May 2014

AGGLOMERATION ECONOMIES, INVESTMENT IN EDUCATION, AND REGIONAL DEVELOPMENT

Shimeng Liu
Syracuse University

Abstract

This dissertation consists of two essays that study the linkages among agglomeration economies, investment in education, and regional development. In the first essay, I study the impact of a federal educational investment on various aspects of local economies. In the second essay, I examine the spillover effects among workers with different skills, which are identified by their college majors.

The first essay presents evidence of direct spillovers from universities and examines the short- and long-run effects of university activities on geographic clustering of economic activity, labor market composition and local productivity. I treat the designation of land-grant universities as a natural experiment after controlling for the confounding factors with a combination of synthetic control methods and event-study analyses. Three key results are obtained. First, the designation substantially increased local population density. Second, the share of manufacturing workers in the population, an indicator of labor market composition, was not affected by the designation. Third, the designation greatly enhanced local manufacturing productivity, as measured by local manufacturing output per worker, especially in the long run. This positive effect on the productivity in non-education sectors suggests the existence of spillovers from universities. Over an 80-year horizon, I estimate that most of the increase in manufacturing productivity was because of direct spillovers from universities instead of induced agglomeration economies that arise from the increase in population.

The second essay studies the manner and extent to which worker skill type affects agglomeration economies that contribute to productivity in cities. I use college major to proxy for skill type among workers with a Bachelor's degree. Workers with college training in information-oriented and technical fields (e.g. STEM areas such as Engineering, Physical

Sciences, and Economics) are associated with economically important within-field agglomeration economies and also generate sizeable spillovers for workers in other fields. In contrast to related work by Florida (2002a, 2002b), within-field and across-field spillovers for workers with college training in the arts and humanities are much smaller and often non-existent. While previous research suggests proximity to college-educated workers enhances productivity, these findings suggest that not all college educated workers are alike. Instead, positive spillover effects appear to derive mostly from proximity to workers with training in information-oriented and technical fields.

AGGLOMERATION ECONOMIES, INVESTMENT IN EDUCATION, AND REGIONAL DEVELOPMENT

By

Shimeng Liu
B.S. Huazhong University of Science & Technology, 2009
M.A. Syracuse University, 2011

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate School of Syracuse University

May 2014

Acknowledgements

I would like to express my special gratefulness for advice from Stuart Rosenthal who gave direction to these papers. I would like to thank Robert Bifulco, William Horrace, Chihwa Kao, Eleonora Patacchini, John Yinger for their suggestions and participation in my oral examination. I also want to thank Jeffrey Kubik, Jeffrey Weinstein, Jing Li, Qianqian Cao, and seminar participants at Syracuse University for their comments and suggestions. I am also grateful for comments and advice received from participants at the American Real Estate and Urban Economics Association (AREUEA) poster session at the 2014 ASSA. Last but not least, I am sincerely thankful for the love and support from my parents, my mother Hongyan Zhou and my father Yong Liu while I was writing this dissertation. All remaining errors are my own.

Table of Contents

List of Tables

List of Figures

Chapter 1 Spillovers from Universities: Evidence from the Land-Grant Program

1.1 Introduction

Universities are widely believed to boost growth and productivity. It is conventional wisdom that "Silicon Valley" near San Jose and Route 128 around Boston owe their status as economic centers to their proximity to Stanford and MIT (Jaffe, 1989). To date, a large literature has sought to provide evidence of the linkage between academic investment, potential spillovers and economic agglomerations.[1] However, much of the literature focuses on the spillover effects from colleges and universities on patents, innovations and business start-ups.[2] Also, the feedback effects from business activity and the common factors that affect both universities and business environment make the causal impact of colleges and universities difficult to measure. The recent literature is paying more attention to the identification of causal effects. Andersson, Quigley and Wilhelmsson (2004, 2009) employ the decentralization policy of higher education in Sweden to investigate the economic impact of educational investment on productivity and innovation. Using an instrumental variables technique, Kantor and Whalley (2012) study the local spillovers from research universities.

Using a new identification strategy, this paper presents evidence of direct spillovers from universities and examines the short- and long-run effects of university activities on geographic clustering of economic activity, labor market composition and local productivity.[3] The identification strategy is that I treat the designation of land-grant universities in the United States

---

[1] See Moretti (2004) for a review of the literature on local social return of education.

[2] See, for example, Jaffe (1989), Acs, Audretsch, and Feldman (1992), Bania, Eberts, and Fogerty (1993), Beeson and Montgomery (1993), Audretsch and Feldman (1996), Anselin, Varga, and Acs (1997), Varga (2000), Adams (2002), Cohen, Nelson, and Walsh (2002), Woodward, Figueiredo, and Guimarães (2006), Abramovsky, Harrison, and Simpson (2007), Andersson, Quigley, and Wilhelmsson (2004, 2009), Aghion, Boustan, Hoxby and Vandenbussche (2009) and Hausman (2011).

[3] Universities can generate spillovers to communities through two possible mechanisms, direct interaction between faculty and local business establishments and training of students who remain in the area and enhance the quality of the labor pool. In this paper, I do not distinguish between these two mechanisms, although I present evidence suggesting the latter mechanism is less likely to be driving my results.

in the 1860s as a natural experiment after controlling for the confounding factors with synthetic control methods (Abadie and Gardeazabal, 2003; Abadie, Diamond and Hainmueller, 2010, 2012; Billmeier and Nannicini, 2013) and event-study analyses (Jacobson, LaLonde and Sullivan, 1993; McCrary, 2007; Kline, 2012).

The Morrill Act, which facilitates my identification strategy, was signed into law in 1862. Within several years, a land-grant university was designated in each state.[4] Large amounts of federal and state dollars were distributed to the land-grant universities annually, and people believed that spillovers from the land-grant universities had caused a strong concentration of economic activities around the universities. The historical documents suggest the designation of land-grant universities was affected by many factors other than economic considerations. According to Williams (1991), "there were no foregone conclusions as to which institution, or institutions, would receive the funds." Moretti (2004) also suggests the locations of land-grant universities were not dependent on natural resources or other factors that could make an area wealthier. Thus, this federal program provides the exogenous variation that is vital to identify spillovers from universities and the causal effects of universities on local economies.[5]

One remaining issue is that the land-grant colleges and universities were usually located in rural counties, because "the vocation in which the majority of Americans were engaged, and with which the land-grant colleges were most strongly identified, was agriculture" (Williams, 1991). Rural counties do not necessarily share the same economic attributes and trends with other counties. As a result, a comparison between the designated counties and the rest of the

[4] In 1890 and 1994, the "1890 land-grant universities" and "1994 land-grant universities" were designated. In this paper, I focus on the "1862 land-grant universities". I conduct a robust check based on the "1890 land-grant universities" in later sections.

[5] Even though the land-grant colleges and universities were different from other colleges and universities in some aspects initially, they became less so over time. Thus, my results can be interpreted as the impact of colleges and universities on local economies in the long-run.

counties in the United States is likely to generate a biased estimate. Thus, I apply a new matching technique, the synthetic control method, to obtain a more reliable estimate of the impact. This method constructs a synthetic control county, which is a weighted average of potential control counties where the weights are chosen to ensure that the "synthetic" county created is closely matched to the treated county on pre-treatment attributes including pre-treatment trends of outcome variables, for each designated county. A synthetic control county reproduces the outcome trajectory that a designated county would have experienced in the absence of the land-grant university. Once a treated county and a synthetic control county are matched on a series of outcome variables and matching variables before the designation, a discrepancy in the outcome variable following the designation is interpreted as the impact of the land-grant university on the specific county.

The synthetic control method has several advantages over regression analysis. First, it requires the selection of an appropriate donor pool to ensure the treated and comparison counties share a common economic environment.[6] In particular, as each land-grant university was designated within each state by the state authority, I ensure the designated county and the potential control counties are in the same state and share similar economic characteristics. Second, it precludes the possibility of extrapolation that regression results are often based on. Third, it accounts for the existence of time-varying unobservable confounders, which improves on panel models that account for only time-invariant unobservable confounders.

After creating a synthetic control county for each designated county, I use an event-study analysis to obtain estimates of the average impact of land-grant universities on geographic

[6] This requirement is called the "common support condition" in matching literature. Heckman, Ichimura, Smith, and Todd (1996) and Heckman, Ichimura, and Todd (1997) point out the failure of this condition can result in a substantial bias of matching estimator. Recently, Billmeier and Nannicini (2009) show the failure of standard cross-sectional estimators to control for the existence of a common support can lead to quite far-fetched estimates. A donor pool is the set of potential control counties out of which the synthetic control county is constructed.

clustering of economic activity, labor market composition and local manufacturing productivity. Event-study analyses can recover any dynamics of the impact of land-grant universities and test whether the land-grant designation followed any county-specific trends in outcome variables.[7]

A related line of research seeks to understand more generally the economic role of agglomeration and spillovers in affecting regional growth and enhancing productivity. Marshall (1890) points out three channels through which agglomeration can enhance productivity: intermediate input sharing, labor market pooling and knowledge spillovers. Rosenthal and Strange (2008) find evidence of human capital spillovers and the attenuation pattern of such spillovers.[8] This paper also contributes to this parcel of the literature by presenting evidence of spillovers from colleges and universities.

Three key results are obtained. First, the designation of land-grant universities substantially increased population density in the designated counties, relative to the synthetic control counties. Within ten years after the designation, population density in the designated counties grew by around 6 percent. The long-run effects are more profound. From the designation to 1940 (80-year impact), population density in the designated counties increased by almost 45 percent.

Second, the share of manufacturing workers in the population, an indicator of local labor market composition, was not affected by the designation.[9] My estimates of the impact on the share of manufacturing workers in the population are small and insignificant over all periods. Although the initial goal of the land-grant program was to provide accessible education to

[7] Severnini (2012) also uses a combination of synthetic control methods and event-study designs to uncover the impact of hydroelectric dams and agglomeration spillovers from the dams.

[8] See Quigley (1998), Rosenthal and Strange (2004) and Head and Mayer (2004) for a comprehensive review of the related literature.

[9] The share of manufacturing workers in the population is the number of manufacturing workers divided by county population. The manufacturing share of employment is potentially a better indicator of labor market composition. However, I do not have data on total labor force.

agricultural and industrial society, it seems that investing in such universities did not generate a detectable impact on the relative size of manufacturing sector and may not be the best way to establish an industrial city.[10]

Third, the land-grant designation greatly enhanced local manufacturing productivity, especially in the long run. On average, manufacturing output per worker increased by around $2,136 (57 percent) from the designation in the late 1860s to 1940. This positive effect of university activities on the productivity in non-education sectors suggests the existence of spillovers from knowledge production centers—colleges and universities. To be sure, one caveat of the analysis is that I cannot separately estimate the direct spillovers from universities and the induced agglomeration economies that arise from the concentration of population. However, over an 80-year horizon, I successfully show that most of the increase in manufacturing productivity was a result of direct spillovers from universities instead of induced agglomeration economies that arise from the increase in population. The fact that manufacturing output per worker rose substantially in response to the land-grant program while the share of manufacturing workers in the population did not is somewhat surprising. One possible explanation is that the land-grant universities generated spillovers for all sectors nearby, and did not disproportionately affect any given sector. Also, the data only allows me to identify sectors at a relatively rough scale, and the differential effects of the land-grant institutions across sectors may simply not be captured by the aggregated measures. However, the fact that the designated counties did not often become industrial cities is consistent to the historical documents.

Additionally, my estimates show a substantial difference between the short- and long-run effects of a large government investment project, which emphasizes the importance of

[10] This result suggests the size of local manufacturing sector was not disproportionately affected by the land-grant designation. The absolute size of local manufacturing sector can still increase as the total population actually increased.

understanding the long-run effects of such events, as advocated by Kline (2010). I also conduct robust checks, such as estimating the impact of the Second Morrill Act and using additional matching variables, as well as placebo tests. All robust checks and placebo tests suggest my results are robust to many potential concerns.

The rest of the paper is organized as follows. The next section provides the historical background of the Morrill Act and the designation of land-grant colleges and universities. Section 3 lays out my research methodology and empirical issues. Section 4 describes the data sources and variable construction. Empirical results are presented in Section 5, including robust checks and placebo tests. Section 6 concludes.

1.2 Historical Background

In the colonial era, higher education was only available at a few privately controlled institutions, such as Harvard and Yale, in the United States. After the Revolutionary War, the country began to organize universities as publicly controlled institutions, which were not essentially different in academic orientation from the privately controlled ones at that period. During the first half of the 19th century, the two types of American colleges and universities, publicly and privately controlled, developed side by side.

These institutions were greatly influenced by the European universities, which were organized to serve a society not predominantly democratic. Colleges and universities offered chiefly classical and professional curricula. During the same period, the importance of science gained recognition gradually. Agricultural colleges started to emerge. The Gardiner Lyceum, the first institution to offer scientific agriculture courses in the United States, was founded in 1823. However, higher education was still unavailable to most agricultural and industrial workers. The American higher education system needed to make a change.

Under this environment, Vermont Representative Justin Smith Morrill introduced the land-grant bill in Congress and the first Morrill Act was passed by the Congress and signed by President Lincoln on July 2, 1862. This act was the first major federal program to support higher education in the United States. It donated public lands to the states, the sale of which was for the "endowment, support, and the maintenance of at least one college where the leading object shall be, without excluding other scientific and classical studies and including military tactics, to teach such branches of learning as are related to agricultural and the mechanic arts, in order to promote the liberal and practical education of the industrial classes in the several pursuits and professional life." Fifty seven land-grant universities were established as a result of the first Morrill Act. The goal of these universities was "to develop at the college level instruction relating to the practical realities of an agricultural and industrial society and to offer to those belonging to the industrial classes preparation for the professions of life." (Association of Public and Land-grant universities, 2012) At the time, agriculture was the vocation in which the majority of Americans were engaged and with which the land-grant universities were identified. Therefore, land-grant universities were usually located in rural settings.0

The factors that affected the designation of land-grant universities were complicated. The historical documents suggest each college's founding was uniquely determined by a complex set of conditions and circumstances within its respective state. There is little evidence that suggests economic considerations played a vital part in determining the designation. According to Williams (1991), "there were no foregone conclusions as to which institution, or institutions, would receive the funds." Pennsylvania provided a case in point. Although many large colleges and universities asked for a share in the land-grant endowment, those universities were excluded from consideration because the lower house of the state believed a land-grant college could not survive as an appendage to a literary college. The Farmers' High School, founded in 1855, changed its name to the Agricultural College of Pennsylvania to stake a stronger claim on the land-grant designation approximately only two months before Lincoln signed the act. In the end, it became the sole recipient of the land-grant funds.

In the first few decades after the designation, the development of the land-grant universities was relatively slow. State support was slim, enrollments grew slowly and student attrition remained high. The situation did not change until the end of the 1880s when the Hatch Act in 1887 made new federal appropriations to the land-grant universities. In 1890, the second Morrill Act was passed, making new appropriations to the land-grant universities. To receive the money, a state had to show that race was not an admission criterion or designate a separate land-grant college for blacks to receive a portion of the funds. Eighteen new land-grant universities, known as the "1890 land-grant universities," were established in the then-segregated south. In 1994, 29 Native American colleges were designated as "the 1994 land-grant universities."[11] Although the land-grant universities started as agricultural and technical schools, many have grown into large public universities that have educated almost one-fifth of all students seeking degrees in the United States.

From the above discussion, it is clear the land-grant designation is relatively exogenous and can be viewed as a federal investment shock to local economies. The uniqueness of such a profound federal endowment program and the knowledge creation and dissemination role of universities make it especially interesting to investigate the impact of the land-grant universities on local economies.

[11] See Figure 1 for a detailed map of the distribution of land-grant colleges and universities.

1.3 Research Design and Methodology

The Morrill Act provides the exogenous variation that helps identify the causal impact of colleges and universities. However, because land-grant colleges and universities were usually located in rural counties, a simple comparison of the counties with land-grant universities and the rest of the counties in the United States will most likely generate a biased estimate. Those rural counties do not necessarily share the same economic attributes and trends with other counties. Thus, I first use a novel econometric technique, the synthetic control method, to construct a counterfactual for each treated county. Comparing the treated county with the synthetic control county provides a county-specific estimate of the impact of a land-grant university. Then, I employ an event-study analysis to obtain estimates of the average impact. This two-step procedure can be thought of as a reweighting/matching strategy to estimate treatment effects that accounts for time-varying unobserved heterogeneity.[12]

1.3.1 Synthetic control method

As discussed in Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010, 2012), a synthetic control county is intended to reproduce the counterfactual of the case of interest in the absence of the event or intervention under scrutiny. A synthetic control county is a weighted average of potential control counties where the weights are chosen to ensure that the "synthetic" county created is closely matched to the treated county on pre-treatment attributes including pre-treatment trends of outcome variables. Once treated and synthetic control counties are matched on outcome variables and matching variables over extended time periods before the intervention, a discrepancy in the outcome variable at post-intervention periods is interpreted as treatment effects.

