Department of
Geography and Environment

Papers in Economic Geography and Spatial Economics

Spatial Diffusion of Local Economic Shocks in Social Networks:

Evidence from the US Fracking Boom

Andreas Diemer

Paper No. 13
Geography and Environment Discussion Paper Series

August 2020

Published by
Department of Geography and Environment
London School of Economics and Political Science
Houghton Street
London
WC2A 2AE

geog.comms@lse.ac.uk
www.lse.ac.uk/Geography-and-Environment

Spatial Diffusion of Local Economic Shocks in Social Networks: Evidence from the US Fracking Boom

Andreas Diemer*

July 31, 2020

Abstract

There is little evidence on the relevance of social networks in the aggregate spatial diffusion of localised economic shocks. This paper uses novel data on the universe of online friendships in the US to uncover how 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. Although most of the diffusion is limited to geographically proximate areas, social networks matter too. According to 2SLS estimates, a million dollar per capita increase in oil and gas extraction raises per capita wages by over 5,000 dollars for workers reporting their incomes in counties located as far as 1,200 km away from the drilling site, but strongly socially connected to it. 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.

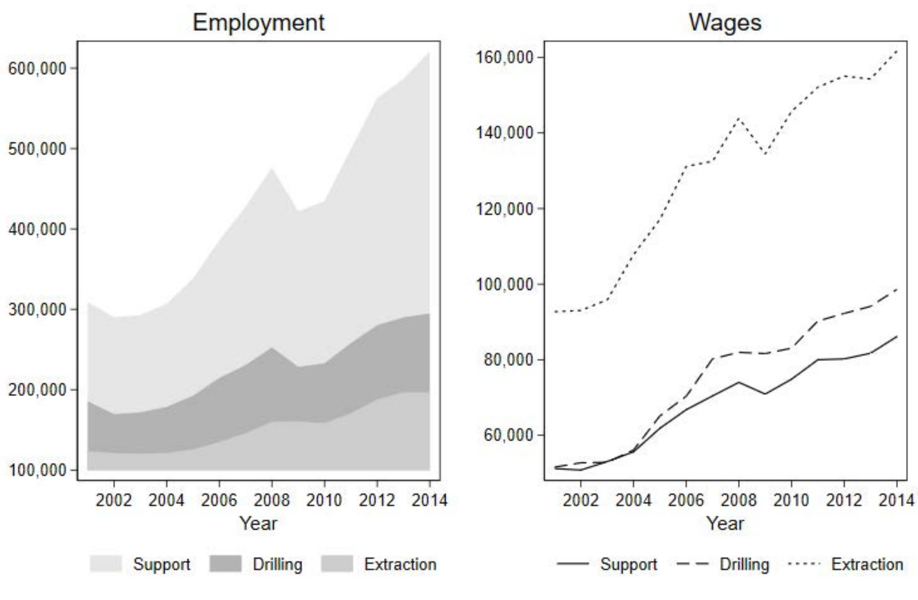
Keywords: Social networks; Fracking; Spatial Diffusion; Job search.
JEL Codes: J61; J64; L71; Q33; R12; R23; Z13.

*London School of Economics and Political Science, Department of Geography and Environment, Houghton Street, London WC2A 2AE, United Kingdom. Email: a.diemer@lse.ac.uk. I am grateful to Simona Iammarino and Michael Storper for their supervision and excellent guidance. This paper also benefited from helpful comments and suggestions received at the LSE PhD Work in Progress Seminars, the 7th Central European Conference in Regional Science in Sopron, the 24th 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 also thank Mike Bailey at Facebook for access to the Social Connectedness data.

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 1). Feyrer et al. (2017) emphasise that this activity is highly localised, making it a suitable case for the analysis proposed herein.

Figure 1: Growth in employment and annual average pay in private sector firms



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

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 (2012) for Italy, and by Gibbons et al. (2016) for the UK. The latter paper offers evidence on fracking, which is rare for Europe

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 con-

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.

tiguous 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]

Empirically, this can also be seen in the aggregate data used for the analysis in this paper. The binned scatterplot in Figure 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 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

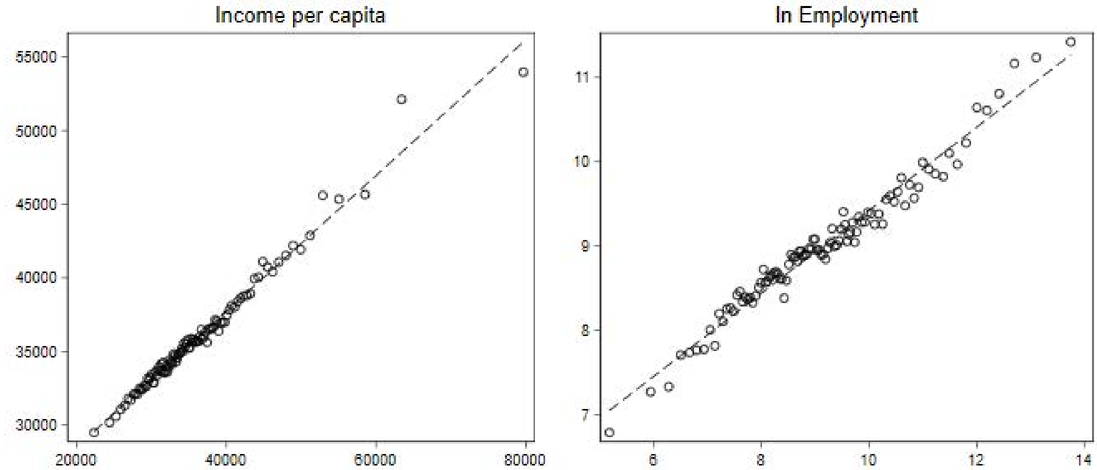
[2]In a related paper, Bailey et al. (2018a) use micro-data on social connections to show that exposure to fluctuations in housing prices via one's network influences beliefs about attractiveness of property investments and ultimately housing market activity of this individual.

[3]See Topa (2011) and Topa and Zenou (2015) for recent reviews of this literature.

[4]Social connections are defined in terms of number of online friendships the counties share. More information on these data are available in Section 3.

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.

Figure 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

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 2, and building on the existing literature 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 2 conceptualises the role of social networks in the transmission of economic shocks across local markets. Section 3 outlines the empirical strategy adopted by this paper and presents the econometric model. Section 4 discusses the main results. Section 5 concludes.

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

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-

[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.

[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.

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 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.

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

[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.

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 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.

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

[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.

there is now 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:

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

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:

$$F(z_{ij,o}) = e^{-T_i E_j z_{ij,o}^{-\epsilon}}, \qquad T_i, E_j > 0, \epsilon > 1 \tag{2}$$

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:

$$\pi_{ij} = \frac{T_i E_j (d_{ij} Q_i^{1-\beta})^{-\epsilon} (B_i w_j)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s (d_{rs} Q_r^{1-\beta})^{-\epsilon} (B_r w_s)^{\epsilon}} \tag{3}$$

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 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:

$$\pi_{ij|i} = \frac{E_j(w_j/d_{ij})^{\epsilon}}{\sum_{s=1}^{S} E_s(w_s/d_{is})^{\epsilon}} \tag{4}$$

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:

$$\mathbb{E}[w_j|i] = \sum_{j=1}^{S} \pi_{ij|i} w_j \tag{5}$$

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) implies a semi-log gravity commuting equation:

$$\ln \pi_{ij} = -\delta \tau_{ij} + \gamma \sigma_{ij} + \varsigma_j, \qquad \delta = \epsilon\kappa, \gamma = \epsilon\eta \tag{6}$$

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 (5) consistent with the relationship highlighted in (6) expresses wages observed in region i as a weighted average of wages in all other connected locations:

$$\Delta w_{i,t} = \gamma \times m(\Delta w, s)_{i,t} + \epsilon_{i,t} \tag{7}$$

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 (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.

3 Data and Empirical Methods

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.

3.1.1 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) 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

[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).

commuting and transient workers, who indeed appear to represent the bulk of earners in the industry.

3.1.2 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]

$$\Delta X_{i,t} = \frac{\Delta Q_{i,t}^{oil} \times P_{i,t}^{oil} + \Delta Q_{i,t}^{gas} \times P_{i,t}^{gas}}{L_{i,t-1}} \tag{8}$$

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 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 3 maps quintiles of the total value of new production of oil and gas per capita over the 2005-2012 period.

3.1.3 Social Network Matrices and Socially Lagged Shocks

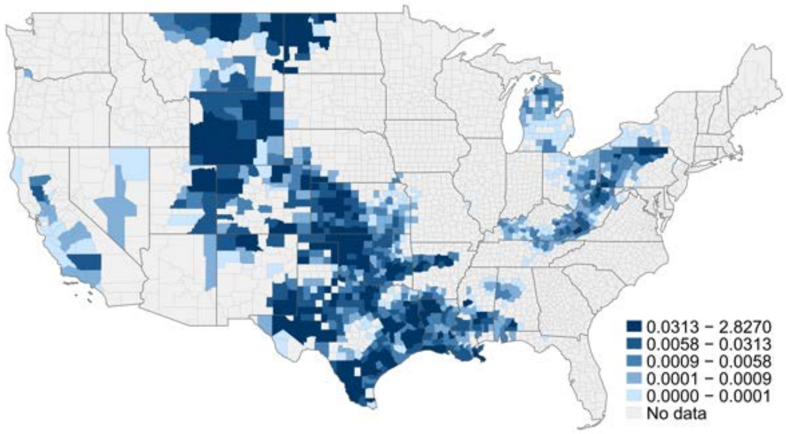
The proposed measure of social networks, or social connectedness, relies on an index developed by Bailey et al. (2018b): the Social Connectedness Index (SCI). This index essentially captures the social graph for the universe of *active* US Facebook users as of April 2016, aggregated up to the level of 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

[12]Per capita and per worker are used interchangeably in what follows.

[13]Appendix A includes all additional figures, Appendix B includes all additional tables.

[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.

Figure 3: Total value of new production per capita in 2005-2012 (in millions)



users m and n and for each pair of counties i and j, the index is constructed as:

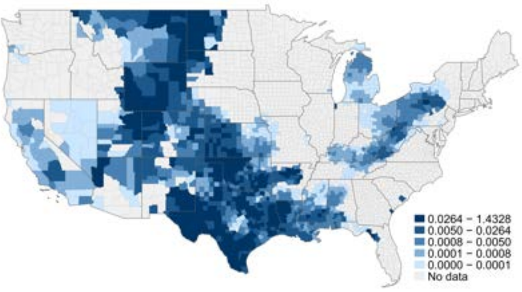
$$SCI_{ij} = \sum_{m \neq n} \sum_{n} \mathbb{1}_{mn}, \text{ for } m \in i \text{ and } n \in j \tag{9}$$

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 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.

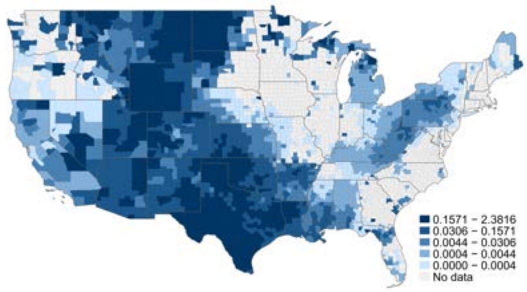
[15] Information obtained from Facebook's 2016 quarterly results report, retrieved at: `https://s21.q4cdn.com/399680738/files/doc_presentations/FB-Q4'16-Earnings-Slides.pdf`. Unfortunately, Facebook would not release covariates for these data. However, it is possible to gauge some descriptive facts from 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 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).

