



The Review of Regional Studies

The Official Journal of the Southern Regional Science Association



Universities, Agglomeration, and Regional Innovation*

Michael J. Orlando,^a Michael Verba,^b and Stephan Weiler^c

^a *Econ One Research, Inc. and University of Colorado Denver, USA*

^b *Department of Public Governance, Tilburg University, NL*

^c *Department of Economics, Colorado State University, USA*

Abstract: Urban agglomeration is an important correlate to regional innovation. Large population centers pool knowledge workers and facilitate spillovers essential to innovative activity. And large populations provide more cost-effective locations for non-labor inputs to innovation, including local infrastructure that may facilitate innovative activity. However, university locations may also agglomerate these innovative inputs, even absent the agglomerative effects of large populations. Regional policymakers may find it useful to differentiate between various correlates to innovation. This paper exploits the collinearity of universities and population with regional human capital to apportion the relationship between these regional correlates of innovation into human-capital related and non-human-capital related channels. We identify a correlation between universities and regional innovation that reflects a relationship between innovation and regional human capital correlated with university presence. None of this relationship can be apportioned to factors correlated with university presence and uncorrelated with local human capital. A key methodological contribution of this paper is the analytical framework, which can be extended to a larger number of aggregate factors and causal channels.

Keywords: universities, agglomeration, geography, innovation, knowledge spillovers, patents

JEL Codes: R3, R1, H4, L3

1. INTRODUCTION

A number of studies have explored the determinants of regional innovation. Several studies highlight the role of universities as a key factor behind high regional innovative performance. They show that universities contribute to regional innovation both directly, through innovations developed inside university labs, and indirectly, through localized knowledge spillovers—by enhancing research and development activities of firms in close proximity to

*We wish to thank the anonymous reviewers for comments that greatly improved this manuscript.

Michael J. Orlando is Managing Director of Econ One Research, Inc. and Lecturer in the Global Energy Management Program at the University of Colorado Denver. Michael Verba is an Assistant Professor of Public Governance at Tilburg University. Stephan Weiler is a Professor of Economics at Colorado State University. *Corresponding Author:* Michael J. Orlando, Email: morlando@econone.com

universities (Anselin et al., 1997; Cowan and Zinovyeva, 2013; Jaffe, 1989; Ponds et al., 2010). Another set of studies demonstrate the importance of urban agglomeration in supporting high rates of regional innovation. Highly populated geographic areas are shown to support significantly more innovations per resident than less populated areas (Audretsch and Feldman, 1996; Carlino et al., 2015; Jaffe et al., 1993; Ó hUallicháin, 1999; Orlando and Verba, 2005). However, the interrelation between pathways through which innovative factors may contribute to regional innovation has not been sufficiently explored.

This paper considers, simultaneously, how university presence and urban agglomeration are related to regional innovation. Moreover, we exploit a channel through which these factors are related to innovative output—human capital—in order to apportion factor correlations with innovation into human-capital and non-human-capital channels. We find that a positive relationship between universities and regional innovation reflects the relationship between innovation and the regional human capital correlated with university locations. None of the correlation between university presence and innovation can be attributed to non-human capital channels. The relationship between innovation and university presence reflects a correlation between innovation and locations with doctoral degree granting programs at all population levels excepting the 11 percent of most congested county areas in the United States. This relationship also reflects a correlation between innovation and locations with Master's degree granting programs at all population levels except 11 percent of the most congested and 33 percent of the least populated county areas in the U.S.

We also identify a correlation between population and regional innovation across the entire sample. The relationship between innovation and population reflects correlations in 11 percent of the most congested and 33 percent of the least congested county areas in the U.S. The former corresponds to a positive correlation between human capital and population. The latter corresponds to a relationship between non-human capital correlates to population and innovation. An important contribution of this paper is the analytical framework, which can be extended to consider a larger number of hypothetically causal factors related to innovation.

The next section motivates the study in the context of the relevant literature. Section 3 discusses the analytical model and data used in the analysis. Section 4 presents the results. The concluding section summarizes the implications of these findings.

2. LITERATURE REVIEW AND MOTIVATION

A number of empirical studies establish that innovative activity is more prevalent in highly populated regions. Ó hUallicháin (1999), for example, finds that the largest metropolitan areas tend to be the most innovative. Carlino et al. (2015) compare patent productivity in 280 metropolitan areas and find that the number of patents issued per capita increases with the size of the metropolitan labor market. According to data from the 1990's from the U.S. Census Bureau and the U.S. Patent and Trademark Office (USPTO), the top third of counties ranked by regional population account for less than 68 percent of total population but over 80 percent of total patent activity. The bottom third of counties account for nearly 11 percent of total population but less than five percent of total patent activity.

The correlation between population and innovation is widely believed to reflect several aspects of greater population that enhance innovative activity. First, populous places create greater opportunities for knowledge spillovers. The term “knowledge spillovers” refers to the ideas acquired by one person or organization that are attributable to another person’s or organization’s effort. Innovators learn about each other’s ideas through professional and social interaction. Since many such interactions take place locally, much of the knowledge that spills beyond firm boundaries is confined to a particular region. Such face-to-face learning is what Storper and Venables (2004) call “the buzz” that facilitates agglomeration in big cities.¹

Populous places also provide thick markets for specialized inputs to innovation (Glaeser, 1998). The innovative process often requires specialized equipment and services. Buyers and suppliers of these specialized inputs benefit from larger markets because they reduce uncertainty and allow market participants to leverage fixed costs. As a result, thick markets allow innovators to acquire specialized inputs more reliably and at a lower cost. As with knowledge spillovers, thick markets also enhance the innovation output of human capital in more populous places.

Finally, greater population can directly enhance the level of human capital necessary for innovation. Innovation often requires high levels of human capital. Moreover, innovative occupations can be highly specialized. Thick markets for specialized, innovative occupations reduce workers’ cost of job search and increase the number of suitable candidates in demand by innovative firms. As a result, workers with high levels of human capital and firms with specialized labor demands will agglomerate in more populous places.

Educational institutions may substitute for traditional agglomeration economies by bringing together, in a particular place, innovative inputs (Goldstein and Drucker, 2006), even in a location that is not densely populated. Universities increase local human capital directly by employing a large number of specialized, skilled occupations. In addition, universities enhance the level of human capital by educating students, some of whom remain in the same area (Abel and Deitz, 2011). In fact, scientific and engineering employment is more heavily concentrated around universities. For example, Table 1 shows that human capital inputs to innovation tend to be more highly concentrated in metropolitan areas with a large number of universities.² Furthermore, through their research labs, universities embody capital inputs in the production of innovation. Data from the National Science Foundation indicate that as much as 14 percent of all research and development in the U.S. is carried out by universities. Finally, together with high levels of human capital around universities, knowledge infrastructure in the form of libraries and scientific organizations may increase knowledge spillovers.

A number of studies suggest that university presence is associated with a higher level of innovation in the surrounding region. According to data from the National Council of Education Statistics and the U.S. Patent & Trademark Office, 42 percent of U.S. counties that

¹Also, see Jaffe et al. (1993), Mowery and Ziedonis (2001), and Orlando (2004).

²Includes research performed by colleges and universities (12 percent of U.S. R&D) and by federally funded college and university research and development centers (2 percent of U.S. R&D). Data from the National Science Foundation, “State distribution of expenditures for R&D, by performing sector and source of funding: 1999.”

Table 1: High-Tech Employment in Metropolitan Areas, by University Presence, 1999

| Number of Universities | Scientists and Engineers per Thousand Population |
|--------------------------|---|
| Equal or greater than 50 | 8.8 |
| Between 25 and 49 | 11.8 |
| Between 5 and 24 | 6.5 |
| Equal to or less than 4 | 2.1 |

Notes: Data is presented for 310 metropolitan areas. Metropolitan areas outside New England are based on the U.S. Census Bureau's 1993 MSA/PMSA definitions. Metropolitan areas in New England are based on the U.S. Census Bureau's 1993 NECMA definitions.

