Abstract—In this paper, we analyze how the physical layout of cities affects innovation by influencing the organization of knowledge exchange. We exploit a novel data set covering all census block groups in the contiguous United States with information on innovation outcomes, street infrastructure, as well as population and workforce characteristics. To deal with concerns of omitted variable bias, we apply commuting zone fixed effects and construct instruments based on historic city planning. The results suggest that variation in street network density may explain regional innovation differentials beyond the traditional location externalities found in the literature.

I. Introduction

Streets and their sidewalks, the main public places of a city, are its most vital organs.

—Jane Jacobs, *The Economy of Cities*

The geographic concentration of innovation in metropolitan areas across space is well documented in the literature (Acemoglu, Ansell, & Varga, 2002; Carlino & Kerr, 2014; Jaffe, Trajtenberg, & Hall, 1993; Rosenthal & Strange, 2003), and a wide stream of research identifies the importance of geographic environments in organizing and supporting innovation (Porter, 1996; Saxenian, 1996; Scott & Storper, 2003). But not all geographic environments such as cities, and the neighborhoods within, are equally equipped to do so. Cities and their neighborhoods vary in size and scope, their density of activities and amenities, and the manner in which they facilitate or impede the movement of individuals.

In order to understand the roots of innovation differentials across cities, much work has focused on modeling the innovative output from cities as a function of agglomeration economies and, more specifically, urban size and density. The empirical evidence suggests that by facilitating exchange, urban density helps knowledge spread (Arzaghi & Henderson, 2008; Carlino, Chatterjee, & Hunt, 2007; Kantor & Whalley, 2014; Lin, 2011; Rosenthal & Strange, 2008). However, it remains puzzling that regions with similar population and even inventor density differ so starkly in innovation output (Agrawal et al., 2014).

A further stream of research suggests that urban efficiencies are not only a function of agglomeration economies, but can also be attributed to nonagglomeration channels. Early research points out that these efficiencies may depend on the nature of urban exchange (Chinitz, 1961; Jacobs, 1969) that is partially determined by social structures and industrial practices (Saxenian, 1996). More recent findings highlight the importance of physical structures for growth and innovation within a region. For instance, by supporting the circulation of local knowledge, regional transportation infrastructure has been found to increase patenting output (Agrawal, Galasso, & Oettl, 2017).

In this paper, we introduce an additional factor that has received little attention in the innovation literature thus far, but may have important implications for innovation: the physical layout of neighborhoods. We specifically examine the effect of neighborhoods’ street infrastructure on innovation. Our main notion is that a more physically connected infrastructure, as determined by its street network, positively affects the extent to which interpersonal exchange can take place and is organized. For one, within a strongly connected environment, the number of potential contacts is high, thereby increasing the likelihood of more serendipitous knowledge exchange. For another, a strongly connected environment enables more efficient time allocation between travel and planned interpersonal knowledge exchange. For instance, shorter travel distance between economic partners, to formal knowledge centers, and to places where social activity is hosted both reduces the costs associated with interpersonal knowledge exchange and increases the available amount of time for interaction.

Providing more potential contacts and higher levels of interaction efficiency are important for innovation given that interpersonal exchange facilitates the recombination of existing knowledge and creation of new knowledge (Fleming & Sorenson, 2004; Hargadon, 1998; Simonton, 2003; Singh & Fleming, 2010). Moreover, knowledge production is increasingly a collaborative endeavor between multiple individuals (Wuchty, Jones, & Uzzi, 2007). This is why we expect any physical infrastructure that more efficiently organizes the circulation of individuals to also have a positive effect on innovation.

The data for the analyses come from various publicly available data sources that we have collected on the neighborhood level. Our definition of a neighborhood encompasses the most microgeographic unit of analysis available for street infrastructure: the census block group (hereinafter BG). In this paper, we measure the physical connectivity of a BG using street network density and proxy innovation with the number...
of U.S. patents applied for in a BG. We retain those patents only where assignees are located and inventors reside within the same larger metropolitan area. In doing so, we ensure that the location of each patent indeed reflects the place where the creation of the underlying idea most likely took place. To control for traditional location externalities, we include measures that capture both historic and contemporaneous employment and population density, as well as characteristics of the workforce employed in a BG.

The decision where to locate entails a long-term commitment. As such, individuals and organizations pay particular attention to the features of a place when deciding where to settle down. Consequently, location choices are likely endogenous to economic outcomes, making it difficult, from an empirical standpoint, to identify causal relationships (Hanson, 2001). To address endogeneity concerns, we apply a fixed effects and instrumental variable approach. One aggregate geographic boundary we use for the fixed effects estimation is the commuting zone, a natural boundary definition determined by places of residence and work of employees. Access to amenities, fiscal policies, exposure to a certain culture or lifestyle, and other unobservable features will be similar within the commuting zone boundaries. By holding this type of general environment constant, we can exploit variation within commuting zones. To deal with concerns about simultaneity, we construct instruments based on historic city planning. The instruments we use are the percent of housing units built prior to 1940 and 1940–1949, which (conditional on controls) should have little effect on innovation today other than through their effect on street layout. First-stage results show that both instruments together strongly predict contemporary street infrastructure. Our second-stage results from the instrumental variable estimation reveal a positive causal relationship between physical connectivity and innovation.

In order to provide more insight into the mechanisms that might be driving our results, we first analyze citation patterns at the neighborhood level. Here, the results indicate that physical connectivity to some extent influences local interorganizational knowledge exchange within a neighborhood. Next, we interact physical connectivity with measures for social activity, finding that the physical layout of neighborhoods bolsters the impact of social factors on innovation. Taken together, these results provide suggestive evidence that higher physical connectivity has a positive impact on innovation by increasing local knowledge circulation within a neighborhood.

Our findings are relevant to our understanding of what urban structures best support local innovation, offering useful insights for the (micro-)geography of innovation literature, organizations faced with the decision where to locate, and regional policymakers wishing to influence local conditions. The geography of innovation literature has shown mixed results with regard to proximity and innovation outcomes. For example, empirical studies find that large corporate plants are relatively isolated from knowledge externalities, whereas small, single-plant firms are those that seem to benefit most from proximity (Beardsell & Henderson, 1999). Based on the evidence we provide in this paper, this could possibly be explained by the fact that certain organizations select into places where local infrastructure is not as conducive to facilitating interorganizational knowledge flows. Our findings are also informative for organizations faced with location choice that should be aware of how their most immediate environment may influence knowledge flows (Moretti, 2004). Like industry structure, local infrastructure presents a factor that can either promote or prevent knowledge from spilling over. Similarly, we highlight that street infrastructure may represent an important asset for policymakers and regions can leverage as a source of competitive advantage.

The paper is structured as follows. In section II, we develop the basic theoretical framework to guide empirical predictions and interpretations of the findings. Section III provides an overview of how the data were constructed, followed by the main results. We conclude with a discussion of the results, including limitations, implications, and opportunities for future research.