As will be discussed in Section 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^{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 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).

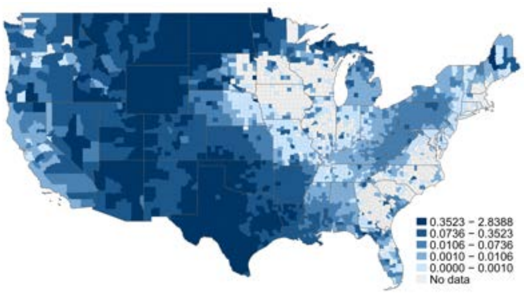
Figure 4: Cumulative socially lagged new production in 2005-2012



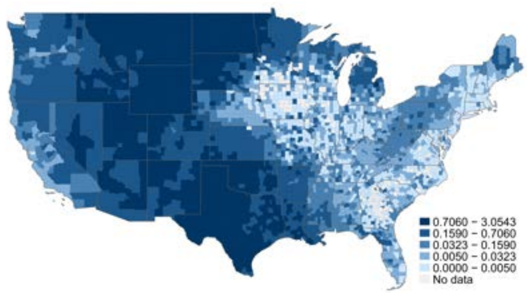
(a) First bin



(b) Up to 5^{th} bin



(c) Up to 10^{th} bin



(d) Up to 20^{th} bin

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 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 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^{th} bin, nearly every county in the contiguous US is exposed to fracking shocks through one of its social neighbours.

3.2 Identification Strategy

3.2.1 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 2, the following empirical model in first differences is estimated using ordinary least squares (OLS):

$$\Delta Y_{i,t} = \beta \times \Delta X_{i,t} + \sum_{d=1}^{20} \gamma_d \times G_{d,i}\Delta X_t + \sum_{d=1}^{20} \delta_d \times W_{d,i}\Delta X_t + \theta_t + \epsilon_{i,t} \tag{10}$$

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 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 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.

3.2.2 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 codetermined. 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:

$$RSCI_{ij} = f(d_{ij}, M_{ij}, \upsilon_{ij}) \tag{11}$$

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:

$$\ln RSCI_{ij} = \beta \times \ln d_{ij} + \gamma \times \ln(M_{ij} + 1) + \upsilon_{ij} \tag{12}$$

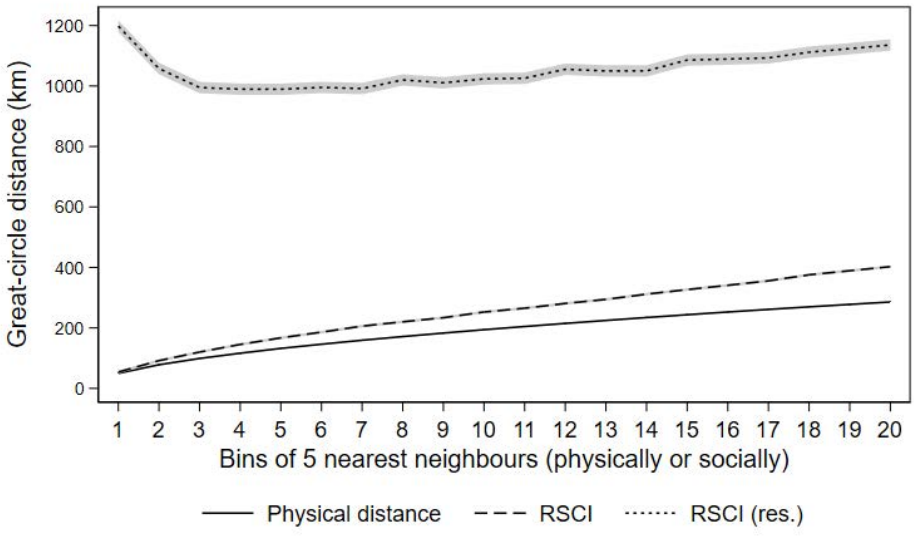
Where M_{ij} captures cumulative gross migration flows between all county pairs in the 2002-2016 period. This variable is constructed using counts of yearly county-to-county migration flows, obtained from the IRS Statistics of Income Division (SOI).[16] I predict the residuals $\hat{\upsilon}_{ij}$ and 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 3.1. The third column in the set of maps in Appendix Figure A.1 shows the resulting bins. I then compute alternative measures of socially lagged shocks based on these matrices. Appendix Figure A.3 maps them. Note how there is no clearly emerging spatial pattern over

[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

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 5 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.

Figure 5: 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.

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

3.2.3 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:

$$\ln(\Delta Q_{i,t}^{oil} \times P_{i,t}^{oil} + \Delta Q_{i,t}^{gas} \times P_{i,t}^{gas} + 1) = \alpha_i + \pi_{p,t} + \epsilon_{i,t} \tag{13}$$

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:

$$\Delta Z_{i,t} = \frac{\exp(\hat{\alpha}_i + \hat{\pi}_{p,t}) - 1}{L_{i,t}} \tag{14}$$

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 A.2 in Appendix maps these plays. Note that, for every bin defined by the matrices G_d, G_d^{res} 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:

$$\Delta Y_{i,t} = \beta \times \Delta \hat{X}_{i,t} + \sum_{d=1}^{20} \gamma_d \times G_{d,i}^{res} \Delta \hat{X}_t + \sum_{d=1}^{20} \delta_d \times W_{d,i} \Delta \hat{X}_t + \theta_t + \epsilon_{i,t} \tag{15}$$

Note that the set of matrices G_d^{res} 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 B.2 in Appendix gives summary statistics for all the main variables used in the analysis.

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

[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.

of diffusion using the plain measure of social connectedness ($RSCI$) and the one obtained by partialling-out physical distance and migration.[18] Appendix 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 B.

4.1 Effects on Wages

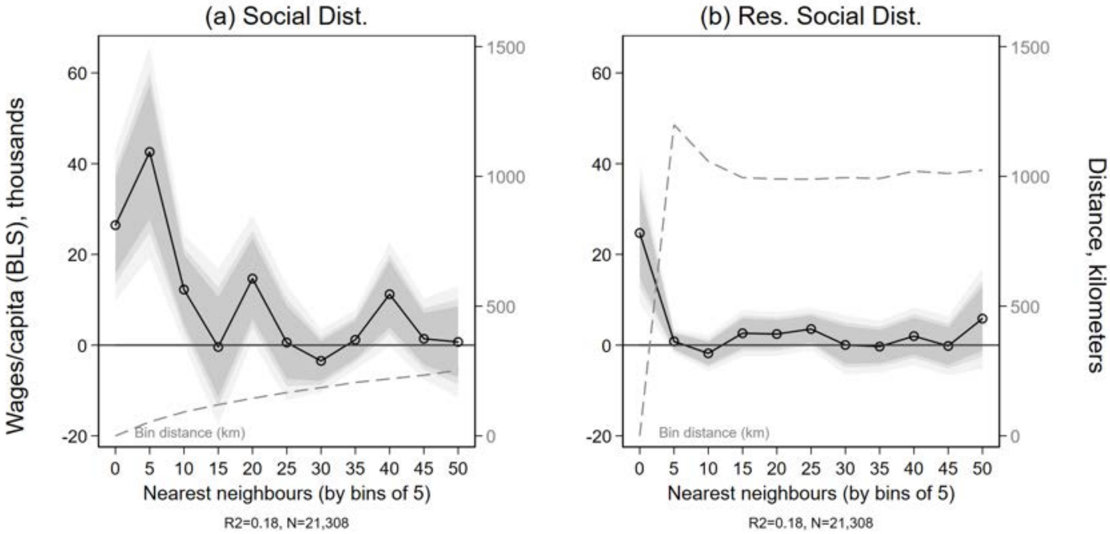
I begin by showing effects on wages. The conceptual framework predicts that in response to positive shocks to a social neighbour, wages should increase by a larger amount the more socially connected two counties are. In this specific application, a fracking-related shock to a county with which many friendship connections are shared should matter more for local outcomes than one taking place in a more 'socially distant' county. Figure 6 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 B.3 in Appendix reports exact estimates for all coefficients.

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 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

[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.

[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.

Figure 6: Coefficients plot for wages (BLS) using OLS (main)



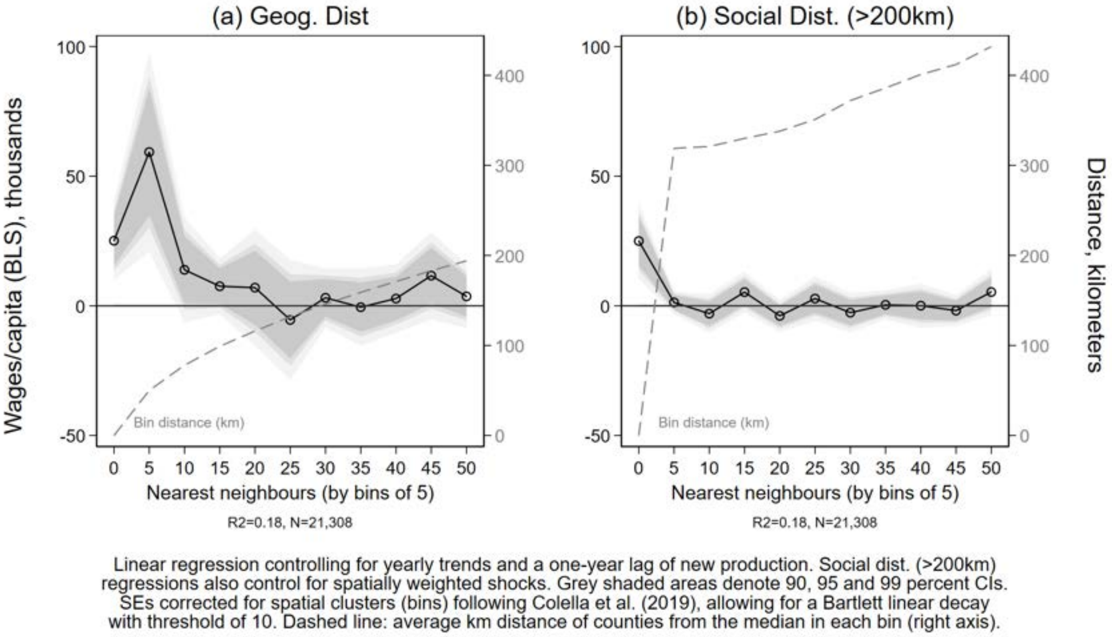
Linear regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a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

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 7. 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.