[12] See Severnini (2012) for a discussion of this reweighting/matching method.

To provide a formal discussion of this method, suppose there is a sample of J + 1 counties indexed by j, among which unit j = 1 is the case of interest and units j = 2 to j = J + 1 are potential comparisons.[13] Units $j = 2$ to j = J + 1 constitute the donor pool, from which the synthetic control unit is constructed. Thus, it is crucial to restrict the donor pool to counties with outcomes that are thought to be driven by the same structural process as the treated unit and that were not subject to structural shocks during the sample period of this study. In my analysis, as the land-grant designation was determined within each state, I used the rest of counties in each state as the set of potential comparisons.

I also assume a balanced panel, which includes a positive number of pre-intervention periods, T_0, as well as a positive number of post-intervention periods, T_1, with $T = T_0 + T_1$. $W = (w_2, \dots, w_{J+1})'$ is a (J + 1) weight vector, with $0 \leq w_j \leq 1$ for j = 2 to J and $w_2 + \cdots + w_{J+1} = 1$. X_1 is a (k +1) vector containing the values of pre-intervention characteristics of the treated county we aim to match as closely as possible, and X_0 is the $(k \times J)$ matrix collecting the values of the same variables for the counties in the donor pool. Y_{jt} is the outcome of county j at time t. The synthetic control estimator of the impact of the intervention at time t is given by the comparison between the outcome of the treated unit and its synthetic control unit,

$$Y_{1t} - \sum_2^{J+1} w_j^* Y_{jt}. \qquad (1)$$

Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010, 2012) choose the optimal weight w^* that minimizes

$$\sum_{m=1}^{k} v_m (X_{1m} - X_{0m}W)^2, \qquad (2)$$

[13] To assume only one unit is exposed to the intervention is for expositional simplicity. In cases where multiple units are treated, one can apply this method to each treated unit separately.

where v_m is a weight that reflects the relative importance of the matching variables in accordance to their predictive power on the outcome. An optimal choice of the vector V minimizes the mean squared error of the synthetic control estimator.

The matching variables are meant to be predictors of post-intervention outcomes, which are not themselves affected by the event. The matching variables I use are a set of pre-intervention county-specific attributes and pre-intervention outcome variables.[14]

This method extends upon traditional panel models, which only allow for time-invariant unobservable factors, by allowing the effect of unobservable confounding factors to vary with time. Using this approach, I create a synthetic control county for each designated county. The comparison within each pair is the synthetic control estimate of the impact of the land-grant designation on the specific county.

1.3.2 Event-study design

An event-study analysis can recover the dynamics of the impact of the event and test if such an event happened in response to any county-specific trends in the outcome variables. I pool all pairs of designated and synthetic control counties, and use this method to obtain estimates of the average economic impact of land-grant universities.

Following the model used in Jacobson, LaLonde and Sullivan (1993), McCrary (2007) and Kline (2012), I consider the following econometric model:

$$Y_{jt} = \sum_n \beta_n D_{jt}^n + d_j + d_t + \epsilon_{jt}, \tag{3}$$

where Y_{jt} is the value of outcome variable, e.g. log of population density, in county j in calendar year t, d_j is a county fixed effect, d_t is a year fixed effect, and ϵ_{jt} is an error term that may exhibit arbitrary dependence within a case but is uncorrelated with other right-hand side

[14] See data section for details.

variables.[15] The county and year fixed effects ensure that my research design is not subject to contamination from state-wide temporal shocks.

The variables D_{jt}^{n} are a series of event-time dummies that equal 1 when the land-grant designation is *n* years away in a given county. Formally, it is

$$D_{jt}^{n} = I[t - e_j = n], \tag{4}$$

where I[.] is an indicator function for the expression in brackets being true, and e_j is the event time (in this case, the year of the land-grant designation in county *j*).

Based on the model, the β_n coefficients represent the time path of the outcome variable relative to the date of intervention for the treated counties conditional on the three unobserved variance components, d_j, d_t and ϵ_{jt}. If land-grant universities are randomly assigned between the treated counties and synthetic control counties, the restriction $\beta_n = 0$ should hold for all pre-intervention periods. In other words, the land-grant designation should not be, on average, preceded by county-specific trends in outcome variables. Also, because not all of the βs can be identified due to the collinearity of event-time dummies and county fixed effects, I normalize $\beta_{-1} = 0$, so all post-intervention coefficients can be thought of as treatment effects.[16] Each synthetic control unit is intrinsically associated with its treated counterpart, so I cluster the standard errors at the case level.

[15] In the empirical analysis, I also experiment with case fixed effects, region-by-year fixed effects and division-by-year fixed effects. A case is a pair of a treated county and its corresponding synthetic control county. As will be apparent, the results are robust to the change in fixed effects.

[16] In my analysis, the first Morrill Act was passed in 1862, and the designation of land-grant universities was mostly determined within the next several years. Because I use a decennial data set, I set year 1870 as the intervention period and normalize the coefficient of event-time dummy for 1860 to 0. In Jacobson, LaLonde, and Sullivan (1993), McCrary (2007) and Kline (2012), certain endpoint restrictions are applied, which simply state that any dynamics wear off after certain years. Because the intervention time in my analysis is the same for all treatment units, I implicitly have such endpoint restrictions in my analysis.

1.4 Data Description

This section describes the data sets used in this paper. County-level data on population, number of manufacturing workers, manufacturing output and other county-specific attributes are drawn from the U.S. census of population (Haines and ICPSR, 2010). County level geographic information, such as county area, latitude and longitude, comes from The National Historical Geographic Information System (NHGIS). The information on the land-grant designation is obtained from Integrated Postsecondary Education Data System (IPEDS) and Association of Public and Land-grant universities. The market access data from 1870 are from Donaldson and Hornbeck (2012).[17]

The sample is restricted to counties for which data are available in each decennial census from 1840 to 1940. As a result, my basic dataset is a balanced panel of 1180 U.S. counties from 1840 through 1940. This large sample ensures most of my synthetic control counties are not constructed based on a thin donor pool. Some county boundaries changed over this time period; therefore, data are adjusted in later periods to maintain the 1840 county definition (Hornbeck, 2010). All dollar variables, such as manufacturing output, are reported in 1840 dollars (inflation data comes from The Federal Reserve Bank of Minneapolis).

A natural measure of economic concentration is population density. This outcome variable is intended to capture the overall impact of the land-grant designation on local economies. Other outcome variables include the share of manufacturing workers in the population and manufacturing output per worker. The share of manufacturing workers in the population is an indicator of labor market composition. The manufacturing share of employment is potentially a better indicator of labor market composition. However, I do not have data on total

[17] A market access can be viewed as a measure of how easily a county can trade with other counties, it is a reduced-form expression derived from general equilibrium trade theory by Donaldson and Hornbeck (2012).

labor force. Manufacturing output per worker is the dollar value of manufacturing output produced in the county divided by the number of manufacturing workers in the county. It indicates local manufacturing productivity. An increase in manufacturing output per worker in the treated counties at post-intervention periods is potentially caused by spillovers from the land-grant universities.

The matching variables I use in the construction of synthetic control counties include percentage of urban population, percentage of white population, per capita agricultural output, per capita farm value, percentage of college students in the population and all pre-intervention outcome variables.[18] These variables are considered the predictors of post-intervention outcomes. Other variables, such as the market access for each county and counties' latitudes and longitudes, are used as additional matching variables in robust checks.

1.5 Results

1.5.1 The impact on population density

In this section, I present the estimates of the impact of land-grant universities on population density. I first show the impact of the land-grant designation case by case for a representative group of counties. This is the county-specific estimate, obtained with synthetic control methods. I then present the estimates of the average impact of land-grant universities on population density for all treated counties in my sample. This part of the results is obtained through event-study analyses.

[18] Per capita agricultural output is the total agricultural output in the county divided by the total population in the county. The data on number of workers working in agricultural sector is not available.

1.5.1.1 Synthetic control method: County-specific estimates

The synthetic control method constructs a counterfactual for each treated county. Thus, I can estimate the county-specific impact of the land-grant designation for each treated county. I show several representative cases here and the others are presented graphically in Appendix A.

Immediate Impact

Figure 2, panel A, displays a case of immediate impact of the land-grant designation on population density. In the figure, the time path of population density in Knox County, Tennessee and the synthetic Knox County matches very well from 1840 to the late 1860s. However, after East Tennessee College[19] was designated to receive the land-grant funds in 1869, population density in Knox County grew much faster than the synthetic Knox County. This trend continued to 1940, the end of my sample period. The impact of the land-grant designation from the late 1860s through 1940 was approximately 1.16 log points (219 percent).[20] To show how these numbers are calculated, I also present the comparison between Knox County and the synthetic Knox County numerically in Table 2.

Table 1, panel A, compares the pre-treatment characteristics of Knox County and the synthetic Knox County, as well as the state average. The state average does not appear to provide a suitable control. In particular, the state average of pre-intervention population density is substantially lower than Knox County. In contrast, the synthetic Knox County accurately

[19] It was renamed The University of Tennessee in 1879.

[20] I calculate the impact of the land-grant designation from the late 1860s to 1940 as the difference of population density between the treated county and its synthetic control county in 1940, minus the difference of population density between the treated and its synthetic control county in 1860. The latter difference is almost zero, which suggests the synthetic control county simulates the treated county well.

reproduces the values of pre-intervention population density and most other matching variables for Knox County.[21]

Lagged Impact

A case of lagged impact of the land-grant designation is presented in Figure 2, panel B. The University of Maine was established in 1865 as a land grant college in Penobscot County, Maine. In the first 30 years, the new land-grant university had no obvious impact. The population density in Penobscot County did not appear to differ from the synthetic Penobscot County until 1890. After 1890, Penobscot County displayed faster growth in population density relative to the synthetic Penobscot County. The impact of the land-grant designation from the late 1860s through 1940 was around 0.44 log points (55 percent), which all happened between 1890 and 1940.

Indifference

An unattractive case from a policymaking point of view is displayed in Figure 2, panel C. The time path of population density in Ingham, Michigan, and the synthetic Ingham County did not differ significantly despite the designation of Michigan State University in the 1860s. From the figure and Table 2, the impact of Michigan State University was only 0.15 log points (16 percent) until 1920 and was slightly larger after that. The county would not be much worse off without the new land-grant university.

Reversion

Figure 2, panel D portrays a disturbing case of a public investment. After the Agricultural College of Pennsylvania was designated as a land-grant university in 1863, Centre County,

[21] The comparisons of pre-treatment characteristics between other representative counties and their corresponding synthetic control counties are presented in Table 1 and Appendix B. The general pattern is the same: The synthetic control counties match the treated counties better than the state average.

Pennsylvania, experienced a growth of 0.10 log points (11 percent) from the 1860s to 1890 in population density, relative to the synthetic Centre County. Nevertheless, the trend reversed after 1890 and the county experienced a drop of 0.48 log points (62 percent) in population density from 1890 through 1940.

1.5.1.2 Event-study analysis: Pooled estimates

To estimate the average impact of the land-grant universities, I pool all pairs of treated and synthetic control counties and estimate equation (3). The synthetic control method takes into account both observed and unobserved county level heterogeneity. Meanwhile, the Event-study analysis recovers the dynamics of the impact of the land-grant designation and tests if such an event happened in response to any county-specific trends.

The coefficient estimates on the event-time dummies are presented in Table 3. I estimate four different models. Model 1 includes case fixed effects and year fixed effects, Model 2 county fixed effects and year fixed effects, Model 3 county fixed effects and region-by-year fixed effects, and Model 4 county fixed effects and division-by-year fixed effects. The results are robust to the change in fixed effects. Model 2 is sufficient to eliminate all pre-treatment trends: the coefficient estimates on the event-time dummies for 1840 and 1850 are small and highly insignificant. Therefore, I focus on discussing this model.

The results are quite interesting. First, the magnitude of the impact from the land-grant universities is remarkably large. On average, population density in designated counties grew by around 0.06 log points (6 percent) within only ten years after the designation, compared to the synthetic control counties. This short-run impact could be caused by the fact that new jobs were created and more students enrolled in the county as large federal and state endowments poured into the designated counties. From the 1860s to 1940 (80-year impact), population density in designated counties increased by 0.37 log points (45 percent). This long-term impact is more likely to be caused by potential spillovers from university activities. All these estimates are highly significant.

Second, the difference between the short- and long-run effects is revealing. The 40-year estimate is around 0.10 log points (11 percent), only around one-quarter of its 80-year counterpart (0.37 log points, or 45 percent). This is consistent with the history of land-grant universities that their development was relatively slow in the first several decades. It may also imply that the impact can re-enforce itself in the long-run, which is consistent with the predictions of the theory of agglomeration economies. This suggests the assessment of large government projects require understanding of both short- and long-run effects.

1.5.2 The impact on share of manufacturing workers

The initial target of land-grant universities was agricultural and industrial society. Also, at that time, the development of manufacturing sector was a leading factor in city development. Thus, a natural question to ask is how the land-grant universities affect manufacturing sector.[22] Although the land-grant funds were poured into education sector, spillovers from university activities can still generate important impact on manufacturing sector.

Thus, using the same two-step procedure, I also look at the impact of the land-grant designation on the share of manufacturing workers in the population, an indicator of local labor market composition. In this and later sections, I only present the results from event-study

[22] It is also interesting to see how the land-grant program affects agricultural sector. However, the data forbids me to do further investigation toward that direction. Also, the emphasis of this paper is to provide evidence of spillovers from university activities.

analyses—estimates of the average impact from the land-grant designation. The county-specific estimates are available upon request.

Table 4 presents the estimates of the short- and long-run effects of the land-grant designation on the share of manufacturing workers in the population. Similarly, four different specifications are estimated. The results are robust to the specification adjustments and I focus on discussing Model 2. The 1910 event-time dummy is omitted because data on manufacturing workers in 1910 is missing. The coefficient of 1860 event-time dummy is normalized to 0.

The estimates suggest the land-grant designation did not substantially affect the percentage of manufacturing workers in the population. All post-intervention coefficient estimates on the event-time dummies are very small and highly insignificant. On average, the share of manufacturing workers in the population grew only 0.2 percentage points within ten years after the designation. The largest impact in my sample period was only 0.6 percentage points, which occurred 70 years after the designation.

These results are important from a policy-making point of view. A particular goal of the land-grant universities is "to develop at the college level instruction relating to the practical realities of an agricultural and industrial society" and the initial focus of the curriculum in those universities is agricultural and engineering related (Association of Public and Land-grant universities, 2012). However, my results suggest local manufacturing sector was not disproportionately affected by the land-grant designation, despite the strong manufacturing orientation of the land-grant program. This yields potential implications for policy makers who try to develop an industrial town by investing in higher education. However, as will be apparent in later sections, local manufacturing productivity was substantially enhanced by the land-grant program.

1.5.3 The impact on manufacturing output per worker

In this section, I present my estimates of the impact of the land-grant designation on local manufacturing productivity, as measured by manufacturing output per worker.[23] Because manufacturing workers were not directly affected by the Morrill Act, enhancement of productivity in manufacturing sector was potentially caused by spillovers from land-grant universities.[24]

Table 5 shows the short- and long-run effects of the land-grant designation on local manufacturing output per worker. Four different models are estimated and I only focus on Model 2 for aforementioned reasons. The estimated short-term impact is not significant. On average, manufacturing output per worker increased by only around $102 (7 percent) from the designation to 1890.[25] However, the enormous magnitude of the estimated impact in the long-run seems remarkable. In 1940, the estimated impact (80-year impact) increased to $2,136 (57 percent). This large impact caused by land-grant universities seems to re-enforce itself in the long run as the local industries evolve over time to take advantage of the spillovers (Kantor and Whalley, 2012).

As aforementioned, it is difficult to estimate separately the direct spillovers from university activities and the induced agglomeration economies that arise from the concentration of population. However, over an 80-year horizon, I show that most of the increase in

[23] Manufacturing output per worker is the dollar value of manufacturing output produced in the county divided by the number of manufacturing workers in the county. At the time, multi-site companies were not as common as today. Therefore, this measure can be a good indicator of productivity.

[24] Although I cannot estimate separately the two aforementioned mechanisms through which universities affect local productivity and it is not the emphasis of this paper, I argue the driving force of my results should be knowledge spillovers rather than training of students who stay locally. The coverage of a large university usually spans larger than a county. Thus, if the increase in productivity is caused by the increase in workers' education levels, the nearby counties should experience the same (or slightly less) productivity gain. Then, if I construct my synthetic control counties based on a donor pool that is near the treated county, the effect of the land-grant designation should fade away. However, I show the effect does not fade away when the counties in the donor pool are near the treated county in the robust checks.