II. A Physical Environment That Connects

Agglomeration economies, and especially knowledge spillovers, until now have mainly been viewed as a function of urban size or density (Arzaghi & Henderson, 2008; Glaeser et al., 1992; Lin, 2011). The empirical evidence suggests that by facilitating exchange, urban density helps knowledge spread (Carlino et al., 2007). In light of these findings, it remains a puzzle that regions with similar population and even inventor density are found to differ in innovation output (Agrawal et al., 2014).

More recently, as an explanation for these disparities, the impact of transportation infrastructure on innovation has been receiving attention. The evidence provides valuable insights into the implications that transportation infrastructure has for reducing costs associated with knowledge exchange. One example is Agrawal et al. (2017), who exploit interstate highway system plans, railroads, and exploration maps as instruments to study the impact of highways on patenting.¹

In this paper, we combine previous findings on the effects of urban density and interurban transportation infrastructure on innovation and highlight a novel dimension that goes in line with these findings. Moving from regional to neighborhood-level infrastructure, we analyze the impact of physical connectivity on innovation via its effect on increasing both the potential for and efficiency of interaction. Interaction between individuals is especially crucial for innovation as it enables the exchange and recombination of existing knowledge necessary to create new or improve existing technologies, processes, or products (Fleming & Sorenson, 2004; ¹Similarly, in the most recent work, Davis and Dingel (2019) propose a system-of-cities model with costly knowledge exchange as the primary agglomeration force. The authors stress the important role that transportation infrastructure plays in determining at what frequency interactions can feasibly occur in the first place.
Gaspar & Glaeser, 1998; Hargadon, 1998; Simonton, 2003). Moreover, over the past decades, knowledge production has increasingly become a team process involving multiple individuals (Wuchty et al., 2007), making it ever more important to understand what structures best support collaboration.

The individuals involved in knowledge production operate in different physical environments that are organized in distinct ways. Research has shown that the physical structure of the environment has strong implications for the frequency and likelihood of interaction (Allen, 1977; Estabrook & Sommer, 1972; Festinger, Schachter, & Back, 1950). One aspect of the physical environment influencing the frequency and likelihood of interaction between actors is street network structure (Levinson, 2010), which thereby determines the physical connectivity within a given area. Most important, denser street networks have been found to be strongly correlated with lower car usage, increased nonauto travel, and more direct trips (Parthasarathi, 2014), factors that make trips both shorter and faster. Generally, we are more likely to find elevated levels of street density in metropolitan areas, rather than rural areas. But even within metropolitan areas not all street networks are created equally, there being much heterogeneity between and within regions and agglomerations.

A more strongly physically connected environment creates greater potential for interpersonal encounters and enables a more efficient organization of interaction. This should positively affect the extent to which interpersonal knowledge exchange occurs since both the number of contacts and the amount of time spent with partners related to the time spent traveling increase with higher physical connectivity. In other words, higher physical connectivity should reduce both execution costs (e.g., the cost of face-to-face meetings, coordination costs, monitoring costs, and costs incurred for the transfer of tacit knowledge) and search costs (e.g., finding collaborators, suitable technologies, and identifying facilities that provide certain instruments) associated with knowledge production (Agrawal et al., 2006; Catalini, 2018; Mors, 2010). This difference in organization could translate into important innovation differentials found between cities, neighborhoods within a city, and the organizations located there. From this, and capturing physical connectivity through street network density, we expect that with increasing street network density, innovative output will rise.

### III. Estimation Strategy

#### A. The Unit of Analysis

The unit of analysis for this study is the neighborhood, which represents the natural boundary of most individuals’ daily work activities and routines (excluding the commute to work). In this paper, we define a neighborhood as a BG and use it to probe deeper into the impact of the immediate environment on innovative outcomes. We base our definition on prior literature providing evidence that social interactions are notably local in nature. Studies in this body of research suggest that localization effects may indeed be strongest within 500 meters or less (Arzaghi & Henderson, 2008) and decay rapidly with distance (Rice, Venables, & Patacchini, 2006; Rosenthal & Strange, 2003, 2008). Considering that a standard block in Manhattan is 200 × 500 feet (roughly 61 × 152 meters), walking along the one or the other block side, 500 meters is the equivalent of three to eight blocks. This is slightly less than the number of blocks in the average Manhattan BG.²

#### B. Threats to Identification

What would be the best way to measure the effect of physical connectivity on innovation? In a utopian world, a neighborhood would be randomly assigned to one of two groups with different conditions. The two possible conditions would be having high or having low street network density. Through randomization, neighborhoods would not choose different street infrastructure based on their characteristics, nor would simultaneously occurring events influence this decision, allowing the researcher to cleanly estimate the effect of street network density on innovation. Although a randomized trial solves this type of identification problem, it is evidently not possible and extremely impractical in the real world given an array of associated economic and social costs. Nonetheless, this thought experiment highlights two major threats to identification that we must be aware of and address as far as possible: omitted variable bias and selection.

With regard to omitted variable bias, urban growth, economic activity, and infrastructure may be simultaneously determined, and regions that were developed earlier may have attracted more people and created more employment than younger regions. In the case of innovation, many amenities such as laboratories or even a scientific culture take time to establish. As such, it is likely that amenities necessary for innovation are found in older neighborhoods and locations. These locations may then have also continuously attracted more people who need access to such amenities: inventors. Additionally, it is feasible that some areas may have historically been more suitable for development than others even within one commuting zone. These factors may still persist today and affect both street infrastructure and other infrastructure that supports innovation. For example, a reason that a place may have been or remains more suitable for development could be access to water (Duranton & Turner, 2012).

With regard to selection, firms may choose or be forced to locate in certain areas because of their characteristics. Especially large firms might find it difficult to acquire or rent enough space to house their operations within denser metropolitan areas given geographic boundaries or restrictions imposed by the built environment. It is also feasible

---

² Until recently, applying this level of analysis has not been possible without sacrificing geographic scope or depth. Refer to section A1 in the appendix for a description of U.S. geographic boundaries and how their documentation has improved over time.
that the most innovative firms move to areas with high levels of connectivity because they value connectivity more than less innovative firms that do not rely on knowledge exchange. Alternatively, the most innovative firms could locate in less densely connected areas to avoid outward knowledge spillovers (Alcácer & Chung, 2007). In either case, selection poses a threat to identifying the actual effect physical connectivity has on innovation.

C. Addressing Threats to Identification

To best address issues of omitted variable bias, we apply a fixed effects approach on the commuting zone level. Commuting zones are clusters of counties that are characterized by strong commuting ties within commuting zones and weak commuting ties across commuting zones (Autor, Dorn, & Hanson, 2013; Tolbert & Sizer, 1996). Using commuting zone fixed effects, we can keep unobservable features of a place, such as culture, or access to amenities constant since individuals located within the same workplace and residential boundaries should be equally affected by these unobservable factors. The equation we estimate on the BG level is

\[
I_{c,b} = \alpha \text{Connectivity}_{c,b,2010} + \eta \text{Social Activity CONTROLS}_{c,b} + \theta \text{Formal Knowledge CONTROLS}_{c,b} + \beta \text{Human Capital CONTROLS}_{c,b} + \delta \text{Social Activity CONTROLS}_{c,b} + \gamma \text{Physical Geography CONTROLS}_{c,b} + \epsilon_{c,b}.
\]

In the equation, \(\epsilon_{c,b}\) represents the commuting zone fixed effects, and standard errors are clustered on the commuting zone level to account for intragroup correlation.

