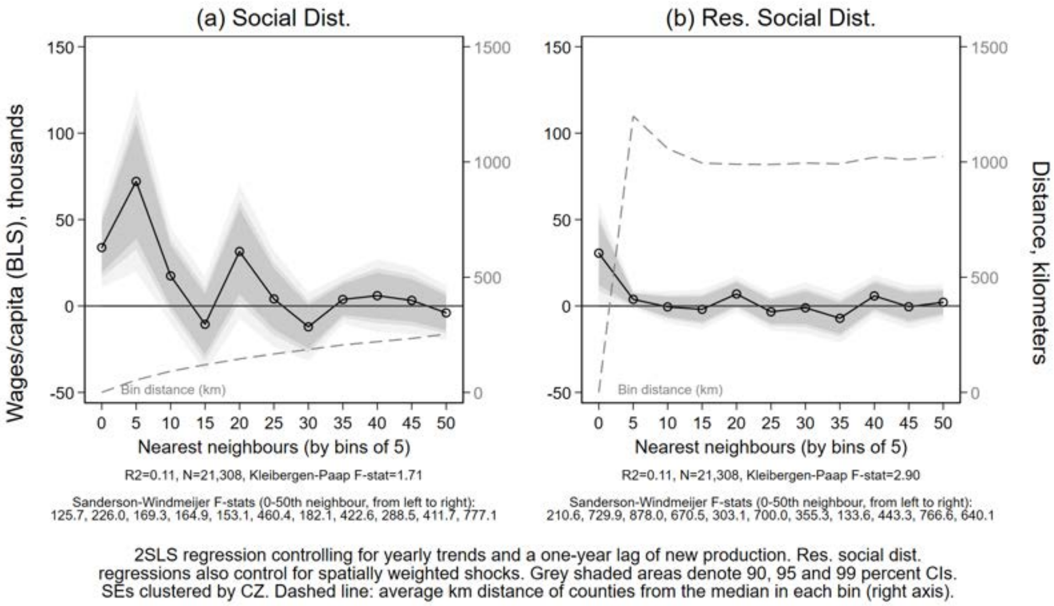


Figure 4.A.7: Coefficients plot for wages (BLS) using OLS with county FEs

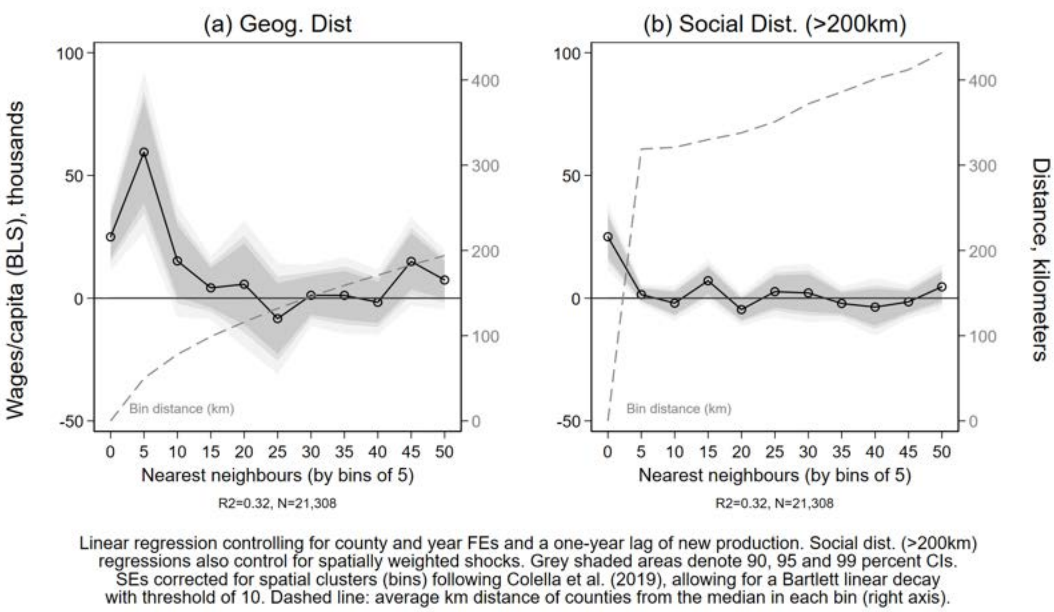
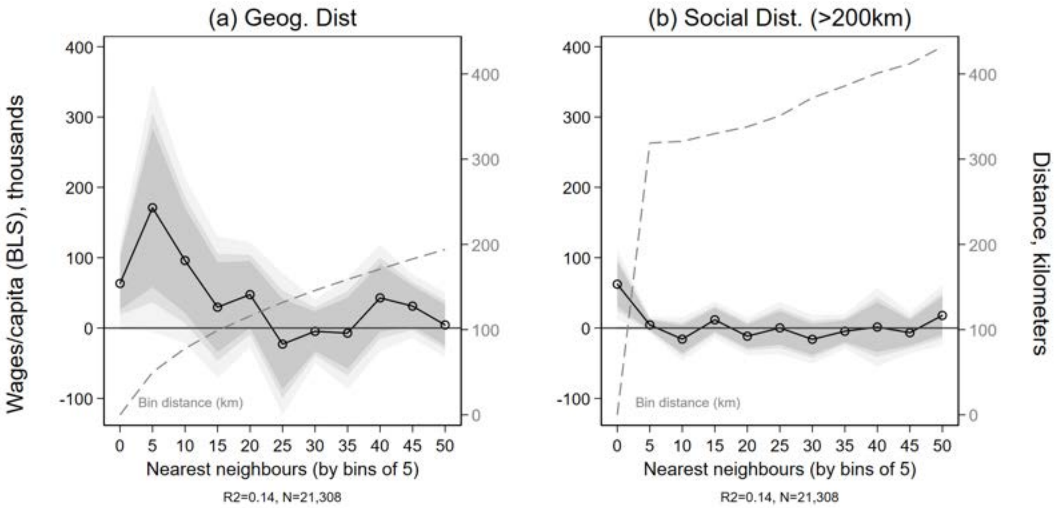
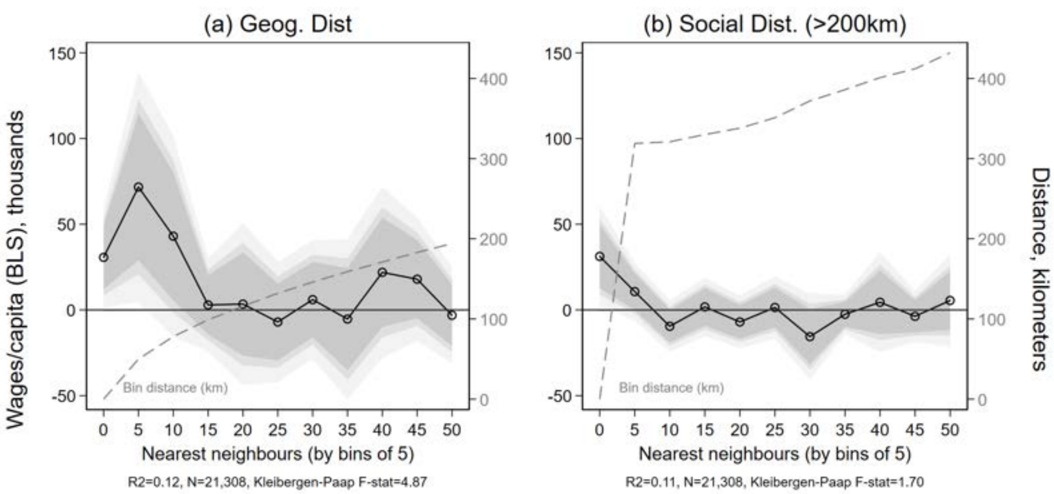


Figure 4.A.8: Coefficients plot for wages (BLS), reduced form of 2SLS



Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.9: Coefficients plot for wages (BLS) using 2SLS

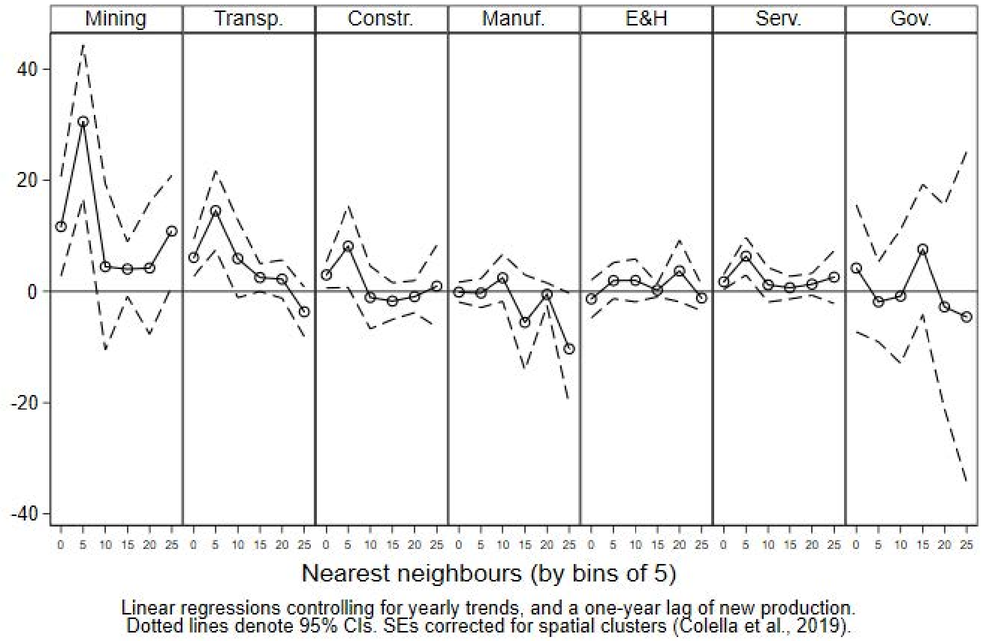


Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right): 110.8, 118.0, 92.7, 247.2, 245.7, 145.6, 81.7, 127.3, 172.3, 212.8, 208.3

Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right): 273.2, 301.6, 436.9, 387.5, 771.1, 370.8, 814.8, 639.5, 329.8, 437.9, 285.2

2SLS regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.10: Coefficient plot for wages (BLS) by industry using OLS (geog. dist.)



4.A.2 Regression Coefficients Plots for Wages (IRS)

Figure 4.A.11: Coefficients plot for wages (IRS) using OLS with county FEs

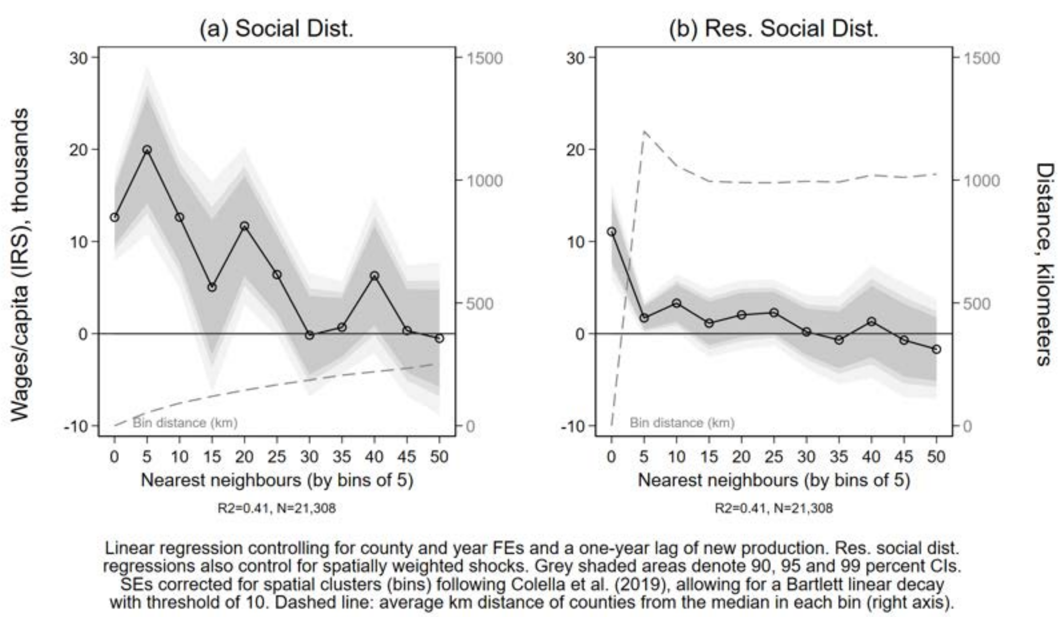
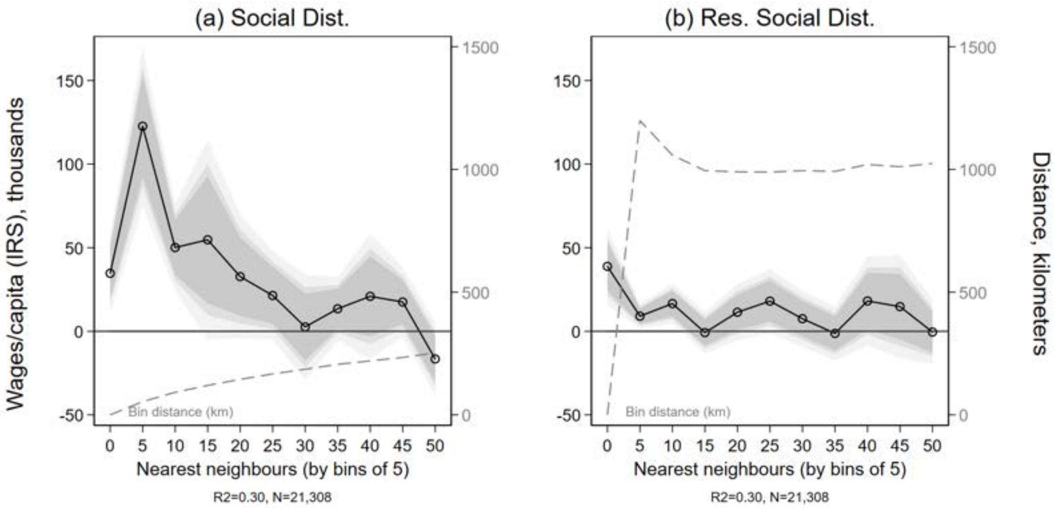
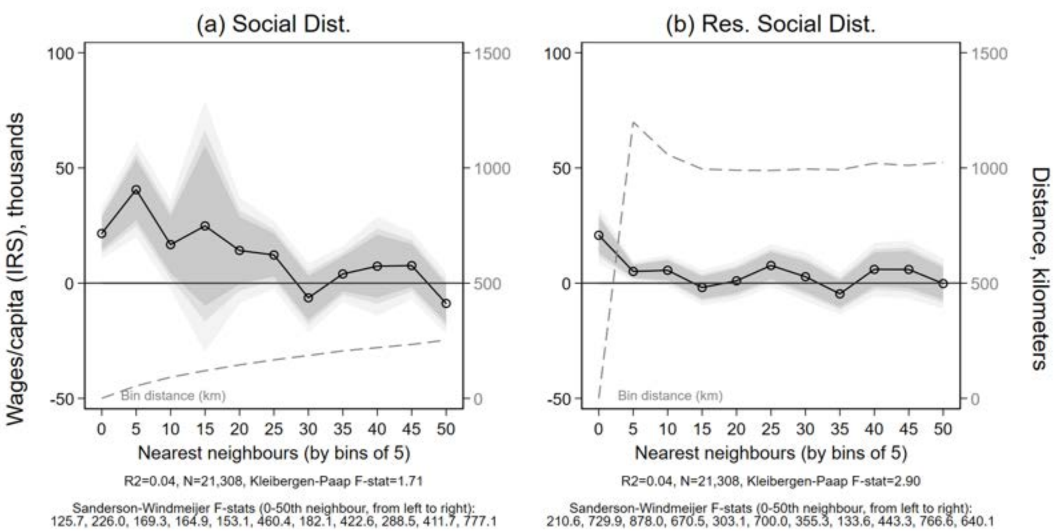