Source: Author calculations based on data from the U.S. Department of Labor Occupational Employment Statistics for 1999.

were host to a degree-granting educational institution produced 92 percent of patents issued during the 1990s.³ Prior research has linked the role of universities in regional innovation to localized knowledge spillovers. Much of this research has focused on knowledge transfers from universities to local private firms (Jaffe, 1989; Acs et al., 1992, 1994; Anselin et al., 1997; Varga, 2000; Ghinamo, 2012).⁴ Kim et al. (2005) find that university researchers play an important role as pathways for the diffusion of ideas from university labs to industry.

The diffusion of ideas appears to be particularly strong in areas immediately surrounding universities. Mowery and Ziedonis (2001), for example, conclude that knowledge flows from universities to industry (measured by licensing agreements) tend to be geographically localized. Though the direct contribution of universities (as opposed to the contribution through firms) to innovation is generally outside the focus of studies of university-to-industry knowledge transfers, the previous discussion suggests that the role of universities in innovation should not be ignored. This is particularly true in the era following the passage of the Bayh-Dole Act, which opened the door for universities to become owners of intellectual property resulting from federally-funded research.

The analytic framework developed in the next section allows us to explore the complex relationship between population, universities, and regional innovative activity. The methodology decomposes these aggregate factor inputs into a pathway associated with human capital agglomeration as well as residual pathways.

3. EXPERIMENTAL DESIGN

This study models innovation using a regional knowledge production function framework. The estimating strategy identifies the importance of population, universities, and human

³Counties with degree-granting educational institutions were defined as those that were host to an active educational institution granting Associates, Bachelors, Masters, or Doctoral degrees between 1996 and 2001.

⁴For a review and meta-analysis of this literature, see Ghinamo (2012).

capital for innovation. However, because population and university locations are collinear with human capital, individual coefficient estimates will not capture the full extent to which regressors covary with innovation.⁵ Because multicollinearity is plausibly an artifact of human capital being endogenous to population and university location, we use a multi-stage estimation approach.

The first stage models human capital as a function of population and universities. The residual from the first stage provides a measure of human capital uncorrelated with both population and universities. The final stage models innovation as a function of population, university presence, and the regional measure of human capital uncorrelated with population or universities. Parameter estimates from the two stage model allow us to decompose the relationship between population and innovation into that which is correlated with human capital and that which is uncorrelated with human capital. Similarly, the relationship between universities and innovation is also decomposed into human capital and non-human capital channels. Estimation of a naïve model illustrates the downward bias in coefficients due to multicollinearity among regressors.

Our notional model assumes that population and universities facilitate regional innovation through a number of pathways, including by facilitating the formation of high levels of human capital. Population and universities also serve as bases for agglomeration of innovation inputs that are not embodied in human capital. These human capital and non-human capital channels can be identified formally. Assume:

$$H = \lambda + \lambda_N N + \sum_{u \in [P, M, B, A]} \lambda_u D_u + \epsilon_H \quad (1)$$

where H is a measure of regional human capital, N is the region population, D is a vector indicating the highest level of degree granting program in a region - PhD, Master's, Bachelor's, or Associates—indexed by u , and ϵ_H is a residual term. The λ parameters summarize the relationship between the levels of human capital and population and universities in a region. Similarly, let:

$$G = \kappa + \kappa_N N + \sum_{u \in [P, M, B, A]} \kappa_u D_u + \epsilon_G \quad (2)$$

where G is a region's level of non-human-capital related resources employed in the production of innovation. In Equation (2), N and D are defined as in Equation (1), and ϵ_G is, likewise, a residual term. The parameters summarize the relationship between regional levels of innovative non-human capital and population and universities. Most generally, if regional innovation is attributable to regional human capital, other innovative capital in a region, and other attributes of population and universities that facilitate innovation, we can write:

$$I = \beta + \beta_H H + \beta_G G + \beta_N N + \sum_{u \in [P, M, B, A]} \beta_u D_u + \epsilon_I \quad (3)$$

⁵In the presence of multicollinearity, overall model fit and model-level statistics will be unaffected. However, individual variable parameter estimates will be affected, reflecting only the portion of correlation with the dependent variable that is independent of other regressors.

where I is a region's level of innovative production, H , G , N , and D are defined in Equations (1) and (2), and ϵ_I is a residual term. The β parameters summarize the relationship between regional levels of innovation and factors represented by H , G , N , and D .⁶

$$I = \beta + \beta_H \lambda + \beta_G \kappa + (\beta_N + \beta_H \lambda_N + \beta_G \kappa_N)N \sum_{u \in [P, M, B, A]} (\beta_u + \beta_H \lambda_u + \beta_G \kappa_u) D_u + \beta_H \epsilon_H + \beta_G \epsilon_G + \epsilon_I \quad (4)$$

Equation (4) summarizes the relationship between regional innovation and only the exogenous factors. Population (N) can contribute to regional innovation through human capital agglomeration ($\beta_H \lambda_N$) and thick markets for non-human-capital inputs to innovation ($\beta_G \kappa_N$), as well as through other potential factors that are uncorrelated with human and non-human capital, (β_N). University presence (D_u for $u \in [P, M, B, A]$) can make analogous contributions to regional innovation ($\beta_H \lambda_u$, $\beta_G \kappa_u$, and β_u , for $u \in [P, M, B, A]$). Finally, regional innovation may be enhanced by levels of human capital ($\beta_H \epsilon_H$) and other specialized but non-human-capital-correlated inputs ($\beta_G \epsilon_G$) that are also uncorrelated with population or university presence.

Relabeling parameters in Equation (4) for simplicity, we obtain Equation (5):

$$I = \rho + \rho_N N + \sum_{u \in [P, M, B, A]} \rho_u D_u + \rho_H \epsilon_H + \rho_G \epsilon_G + \epsilon_I \quad (5)$$

the reduced-form relationship quantifying total correlation between population and university location and regional innovation.

Estimating Equations (1) and (5) allow us to identify many of the factors contributing to innovation illustrated in Equation (4). For example, the innovation contribution of human capital uncorrelated with population and university presence can be recovered directly; $\beta_H = \rho_H$. Consequently, the innovative contribution of human capital attributable to greater population can also be estimated; $\beta_H \rho_N = \rho_H \lambda_N$. In the absence of data on regional levels of innovative capital resources or expenditures, we can recover the component of population associated with non-human capital channels, as a residual; $\beta_N + \beta_G \kappa_N = \rho_N - \rho_H \lambda_N$.⁷ The innovation contribution of universities through human capital and non-human capital pathways can be estimated in a similar fashion; $\beta_H \lambda_u = \rho_H \lambda_u$ and $\beta_u + \beta_G \kappa_u = \rho_u - \rho_H \lambda_u$; where $u \in [P, M, B, A]$.

⁶Recall that because H and G are endogenous to N and D , actual parameter estimates from Equation (3) will only quantify that portion of the relationship between I and each regressor that is independent of collinearity among the regressors.

⁷Alternatively, we can obtain this value directly from Equation (3) estimated under multicollinearity. As expected, we obtain the same point estimate for this parameter when estimated under either multicollinearity or calculated from orthogonal residual regressors.

4. DATA

Studies of innovation employ various sources of data for measuring the intensity of innovative activity in an area. Common proxies for innovation include patents, licensing agreements, and employment in R&D-intensive industries and occupations. As we are interested in outputs of innovative activity, our analysis relies on commonly-used patent data from the National Bureau of Economic Research (NBER).⁸ The NBER patent dataset contains information on all patents issued by the U.S. Patent & Trademark Office between 1963 and 1999. The dataset includes information on the geographic location of the first-named investigator to whom the patent was granted. County location coordinates from the U.S. Geological Survey (USGS), matched to NBER information by location of the first inventor listed on each patent, were used to construct patent totals for 3,141 counties or county-equivalent entities for the period 1990-1999. Patent counts are an imperfect measure of innovative activity because they exclude productivity-enhancing activities that do not lead to patentable inventions. Nevertheless, we consider it a suitable proxy for a wider array of innovative activity assuming patentable innovations are correlated with other forms of innovation.⁹ For the purpose of model estimation, our innovation measure is the total number of patents awarded to local county resident inventors from 1990 through 1999, in units of patents per 1 million residents.