Our measure for innovation \(I_{c,b}\) is the number of granted patents the assignees located in a BG applied for from 2011 to 2013. By using patent application dates, we measure as much as possible the timing of innovation produced in a BG, and by counting only patents that were granted from such applications, we condition on valuable technologies (Conti & Graham, 2020).

The main independent variable of interest is Connectivity, which we measure using street density in every BG \(c\) within a commuting zone \(c\). This variable includes streets where pedestrians and automobiles are both permitted (as of 2010) and other modes of street transportation are possible. An important feature of these streets is that they are inclusive to distinct forms of movement and extend prior findings on the effect of automobiles and highways to other means of transportation and contact.

We control for components that may influence the gains from both higher physical connectivity and innovation. The SocialActivityCONTROLS include the number of bars, restaurants, and hotels in a BG. This is in line with the Saxenian argument that establishments where social interaction takes place, such as the much acclaimed Wagon Wheel bar in Silicon Valley, contribute to informal transfer of knowledge (Saxenian, 1996). Similarly, the exposure to relevant social events, and thus the locations where they take place, has been found to reduce the costs of building social ties, which in turn affect collaboration and innovation (Agrawal, 2006). We include an indicator equal to 1 if the BG has a postsecondary education campus (FormalKnowledgeCONTROLS) given that campuses are usually designed to have dense street structures and proximity to universities and other formal knowledge centers has been found to have a profound effect on the rate and direction of local research activity (Belzoni & Schankerman, 2013; Kantor & Whalley, 2014). The HumanCapitalCONTROLS consist of historic inventor counts from 2000 and 2005, as well as employment levels for 2005 and 2010, the number of college degree holders in 2000 and 2010 (by work location), and the number of working-age population that is within a 45-minute commute from a focal BG (in 2010). By holding these factors constant in the main specification, we can determine if the effect of physical connectivity persists beyond traditional measures of human capital density (Arzaghi & Henderson, 2008; Glaeser et al., 1992; Lin, 2011; Rosenthal & Strange, 2008). Similarly, we include Socio-DemographicCONTROLS to account for an explanation that could be linked to a pure agglomeration of people regardless of infrastructure (Carlino et al., 2007). These controls are population counts for 2000 and 2010. We include PhysicalGeographyCONTROLS to ensure that the effect of physical connectivity on innovation is not based on natural geographic conditions (Duranton & Turner, 2012; Hoxby, 2000). These are the areas covered by water, the area of developable land, and total land area.

In addition to the standard OLS model with fixed effects described in equation (1), we apply an instrumental variable (IV) estimation approach to address endogeneity concerns (Angrist & Pischke, 2008). In this case, an appropriate instrument to detect the causal relationship between street layout and innovation would have to be strongly related to current street networks but have little influence on today’s innovative activity other than through its effect on street layout.\(^6\)

\(^3\)Please refer to section A2 for a closer description of how commuting zones boundaries are determined.

\(^4\)An exact description of how all variables were constructed and what restrictions apply follows in the next section.

\(^5\)Note that “population” refers to the place of an individual’s residence, and “employment” refers to an individual’s place of work. Given that in most metropolitan areas across the United States, workers live in places different from where they work, and either employment or residential areas can vary from purely employment/residential to mixed use, we control for both.

\(^6\)Similar instruments used in the urban economics literature that are related to transportation infrastructure are railway lines, rivers, and highways (Agrawal et al., 2017; Duranton & Turner, 2012; Hoxby, 2000).
The main instrumental variable we use for the IV estimation is the percentage of housing units in a given BG that were built before 1940 (we also use the percentage of housing units built between 1940 and 1949 to test our model specification). A typical feature of neighborhoods built in the first half of the twentieth century is a grid-like street network structure that was constructed under the intention to grant city dwellers access to the main means of public transportation: the streetcar (Montgomery, 2013; Wells, 2013). Though built over 100 years ago, streetcar lines still have a profound effect on the local circulation of people and information given the major impact streetcars had on urban street network development (Jackson, 1985). Historically, streetcar lines were built and run by private companies anticipating short- and medium-term profits. These lines initially led to recreational sites or largely undeveloped land (Young, 2016). So-called streetcar neighborhoods were developed around the lines, usually by the same companies that ran them. The goal was to make the streetcar accessible within a short walk from all points in the neighborhood, leading to the construction of many side and connecting streets (Wells, 2013). Therefore, a typical feature of districts built in the early twentieth century was the high density of streets oriented toward transit and pedestrian traffic (Montgomery, 2013). After the world wars and with the introduction of affordable, privately owned fuel-driven vehicles, automobiles, also came the demise of the streetcar in the United States and a drastic shift in street network design. By the mid-twentieth century, most of the original streetcar companies had shut down their operations for good, and streets built after that time were no longer devised for pedestrian travel, streetcars, or other transit but primarily to accommodate cars (Wells, 2013).

Taking the percentage of housing units built before 1940 in a BG as an instrument and including relevant controls, our IV estimation can be written as follows:

First stage:

\[
Connectivity_{c,b,2010} = \theta(HU_{pre1940,c,b,2010}) + \eta(Social \ Activity \ CONTROLS_{c,b}) + \theta(Formal \ Knowledge \ CONTROLS_{c,b}) + \beta(Human \ Capital \ CONTROLS_{c,b}) + \delta(Socio-Demographic \ CONTROLS_{c,b}) + \gamma(Physical \ Geography \ CONTROLS_{c,b}) + f_c + \omega_{c,b},
\]

with \( \epsilon_{c,b} \) from equation (1), only identified if

\[
\theta \neq 0 \quad (c.1)
\]

and

\[
Cov(HU_{pre1940}, \epsilon_{c,b}) = 0. \quad (c.2)
\]

In equation (2), \( f_c \) represents the commuting zone \( (c) \) fixed effects and \( \epsilon_{c,b} \) is the error term. Condition c.1 requires that, conditional on controls, the instrument predicts the endogenous dependent variable (relevance condition). Condition c.2 denotes the exclusion restriction. In this case, the exclusion restriction entails that the percentage of housing units built prior to 1940 does not directly affect innovative output today.

Our IV estimation approach is only credible if we can make a plausible argument that condition c.2 is not violated. One reason we believe the exclusion restriction is valid is based on the changes in and spatial movement of economic and innovative activity the United States has experienced over the past century (Agrawal et al., 2017; Carlino & Kerr, 2014). Especially after World War II, many U.S. cities experienced major population shifts, as well as technology booms and busts (Klepper, 2010). In the case of individual neighborhoods, these types of trends are arguably even more volatile. What was considered a great neighborhood to live, work, or innovate in in the early twentieth century is unlikely to be so today.