Introducing county fixed effects to the BLS wage regressions leaves the results virtually unchanged (Appendix Figure 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 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 8 offers a sector breakdown of the OLS estimates obtained using the simple

Figure 7: Coefficients plot for wages (BLS) using OLS (additional)

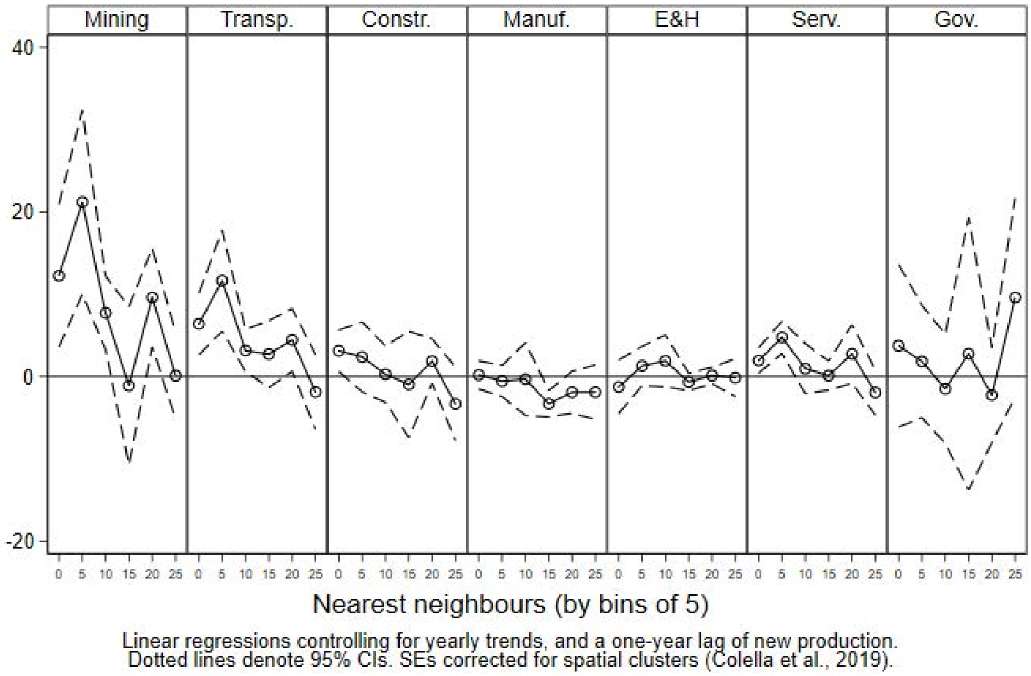


social connectedness measure.[20] The diagram shows that, in relative terms, the largest surge in wages is observed in mining activities and extraction activities, followed by transportation. In addition, it appears that services benefit somewhat from fracking shocks, although only in the most closely socially connected counties. This aligns to previous findings in the literature and with what discussed in Section 2.1. A corresponding set of results for geographical distance is available in Appendix (Figure A.10).

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 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

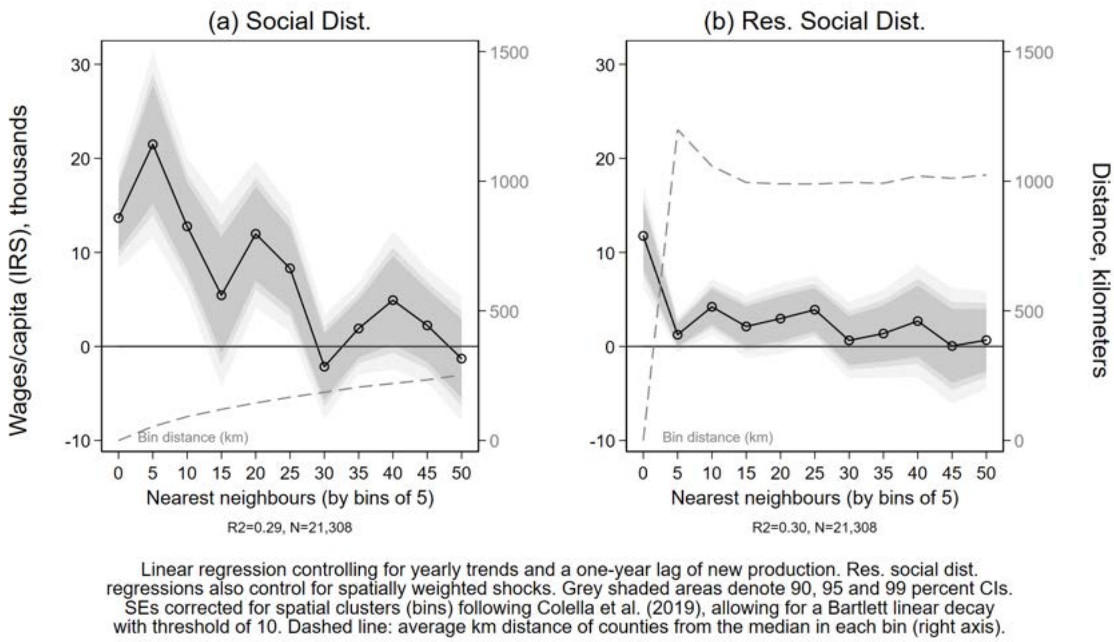
[20] I only report results based on the plain measure of social connectedness, rather than the partialled-out one, because there were no detectable effects on BLS wages in the latter. Feyrer et al. (2017) provide more details on effects across industries, space and time. An alternative way of studying 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.

Figure 8: Coefficients plot for wages (BLS) by industry using OLS (soc. dist)



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 9 reports OLS estimates of the effects for IRS wages. Table B.7 in Appendix reports exact estimates for all coefficients.

Figure 9: Coefficients plot for wages (IRS) using OLS



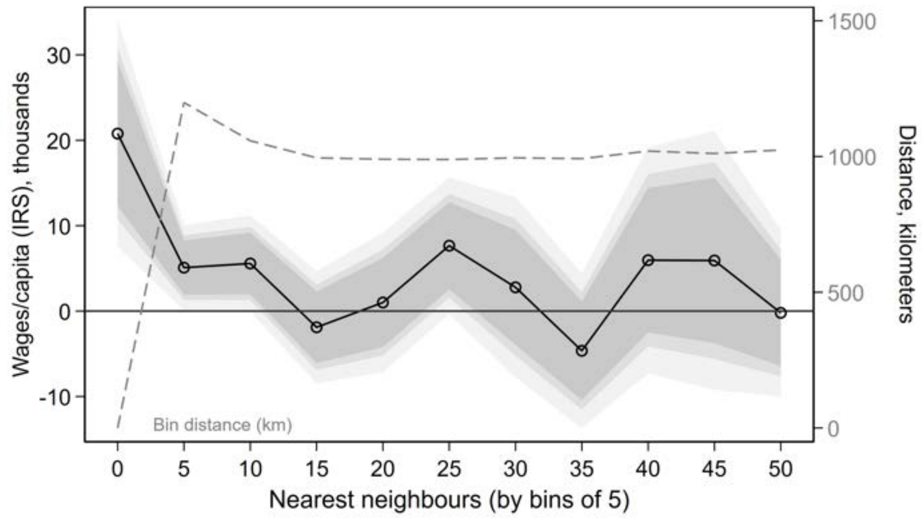
Panel (a) largely confirms previous findings, although the relevance of social connectedness ($RSCI$) decays more slowly, with effects diffusing up to the 25$^{\text{th}}$ nearest social neighbour, or about 170 kilometres away. The key take-away from this diagram, however, is in panel (b), which uses 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 10 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 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.

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 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]

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

[22] Unfortunately, the IRS wage measure does not provide an industrial breakdown. This would have

Figure 10: Coefficients plot for wages (IRS) using 2SLS (res. soc. dist.)

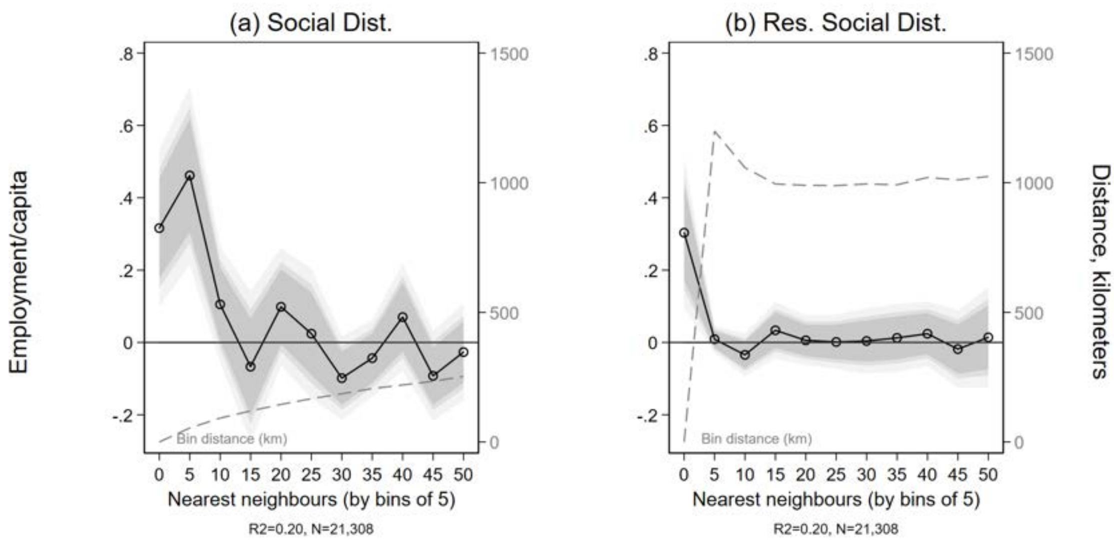


2SLS regression controlling for yearly trends, a one-year lag of new production, and spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for a Bartlett linear decay with threshold of 10. Dashed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

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 11 provides baseline OLS estimates for this relationship (Table B.11 in Appendix reports exact estimates for all coefficients).

Figure 11: 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).

allowed to test whether surges in wages occur in extraction related sectors despite the geographical distance from the sites.

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 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 (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.

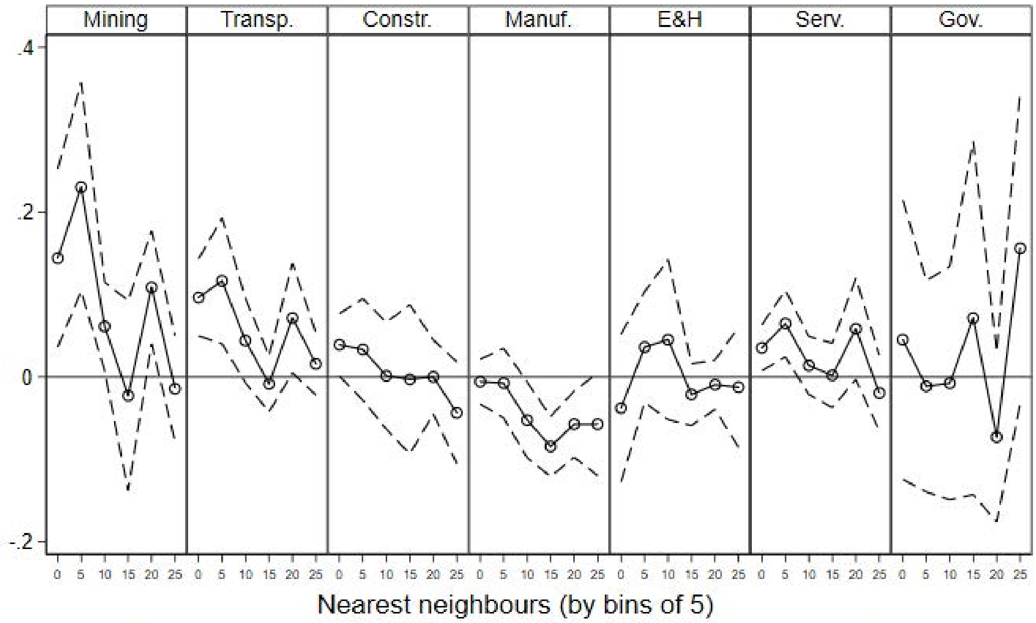
Similarly to what done with wages, we can look at a sector breakdown for employment creation. I report OLS estimates for the plain measure of social connectedness only, since there were no clearly discernible effects for the partialled-out measure of networks.[24] Consistent with the findings on wage-gains, Figure 12 shows that most job creation occurs directly in mining and transportation, with some new employment also being generated in 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]

[23] See Appendix, Figures A.18 and A.20, and Tables B.12 and B.14, respectively.