[25] I use 1840 dollars values throughout the paper.

manufacturing productivity was a result of direct spillovers from universities instead of induced agglomeration economies that arise from the increase in population. In the literature, the range of estimated urbanization elasticities is between 2 percent to 5 percent.[26] Combes *et al.* (2008) report urbanization elasticities in France that range from 2.5 to 4.7 percent depending on the number of controls included in the model. Ciccone (2002) estimates an elasticity of 4.5 percent drawing on data from several countries in Europe. Ciccone and Hall (1996) estimate an elasticity of 5 percent based on state-level data in the United States. Rosenthal and Strange (2008) estimate urbanization elasticities that are in the range of 3 to 5 percent. I take the upper bound of the estimated urbanization elasticities in the literature, 5 percent, to do a simple calculation. The estimated 80-year impact of the land-grant designation on population density is 45 percent. Thus, the implied productivity gain from induced agglomeration economies that is caused by the increased population is only 2.25 percent. This is only a small fraction of the estimated 80-year productivity gain in the manufacturing sector caused by the land-grant designation. It suggests the impact of the land-grant designation on manufacturing productivity comes mostly from the direct spillovers from university activities.

These findings are especially important given the results in the last section that the land-grant designation had no substantial effects on local labor market composition. It explains the existence of many college towns in the United States. College towns are the beneficiaries of spillovers from universities; however, they do not often develop as industrial cities. Cornell University, one of the most famous land-grant universities, stimulated a small but active industrial sector in Ithaca. But Ithaca never developed into a large industrial city. The fact that manufacturing output per worker rose substantially in response to the land-grant program while

[26] An urbanization elasticity of 1 percent means doubling the nearby population increases productivity by 1 percent. This is called the urbanization effect in the agglomeration literature.

the share of manufacturing workers in the population did not is somewhat surprising. One possible explanation is that the land-grant universities generated spillovers for all sectors nearby, and did not disproportionately affect any given sector. Also, the data only allows me to identify sectors at a relatively rough scale, and the differential effects of the land-grant institutions across sectors may simply not be captured by the aggregated measures.

1.5.4 Robust checks and specification issues

Although my results are robust to various specifications of fixed effects, it is still important to conduct additional robust checks. In this section, I present the results from robust checks and placebo tests for log population density.[27] In certain cases, I also present the results for manufacturing output per worker.

First, I use the same procedure to estimate the impact of the “1890 land-grant universities.” The 1890 land-grant universities were established based the Second Morrill Act in 1890. The Second Morrill Act is quite different from the first act on policy target, appropriation amount and selection criteria. Thus, it is not appropriate to simply pool the “1862 land-grant universities” and the “1890 land-grant universities” and estimate an average impact. However, using the “1890 land-grant universities” to conduct my two-step procedure has the advantage of a longer pre-intervention period. Abadie, Diamond and Hainmueller (2010) suggest a long pre-intervention period helps control for unobserved factors affecting the outcome of interest as well as for heterogeneity of the effect of the observed and unobserved factors.

The results are presented in Table 6. In Model 2, on average, the impact of 1890 land-grant universities on population density in ten years was around 0.06 log points (6 percent).[28] To

[27] The results for other outcome variables are available upon request. They all suggest my conclusions are robust.
[28] The Second Morrill Act was passed in 1890, so the coefficient estimate of the event-time dummy for 1890 can be viewed as an immediate impact and the coefficient estimate of event-time dummy for 1900 is the 10-year impact.

1940, the impact was 0.27 log points (31 percent). These estimates are qualitatively the same as my previous estimates based on the "1862 land-grant universities," although less significant.[29] It suggests the length of the pre-intervention periods in my main specifications is not a concern.

Second, I add counties' market access in 1870 as an additional matching variable and conduct my two-step procedure to estimate the impact of "1862 land-grant universities."[30] This measure of market access (Donaldson and Hornbeck, 2012) is a novel measure that summarizes how easily a county can trade with other counties. Although the set of matching variables in my main specification seems comprehensive enough, it is helpful to see whether the inclusion of additional matching variables changes my conclusions. The results are in Table 7. It shows my estimates are robust to the inclusion of this additional matching variable.

Third, I include the cubic function in counties' latitudes and longitudes as additional controls and estimate the impact of "1862 land-grant universities." Some may argue that counties near the treated county geographically are potentially better control units than the rest of the counties in the state. Matching on the cubic function in counties' latitudes and longitudes ensures the treated and the synthetic control counties are geographically close. The results are in Table 8. The estimates are consistent with my previous findings. I also present the results for manufacturing output per worker when latitudes and longitudes are controlled for in Table 9. Although I do not try to identify separately the two potential mechanisms through which universities affect local productivity, I argue that the driving force of my results should be knowledge spillovers rather than training of students who stay locally. The coverage of a large university usually spans larger than a county. If the increase in productivity is caused by the increase in workers' education levels, the nearby counties should experience the same (or

[29] This may be because of the smaller sample size in these regressions.
[30] The land-grant universities were generally designated before 1870. Thus, an implicit assumption here is that counties' market access had not changed from the land-grant designation to 1870.

slightly less) productivity gain. Then, the effects should fade away when I construct the synthetic control counties based on the nearby counties. However, the effects do not fade away at all in Table 9.

Finally, to ensure my research design captures the impact of the land-grant universities rather than some random factors or unobserved interventions, I run placebo tests. I run the same two-step procedure except now I choose the treated county randomly. My previous findings would be undermined if I obtained a similar or even greater effect when the treated counties are randomly selected (where the intervention did not take place). I run the two-step procedure 20 times. The estimated effects of artificial treatments are shown in Figure 3. The heavy solid black line is the impact of the real treatment, which is plotted for comparison purpose. It is obvious the effect of the real treatment is larger than any placebo effects. Because I conduct the placebo tests 20 times, the probability of estimating a placebo effect as large as the true effect is 5 percent, a test level typically used in conventional tests of statistical significance.

1.6 Conclusions

The success of "Silicon Valley" and Route 128 is glaring. The attempt to mimic such success has never stopped. Most recently, Cornell University, and its partner, Technion-Israel Institute of Technology, won the right to build a facility for job-spinning engineering research on Roosevelt Island in New York City, aiming to increase entrepreneurship and job growth in the city's technology sector. However, the precise linkage among educational investment, potential spillovers and regional development remains unclear because of the identification challenges that arise from the feedback effects from business activity and the common factors affecting both universities and business environment. In this paper, I seek to fill part of this gap.

My identification strategy is that I treat the designation of land-grant universities in the United States in the 1860s as a natural experiment after controlling for the confounding factors with synthetic control methods and event-study analyses. Using this strategy, I present evidence of direct spillovers from universities and examine the short- and long-run effects of university activities on geographic clustering of economic activity, labor market composition and local productivity.

Several key conclusions are obtained. First, population density in the designated counties grew substantially as a result of the land-grant designation. On average, population density in the designated counties rose by around 6 percent within ten years and grew by 45 percent in 80 years after the designation, relative to the synthetic control counties. Second, the land-grant designation did not appear to affect the share of manufacturing workers in the population, an indicator of local labor market composition. Third, manufacturing productivity in the designated counties, as captured by manufacturing output per worker, was greatly enhanced by the designation in the long-run. Within 80 years after the designation, manufacturing output per worker climbed by around $2136 (57 percent) in the designated counties. This impact on the productivity in non-education sectors suggests the existence of spillovers from universities. Over an 80-year horizon, I estimate that most of the increase in manufacturing productivity was a result of direct spillovers from universities instead of induced agglomeration economies that arise from the increase in population. The robust checks and placebo tests suggest these results are robust to many potential concerns.

There are broad policy implications of these results. My results suggest that investing in higher education may not serve well as a policy tool to develop an industrial city because the land-grant universities had no substantial impact on the size of local manufacturing sector compared to other sectors. However, the land-grant universities greatly enhanced local manufacturing productivity. A possible explanation is that universities have an equal impact on all sectors in a city. These findings are also consistent with the existence of many college towns in the United States, such as Ithaca, New York. Cornell University, the world famous land-grant university, stimulated an active but small industrial sector, but the town never became a large industrial city.

References

Abadie, A., Diamond, A., Hainmueller, J., 2007. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, Journal of The American Statistical Association 105(490), 493-505.

Abadie, A., Diamond, A., Hainmueller, J., 2012. Comparative Politics and The Synthetic Control Method, Working Paper.

Abadie, A., Gardeazabal, J., 2003. The Economic Costs of Conflict: A Case Study of the Basque Country, The American Economic Review 93(1), 113-132.

Abramovsky, L., Harrison, R., Simpson, H., 2007. University Research and the Location of Business R&D, Economic Journal 117(519), C114-C141.

Acs, Z.J., Audretsch, D.B., Feldman, M.P., 1992. Real Effects of Academic Research: Comment, The American Economic Review 82(1), 363-367.

Adams, J.D., 2002. Comparative localization of academic and industrial spillovers, Journal of Economic Geography 2(3), 253-278.

Aghion, P., Boustan, L., Hoxby, C., Vandenbussche, J., 2009. The Causal Impact of Education on Economic Growth: Evidence from U.S., Working Paper.

Andersson, R., Quiley, J.M., Wilhelmsson, M., 2004. University decentralization as regional policy: the Swedish experiment. Journal of Economic Geography 4(4), 371-388.

Andersson, R., Quiley, J.M., Wilhelmsson, M., 2009. Urbanization, productivity, and innovation: Evidence from investment in higher education. Journal of Urban Economics 66(1), 2-15.

Anselin, L., Varga, A., Acs, Z., 1997. Local Geographic Spillovers between University Research and High Technology Innovations, Journal of Urban Economics 42(3), 422-448.

Association of Public and Land-grant universities, The Land Grant Tradition, Washington, D.C. 2012.

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

Bania, N., Eberts, R.W., Fogarty, M.S., 1993. Universities and the Startup of New Companies: Can We Generalize from Route 128 and Silicon Valley?, The review of economics and statistics 75(4), 761-766.

Beeson, P., Montgomery, E., 1993. The Effects of Colleges and Universities on Local Labor Markets, The review of economics and statistics 75(4), 753-761.

Billmeier, A., Nannicini, T., 2009. Trade Openness and Growth: Pursuing Empirical Glasnost, IMF Staff Papers, Palgrave Macmillan, 56(3), 447-475.

Billmeier, A., Nannicini, T., 2013. Assessing Economic Liberalization Episodes: A Synthetic Control Approach, Review of Economics and Statistics 95(3), 983-1001.

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and Impacts: The Influence of Public Research on Industrial R&D, Management Science 48(1, Special Issue on University Entrepreneurship and Technology Transfer), 1-23.

Donaldson, D., Hornbeck, R., 2012. Railroads and American economic growth: a "market access" approach, Working paper.

Haines, Michael R., and Inter-university Consortium for Political and Social Research. Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896.v3

Hausman, N., University innovation, local economic growth, and entrepreneurship, Harvard University working paper.

Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. Handbook of Regional and Urban Economics 4, 2609-2669.

Heckman, J.J., Ichimura, H., Smith, J., Todd, P.E., 1996. Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method, Proceedings of the National Academy of Sciences 93, 13416-13420.

Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, Review of Economic Studies 64(4), 605-54.

Hornbeck, R., 2010. Barbed wire: Property rights and agricultural development, Quarterly Journal of Economics 125(2), 767-810.

Jacobson, L.S., LaLonde, R.J., Sullivan, D.G., 1993. Earnings Losses of Displaced Workers, The American Economic Review 83(4), 685-709.

Jaffe, A.B., 1989. Real Effects of Academic Research, The American Economic Review 79(5), 957-970.

Kantor, S., Whalley, A., 2012. Knowledge spillovers from research universities: Evidence from endowment value shocks, Working paper.

Kline, P., 2010. Place Based Policies, Heterogeneity, and Agglomeration, The American Economic Review 100(2, PAPERS AND PROCEEDINGS OF THE One Hundred Twenty Second Annual Meeting OF THE AMERICAN ECONOMIC ASSOCIATION), 383-387.

Kline, P., 2012. The Impact of Juvenile Curfew Laws on Arrests of Youth and Adults, American Law and Economic Review 14(1), 44-67.

Marshall, A., 1890. Principles of Economics. Macmillan, London.

McCrary, J., 2007. The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police, The American Economic Review 97(1), 318-353.

Moretti, E., 2004. Human capital externalities in cities, Handbook of Regional and Urban Economics 4(51), 2243-2291.

Quigley, J.M., 1998. Urban Diversity and Economic Growth. The Journal of Economic Perspectives 12(2), 127-138.

Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. Handbook of Regional and Urban Economics 4, 2119-2171.

Rosenthal, S.S., Strange, W.C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64(2), 373-389.

Saxenian, A., 1994. Regional advantage: Cultural and competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

Severnini, E.R., 2012. The power of hydroelectric dams: Agglomeration spillovers. Working paper.

Varga, A., 2000. Local Academic Knowledge Transfers and the Concentration of Economic Activity, Journal of Regional Science 40(2), 289-309.

Williams, R.L., 1991. The Origins of Federal Support for Higher Education, The Pennsylvania State University Press, Pennsylvania.

Woodward, D., Figueiredo, O., Guimaraes, P., 2006. Beyond the Silicon Valley: University R&D and high-technology location, Journal of Urban Economics 60(1), 15-32.

Table 1-1: Population Density Predictor Means

Panel A. Population Density Predictor Means --- Knox, Tennessee

	Knox, Tennessee	Synthetic Control	The State Average
Log(Population density), 1840	3.3056	3.3014	2.8562
Log(Population density), 1850	3.4999	3.5105	3.0605
Log(Population density), 1860	3.7389	3.7298	3.1809
Percent of Manufacturing Workers	0.0193	0.0161	0.0151
Manufacturing Output Per Worker	1460.45	1235.61	1265.66
Per Capita Agricultural Output	44.2339	52.4448	58.9446
Percent of Urban Population	0.0000	0.0556	0.0088
Percent of White Population	0.8721	0.7740	0.8052
Per Capita Farm Value	167.255	218.273	174.813
Per Capita College Students	0.0051	0.0058	0.0007

Panel B. Population Density Predictor Means --- Penobscot, Maine

	Penobscot, Maine	Synthetic Control	The State Average
Log(Population density), 1840	2.3044	2.3131	2.8019
Log(Population density), 1850	2.6249	2.6325	2.9225
Log(Population density), 1860	2.7529	2.7430	3.0347
Percent of Manufacturing Workers	0.0551	0.0547	0.0440
Manufacturing Output Per Worker	1505.154	919.885	1121.29
Per Capita Agricultural Output	32.3850	21.4897	42.2654
Percent of Urban Population	0.2070	0.0777	0.0895
Percent of White Population	0.9983	0.9982	0.9982
Per Capita Farm Value	89.4920	60.2134	123.7295
Per Capita College Students	0	0	.0003

Note. This table shows the mean values of population density predictors for two counties: Knox, Tennessee and Penobscot, Maine. All variables except log population density are averaged for the 1840-1860 period. Dollar variables are reported in 1840 dollars. Percent of Manufacturing Workers is the percentage of manufacturing workers in the whole population. Per Capita Agricultural Output, Per Capita Farm Value and Per Capita College Students are calculated as the total agricultural output, total farm value and total college students in the county divided by county population.

Table 1-2: Population Density Trend Comparisons

Year	Knox, Tennessee	Synthetic Control	Penobscot, Maine	Synthetic Control
1840	3.3056	3.3014	2.3044	2.3131
1850	3.4999	3.5105	2.6249	2.6325
1860	3.7389	3.7298	2.7529	2.7430
1870	3.9756	3.9599	2.7998	2.7640
1880	4.2754	4.0962	2.7692	2.7957
1890	4.6843	4.1987	2.8180	2.7990
1900	4.9025	4.3504	2.8822	2.8197
1910	5.1311	4.3797	3.0061	2.7804
1920	5.3095	4.4117	3.0427	2.7488
1930	5.6271	4.4935	3.0970	2.6721
1940	5.7587	4.5923	3.1501	2.6796

Year	Ingham, Michigan	Synthetic Control	Centre, Pennsylvania	Synthetic Control
1840	1.4936	1.4919	2.9140	2.8722
1850	2.7335	2.7316	3.0448	3.0891
1860	3.4366	3.4340	3.1898	3.2085
1870	3.8077	3.9013	3.4326	3.2840
1880	4.0949	4.1774	3.5295	3.4453
1890	4.2069	4.3442	3.6614	3.5656
1900	4.2625	4.4447	3.6527	3.8039
1910	4.5543	4.5783	3.6650	4.0721
1920	4.9794	4.8319	3.6851	4.2381
1930	5.3368	5.0723	3.7290	4.2253
1940	5.4504	5.1556	3.8568	4.2787

Note. This table presents the comparison of log population density between the representative counties and their corresponding synthetic control counties in the sample period. These results are also showed graphically in Figure 2.

.