Another reason we think the exclusion restriction should hold is that the percentage of housing units built prior to the 1940s is unlikely to have a direct effect on innovation other than through the indirect channels (social activity, formal knowledge, human capital, sociodemography, and physical geography) we control for. It is important to note that the exclusion restriction requires orthogonality of innovation and the percentage of housing units built before 1940 conditional on these control variables and not unconditional orthogonality (Duranton & Turner, 2012). In other words, conditional on controls, the instrument should only affect innovation today through its effect on the street network.\(^7\)

IV. Data Set Construction

The data we use in this paper come from various sources. They can be divided into two main components: location and innovation. For a description of all the variables used for estimation and their original source, refer to appendix table a1.

A. Measuring Features of a Location

To construct the variables measuring location efficiency, we use the Smart Location Database (SLD) provided by

\(^7\)For example, some areas may have historically been more suitable for development than others even within one commuting zone. This raises concerns that the percent of housing units pre-1940 and street network density depend on an omitted variable. For example, areas with water access were often developed earlier than those without (Duranton & Turner, 2012). If this type of omitted variable is important to our estimation, we would detect that including such observable physical characteristics strongly affects the results. The IV results remain qualitatively unchanged when we include or exclude water access.
the U.S. Environmental Protection Agency (EPA). This database was developed as a tool to consistently compare the attributes of locations across the country. The SLD includes demographic, employment, and built environment measures for every BG in the United States for 2010. These variables were constructed using BG boundaries from the 2010 Census TIGER shapefiles (Topologically Integrated Geographic Encoding and Referencing), data from the U.S. Census, the American Community Survey (ACS), and the U.S. Census Longitudinal Employer-Household Dynamics (LEHD) Statistics. The spatially derived variables, such as street network density, were built using the NAVTEQ (now part of the HERE Group) NAVSTREETS data set. This U.S.-wide street network includes information such as pedestrian restrictions and accessibility metrics. In order to determine the amount of land that is protected from development, information from the U.S. Geological Survey (USGS) on the protection status of public lands was included in the SLD as well as additional NAVTEQ geographic information system (GIS) layers that include water features and land use layers (Ramsey & Bell, 2014).

Connectivity is constructed using the total miles of multimodal streets in a BG, in 2010, divided by total BG area (in square miles). Following the SLD, multimodal streets are roads that can be accessed by at least two different modes of transportation (e.g., pedestrians and automobiles, hereinafter referred to as Streets). We use this category of streets since it most closely reflects the features we expect to support knowledge exchange on the microgeographic level: being inclusive to pedestrian travel and enabling auto travel that is sufficiently fast and unobstructed. The other mutually exclusive road categories are those intended primarily for pedestrian travel (referred to as Pathways and Trails) and where only auto travel (referred to as Auto Only Roads) is permitted. In figure 1 we provide a visual example of how Connectivity (= Street Miles/Total Area) is constructed using a snapshot from New York County, New York. We provide each BG’s corresponding Connectivity value in the table placed under the map and shade the BG according to these values.

An important precondition for the inclusion of commuting zone fixed effects is within commuting zone variation in Connectivity. Figure 2 depicts the 99th percentile, the 90th percentile, median, and 10th percentile for Connectivity within all commuting zones (rank ordered from the lowest to highest value of the 99th Connectivity percentile). As displayed, we can identify substantial variation within commuting zones in terms of street network density. The range of values by percentile are relatively uniformly distributed among most commuting zones, though there are strong differences between the lowest fifty and upper forty commuting zones with regard to the values representing the local 90th and 99th percentiles.

We further use the variable Accessibility provided by the SLD that captures the size of the working-age population that is within a 45-minute commute from the focal BG. To measure attributes of the physical geography of a BG, we use variables provided by the SLD that measure the area of developable land and the area covered by water. We exclude very large rural areas following the transportation literature that proposes to use a ceiling of 1 square mile (640 acres or 2.6 km²), the size of a large superblock, when analyzing street networks, since huge rural tracts are unrepresentative of the places most residents live and work and can distort averages (Ewing, Pendall, & Chen, 2003).

To construct the variable equal to 1 if the BG has a post-secondary education campus, Campus, we geolocate all campuses listed on the U.S. Department of Education’s database of accredited postsecondary institutions and programs in 2010 (U.S. Department of Education, 2018). In the first step, we search for their geocoordinates via the Google Maps Geocoding API and then join them with 2010 Census TIGER shapefiles in order to assign the corresponding BG.

We collect data from the U.S. Census County Business Pattern series to construct a variable measuring the number of bars, restaurants, and hotels (NAICS 72) in a BG and to create firm size measures (U.S. Census Bureau, 2017a). The lowest level of geography provided is the ZIP code level. Using crosswalks provided by the Missouri Census Data Center via the Geographic Correspondence Engine and the Census Bureau ZIP Code Tabulation Area (ZCTA) Relationship files (Missouri Census Data Center, 2012), we map ZIP codes to the BG level and weight accordingly since ZIP codes and BGs do not correspond perfectly. A BG boundary may encompass entire ZIP codes, and in turn, a ZIP code may cross multiple BGs. Due to inconsistencies in the ZIP code boundaries, we are missing information for historic employment and number of bars, restaurants, and hotels for some BGs.

We further include data from the Integrated Public Use Microdata Series (IPUMS) Census Demographics on the age of housing structures in a given BG, the instrument in the IV estimation approach. From IPUMS, we also collect historic decennial population counts for every BG and the level of educational attainment for workers. At the time our data were collected and assembled, information on educational attainment was publicly available only on the census tract level (Manson et al., 2017) and is missing for some BGs.

The results remain robust when excluding the forty commuting zones with the highest values.
This figure is a snapshot of an area in Harlem, New York. The white, bold lines represent BG boundaries (with corresponding numbers in black surrounded by white). The thin black lines represent all roads and streets with no further special characteristics. Trails are not displayed. Major roads/freeways, secondary roads, connecting, and important local roads are identified as shown in the legend (bottom left). The corresponding map scale can be found in the bottom right corner of the image. Each BG’s Connectivity level for the census tracts that appear fully in the image are reported in the table below. Connectivity is calculated using the total street miles in a BG oriented to both pedestrian and automobile use (Street Miles); other mutually exclusive facility categories are auto-only oriented roads and pedestrian-only oriented pathways and trails) divided by total block group area (in square miles). As a reference, the last column displays the total miles of all types of roads (All Road Types; including auto-only roads, streets, pathways, and trails). The color shading of each BG reflects the Connectivity value of the corresponding BG (Street Miles/Total Area). Image created by authors in ArcGIS using Census TIGER shapefiles and values from the EPA SLD.