[24] A corresponding set of results for geographical distance is available in Appendix (Figure A.25).

[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

Figure 12: Coefficients plot for employment by industry using OLS (soc. dist.)



Linear regressions controlling for yearly trends, and a one-year lag of new production.
Dotted lines denote 95% CIs. SEs corrected for spatial clusters (Colella et al., 2019).

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.

when a measure based on social rather than physical distance is considered.

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 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.

References

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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

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.

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.

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*.

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, 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.

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.

Koster, H. R. and van Ommeren, J. (2015). A shaky business: Natural gas extraction, earthquakes and house prices. *European Economic Review*, 80:120–139.

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.

Michaels, G. (2011). The Long Term Consequences of Resource-Based Specialisation. *The Economic Journal*, 121(551):31–57.

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.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708.

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.

Patacchini, E. and Zenou, Y. (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6):938–949.

Percoco, M. (2012). Mining Activity and the Setting up of New Businesses in Basilicata. *Rivista economica del Mezzogiorno*, XXVI(4/2012):881–898.

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.

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.

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.

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.

Sonis, M. and Hewings, G. J. (1992). Coefficient Change in Input-Output Models: Theory and Applications. *Economic Systems Research*, 4(2):143–158.

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.

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.

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.

Appendices

A Figures

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

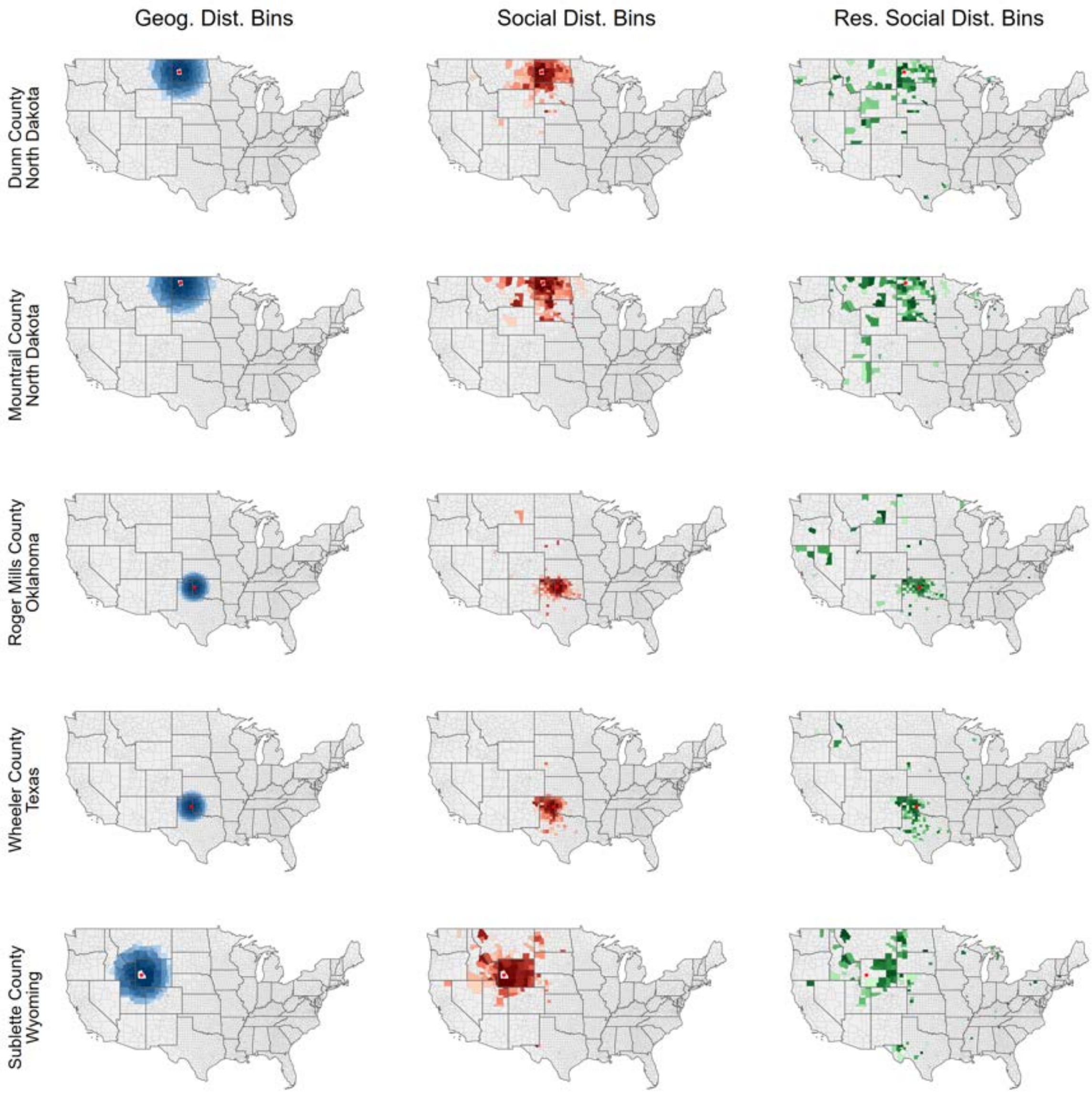
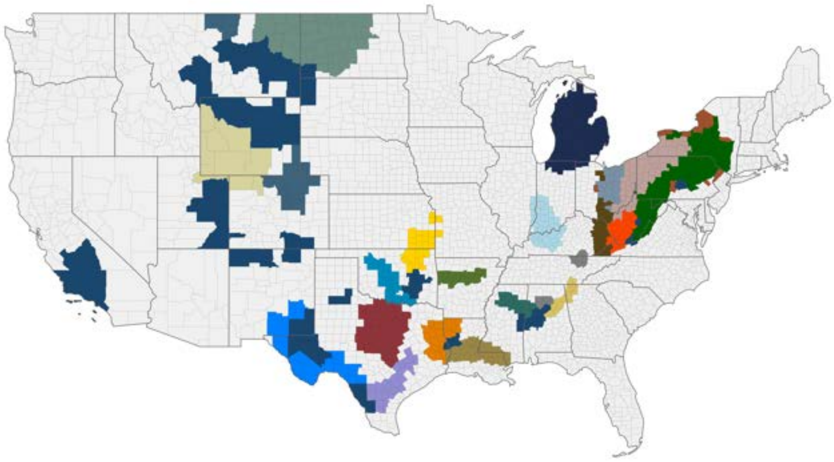
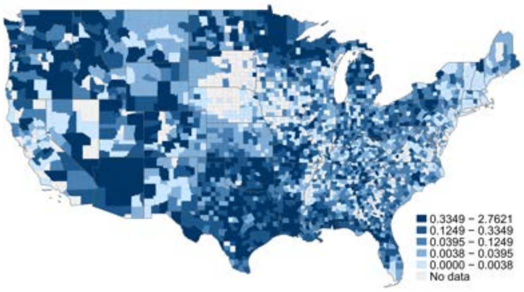


Figure A.2: US Shale plays, 23 designations and one ‘other’ category

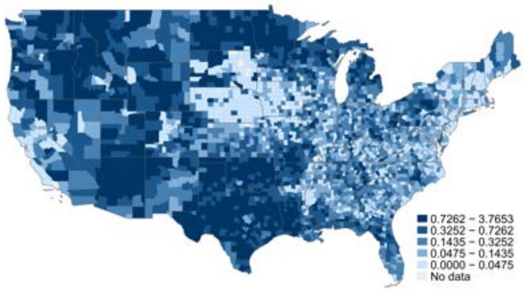


Source: EIA Shapefile “Major Tight Oil and Shale Gas Plays in Lower 48 States”

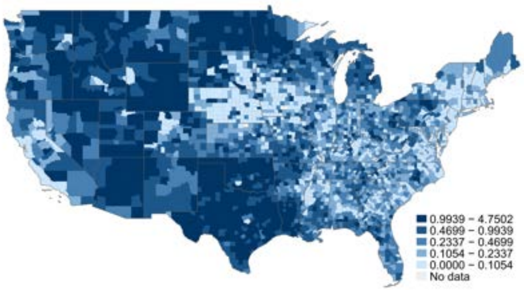
Figure A.3: Cumulative socially lagged new production (using the partialled-out $RSCI$)



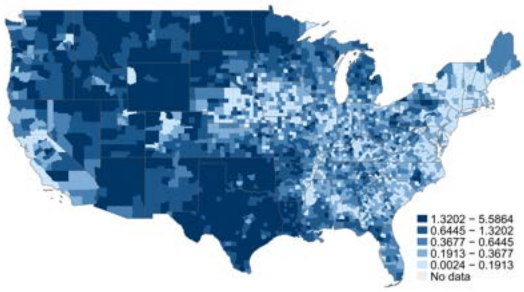
(a) First bin



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



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



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

A.1 Regression Coefficients Plots for wages (BLS)