A number of related studies either worked with data aggregated at the state level or focused on innovative activity in metropolitan areas only. Either approach presents challenges for estimation. Higher levels of aggregation lead to less precise estimates, while data truncation (such as exclusion of non-MSA areas) can bias estimates (Ghinamo, 2012). A large number of areas in the U.S. are sparsely populated and not part of any Metropolitan or Micropolitan Statistical Area; excluding these places from analysis leaves out a large part of the picture of innovation in the U.S. Previous research also provides support for preferring lower levels of aggregation. Ghinamo notes that academic spillovers are “strictly local and geographically bounded, disseminating within small areas in particular”, which implies that “[l]ower geographical levels are consequently better suited to study the knowledge spillover effect” (Ghinamo, 2012, p. 614).

We work with a county-level dataset that allows us to gain the benefits of disaggregation, and include in the analysis the entire geographic area of the US, including sparsely populated rural areas not part of an MSA. County population data are obtained from the U.S. Census Bureau Gazetteer, 2000. All measures of county populations are 1990's averages, in millions. The Census Bureau also provides statistics on the fraction of adult-aged population holding a Bachelor's degree or higher, which we use as a measure of local human capital.

The USGS provides latitude and longitude coordinates of the geographic centroid of each county, which allows us to estimate the distance between all county pairs as measured from their centroids. In calculating the area population of a given county, we started with that county's population and added the population of all neighboring counties with a geographic centroid located within a specified radius. Four alternative area population estimates were considered, based on 25, 50, 75 and 100 mile radii from county centroids. We identified

⁸See Hall et al. (2001).

⁹For a comprehensive analysis of the use of patent data as a measure of innovation see Griliches (1998).

evidence of positive spatial effects reaching as far as 50 miles from the central county. From this result and the intuition that county residents interact within a more general commuting area, our analysis uses an area population variable (N) defined as the total population of each county plus the populations of those counties within 50 miles of the geographic centroid of the subject county. We acknowledge that this definition of area population will only approximate the actual population within 50 miles of each county centroid. We expect associated mismeasurement to introduce random error into the population variable, leading to attenuated and downward-biased parameter estimates.¹⁰

Information on U.S. universities comes from the National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS). The NCES provides a variety of information on educational institutions, including detailed geographic information and full-time-equivalent registrants in various levels of degree-granting programs, which we use to identify the maximum level of degree-granting program in each county. Analysis was limited to degree-granting institutions active between 1996 and 2001 that were eligible to participate in Title IV federal financial aid programs, and offered Associate's, Bachelor's, Master's, and Doctoral degrees.¹¹ We include in our analysis all major educational institutions in the U.S. documented by the NCES.

The university vector measures the highest level of degree-granting institution in each county. For example, a county with at least one university offering Master's degree programs, no universities offering Doctoral degree programs, and any number of universities offering Bachelor's and Associate's degree programs would have a university indicator vector $D = [0, 1, 0, 0]$. Having more universities at the same level of degree offering does not alter a county's degree-offering category.

This relatively simple measure of university inputs is most consistent with the non-rival nature of knowledge presumably produced at universities. While the NCES data makes possible construction of various measures of university presence in a region, such as the total number of universities in a county or different measures of university size, many of the advantages universities offer to a region are not necessarily dependent on university size or the total number of universities. One research university could mean more than a large number of smaller liberal arts colleges for local R&D. In addition, on many dimensions the importance of access to the first university of a given degree-granting level is greater than the marginal importance of each additional such institution. For example, a local university library is much better than no library at all. But access to a second library is likely to provide little additional benefit due to the low cost of scaling use of one facility and duplication of resources across institutions.¹²

¹⁰In considering the role of population agglomeration on county innovation and human capital, we also considered alternative geographic units of analysis in order to capture spatial spillover effects of population. We estimated regression specifications where the "area population" variable was based on county population alone and broader measures based on the sum of the local county population plus the population of bordering counties, in addition to those radii noted in the discussion. These results were not inconsistent with the analysis presented here, based upon defining "area population" as a commuting space, i.e. according to a 50 mile radius from geographic centroids of each county.

¹¹Institutions granting first-professional degrees were categorized as Master's-level degrees in this study.

¹²The one-dimensional scope of the university vector used here also yields an easy-to-interpret coefficient. In most cases, a location with a university of a particular degree granting level will also offer degrees of

Nevertheless, it is important to keep in mind exactly what our university presence variables mean. A university dummy variable captures the effect on the county stemming from the presence of institutions of higher learning, due to access to educational opportunities to the local population and provision of non-human-capital factors of import for innovation. Our analysis design biases against finding significance in human capital associated with universities if the benefits of access to educational opportunities accrue to an area larger than the county, while the benefit of non-human-capital (presumably physical) assets is expected to be more localized.

Due to data limitations, we cannot directly estimate the extent to which the role of aggregate factors in regional innovation is associated with physical capital channels (G). In identifying the channels through which universities and population might impact innovation, we focus on the human capital channel, estimating the non-human capital channels as a residual. This approach to estimating non-human-capital factor productivity is similar to a residual growth-theoretic estimate of total factor productivity (Solow, 1956, 1957).

Table 2 presents summary statistics for data used in the analysis: area population (total population within a 50 mile radius from county centroids, as described above), values of the degree-granting institution vector components, human capital, and innovation. Total panel statistics are presented at the top of the table. Statistics are also presented for total sample tertiles and for the largest-population tertile decomposed into areas of population greater than two million and those of population below approximately 2 million.

Total area population ranges from 18.276 million in West Chester County, New York to 830 residents in Yukutat County, Alaska. All population tertiles and sub-units of the top tertile include at least one county with each level of degree granting program as the maximum level of degree-granting program in a county. The mean value of the degree-granting program indicator is increasing with population by tertile for PhD, Master's, and Bachelor's. The mean value for the Associate's degree program indicator is decreasing (though not statistically significantly) with population by tertile. Innovation is also increasing with population by tertile, as measured by maximum and mean number of patents per county from 1990 to 1999. All population tertiles include at least some counties with no patent activity in the decade.

5. RESULTS

Table 3 presents estimates for the two-stage model (Equations 1 and 5) and the naïve model (Equation 3) for purposes of comparison. In the first stage (column A), human capital is regressed against population and the university vector. In the second stage (column B), innovation is regressed on population, the university vector, and the residual from the first stage regression. In the naïve model (column C), parameter estimates are affected by

all lower levels. Using a uni-dimensional vector implies parameter estimates be interpreted as total effects of each level of degree-granting institution. As a result, the four parameter estimates from the university vector are directly comparable. In contrast, parameter estimates from a multi-dimensional vector would be interpretable as marginal effects of each degree-granting institution level, with the total effect of a particular degree-granting institution level obtained by adding together its coefficient and those of all lower degree-granting levels.