<table>
<thead>
<tr>
<th>Census Tract #, Block Group #</th>
<th>Street Miles (miles)</th>
<th>Total Area (square miles)</th>
<th>Connectivity (miles/square miles)</th>
<th>All Road Types (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Tract 190, Block Group 1</td>
<td>0.20</td>
<td>0.040</td>
<td>5.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Census Tract 194, Block Group 1</td>
<td>0.01</td>
<td>0.026</td>
<td>0.31</td>
<td>0.70</td>
</tr>
<tr>
<td>Census Tract 194, Block Group 2</td>
<td>0.00</td>
<td>0.014</td>
<td>0.00</td>
<td>0.46</td>
</tr>
<tr>
<td>Census Tract 194, Block Group 3</td>
<td>0.00</td>
<td>0.014</td>
<td>0.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Census Tract 194, Block Group 4</td>
<td>0.00</td>
<td>0.014</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Census Tract 196, Block Group 1</td>
<td>0.10</td>
<td>0.010</td>
<td>10.10</td>
<td>0.30</td>
</tr>
<tr>
<td>Census Tract 196, Block Group 2</td>
<td>0.10</td>
<td>0.024</td>
<td>4.10</td>
<td>0.86</td>
</tr>
<tr>
<td>Census Tract 196, Block Group 3</td>
<td>0.15</td>
<td>0.035</td>
<td>4.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Census Tract 198, Block Group 1</td>
<td>0.35</td>
<td>0.050</td>
<td>7.00</td>
<td>1.80</td>
</tr>
<tr>
<td>Census Tract 198, Block Group 2</td>
<td>0.00</td>
<td>0.040</td>
<td>0.00</td>
<td>2.50</td>
</tr>
<tr>
<td>Census Tract 200, Block Group 1</td>
<td>0.10</td>
<td>0.015</td>
<td>6.70</td>
<td>0.35</td>
</tr>
<tr>
<td>Census Tract 200, Block Group 2</td>
<td>0.25</td>
<td>0.035</td>
<td>7.10</td>
<td>0.93</td>
</tr>
</tbody>
</table>

**B. Measuring Innovation**

Our measure for innovation is the number of U.S. granted patents located in a BG that were applied for between 2011 and 2013. To construct this variable, we use the Morrison, Riccaboni, and Pammolli (2017) disambiguated patent data set, which contains geocoordinates of all patent assignees and inventors registered in the USPTO, WIPO, and EP patent databases, from 1975 to 2013. The information for this data set is sourced from Harvard’s Dataverse Project for USPTO patents and from both the RegPat and Citation databases of the OECD. By joining the geocoordinates of both assignees and inventors with 2010 Census TIGER shapefiles of USA Block Group Boundaries\(^{13}\) in ArcGIS, we are able to obtain the corresponding BGs for every inventor’s and assignee’s location.

\(^{13}\)This data can be found at https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.2010.html.
In the next step, we identify, as far as possible, where the creation of an idea took place. Since inventors may use their residential address and assignees may use one central address handling all intellectual property, we do not know for certain from the data where the idea actually originated. We apply a conservative approach to determine the most likely set of patents that were created in a specific place. To do so, we take both the location of all inventors and all assignees of a patent and determine their corresponding commuting zone (Autor et al., 2013). If all inventors are in the same commuting zone as an assignee, we include that patent in the sample and link it to the matched assignees’ location.\(^{14}\)

The Morrison et al. (2017) disambiguated patent data set locates assignees and inventors using the Yahoo Geocoding API and is missing exact street-level information for a number of assignees. To increase the sample of patents from inventors and assignees in the same commuting zone, we conduct a further search for the assignees who had not been successfully located on the exact street level in the original data set. Using the Google Maps Geocoding API, we query the geocoordinates of the assignees by specifically searching for the assignee name, conditional on being in the specific state and county code listed in the Morrison et al. (2017) disambiguated patent data set. Our search approach returns coordinates only if the name queried is found within the strict geographic boundary conditions provided. The end sample consists of 42,259 assignees located on the exact street address for 2011 to 2013.

From the Morrison et al. (2017) data set, we further attain information on the number of inventors and citations. From this information, we construct variables that capture the number of historic inventors in a BG for 2000 and 2005. These are determined using the exact geocoordinates of all inventors and matching these to TIGER Census Boundaries in ArcGIS.

Table 1 displays summary statistics of all variables for the 122,899 BGs in the end sample. The number of aggregate patents in a BG is highly skewed, with an average of 0.18 and a maximum value of 1,025. Of the BGs in our sample, 4,916 applied for at least one patent between 2011 and 2013. Similarly, the physical network structure of BGs across the United States also varies strongly, with an average of 2.83 miles of streets, to a maximum of 466.98 miles of streets divided by BG area (the next highest value is 144.72). The average percentage of housing units that were built before 1940 is 20. We also report key descriptives for all the control variables.\(^{15}\)

V. Results

A. OLS Regression Results

As laid out in an earlier section, threats to identification are a serious concern, a reason that the OLS results serve primarily as a description of the relationship between Connectivity and patenting output with no claims to causality. We first estimate equation (1) with commuting zone fixed effects and with county fixed effects for robustness. The dependent variable is the number of granted patents applied for by assignees located in a BG. In order to estimate the OLS fixed effects model, we log-transform the dependent variable.

Table 2 presents the results of the regressions predicting the change in log patenting as a function of physical connectivity (Connectivity), social activity, formal knowledge, human capital, sociodemographic, and physical geography controls.\(^{16}\) The reported standard errors are robust and clustered on the commuting zone (county) level. Column 1 reports the relationship of patenting with Connectivity only, and column 2 includes the number of bars, restaurants, and hotels in a BG. Column 3 presents the results with an indicator equal to 1 if the BG encompasses a postsecondary education.

\(^{14}\)We follow the same procedure using looser constraints. In table A3, we report the main results using all patents where at least one inventor is in the same commuting zone as the assignee and locating them in the BG of the assignee. The point estimate and standard errors are slightly larger. In the subsequent analysis, we use the stricter approach described in the main text. When there are multiple assignees, we keep the assignee that matches all inventors on the commuting zone. See figure A1 for a stylized depiction of the approach just described.

\(^{15}\)See figure A2 for a visual depiction of the relationship between patenting and Connectivity where we label outliers across the United States. We provide separate plots for California and Massachusetts to provide more evidence that our main results are not driven by one state or region alone and to support our choice of the BG level given extreme within-city variation (which is especially visible for Boston and San Francisco/Bay Area).

\(^{16}\)Note that the number of observations vary depending on the variables included in the regression model. This is due to missing data for the variable Number Bars (short form of the measure for number of bars, restaurants, and hotels in a BG) and College, as described earlier. In addition, we run the models using only commuting zones where there was at least one patent to ensure comparability across models.
Table 1.—Summary Statistics

<table>
<thead>
<tr>
<th>Census Block Group Level</th>
<th>Minimum</th>
<th>p25</th>
<th>Mean</th>
<th>p50</th>
<th>p95</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Patents (2011–2013)</td>
<td>0.00 0.00 0.18 0.00 0.00 1,025.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patent (⇐ 0/1)</td>
<td>0.00 0.00 0.04 0.00 0.00 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge Exchange</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Same BG Citations</td>
<td>0.00 0.00 0.03 0.00 0.00 612.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Same BG Citations</td>
<td>0.00 0.00 0.03 0.00 0.00 607.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Self Same BG Citations</td>
<td>0.00 0.00 0.002 0.00 0.00 33.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical Network Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connectivity (in miles per square mile)</td>
<td>0.00 0.86 2.83 2.20 7.94 466.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HU/pre1940</td>
<td>0.00 0.00 0.20 0.06 0.73 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HU/1940-1949</td>
<td>0.00 0.00 0.08 0.04 0.31 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social Activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Bars, Restaurants, and Hotels</td>
<td>0.00 1.00 2.54 2.00 6.00 143.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Formal Knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campus</td>
<td>0.00 0.00 0.01 0.00 0.00 1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Human Capital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessibility (in thousands)</td>
<td>0.00 97.74 292.91 194.05 1,039.11 1,598.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment 2010 (in thousands)</td>
<td>0.00 0.05 0.55 0.16 2.03 232.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment 2005 (in thousands)</td>
<td>0.00 0.00 0.42 0.07 1.70 167.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inventors 2005</td>
<td>0.00 0.00 0.02 0.00 0.00 18.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inventors 2000</td>
<td>0.00 0.00 0.02 0.00 0.00 10.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sociodemographic (in thousands)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population 2010</td>
<td>0.00 0.88 1.32 1.19 2.48 19.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population 2000</td>
<td>0.00 0.88 1.28 1.17 2.33 12.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical Geography (in hundred acres)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area Water</td>
<td>0.00 0.00 0.05 0.00 0.23 5.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area Developable Land</td>
<td>0.00 0.73 1.79 1.38 4.85 6.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area Land</td>
<td>0.004 0.77 1.86 1.45 5.02 6.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>122,899</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