Figure 4.A.12: Coefficients plot for wages (IRS), reduced form of 2SLS



Linear regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.13: Coefficients plot for wages (IRS) using 2SLS



2SLS regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.14: Coefficients plot for wages (IRS) using OLS

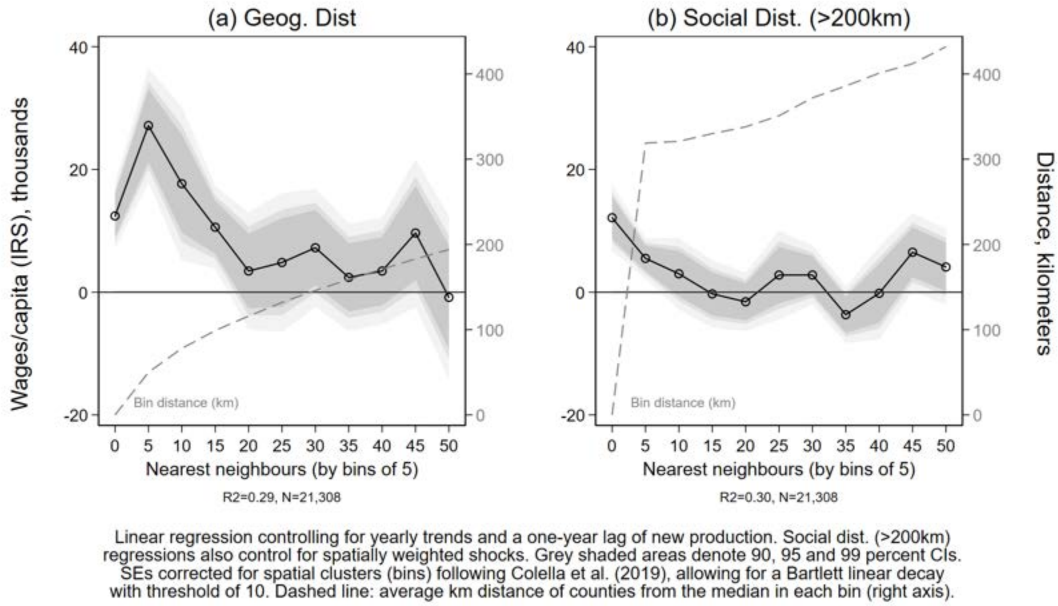


Figure 4.A.15: Coefficients plot for wages (IRS) using OLS with county FEs

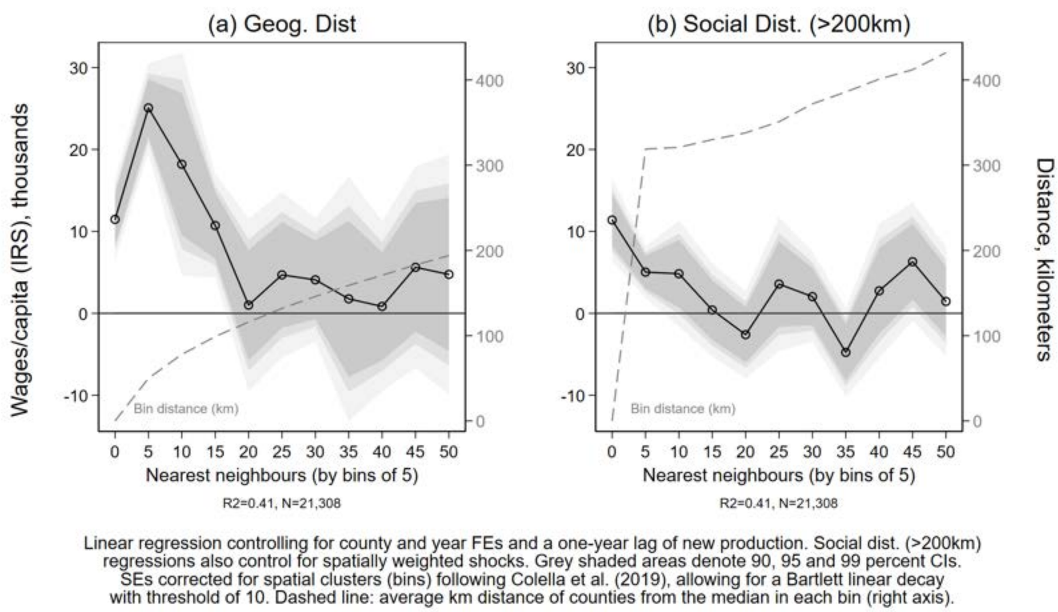
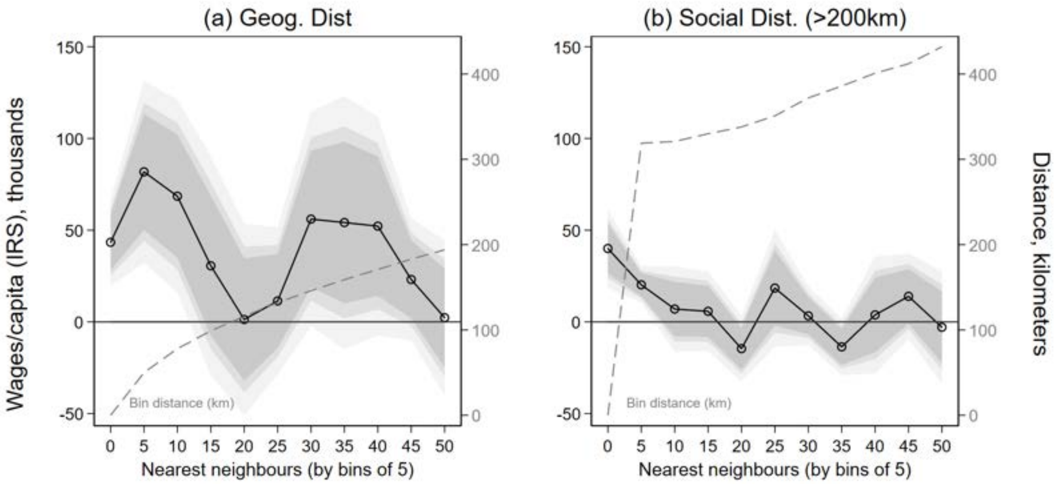


Figure 4.A.16: Coefficients plot for wages (IRS), reduced form of 2SLS

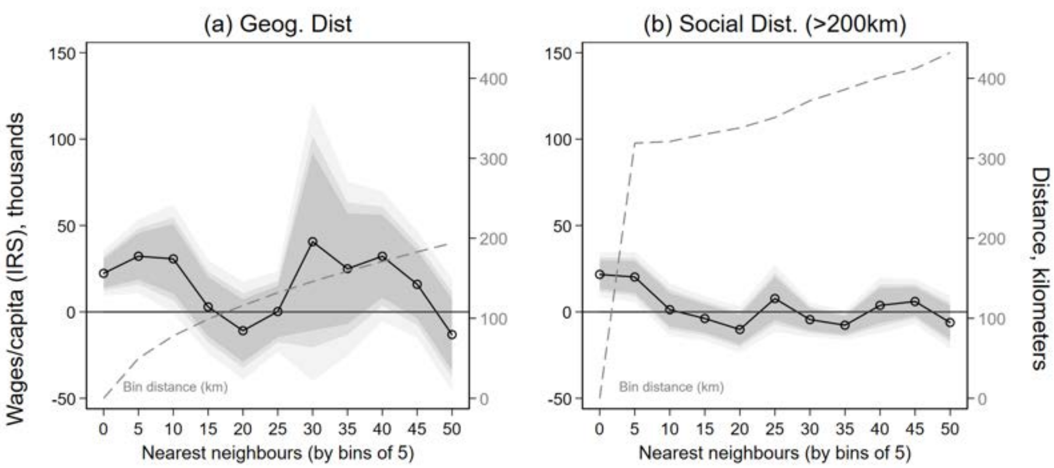


R2=0.29, N=21,308

R2=0.30, N=21,308

Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.17: Coefficients plot for wages (IRS) using 2SLS



R2=0.04, N=21,308, Kleibergen-Paap F-stat=4.87

Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right):
110.8, 118.0, 92.7, 247.2, 245.7, 145.6, 81.7, 127.3, 172.3, 212.8, 208.3

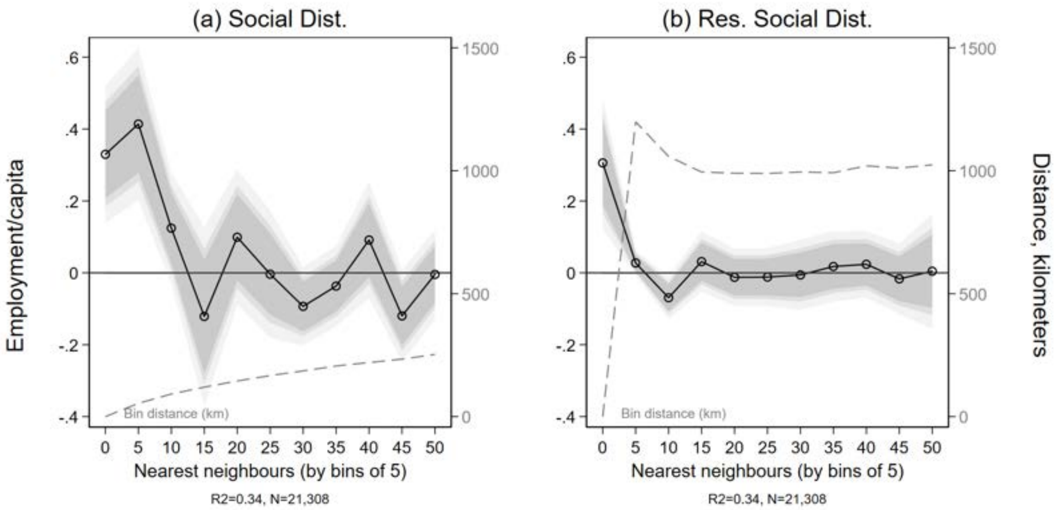
R2=0.04, N=21,308, Kleibergen-Paap F-stat=1.70

Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right):
273.2, 301.6, 436.9, 387.5, 771.1, 370.8, 814.8, 639.5, 329.8, 437.9, 285.2

2SLS regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin (right axis).

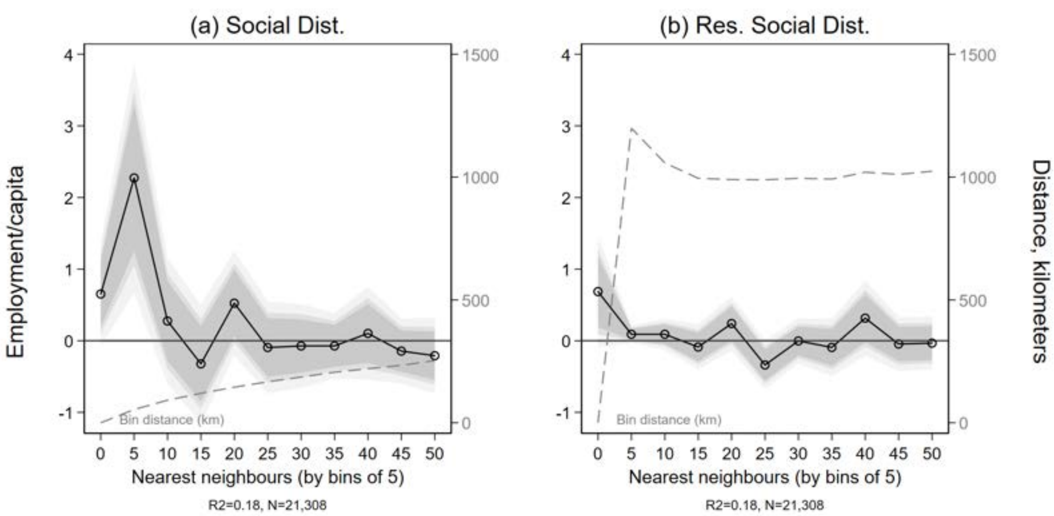
4.A.3 Regression Coefficients Plots for Employment

Figure 4.A.18: Coefficients plot for employment using OLS with county FEs



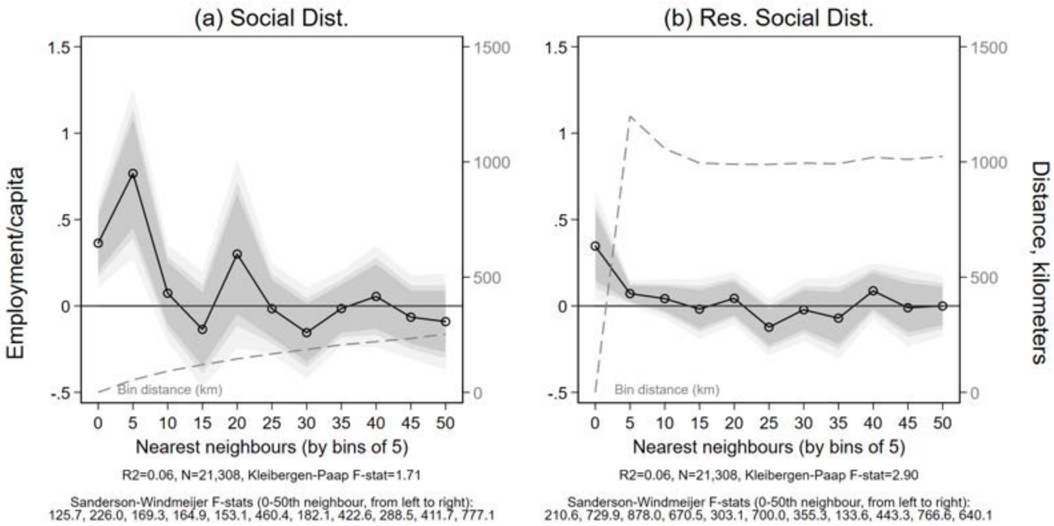
Linear regression controlling for county and year FEs and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.19: Coefficients plot for employment, reduced form of 2SLS



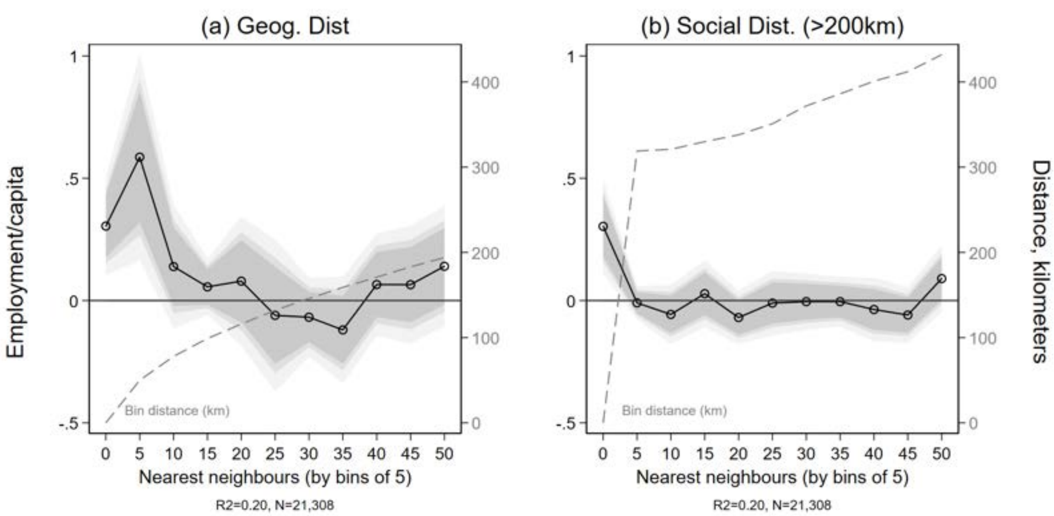
Linear regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.20: Coefficients plot for employment using 2SLS