Table 2: Data Summary Statistics

| | N | D_P | D_M | D_B | D_A | H | I |
|---|-----------------------------------|--|---|---|--|--|--|
| | (area population, millions) | (PhD is max degree in county, flag) | (Master is max degree in county, flag) | (Bachelor is max degree in county, flag) | (Associate is max degree in county, flag) | (human capital, fraction adult population w/ bachelor degree or more) | (innovation, total patents per million county residents, 1990 - 1999) |
| Total Panel (n = 3,141) | | | | | | | |
| max | 18.276 | 1.00 | 1.00 | 1.00 | 1.00 | 0.637 | 43,947 |
| min | 0.001 | 0.00 | 0.00 | 0.00 | 0.00 | 0.049 | 0 |
| mode | 18.045 | 0.00 | 0.00 | 0.00 | 0.00 | 0.120 | 0 |
| median | 0.502 | 0.00 | 0.00 | 0.00 | 0.00 | 0.145 | 534 |
| mean | 0.985 | 0.10 | 0.12 | 0.06 | 0.14 | 0.165 | 1,044 |
| std dev | 1.774 | 0.30 | 0.32 | 0.24 | 0.35 | 0.078 | 2,016 |
| Top Tercile (n = 1,047) | | | | | | | |
| max | 18.276 | 1.00 | 1.00 | 1.00 | 1.00 | 0.637 | 43,947 |
| min | 0.781 | 0.00 | 0.00 | 0.00 | 0.00 | 0.056 | 0 |
| mode | 18.045 | 0.00 | 0.00 | 0.00 | 0.00 | 0.120 | 0 |
| median | 1.438 | 0.00 | 0.00 | 0.00 | 0.00 | 0.164 | 861 |
| mean | 2.324 | 0.21 | 0.17 | 0.09 | 0.13 | 0.191 | 1,754 |
| std dev | 2.580 | 0.40 | 0.38 | 0.28 | 0.34 | 0.095 | 2,885 |
| Top Tercile, Upper Segment (n = 345) | | | | | | | |
| max | 18.276 | 1.00 | 1.00 | 1.00 | 1.00 | 0.637 | 31,543 |
| min | 2.005 | 0.00 | 0.00 | 0.00 | 0.00 | 0.076 | 0 |
| mode | 18.045 | 0.00 | 0.00 | 0.00 | 0.00 | n/a | n/a |
| median | 3.128 | 0.00 | 0.00 | 0.00 | 0.00 | 0.227 | 1,702 |
| mean | 4.542 | 0.35 | 0.21 | 0.08 | 0.10 | 0.242 | 2,709 |
| std dev | 3.559 | 0.48 | 0.41 | 0.26 | 0.30 | 0.107 | 3,326 |
| Top Tercile, Lower Segment (n = 703) | | | | | | | |
| max | 2.005 | 1.00 | 1.00 | 1.00 | 1.00 | 0.515 | 43,947 |
| min | 0.781 | 0.00 | 0.00 | 0.00 | 0.00 | 0.056 | 0 |
| mode | 1.930 | 0.00 | 0.00 | 0.00 | 0.00 | 0.120 | 0 |
| median | 1.190 | 0.00 | 0.00 | 0.00 | 0.00 | 0.144 | 712 |
| mean | 1.235 | 0.13 | 0.15 | 0.10 | 0.14 | 0.166 | 1,286 |
| std dev | 0.328 | 0.34 | 0.36 | 0.29 | 0.35 | 0.077 | 2,512 |
| Middle Tercile (n = 1,047) | | | | | | | |
| max | 0.781 | 1.00 | 1.00 | 1.00 | 1.00 | 0.477 | 29,602 |
| min | 0.289 | 0.00 | 0.00 | 0.00 | 0.00 | 0.049 | 0 |
| mode | 0.736 | 0.00 | 0.00 | 0.00 | 0.00 | 0.188 | 0 |
| median | 0.502 | 0.00 | 0.00 | 0.00 | 0.00 | 0.122 | 406 |
| mean | 0.509 | 0.07 | 0.11 | 0.06 | 0.14 | 0.144 | 776 |
| std dev | 0.139 | 0.26 | 0.31 | 0.23 | 0.35 | 0.066 | 1,504 |
| Bottom Tercile (n = 1,047) | | | | | | | |
| max | 0.289 | 1.00 | 1.00 | 1.00 | 1.00 | 0.605 | 19,627 |
| min | 0.001 | 0.00 | 0.00 | 0.00 | 0.00 | 0.049 | 0 |
| mode | 0.210 | 0.00 | 0.00 | 0.00 | 0.00 | 0.222 | 0 |
| median | 0.109 | 0.00 | 0.00 | 0.00 | 0.00 | 0.150 | 412 |
| mean | 0.123 | 0.02 | 0.07 | 0.04 | 0.15 | 0.161 | 603 |
| std dev | 0.087 | 0.15 | 0.25 | 0.20 | 0.36 | 0.060 | 924 |

collinearity between human capital and the two other regressors.

The first stage regression shows that county fraction of adult age population holding a Bachelor's degree or higher, our proxy for county human capital, is positively and significantly correlated with area population as well as with proximity to universities. An area with a population of 1 million more than the panel mean will have human capital approximately 4.2 percent higher than panel mean human capital.

Counties that host at least one university with maximum degree granting programs at the associates level have 1.1 percentage points higher human capital than counties without

Table 3: Complete Data Estimates

| | (A) | (B) | (C) |
|--------------------|-----------------------|-------------------------|-------------------------|
| <i>Dep. Var.:</i> | Equation (1) | Equation (5) | Equation (3) |
| <i>Indep. Var.</i> | Human Capital (H) | Innovation (I) | Innovation (I) |
| Population (N) | 0.007*** (0.00) | 267.6*** (34.34) | 193.0*** (32.52) |
| D_P | 0.132*** (0.01) | 1172.8*** (151.08) | -189.1 (226.26) |
| D_M | 0.060*** (0.00) | 332.6*** (110.01) | -282.4** (136.14) |
| D_B | 0.031*** (0.00) | 97.0 (98.62) | -218.1** (109.28) |
| D_A | 0.011*** (0.00) | 72.6 (79.81) | -38.4 (77.53) |
| H | | | 10331.3*** (1147.51) |
| ϵ_H | | 10331.3*** (1147.51) | |
| Constant | 0.135*** (0.00) | 608.3*** (42.60) | -782.2*** (154.68) |
| R^2 | 0.367 | 0.217 | 0.217 |
| df | 3135 | 3134 | 3134 |
| F | 214.5 | 48.1 | 48.1 |

Notes: The human capital variable (H) is in the [0,1] range and represents the fraction of county population with at least a bachelors-level degree. Population (N) represents the total number of residents in the county and its surrounding areas within 50 miles of the county centroid, in units of millions. Innovation (I) is the total number of patents issued to county residents, per million residents. The D variables are dummies indicating the highest level of degree offering at universities located within the county, where the subscript $u \in [P, M, B, A]$ represents PhD, Masters, Bachelors or Associates degrees, respectively. The ϵ_H variable is a residual term from Equation (1) representing human capital not explained by population or university presence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

universities. Universities of higher degree-granting levels are associated with an even greater effect on local human capital. Counties host to Bachelor's, Master's, and PhD degree granting programs have, respectively, 3.1, 6.0 and 13.2 percentage points higher human capital than counties without universities.

As compared to the mean population county area without any universities, the incremental human capital associated with hosting no higher than Associate's degree granting programs is equivalent to the incremental human capital associated with adding 1.57 million residents to a county area, a population increase of 159 percent of the mean total area population. The incremental human capital associated with hosting no more than Bachelor's,

Master's, or PhD degree granting programs is equivalent to the incremental human capital associated with an additional 4.43, 8.57, or 18.9 million residents, respectively, as compared to the mean county area population.

The second stage parameter estimates in Table 3 (column B) are obtained from a regression of innovation on population, the university vector, and the human capital residual (ϵ_H) from the first-stage regression—a measure of county human capital not associated with universities or area population. Inclusion of the human capital residual allows us to estimate the elasticity of innovation to human capital, while maintaining strict orthogonality between human capital and other variables in the innovation regression.

Population agglomeration is positively and significantly associated with innovation, as is being host to PhD and Master's degree granting programs. In contrast, counties home to universities for which Bachelor's and Associate's are the highest level of degree granting programs are not associated with a higher level of innovative activity than counties without universities.