campus. Column 4 presents regression results with employment in 2005 and 2010. Column 5 includes population controls as measured in 2010 and 2000. The results reported in column 6 display the relationship between Connectivity and patent output, including a measure for the number of workers who live within a 45-minute driving distance from a focal BG (Accessibility). Column 7 presents the relationship with historic inventor counts and the number of workers by work location with an undergraduate college degree or higher. The model in column 8 consists of Connectivity and the physical geography control variables, and column 9 shows the full model with all controls. Column 10 reports the results of the full model using county fixed effects. Overall, there is little change to any of the coefficients comparing the commuting zone and county fixed effects models. The coefficient on Connectivity in the full model suggests that a 1% increase in Connectivity is associated with a 0.004% increase in patenting. The results indicate that all of the included types of controls explain some of the relationship of Connectivity and patent output. Individually, the strongest control variable is contemporaneous employment. Including employment measures (2010, 2005) quarters the Connectivity coefficient.

In a next step, we analyze if the relationship between physical connectivity is linear or could possibly be driven by a few outliers. To do so, we run the full model using deciles of the connectivity measure as our main independent variable of interest. Figure 3 presents the results from this estimation displaying the coefficients of the connectivity measure by decile. The figure clearly illustrates a nonlinear trend and highlights that most of the effect seems to be driven by the upper two deciles. The BGs in these deciles are not concentrated in one state, but in fact are dispersed among various states. The ten states with the highest number of BGs in the top decile are, in descending order, California, New York, Texas, Massachusetts, Pennsylvania, New Jersey, Illinois, Florida, Maryland, and Tennessee.

We further examine the relationship between Connectivity and patenting using an indicator equal to 1 if a BG had

---

17 Given that the accessibility coefficient is very small, we exclude this variable in the IV estimation models to increase degrees of freedom.

18 Note that in the model with county fixed effects, additional singleton observations are dropped. As such, the sample is slightly different, leading to distinct point estimates.

19 We run the full model using different time spans of the dependent variable and patents from the USPTO only. The results remain robust. We further use a fixed effects Poisson model (see table A4) and count data to see if the directionality and statistical significance hold. Both models confirm the findings of the OLS estimation.

20 In figure A3, we report the relationship between Connectivity and all continuous controls together, as well as the individual relationship between Connectivity and the control variables used in the fully specified model. Note that here we can identify that the employment measures and Connectivity are the most strongly correlated of all controls.

21 In Table A5, we provide the results from estimating columns 1, 8, and 9 using alternative measures of Connectivity: Pathway and Trail Density, Connectivity including pathways and trails, Intersection Density, and Transit Frequency. With the exception of Pathway and Trail Density, the full model holds using these alternative measures. A likely reason is that areas with many pathways and trails are parks and other recreational areas with no or little economic activity.
at least one patent and 0 otherwise. The results reported in table 3 show similar magnitudes of the coefficients as compared to table 2, where we use the natural log amount of patenting as the dependent variable. Column 1 shows the reduced model without any controls, and columns 2 and 3 present the fully saturated model with commuting zone and county fixed effects. Together, these results provide suggestive evidence that, albeit small in magnitude, there is a robust relationship between physical connectivity and innovation.

Theoretically, we may have expected the increase in patenting to be larger on the intensive margin than on the extensive margin. Although a comparison of the results from tables 2 and 3 may suggest a larger effect on the extensive margin of patenting, it is still plausible that our theoretical argument holds. One possible scenario is that by increasing knowledge exchange, physical connectivity leads to more ideas or induces latent ideas to improve in quality, and this, in turn, makes it more likely that ideas cross the threshold of patentability.