Table 1-3: Short- and Long-Run Effects of 1862 Land-Grant Universities on Population Density (Dependent variable: log of population density; cluster-robust t-ratios in the parentheses)				
	Model 1	Model 2	Model 3	Model 4
Year 1840	0.0350	0.0367	0.0367	0.0367
	(1.35)	(1.01)	(0.97)	(0.93)
Year 1850	0.0120	0.0137	0.0137	0.0137
	(1.01)	(0.69)	(0.67)	(0.64)
Year 1870	0.0595	0.0613	0.0613	0.0613
	(2.23)	(2.38)	(2.29)	(2.19)
Year 1880	0.0527	0.0545	0.0545	0.0545
	(1.59)	(1.75)	(1.68)	(1.61)
Year 1890	0.0905	0.0923	0.0923	0.0923
	(1.68)	(1.72)	(1.65)	(1.59)
Year 1900	0.0945	0.0963	0.0963	0.0963
	(1.40)	(1.42)	(1.37)	(1.31)
Year 1910	0.1412	0.1430	0.1430	0.1430
	(1.61)	(1.65)	(1.58)	(1.52)
Year 1920	0.1989	0.2007	0.2007	0.2007
	(1.99)	(2.04)	(1.96)	(1.88)
Year 1930	0.3143	0.3161	0.3161	0.3161
	(2.94)	(2.96)	(2.85)	(2.73)
Year 1940	0.3671	0.3689	0.3689	0.3689
	(3.21)	(3.24)	(3.11)	(2.98)
Observations	462	462	462	462
Case FE	21	-	-	-
County FE	-	42	42	42
Year FE	11	11	-	-
Region by Year FE	-	-	44	-
Division by Year FE	-	-	-	77
R-squared	0.849	0.863	0.898	0.921

Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on population density. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects.

Table 1-4: Short- and Long-Run Effects of 1862 Land-Grant Universities on Percentage of Manufacturing Workers
(Dependent variable: Percentage of Manufacturing Workers; cluster-robust t-ratios in the parentheses)

	Model 1	Model 2	Model 3	Model 4
Year 1840	-0.0002	-0.0004	-0.0004	-0.0004
	(-0.25)	(-0.55)	(-0.53)	(-0.50)
Year 1850	0.0009	0.0006	0.0006	0.0006
	(0.96)	(0.48)	(0.46)	(0.44)
Year 1870	0.0021	0.0019	0.0019	0.0019
	(0.42)	(0.36)	(0.35)	(0.33)
Year 1880	0.0020	0.0018	0.0018	0.0018
	(0.41)	(0.35)	(0.34)	(0.32)
Year 1890	0.0034	0.0032	0.0032	0.0032
	(0.64)	(0.60)	(0.57)	(0.55)
Year 1900	-0.0006	-0.0008	-0.0008	-0.0008
	(-0.11)	(-0.14)	(-0.14)	(-0.13)
Year 1920	0.0034	0.0031	0.0031	0.0031
	(0.32)	(0.30)	(0.29)	(0.27)
Year 1930	0.0066	0.0059	0.0059	0.0059
	(0.60)	(0.53)	(0.51)	(0.48)
Year 1940	-0.0036	-0.0030	-0.0030	-0.0030
	(-0.46)	(-0.38)	(-0.36)	(-0.35)
Observations	394	394	394	394
Case FE	20	-	-	-
County FE	-	40	40	40
Year FE	10	10	-	-
Region by Year FE	-	-	40	-
Division by Year FE	-	-	-	70
R-squared	0.726	0.772	0.806	0.843

Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on the percentage of manufacturing workers in the whole population. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects. Data on number of manufacturing workers in 1910 is missing.

Table 1-5: Short- and Long-Run Effects of 1862 Land-Grant Universities on Manufacturing Output Per Worker (Dependent variable: Manufacturing Output Per Worker (in 1840 dollars); cluster-robust t-ratios in the parentheses)				
	Model 1	Model 2	Model 3	Model 4
Year 1850	4.7585	0.1642	0.1642	0.1642
	(1.25)	(0.33)	(0.32)	(0.30)
Year 1870	-95.6020	-100.1964	-100.1964	-100.1964
	(-1.30)	(-1.32)	(-1.26)	(-1.20)
Year 1880	-72.6298	-77.2241	-77.2241	-77.2241
	(-0.64)	(-0.67)	(-0.64)	(-0.61)
Year 1890	106.3597	101.7654	101.7654	101.7654
	(0.96)	(0.88)	(0.85)	(0.80)
Year 1900	47.1020	42.5077	42.5077	42.5077
	(0.19)	(0.17)	(0.16)	(0.15)
Year 1920	737.7035	733.1091	733.1091	733.1091
	(1.58)	(1.52)	(1.45)	(1.38)
Year 1930	1,265.8356	1,261.7632	1,261.7632	1,261.7632
	(2.09)	(2.03)	(1.94)	(1.85)
Year 1940	2,115.9866	2,136.4186	2,136.4186	2,136.4186
	(1.80)	(1.80)	(1.73)	(1.64)
Observations	332	332	332	332
Case FE	19	-	-	-
County FE	-	38	38	38
Year FE	9	9	-	-
Region by Year FE	-	-	36	-
Division by Year FE	-	-	-	63
R-squared	0.557	0.618	0.694	0.776

Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on manufacturing output per worker. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects. Data on manufacturing output in 1840 and 1910 is missing.

Table 1-6: Short- and Long-Run Effects of 1890 Land-Grant Universities on Population Density
(Dependent variable: log of population density; cluster-robust t-ratios in the parentheses)

	Model 1	Model 2	Model 3	Model 4
Year 1840	0.0491	0.0032	0.0032	0.0032
	(1.18)	(0.10)	(0.09)	(0.09)
Year 1850	0.0427	-0.0032	-0.0032	-0.0032
	(1.44)	(-0.07)	(-0.07)	(-0.07)
Year 1860	0.0171	-0.0288	-0.0288	-0.0288
	(0.63)	(-1.18)	(-1.12)	(-1.09)
Year 1870	-0.0003	-0.0461	-0.0461	-0.0461
	(-0.01)	(-0.89)	(-0.85)	(-0.82)
Year 1890	0.0797	0.0339	0.0339	0.0339
	(1.35)	(1.00)	(0.94)	(0.92)
Year 1900	0.1101	0.0642	0.0642	0.0642
	(1.80)	(1.39)	(1.32)	(1.28)
Year 1910	0.1897	0.1438	0.1438	0.1438
	(2.06)	(1.58)	(1.50)	(1.45)
Year 1920	0.2200	0.1742	0.1742	0.1742
	(1.77)	(1.38)	(1.31)	(1.27)
Year 1930	0.3067	0.2609	0.2609	0.2609
	(1.84)	(1.51)	(1.43)	(1.39)
Year 1940	0.3111	0.2652	0.2652	0.2652
	(1.70)	(1.39)	(1.32)	(1.28)
Observations	242	242	242	242
Case FE	11	-	-	-
County FE	-	22	22	22
Year FE	11	11	-	-
Region by Year FE	-	-	33	-
Division by Year FE	-	-	-	44
R-squared	0.818	0.836	0.850	0.927

Notes. This table presents the short- and long-run effects of the 1890 land-grant universities on population density. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1880 event-time dummy is normalized to 0, so all coefficients after 1890 can be thought of as treatment effects.

Table 1-7: Effects of 1862 Land-Grant Universities on Population Density-with Market Access Controls
(Dependent variable: log of population density; cluster-robust t-ratios in the parentheses)

	Model 1	Model 2	Model 3	Model 4
Year 1840	0.0340	0.0369	0.0369	0.0369
	(1.31)	(1.01)	(0.98)	(0.93)
Year 1850	0.0109	0.0138	0.0138	0.0138
	(0.93)	(0.70)	(0.67)	(0.64)
Year 1870	0.0574	0.0603	0.0603	0.0603
	(2.19)	(2.36)	(2.27)	(2.17)
Year 1880	0.0479	0.0508	0.0508	0.0508
	(1.53)	(1.71)	(1.65)	(1.58)
Year 1890	0.0864	0.0893	0.0893	0.0893
	(1.64)	(1.69)	(1.63)	(1.56)
Year 1900	0.0895	0.0924	0.0924	0.0924
	(1.34)	(1.38)	(1.33)	(1.27)
Year 1910	0.1351	0.1380	0.1380	0.1380
	(1.55)	(1.60)	(1.54)	(1.48)
Year 1920	0.1923	0.1952	0.1952	0.1952
	(1.94)	(1.99)	(1.92)	(1.84)
Year 1930	0.3071	0.3100	0.3100	0.3100
	(2.86)	(2.90)	(2.79)	(2.67)
Year 1940	0.3585	0.3614	0.3614	0.3614
	(3.12)	(3.16)	(3.04)	(2.91)
Observations	462	462	462	462
Case FE	21	-	-	-
County FE	-	42	42	42
Year FE	11	11	-	-
Region by Year FE	-	-	44	-
Division by Year FE	-	-	-	77
R-squared	0.849	0.863	0.898	0.921

Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on population density. The log of market access by county in 1870 is used as an additional matching variable. Market access is estimated by Donaldson and Hornbeck (2012). The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects.

Table 1-8: Effects of 1862 Land-Grant Universities on Population Density-with Latitudes and Longitudes Controls
(Dependent variable: log of population density; cluster-robust t-ratios in the parentheses)

	Model 1	Model 2	Model 3	Model 4
Year 1840	0.0348	0.0367	0.0367	0.0367
	(1.35)	(1.01)	(0.97)	(0.93)
Year 1850	0.0118	0.0137	0.0137	0.0137
	(1.00)	(0.69)	(0.67)	(0.64)
Year 1870	0.0543	0.0562	0.0562	0.0562
	(2.02)	(2.17)	(2.08)	(2.00)
Year 1880	0.0443	0.0462	0.0462	0.0462
	(1.35)	(1.50)	(1.44)	(1.38)
Year 1890	0.0846	0.0865	0.0865	0.0865
	(1.55)	(1.60)	(1.53)	(1.47)
Year 1900	0.0885	0.0904	0.0904	0.0904
	(1.29)	(1.32)	(1.27)	(1.22)
Year 1910	0.1385	0.1404	0.1404	0.1404
	(1.56)	(1.61)	(1.55)	(1.48)
Year 1920	0.1992	0.2011	0.2011	0.2011
	(1.99)	(2.04)	(1.96)	(1.88)
Year 1930	0.3144	0.3163	0.3163	0.3163
	(2.94)	(2.98)	(2.87)	(2.75)
Year 1940	0.3645	0.3664	0.3664	0.3664
	(3.23)	(3.27)	(3.15)	(3.02)
Observations	462	462	462	462
Case FE	21	-	-	-
County FE	-	42	42	42
Year FE	11	11	-	-
Region by Year FE	-	-	44	-
Division by Year FE	-	-	-	77
R-squared	0.849	0.864	0.898	0.921

Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on population density. The cubic function in latitude and longitude is used as additional matching variables. Matching on latitudes and longitudes ensures the treated and its synthetic control near each other geographically. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects.

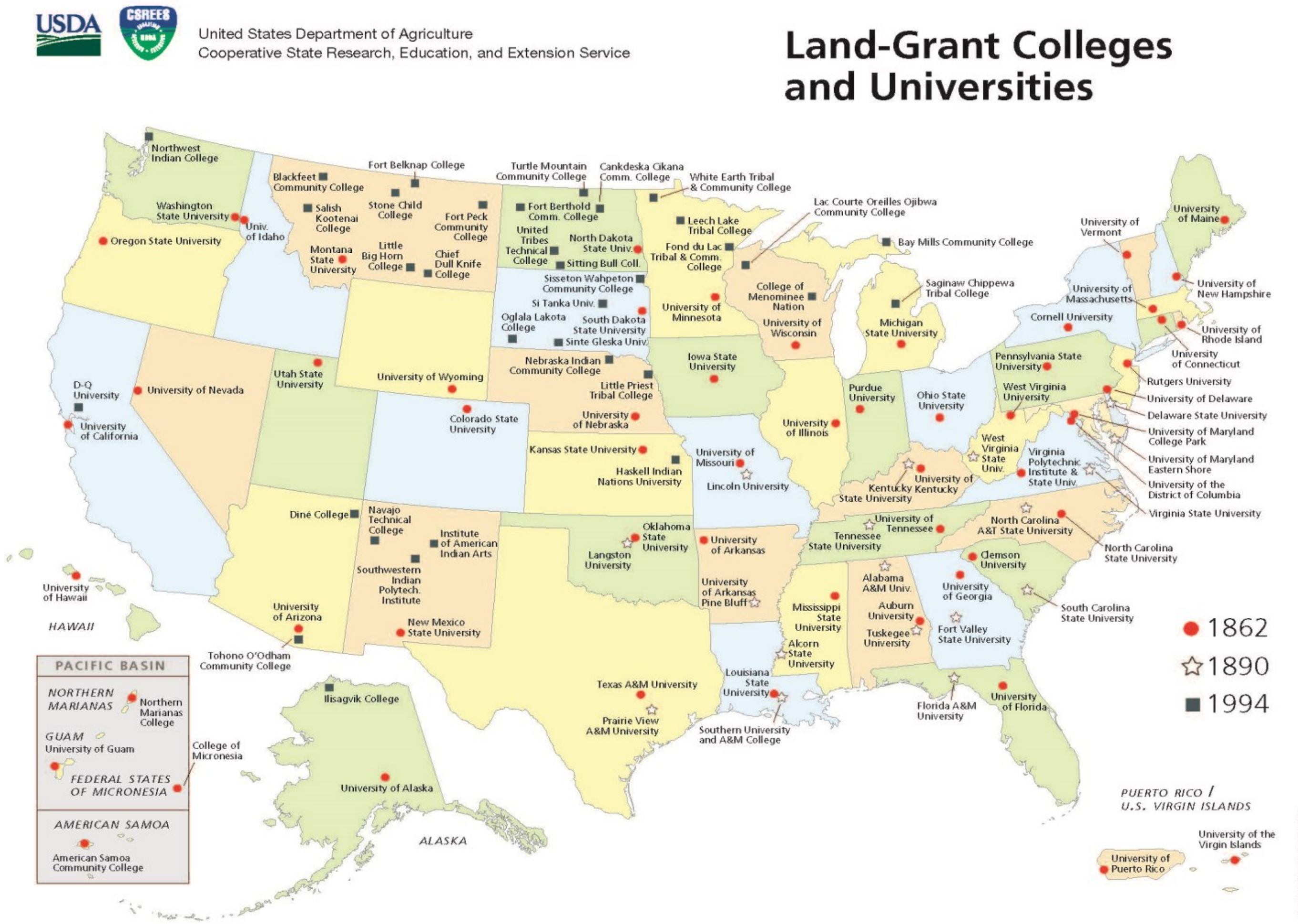
Table 1-9: Effects of 1862 Land-Grant Universities on Manufacturing Output Per Worker-with Latitudes and Longitudes Controls

(Dependent variable: Manufacturing Output Per Worker (in 1840 dollars); cluster-robust t-ratios in the parentheses)

	Model 1	Model 2	Model 3	Model 4
Year 1850	4.419716	0.018493	0.018493	0.018493
	(1.16)	(0.04)	(0.03)	(0.03)
Year 1870	-94.678778	-99.080002	-99.080002	-99.080002
	(-1.37)	(-1.39)	(-1.33)	(-1.27)
Year 1880	-106.835582	-111.236806	-111.236806	-111.236806
	(-0.90)	(-0.93)	(-0.89)	(-0.84)
Year 1890	74.126176	69.724953	69.724953	69.724953
	(0.64)	(0.58)	(0.55)	(0.53)
Year 1900	-3.408306	-7.809530	-7.809530	-7.809530
	(-0.01)	(-0.03)	(-0.03)	(-0.03)
Year 1920	632.533024	628.131800	628.131800	628.131800
	(1.34)	(1.29)	(1.23)	(1.17)
Year 1930	1,231.490670	1,229.378392	1,229.378392	1,229.378392
	(2.02)	(1.96)	(1.88)	(1.79)
Year 1940	2,103.691578	2,125.749849	2,125.749849	2,125.749849
	(1.78)	(1.78)	(1.70)	(1.62)
Observations	332	332	332	332
Case FE	19	-	-	-
County FE	-	38	38	38
Year FE	9	9	-	-
Region by Year FE	-	-	36	-
Division by Year FE	-	-	-	63
R-squared	0.557	0.620	0.695	0.777