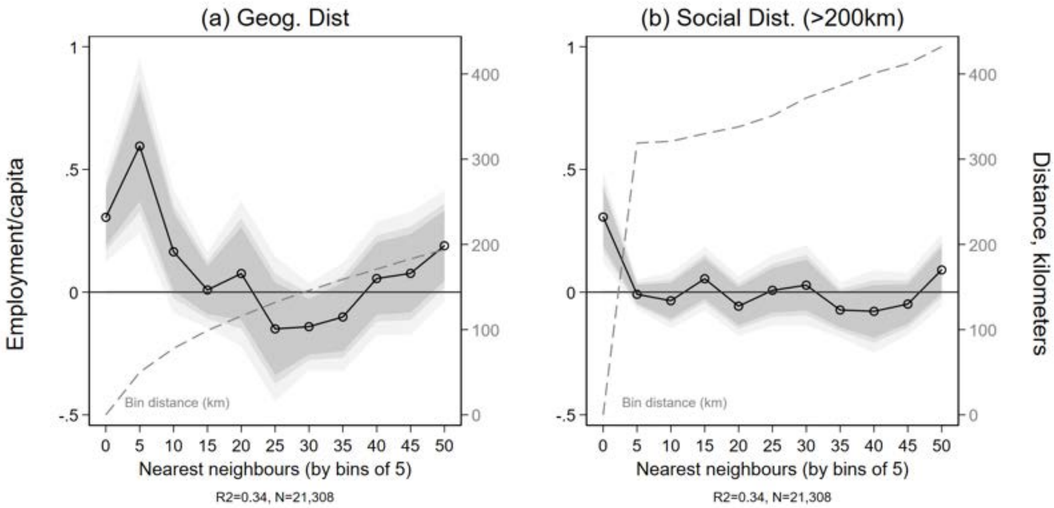
2SLS regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.21: Coefficients plot for employment using OLS



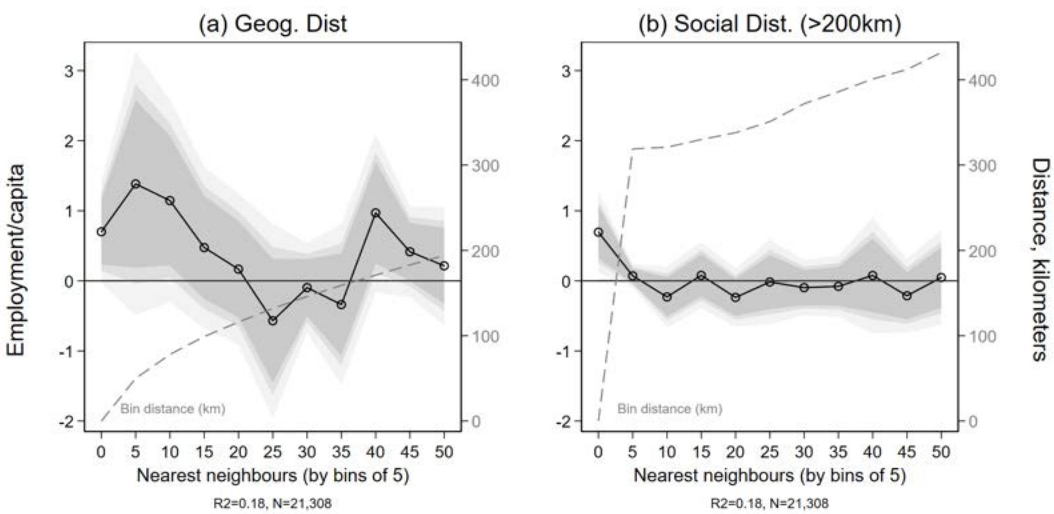
Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.22: Coefficients plot for employment using OLS with county FEs



Linear regression controlling for county and year FEs and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.23: Coefficients plot for employment, reduced form of 2SLS



Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed line: average km distance of counties from the median in each bin (right axis).

Figure 4.A.24: Coefficients plot for employment using 2SLS

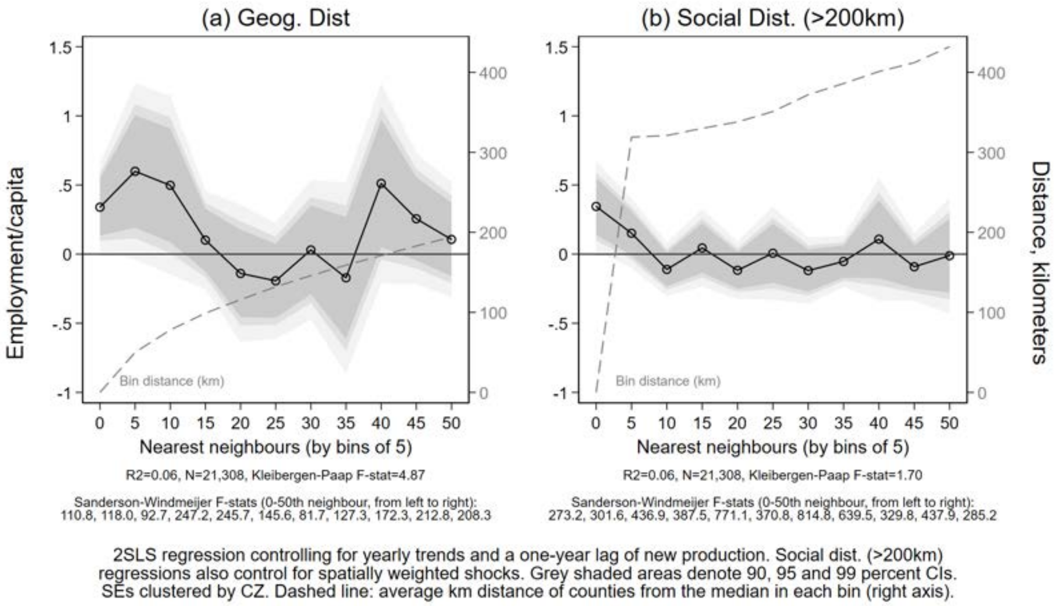
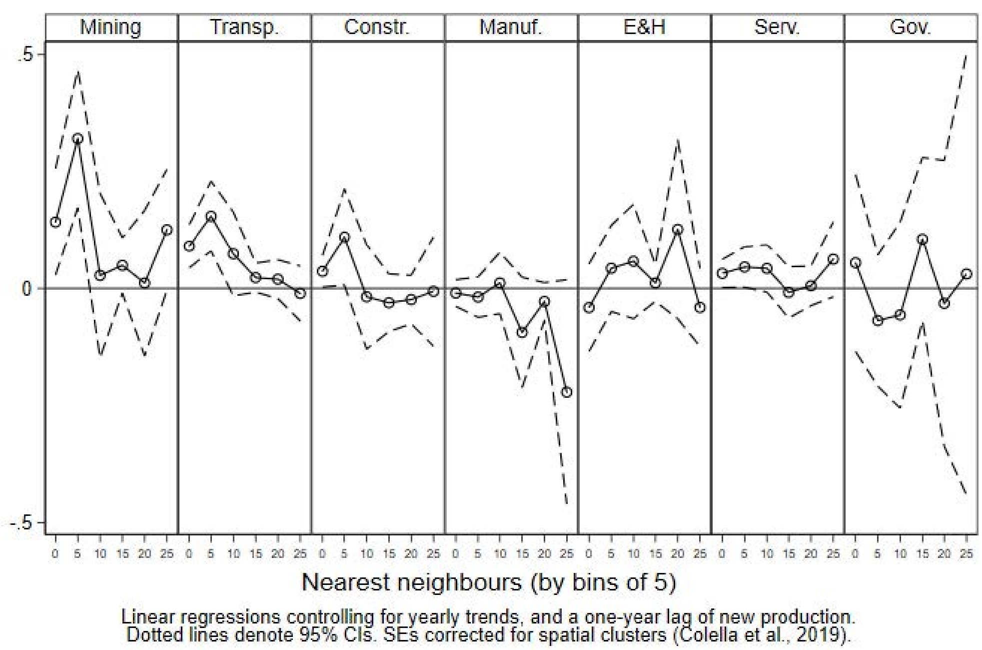


Figure 4.A.25: Coefficients plot for employment by industry using OLS (geog. dist.)



4.B Additional Tables

4.B.1 Summary Statistics

Table 4.B.1: Top 20 producing states over 2005-2012

State	Rank by new prod. Per capita	Total	New production Per capita	Total	Δ Empl. capita	Δ Wages capita
North Dakota	1	2	0.1189	45,889.67	0.33	29,005.43
Wyoming	2	6	0.0757	20,026.82	0.14	13,542.86
New Mexico	3	7	0.0243	18,860.81	0.04	3,975.75
Oklahoma	4	3	0.0195	28,878.77	0.11	9,490.23
Texas	5	1	0.0148	149,962.50	0.19	14,394.42
Louisiana	6	4	0.0148	27,212.29	0.02	6,302.91
Colorado	7	5	0.0109	24,277.54	0.12	8,490.79
Montana	8	14	0.0090	3,683.46	0.10	7,650.65
Arkansas	9	9	0.0084	9,510.96	0.02	2,616.08
Utah	10	10	0.0079	9,320.16	0.19	10,665.65
West Virginia	11	12	0.0070	4,785.64	0.00	2,735.08
Kansas	12	13	0.0036	4,638.46	0.04	3,381.15
Pennsylvania	13	8	0.0031	16,849.00	0.02	2,552.36
Mississippi	14	16	0.0026	2,796.24	0.00	1,004.55
Alabama	15	17	0.0006	1,090.46	0.01	1,536.05
Ohio	16	15	0.0006	2,858.24	-0.03	-1,408.19
California	17	11	0.0005	7,535.82	0.06	6,212.40
Kentucky	18	20	0.0003	583.81	0.04	2,126.07
Nebraska	19	21	0.0002	201.91	0.07	4,009.81
Virginia	20	19	0.0002	605.41	0.04	4,056.77

Note: The table excludes the smallest 2% of counties in terms of population.

Table 4.B.2: Summary statistics for the main variables in the analysis (2005-2012)

	Mean	Std. Dev.	Min.	25^{th} Pct.	Median	75^{th} Pct.	Max.
Δ Empl. pc	0.0010	0.0452	-0.5603	-0.0184	0.0017	0.0200	1.6244
Δ Wages pc	249.4034	2,546.7273	-36,281.6250	-717.5019	132.0744	1,031.8842	71,280.0703
Δ IRS wages pc	440.9627	2,552.1993	-114996.1484	-669.6421	266.3925	1,366.8599	107,233.7578
Δ IRS oth. inc. pc	996.4099	6,369.1418	-343888.8438	-540.6148	656.9911	2,148.3064	316,529.9688
Δ IRS AGI pc	1,393.9108	6,894.9531	-244675.6719	-1,294.9114	1,100.8493	3,516.9590	258,477.3594
Δ New prod. pc	0.0022	0.0206	0.0000	0.0000	0.0000	0.0000	0.7588
G_1 New prod. pc	0.0019	0.0119	0.0000	0.0000	0.0000	0.0001	0.3926
G_1 New prod. pc (res.)	0.0226	0.0489	0.0000	0.0002	0.0040	0.0204	0.6637
G_2 New prod. pc	0.0020	0.0116	0.0000	0.0000	0.0000	0.0001	0.3207
G_2 New prod. pc (res.)	0.0104	0.0286	0.0000	0.0000	0.0008	0.0066	0.5457
G_3 New prod. pc	0.0020	0.0115	0.0000	0.0000	0.0000	0.0002	0.4389
G_3 New prod. pc (res.)	0.0073	0.0228	0.0000	0.0000	0.0004	0.0040	0.5376
G_4 New prod. pc	0.0019	0.0094	0.0000	0.0000	0.0000	0.0002	0.3516
G_4 New prod. pc (res.)	0.0063	0.0203	0.0000	0.0000	0.0003	0.0031	0.3635
G_5 New prod. pc	0.0020	0.0103	0.0000	0.0000	0.0000	0.0003	0.4310
G_5 New prod. pc (res.)	0.0054	0.0185	0.0000	0.0000	0.0002	0.0024	0.3447

Note: The table excludes the smallest 2% of counties in terms of population.

4.B.2 Regression Tables for Wages (BLS)

Table 4.B.3: Regression table for wages (BLS) using OLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	26.44^a	(6.459)	25.12^a	(5.898)	24.72^a	(6.043)	25.01^a	(5.694)
Social neighbours								
1 to 5th	42.59^a	(9.011)			0.826	(1.047)	1.318	(1.706)
6 to 10th	12.26^a	(4.668)			-1.800	(1.527)	-3.005	(2.880)
11 to 15th	-0.397	(6.704)			2.629	(1.985)	5.296^c	(3.043)
16 to 20th	14.67^a	(5.392)			2.454	(1.873)	-3.853	(2.393)
21 to 25th	0.580	(4.901)			3.550^b	(1.796)	2.800	(3.312)
26 to 30th	-3.474	(2.682)			0.0418	(2.522)	-2.634	(2.956)
31 to 35th	1.178	(2.546)			-0.314	(2.213)	0.392	(2.192)
36 to 40th	11.21^b	(4.446)			1.988	(2.458)	0.0639	(3.286)
41 to 45th	1.409	(3.451)			-0.178	(2.522)	-1.843	(2.321)
46 to 50th	0.703	(4.756)			5.851	(4.295)	5.339	(3.444)
Geog. neighbours								
1 to 5th			59.31^a	(14.86)	58.16^a	(13.39)	59.11^a	(11.22)
6 to 10th			13.90^c	(7.870)	12.52	(9.445)	13.99	(10.03)
11 to 15th			7.564^c	(4.180)	7.038^c	(3.795)	7.562^b	(3.772)
16 to 20th			7.012	(8.674)	5.476	(8.462)	6.467	(8.912)
21 to 25th			-5.422	(8.988)	-6.577	(8.854)	-5.457	(8.571)
26 to 30th			3.098	(4.424)	2.129	(4.453)	2.671	(4.933)
31 to 35th			-0.507	(5.788)	-1.861	(5.377)	0.396	(5.223)
36 to 40th			2.807	(5.106)	1.449	(5.414)	1.818	(5.761)
41 to 45th			11.65^c	(6.538)	10.72^c	(6.319)	10.77^c	(6.424)
46 to 50th			3.647	(4.739)	3.557	(4.474)	4.827	(4.387)
R^2	0.18		0.18		0.18		0.18	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.4: Regression table for wages (BLS) using OLS with county FEs