### B. IV Results

Next, we estimate equation (2). First-stage results as presented in table 4 indicate that the instrument, the percentage of housing units built pre-1940 (HUpre1940) taken alone, is strong with $F$-statistics of over fifty in the model, including all geographic area variables. The results reported in table 4 show similar magnitudes of the coefficients as compared to table 3, where we use the natural log amount of patenting as the dependent variable. Column 1 shows the reduced model without any controls, and columns 2 and 3 present the fully saturated model with commuting zone and county fixed effects. Together, these results provide suggestive evidence that, albeit small in magnitude, there is a robust relationship between physical connectivity and innovation.
controls. Table 4 further presents the second-stage results obtained from estimating equation (1) instrumenting Connectivity with HUpre1940. We apply commuting zone fixed effects and report robust standard errors. As in the OLS regressions, we log-transform the dependent variable. Column 1 displays the IV estimation results without controls except PhysicalGeographyCONTROLS, column 2 presents the results adding SocialActivityCONTROLS, and an indicator equal to 1 if the BG encompasses a postsecondary education campus, column 3 adds Socio-Demogr.CONTROLS to the equation, and column 4 excludes Socio-Demogr.CONTROLS, but adds HumanCapitalCONTROLS to the equation. Including all controls (column 5) almost halves the magnitude of the Connectivity coefficient in comparison to the results in column 1. These results suggest that the IV estimation is highly sensitive to the inclusion of distinct controls. In order to provide more support that our model is correctly specified, we include a further instrument in column 6: the percentage of housing units built before 1940 and 1949 (HU1940-1949). Using two instruments allows us to apply Hansen’s J-test evaluating overidentification assumptions. Although we cannot test the exclusion restriction directly, a nonstatistically significant Hansen’s J-statistic gives us more confidence in the validity of the model. It is, however, important to point out that this approach only tests overidentifying assumptions conditional on having one correctly identified instrument. The first-stage F-statistics remain strong, albeit slightly weaker than in the model with one instrument. Taken together, the results of the full model (column 6) can be interpreted such that, conditional on controls, a 1% increase in Connectivity causes a 0.04 percent increase in patenting output.\footnote{To confirm the choice of the age categories used in the IV estimation, we run the full IV model with all possible age categories. The results are presented in table A6. Note that more recent Housing Age Categories are far more likely to violate the exclusion restriction (closer in time) and negatively affect street density given the rise of automobile transportation post-1950s. In table A7, we report the IV results including pathways and trails in our measure of Connectivity. The results remain robust, with slightly smaller point estimates. We report the results from implementing a control function approach (CF), using a Poisson model in the second stage in table A8. The outcome we report is the number of patents in a BG, and the coefficients reflect a much larger local average treatment effect than an average treatment effect (e.g., through negative selection). We cannot empirically rule out that the exclusion restriction is violated, and our estimation relies heavily on the assumption that the percentage of housing units built before 1940 (and from 1940 to 1949) only affects innovation via its effect on the street network conditional on controls. However, the results from Hansen’s J overidentification test provide a reconciliation of our results with those provided by previous research is useful. Note, however, that to date, most studies examining the relationship between urban features and patenting have largely been on the MSA level. One example is Agarwal et al. (2017), who find that a 10% increase in highways leads to a 1.7% increase in patenting over five years on the MSA level. Similarly, Carlino et al. (2007) find that a 10% increase in employment density results in a 2% increase in patent intensity (patents per capita) over a ten-year period. In addition, over the past years, much research effort as been put into applying a more microgeographic lens on the foundations of agglomerations. For instance, Rosenthal and Strange (2008) determine their unit of analysis based on concentric rings, finding strong evidence of an urban wage premium. Their results suggest that the elasticity of wage to the number of workers within 5 miles is about 4.5%. Also using distance rings, Arzaghi and Henderson (2008) find that a 1 unit increase in the number of neighboring advertising agencies within 250 meters results in an increase of new establishment births by 2%. In comparison to these studies, our estimates are smaller. This is not unexpected since the time period we examine is shorter and the unit of analysis is at a very microlevel. Larger geographic areas tend to confine direct responses and are therefore likely to overstate the size of local point estimates. We find that a 10% increase in Connectivity is associated with a 0.05% to 0.2% increase in patenting in the OLS model and results in a 0.4% to 0.96% increase in the IV model. Using IV estimates from the fully saturated model are used, an increase from the 25th percentile of Connectivity (0.86) to the 95th percentile (7.94), would roughly translate into a 35% increase in patenting. The magnitude of the coefficient on Connectivity is larger in the IV than in the OLS model. Three possible reasons why this is the case are that (a) the exclusion restriction is violated, (b) there may be reverse causation, or (c) the results reflect a much larger local average treatment effect than an average treatment effect (e.g., through negative selection). We cannot empirically rule out that the exclusion restriction is violated, and our estimation relies heavily on the assumption that the percentage of housing units built before 1940 and 1949 (HU1940-1949). Using two instruments allows us to apply Hansen’s J-test evaluating overidentification assumptions. Although we cannot test the exclusion restriction directly, a nonstatistically significant Hansen’s J-statistic gives us more confidence in the validity of the model. It is, however, important to point out that this approach only tests overidentifying assumptions conditional on having one correctly identified instrument. The first-stage F-statistics remain strong, albeit slightly weaker than in the model with one instrument. Taken together, the results of the full model (column 6) can be interpreted such that, conditional on controls, a 1% increase in Connectivity causes a 0.04 percent increase in patenting output.\footnote{To confirm the choice of the age categories used in the IV estimation, we run the full IV model with all possible age categories. The results are presented in table A6. Note that more recent Housing Age Categories are far more likely to violate the exclusion restriction (closer in time) and negatively affect street density given the rise of automobile transportation post-1950s. In table A7, we report the IV results including pathways and trails in our measure of Connectivity. The results remain robust, with slightly smaller point estimates. We report the results from implementing a control function approach (CF), using a Poisson model in the second stage in table A8. The outcome we report is the number of patents in a BG, and the coefficients represent incidence rate ratios. A special feature of the CF is that it enables us to study the nature of self-selection (Wooldridge, 2015). As reported, the residuals suggest that there is negative selection into places with high levels of Connectivity. Coefficients smaller than 1 indicate a lower incidence rate ratio (the equivalent of a negative sign in the OLS regressions).
some support that our model is correctly specified. With regard to reverse causation, it may be that BGs that experienced negative shocks to innovation, conditional on controls, also experience positive shocks to the number of streets. Since our data only include details on street infrastructure from 2010, we cannot test this hypothesis. Viewing the third explanation, it is possible that the IV is shifting the “behavior” of a subgroup of BGs for which the returns to Connectivity are larger than average, such as central business districts and BGs in the upper quintile of Connectivity. If the local average treatment effect is larger than the average treatment effect, it is plausible that IV estimates are larger than OLS estimates because of heterogeneity in the sample we are analyzing.

As discussed earlier, one manifestation of selection—in this case, negative selection—could be that large firms are opting out of locating in areas with high levels of Connectivity. Compared to smaller firms, it is likely that large firms would benefit less from physical connectivity in the first place (which could explain differences between average and local average treatment effects). We base this on the assumption that larger firms already have access to an abundance of skills and knowledge sources in-house and may therefore not rely on external exchange as much.\footnote{For further discussion of this potential explanation based on BG heterogeneity in firm size composition, see section A4. Results displayed in figure A4 and table A9 provide support for this explanation.}

C. Is There Any Knowledge Exchange?

Our main set of results provides evidence that physical connectivity increases innovation. The question remains if increased knowledge exchange is a possible channel through which denser street networks affect innovation. One way to test if Connectivity indeed affects innovation via its effect on knowledge exchange is to examine knowledge flows of actors in a BG. A conservative approach to measuring such knowledge flows is using patent citations (Belenzon & Schankerman, 2013; Jaffe et al., 1993; Thompson, 2006). Naturally, not all citations represent knowledge flows, but studies comparing citation data with surveys of inventors have detected a strong correlation between patent citations and knowledge flows (Duguet & MacGarvie, 2005; Hall, Jaffe, & Trajtenberg, 2005). In order to examine the relationship between physical connectivity and knowledge flows, we create (a) a count of citation pairs in a BG, excluding self-citations, and (b) a count of self-citation pairs within a BG. We construct citation pairs using patents that were applied for between 2010 and 2013, the patents these cite, and counting and struct citation pairs using patents that were applied for between 2010 and 2013, the patents these cite, and counting

\[ \text{Connectivity} = \beta_1 \text{Connectivity} + \beta_2 \text{Sociodemographic Controls} + \beta_3 \text{Formal Knowledge Controls} + \beta_4 \text{Social Activity Controls} + \epsilon \]

This table reports the results obtained from instrumenting Connectivity with HUpre1940 (columns 1–6) and HU1940-1949 (column 6). In the First Stage, the outcome variable is Connectivity (log). HUpre1940 is the percentage of housing units built before 1940 and HU1940-1949 is the percentage of housing units built between 1940-1949. The SocialActivityCONTROLS include the number of bars, restaurants, and hotels in a BG. FormalKnowledgeCONTROLS is an indicator equal to 1 if the BG has a postsecondary education campus. HumanCapitalCONTROLS consist of historic inventor counts from 2000 and 2005, as well as employment levels for 2005, and 2010, and the number of college degree holders in 2000 and 2010 (by work location). Socio-DemographicCONTROLS are population counts for 2000 and 2010. PhysicalGeographyCONTROLS are the area covered by water, the area of developable land, and total land area. Variation in the number of observations depending on the included controls is due to missing values for the number of bars, restaurants, and hotels, as well as college education. We report First Stage F-statistics in all columns and the p-value obtained from Hansen’s J-statistic, which tests the validity of the overidentifying restrictions in column 6. We cluster standard errors (in parentheses) at the commuting zone level. *p < 0.10, **p < 0.05, ***p < 0.01.
imply that Connectivity increases competitive pressures pushing organizations to patent strategically (Singh, 2005) or create patent thickets (Shapiro, 2000).