The role of universities in local innovation is substantial. On average, counties host to universities with at least one doctoral degree granting program were associated with 1,172.8 more patents per million residents (ppm) over the decade of the 1990s, as compared to counties with no universities. Counties host to universities without PhD degree programs but at least one Master's degree program are associated with 332.6 more ppm than counties with no universities.

In comparison, population agglomeration in the county area has a more subtle association to county innovation. An increase in area population by 1 million is associated with 267.6 extra patents per million county residents, about 80 percent of the marginal innovation associated with hosting a masters degree granting program, and 23 percent of the marginal innovation associated with hosting a PhD degree granting program. Alternatively, the marginal innovation associated with hosting the maximum of a Master's or PhD degree granting program is equivalent to an additional population of approximately 1.2 million or 4.4 million in a county area, respectively.

Human capital, as measured by fraction of population with at least a Bachelor's degree, is also significantly correlated to innovation. On average, a county with 1 percent more human capital is associated with 103.3 more patents per million residents per decade. This relationship implies a marginal innovation equivalence between a Master's degree granting program and a 3.2 percent increase in human capital, or a PhD degree granting program and an 11 percent increase in human capital.

The final regression in Table 3 illustrates how including a human capital measure into a naïve innovation estimation can mask the importance of population and universities. Due to collinearity with overall measures of human capital, correlations to innovation may be downward biased (as in the case of population) or even reverse sign (as in the case of components of the university vector.) This evidence of collinearity between universities and human capital suggests that the correlation between universities and innovation is due to the human capital channel.

Comparing coefficients and standard errors for human capital obtained from Equation 5 (ϵ_H) and Equation 3 (H) provides further support for the 2-stage estimation approach. The

naïve estimation of correlation between innovation and population, universities and human capital (Equation 3) results in less precise point estimates for population and universities (except in the case of the bachelors degree program indicator). In addition, coefficient estimates are downward-biased in every case because those variables are not independent of human capital. More specifically, coefficient estimates from the naïve model represent only that part of the correlation between innovation and the independent variable that is independent of human capital. For example, the population coefficient estimate from Equation 3 would exclude any correlation between innovation and population associated with human capital agglomeration.

Utilizing a human capital variable that is uncorrelated with the other regressors (ϵ_H in Equation 5) results in a greater precision of parameter estimates. Coefficients on the human capital residual in Equation 5 (ϵ_H) and human capital in Equation 3 (H) are identical because both are measures of the correlation between human capital and innovation uncorrelated with other included regressors.¹³ However, population and university coefficient estimates from the two-stage model represent the total correlation between innovation and independent variables.

For example, the population coefficient estimate from Equation 5 would include both correlation between innovation and population associated with human capital agglomeration and that associated with the agglomeration of factors uncorrelated with human capital. Consequently, and as discussed in the previous section, we can decompose independent variable coefficients from Equation (5) into human capital and residual components. This allows us to distinguish the correlation between innovation and the independent variables (universities and population) into the covariation associated with human capital and that associated with unobserved factors uncorrelated with human capital.¹⁴ This decomposition of dependent variable correlation to innovation is presented in Table 4. Column (D) reproduces the regression coefficients from equation (5) (Table 3, column (B)). Column (E) shows the magnitude of the independent variable coefficient attributable to the relationship between the independent variable and human capital. Likewise, column (F) presents the non-human capital portion of the independent variables' overall correlation with innovation. Coefficients in columns (E) and (F) add to the coefficient in column (D) because values in column (F) represent residual measures.¹⁵

¹³Identical regression level statistics R^2 and F between Equation 5 and Equation 3 provide an additional check that the constructed human capital residual captures all of the independent variation quantified in the naïve model while avoiding downward-biased parameter estimated due to collinearity between H and the other regressors.

¹⁴Recall from the concluding discussion in Section 3 that some of the parameters represented in Equation (4) can be recovered from parameters estimated using Equations (5) and (1). For example, the increase in innovation associated with human capital agglomeration attributable to an increase in population is $\beta_H \lambda_N = \rho_H \lambda_N$. The remainder of the total relationship between innovation and population is interpreted as the increase in innovation associated with the extent to which greater population also enhances innovation through non-human-capital channels $\beta_N + \beta_G \kappa_N = \rho_N - \rho_H \lambda_N$. Parameters and their standard errors were recovered using rules of addition and multiplication of random variables. For example, see Goodman (1960).

¹⁵Also of note is that our naïve model coefficient estimates (Table 3, Column C) are identical to our non-human-capital channel measures computed manually from our two-stage regression estimates (Table 4, Column C). This is expected because Equation (3) includes a human capital variable that is not orthogonal

Table 4: Decomposition of Dependent Variable Correlation to Innovation

| <i>Indep. Var.</i> | (D) | (E) | (F) |
|-------------------------|-----------------------|---------------------------|--------------------------------|
| | Total Impact | Via Human Capital Channel | Via Non-Human Capital Channels |
| Population (<i>N</i>) | 267.6*** (34.34) | 74.6*** (14.69) | 193.0*** (37.35) |
| <i>D_P</i> | 1172.8*** (151.08) | 1361.9*** (161.84) | -189.1 (221.4) |
| <i>D_M</i> | 332.6*** (110.01) | 615.1*** (78.87) | -282.4** (135.37) |
| <i>D_B</i> | 97.0 (98.62) | 315.1*** (58.16) | -218.1* (114.5) |
| <i>D_A</i> | 72.6 (79.81) | 111.1*** (32.66) | -38.4 (86.24) |

Notes: Column (D) represents the total correlation between population (*N*) and level of university degree offering (*D_u*: $u \in [P, M, B, A]$) on innovation. Columns (E) and (F) decompose the overall impact into human capital and non-human capital components. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

The correlation between population and innovative activity is attributable to both human capital and non-human-capital linkages, although the association with non-human-capital predominates. The magnitude of the non-human-capital association with innovation (193.0*** ppm) is more than twice as great as that of the human capital association (74.6*** ppm). More populous places appear to facilitate innovation through thicker markets for non-labor inputs to innovation and, to a lesser degree, human capital agglomeration.

A positive relationship between innovation and degree granting programs is identified at the PhD and Master's degree levels. In contrast to the correlation between innovation and population, the positive relationship between these types of universities and local innovation is associated exclusively with greater levels of human capital in university locations. No portion of the correlation between innovation and university presence can be attributed to factors uncorrelated with human capital.

Proximity to universities offering a maximum of Bachelor's and Associate's degree granting programs has a statistically insignificant relation to innovation. However, the human capital in such locations is disproportionately high, perhaps conferring benefits unrelated to innovation.

For counties where Master's and Bachelor's are the highest level of university degree of-

to the other regressors. As a result, regression coefficients will reflect only independent variable variation that is uncorrelated with human capital. Our manually-calculated decomposition by channel serves as a check on the integrity of regression estimates and validates the concerns motivating this study.