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	26.83^a	(5.526)	24.98^a	(5.220)	24.62^a	(5.323)	25.02^a	(5.162)
Social neighbours								
1 to 5th	40.07^a	(7.978)			2.058^b	(0.920)	1.471	(1.544)
6 to 10th	12.46^b	(5.089)			-2.445	(1.678)	-2.009	(2.742)
11 to 15th	-4.208	(8.647)			2.533	(2.164)	7.108^b	(3.281)
16 to 20th	15.46^a	(5.717)			1.823	(2.165)	-4.630^c	(2.575)
21 to 25th	-0.386	(4.926)			2.177	(1.759)	2.670	(3.944)
26 to 30th	-2.471	(2.865)			-1.483	(2.789)	2.140	(4.672)
31 to 35th	-0.103	(2.576)			0.0364	(2.142)	-2.146	(2.639)
36 to 40th	12.74^b	(5.052)			3.370	(2.544)	-3.626	(4.549)
41 to 45th	0.374	(3.779)			-0.359	(2.455)	-1.485	(2.617)
46 to 50th	0.986	(4.539)			5.254	(5.120)	4.640	(3.566)
Geog. neighbours								
1 to 5th			59.49^a	(12.68)	57.80^a	(12.15)	58.63^a	(10.96)
6 to 10th			15.19^c	(8.832)	13.57	(8.748)	14.60	(10.96)
11 to 15th			4.211	(4.960)	3.789	(4.334)	3.699	(3.976)
16 to 20th			5.708	(10.02)	4.361	(9.534)	5.093	(9.436)
21 to 25th			-8.313	(8.758)	-9.885	(8.951)	-10.40	(9.475)
26 to 30th			1.179	(4.817)	-0.102	(5.056)	0.0827	(5.146)
31 to 35th			1.138	(6.046)	0.0416	(5.411)	2.178	(5.385)
36 to 40th			-1.652	(5.099)	-2.827	(5.565)	-2.619	(5.531)
41 to 45th			14.97^b	(6.977)	13.46^b	(6.684)	12.79^b	(6.331)
46 to 50th			7.429	(4.629)	6.560	(4.421)	7.062	(4.323)
R^2	0.31		0.32		0.32		0.32	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.5: Regression table for wages (BLS), reduced form of 2SLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	59.07^{a}	(22.86)	63.42^{a}	(22.64)	61.26^{b}	(24.53)	62.27^{a}	(19.31)
Social neighbours								
1 to 5th	213.8^{a}	(56.03)			4.394	(3.209)	4.519	(4.510)
6 to 10th	51.08^{a}	(18.56)			2.665	(6.065)	-15.79	(11.97)
11 to 15th	-21.74	(23.39)			-6.084	(6.716)	11.69	(10.11)
16 to 20th	58.47^{a}	(20.07)			26.00^{b}	(12.00)	-11.57	(10.15)
21 to 25th	1.691	(19.51)			-7.995	(8.028)	0.197	(14.43)
26 to 30th	0.854	(16.46)			2.511	(10.27)	-16.04	(13.43)
31 to 35th	8.025	(11.92)			-8.125	(8.877)	-4.547	(9.447)
36 to 40th	15.33	(17.34)			19.20	(15.83)	1.398	(21.65)
41 to 45th	8.040	(11.56)			-2.022	(8.590)	-6.633	(11.17)
46 to 50th	-11.06	(15.17)			3.205	(9.149)	17.95	(16.63)
Geog. neighbours								
1 to 5th			171.0^{b}	(68.60)	166.7^{a}	(62.64)	171.1^{a}	(55.03)
6 to 10th			96.03^{b}	(45.38)	87.28^{b}	(42.18)	93.45^{b}	(37.54)
11 to 15th			29.64	(38.88)	29.78	(38.97)	30.35	(37.97)
16 to 20th			47.57	(28.95)	49.19	(39.88)	48.19	(37.30)
21 to 25th			-22.87	(39.05)	-25.73	(36.75)	-18.58	(39.14)
26 to 30th			-4.693	(17.35)	-9.766	(19.31)	-4.202	(18.64)
31 to 35th			-7.189	(30.73)	-7.886	(29.37)	-5.511	(30.34)
36 to 40th			42.86	(29.37)	36.77	(29.46)	39.45	(29.24)
41 to 45th			31.18^{c}	(17.46)	32.51	(20.10)	34.28	(22.86)
46 to 50th			4.426	(17.78)	0.870	(18.26)	8.502	(18.56)
R^2	0.15		0.14		0.14		0.14	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.6: Regression table for wages (BLS) using 2SLS

	(1) Social Dist.			(2) Geog. Dist.			(3) Res. Social Dist.			(4) Alt. Social Dist.		
	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F
Own county	33.8^{a}	(8.83)	[125.7]	30.7^{a}	(11.3)	[110.8]	30.6^{a}	(11.2)	[210.6]	31.2^{a}	(11.3)	[273.2]
Social neighbours												
1 to 5th	72.1^{a}	(20.1)	[226]				3.84^{c}	(2.05)	[729.9]	10.7	(6.53)	[301.6]
6 to 10th	17.5	(10.7)	[169.3]				-0.53	(3.48)	[878]	-9.52^{c}	(5.77)	[436.9]
11 to 15th	-10.6	(10.6)	[164.9]				-1.95	(4.43)	[670.5]	1.76	(6.64)	[387.5]
16 to 20th	31.5^{b}	(15.0)	[153.1]				6.94^{c}	(4.04)	[303.1]	-7.00	(5.95)	[771.1]
21 to 25th	4.09	(10.7)	[460.4]				-3.38	(4.18)	[700]	1.44	(7.03)	[370.8]
26 to 30th	-12.1	(7.65)	[182.1]				-1.04	(5.72)	[355.3]	-15.6	(9.73)	[814.8]
31 to 35th	3.76	(5.44)	[422.6]				-7.12	(5.45)	[133.6]	-2.54	(4.47)	[639.5]
36 to 40th	5.98	(8.10)	[288.5]				5.81	(4.62)	[443.3]	4.49	(11.2)	[329.8]
41 to 45th	3.22	(7.45)	[411.7]				-0.45	(4.88)	[766.6]	-3.71	(5.72)	[437.9]
46 to 50th	-3.96	(6.09)	[777.1]				2.17	(4.02)	[640.1]	5.50	(10.4)	[285.2]
Geog. neighbours												
1 to 5th				71.7^{a}	(25.9)	[118]	70.4^{a}	(27.0)	[340.8]	74.2^{a}	(25.9)	[418.6]
6 to 10th				43.0^{c}	(22.7)	[92.7]	40.8^{c}	(22.6)	[236.1]	41.8^{c}	(23.3)	[169.7]
11 to 15th				2.92	(10.5)	[247.2]	5.34	(10.4)	[419.1]	3.12	(10.4)	[374.7]
16 to 20th				3.44	(18.3)	[245.7]	6.31	(18.3)	[431.4]	4.29	(18.7)	[375.8]
21 to 25th				-7.02	(13.6)	[145.6]	-6.38	(13.0)	[315]	-3.87	(14.3)	[264.5]
26 to 30th				6.02	(13.3)	[81.7]	1.43	(12.4)	[139.5]	2.54	(14.3)	[139.9]
31 to 35th				-5.31	(18.1)	[127.3]	-7.18	(18.1)	[227.8]	-4.45	(18.6)	[329]
36 to 40th				21.9	(19.3)	[172.3]	19.4	(19.4)	[373.2]	20.0	(20.3)	[408.3]
41 to 45th				17.9	(13.9)	[212.8]	20.2	(13.4)	[464.8]	19.9	(13.9)	[342.2]
46 to 50th				-3.10	(10.9)	[208.3]	-0.016	(11.3)	[310.4]	0.75	(10.9)	[316]
R^2	0.11			0.12			0.11			0.11		
N	21,308			21,308			21,308			21,308		
First stage KP F	1.71			4.87			2.90			1.70		

SEs clustered by commuting zone. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

4.B.3 Regression Tables for Wages (IRS)

Table 4.B.7: Regression table for wages (IRS) using OLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	13.66[a]	(2.098)	12.43[a]	(2.083)	11.75[a]	(2.124)	12.13[a]	(2.088)
Social neighbours								
1 to 5th	21.49[a]	(3.831)			1.244[c]	(0.736)	5.511[a]	(1.288)
6 to 10th	12.78[a]	(2.835)			4.215[a]	(1.145)	3.002	(2.247)
11 to 15th	5.442	(3.771)			2.126	(1.294)	-0.274	(2.122)
16 to 20th	11.97[a]	(3.029)			2.960[b]	(1.464)	-1.558	(1.839)
21 to 25th	8.305[a]	(2.607)			3.905[a]	(1.410)	2.793	(2.808)
26 to 30th	-2.148	(2.134)			0.641	(1.580)	2.800	(1.864)
31 to 35th	1.910	(1.892)			1.376	(1.814)	-3.667[b]	(1.797)
36 to 40th	4.921[c]	(2.848)			2.696	(2.308)	-0.176	(2.929)
41 to 45th	2.238	(2.356)			0.0507	(2.386)	6.509[a]	(2.471)
46 to 50th	-1.289	(2.567)			0.669	(2.022)	4.109[c]	(2.406)
Geog. neighbours								
1 to 5th			27.16[a]	(3.616)	25.43[a]	(3.336)	26.55[a]	(2.962)
6 to 10th			17.69[a]	(4.865)	16.27[a]	(4.995)	17.73[a]	(5.249)
11 to 15th			10.59[a]	(2.627)	9.781[a]	(2.251)	10.70[a]	(2.266)
16 to 20th			3.456	(3.704)	1.771	(3.474)	2.925	(3.448)
21 to 25th			4.840	(4.395)	4.478	(4.137)	3.003	(4.092)
26 to 30th			7.240[c]	(3.741)	6.199[c]	(3.236)	6.376[c]	(3.610)
31 to 35th			2.411	(3.382)	1.285	(3.063)	2.489	(3.201)
36 to 40th			3.448	(3.377)	1.289	(3.974)	-0.314	(4.003)
41 to 45th			9.648[b]	(4.682)	9.031[b]	(4.187)	7.960[c]	(4.170)
46 to 50th			-0.842	(5.121)	-0.755	(5.393)	-1.373	(5.414)
R^2	0.29		0.29		0.30		0.30	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.8: Regression table for wages (IRS) using OLS with county FEs

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	12.63[a]	(1.863)	11.47[a]	(1.908)	11.07[a]	(2.007)	11.39[a]	(1.939)
Social neighbours								
1 to 5th	19.97[a]	(3.533)			1.705[b]	(0.755)	5.034[a]	(1.225)
6 to 10th	12.64[a]	(2.995)			3.303[a]	(1.231)	4.838[c]	(2.507)
11 to 15th	5.040	(4.442)			1.131	(1.433)	0.407	(2.209)
16 to 20th	11.69[a]	(3.298)			2.041	(1.455)	-2.606	(2.056)
21 to 25th	6.416[b]	(2.602)			2.278[c]	(1.381)	3.573	(3.164)
26 to 30th	-0.170	(2.594)			0.200	(1.528)	2.040	(2.124)
31 to 35th	0.688	(1.932)			-0.683	(1.868)	-4.762[b]	(2.098)
36 to 40th	6.295[c]	(3.282)			1.316	(2.368)	2.749	(3.179)
41 to 45th	0.324	(2.750)			-0.706	(2.397)	6.299[b]	(2.803)
46 to 50th	-0.511	(3.200)			-1.692	(2.096)	1.456	(2.587)
Geog. neighbours								
1 to 5th			25.08[a]	(2.124)	23.24[a]	(2.872)	23.90[a]	(3.099)
6 to 10th			18.20[a]	(5.287)	16.89[a]	(4.998)	17.45[a]	(5.269)
11 to 15th			10.72[a]	(2.463)	9.560[a]	(2.316)	10.34[a]	(2.111)
16 to 20th			1.004	(4.087)	0.148	(3.740)	0.520	(3.614)
21 to 25th			4.692	(3.919)	3.214	(4.048)	1.842	(3.946)
26 to 30th			4.087	(2.923)	2.851	(2.615)	3.067	(2.703)
31 to 35th			1.769	(5.787)	1.754	(5.494)	1.745	(5.482)
36 to 40th			0.844	(4.014)	0.126	(4.448)	-2.885	(4.624)
41 to 45th			5.617	(4.771)	3.689	(4.749)	2.564	(4.258)
46 to 50th			4.750	(5.677)	3.747	(5.686)	2.097	(5.713)
R^2	0.41		0.41		0.41		0.41	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.9: Regression table for wages (IRS), reduced form of 2SLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	34.69[a]	(9.303)	43.38[a]	(9.173)	38.83[a]	(8.941)	40.01[a]	(8.292)
Social neighbours								
1 to 5th	122.7[a]	(18.46)			9.050[a]	(2.821)	20.22[a]	(3.857)
6 to 10th	50.08[a]	(10.06)			16.58[a]	(4.675)	7.001	(9.037)
11 to 15th	54.74[b]	(23.10)			-0.791	(4.844)	5.744	(8.459)
16 to 20th	32.73[b]	(14.28)			11.40[c]	(6.497)	-14.58[b]	(6.929)
21 to 25th	21.44[b]	(10.35)			18.04[b]	(7.467)	18.42	(12.40)
26 to 30th	2.606	(12.16)			7.538	(6.488)	3.294	(6.162)
31 to 35th	13.49[c]	(7.289)			-1.258	(6.240)	-13.62[b]	(6.005)
36 to 40th	20.90	(14.57)			18.14[c]	(10.27)	3.781	(12.36)
41 to 45th	17.49[b]	(8.018)			14.78	(11.98)	13.92	(9.101)
46 to 50th	-16.56[b]	(8.368)			-0.469	(7.433)	-2.922	(11.87)
Geog. neighbours								
1 to 5th			81.81[a]	(19.22)	76.21[a]	(19.44)	73.14[a]	(19.37)
6 to 10th			68.51[a]	(20.40)	58.60[a]	(19.39)	66.77[a]	(18.35)
11 to 15th			30.59	(23.19)	29.55	(20.01)	31.94[c]	(17.79)
16 to 20th			1.300	(20.28)	-1.008	(18.31)	0.657	(17.90)
21 to 25th			11.39	(15.58)	9.575	(14.76)	7.914	(15.72)
26 to 30th			56.01[b]	(22.69)	46.15[b]	(22.42)	47.98[b]	(21.62)
31 to 35th			54.15[b]	(26.75)	48.01[c]	(27.95)	56.20[c]	(29.00)
36 to 40th			52.21[b]	(23.11)	42.68[c]	(21.82)	34.94	(21.37)
41 to 45th			23.10[c]	(12.92)	21.26[c]	(12.68)	24.53[c]	(13.33)
46 to 50th			2.290	(16.35)	-6.821	(16.83)	0.824	(16.28)
R^2	0.30		0.29		0.30		0.30	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.10: Regression table for wages (IRS) using 2SLS

	(1) Social Dist.			(2) Geog. Dist.			(3) Res. Social Dist.			(4) Alt. Social Dist.		
	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F
Own county	21.6[a]	(4.32)	[125.7]	22.3[a]	(5.05)	[110.8]	20.8[a]	(4.68)	[210.6]	21.7[a]	(4.99)	[273.2]
Social neighbours												
1 to 5th	40.6[a]	(8.03)	[226]				5.09[a]	(1.70)	[729.9]	20.2[a]	(5.64)	[301.6]
6 to 10th	16.7[b]	(7.54)	[169.3]				5.57[b]	(2.61)	[878]	1.28	(5.90)	[436.9]
11 to 15th	24.9	(21.1)	[164.9]				-1.89	(2.93)	[670.5]	-3.82	(4.98)	[387.5]
16 to 20th	14.2	(8.90)	[153.1]				1.01	(3.26)	[303.1]	-10.2[b]	(5.16)	[771.1]
21 to 25th	12.2[b]	(5.65)	[460.4]				7.66[b]	(3.60)	[700]	7.74	(7.54)	[370.8]
26 to 30th	-6.37	(5.81)	[182.1]				2.78	(4.02)	[355.3]	-4.45	(3.90)	[814.8]
31 to 35th	3.99	(4.64)	[422.6]				-4.64	(3.51)	[133.6]	-7.70[b]	(3.25)	[639.5]
36 to 40th	7.34	(8.31)	[288.5]				5.96	(4.50)	[443.3]	3.76	(6.19)	[329.8]
41 to 45th	7.61	(5.66)	[411.7]				5.92	(4.88)	[766.6]	6.04	(5.03)	[437.9]
46 to 50th	-8.80[c]	(5.07)	[777.1]				-0.21	(4.21)	[640.1]	-6.15	(5.90)	[285.2]
Geog. neighbours												
1 to 5th				32.2[a]	(8.24)	[118]	31.8[a]	(9.06)	[340.8]	33.3[a]	(8.10)	[418.6]
6 to 10th				30.7[b]	(12.2)	[92.7]	29.6[b]	(11.7)	[236.1]	30.0[b]	(12.2)	[169.7]
11 to 15th				2.93	(10.4)	[247.2]	5.23	(9.93)	[419.1]	4.31	(9.83)	[374.7]
16 to 20th				-10.9	(11.0)	[245.7]	-9.42	(10.3)	[431.4]	-9.63	(10.9)	[375.8]
21 to 25th				0.22	(9.02)	[145.6]	2.64	(8.59)	[315]	0.051	(9.03)	[264.5]
26 to 30th				40.6	(31.1)	[81.7]	34.0	(29.7)	[139.5]	29.5	(27.5)	[139.9]
31 to 35th				25.0	(19.5)	[127.3]	21.0	(19.2)	[227.8]	26.3	(19.1)	[329]
36 to 40th				32.2[b]	(14.6)	[172.3]	29.1[b]	(14.3)	[373.2]	21.7	(15.6)	[408.3]
41 to 45th				16.0	(12.0)	[212.8]	17.8	(11.1)	[464.8]	19.7	(12.1)	[342.2]
46 to 50th				-13.1	(12.6)	[208.3]	-14.3	(11.9)	[310.4]	-12.1	(12.3)	[316]
R^2	0.04			0.04			0.04			0.04		
N	21,308			21,308			21,308			21,308		
First stage KP F	1.71			4.87			2.90			1.70		