Table 5 presents the results for the two citation outcomes we estimate using equations (1) and (2) instrumenting Connectivity with the percentage of housing units that were built before 1940 and 1940 to 1949, and including the number of patents applied for between 2011 and 2013 in a given BG as a control. Column 1 presents the main effect of Connectivity on non-self-citations within a BG (mean of 0.03) only including PhysicalGeographyCONTROLS. Column 2 reports the results for non-self-citations using the full model, column 3 presents the IV model with all controls, and column 4 includes the log number of patents as a further control. Columns 5 to 8 present the corresponding models using self-citations (mean of 0.002) as the outcome variable.

Table 5 includes Pathways and Trails in our measure for physical connectivity and patenting we detect in the previous set of results and that knowledge exchange is a feasible channel. Our findings indicate that actors use relatively more local external knowledge sources in BGs with denser street networks. Besides the specific type of technical knowledge exchange, which occurs and is captured by patent citations, it is likely that at least some of the innovation productivity advantage found in the earlier set of results is also related to the exchange of other types of knowledge and increased interaction efficiency. Higher productivity could, for example, stem from knowledge on how to better organize a lab that individuals learn about via informal conversations (e.g., at bars) or from knowledge on how to better organize a lab that individuals learn about via informal conversations (e.g., at bars) or from knowledge on how to better organize a lab that individuals learn about via informal conversations (e.g., at bars).

D. Interactions between the Physical and Social Space

Previous research has provided evidence of the importance of social factors, such as population and employment density, as well as local meeting points in explaining regional innovation differences. The main channel underlying the relationship between these social factors and inventive activity is similar to what we propose in this paper: density influences interpersonal exchange. As such, examining the interaction of social factors with physical connectivity could provide more insight into the role of interpersonal exchange as a channel driving our main results.
To do so, we run OLS regressions including interaction terms of the physical and social space. The corresponding results are reported in table 6. In columns 1 and 2, we interact Connectivity with an indicator equal to 1 if population is over 1,650 and equal to 0 otherwise. In columns 3 and 4 we use an indicator equal to 1 if the number of bars, restaurants, or hotels is over 5 and is 0 otherwise, interacting this with Connectivity. Columns 5 and 6 report the results from including the interaction between physical connectivity and an indicator equal to 1 if employment is over 950, and is 0 otherwise. The even-numbered columns present the estimates only including PhysicalGeographyCONTROLS, and the uneven numbered columns display the results using the full model with all controls.

The main effect of Connectivity remains statistically significant on conventional levels and is positive across all models (meaning when population is equal to or below 1,650, employment is equal to or under 950, and there are five or fewer bars, there is still an effect of Connectivity). Overall, there seems to be a positive interaction relationship between high levels of population and Connectivity, although the main effect of High Population is negative. The results further indicate a positive interaction relationship between high levels of employment and Connectivity. In the case of High Employment, the main effect is also positive and statistically significant on conventional levels, meaning that with no Connectivity, having a high level of employment has a positive impact on innovative output. The interaction between High No. Bars and Connectivity is positive and statistically significant on conventional levels. However, the main effect of High No. Bars no longer holds when including the interaction term. This suggests that without Connectivity, there may be no additional effect of elevated numbers of bars, restaurants, and hotels in a BG on patenting.25

Taken together, these findings indicate that high levels of population, employment, as well as bars, restaurants, and hotels may be complementary to physical connectivity. This backs the idea that the physical layout of a place quite plausibly affects innovation by facilitating exchange among individuals.

E. Limitations

There are several limitations to this study. One is that patents are not the ideal measure of innovation given that not all types of innovation are patentable. In fact, in some industries, inventors rarely seek patent protection, but resort to other mechanisms such as secrecy or first-to-market advantages instead (Cohen, Nelson, & Walsh, 2000). A reason why inventors do not seek patent protection are the high costs associated with patent filing (Graham et al., 2009). From this, it could be that we are measuring a specific type of innovation only, or it could be that we are possibly capturing a BG culture of patenting or propensity to patent. For example, it is plausible that the patenting behavior of one or more actors in close proximity makes it necessary for all actors to patent.

25To add more transparency about the data and to show that outliers are not driving our results, we also present the interactions with high levels of population, employment, and bars in visual form (see figure A5).
In addition, we only include those patents where we locate inventors and assignees in the same commuting zone. It is possible that this sample selection approach introduces a bias that goes toward underestimating the patenting output of large corporations. Large corporations may tend to centrally organize the handling of their intellectual property at established headquarters. Consequently, it could be that the assignees in our sample are, on average, smaller than the general population of firms. Though possible, our robustness checks do not indicate a large systematic size bias. 26

A further limitation to this study is that we can only proxy knowledge exchange and do not directly measure it. Capturing actual interaction between actors is a tedious and difficult endeavor. Not only does it require very microlevel data, but also close observation of the behavior of individuals. There have been advances in tracking the possible ways that interaction takes place within larger human agglomeration, such as Williams and Currid-Halkett (2014). Over two weeks, the authors tracked 77 fashion designers working in the Garment District and the larger New York region. Using cell phone data and social-media tools, they captured geographical movements and documented exact real-time data. A similar approach covering a larger geographic area may be possible in the future.

Another limitation is that we base our analysis on cross-sectional data. To get closer to understanding actual selection processes, we would need a panel data set. Although most of the variables in this current data set are available for multiple years, we only have access to information on infrastructure for 2010.

VI. Discussion and Conclusion

In a very literal sense, this paper is taking innovation from being up "in the air" (Marshall, 1890) to the streets and makes two main contributions to the empirical literature on geography and innovation. First, we use a unique data set covering the entire contiguous United States on the smallest geographic entity for which information on street infrastructure is available. Previous research has not been able to apply such a microgeographic lens to assess innovation outcomes. Second, we go beyond the traditional location externalities found in the literature. For example, including city layout may help us understand why certain regions and firms can exploit diverse or specialized knowledge better than others. Second, it would be interesting to assess the effect of physical connectivity on other measures of innovation such as trademarks. Third, in order to make recommendations for firm location choice, it would also be relevant to better comprehend who benefits (and loses) from proximity and the capacity of a place for connecting people.

In 1922, Henry Ford stated that “the modern city is probably the most unlovely and artificial site this planet affords. The ultimate solution is to abandon it . . . We shall solve the City Problem by leaving the city” (in Wells, 2013, 63). About 100 years later, this statement stands corrected. Leaving the dense street network the city provides is hardly the solution—at least not for innovation.

REFERENCES


26 See figure A6 for the results from comparing kernel density distributions of assignee size (determined by the amount of patents) for the sample and full data set.