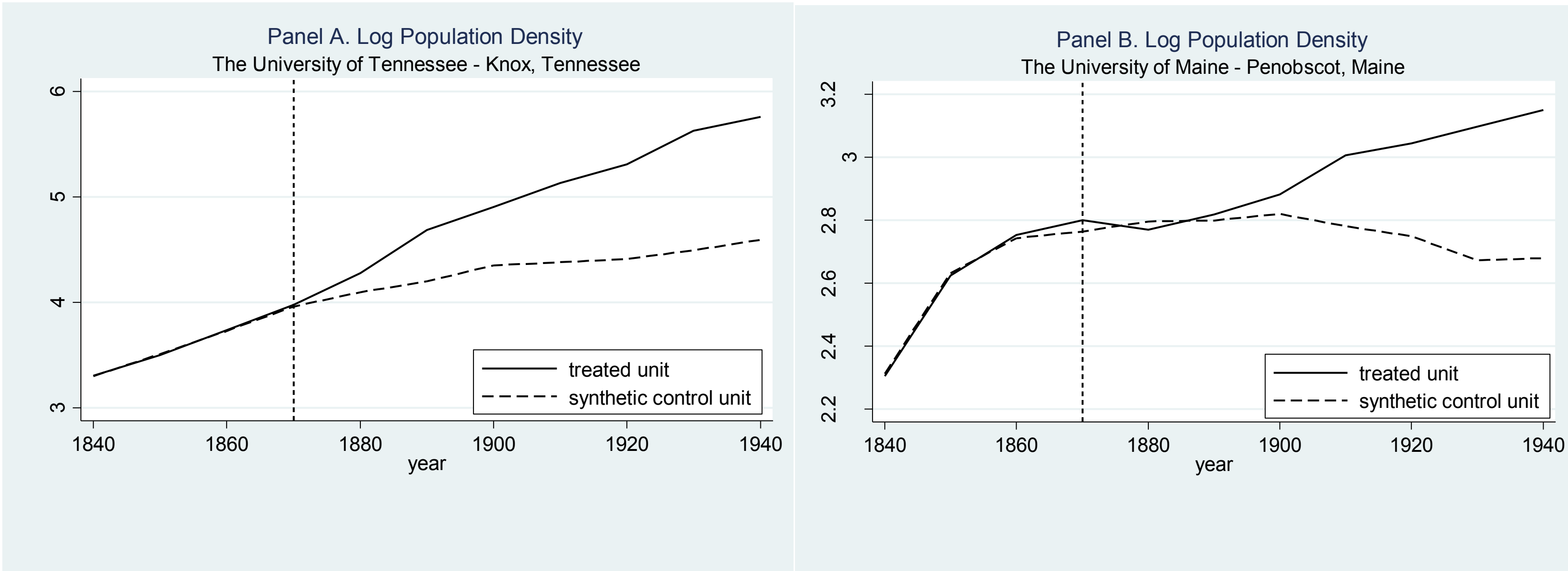
Notes. This table presents the short- and long-run effects of the 1862 land-grant universities on manufacturing output per worker. The cubic function in latitude and longitude is used as additional matching variables. Matching on latitudes and longitudes ensures the treated and its synthetic control near each other geographically. The estimated coefficients are the coefficients of the event-time dummies. T-ratios are based on standard errors clustered at a case level. A case is a pair of a treated county and its corresponding synthetic control county. The coefficient of the 1860 event-time dummy is normalized to 0, so all coefficients after 1870 can be thought of as treatment effects. Data on manufacturing output in 1840 and 1910 is missing.

Figure 1-1. U.S. Land-Grant Colleges and Universities



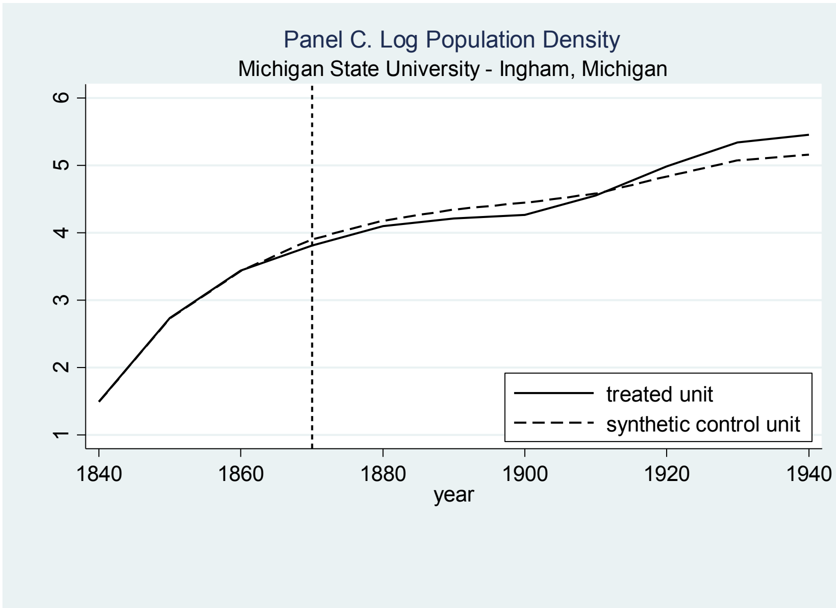
Note. Picture source: Association of Public and Land-grant universities, The Land Grant Tradition, Washington, D.C. 2012

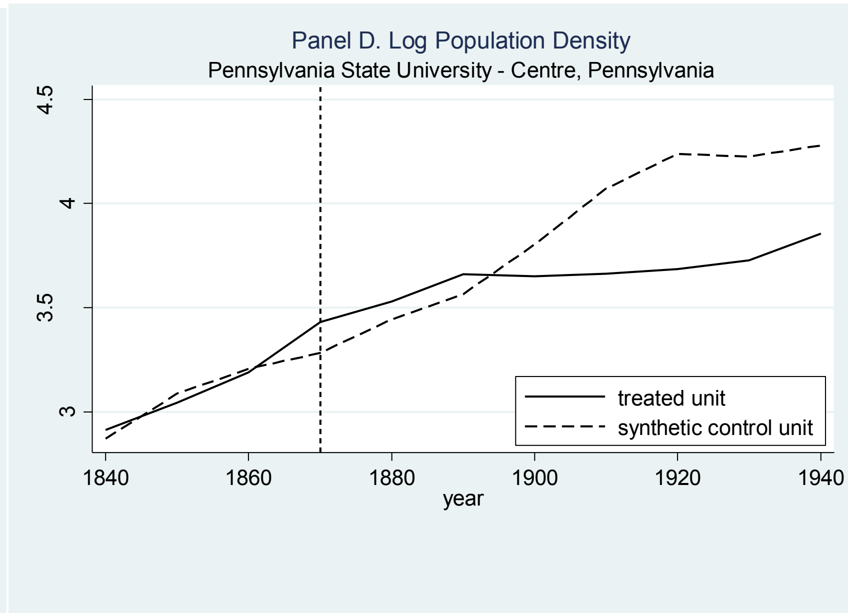
Figure 1-2. Impact of Land-Grant Universities on Population Density



Note. These two panels plot the log population density from 1840 to 1940 for two counties: Knox, Tennessee (Panel A) and Penobscot, Maine (Panel B). The dashed vertical line shows the decade in which the new land-grant university was designated. The solid line displays the observed time series of log population density for the designated county and the dashed line presents the predicted time series of the corresponding synthetic control county. In Panel A, the new land-grant university has an immediate positive impact on population density. In Panel B, the new land-grant university has a lagged positive impact on population density.

Figure 1-2-Continued. Impact of Land-Grant Universities on Population Density

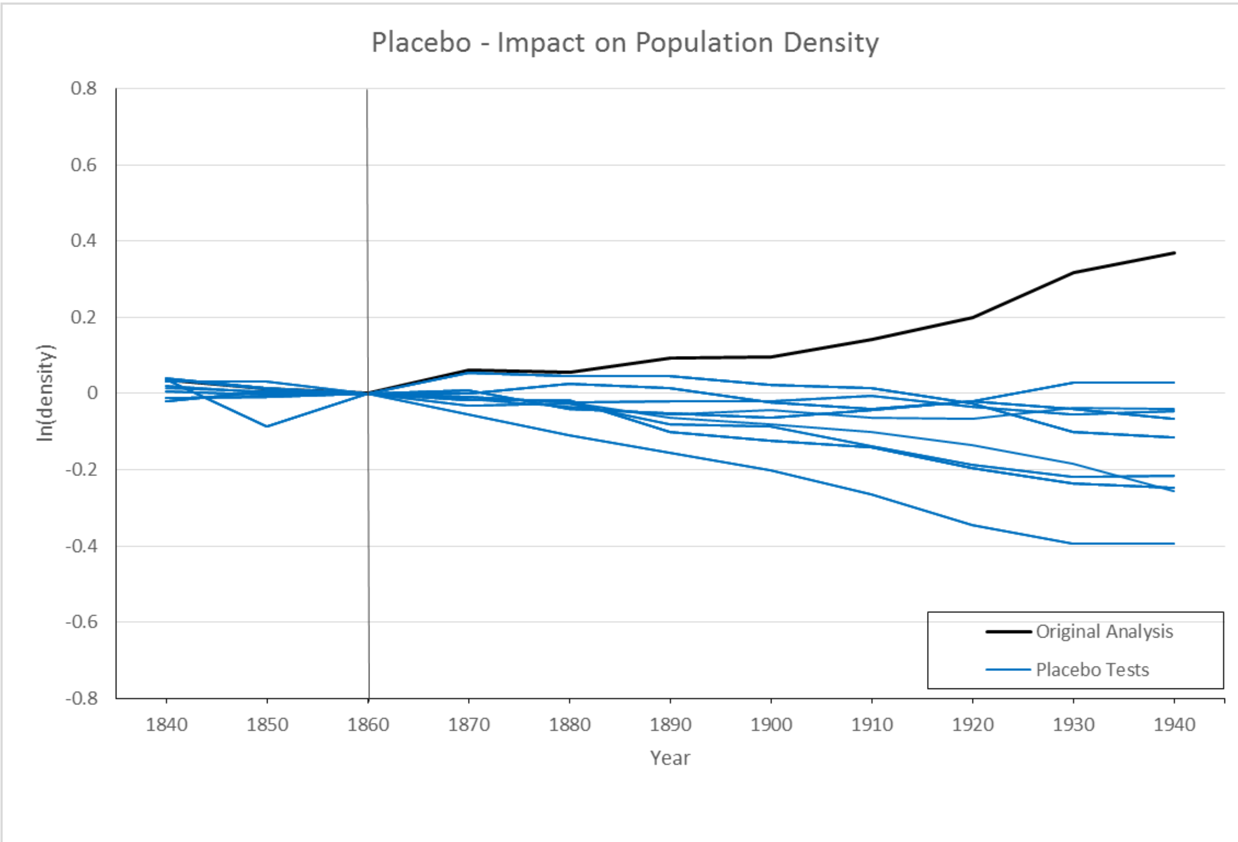




Note. These two panels plot the log population density from 1840 to 1940 for two counties: Ingham, Michigan (Panel C) and Centre, Pennsylvania (Panel D). The dashed vertical line shows the decade in which the new land-grant university was designated. The solid line displays the observed time series of log population density for the designated county and the dashed line presents the predicted time series of the corresponding synthetic control county. In Panel C, the new land-grant university does not appear to have huge impact on population density. In Panel D, the new land-grant university has a slightly positive impact on population density at first, but the trend reverses at later periods.

Figure 1-3. Impact of Land-Grant Colleges and Universities on Population Density

Placebo Tests



Note. This figure plots the estimates of placebo tests with 1862 land-grant colleges and universities. It graphs estimated coefficients of the event-time dummies with actual and artificially assigned treatments. The vertical solid line at 1860 facilitates the comparison of the dynamics before and after the treatment. The thick solid black line displays the actual effects of a new land-grant university. The thin solid blue lines present the effects with artificially treated counties. I run placebo tests 20 times (20 thin solid blue lines).