SEs clustered by commuting zone. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

4.B.4 Regression Tables for Employment

Table 4.B.11: Regression table for employment using OLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	0.316^a	(0.0839)	0.304^a	(0.0779)	0.303^a	(0.0806)	0.303^a	(0.0749)
Social neighbours								
1 to 5th	0.462^a	(0.0955)			0.00888	(0.0145)	-0.00930	(0.0266)
6 to 10th	0.105^c	(0.0617)			-0.0344	(0.0228)	-0.0566	(0.0467)
11 to 15th	-0.0669	(0.0809)			0.0335	(0.0300)	0.0281	(0.0527)
16 to 20th	0.0984	(0.0630)			0.00574	(0.0265)	-0.0692	(0.0427)
21 to 25th	0.0241	(0.0695)			0.00126	(0.0291)	-0.0104	(0.0511)
26 to 30th	-0.0987^b	(0.0447)			0.00397	(0.0347)	-0.00443	(0.0447)
31 to 35th	-0.0434	(0.0413)			0.0125	(0.0362)	-0.00424	(0.0395)
36 to 40th	0.0700	(0.0576)			0.0240	(0.0345)	-0.0366	(0.0505)
41 to 45th	-0.0923^c	(0.0483)			-0.0184	(0.0414)	-0.0591	(0.0439)
46 to 50th	-0.0267	(0.0515)			0.0141	(0.0538)	0.0899^c	(0.0527)
Geog. neighbours								
1 to 5th			0.586^a	(0.163)	0.586^a	(0.146)	0.593^a	(0.123)
6 to 10th			0.139	(0.0978)	0.140	(0.128)	0.148	(0.117)
11 to 15th			0.0561	(0.0446)	0.0569	(0.0474)	0.0571	(0.0483)
16 to 20th			0.0794	(0.102)	0.0756	(0.0998)	0.0772	(0.103)
21 to 25th			-0.0604	(0.121)	-0.0572	(0.123)	-0.0499	(0.119)
26 to 30th			-0.0679	(0.0626)	-0.0669	(0.0599)	-0.0647	(0.0598)
31 to 35th			-0.120	(0.0845)	-0.128^c	(0.0750)	-0.0912	(0.0724)
36 to 40th			0.0654	(0.0810)	0.0586	(0.0840)	0.0766	(0.0891)
41 to 45th			0.0650	(0.0936)	0.0700	(0.0895)	0.0736	(0.0903)
46 to 50th			0.140	(0.0952)	0.152	(0.0937)	0.161^c	(0.0864)
R^2	0.20		0.20		0.20		0.20	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.12: Regression table for employment using OLS with county FEs

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	0.330^a	(0.0741)	0.305^a	(0.0703)	0.306^a	(0.0717)	0.306^a	(0.0708)
Social neighbours								
1 to 5th	0.414^a	(0.0823)			0.0274^b	(0.0137)	-0.00862	(0.0234)
6 to 10th	0.124^b	(0.0602)			-0.0692^a	(0.0229)	-0.0348	(0.0442)
11 to 15th	-0.121	(0.0962)			0.0313	(0.0324)	0.0545	(0.0501)
16 to 20th	0.0991	(0.0730)			-0.0123	(0.0308)	-0.0570	(0.0472)
21 to 25th	-0.00377	(0.0688)			-0.0119	(0.0313)	0.00739	(0.0555)
26 to 30th	-0.0932^b	(0.0422)			-0.00584	(0.0378)	0.0285	(0.0630)
31 to 35th	-0.0369	(0.0433)			0.0177	(0.0379)	-0.0725^c	(0.0434)
36 to 40th	0.0914	(0.0628)			0.0238	(0.0359)	-0.0782	(0.0651)
41 to 45th	-0.120^b	(0.0499)			-0.0167	(0.0380)	-0.0486	(0.0490)
46 to 50th	-0.00439	(0.0481)			0.00451	(0.0620)	0.0905	(0.0569)
Geog. neighbours								
1 to 5th			0.595^a	(0.137)	0.593^a	(0.122)	0.596^a	(0.116)
6 to 10th			0.165^c	(0.0972)	0.167	(0.106)	0.166	(0.121)
11 to 15th			0.00904	(0.0598)	0.0180	(0.0562)	0.00247	(0.0547)
16 to 20th			0.0758	(0.115)	0.0783	(0.110)	0.0688	(0.107)
21 to 25th			-0.150	(0.114)	-0.148	(0.120)	-0.162	(0.118)
26 to 30th			-0.141^b	(0.0689)	-0.144^b	(0.0733)	-0.148^b	(0.0671)
31 to 35th			-0.101	(0.0851)	-0.107	(0.0740)	-0.0757	(0.0719)
36 to 40th			0.0556	(0.0896)	0.0545	(0.0926)	0.0646	(0.0880)
41 to 45th			0.0766	(0.0969)	0.0781	(0.0985)	0.0698	(0.0989)
46 to 50th			0.189^b	(0.0873)	0.189^b	(0.0854)	0.204^b	(0.0817)
R^2	0.34		0.34		0.34		0.34	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.13: Regression table for employment, reduced form of 2SLS

	(1) Social Dist.		(2) Geog. Dist.		(3) Res. Social Dist.		(4) Alt. Social Dist.	
Own county	0.652^b	(0.281)	0.700^b	(0.283)	0.686^b	(0.303)	0.695^a	(0.226)
Social neighbours								
1 to 5th	2.274^a	(0.624)			0.0902^c	(0.0477)	0.0720	(0.0645)
6 to 10th	0.276	(0.339)			0.0905	(0.0827)	-0.229	(0.169)
11 to 15th	-0.322	(0.320)			-0.0881	(0.122)	0.0785	(0.184)
16 to 20th	0.524^c	(0.283)			0.240^c	(0.146)	-0.236	(0.161)
21 to 25th	-0.0943	(0.249)			-0.339^b	(0.132)	-0.0162	(0.235)
26 to 30th	-0.0730	(0.227)			-0.00257	(0.122)	-0.0966	(0.153)
31 to 35th	-0.0724	(0.177)			-0.0936	(0.155)	-0.0784	(0.169)
36 to 40th	0.102	(0.251)			0.315	(0.203)	0.0772	(0.319)
41 to 45th	-0.145	(0.176)			-0.0459	(0.146)	-0.212	(0.205)
46 to 50th	-0.209	(0.206)			-0.0335	(0.143)	0.0475	(0.261)
Geog. neighbours								
1 to 5th			1.383^c	(0.728)	1.364^b	(0.667)	1.410^b	(0.575)
6 to 10th			1.145^b	(0.561)	1.077^b	(0.510)	1.129^b	(0.450)
11 to 15th			0.475	(0.447)	0.515	(0.482)	0.519	(0.462)
16 to 20th			0.168	(0.419)	0.256	(0.521)	0.197	(0.469)
21 to 25th			-0.569	(0.538)	-0.550	(0.520)	-0.494	(0.534)
26 to 30th			-0.0968	(0.248)	-0.123	(0.264)	-0.0753	(0.261)
31 to 35th			-0.338	(0.443)	-0.305	(0.416)	-0.275	(0.440)
36 to 40th			0.970^b	(0.439)	0.896^b	(0.422)	0.948^b	(0.440)
41 to 45th			0.415^c	(0.251)	0.476^c	(0.274)	0.491	(0.304)
46 to 50th			0.215	(0.328)	0.233	(0.309)	0.286	(0.308)
R^2	0.18		0.18		0.18		0.18	
N	21,308		21,308		21,308		21,308	

SEs corrected for spatial clusters (Colella et al., 2019). Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.14: Regression table for employment using 2SLS

	(1) Social Dist.			(2) Geog. Dist.			(3) Res. Social Dist.			(4) Alt. Social Dist.		
	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F	Coef.	SE	SW F
Own county	0.36^a	(0.099)	[125.7]	0.34^a	(0.12)	[110.8]	0.35^a	(0.12)	[210.6]	0.35^a	(0.13)	[273.2]
Social neighbours												
1 to 5th	0.77^a	(0.19)	[226]				0.071^b	(0.029)	[729.9]	0.15	(0.097)	[301.6]
6 to 10th	0.074	(0.11)	[169.3]				0.042	(0.043)	[878]	-0.11	(0.074)	[436.9]
11 to 15th	-0.14	(0.13)	[164.9]				-0.018	(0.067)	[670.5]	0.045	(0.11)	[387.5]
16 to 20th	0.30	(0.21)	[153.1]				0.044	(0.059)	[303.1]	-0.12	(0.080)	[771.1]
21 to 25th	-0.016	(0.10)	[460.4]				-0.12^c	(0.064)	[700]	0.0067	(0.13)	[370.8]
26 to 30th	-0.15	(0.10)	[182.1]				-0.022	(0.070)	[355.3]	-0.12	(0.093)	[814.8]
31 to 35th	-0.015	(0.085)	[422.6]				-0.071	(0.091)	[133.6]	-0.053	(0.071)	[639.5]
36 to 40th	0.055	(0.11)	[288.5]				0.087	(0.063)	[443.3]	0.11	(0.17)	[329.8]
41 to 45th	-0.065	(0.094)	[411.7]				-0.010	(0.087)	[766.6]	-0.091	(0.095)	[437.9]
46 to 50th	-0.091	(0.11)	[777.1]				-0.00022	(0.068)	[640.1]	-0.011	(0.16)	[285.2]
Geog. neighbours												
1 to 5th				0.60^b	(0.25)	[118]	0.58^b	(0.26)	[340.8]	0.64^a	(0.24)	[418.6]
6 to 10th				0.50^b	(0.25)	[92.7]	0.50^b	(0.25)	[236.1]	0.49^c	(0.26)	[169.7]
11 to 15th				0.10	(0.14)	[247.2]	0.14	(0.14)	[419.1]	0.11	(0.14)	[374.7]
16 to 20th				-0.14	(0.19)	[245.7]	-0.089	(0.19)	[431.4]	-0.11	(0.20)	[375.8]
21 to 25th				-0.19	(0.16)	[145.6]	-0.17	(0.16)	[315]	-0.15	(0.17)	[264.5]
26 to 30th				0.030	(0.19)	[81.7]	-0.023	(0.18)	[139.5]	0.011	(0.21)	[139.9]
31 to 35th				-0.17	(0.27)	[127.3]	-0.18	(0.27)	[227.8]	-0.15	(0.27)	[329]
36 to 40th				0.51^c	(0.28)	[172.3]	0.48^c	(0.28)	[373.2]	0.51	(0.32)	[408.3]
41 to 45th				0.26	(0.18)	[212.8]	0.29	(0.18)	[464.8]	0.30	(0.19)	[342.2]
46 to 50th				0.11	(0.16)	[208.3]	0.16	(0.17)	[310.4]	0.13	(0.16)	[316]
R^2	0.06			0.06			0.06			0.06		
N	21,308			21,308			21,308			21,308		
First stage KP F	1.71			4.87			2.90			1.70		

SEs clustered by commuting zone. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

4.B.5 Regression Tables, First Stage Regressions for 2SLS

Table 4.B.15: New production per capita in bins of geographical neighbours

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	0	5	10	15	20	25	30	35	40	45	50
0	1.921^{a}	-0.0102	0.142	-0.0392	-0.106^{c}	-0.0222	-0.0654	0.00478	-0.00957	0.00811	-0.00467
	(0.405)	(0.0851)	(0.139)	(0.0654)	(0.0594)	(0.0194)	(0.0418)	(0.0200)	(0.00941)	(0.00732)	(0.0144)
5	-0.0968	2.399^{a}	-0.0390	0.00145	0.102	0.0470	0.105	-0.0755	-0.0303	0.0459^{c}	-0.0429
	(0.170)	(0.434)	(0.0984)	(0.188)	(0.111)	(0.0799)	(0.163)	(0.0503)	(0.0348)	(0.0271)	(0.0272)
10	0.841	-0.123	2.113^{a}	0.152	0.0551	0.0847	0.0616	0.0702	-0.0257	-0.0302	0.0267
	(0.610)	(0.152)	(0.431)	(0.455)	(0.233)	(0.219)	(0.173)	(0.144)	(0.0550)	(0.0376)	(0.0413)
15	-0.309	0.541^{b}	0.0287	3.732^{a}	0.310	-0.0693	0.294	-0.0842	0.0661	-0.101	-0.0267
	(0.368)	(0.240)	(0.165)	(0.379)	(0.256)	(0.164)	(0.201)	(0.116)	(0.0569)	(0.0717)	(0.0437)
20	1.022	-0.0760	0.518^{c}	-0.314	3.138^{a}	0.129	-0.00670	-0.167	0.214	-0.0513	-0.0681
	(0.725)	(0.131)	(0.304)	(0.243)	(0.581)	(0.149)	(0.0916)	(0.215)	(0.202)	(0.0981)	(0.0509)
25	-0.324	-0.245^{b}	0.267	-0.276^{c}	0.326	3.008^{a}	0.192	0.505	0.0674	-0.0301	-0.0672
	(0.277)	(0.123)	(0.343)	(0.161)	(0.400)	(0.307)	(0.189)	(0.374)	(0.108)	(0.0778)	(0.0701)
30	-0.0713	-0.215^{c}	0.00244	0.171	-0.284^{b}	-0.0106	1.759^{a}	-0.0106	-0.0287	-0.0503	0.0377
	(0.0874)	(0.124)	(0.0657)	(0.200)	(0.137)	(0.0479)	(0.422)	(0.127)	(0.0387)	(0.0591)	(0.0840)
35	0.0467	0.160	-0.150	0.325	-0.231	0.227	0.106	2.521^{a}	-0.193^{c}	0.0899	0.242
	(0.132)	(0.112)	(0.139)	(0.287)	(0.193)	(0.159)	(0.203)	(0.474)	(0.104)	(0.0945)	(0.215)
40	-0.0877	0.165^{c}	-0.247^{c}	-0.209^{b}	-0.252	-0.312^{b}	-0.139^{c}	-0.218^{b}	2.144^{a}	0.0875	0.111
	(0.173)	(0.0912)	(0.128)	(0.0935)	(0.173)	(0.147)	(0.0755)	(0.108)	(0.419)	(0.0698)	(0.118)
45	-0.0637	0.0752	0.0703	-0.0297	-0.00988	0.0376	-0.186^{c}	-0.0916	0.0633	1.634^{a}	0.0853
	(0.101)	(0.0474)	(0.118)	(0.0812)	(0.0974)	(0.127)	(0.108)	(0.0980)	(0.0730)	(0.266)	(0.187)
50	-0.0205	0.0133	0.0601	-0.122	0.0439	-0.0457	0.241^{c}	0.612^{c}	0.150	0.106	2.468^{a}
	(0.125)	(0.0566)	(0.0627)	(0.0847)	(0.118)	(0.102)	(0.131)	(0.328)	(0.159)	(0.141)	(0.420)
Lagged IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geog. IVs	No	No	No	No	No	No	No	No	No	No	No
R^2	0.5833	0.7590	0.6765	0.7162	0.6758	0.6934	0.5618	0.6679	0.6115	0.7119	0.6980
R^2 adj.	0.5827	0.7586	0.6761	0.7158	0.6754	0.6929	0.5612	0.6674	0.6110	0.7114	0.6976
F Stat.	78.28	42.34	48.09	34.00	40.16	44.82	42.33	44.19	55.53	73.44	62.13
SW F stat.	110.82	118.00	92.74	247.23	245.65	145.62	81.68	127.30	172.33	212.80	208.31
AP F stat.	327.24	280.36	109.02	384.51	191.78	363.40	57.17	194.44	300.00	195.45	190.35
N	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308