Table 5: Estimates by Area Population Tercile

| Range: | Top Tercile 780,900 through 18,276,400 | | | Middle Tercile 289,100 through 780,600 | | | Bottom Tercile 800 through 288,900 | | |
|-----------------------|---|-------------------------|-------------------------|---|------------------------|------------------------|---------------------------------------|------------------------|------------------------|
| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) |
| <i>Dep. Var.:</i> | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) |
| <i>Indep. Var.</i> | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> |
| <i>N</i> | 0.009*** (0.00) | 211.3*** (35.45) | 75.5** (33.35) | -0.001 (0.01) | -245.4 (376.74) | -240.8 (376.70) | -0.150*** (0.02) | 583.2* (329.56) | 1571.6** (647.25) |
| <i>D_P</i> | 0.125*** (0.01) | 1037.6*** (232.86) | -825.1** (369.45) | 0.150*** (0.01) | 1436.4*** (265.92) | 190.3 (287.05) | 0.144*** (0.02) | 648.4*** (221.62) | -298.4 (421.85) |
| <i>D_M</i> | 0.052*** (0.01) | 416.4* (240.65) | -361.9 (276.05) | 0.078*** (0.01) | 193.5** (85.36) | -450.1*** (123.29) | 0.064*** (0.01) | 11.6 (59.03) | -410.0** (180.24) |
| <i>D_B</i> | 0.030*** (0.01) | -58.8 (222.43) | -498.7** (242.01) | 0.030*** (0.01) | 124.2 (109.12) | -123.5 (114.90) | 0.049*** (0.01) | 35.3 (84.38) | -287.4* (156.00) |
| <i>D_A</i> | 0.010 (0.01) | 86.5 (223.20) | -67.7 (223.13) | 0.015*** (0.00) | -94.3 (80.29) | -216.8*** (83.28) | 0.011* (0.00) | 159.0 (106.13) | 84.7 (82.80) |
| <i>H</i> | | | 14845.3*** (1919.96) | | | 8297.9*** (1026.01) | | | 6590.4*** (2509.00) |
| <i>ε_H</i> | | 14845.3*** (1919.96) | | | 8297.9*** (1026.01) | | | 6590.4*** (2509.00) | |
| Constant | 0.131*** (0.00) | 971.2*** (151.22) | -974.1*** (261.34) | 0.121*** (0.01) | 781.9*** (233.95) | -220.3 (249.50) | 0.168*** (0.00) | 491.2*** (42.61) | -616.8 (447.10) |
| <i>R</i> ² | 0.399 | 0.219 | 0.219 | 0.418 | 0.140 | 0.140 | 0.220 | 0.162 | 0.162 |
| df | 1041 | 1040 | 1040 | 1041 | 1040 | 1040 | 1041 | 1040 | 1040 |
| F | 94.0 | 25.6 | 25.6 | 83.5 | 16.2 | 16.2 | 44.7 | 3.3 | 3.3 |

Notes: The human capital variable (*H*) is in the [0,1] range and represents the fraction of county population with at least a Bachelor's-level degree. Population (*N*) represents the total number of residents in the county and its surrounding areas within 50 miles of the county centroid, in units of millions. Innovation (*I*) is the total number of patents issued to county residents, per million residents. The *D* variables are dummies indicating the highest level of degree offering at universities located within the county, where the subscript $u \in [P, M, B, A]$ represents PhD, Master's, Bachelor's, or Associate's degrees, respectively. The ϵ_H variable is a residual term from Equation (1) representing human capital not explained by population or university presence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

ferred, the coefficient representing the relationship between universities and local innovation uncorrelated with human capital is negative and significant. In the cases of PhD and Associate's degree granting programs, these correlations are also negative, though not statistically significant. This result suggests that counties with universities of any degree granting level have a rate of innovation that is disproportionately low in relation to the level of human capital typically associated with local degree granting programs.

One interpretation of these findings is that degree-granting institutions are associated with an impact on human capital that is disproportionate to their impact on innovation. As a result, the non-human-capital channel appears as a drag on the level of innovation that should otherwise be observed with high levels of human capital in university locations. We interpret negative non-human-capital factor values as an indication that high levels of human capital associated with university locations may be less-intensively directed to patentable innovations, as compared to similar levels of human capital when observed in non-university locations. This could reflect the fact that university-related human capital is disproportionately directed toward vocational education and/or education and research from which knowledge development and discovery is less likely to be codified in the form of a patent.

Next, we consider whether the correlation between innovation and the independent variables is conditional on population. Table 5 reports model parameter estimates by county

Table 6: Decomposition of Dependent Variable Pathway to Innovation

| Range: | Top Tercile | | | Middle Tercile | | | Bottom Tercile | | |
|----------------------|----------------------------|-----------------------|----------------------|-------------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|
| | 780,900 through 18,276,400 | | | 289,100 through 780,600 | | | 800 through 288,900 | | |
| | (J) | (K) | (L) | (M) | (N) | (O) | (P) | (Q) | (R) |
| | | Via | Via Non- | | Via | Via Non- | | Via | Via Non- |
| | Total | Human | Human | Total | Human | Human | Total | Human | Human |
| | Impact | Capital | Capital | Impact | Capital | Capital | Impact | Capital | Capital |
| <i>Indep. Var.</i> | | Channel | Channel | | Channel | Channel | | Channel | Channel |
| <i>N</i> | 211.3*** (35.45) | 135.8*** (27.94) | 75.5* (45.14) | -245.4 (376.74) | -4.6 (97.79) | -240.8 (389.23) | 583.2* (329.56) | -988.4** (397.69) | 1571.6*** (516.5) |
| <i>D_P</i> | 1037.6*** (232.86) | 1862.8*** (266.73) | -825.1** (354.08) | 1436.4*** (265.92) | 1246.1*** (172.14) | 190.3 (316.77) | 648.4*** (221.62) | 946.8** (378.89) | -298.4 (438.94) |
| <i>D_M</i> | 416.4* (240.65) | 778.3*** (139.72) | -361.9 (278.27) | 193.5** (85.36) | 643.6*** (94.2) | -450.1*** (127.13) | 11.6 (59.03) | 421.7** (166.97) | -410.0** (177.1) |
| <i>D_B</i> | -58.8 (222.43) | 439.8*** (122.45) | -498.7** (253.91) | 124.2 (109.12) | 247.7*** (57.96) | -123.5 (123.56) | 35.3 (84.38) | 322.7** (138.12) | -287.4* (161.85) |
| <i>D_A</i> | 86.5 (223.20) | 154.2 (98.41) | -67.7 (243.93) | -94.3 (80.29) | 122.5*** (34.36) | -216.8** (87.33) | 159.0 (106.13) | 74.3* (42.45) | 84.7 (114.3) |

Notes: Columns (J, M, P) represent the total correlation between population (*N*) and level of university degree offering (*D_u*: $u \in [P, M, B, A]$) on innovation. Columns (K, N, Q) and (L, O, R) decompose the overall impact into human capital and non-human-capital components, respectively. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

area population tercile.¹⁶ Table 6 decomposes those parameter estimates into covariation associated with human capital and that not associated with correlates to human capital. The relationships between innovation and the independent variables vary by population tercile.

Table 5 shows population is positively associated with innovation in the most and the least populated terciles of counties (Table 5, row *N*, columns B and H), while it appears to have no statistically significant correlation with innovation in middle tercile counties (Table 5, row *N*, column E).

In the top tercile counties, population is correlated with innovation primarily through a correlation between population and human capital (Table 5, row *N*, column K vs L), in contrast to the total panel estimates for which the innovation-population relationship is predominantly driven by non-human-capital correlates to innovation (Table 4, row *N*, column E vs F).

Among the least populated county areas, human capital is negatively correlated with population (Table 5, row *N*, column G). Hence, the positive correlation between innovation and population (Table 5, row *N*, column H) is attributable to factors correlated with population but uncorrelated with human capital (Table 6, row *N*, column R). Being host to PhD degree granting programs is positively correlated to innovation in all three population terciles (Table 5, row *D_P*, columns B, E, and H). Interestingly, the highest coefficient estimate is obtained from the middle population tercile areas. In all population terciles, the coefficient decomposition (Table 6 row *D_P*) indicates that the correlation between innovation and PhD degree program is entirely attributable to degree program correlates to human capital.

Master's degree granting programs are significantly correlated with innovation only across areas in the highest and middle population terciles (Table 5, row *D_M*, columns B vs E).

¹⁶Recall, that county area population is defined as the population sum of the subject county and all other counties within a 50 mile radius of the county centroid.