Appendix

Figure 1-4. Impact of Land-Grant Universities on Population Density

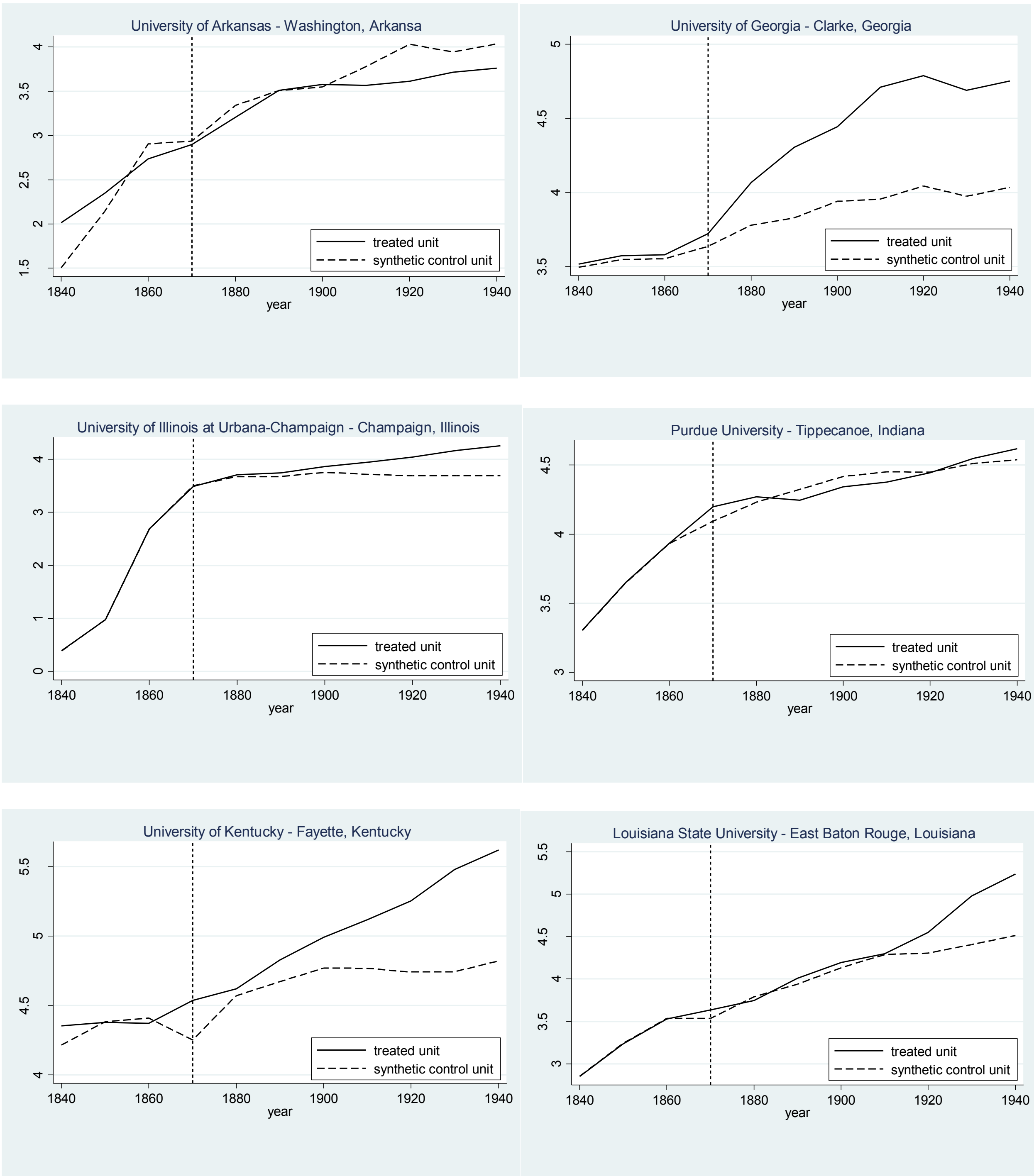


Figure 1-4-Continued. Impact of Land-Grant Universities on Population Density

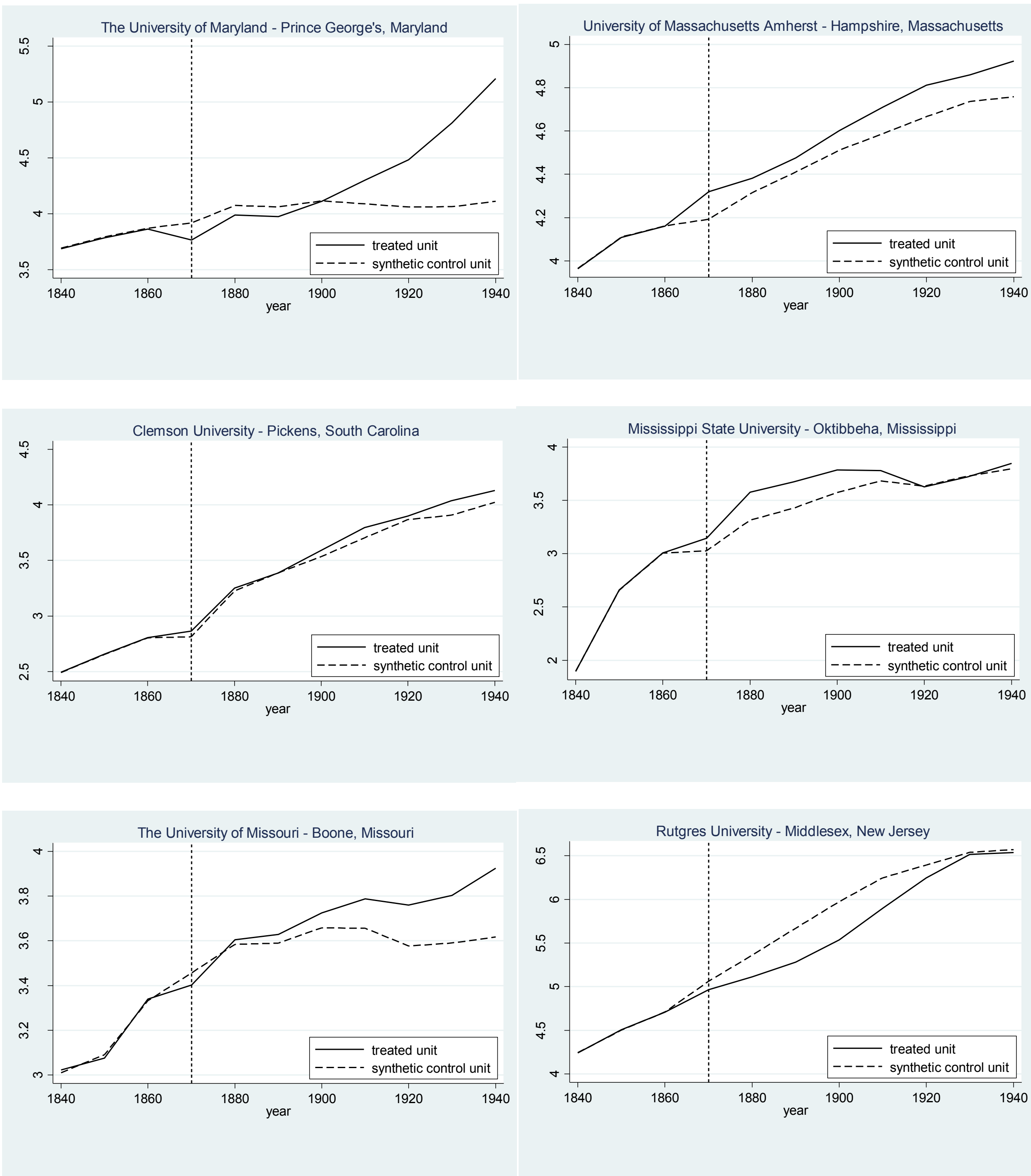


Figure 1-4-Continued. Impact of Land-Grant Universities on Population Density

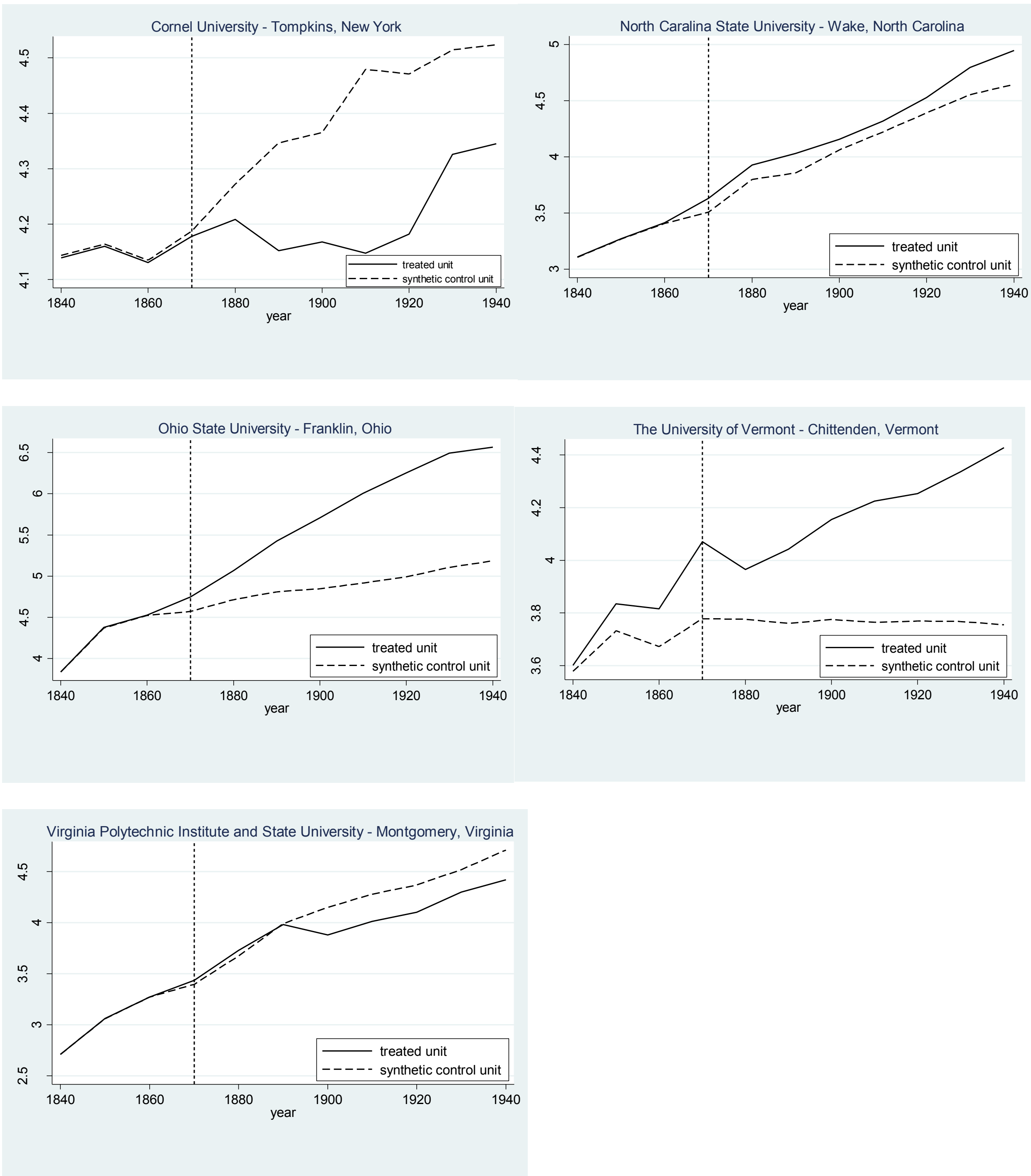


Table 1-10: Population Density Predictor Means

Panel A. Population Density Predictor Means --- Ingham, Michigan

	Ingham, Michigan	Synthetic Control	The State Average
Log(Population density), 1840	1.4936	1.4919	1.8064
Log(Population density), 1850	2.7335	2.7316	2.6464
Log(Population density), 1860	3.4366	3.4340	3.3702
Percent of Manufacturing Workers	0.0108	0.0420	0.0287
Manufacturing Output Per Worker	1579.30	1657.48	1822.13
Per Capita Agricultural Output	39.42	39.43	44.57
Percent of Urban Population	0.0588	0.1251	0.0444
Percent of White Population	0.9982	0.9894	0.9916
Per Capita Farm Value	182.76	165.80	198.35
Per Capita College Students	0	0.0005	0.0006

Panel B. Population Density Predictor Means --- Centre, Pennsylvania

	Centre, Pennsylvania	Synthetic Control	The State Average
Log(Population density), 1840	2.9140	2.8722	3.4223
Log(Population density), 1850	3.0448	3.0891	3.7155
Log(Population density), 1860	3.1898	3.2085	3.9396
Percent of Manufacturing Workers	0.0495	0.0252	0.0460
Manufacturing Output Per Worker	1276.65	1460.35	1473.41
Per Capita Agricultural Output	58.1829	62.0971	52.23
Percent of Urban Population	0	0.0036	0.0736
Percent of White Population	0.9884	0.9956	0.9847
Per Capita Farm Value	274.6277	266.6617	258.99
Per Capita College Students	0	0	.0008

Note. This table shows the mean values of population density predictors for two counties: Ingham, Michigan and Centre, Pennsylvania. All variables except log population density are averaged for the 1840-1860 period. Dollar variables are reported in 1840 dollars. Percent of Manufacturing Workers is percentage of manufacturing workers in whole population. Per Capita Agricultural Output, Per Capita Farm Value and Per Capita College Students are calculated as the total agricultural output, total farm value and total college students in the county divided by county population.

Chapter 2 Agglomeration, Urban Wage Premiums, and College Majors

2.1 Introduction

Productivity and wages are higher in more densely populated areas—cities. This phenomenon has been documented by a vast empirical literature.[31] For example, Glaeser and Mare (2001) document that the unadjusted urban wage premium between dense metropolitan areas and nonmetropolitan places is around 33 percent.[32] One strand of the literature tries to provide evidence on the micro-foundations of urban increasing productivity and returns, also known as agglomeration economies.[33] Studies of this kind include Holmes (1999) on intermediate input sharing, Costa and Kahn (2001) on labor market pooling, and Jaffe, Trajtenberg and Henderson (1993) on knowledge spillovers.

Another strand of the literature explores the pattern of urban wage premium and investigates what types of people play the central role in enhancing cities' productivity. For instance, Rosenthal and Strange (2008a) look at the attenuation pattern of human capital spillovers and conclude that college-educated workers exhibit more productivity spillovers to nearby workers. Bacolod, Blum and Strange (2009) find that cities enhance productivity most for workers with strong cognitive and interpersonal skills. These studies seem to imply that workers with high human capital, cognitive skills and interpersonal skills are more important in making a city more productive. Yet, some other studies have a different point of view. Florida (2002a, 2002b) argues that technology workers, artists, musicians, lesbians and gay men, and bohemians,

---

[31] See Quigley (1998), Rosenthal and Strange (2004) and Head and Mayer (2004) for a comprehensive review of the literature on urban wage premium and agglomeration economies.

[32] Their definition of dense metropolitan areas is Metropolitan Statistical Area (MSA) with a city of over 500,000 people.

[33] Marshall (1890) provides the three best known sources of agglomeration economies: Intermediate input sharing, labor market pooling, and knowledge spillovers.

the so-called "creative class," enhance urban productivity and thus, boost city prosperity.[34] This paper seeks to resolve part of that discrepancy.

The primary goal of this paper is to examine the manner and extent to which worker skill type affects agglomeration economies that contribute to productivity in cities. This paper extends the existing literature in a number of important ways, foremost by using college major to proxy for skill type and studying the heterogeneity in within-field agglomeration economies and across-field spillovers.[35] Central to my analysis is the comparison of within-field and across-field spillovers for workers with different Bachelor's degrees. The results of this comparison help to identify which fields and associated skills are more important in making a city more productive. The across-field spillovers also point to the nature of complementarity between skills. Ellison, Glaeser and Kerr (2010) study the complementarity between industries. However, the complementarity between skills in cities has not yet been studied in the agglomeration literature.

There are two closely related identification challenges that arise when attempting to estimate the causal impact of agglomeration on local wage rates. The first is the standard concern that unobserved individual characteristics may be correlated with indicators of agglomeration causing estimates of the impact of agglomeration on wage rates to be biased. This could arise, for example, if unusually talented workers may endogenously choose to locate in areas with valuable local attributes, including possibly existing concentrations of other talented workers. The second concern is that city-specific unobserved amenities may be correlated with agglomeration measures and wage rates.[36] It is possible that workers in cities with valuable unobserved

[34] Florida (2002a, 2002b) defines bohemians as artists, writers, musicians and people living an unconventional life. This conclusion is viewed as controversial in the literature. See, for example, Glaeser (2005).
[35] Within-field agglomeration economies are defined as how workers' productivity in a given degree field is enhanced by the concentration of workers in the degree field.
[36] Citi-specific unobserved amenities include both amenities that are hard to measure, like community friendliness, and amenities not readily available in the data, like cultural activities.

attributes are compensated by the amenities, and are willing to accept lower wage rates than their counterparts in cities without such amenities.

I respond to the identification challenges in four ways. First, I control for a rich set of individual characteristics in the regressions.[37] This helps to reduce unobserved worker ability in the error term and mitigate bias. Second, each of the regressions includes both state fixed effects and occupation fixed effects. The state fixed effects control for state-specific characteristics, such as weather and geographic locations, and the occupation fixed effects control for occupation-specific characteristics including the type of skills needed to perform a given set of tasks. Inclusion of these fixed effects helps to ensure that any bias arising from endogenous sorting of workers across locations comes from sorting within states and occupations. Third, a differencing strategy is employed when I compare within-field and across-field spillovers for workers in different degree fields.[38] Differencing in this manner helps to further remove the impact of the unobserved attributes that are common across fields.

A fourth and final feature of my modeling strategy is that I instrument for various indicators of agglomeration in a series of generalized method of moments (GMM) wage equations. In these models, two different sets of agglomeration variables are treated as endogenous. The first of these is the MSA population.[39] The second is the field-specific concentration of workers. Two sets of instrumental variables are used to identify these models. I instrument for the MSA population with several geological variables reflecting cities' underlying geology. Geology is an important determinant of settlement pattern. For example, some soils are

[37] The control variables include the worker's education, whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker, and the number of years the worker has been in the United States.
[38] Of course, adding state and occupation fixed effects can also be considered as differencing strategies.
[39] I use Metropolitan Statistical Area (MSA) as a local labor market. When a city is referred to in this paper, it means an MSA.

more fertile and thus, can support a greater density of population. Thus, cities' underlying geology potentially constitutes a strong instrument for city population. The geological features that are used as instruments in this paper include seismic hazard, landslide hazard, and underlying sedimentary rock.[40] For the field-specific concentration of workers in MSAs, I draw on the field-specific concentration of post-secondary faculty in MSAs in 1980 as instruments.[41] More faculty members in a given degree field in the past tend to raise the number of workers in the field today.[42] Meanwhile, the stock of faculty in the past is largely determined by historical and political factors, which are exogenous to contemporaneous labor market conditions.

The empirical analysis reaches several important conclusions. First, I find strong heterogeneity in within-field agglomeration economies for workers with college training in different degree fields. More importantly, workers with college training in information-oriented and technical fields (e.g. STEM areas such as Engineering, Physical Sciences, and Economics) benefit more from proximity to human capital in their own fields, compared to workers in less information-oriented and technical fields (e.g. arts and humanities).[43] For example, the GMM estimates suggest that doubling the number of workers with the same college degrees in the city increases hourly wage rates by 32.9 percent and 22.1 percent, respectively for workers in Health Services and Economics & Business, while increasing hourly wage rates by only 5.9 percent and 4.8 percent, respectively for workers in Social Sciences and History & Arts. The nature of skills

[40] See Combes, Duranton, Gobillon and Roux (2010) and Rosenthal and Strange (2008a) for examples of papers using geological variables as instruments for distribution of population.

[41] The primary data used in the regressions is from 2009. The lagged values of the instruments are used to avoid any simultaneity bias caused by contemporaneous local shocks. A similar instrument is used in Moretti (2004), who instruments for the percentage of college graduates in MSAs with lagged presence of land-grant colleges.

[42] First, more faculty members in a given field in the past educate more students in that field. Second, college graduates are more likely to stay and work in the MSA where they studied.

[43] Information oriented and technical fields include Computer and Information Sciences, Engineering, Engineering Technologies, Biology and Life Sciences, Mathematics and Statistics, Physical Sciences, Medical and Health Sciences and Services, Business and Economics; less information oriented and technical fields include Liberal Arts and Humanities, Psychology, Public Affairs, Policy, and Social Work, Social Sciences (excluding Economics), Fine Arts, and History. See data section for details.

in information-oriented and technical fields seems to allow workers with college training in those fields to benefit more from each other, as opposed to the skills in less information-oriented and technical fields.

Second, I also find strong heterogeneity in across-field spillovers. Workers in information-oriented and technical fields often exhibit economically important spillovers for workers in other fields. On the contrary, the spillover effects from workers in less information-oriented and technical fields are much smaller and often non-existent. For instance, the GMM models suggest that doubling the workers in Computer Sciences & Engineering enhances the productivity of workers in History & Arts and Economics & Business by 19.2 percent and 17.5 percent respectively, while workers in Social Sciences do not appear to enhance the productivity for workers in any other fields. These results seem to suggest that the skills in information-oriented and technical fields are more likely to complement the skills in other fields.