SEs clustered by commuting zone. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.16: New production per capita in bins of social neighbours

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	0	5	10	15	20	25	30	35	40	45	50
0	1.849^{a}	-0.0694	-0.120^{b}	-0.102^{c}	-0.0562^{c}	-0.0101	-0.0140	-0.00263	-0.0535	-0.0164	-0.0131
	(0.437)	(0.0754)	(0.0564)	(0.0585)	(0.0300)	(0.0371)	(0.0402)	(0.0240)	(0.0544)	(0.0344)	(0.0226)
5	0.999^{c}	2.919^{a}	-0.0356	0.122	0.0713	0.120	-0.0305	-0.153^{b}	-0.0277	-0.0542	-0.0899^{b}
	(0.515)	(0.385)	(0.104)	(0.0942)	(0.114)	(0.143)	(0.0710)	(0.0718)	(0.0436)	(0.0570)	(0.0354)
10	-0.161	0.157	2.673^{a}	0.0934	0.107	0.0576	-0.0151	-0.00604	-0.0544	-0.193^{a}	-0.0439
	(0.192)	(0.122)	(0.419)	(0.0745)	(0.0784)	(0.0704)	(0.0771)	(0.0533)	(0.0470)	(0.0529)	(0.0415)
15	-0.0156	-0.00106	0.320^{b}	2.498^{a}	0.0623	0.317^{a}	0.152	0.252^{c}	-0.0829	-0.115	0.0153
	(0.152)	(0.115)	(0.156)	(0.426)	(0.0513)	(0.120)	(0.103)	(0.146)	(0.0520)	(0.0888)	(0.0654)
20	0.00151	-0.0411	0.180	0.0408	2.363^{a}	-0.151^{b}	0.122	-0.104	0.0467	0.0412	-0.0362
	(0.0823)	(0.112)	(0.167)	(0.0772)	(0.287)	(0.0621)	(0.0839)	(0.0635)	(0.0682)	(0.0775)	(0.0387)
25	-0.214	-0.00662	-0.0267	-0.108	-0.0201	2.699^{a}	0.0173	0.0396	0.0737	-0.0605	-0.00176
	(0.165)	(0.100)	(0.0610)	(0.101)	(0.0430)	(0.292)	(0.0498)	(0.0589)	(0.0915)	(0.0505)	(0.0595)
30	0.466	0.129	0.00764	0.0419	0.0645	0.0702	2.537^{a}	0.121	-0.0303	0.386^{b}	-0.0870
	(0.334)	(0.0943)	(0.0834)	(0.0610)	(0.0435)	(0.0699)	(0.295)	(0.0987)	(0.0540)	(0.186)	(0.0711)
35	0.0331	0.000340	0.0438	0.0613	0.0202	0.0834	0.0747	3.028^{a}	0.0983	0.00147	0.0457
	(0.0649)	(0.0399)	(0.0553)	(0.0579)	(0.0612)	(0.0759)	(0.0902)	(0.276)	(0.0729)	(0.0755)	(0.0562)
40	-0.178	-0.0140	-0.0111	0.155	-0.00313	-0.0341	0.131	0.0641	2.913^{a}	0.0187	0.131
	(0.130)	(0.0353)	(0.0553)	(0.0977)	(0.0408)	(0.0366)	(0.117)	(0.0988)	(0.270)	(0.0657)	(0.0819)
45	-0.0433	-0.0754^{c}	-0.0415	-0.0695	0.0244	-0.00190	0.0786	0.373	0.0162	2.781^{a}	-0.0254
	(0.125)	(0.0395)	(0.0543)	(0.0708)	(0.0589)	(0.0476)	(0.0806)	(0.241)	(0.0444)	(0.283)	(0.0809)
50	0.0349	-0.0720^{c}	-0.0732^{b}	-0.0495	0.0294	0.0339	0.0474	0.0619	0.0895^{c}	0.135	2.919^{a}
	(0.0804)	(0.0379)	(0.0316)	(0.0310)	(0.0483)	(0.0493)	(0.0560)	(0.110)	(0.0493)	(0.119)	(0.259)
Lagged IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geog. IVs	No	No	No	No	No	No	No	No	No	No	No
R^2	0.5835	0.7087	0.7011	0.7066	0.6625	0.6894	0.6719	0.6895	0.7254	0.7023	0.6970
R^2 adj.	0.5829	0.7083	0.7007	0.7061	0.6620	0.6890	0.6715	0.6891	0.7251	0.7019	0.6965
F Stat.	52.96	37.58	58.96	52.26	86.17	66.35	59.19	73.12	73.67	54.43	86.91
SW F stat.	125.68	226.02	169.29	164.86	153.11	460.37	182.12	422.59	288.55	411.65	777.09
AP F stat.	208.39	442.00	151.98	301.14	114.69	467.88	159.99	687.42	281.43	324.55	561.81
N	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308

SEs clustered by commuting zone. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.17: New production per capita in bins of social neighbours (res.)

	(1) 0	(2) 5	(3) 10	(4) 15	(5) 20	(6) 25	(7) 30	(8) 35	(9) 40	(10) 45	(11) 50
0	1.916^{a}	-0.0197	-0.0521	0.0419	-0.0630	-0.0487	0.00655	-0.00366	-0.127	-0.0440^{b}	-0.0173
	(0.411)	(0.106)	(0.0884)	(0.0354)	(0.129)	(0.0338)	(0.0543)	(0.0307)	(0.107)	(0.0205)	(0.0190)
5	-0.00467	2.091^{a}	0.0700^{a}	0.0106	0.0206	0.0304^{b}	0.0180	0.0226^{b}	-0.00605	-0.0145	0.0120
	(0.0134)	(0.131)	(0.0264)	(0.0171)	(0.0139)	(0.0124)	(0.0111)	(0.0109)	(0.0108)	(0.00895)	(0.00867)
10	0.0193	0.133^{c}	2.829^{a}	0.0509	0.0857^{b}	0.0301	0.0873^{a}	0.0401^{c}	0.00667	0.0235	0.0329^{c}
	(0.0343)	(0.0783)	(0.137)	(0.0359)	(0.0376)	(0.0223)	(0.0338)	(0.0208)	(0.0254)	(0.0175)	(0.0193)
15	-0.0361	0.245^{a}	0.118^{c}	2.966^{a}	0.0482	0.109^{b}	0.00269	0.0453	-0.00605	-0.00764	-0.0531^{c}
	(0.0404)	(0.0936)	(0.0656)	(0.138)	(0.0574)	(0.0511)	(0.0376)	(0.0310)	(0.0374)	(0.0299)	(0.0290)
20	0.196	0.168	0.0786	0.0113	3.194^{a}	0.0397	-0.0328	0.0774^{c}	0.184^{b}	0.0438	0.0408
	(0.144)	(0.128)	(0.0920)	(0.0713)	(0.158)	(0.0345)	(0.0494)	(0.0411)	(0.0791)	(0.0277)	(0.0407)
25	-0.0818	-0.00428	0.0175	0.190^{c}	0.0204	3.427^{a}	0.0679	0.0903^{c}	0.0947	-0.0291	0.0136
	(0.0589)	(0.125)	(0.0921)	(0.0971)	(0.0463)	(0.176)	(0.0889)	(0.0496)	(0.0588)	(0.0408)	(0.0511)
30	-0.0266	-0.0207	0.105	0.168^{b}	0.0343	-0.0373	2.644^{a}	0.00556	-0.0419	0.0357	0.0228
	(0.0579)	(0.0903)	(0.0862)	(0.0819)	(0.0699)	(0.0682)	(0.287)	(0.0466)	(0.0315)	(0.0310)	(0.0283)
35	0.0235	0.0918	0.00510	0.0680	0.0407	0.0117	0.00504	3.123^{a}	0.114	0.0524	-0.0482^{c}
	(0.0852)	(0.147)	(0.103)	(0.0842)	(0.0707)	(0.0600)	(0.0384)	(0.310)	(0.0711)	(0.0393)	(0.0273)
40	0.151	0.295^{c}	-0.0650	0.0554	0.00797	0.150	-0.111^{a}	0.0168	3.373^{a}	-0.0154	0.0673
	(0.159)	(0.166)	(0.0727)	(0.110)	(0.0695)	(0.0956)	(0.0387)	(0.0372)	(0.213)	(0.0561)	(0.0510)
45	-0.0128	-0.144	-0.112^{c}	-0.0357	-0.0107	0.0487	-0.000547	-0.0713	0.0311	3.222^{a}	0.0914
	(0.0407)	(0.135)	(0.0642)	(0.0854)	(0.0533)	(0.101)	(0.0530)	(0.0649)	(0.0509)	(0.174)	(0.113)
50	-0.109	-0.148	-0.0486	-0.00119	0.0411	0.0587	0.00858	0.0355	0.000109	0.113	2.852^{a}
	(0.0726)	(0.106)	(0.0823)	(0.0757)	(0.0694)	(0.0754)	(0.0508)	(0.0499)	(0.0716)	(0.0965)	(0.212)
Lagged IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geog. IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.5848	0.7365	0.7266	0.6960	0.6959	0.7033	0.6845	0.6949	0.7088	0.6759	0.7052
R^2 adj.	0.5838	0.7359	0.7260	0.6953	0.6951	0.7026	0.6837	0.6942	0.7081	0.6751	0.7045
F Stat.	69.83	109.90	97.63	66.58	60.37	84.27	80.39	59.88	79.88	54.92	78.73
SW F stat.	210.60	729.90	878.00	670.50	303.10	700.00	355.30	133.60	443.30	766.60	640.10
AP F stat.	701.00	548.70	954.40	733.60	567.10	931.40	328.30	453.50	924.60	1,178.10	691.50
N	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308

SEs clustered by commuting zone. Sig. lev.: $^{a}p < 0.01$; $^{b}p < 0.05$; $^{c}p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Table 4.B.18: New production per capita in bins of social neighbours (>200km)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	0	5	10	15	20	25	30	35	40	45	50
0	1.918^a	-0.00993	0.0146	-0.0320	-0.0171	0.0484	-0.0434	0.00152	-0.0173	-0.0232	-0.0268
	(0.408)	(0.0338)	(0.0156)	(0.0445)	(0.0237)	(0.0424)	(0.0313)	(0.0159)	(0.0192)	(0.0143)	(0.0222)
5	-0.0107	1.432^a	0.0897^a	0.0290	-0.0643^c	0.0375	-0.00107	-0.00146	-0.0158^c	-0.0246^c	-0.0273^c
	(0.0275)	(0.145)	(0.0246)	(0.0289)	(0.0365)	(0.0298)	(0.0153)	(0.0141)	(0.00942)	(0.0137)	(0.0165)
10	-0.0568	0.244	2.513^a	0.201^c	0.120^b	0.0362	-0.0511	-0.0347	0.0242	-0.0512	-0.00932
	(0.0437)	(0.189)	(0.217)	(0.107)	(0.0576)	(0.0604)	(0.0414)	(0.0464)	(0.0276)	(0.0459)	(0.0510)
15	0.0313	0.173	0.0789	3.145^a	0.106	0.0597	0.111^c	0.0202	0.0560	0.0861	0.00393
	(0.0590)	(0.107)	(0.0551)	(0.279)	(0.0698)	(0.0506)	(0.0610)	(0.0567)	(0.0419)	(0.0558)	(0.0307)
20	0.0673	-0.0718	0.242^b	0.0211	3.236^a	0.0643	0.100^b	0.152^b	-0.0520	0.115	-0.0748
	(0.0783)	(0.141)	(0.0950)	(0.107)	(0.303)	(0.0473)	(0.0465)	(0.0693)	(0.0410)	(0.0975)	(0.0482)
25	-0.150^c	-0.0187	0.102	0.0635	-0.0429	2.964^a	0.191^c	-0.145	-0.0226	0.0939	0.105
	(0.0886)	(0.0977)	(0.0941)	(0.0622)	(0.0692)	(0.230)	(0.0990)	(0.0955)	(0.0527)	(0.0819)	(0.0657)
30	0.136	0.0748	0.0558	0.0747	0.155	0.0778	2.512^a	-0.0487	0.0902^c	0.0284	0.0635
	(0.116)	(0.0868)	(0.0602)	(0.0691)	(0.107)	(0.0542)	(0.197)	(0.0772)	(0.0503)	(0.0299)	(0.0503)
35	0.00329	0.00556	-0.0290	0.182	-0.00523	0.0183	0.0332	2.837^a	0.0937	-0.0454^c	-0.00829
	(0.0225)	(0.0493)	(0.0334)	(0.118)	(0.0672)	(0.0300)	(0.0559)	(0.274)	(0.0876)	(0.0238)	(0.0274)
40	-0.0698	-0.0497	-0.0246	0.188^c	-0.0456	0.0793	0.0937	0.0879	2.974^a	0.0553	0.132
	(0.0948)	(0.0706)	(0.0520)	(0.103)	(0.0514)	(0.0616)	(0.0743)	(0.148)	(0.272)	(0.0763)	(0.105)
45	-0.00367	0.153	-0.0528	-0.0453	0.0288	0.0141	-0.0994^b	0.0271	0.0244	2.940^a	0.0240
	(0.0373)	(0.0989)	(0.0540)	(0.0542)	(0.0501)	(0.0290)	(0.0459)	(0.0464)	(0.0498)	(0.265)	(0.0462)
50	0.0206	-0.00656	0.00339	-0.0327	0.0723	0.0125	0.0291	-0.0452	0.0257	0.248	3.074^a
	(0.0803)	(0.0847)	(0.0815)	(0.0452)	(0.0921)	(0.0514)	(0.0790)	(0.0341)	(0.0320)	(0.181)	(0.263)
Lagged IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geog. IVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.5839	0.7765	0.7128	0.6827	0.7065	0.6964	0.7218	0.7029	0.6879	0.6925	0.6789
R^2 adj.	0.5829	0.7759	0.7121	0.6819	0.7057	0.6956	0.7211	0.7022	0.6872	0.6917	0.6781
F Stat.	80.97	277.37	129.03	66.56	62.29	65.97	73.76	66.03	38.24	67.10	54.12
SW F stat.	273.20	301.60	436.90	387.50	771.10	370.80	814.80	639.50	329.80	437.90	285.20
AP F stat.	829.60	434.80	842.50	716.70	855.60	471.30	625.00	330.60	387.30	619.30	308.90
N	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308	21,308

SEs clustered by commuting zone. Sig. lev.: $^a p < 0.01$; $^b p < 0.05$; $^c p < 0.1$.

Note: Regressions also control for one-year lags of all variables and for 50-100th neighbours.

Bibliography

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016.

Aghion, P. and Howitt, P. W. (1992). A Model of Growth through Creative Destruction. *Econometrica*, 60(2):323–351.

Agrawal, A. K., Cockburn, I., and McHale, J. (2006). Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6(5):571–591.

Agrawal, A. K., Cockburn, I., and Rosell, C. (2010). Not Invented Here? Innovation in Company Towns. *Journal of Urban Economics*, 67(1):78–89.

Agrawal, A. K., Galasso, A., and Oettl, A. (2017). Roads and Innovation. *The Review of Economics and Statistics*, 99(3):417–434.

Agrawal, A. K., Kapur, D., and McHale, J. (2008). How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2):258–269.