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

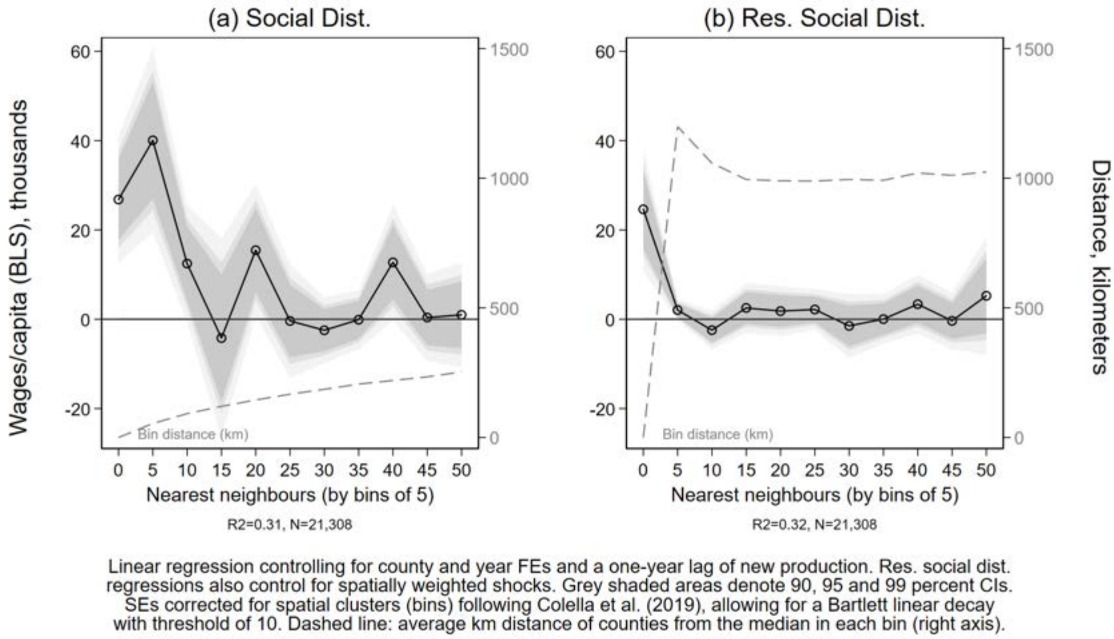


Figure A.5: Coefficients plot for wages (BLS), reduced form of 2SLS

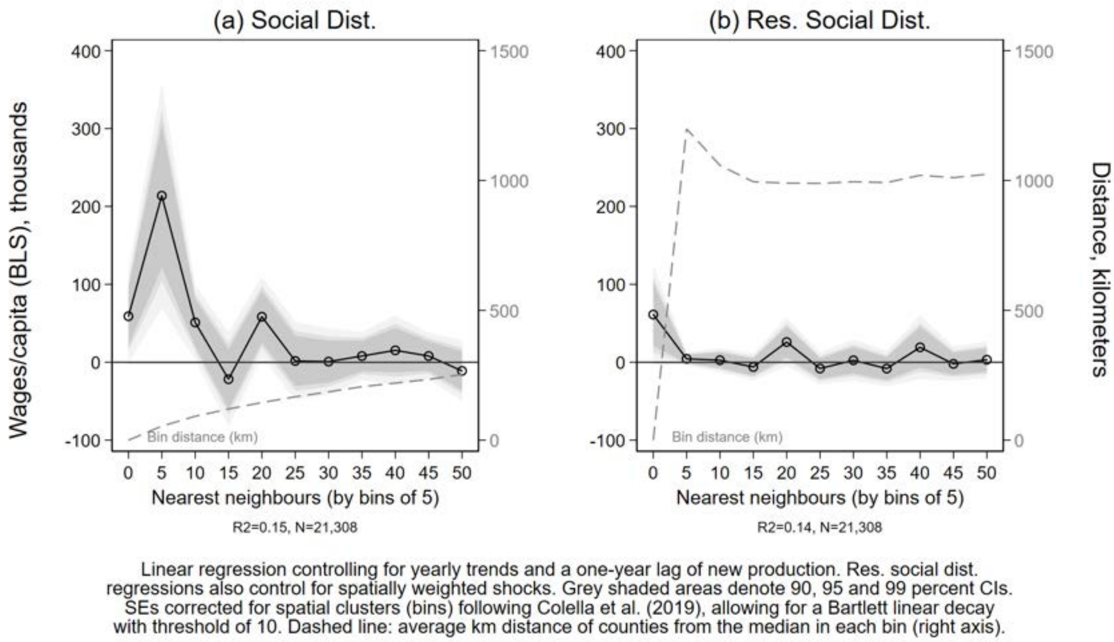
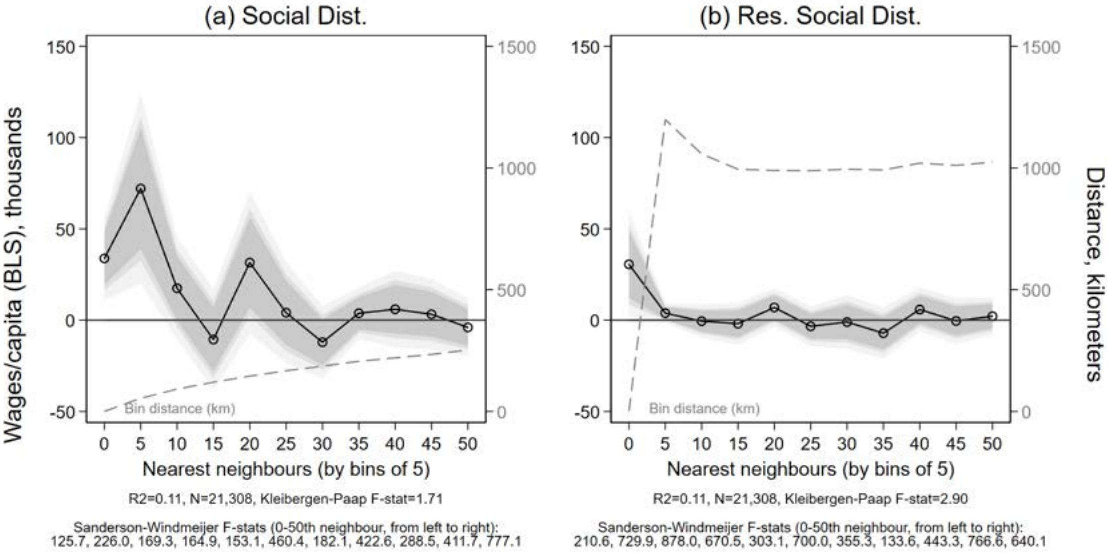
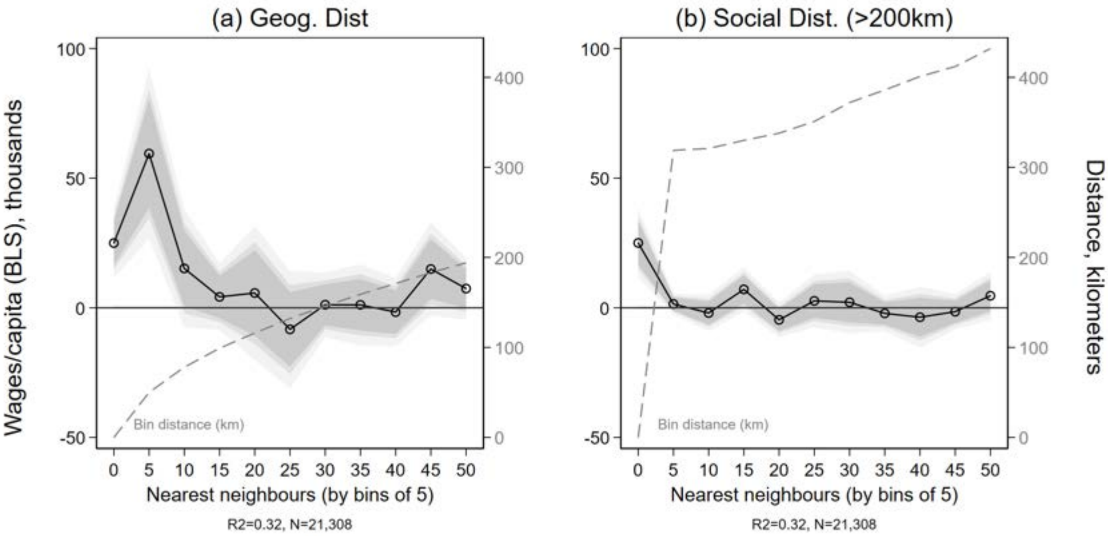


Figure A.6: Coefficients plot for wages (BLS) using 2SLS



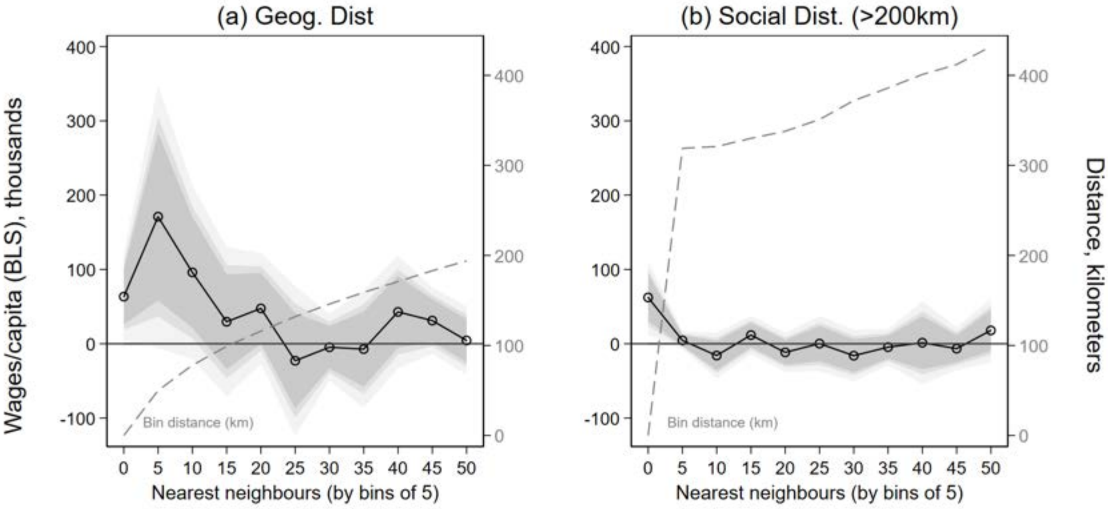
2SLS regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin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

Figure A.7: Coefficients plot for wages (BLS) 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).

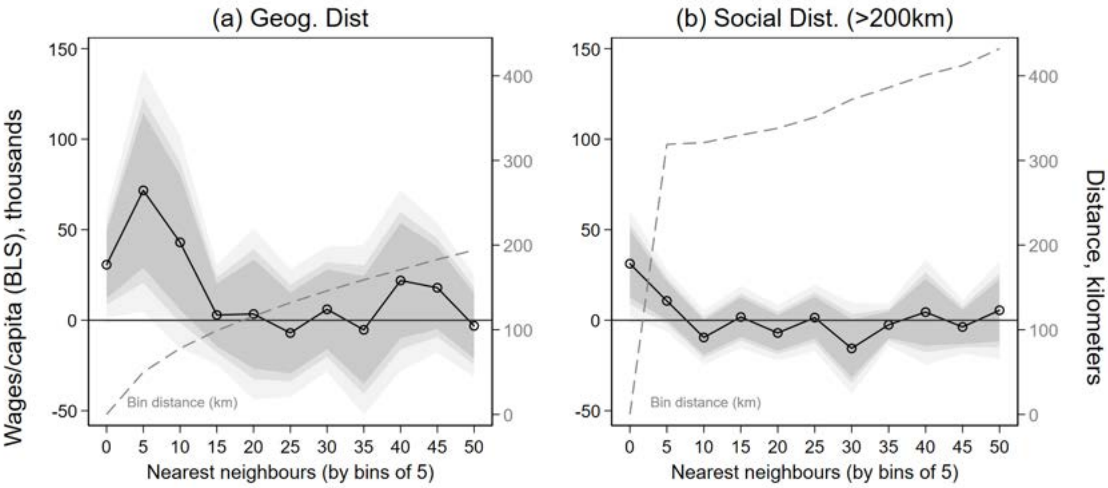
Figure A.8: Coefficients plot for wages (BLS), reduced form of 2SLS



R2=0.14, N=21,308 | R2=0.14, 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 A.9: Coefficients plot for wages (BLS) using 2SLS



R2=0.12, N=21,308, Kleibergen-Paap F-stat=4.87

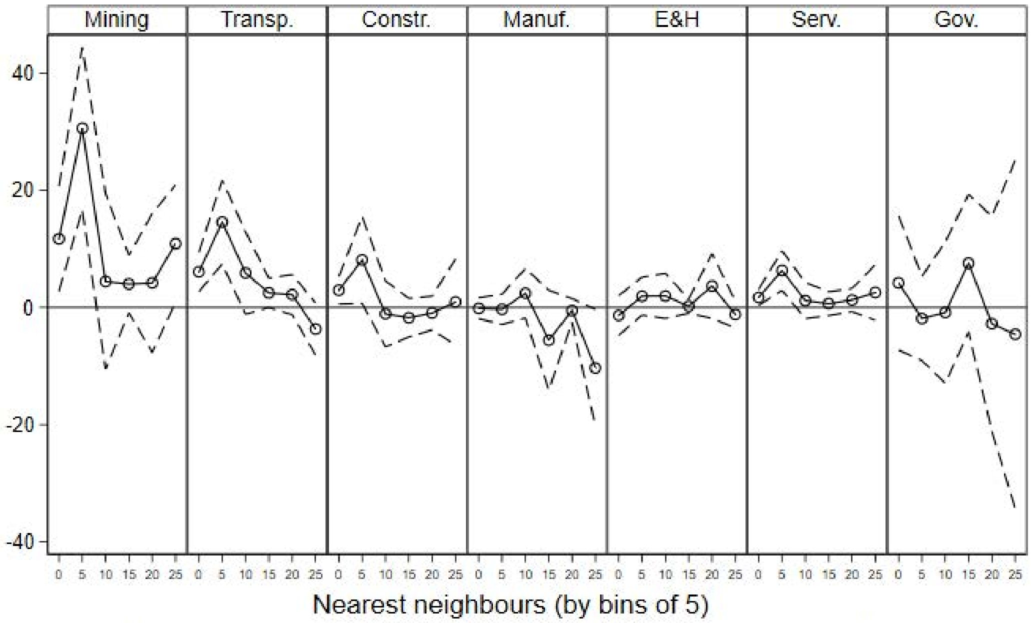
Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right): 110.8, 118.0, 92.7, 247.2, 245.7, 145.6, 81.7, 127.3, 172.3, 212.8, 208.3

R2=0.11, 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).