Table 7: Top Tercile Estimates

| | Top Tercile | | | Upper Segment | | | Lower Segment | | |
|----------------------|----------------------------|-------------------------|-------------------------|------------------------------|-------------------------|-------------------------|---------------------------|-------------------------|-------------------------|
| Range: | 780,900 through 18,276,400 | | | 2,000,000 through 18,276,400 | | | 780,800 through 2,000,000 | | |
| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) |
| Dep. Var.: | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) |
| Indep.Var. | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> |
| <i>N</i> | 0.009*** (0.00) | 211.3*** (35.45) | 75.6** (33.35) | 0.007*** (0.00) | 136.0*** (39.39) | 10.8 (41.74) | 0.029*** (0.01) | -9.3 (268.06) | -300.6 (281.06) |
| <i>D_P</i> | 0.125*** (0.01) | 1037.6*** (232.86) | -825.1** (369.45) | 0.092*** (0.01) | 693.1 (474.89) | -1015.2 (629.72) | 0.141*** (0.01) | 1048.1*** (268.57) | -371.8 (331.00) |
| <i>D_M</i> | 0.052*** (0.01) | 416.4* (240.65) | -361.9 (276.05) | 0.028* (0.01) | -231.9 (471.44) | -749.4 (503.01) | 0.056*** (0.01) | 656.5** (288.32) | 94.0 (313.88) |
| <i>D_B</i> | 0.030*** (0.01) | -58.8 (222.43) | -498.7** (242.01) | -0.018 (0.02) | -1172.7*** (429.04) | -841.5** (405.58) | 0.044*** (0.01) | 280.3 (263.90) | -166.7 (285.02) |
| <i>D_A</i> | 0.010 (0.01) | 86.5 (223.20) | -67.7 (223.13) | 0.017 (0.02) | 769.2 (693.33) | 460.6 (693.64) | 0.006 (0.01) | -151.1 (174.09) | -213.4 (177.07) |
| <i>H</i> | | | 14845.3*** (1919.96) | | | 18471.5*** (3042.45) | | | 10053.8*** (1469.31) |
| <i>c_H</i> | | 14845.3*** (1919.96) | | | 18471.5*** (3042.45) | | | 10053.8*** (1469.31) | |
| Constant | 0.131*** (0.00) | 971.2*** (151.22) | -974.1*** (261.34) | 0.173*** (0.01) | 1907.9*** (361.64) | -1282.4*** (454.74) | 0.098*** (0.01) | 1050.4** (451.41) | 68.6 (427.79) |
| <i>2</i> | 0.399 | 0.219 | 0.219 | 0.258 | 0.325 | 0.325 | 0.404 | 0.083 | 0.083 |
| df | 1041 | 1040 | 1040 | 339 | 338 | 338 | 696 | 695 | 695 |
| F | 94.0 | 25.6 | 25.6 | 23.7 | 15.4 | 15.4 | 66.5 | 11.9 | 11.9 |

Notes: The human capital variable (*H*) is in the [0,1] range and represents the fraction of county population with at least a bachelors level degree. Population (*N*) represents the total number of residents in the county and its surrounding areas within 50 miles of the county centroid, in units of millions. Innovation (*I*) is measured in total number of patents issued to county residents over the total number of county residents, in units of patents per million residents. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

As was the case at the PhD level, coefficients reflect correlation between Master's degree programs and correlates to human capital (Table 6, row *D_M*). In the lowest population tercile, the overall rate of innovation is uncorrelated with being host to a Master's degree program, even though human capital is positively correlated. As was the case with the total panel, the overall rate of innovation in each county population tercile is uncorrelated with the bachelors and associates degree granting program indicator variables.

The top tercile of county areas arguably combines very different types of agglomerations. County areas in this population tercile range from 780,900 (Lincoln County, TN, adjacent to Madison County and including the city of Huntsville, AL) through more than 18 million (Westchester County, NY). To investigate how correlates to human capital and innovation vary across this tercile, we provide parameter estimates for upper and lower segments. The upper segment contains counties located in large metropolitan areas where the area population exceeds 2 million residents (e.g among the smallest being Fayette County, IN near Indianapolis, or Marion County, OH near Columbus). The lower segment contains top tercile counties in areas with a population below 2 million (e.g. among the largest being Shelby County, IN near Indianapolis, or Licking County, OH near Columbus). These results are presented in Tables 7 and 8.

Area population is positively correlated with human capital in the total and both subsets of the top tercile (Table 7, row *N*, columns A, D, G). However, innovation is correlated with population only in the total and the most populous segment of top tercile agglomerations (Table 7, row *N*, columns B, E).

Table 8: Top Tercile Decomposition of Dependent Variable Pathway to Innovation

| | Top Tercile | | | Upper Segment | | | Lower Segment | | |
|----------------------|----------------------------|-----------------------|----------------------|------------------------------|-----------------------|---------------------|---------------------------|-----------------------|--------------------|
| Range: | 780,900 through 18,276,400 | | | 2,000,000 through 18,276,400 | | | 780,800 through 2,000,000 | | |
| | (J) | (K) | (L) | (M) | (N) | (O) | (P) | (Q) | (R) |
| Dep. Var.: | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) | Eq. (1) | Eq. (5) | Eq. (3) |
| Indep.Var. | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> | <i>H</i> | <i>I</i> | <i>I</i> |
| <i>N</i> | 211.3*** (35.45) | 135.8*** (27.94) | 75.5* (45.14) | 136.0*** (39.39) | 125.2*** (37.87) | 10.8 (54.64) | -9.3 (268.06) | 291.3*** (89.68) | -300.6 (282.66) |
| <i>D_P</i> | 1037.6*** (232.86) | 1862.8*** (266.73) | -825.1** (354.08) | 693.1 (474.89) | 1708.3*** (396.23) | -1015.2 (618.49) | 1048.1*** (268.57) | 1419.9*** (226.32) | -371.8 (351.22) |
| <i>D_M</i> | 416.4* (240.65) | 778.3*** (139.72) | -361.9 (278.27) | -231.9 (471.44) | 517.5* (287.2) | -749.4 (552.03) | 656.5** (288.32) | 562.5*** (106.64) | 94 (307.41) |
| <i>D_B</i> | -58.8 (222.43) | 439.8*** (122.45) | -498.7** (253.91) | -1172.7*** (429.04) | -331.2 (289.67) | -841.5 (517.67) | 280.3 (263.90) | 446.9*** (106.95) | -166.7 (284.75) |
| <i>D_A</i> | 86.5 (223.20) | 154.2 (98.41) | -67.7 (243.93) | 769.2 (693.33) | 308.5 (349.39) | 460.6 (776.39) | -151.1 (174.09) | 62.3 (58.29) | -213.4 (183.59) |

Notes: Columns (J, M, P) represents the total correlation between population (*N*) and level of university degree offering (*D_u*: $u \in [P, M, B, A]$) on innovation. Remaining columns decompose the overall impact into human capital and non-human capital components. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

As in the overall sample and in the top tercile, being host to PhD and Master's degree granting programs is associated with higher levels of human capital (Table 7, rows *D_P*, *D_M*, columns A, D, G). Being host to Bachelor's degree granting programs is correlated with human capital in the overall sample, the top tercile, and the bottom segment of the top tercile (Table 7, row *D_B*, column D). However, it is uncorrelated with human capital in the upper segment of the top tercile. Being host to Associate's degree programs is uncorrelated with human capital in the top tercile and subsets thereof (Table 7, row *D_A*, columns A, D, G).

Innovation is positively correlated with hosting degree granting programs in only the lower segment of the largest population tercile (Table 7, rows *D_P*, *D_M*, column H). Innovation is not positively correlated with any level of degree granting program in the largest population areas (Table 7, column E). Indeed, we estimate a negative correlation between innovation and being host to a maximum of bachelors degree granting programs in the highest population segment of the data (Table 7, row *D_B*, column E).

6. CONCLUSION

Prior literature demonstrates that the innovative output of cities is positively correlated with population size. A positive relationship between universities and regional innovation has also been shown. This paper considers the role of both population and institutions of higher education as factors related to the degree of regional innovation. We employ an analytical structure that allow us to decompose correlates to innovation into a share associated with human capital and a share associated with factors uncorrelated with human capital. Finally, we consider whether the relationships between population, universities, human capital, and innovation vary with population.