Ahlfeldt, G. M., Bald, F., Roth, D., and Seidel, T. (2020). The stationary spatial equilibrium with migration costs. *Working Paper*.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., and Sterzi, V. (2018). Dancing with the Stars: Innovation Through Interactions. *National Bureau of Economic Research Working Paper Series*, No. 24466.

Akçomak, . S. and ter Weel, B. (2012). The impact of social capital on crime: Evidence from the Netherlands. *Regional Science and Urban Economics*, 42(1):323–340.

Akçomak, S. . and ter Weel, B. (2009). Social capital, innovation and growth: Evidence from Europe. *European Economic Review*, 53(5):544–567.

Alcácer, J. and Gittelman, M. (2006). Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations. *The Review of Economics and Statistics*, 88(4):774–779.

Alcácer, J., Gittelman, M., and Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, 38(2):415–427.

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

Alesina, A. and La Ferrara, E. (2000). Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, 115(3):847–904.

Allcott, H. and Keniston, D. (2017). Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, 85(2):695–731.

Amarasinghe, A., Hodler, R., Raschky, P. A., and Zenou, Y. (2018). Spatial Diffusion of Economic Shocks in Networks. *CESifo Working Papers*, No. 7001.

Amior, M. and Manning, A. (2018). The persistence of local joblessness. *American Economic Review*, 108(7):1942–1970.

Arribas-Bel, D. and Tranos, E. (2018). Big Urban Data: Challenges and Opportunities for Geographical Analysis. *Geographical Analysis*, 50(2):123–124.

Atkin, D., Chen, K., and Popov, A. (2019). The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley. *Working Paper*.

Audretsch, D. B. and Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, 86(3):630–640.

Audretsch, D. B. and Feldman, M. P. (2004). Knowledge Spillovers and the Geography of Innovation. In Henderson, J. V. and Thisse, J.-F. B. T., editors, *Handbook of Regional and Urban Economics: Cities and Geography*, volume 4, chapter 61, pages 2713–2739. Elsevier.

Autant-Bernard, C., Billand, P., Frachisse, D., and Massard, N. (2007a). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science*, 86(3):495–519.

Autant-Bernard, C., Mairesse, J., and Massard, N. (2007b). Spatial knowledge diffusion through collaborative networks. *Papers in Regional Science*, 86(3):341–350.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2016). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *National Bureau of Economic Research Working Paper Series*, No. 22637.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *American Economic Review*.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013a). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–2168.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013b). The Geography of Trade and Technology Shocks in the United States. *American Economic Review*, 103(3):220–225.

Autor, D. H., Dorn, D., Hanson, G. H., and Song, J. (2014). Trade Adjustment: Worker-Level Evidence. *The Quarterly Journal of Economics*, 129(4):1799–1860.

Autor, D. H., Palmer, C. J., and Pathak, P. A. (2019). Ending Rent Control Reduced Crime in Cambridge. *AEA Papers and Proceedings*, 109:381–384.

Azzi, C. and Ehrenberg, R. (1975). Household Allocation of Time and Church Attendance. *Journal of Political Economy*, 83(1):27–56.

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018a). The Economic Effects of Social Networks: Evidence from the Housing Market. *Journal of Political Economy*, 126(6):2224–2276.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018b). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, 32(3):259–280.

Bala, V. and Goyal, S. (2000). A Noncooperative Model of Network Formation. *Econometrica*, 68(5):1181–1229.

Banfield, E. C. (1958). *The moral basis of a backward society.* Glencoe, Ill., Free Press, Research Center in Economic Development and Cultural Change, University of Chicago, Chicago, IL.

Bartelsman, E., Scarpetta, S., and Schivardi, F. (2005). Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3):365–391.

Bartik, A. W., Currie, J., Greenstone, M., and Knittel, C. R. (2019). The Local Economic and Welfare Consequences of Hydraulic Fracturing. *American Economic Journal: Applied Economics*, 11(4):105–155.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Bathelt, H., Malmberg, A., and Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1):31–56.

Bayer, P., Ross, S. L., and Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, 116(6):1150–1196.

Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *The Review of Economic Studies*, 79(1):128–161.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 75(299):493.

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy*, 82(6):1063–1093.

Belenzon, S. and Schankerman, M. (2012). Spreading the Word: Geography, Policy, and Knowledge Spillovers. *The Review of Economics and Statistics*, 95(3):884–903.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J. (2018). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *The Quarterly Journal of Economics*, 134(2):647–713.

Berkes, E. and Gaetani, R. (2020). The Geography of Unconventional Innovation. *Working Paper - Conditionally accepted at The Economic Journal.*

Bessen, J. (2008). The value of U.S. patents by owner and patent characteristics. *Research Policy*, 37(5):932–945.

Beugelsdijk, S., de Groot, H. L. F., and van Schaik, A. B. T. M. (2004). Trust and economic growth: a robustness analysis. *Oxford Economic Papers*, 56(1):118–134.

Black, D., McKinnish, T., and Sanders, S. (2005). The Economic Impact Of The Coal Boom And Bust. *The Economic Journal*, 115(503):449–476.

Blanchard, O. J., Katz, L. F., Hall, R. E., and Eichengreen, B. (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, 1992(1):1–75.

Blind, I., Dahlberg, M., and Engströ, G. (2018). Introduction to the Special Issue 'On the Use of Geo-Coded Data in Economic Research'. *CESifo Economic Studies*, 64(2):123–126.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are Ideas Getting Harder to Find? *American Economic Review*, 110(4):1104–1144.

Borge, L. E., Parmer, P., and Torvik, R. (2015). Local natural resource curse? *Journal of Public Economics*, 131:101–114.

Bound, J. and Holzer, H. J. (2000). Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. *Journal of Labor Economics*, 18(1):20–54.

Bourdieu, P. (1986). The forms of capital. In *Cultural theory: An anthology*, chapter 8, pages 81–93. Wiley-Blackwell Hoboken, NJ.

Bowles, S. and Gintis, H. (2002). Social Capital And Community Governance. *The Economic Journal*, 112(483):F419–F436.

Breschi, S. and Lenzi, C. (2016). Co-invention networks and inventive productivity in US cities. *Journal of Urban Economics*, 92:66–75.

Breschi, S. and Lissoni, F. (2001). Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Industrial and Corporate Change*, 10(4):975–1005.

Breschi, S. and Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*, 9(4):439–468.

Breschi, S., Lissoni, F., and Montobbio, F. (2005). The Geography of Knowledge Spillovers: Conceptual Issues and Measurement Problems. In Breschi, S. and Malerba, F., editors, *Clusters, Networks and Innovation*, chapter 13, pages 343–378. Oxford University Press, Oxford, UK.

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

Brueckner, J. K. and Largey, A. G. (2008). Social interaction and urban sprawl. *Journal of Urban Economics*, 64(1):18–34.

Buechel, K., von Ehrlich, M., Puga, D., and Viladecans-Marsal, E. (2019). Calling from the Outside: The Role of Networks in Residential Mobility. *CEPR Discussion Paper Series*, No. DP1361.

Buera, F. J. and Oberfield, E. (2020). The Global Diffusion of Ideas. *Econometrica*, 88(1):83–114.

Burt, R. S. (1992). *Structural holes: The social structure of competition.* Harvard University Press, Cambridge, MA.

Burt, R. S. (2000). The Network Structure Of Social Capital. *Research in Organizational Behavior*, 22:345–423.

Calvó-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *American Economic Review*, 94(3):426–454.

Calvó-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3):389–419.

Carlino, G. A. and Kerr, W. R. (2015). Agglomeration and Innovation. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 6, pages 349–404. Elsevier.

Carvalho, V. M. (2014). From Micro to Macro via Production Networks. *Journal of Economic Perspectives*, 28(4):23–48.

Cascio, E. U. and Narayan, A. (2015). Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change. *National Bureau of Economic Research Working Paper Series*, No. 21359.

Caselli, F. and Michaels, G. (2013). Do oil windfalls improve living standards? Evidence from brazil. *American Economic Journal: Applied Economics*, 5(1):208–238.

Catalini, C. (2018). Microgeography and the Direction of Inventive Activity. *Management Science*, 64(9):4348–4364.

Catalini, C., Fons-Rosen, C., and Gaulé, P. (2020). How Do Travel Costs Shape Collaboration? *Management Science*.

Caulton, D. R., Shepson, P. B., Santoro, R. L., Sparks, J. P., Howarth, R. W., Ingraffea, A. R., Cambaliza, M. O., Sweeney, C., Karion, A., Davis, K. J., Stirm, B. H., Montzka, S. A., and Miller, B. R. (2014). Toward a better understanding and quantification of methane emissions from shale gas development. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17):6237–6242.

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2013). Manufacturing Decline, Housing Booms, and Non-Employment. *National Bureau of Economic Research Working Paper Series*, No. 18949.

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2016). The Masking of the Decline in Manufacturing Employment by the Housing Bubble. *Journal of Economic Perspectives*, 30(2):179–200.

Christopherson, S. and Rightor, N. (2012). How shale gas extraction affects drilling localities: Lessons for regional and city policy makers. *Journal of Town & City Management*, 2(4):350–368.

Cockburn, I. M., Kortum, S., and Stern, S. (2002). Are All Patent Examiners Equal? The Impact of Examiner Characteristics on Patent Statistics and Litigation Outcomes. *National Bureau of Economic Research Working Paper Series*, No. 8980.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1):128–152.

Colborn, T., Schultz, K., Herrick, L., and Kwiatkowski, C. (2014). An Exploratory Study of Air Quality Near Natural Gas Operations. *Human and Ecological Risk Assessment: An International Journal*, 20(1):86–105.

Colella, F., Lalive, R., Orcan Sakalli, S., and Thoenig, M. (2019). Inference with Arbitrary Clustering. *IZA Discussion Paper Series*, No. 12584.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94:S95–S120.

Comin, D. A., Dmitriev, M., and Rossi-Hansberg, E. (2012). The Spatial Diffusion of Technology. *National Bureau of Economic Research Working Paper Series*, No. 18534.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Corden, W. M. and Neary, J. P. (1982). Booming Sector and De-Industrialisation in a Small Open Economy. *The Economic Journal*, 92(368):825–848.

Costa, D. L. and Kahn, M. E. (2001). Understanding the Decline in Social Capital, 1952-1998. *National Bureau of Economic Research Working Paper Series*, No. 8295.

Crescenzi, R., Filippetti, A., and Iammarino, S. (2017). Academic inventors: collaboration and proximity with industry. *The Journal of Technology Transfer*, 42(4):730–762.

Crescenzi, R., Gagliardi, L., and Percoco, M. (2013a). Social capital and the innovative performance of Italian provinces. *Environment and Planning A*, 45(4):908–929.

Crescenzi, R., Gagliardi, L., and Percoco, M. (2013b). The 'Bright' Side of Social Capital: How 'Bridging' Makes Italian Provinces More Innovative. In Crescenzi, R. and Percoco, M., editors, *Geography, Institutions and Regional Economic Performance. Advances in Spatial Science (The Regional Science Series)*, pages 143–164. Springer, Berlin, Heidelberg, DE.

Crescenzi, R., Nathan, M., and Rodríguez-Pose, A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, 45(1):177–194.

Cunningham, S., DeAngelo, G., and Smith, B. (2020). Fracking and risky sexual activity. *Journal of Health Economics*, 72:102322.

Dasgupta, P. (2005). Economics of Social Capital. *The Economic Record*, 81(255):S2–S21.

Davis, D. R. and Dingel, J. I. (2019). A Spatial Knowledge Economy. *American Economic Review*, 109(1):153–170.

De la Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.

de Tocqueville, A. (1835). *Democracy in America: Historical-Critical Edition of "De la démocratie en Amérique"*. Liberty Fund, Inc., Indianapolis, IN.

D'Este, P., Guy, F., and Iammarino, S. (2013). Shaping the formation of university-industry research collaborations: What type of proximity does really matter? *Journal of Economic Geography*, 13(4):537–558.

DiPasquale, D. and Glaeser, E. L. (1999). Incentives and Social Capital: Are Homeowners Better Citizens? *Journal of Urban Economics*, 45(2):354–384.

DOE (2009). Modern Shale Gas Development in the United States: A Primer, prepared for the US Department of Energy. Technical report, U.S. Department of Energy (DOE), Office of Fossil Energy and National Energy Technology Laboratory, Washington, DC.

Duranton, G. and Puga, D. (2001). Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review*, 91(5):1454–1477.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In Henderson, J. V. and Thisse, J.-F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 48, pages 2063–2117. Elsevier B.V.

Durkheim, E. (1893). *The Division of Labour in Society*. Macmillan and Co., Ltd., London, UK.

Durlauf, S. N. (2002). On the Empirics of Social Capital. *The Economic Journal*, 112:459–479.

Durlauf, S. N. and Fafchamps, M. (2005). Social Capital. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 1B, chapter 26, pages 1639–1699. Elsevier B.V.

Dustmann, C., Glitz, A., Schönberg, U., and Brücker, H. (2016). Referral-based Job Search Networks. *The Review of Economic Studies*, 83(2):514–546.

EIA (2018). US Crude Oil and Natural Gas Proved Reserves, Year-End 2017. Technical report, US Energy Information Administration (EIA), Washington, DC.

Faggio, G. and Overman, H. G. (2014). The effect of public sector employment on local labour markets. *Journal of Urban Economics*, 79(December 2011):91–107.

Feldman, M. P. (2002). The Internet revolution and the geography of innovation. *International Social Science Journal*, 54(171):47–56.

Feldman, M. P. and Audretsch, D. B. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2):409–429.

Fetzer, T. (2014). Fracking Growth. *CEP Discussion Paper*, No. 1278.

Feyrer, J., Mansur, E. T., and Sacerdote, B. (2017). Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution. *American Economic Review*, 107(4):1313–1334.

Fontenot, B. E., Hunt, L. R., Hildenbrand, Z. L., Carlton, D. D., Oka, H., Walton, J. L., Hopkins, D., Osorio, A., Bjorndal, B., Hu, Q. H., and Schug, K. A. (2013). An evaluation of water quality in private drinking water wells near natural gas extraction sites in the Barnett shale formation. *Environmental Science and Technology*, 47(17):10032–10040.

Fort, T. C., Pierce, J. R., and Schott, P. K. (2018). New perspectives on the decline of US manufacturing employment. *Journal of Economic Perspectives*, 32(2):47–72.

Fukuyama, F. (2001). Social capital, civil society and development. *Third World Quarterly*, 22(1):7–20.

Gagliardi, L., Iammarino, S., and Rodríguez-Pose, A. (2015). Offshoring and the Geography of Jobs in Great Britain. *CEPR Discussion Paper No. DP10855*.

Gee, L. K., Jones, J., and Burke, M. (2017). Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network. *Journal of Labor Economics*, 35(2):485–518.

Gibbons, S., Heblich, S., Lho, E., and Timmins, C. (2016). Fear of Fracking? The Impact of the Shale Gas Exploration on House Prices in Britain. *National Bureau of Economic Research Working Paper Series*, No. 22859.

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

Gibbons, S., Overman, H. G., and Patacchini, E. (2015). Spatial Methods. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 3, pages 115–168. Elsevier.

Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 211–220, New York, New York, USA. ACM Press.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*, 46(2):254–277.

Glaeser, E. L. and Gyourko, J. (2005). Urban Decline and Durable Housing. *Journal of Political Economy*, 113(2):345–375.

Glaeser, E. L., Laibson, D., and Sacerdote, B. (2002). An Economic Approach to Social Capital. *The Economic Journal*, 112(483):F437–F458.

Goyal, S. (2011). Learning in Networks. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1A, chapter 15, pages 679–727. North-Holland, Amsterdam, NL.

Graham, J., Irving, J., Tang, X., Sellers, S., Crisp, J., Horwitz, D., Muehlenbachs, L., Krupnick, A., and Carey, D. (2015). Increased traffic accident rates associated with shale gas drilling in Pennsylvania. *Accident Analysis and Prevention*, 74:203–209.

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

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1:201–233.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3):481–510.