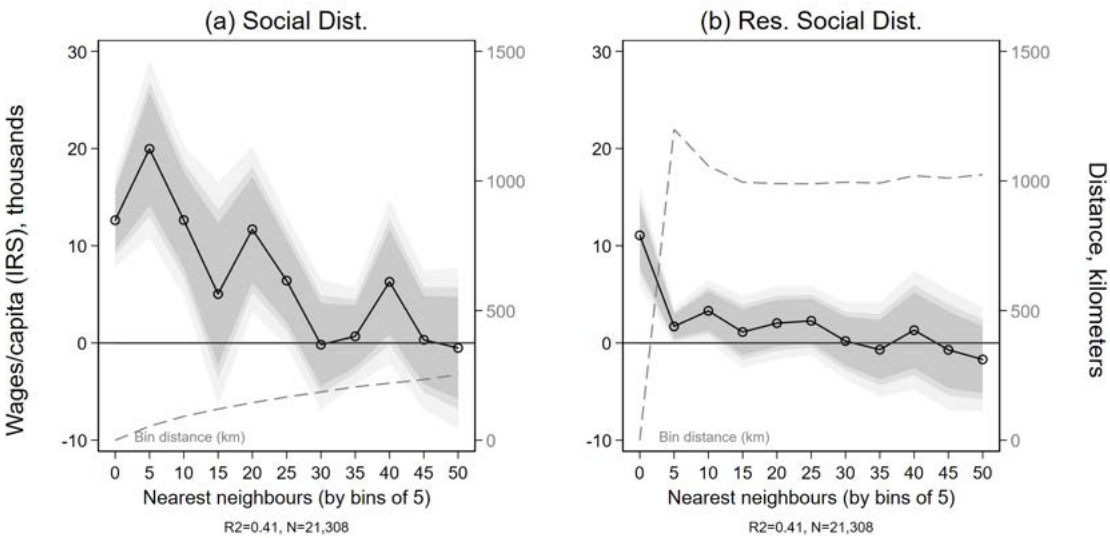
Figure A.10: Coefficient plot for wages (BLS) by industry using OLS (geog. dist.)



Linear regressions controlling for yearly trends, and a one-year lag of new production.
Dotted lines denote 95% CIs. SEs corrected for spatial clusters (Colella et al., 2019).

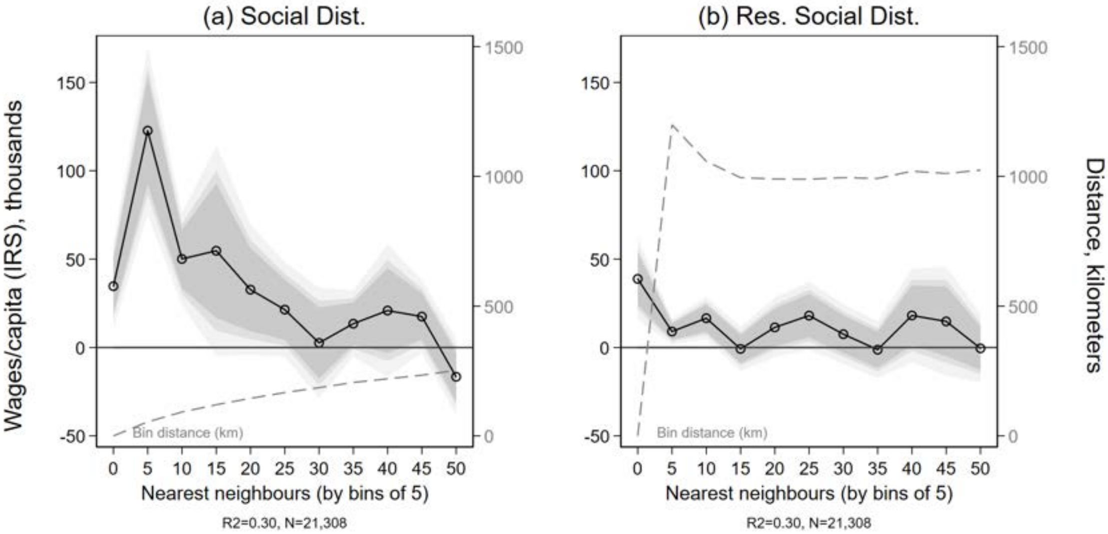
A.2 Regression Coefficients Plots for Wages (IRS)

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



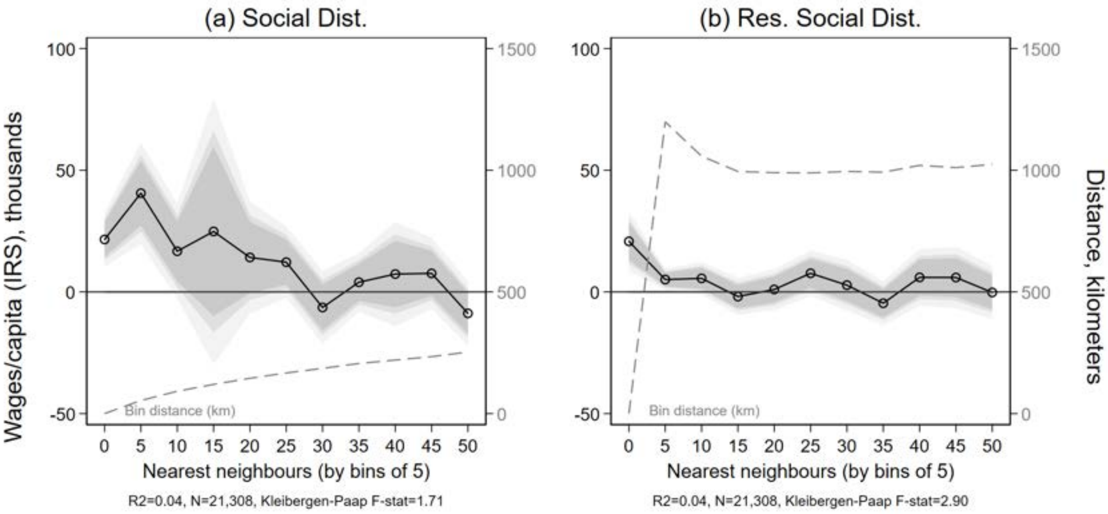
Linear regression controlling for county and year FEs and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing 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 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 A.13: Coefficients plot for wages (IRS) using 2SLS

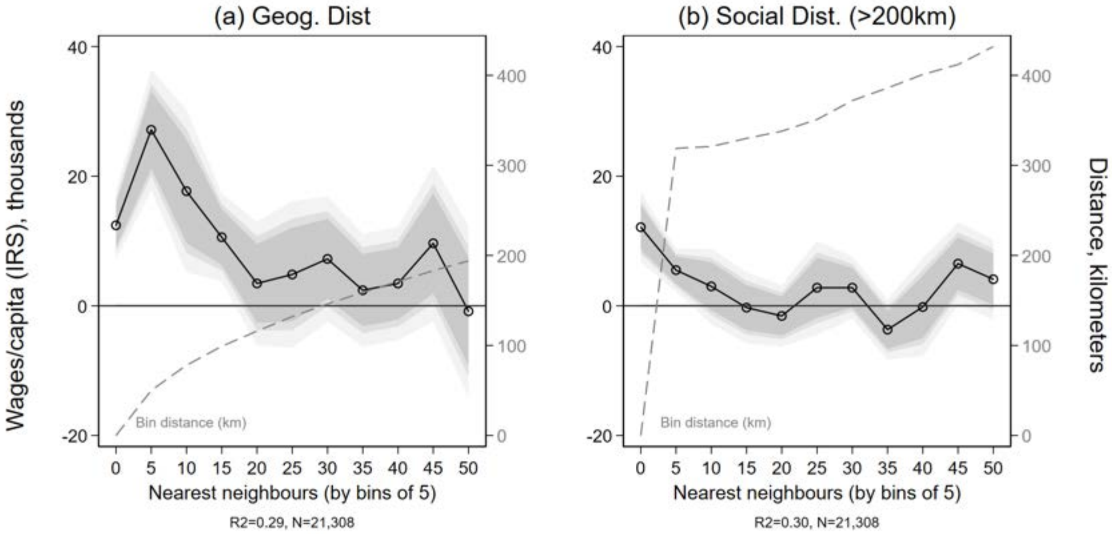


Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right): 125.7, 226.0, 169.3, 164.9, 153.1, 460.4, 182.1, 422.6, 288.5, 411.7, 777.1

Sanderson-Windmeijer F-stats (0-50th neighbour, from left to right): 210.6, 729.9, 878.0, 670.5, 303.1, 700.0, 355.3, 133.6, 443.3, 766.6, 640.1

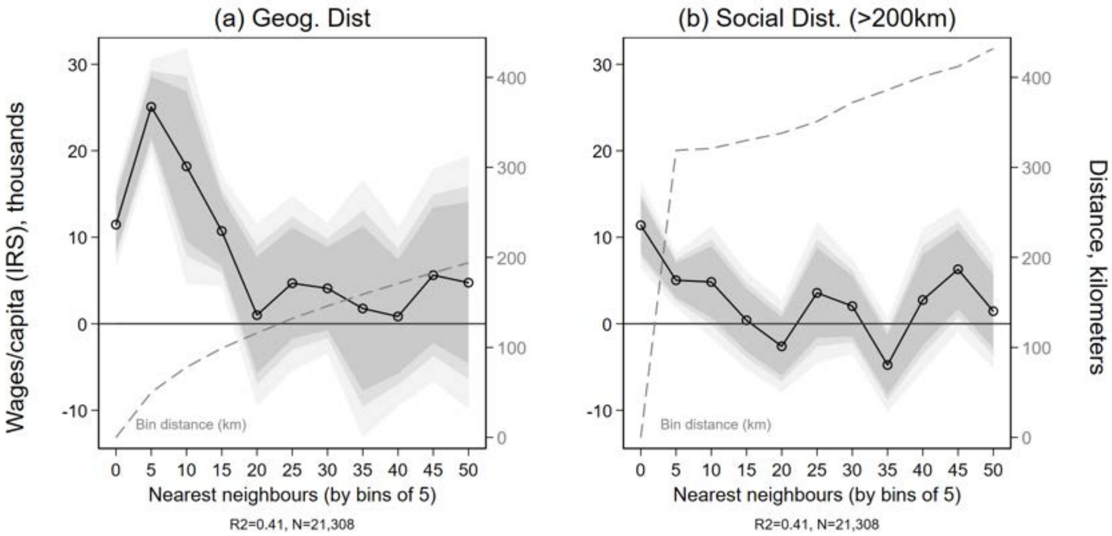
2SLS regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin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