Overall sample parameter estimates indicate that innovation is correlated with population and the location of PhD and Master's degree granting universities. This study shows that a

Table 9: Summary of Coefficients' Significance and Signs

| Sample | pop | Equation 1 | | | | | Equation 5 | | | | |
|----------------|--------------|------------|-------|-------|-------|-------|-----------------|-------------------|-------------------|-------------------|-------------------|
| | | N | D_P | D_M | D_B | D_A | N (hc,nhc) | D_P (hc,nhc) | D_M (hc,nhc) | D_B (hc,nhc) | D_A (hc,nhc) |
| Total Panel | 18M to 1k | + | + | + | + | + | +(+,+) | +(+,0) | +(+,-) | 0(+,-) | 0(+,0) |
| Top Tercile | 18M to 781k | + | + | + | + | 0 | +(+,+) | +(+,-) | +(+,0) | 0(+,-) | 0(0,0) |
| Top Terc. Upp | 18M to 2M | + | + | + | 0 | 0 | +(+,0) | 0(+,0) | 0(+,0) | -(0,0) | 0(0,0) |
| Top Terc. Low | 2M to 781k | + | + | + | + | 0 | 0(+,0) | +(+,0) | +(+,0) | 0(+,0) | 0(0,0) |
| Middle Tercile | 781k to 290k | 0 | + | + | + | + | 0(0,0) | +(+,0) | +(+,-) | 0(+,0) | 0(+,-) |
| Bottom Tercile | 290k to 1k | - | + | + | + | + | +(-,+) | +(+,0) | 0(+,-) | 0(+,-) | 0(+,0) |

Notes: Columns report direction of correlation of population (N) and level of university degree offering (D_u : $u \in [P, M, B, A]$) to human capital (equation 1) and innovation (equation 5). Symbols represent positive significant (+), negative significance (-) or insignificance (0) at a level of 10 percent.

naïve model of regional innovation—including a measure of total human capital—will impose a downward bias on population and university coefficient estimates. This bias is particularly misleading in the case of universities.

The severe bias against identifying a correlation between universities and innovation is due to the significant correlation between universities and human capital at all degree granting program levels. From the decomposition of our parameter estimates, we infer that universities are correlated with innovation entirely through the positive association between universities and regional human capital. In contrast, population is correlated with innovation through human capital and non-human-capital correlates to innovation.

Table 9 provides a qualitative summary of coefficient estimates from the panel when disaggregated by area population. The positive overall correlation between human capital and population is only observed within the most populous tercile of counties. Indeed, among the least populous county areas, we identify a negative correlation between human capital and population.

The positive overall correlation between human capital and PhD and Master's degree granting programs is identified at a minimum of a 10 percent level of significance in every subset of the data considered. The positive overall correlation between human capital and Bachelor's degree granting programs is identified across all county areas except those with population in excess of approximately 2 million. The positive overall correlation between human capital and Associate's degree granting programs is observed in population areas below approximately 780,000.

A positive correlation between innovation and population is observed in the overall panel and in sub-samples of counties with area populations above 2 million and below approximately 290,000. In the most populous areas, the correlation between innovation and population is associated with a correlation between human capital and population. In the least populous areas, this correlation between innovation and population is associated with non-human capital correlates to population. Innovation appears uncorrelated with population in county areas of population between 290,000 and 2 million.

A positive correlation between innovation and PhD degree granting programs is identified in the overall study sample and in all sub-samples of the counties with area population below 2 million. In all cases, the positive correlation between innovation and PhD degree

granting program location is associated with the higher level of human capital geographically coincident with PhD degree granting programs.

The positive overall correlation between innovation and Master's degree granting program is not identified in the most and least populous sub-samples of the data. For county areas with population between 290,000 and 2 million, the positive correlation between innovation and PhD degree granting program location is associated with the higher level of human capital geographically coincident with PhD degree granting programs.

Opportunities for future research include examining if the scale of degree offerings, in a subject county or neighboring counties, is correlated with innovative activity. In addition, explicit measurement and inclusion of non-human-capital factors into the regional knowledge production function is likely to provide greater insights into how population and university location are correlated with regional innovation.

REFERENCES

- Abel, Jaison R. and Richard Deitz. (2011) "Do Colleges and Universities Increase Their Regions Human Capital?," *Journal of Economic Geography*, 12(3), 667–691.
- Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman. (1992) "Real Effects of Academic Research: Comment," *The American Economic Review*, 82(1), 363–367.
- Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman. (1994) "R&D Spillovers and Innovative Activity," *Managerial and Decision Economics*, 15(2), 131–138.
- Anselin, Luc, Attila Varga, and Zoltan Acs. (1997) "Local Geographic Spillovers between University Research and High Technology Innovations," *Journal of Urban Economics*, 42(3), 422–448.
- Audretsch, David B. and Maryann P. Feldman. (1996) "RD Spillovers and the Geography of Innovation and Production," *The American Economic Review*, 86(3), 630–640.
- Carlino, Gerald, Satyajit Chatterjee, and Robert Hunt. (2015) "Matching and Learning in Cities: Urban Density and the Rate of Invention," *Federal Reserve Bank of Philadelphia Working Paper*, 04-16/R. Last Accessed September 2019.
- Cowan, Robin and Natalia Zinovyeva. (2013) "University Effects on Regional Innovation," *Research Policy*, 42(3), 788–80024.
- Ghinamo, Mauro. (2012) "Explaining the Variation in the Empirical Estimates of Academic Knowledge Spillovers," *Journal of Regional Science*, 52(4), 606–634.
- Glaeser, Edward L. (1998) "Are Cities Dying?," *Journal of Economic Perspectives*, 12(2), 139–160.
- Goldstein, Harvey and Joshua Drucker. (2006) "The Economic Development Impacts of Universities on Regions: Do Size and Distance Matter?," *Economic Development Quarterly*, 20(1), 22–43.
- Goodman, Leo A. (1960) "On the Exact Variance of Products," *Journal of the American Statistical Association*, 55(292), 708–713.
- Griliches, Zvi. (1998) "Patent Statistics as Economic Indicators: A Survey," *NBER Working Paper No.8351*.
- Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. (2001) "The NBER Patent Citation Data File: Lessons, Insights, and Methodological Tools," *NBER Working Paper No.8498*.

- Jaffe, Adam B.. (1989) "Real Effects of Academic Research," *The American Economic Review*, 79(5), 957–970.
- Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 103(3), 577–598.
- Kim, Jinyoung, Sangjoon John Lee, and Gerald Marschke. (2005) "The Influence of University Research on Industrial Innovation," *NBER Working Paper No.11447*.
- Mowery, David C. and Arvids A. Ziedonis. (2001) "The Geographic Reach of Market and Non-market Channels of Technology Transfer: Comparing Citations and Licenses of University Patents," *NBER Working Paper No.8568*.
- Ó hUallcháin, Breandán. (1999) "Patent Places: Size Matters," *Journal of Regional Science*, 39(4), 613–636.
- Orlando, Michael J.. (2004) "Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity," *RAND Journal of Economics*, 35(4), 777–786.
- Orlando, Michael J. and Michael Verba. (2005) "Do Only Big Cities Innovate? Technological Maturity and the Location of Innovation," *Federal Reserve Bank of Kansas City Economic Review*, 90(2), 31–57.
- Ponds, Roderik, Frank van Oort, and Koen Frenken. (2010) "Innovation, Spillovers and University-industry Collaboration: An Extended Knowledge Production Function Approach," *Journal of Economic Geography*, 10(2), 231–255.
- Solow, Robert M.. (1956) "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, 70(1), 111–137.
- Solow, Robert M.. (1957) "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39(3), 312–320.
- Storper, Michael and Anthony J. Venables. (2004) "Buzz: Face-to-face Contact and the Urban Economy," *Journal of Economic Geography*, 4(4), 351–370.
- Varga, Attila. (2000) "Local Academic Knowledge Transfers and the Concentration of Economic Activity," *Journal of Regional Science*, 40(2), 289–308.