Granovetter, M. (2005). The Impact of Social Structure on Economic Outcomes. *The Journal of Economic Perspectives*, 19(1):33–50.

Greenwood, S., Perrin, A., and Duggan, M. (2016). Social Media Update 2016. Technical report, Pew Research Center, Washington, DC.

Griliches, Z. (1998). Patent Statistics as Economic Indicators: A Survey. In *R&D and Productivity: The Econometric Evidence*, pages 287–343. University of Chicago Press, Chicago, IL.

Guiso, L., Sapienza, P., and Zingales, L. (2004). The Role of Social Capital in Financial Development. *American Economic Review*, 94(3):526–556.

Hampton, K. N., Goulet, L. S., Rainie, L., and Purcell, K. (2011). Social Networking Sites and Our Lives. Technical report, Pew Research Center, Washington, DC.

Helper, S., Krueger, T., and Wial, H. (2012). Why Does Manufacturing Matter? Which Manufacturing Matters? A Policy Framework. *Metropolitan Policy Program at Brookings*, (February):1–53.

Helsley, R. W. and Strange, W. C. (2004). Knowledge barter in cities. *Journal of Urban Economics*, 56(2):327–345.

Henderson, J. V. (2007). Understanding knowledge spillovers. *Regional Science and Urban Economics*, 37(4):497–508.

Hilber, C. A. L. (2010). New housing supply and the dilution of social capital. *Journal of Urban Economics*, 67(3):419–437.

Hirschman, A. O. (1958). *The Strategy of Economic Development*. Yale University Press, New Haven, CT.

Howarth, R. W. and Ingraffea, A. (2011). Natural gas: Should fracking stop? *Nature*, 477(7364):271–273.

Ioannides, Y. M. (2013). Introduction. In *From Neighborhoods to Nations*, The Economics of Social Interactions, chapter 1, pages 1–10. Princeton University Press, Princeton, NJ.

Iyer, S., Kitson, M., and Toh, B. (2005). Social capital, economic growth and regional development. *Regional Studies*, 39(8):1015–1040.

Jackson, M. O. (2020). A typology of social capital and associated network measures. *Social Choice and Welfare*, 54(2-3):311–336.

Jackson, M. O., Rogers, B., and Zenou, Y. (2017). The Economic Consequences of Social Network Structure. *Journal of Economic Literature*, 55(1):1–47.

Jackson, R. B., Vengosh, A., Darrah, T. H., Warner, N. R., Down, A., Poreda, R. J., Osborn, S. G., Zhao, K., and Karr, J. D. (2013). Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28):11250–11255.

Jacobsen, G. D. and Parker, D. P. (2016). The Economic Aftermath of Resource Booms: Evidence from Boomtowns in the American West. *The Economic Journal*, 126(593):1092–1128.

Jacquet, J. (2011). Workforce development challenges in the natural gas industry. *A comprehensive economic impact analysis of natural gas extraction in the Marcellus Shale - Working Paper Series.*

Jaffe, A. B. and de Rassenfosse, G. (2017). Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, 68(6):1360–1374.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *The American Economic Review*, 90(2):215–218.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2002). The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees. In Jaffe, A. B. and Trajtenberg, M., editors, *Patents, citations, and innovations: A window on the knowledge economy*, chapter 12. MIT press, Cambridge, MA.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3):577–598.

Jahoda, M., Lazarsfeld, P. F., and Zeisel, H. (1971). *Marienthal: The sociography of an unemployed community.* Transaction Publishers, New Brunswick, NJ.

James, A. and Smith, B. (2017). There will be blood: Crime rates in shale-rich U.S. counties. *Journal of Environmental Economics and Management*, 84:125–152.

James, A. G. and Smith, B. (2020). Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution: Comment. *American Economic Review*, 110(6):1905–1913.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *Review of Economic Studies*, 76(1):283–317.

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., and Fowler, J. H. (2013). Inferring Tie Strength from Online Directed Behavior. *PLOS ONE*, 8(1):1–6.

Kaplan, G. (2012). Moving Back Home: Insurance against Labor Market Risk. *Journal of Political Economy*, 120(3):446–512.

Katz, L. F., Summers, L. H., Hall, R. E., Schultze, C. L., and Topel, R. H. (1989). Industry Rents: Evidence and Implications. *Brookings Papers on Economic Activity. Microeconomics*, 1989:209.

Kearney, M. S. and Wilson, R. (2018). Male earnings, marriageable men, and nonmarital fertility: Evidence from the fracking boom. *Review of Economics and Statistics*, 100(4):678–690.

Keefer, P. and Knack, S. (2008). Social Capital, Social Norms and the New Institutional Economics. In Ménard, C. and Shirley, M. M., editors, *Handbook of New Institutional Economics*, chapter 27, pages 701–725. Springer Berlin Heidelberg.

Kerr, W. R. (2008). Ethnic scientific communities and international technology diffusion. *Review of Economics and Statistics*, 90(3):518–537.

Kerr, W. R. and Kominers, S. D. (2014). Agglomerative Forces and Cluster Shapes. *The Review of Economics and Statistics*, 97(4):877–899.

Kim, J. S., Patacchini, E., Zenou, Y., and Picard, P. M. (2020). Spatial Interactions. *Working Paper*, pages 1–56.

Knack, S. (2002). Social Capital and the Quality of Government: Evidence from the States. *American Journal of Political Science*, 46(4):772–785.

Knack, S. and Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics*, 112(4):1251–1288.

Koster, H. R. and van Ommeren, J. (2015). A shaky business: Natural gas extraction, earthquakes and house prices. *European Economic Review*, 80:120–139.

Krugman, P. (1991a). *Geography and trade*. MIT press, Cambridge, MA.

Krugman, P. (1991b). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499.

Krugman, P. and Venables, A. J. (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, 110(4):857–880.

Lederman, D., Loayza, N., and Menéndez, A. M. (2002). Violent Crime: Does Social Capital Matter? *Economic Development and Cultural Change*, 50(3):509–539.

Leip, D. (2016). Dave Leip's Atlas of US Presidential Elections.

Lin, N. (1999). Building a Network Theory of Social Capital. *Connections*, 22(1):28–51.

Lissoni, F. (2018). International migration and innovation diffusion: an eclectic survey. *Regional Studies*, 52(5):702–714.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Lucas, R. E. and Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1–51.

Lychagin, S., Pinkse, J., Slade, M. E., and Van Reenen, J. (2016). Spillovers in Space: Does Geography Matter? *The Journal of Industrial Economics*, 64(2):295–335.

Maclaurin, W. R. (1950). The Process of Technological Innovation: The Launching of a New Scientific Industry. *The American Economic Review*, 40(1):90–112.

Maggioni, M. A., Nosvelli, M., and Uberti, E. T. (2007). Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science*, 86(3):471–493.

Manning, A. and Petrongolo, B. (2017). How local are labor markets? Evidence from a spatial job search model. *American Economic Review*, 107(10).

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.

Marshall, A. (1890). *Principles of Economics*. Macmillan and Co., Ltd., London.

McQuarrie, M. (2017). The revolt of the Rust Belt: place and politics in the age of anger. In *British Journal of Sociology*, volume 68, pages S120–S152. Blackwell Publishing Ltd.

Michaels, G. (2011). The Long Term Consequences of Resource-Based Specialisation. *The Economic Journal*, 121(551):31–57.

Miguel, E., Gertler, P., and Levine, D. I. (2006). Does Industrialization Build or Destroy Social Networks? *Economic Development and Cultural Change*, 54(2):287–317.

Miguélez, E. and Moreno, R. (2015). Knowledge flows and the absorptive capacity of regions. *Research Policy*, 44(4):833–848.

Molloy, R., Smith, C. L., and Wozniak, A. (2011). Internal Migration in the United States. *The Journal of Economic Perspectives*, 25(3):173–196.

Molloy, R., Smith, C. L., and Wozniak, A. K. (2014). Declining Migration within the U.S.: The Role of the Labor Market. *National Bureau of Economic Research Working Paper Series*, No. 20065.

Molloy, R., Trezzi, R., Smith, C. L., and Wozniak, A. (2016). Understanding declining fluidity in the U.S. labor market. *Brookings Papers on Economic Activity*, 2016(SPRING):183–259.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, Migration and Local Employment Elasticities. Technical report.

Moretti, E. (2010). Local Multipliers. *American Economic Review*, 100(2):373–377.

Moretti, E. (2011). Local Labor Markets. *Handbook of Labor Economics*, 4:1237–1313.

Muehlenbachs, L., Spiller, E., and Timmins, C. (2015). The housing market impacts of shale gas development. *American Economic Review*, 105(12):3633–3659.

Murata, Y., Nakajima, R., Okamoto, R., and Tamura, R. (2013). Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *The Review of Economics and Statistics*, 96(5):967–985.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris, FR.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708.

North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, Cambridge, UK.

Notowidigdo, M. J. (2011). The Incidence of Local Labor Demand Shocks. *National Bureau of Economic Research Working Paper Series*, No. 17167.

Olmstead, S. M., Muehlenbachs, L. A., Shih, J. S., Chu, Z., and Krupnick, A. J. (2013). Shale gas development impacts on surface water quality in Pennsylvania. *Proceedings of the National Academy of Sciences of the United States of America*, 110(13):4962–4967.

Olson, M. (1982). *The Rise and Decline of Nations*. Yale University Press, New Haven, CT.

Openshaw, S. and Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, 21:127–144.

Patacchini, E. and Zenou, Y. (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6):938–949.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics*, 7(1-2):77–99.

Paxton, P. (1999). Is Social Capital Declining in the United States? A Multiple Indicator Assessment. *American Journal of Sociology*, 105(1):88–127.

Percoco, M. (2012a). Entrepreneurship, Social Capital and Institutions: Evidence from Italy. *Spatial Economic Analysis*, 7(3):339–355.

Percoco, M. (2012b). Mining Activity and the Setting up of New Businesses in Basilicata. *Rivista economica del Mezzogiorno*, XXVI(4/2012):881–898.

Percoco, M. (2015). Entrepreneurship, Family Ties, and Land Inequality: Evidence from Italy. *Growth and Change*, 46(3):443–457.

Percoco, M. (2016). Highways, local economic structure and urban development. *Journal of Economic Geography*, 16(5):1035–1054.

Polanyi, K. (1945). *The Great Transformation: The Political and Economic Origins of our Time*. Beacon Press, Boston, MA.

Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press, Chicago, IL.

Poorthuis, A. (2018). How to Draw a Neighborhood? The Potential of Big Data, Regionalization, and Community Detection for Understanding the Heterogeneous Nature of Urban Neighborhoods. *Geographical Analysis*, 50(2):182–203.

Portes, A. (1998). Social Capital: Its Origins and Applications in Modern Sociology. *Annual Review of Sociology*, 24(1):1–24.

Powell, W. W. and Grodal, S. (2005). Networks of Innovators. In Fagerberg, J., Mowery, D. C., and Nelson, R. R., editors, *The Oxford Handbook of Innovation*, chapter 3, pages 56–85. Oxford University Press.

Putnam, R. D. (1995). Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America. *PS: Political Science & Politics*, 28(4):664–683.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York, NY.

Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1993). *Making Democracy Work*. Princeton University Press, Princeton, NJ.

Rickman, D. S., Wang, H., and Winters, J. V. (2017). Is shale development drilling holes in the human capital pipeline? *Energy Economics*, 62:283–290.

Righi, C. and Simcoe, T. (2019). Patent examiner specialization. *Research Policy*, 48(1):137–148.

Roche, M. P. (2019). Taking Innovation to the Streets: Microgeography, Physical Structure and Innovation. *The Review of Economics and Statistics*, pages 1–47.

Rodríguez-Pose, A. and Storper, M. (2006). Better Rules or Stronger Communities? On the Social Foundations of Institutional Change and Its Economic Effects. *Economic Geography*, 82(1):1–25.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5):1002–1037.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5, Part 2):S71–S102.

Roy, A. A., Adams, P. J., and Robinson, A. L. (2014). Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. *Journal of the Air and Waste Management Association*, 64(1):19–37.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2020). *IPUMS USA: Version 10.0 [dataset]*. IPUMS, Minneapolis, MN.

Rupasingha, A., Goetz, S. J., and Freshwater, D. (2006). The production of social capital in US counties. *Journal of Socio-Economics*, 35(1):83–101.

Sachs, J. D. and Warner, A. M. (1995). Natural Resource Abundance and Economic Growth. *National Bureau of Economic Research Working Paper Series*, No. 5398.

Sachs, J. D. and Warner, A. M. (2001). The curse of natural resources. *European Economic Review*, 45(4-6):827–838.

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science*, 277(5328):918 LP – 924.

Saxenian, A. (1996). *Regional Advantage*. Harvard University Press, Cambridge, MA.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons. Technical report, World Intellectual Property Organization, Geneva, Switzerland.

Schmutz, B. and Sidibé, M. (2019). Frictional Labour Mobility. *The Review of Economic Studies*, 86(4):1779–1826.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5):80–93.

Smith, A. (1869). *The Theory of Moral Sentiments and Essays on Philosophical Subjects*. Alex. Murray & Son, London, UK, sixth edition.

Sonis, M. and Hewings, G. J. (1992). Coefficient Change in Input-Output Models: Theory and Applications. *Economic Systems Research*, 4(2):143–158.

Sonn, J. W. and Storper, M. (2008). The Increasing Importance of Geographical Proximity in Knowledge Production: An Analysis of US Patent Citations, 1975–1997. *Environment and Planning A: Economy and Space*, 40(5):1020–1039.

Sorenson, O., Rivkin, J. W., and Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, 35(7):994–1017.

Storper, M. (2005). Society, Community, and Economic Development. *Studies in Comparative International Development*, 39(4):30–57.

Storper, M. and Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4):351–370.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, 8(4):677–716.

Thompson, P. (2006). Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations. *The Review of Economics and Statistics*, 88(2):383–388.

Thompson, P. and Fox-Kean, M. (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *The American Economic Review*, 95(1):450–460.

Tolbert, C. M. and Sizer, M. (1996). US commuting zones and labor market areas: A 1990 update. *Economic Research Service Staff Paper*, No. AGES-9.

Tönnies, F. and Loomis, C. P. (1957). *Community and society*. Courier Corporation.

Topa, G. (2011). Labor Markets and Referrals. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1A, chapter 22, pages 1193–1221. North-Holland, Amsterdam, NL.

Topa, G. and Zenou, Y. (2015). Neighborhood and Network Effects. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 9, pages 561–624. Elsevier, Amsterdam, NL.

Tranos, E. (2016). The Internet: its geography, growth and the creation of (digital) social capital. In Shearmu, R., Carrincazeaux, C., and Doloreux, D., editors, *Handbook on the Geographies of Innovation*, chapter 21, pages 356–369. Edward Elgar Publishing, Cheltenham, UK.

Vengosh, A., Warner, N., Jackson, R., and Darrah, T. (2013). The Effects of Shale Gas Exploration and Hydraulic Fracturing on the Quality of Water Resources in the United States. *Procedia Earth and Planetary Science*, 7:863–866.

Wang, Z. and Krupnick, A. (2015). A Retrospective Review of Shale Gas Development in the United States: What Led to the Boom? *Economics of Energy & Environmental Policy*, 4(1):5–18.

Warner, N. R., Christie, C. A., Jackson, R. B., and Vengosh, A. (2013). Impacts of shale gas wastewater disposal on water quality in Western Pennsylvania. *Environmental Science and Technology*, 47(20):11849–11857.

Weber, J. G. (2014). A decade of natural gas development: The makings of a resource curse? *Resource and Energy Economics*, 37:168–183.

Yan, B. and Luo, J. (2017). Measuring technological distance for patent mapping. *Journal of the Association for Information Science & Technology*, 68(2):423–437.

Zacchia, P. (2019). Knowledge Spillovers through Networks of Scientists. *The Review of Economic Studies*, 0:1–30.

Zak, P. J. and Knack, S. (2001). Trust and Growth. *The Economic Journal*, 111(470):295–321.