The estimated within-field and across-field spillovers are broadly in line with prior work. Rosenthal and Strange (2008a) suggest that having 100,000 more college-educated workers nearby increases the wage of a college-educated worker by 16 to 24.5 percent, depending on the specification. If we translate these estimates into elasticities, they roughly fall into the range of my estimates.[44] While they suggest that proximity to college-educated workers enhances productivity, my results suggest that not all college-educated workers are alike. The skills associated with workers in information-oriented and technical fields not only enhance the efficiency of workers in the same fields, but also serve as complements to the skills in other fields. Such valuable attributes are generally not found for the skills associated with workers in the arts and humanities. The positive spillovers in cities appear to derive mostly from proximity

[44] They use log-linear models to estimate human capital spillovers. I use log-log models, with coefficients interpreted as elasticities.

to workers with college training in information-oriented and technical fields. My results support the line of research that suggests highly educated workers, such as engineers and scientists, are important in contributing to agglomeration economies in cities, such as Rosenthal and Strange (2008a) and Bacolod, Blum and Strange (2009), while raising doubts on the claim of Florida (2002a, 2002b) that Bohemians are vital to enhance city productivity.

The remainder of the paper is organized as follows. The next section provides a simple theoretical framework to set out the theory of the agglomeration-wage relationship. Section 3 lays out the econometric issues that bear on estimation and empirical identification strategies. Section 4 describes the data sources and variable construction. Results are presented in Section 5 and Section 6 concludes.

2.2 Theoretical Framework

In this section I present a simple framework to articulate the relationship between agglomeration and wages in the presence of unobserved worker characteristics and unobserved city-specific amenities.[45] It also serves to identify why and how OLS estimates are biased. The agglomeration economies in the model can be viewed as either within-field agglomeration economies or across-field spillovers as long as it is a productivity enhancer. I only discuss the case when only one productivity enhancer is in effect for simplicity.

Suppose there are two types of goods, a nationally traded good and a locally traded good—land. Workers maximize utility by choosing the quantities of the two goods they consume, given a budget constraint and the city amenities Z. If workers and firms are perfectly mobile across all cities, equilibrium can only be achieved when workers obtain equal utility in all cities and firms have the same unit cost at all locations. The equilibrium is presented in Figure 1. The

[45] See Roback (1982), Glaeser and Mare (2001), Moretti (2004) and Rosenthal and Strange (2008) for other papers using this model.

upward sloping curve, labeled as $U(W_A,R_A,Z_A) = U^*$, is workers' iso-utility curve in city A. It is the combinations of rent and wage that give workers the same utility level U^*, given the city amenities. It is upward sloping because workers prefer higher wages and lower rents. The downward sloping curve, labeled as $C(W_A,R_A) = C^*$, is firms' unit cost curve in city A. It sets firms' unit cost to a system-wide level C^*. If wage increases, land rents must fall if firms are to remain at the same level of unit cost. I assume firms and workers are facing one land rent for simplicity.[46] Under these conditions, the equilibrium is at the intersection of the iso-utility curve and the unit cost curve. In Figure 1, the equilibrium wage and rent is (W_A, R_A) in city A.

Consider first the equilibrium in city B without the impact of agglomeration economies. If city B is associated with better unobserved amenities, workers are more compensated by the local amenities in city B and thus, are willing to accept lower wages there. Therefore, workers' iso-utility curve shifts inward in city B and the new equilibrium is (W_B', R_B'). Wages in city B are lower for otherwise similar workers than in city A.

With agglomeration economies in effect, firms' unit cost curve shifts out to $C(W_B,R_B) = C^*$ in city B. If the impact of agglomeration economies in city B is large enough, both wages and rents will be higher in city B, as noted as (W_B, R_B). However, the influence of agglomeration on wages may not exactly reflect the benefits of agglomeration. Some of the productivity gains from agglomeration are capitalized into higher rents, reducing the increase in wage that would otherwise occur. Thus, the impact of agglomeration on wages is a lower bound on the

[46] In reality, firms face commercial land rents and workers are concerned with residential land rents. However, because these rents are highly positively related in a spatial equilibrium, this assumption does not affect the model predictions.

productivity gains from agglomeration even though it is an exact measure of the impact of agglomeration on marginal productivity of labor.[47]

An implication of the model is that OLS estimates can be biased in either direction in a regression of wages on indicators of agglomeration. The shift of iso-utility curve caused by unobserved city amenities induces a downward bias of OLS estimates. In Figure 1, the true effect of agglomeration on wages is $W_B - W_B'$, while the observed effect is $W_B - W_A$. An instrument that is uncorrelated with unobserved city amenities generates a consistent estimate of the true effect. OLS estimates are upward biased if the shift of unit cost curve is caused by the fact that workers' unobserved abilities are higher in city B, other than the influence of agglomeration economies in city B. An instrument that is uncorrelated with unobserved worker abilities generates a consistent estimate of the agglomeration effect.

2.3 Empirical Model and Identification

The basic source of identification in this paper consists in the comparison of wages for otherwise similar workers who work in cities with different stocks of workers in different fields. Bearing the theoretical model in mind, I use the following regression equation:

$$Log\ (w_{iz}) = X_i\beta + A_z^1\varphi + A_z^2\omega + A_z^3\tau + d_s + d_o + u_{iz}, \qquad (1)$$

where w_{iz} is the wage of individual i living at location z; X_i is a vector of individual i's characteristics, including gender, race, age, age squared, marital status, education level, presence of children in the household, and years in the United States; A_z^1 is the total population at location z; A_z^2 is the quantity of workers in individual i's own field at location z; A_z^3 is the quantity of

[47] Nominal wages are used in the empirical models. According to Moretti (2004), higher nominal wages in a city imply greater productivity. If workers weren't more productive, firms producing nationally traded goods would leave high-wage cities and relocate to low-wage cities. Although there are firms that produce locally traded goods and services, firms that produce nationally traded goods sell their products at the same price across the nation. Therefore, as long as there are firms producing nationally traded goods in every city, average productivity has to be higher in cities where nominal wages are higher.

workers in other fields at location z; d_s and d_o represent state fixed effects and occupation fixed effects; u_{iz} is an error term that captures all other factors that affect individual i's wage but are not controlled for. This regression is conducted for workers in each degree field.

This specification controls for many personal attributes and unobserved heterogeneity at the state level and the occupation level. However, some remaining factors in the error term could still cause the estimated agglomeration effects to be biased. For example, to see how estimates of within-field agglomeration economies (the coefficient estimate on A_z^2) can be biased, I write the error term u_{iz} in the following form:

$$u_{iz} = \theta_i + v_z + \varepsilon_{iz}, \tag{2}$$

where θ_i is individual i's unobserved ability that is not captured by X_i, state fixed effects and occupation fixed effects; v_z represents unobserved city amenities at location z that will affect worker i's wage; ε_{iz} is a white noise term, which is assumed to be independently and identically distributed over individuals and locations.

As discussed in the theoretical model, the first remaining factor that can confound identification is unobserved individual ability (θ_i). It is possible that workers in cities with more workers in their own fields are more able. Thus, we have $cov(\theta_i, A_z^2) > 0$. This implies an upward bias of the coefficient estimate on A_z^2 along with the fact that ability is positively correlated with wage ($cov(\theta_i, w_{iz}) > 0$).

Another source of bias in the error term is unobserved city amenities (v_z). Cities with better unobserved amenities attract more workers ($cov(v_z, A_z^2) > 0$). Meanwhile, workers are compensated by the amenities and are willing to accept lower wages in those cities ($cov(v_z, w_{iz}) < 0$). This implies a downward bias of the coefficient estimate on A_z^2.

I propose two methods to deal with the problem. The first is a differencing strategy when I compare the estimated coefficients for different degree fields. With this differencing approach, the two remaining sources of bias do not necessarily affect my conclusions. To see this mathematically, I assume $\widehat{\omega}$ is the consistent coefficient estimate on A_z^2. Thus, the consistent estimate of the difference of within-field agglomeration economies for any two fields is $\widehat{\omega}_1 - \widehat{\omega}_2$. Suppose, because of the two remaining sources of bias, the actual estimates I obtained are $\widehat{\omega}_1 + \hat{\pi}_1$ and $\widehat{\omega}_2 + \hat{\pi}_2$, where $\hat{\pi}_1$ and $\hat{\pi}_2$ are the bias terms. Then the comparison result based on the biased estimates is $(\widehat{\omega}_1 - \widehat{\omega}_2) + (\hat{\pi}_1 - \hat{\pi}_2)$. Therefore, if the impact of the unobserved factors is the same across fields ($\hat{\pi}_1 = \hat{\pi}_2$), the bias terms are differenced out and the final result is $\widehat{\omega}_1 - \widehat{\omega}_2$, a consistent estimate of $\omega_1 - \omega_2$. This is to say, the estimation of $\omega_1 - \omega_2$ is less vulnerable than the estimation of the levels of ω_1 and ω_2.

The second method is to instrument for the indicators of agglomeration. Theoretically, with the two sets of instruments I proposed, I not only have consistent estimates of $\omega_1 - \omega_2$, but also have consistent estimates of ω_1 and ω_2 separately.

2.4 Data and Variables

This section describes the data sets and the construction of variables. The primary data set comes from the 2009 American Community Survey (ACS).[48] In 2009, the ACS began collecting information on the field in which individuals received a Bachelor's degree if the person holds a Bachelor's degree. I draw on this variable to conduct my empirical analysis.

[48] Data are drawn from the Integrated Public Use Microdata Series (IPUMS) project at the University of Minnesota Population Center. See http://usa.ipums.org/usa/ for details.

I leave out the degree fields that do not have enough observations and the degree fields in which the type of skills is difficult to determine.[49] This leaves me 14 fields in total.[50] I further combine the fields associated with similar skills. For example, I view the skills associated with Engineering, Engineering Technologies, and Computer and Information Sciences as similar skills. As a result, six categories of degree fields emerge.[51] They are Computer Sciences & Engineering, History & Arts, Natural Sciences, Health Services, Social Sciences and Economics & Business. Computer Sciences & Engineering consists of Computer and Information Sciences, Engineering and Engineering Technologies. History & Arts consists of Liberal Arts and Humanities, Fine Arts and History. Natural Sciences include Biology and Life Sciences, Mathematics and Statistics, and Physical Sciences. Health Services only consists of Medical and Health Sciences and Services. Social Sciences include Public Affairs, Policy, and Social Work, Social Sciences (excluding Economics) and Psychology. Business is combined with the sub-field Economics from Social Sciences.

These categories cover the majority of Bachelor's degree fields. More importantly, the skill types that different fields imply are idiosyncratic. The STEM areas, Computer Sciences & Engineering, Natural Sciences, Health Services and Economics & Business, are often highly information-oriented and technical fields. In contrast, the skills associated with the other two fields, History & Arts and Social Sciences, are less information-oriented and technical. Thus,

[49] For example, there are only 490 workers with a Bachelor's degree in Library Science and 343 workers with a Bachelor's degree in Cosmetology Services and Culinary Arts; the skill type in Interdisciplinary and Multi-Disciplinary Studies is hard to determine.

[50] These degree fields are Computer and Information Sciences, Engineering, Engineering Technologies, Liberal Arts and Humanities, Biology and Life Sciences, Mathematics and Statistics, Physical Sciences, Psychology, Public Affairs, Policy, and Social Work, Social Sciences, Fine Arts, Medical and Health Sciences and Services, Business and History.

[51] In the results section, I still use the term field instead of category. Indeed, those categories are just bigger fields that are composed of several smaller fields.

focusing on these categories facilitates understanding of the role of different skills in achieving cities' high productivity and is of great policy implications.

The empirical analysis of this paper focuses on full-time workers aged 23 to 65.[52] Hourly wage rates are calculated by dividing annual wage incomes by the usual number of hours worked per week and the number of weeks worked during the previous year. The number of weeks worked in the last year is reported in intervals, in which case mean values are used. Individuals' demographic characteristics as well as state fixed effects and occupation fixed effects are included in the regressions. There are up to 48 state fixed effects and around 300 occupation fixed effects with occupation measured at the 3-digit level, depending on the regression sample. The set of individual demographics includes the worker's education level, whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker, and the number of years the worker has been in the United States. Person weights from ACS are used to make the sample nationally representative.

I consider an MSA as a local labor market. MSA is defined as a large population nucleus, together with adjacent communities that have a high degree of economic and social integration with that nucleus. The total population of an MSA is at least 100,000.[53] Census periodically redefines the component units that comprise MSAs, and I resolve this problem by using the method in Jaeger, Loeb, Turner and Bound (1998) when it is necessary.[54] The number of full-time workers in each field in each MSA is then calculated. Table 1 presents the summary statistics of the MSA level employment variables for the six fields used in the empirical analysis.

---

[52] Full time workers are defined as individuals who report that their usual number of hours worked per week was 30 hours or more and that they worked more than 40 weeks for profit, pay, or as an unpaid family worker during the previous year. I also experimented with samples based on different setup of minimum hours worked and minimum weeks worked. The results are robust.

[53] Census 1990 definition.

[54] Professor Lara Shore-Sheppard generously provided the computer code.

MSA population is also tabulated and included in the regressions to control for urbanization effect.

Two sets of instruments are used in the GMM models. I instrument for MSA population with cities' underlying geological features. The geologic data is drawn from the United States Geological Survey (USGS) as boundary files. For example, the geological variation of seismic hazard in Los Angeles is in Figure 2. The proportional average area of each MSA underlain by each geological feature can be calculated by overlaying the boundary files on top of the MSA boundary map.[55] Summary statistics of the geologic variables in MSA level are reported in Appendix A. The field-specific concentration of faculty in 1980 is used to instrument for the field-specific concentration of workers in 2009. The data on the instruments is obtained from the 5 percent sample of the U.S. 1980 Census. Data from the 5 percent sample of the U.S. 1990 Census is also used in some subordinate regressions.

2.5 Results

This section presents the empirical results. Generally, the coefficient estimates on individual demographic characteristics are consistent with the estimates in labor literature and are not reported in the main tables. The complete results for selected regressions are reported in Appendix B.

2.5.1 Urban wage premium and urban amenities

Urban amenities are important when workers make location decisions. Glaeser, Kolko and Saiz (2001) and Tabuchi and Yoshida (2000) suggest urban consumer amenities are the centripetal force attracting workers into cities. In the theoretic model, I treat urban amenities as one important omitted endogenous factor. Then, I propose an IV strategy to address this problem.

[55] Professor Stuart Rosenthal generously provided the geological variation files for census tracts (Rosenthal and Strange, 2008a). I then use a weighted average method to aggregate census tract level variables to the MSA level, in which the size of census tract is used as the weight.

In this section, I provide some empirical evidence suggesting unobserved urban amenities are one important factor in determining the urban wage premium.

The theory of cities' consumer amenities suggests cities are better places for high income people and thus, are more valued by the high income class. Lee (2010) confirms this theory by showing that the consumer amenities in cities are more important for high skill workers. An implication of the theory is that high income workers, with great consumption power and great demand for urban amenities, are more compensated by urban amenities and thus, are willing to accept lower wages in large cities than their rural counterparts. If this is true, it suggests unobserved urban amenities are an important factor in determining the urban wage premium, and urban wage premiums should actually decrease in income for certain high income classes.

Table 2 presents the mean wage and income by degree field and Table 3 shows the estimates of urban wage premiums by degree field. The coefficient estimates in the first row of Table 3 represent the urban wage premium for each field, which is the correlation between city population and wage rates and should not be interpreted as causal effects. These estimated urban wage premiums are largely consistent with the literature.

Table 2 and Table 3 suggest the degree fields with higher income are actually associated with lower urban wage premiums. Workers in Natural Sciences and Computer Sciences & Engineering have the highest average wage and income, but are associated with the lowest urban wage premiums among the six fields. In contrast, workers in History & Arts and Social Sciences rank at the top by the urban wage premiums they obtain, while having the lowest average wage and income. Figure 3 shows urban wage premiums across degree fields against their mean hourly wage rates. It suggests a negative relationship between urban wage premiums and average hourly wage rates. This is consistent with the theory of cities' consumer amenities and implies that unobserved urban amenities are one important factor in determining urban wage premiums. It also motivates the need to instrument for the agglomeration variables.

2.5.2 Within-field agglomeration economies

Table 4 reports OLS estimates of within-field agglomeration economies—how workers' productivity, as measured by wage rates, can be enhanced by the nearby workers in their own fields. Before looking at the coefficients in detail, it is important to recall that OLS estimates can be biased and thus, need to be interpreted with caution.

The first row of Table 4 represents the estimates of within-field agglomeration economies. The first coefficient estimate in the row suggests doubling the workers in Computer Sciences & Engineering in the MSA increases the wage rates of workers in this field by 13.2 percent. Similarly, the wage rates in History & Arts, Natural Sciences, Health Services, Social Sciences, and Economics & Business increase by 10.3 percent, 1.2 percent, 10.9 percent, 6.9 percent, and 18.3 percent, respectively, as the workers in their own fields in the MSA doubles. These estimates are highly significant except for Natural Sciences. An interesting pattern stands out if we rank these fields by their associated within-field agglomeration economies. The top three in the ranking are Economics & Business, Computer Sciences & Engineering and Health Services, all of which are information-oriented and technical fields. In contrast, the bottom three in the ranking are History & Arts, Social Sciences and Natural Sciences, two of which are less information-oriented and technical fields. This suggests within-field agglomeration economies in information-oriented and technical fields are often economically larger than less information-oriented and technical fields. This conclusion is based on comparison of estimates and is more robust than any conclusions based on individual estimates. The small and insignificant within-field agglomeration economies for Natural Sciences are somewhat unexpected. It could be caused by the endogenous factors because the GMM models suggest within-field agglomeration economies for workers in Natural Sciences are substantial.