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



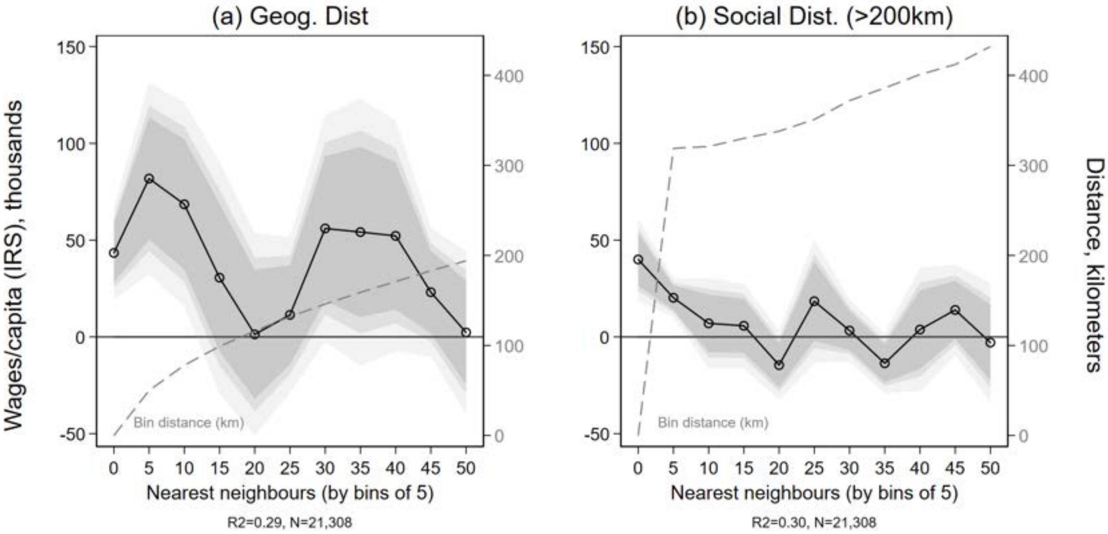
Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for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

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



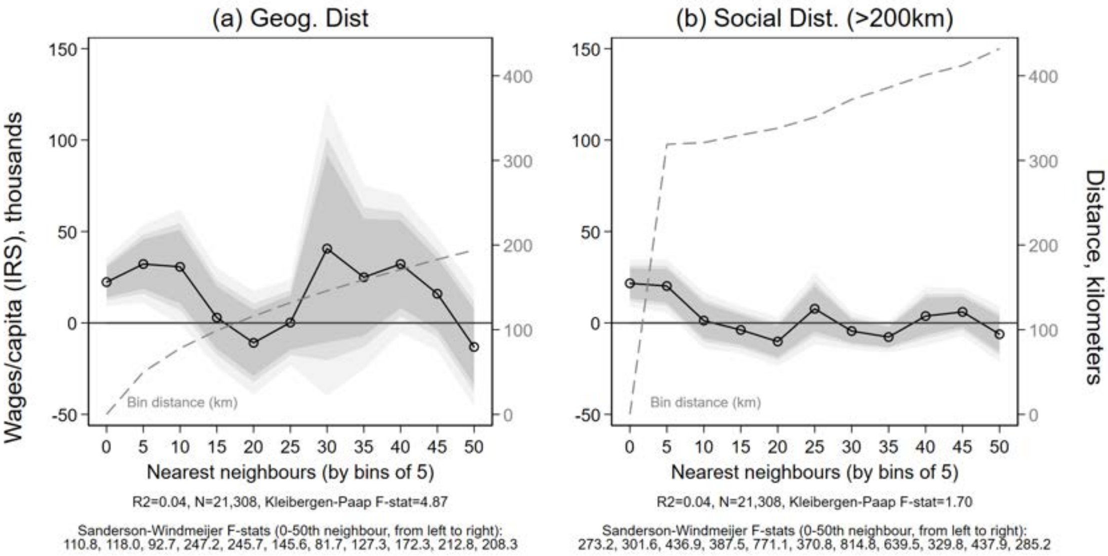
Linear regression controlling for county and year FEs and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing 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 A.16: Coefficients plot for wages (IRS), 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 A.17: Coefficients plot for wages (IRS) using 2SLS



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).

A.3 Regression Coefficients Plots for Employment

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

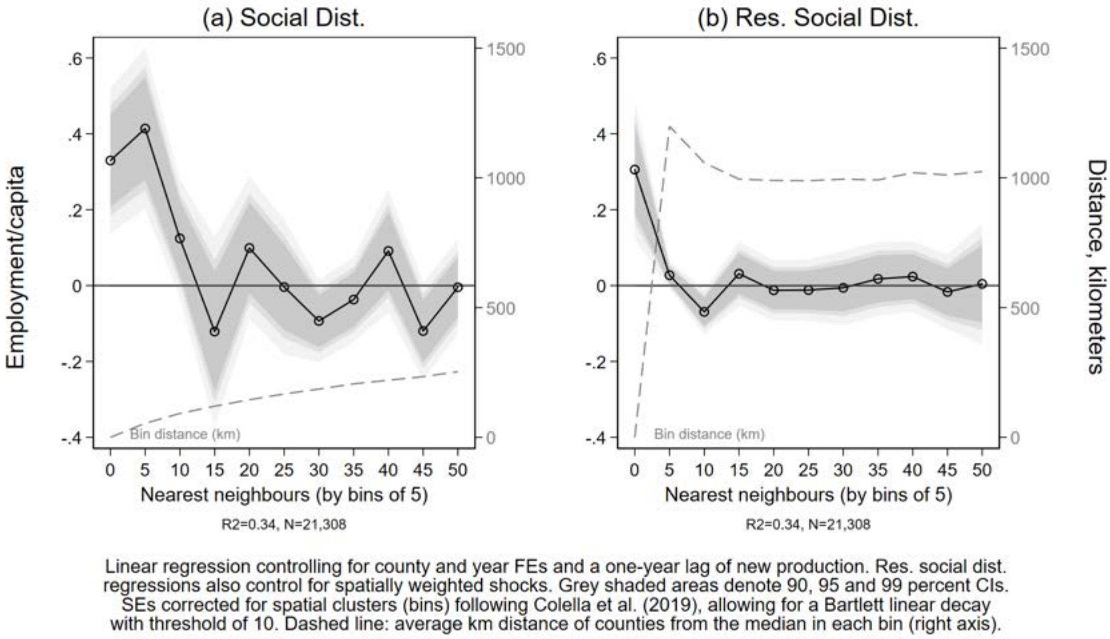


Figure A.19: Coefficients plot for employment, reduced form of 2SLS

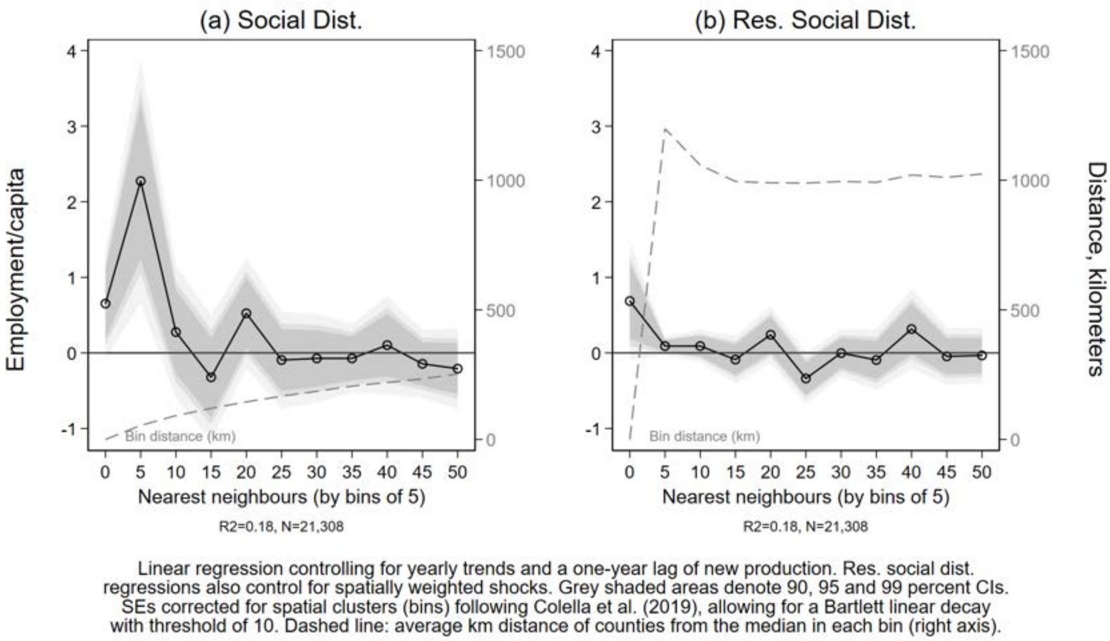
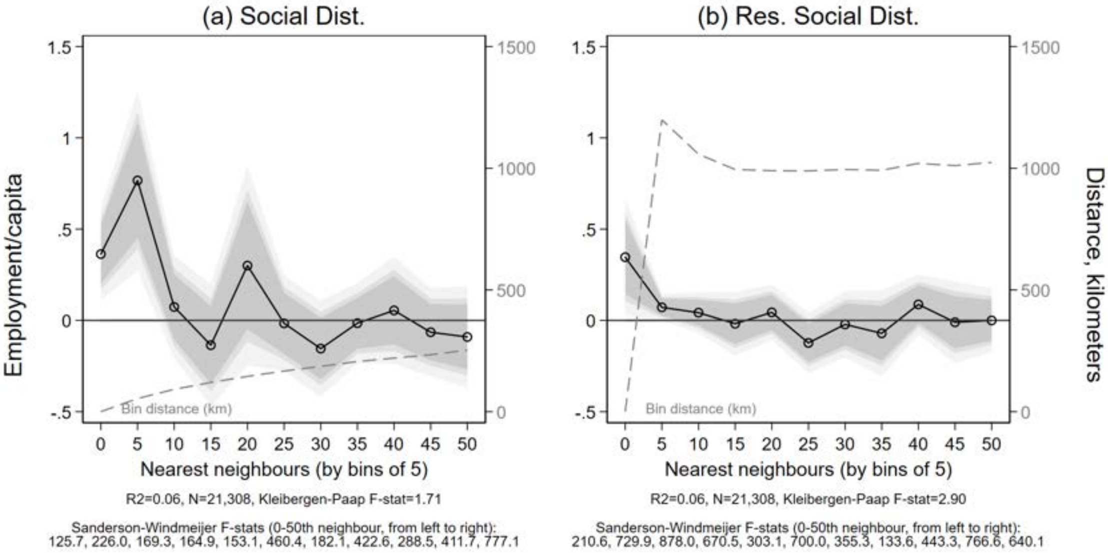
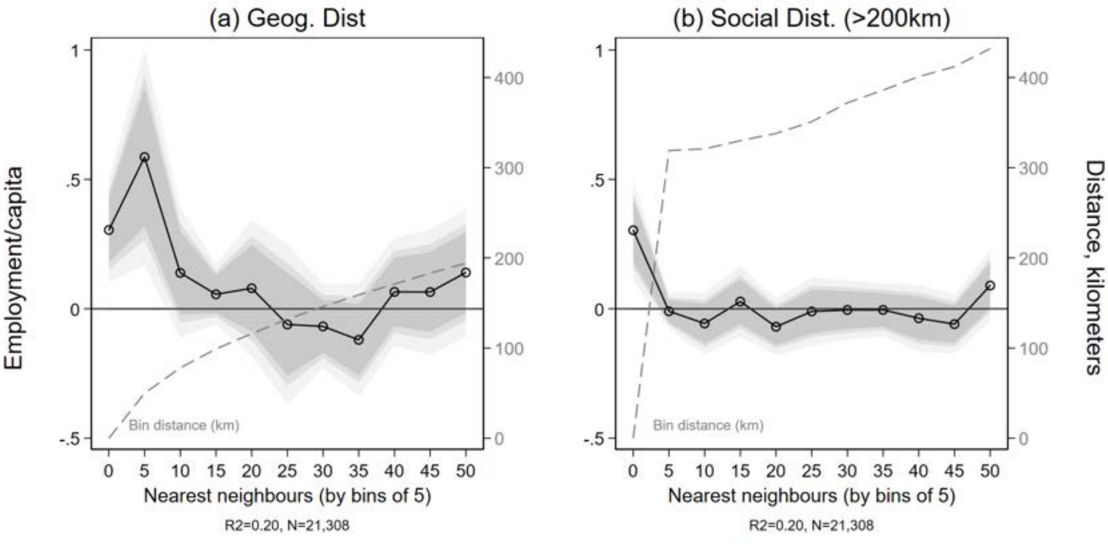


Figure A.20: Coefficients plot for employment using 2SLS



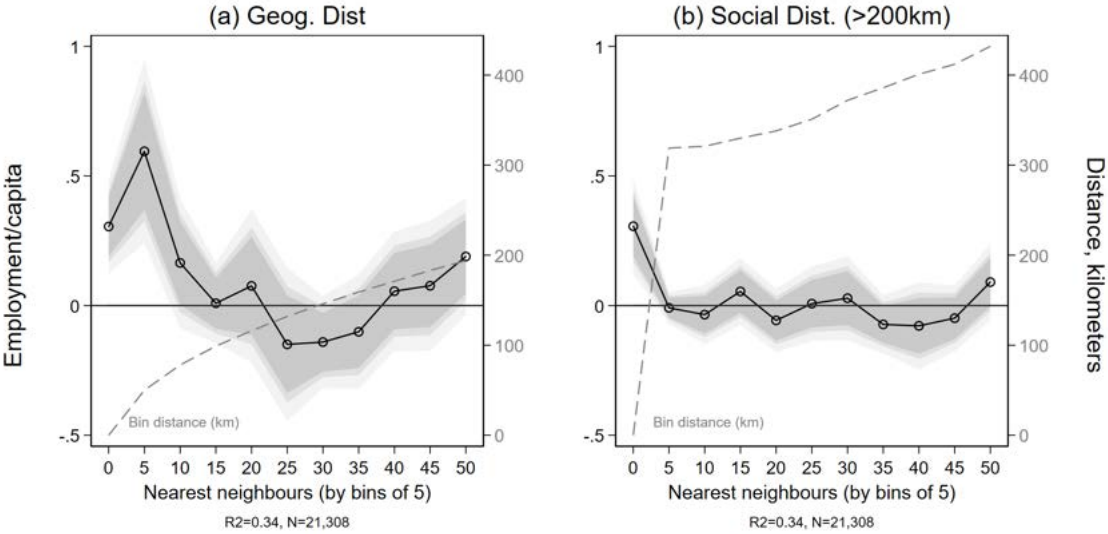
2SLS regression controlling for yearly trends and a one-year lag of new production. Res. social dist. regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs clustered by CZ. Dashed line: average km distance of counties from the median in each bin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

Figure A.21: Coefficients plot for employment using OLS



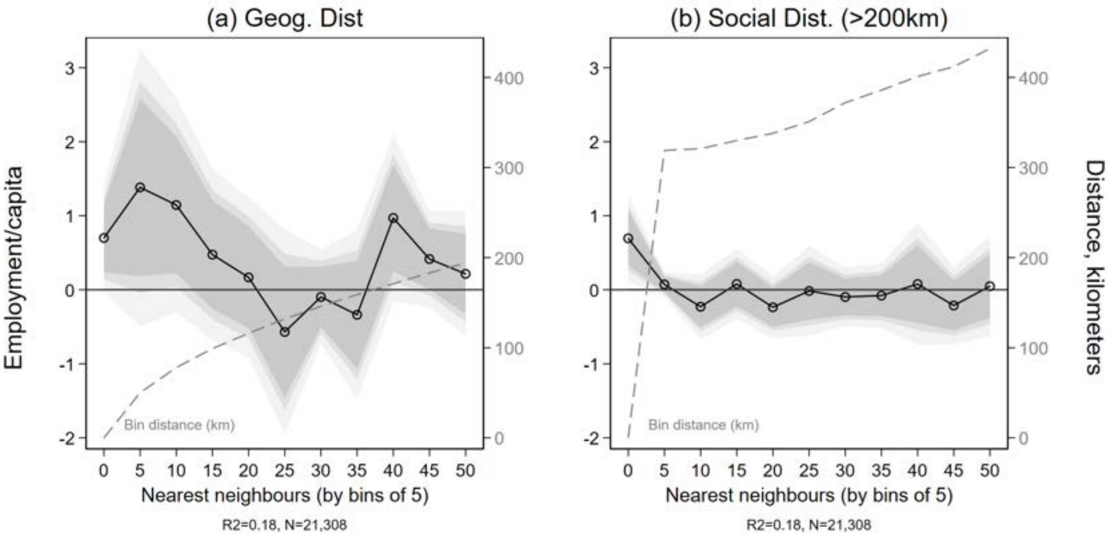
Linear regression controlling for yearly trends and a one-year lag of new production. Social dist. (>200km) regressions also control for spatially weighted shocks. Grey shaded areas denote 90, 95 and 99 percent CIs. SEs corrected for spatial clusters (bins) following Colella et al. (2019), allowing for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

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

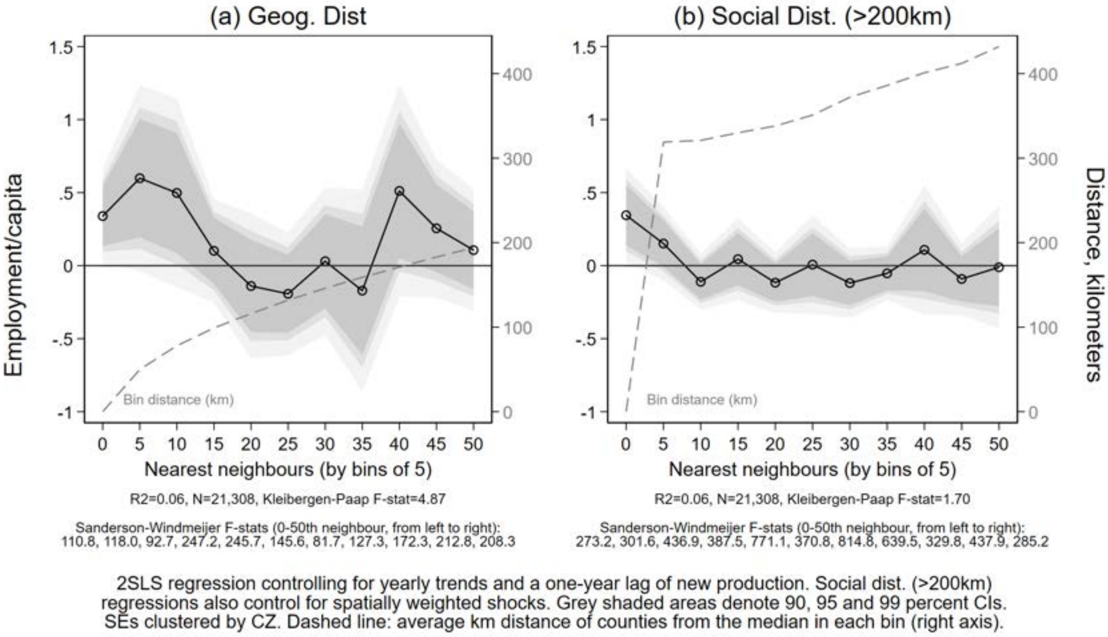
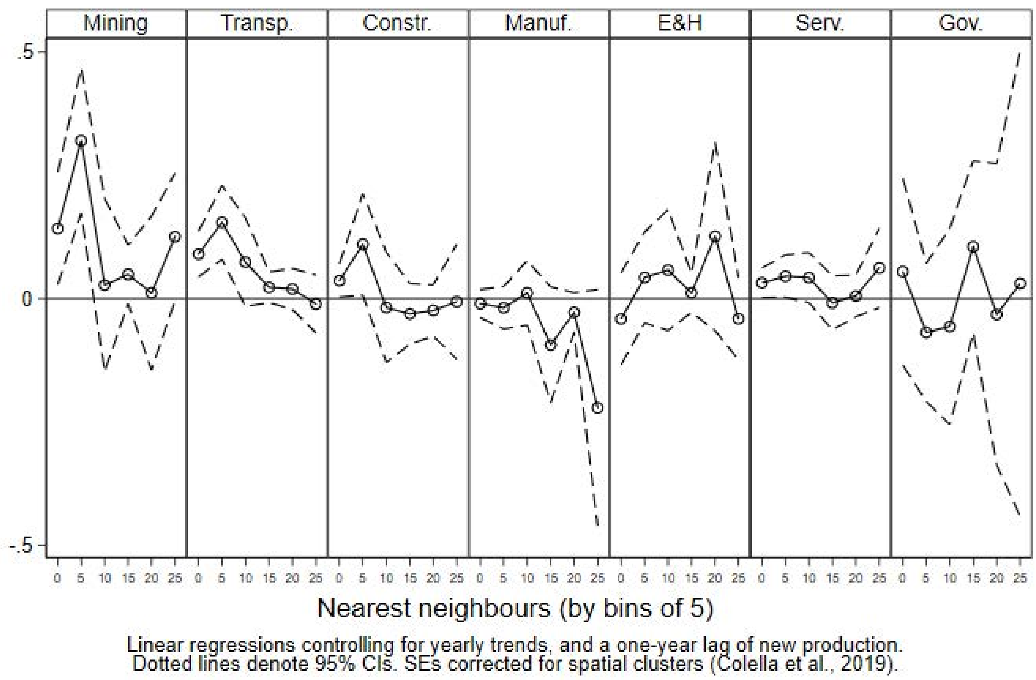


Figure A.25: Coefficients plot for employment by industry using OLS (geog. dist.)



B Tables

B.1 Summary Statistics

Table 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 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.

B.2 Regression Tables for Wages (BLS)

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

	(1)		(2)		(3)		(4)	
	Social Dist.		Geog. Dist.		Res. Social Dist.		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 B.6: Regression table for wages (BLS) using 2SLS

	(1)			(2)			(3)			(4)		
	Social Dist.			Geog. Dist.			Res. Social Dist.			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.

B.3 Regression Tables for Wages (IRS)

Table 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 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 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 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.

B.4 Regression Tables for Employment

Table 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 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 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 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.

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

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

	(1) 0	(2) 5	(3) 10	(4) 15	(5) 20	(6) 25	(7) 30	(8) 35	(9) 40	(10) 45	(11) 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 B.17: New production per capita in bins of social neighbours (res.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	0	5	10	15	20	25	30	35	40	45	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 B.18: New production per capita in bins of social neighbours (>200km)

	(1) 0	(2) 5	(3) 10	(4) 15	(5) 20	(6) 25	(7) 30	(8) 35	(9) 40	(10) 45	(11) 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.