The second row of Table 4 captures city size effect. With the human capital in the workers' own fields being controlled for, city population captures the overall influence from all other nearby population.[56] The estimates are either negative or insignificant. Rosenthal and Strange (2003, 2005) note this effect as urbanization effect and suggest positive impact on productivity comes mostly from human capital that are most similar—what they noted as localization effect. Thus, my estimates are consistent with the idea that once the localization effect is controlled for, the urbanization effect is likely to be either small or negative.

To address the problem of endogenous agglomeration measures, GMM estimates are obtained in Table 5, as well as a series of instrument diagnostic test statistics.[57] Recent research pays increasing attention to instrument performance, but it is still an evolving science. These test statistics can only be suggestive because of their sensitivity to how the model standard errors are clustered. All the test statistics decrease when I cluster at larger groups compared to clustering at smaller groups or not clustering at all. The reduction in Kleibergen-Paap statistics and first state F-statistics increases the tendency to view the instruments as weak and the reduction in Hansen-J statistics decreases the tendency to view the model as mis-specified. This, of course, casts doubt on the power of these tests.[58] So far, there is no consensus in the literature on how standard errors should be clustered. Rogers (1994) shows clustered standard errors have nice asymptotic properties when the largest cluster is less than 5 percent of the sample.[59] In this paper, I cluster the standard errors at state/occupation level. As will be apparent, this is a safe approach. When I

[56] This impact includes, for example, the influence of pollution and congestion.
[57] The test statistics reported include Hansen-J over-identification test statistics, Kleibergen-Paap rk weak identification test statistics, Kleibergen-Paap rk under-identification test statistics, and first stage F-statistics.
[58] The literature calls for more caution on the power of Hansen-J test than others.
[59] This is equivalent to say the number of clusters should not be too small.

cluster at the state level or the MSA level, the models pass all tests; however, some clusters include more than 5 percent of the sample. When I cluster at the state/occupation level, no cluster includes more than 5 percent of the sample; but a few regressions fail the Hansen-J test. However, it is most important that the results are robust to different clustering strategies.

The GMM estimates in the first row of Table 5 suggest that workers' wage rates in Computer Sciences & Engineering, History & Arts, Natural Sciences, Health Services, Social Sciences, and Economics & Business increase by 10.2 percent, 4.8 percent, 8.1 percent, 32.9 percent, 5.9 percent and 22.1 percent, respectively, when the number of workers in their own fields doubles. All estimates are significant at 10 percent level except for History & Arts. The pattern in OLS estimates reappears if we rank these fields by their associated within-field agglomeration economies: the top four degree fields in the ranking are all information-oriented and technical fields, while the bottom two are the fields related to arts and humanities. Again, workers in information-oriented and technical fields benefit more from proximity to human capital in their own fields, compared to less information-oriented and technical fields. The urbanization effects in the second row of Table 5 are uniformly negative. This is consistent with the literature and suggests that GMM estimates are potentially more reliable than OLS estimates.

The test-statistics in Table 5 imply that the instruments have nice properties. The Kleibergen-Paap test statistics and first-stage F-statistics suggest the instruments are sufficiently strong.[60] In three out of six cases, the Hansen-J statistic is high such that we have to reject the over-identification hypothesis. However, this does not necessarily indicate a model misspecification. First, as mentioned above, all models pass the test when I set different clusters. Second, among my instruments, geological variables are arguably exogenous and thus, are less

[60] There are no available critical values for Kleibergen-Paap weak instrument test when model errors are adjusted for heteroskedasticity and intra-cluster correlation. Thus, the critical values developed by Stock and Yogo (2005) when the errors are i.i.d. are used as benchmarks.

vulnerable. However, Rosenthal and Strange (2008a) use these geological variables to instrument for city population and their models fail to pass Hansen-J test in certain cases. Further evidence on the validity of my instruments will be presented in later sections. It is also encouraging that the results are robust when expanded instrument sets are used in the following section as this should be the case if the instruments are valid.

2.5.3 Across-field Spillovers

Workers' productivity can also be enhanced by across-field spillovers. This type of spillovers has been implicitly studied in the literature. For example, Jacobs (1969) believes the most important knowledge spillovers come from outside the core industry. In this section, I examine across-field spillovers and complementarity between skills directly.

Table 6 reports the OLS estimates. The specification is similar to the specification in Table 4 except now the concentration of workers in each of the six fields is included in the regressions. The table is formatted such that the main diagonal coefficients in the first six rows indicate within-field agglomeration economies and the off diagonal coefficients in the first six rows measure across-field spillovers. The seventh row of the table measures urbanization effects. The GMM estimates are reported in the same manner in Table 7.[61]

Two important patterns emerge. First, the main diagonal coefficients reaffirm the conclusion in the previous section: information-oriented and technical fields are associated with larger within-field agglomeration economies, compared to less information-oriented and technical fields. As I discussed above, the GMM estimates are potentially more reliable than the OLS estimates. Thus, I focus on discussing the GMM estimates in Table 7, although the OLS

[61] I also conducted the regressions in Table 4 to Table 7 separately for male workers and female workers. The general pattern of the results is little changed for each subgroup except the estimates are less significant for the female subsample. I show the gender specific regression results of Table 7 in Appendix C and the gender specific regression results of other tables are available upon request.

estimates in Table 6 reveal a similar pattern. The ranking of within-field agglomeration economies in Table 7 is identical to Table 5. The estimates of within-field agglomeration economies for the two less information-oriented and technical fields, Social Sciences and History & Arts, are small and highly insignificant. This further suggests within-field agglomeration economies in less information-oriented and technical fields are small and often non-existent.

Second, workers in information-oriented and technical fields often generate sizable across-field spillovers for workers in other fields, while across-field spillovers from workers in less information-oriented and technical fields are much smaller and often non-existent. Across-field spillovers are represented by the off diagonal coefficients. I only look at the off diagonal coefficients associated with absolute t-ratios larger than 1.[62] A positive (negative) off diagonal coefficient indicates the workers in the field of the corresponding row have positive (negative) spillover effects for the workers in the field of the corresponding column. In Table 7, the off diagonal coefficients are all positive in the row of information-oriented and technical fields. For example, doubling the workers in Computer Sciences & Engineering enhances the productivity of workers in History & Arts and Economics & Business by 19.2 percent and 17.5 percent, respectively. This implies the human capital and associated skills in STEM areas generate positive spillovers for workers in other fields and are more likely to serve as complements. In contrast, the off diagonal coefficients are mostly negative in the row of less information-oriented and technical fields. It is striking that all six negative off diagonal coefficients are in the row of History & Arts and Social Sciences. This suggests across-field spillovers from workers in arts and humanities are small and often non-existent. While previous research (e.g. Rosenthal and Strange, 2008a) suggests proximity to college-educated workers enhances productivity, my

[62] I consider estimates with absolute t-ratios larger than 1 as informative.

findings suggest not all college-educated workers are alike. Positive spillovers appear to derive mostly from proximity to workers with training in information-oriented and technical fields.

Florida (2002a, 2002b) argues that technology workers, artists, musicians, lesbians and gay men, and bohemians, the so-called "creative class," enhance urban productivity and thus, induce city prosperity. Based on my results, technology workers (workers in information-oriented and technical fields) exhibit large within-field agglomeration economies and sizable spillovers for workers in other fields and thus, can potentially boost city development. However, there is not enough evidence, especially in the GMM estimates, that artists, musicians and bohemians (workers in less information-oriented and technical fields) have similar effects. It is not clear those groups are essential in enhancing productivity in cities.

Instrument diagnostic statistics are reported in Table 7 as well. The general pattern remains unchanged. Fewer models fail Hansen-J test now. Additional regressions are provided to further investigate on the validity of city-field concentration of faculty in 1980 as instruments. This set of instruments is valid if it is orthogonal to city-field specific workers' unobserved attributes and city amenities in 2009.[63] It is hard to test it directly; however, some implications of this identification assumption can be tested.

In particular, I test whether city-field specific growth of faculty from 1980 to 1990 is determined by city attributes and city-field specific labor market conditions in year 1980. If the changes in city-field specific quantity of faculty are largely exogenous to city attributes and especially city-field specific labor market conditions, it conforms to the idea that the development of post-secondary education system are mainly determined by exogenous factors,

[63] This identification assumption suggests it should be the number of faculty in a field in the past affects the number of workers in the field now, not vice versa. The determinants of the lagged number of faculty should be some exogenous factors, such as government policies, and should not be any factors that also affect workers' wage. Considering the lagged value I use and the hundreds of college towns in the United States, where the universities are the main determinants of population and labor force composition, this assumption is likely to be true.

such as historical and political factors. Table 8 reports the regression results of city-field specific growth of faculty from 1980 to 1990 on a set of city attributes and city-field specific labor market conditions in 1980.[64] Although a few coefficient estimates are marginally significant, most coefficient estimates are highly insignificant. Especially, city-field specific growth of faculty from 1980 to 1990 is not correlated with city population, city-field specific medium income and city-field specific number of workers in 1980. It implies that city-field specific growth of faculty between 1980 and 1990 is not systematically determined by any city attributes or city-field specific labor market conditions in 1980 listed. This conforms to my identification assumption that the development of post-secondary education system is mainly exogenously determined.

2.6 Conclusions

This paper is the first paper in the literature that employs worker's field of Bachelor's degree to proxy for skill type and looks at spillover effects within and across skills. The newly included variable indicating workers' college majors in American Community Survey makes this strategy feasible. The heterogeneity of within-field agglomeration economies and across-field spillovers are then examined to study the manner and extent to which worker skill type affects agglomeration economies that contribute to productivity in cities. A number of methods, including differencing and instrumental variables technique, are used to address the endogeneity problem associated with agglomeration measures.

I obtain several important conclusions. First, I find strong heterogeneity in within-field agglomeration economies for workers with college training in different fields. More importantly, workers in information-oriented and technical fields generally benefit more from proximity to

[64] Summary statistics of city attributes and city-field specific labor market conditions in 1980 are presented in Appendix D.

workers in their own fields, compared to workers in less information-oriented and technical fields. Second, I also find strong heterogeneity in spillovers across fields. Workers information-oriented and technical fields tend to generate economically important spillovers for workers in other fields. On the contrary, workers in less information-oriented and technical fields do not seem to have such effects. While previous research suggests proximity to college-educated workers enhances productivity, these findings suggest not all college educated workers are alike. Instead, positive spillover effects appear to derive mostly from proximity to workers with college training in information-oriented and technical fields. Workers' skill types are strikingly important in determining the manner and extent to which workers' productivity is affected by nearby workers. In addition to these main conclusions, I also present suggestive evidence implying that unobserved urban amenities are an important factor in determining urban wage premiums. This supports the theoretical model and motivates the need to use IV strategy to identify my models.

These results are relevant to a broad range of public policy issues. Public policies can be the mechanisms through which society achieves city prosperity. To the extent that this study is successful at finding which groups of people and associated skills are critical in enhancing agglomeration economies in cities, seemly then metropolitan areas should adopt policies aimed at attracting and keeping workers in STEM areas (i.e. workers in information-oriented and technical fields).

This study also serves to comment on some previous studies on city development, such as Richard Florida's "creative class" theory. My estimates suggest technology workers are associated with economically important within-field agglomeration economies and generate sizable spillovers for workers in other fields, while artists and musicians do not seem to have such effects. My conclusions conform to the large body of urban literature while raising doubts on the claim that bohemians play a central role in city development.

References

Acemoglu, D., Angrist, J., 2000. How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws. Working Papers, NBER.

Bacolod, M., Blum, B.S., Strange, W.C., 2009. Skills in the city. Journal of Urban Economics 65(2), 136-153.

Black, D., Kolesnikova, N., Taylor, L., 2009. Earnings Functions When Wages and Prices Vary by Location. Journal of Labor Economics 27(1), 21-47.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters!. Journal of Urban Economics 63(2), 723-742.

Combes, P.-P., Duranton, G., Gobillon, L., Roux, S., 2010. Estimating Agglomeration Economies with History, Geology, and Worker Effects. NBER Chapters, 15-66.

Compton, J., Pollak, R.A., 2007. Why Are Power Couples Increasingly Concentrated in Large Metropolitan Areas?. Journal of Labor Economics 25(3), 475-512.

Costa, D.L., Kahn, M.E., 2000. Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990. Quarterly Journal of Economics 115(4), 1287-1315.

Dewey, J., Montes-Rojas, G., 2009. Inter-city wage differentials and intra-city workplace centralization. Regional Science and Urban Economics 39(5), 602-609.

Donald, S.G., Newey, W.K., 2001. Choosing the Number of Instruments. Econometrica 69(5), 1161-1191.

Ellison, G., Glaeser E.L., Kerr, W.R., 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. American Economic Review 100(3), 1195-1213.

Florida, R., 2002a. Bohemia and Economic Geography. The Journal of Economic Geography 2, 55-71.

Florida, R., 2002b. The Rise of the Creative Class. New York, NY: Basic Books.

Glaeser, E., 2005. Edward L. Glaeser, Review of Richard Florida's The Rise of the Creative Class. Regional Science and Urban Economics 35(5), 593-596.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in Cities. Journal of Political Economy 100(6, Centennial Issue), 1126-1152.

Glaeser, E.L., Maré, D.C., 2001. Cities and Skills. Journal of Labor Economics 19(2), 316-342.

Glaeser, E.L., Kolko, J., Saiz, A., 2001. Consumer city, Journal of Economic Geography 1(1), 27-50.

Gumprecht, B., 2003. The American College Town. Geographical Review 93(1), 51-80.

Halfdanarson, B., Heuermann, D.F., Suedekum, J., 2008. Human Capital Externalities and the Urban Wage Premium: Two Literatures and their Interrelations. IZA Discussion Papers.

Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. Handbook of Regional and Urban Economics 4, 2609-2669.

Holmes, T.J., 1999. Localization of Industry and Vertical Disintegration. Review of Economics and Statistics 81(2), 314-325.

Jacobs, J., 1969. The Economy of Cities. Vintage, New York.

Jaeger, D.A., Loeb, S., Turner, S.E., Bound, J., 1998. Coding Geographic Areas Across Census Years: Creating Consistent Definitions of Metropolitan Areas. Working Papers, NBER.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. Quarterly Journal of Economics 108(3), 577-598.

Kolesár, M., Chetty, R., Friedman, J.N., Glaeser, E.L., Imbens, G.W., 2011. Identification and Inference with Many Invalid Instruments. Working Papers, NBER.

Lee, S., 2010. Ability sorting and consumer city. Journal of Urban Economics 68(1), 20-33.

Marshall, A., 1890. Principles of Economics. Macmillan, London.

Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of Econometrics 121(1-2), 175-212.

Quigley, J.M., 1998. Urban Diversity and Economic Growth. The Journal of Economic Perspectives 12(2), 127-138.

Rauch J.E., 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. Journal of Urban Economics 34(3), 380-400.

Roback, J., 1982. Wages, Rents, and the Quality of Life. Journal of Political Economy 90(6), 1257-1278.

Rogers, W., 1994. Regression standard errors in clustered samples. Stata Technical Bulletin 3(13).

Rosenthal, S.S., Strange, W.C., 2003. Geography, Industrial Organization, and Agglomeration. Center for Policy Research Working Papers 56, Center for Policy Research, Maxwell School, Syracuse University.

Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. Handbook of Regional and Urban Economics 4, 2119-2171.

Rosenthal, S.S., Strange, W.C., 2005. The geography of entrepreneurship in the New York metropolitan area. Economic Policy Review, Federal Reserve Bank of New York, issue Dec, 29-53.

Rosenthal, S.S., Strange, W.C., 2008a. The attenuation of human capital spillovers. Journal of Urban Economics 64(2), 373-389.

Rosenthal, S.S., Strange, W.C., 2008b. Agglomeration and Hours Worked. Review of Economics and Statistics 90(1), 105-118.

Smith, A., 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. W. Strahan and T. Cadell, London.

Stock, J.H., Yogo, M., 2005. Testing for Weak Instruments in Linear IV Regression. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, 80–108.

Tabuchi, T., Yoshida, A., 2000. Separating Urban Agglomeration Economies in Consumption and Production. Journal of Urban Economics, 48(1), 70-84.

Weber, A.F., 1899. The Growth of Cities in the Nineteenth Century. Macmillan, New York.

Wheeler, C.H., 2001. Search, Sorting, and Urban Agglomeration. Journal of Labor Economics 19(4), 879-899.

Table 2-1: Summary Statistics for Employment Variables (MSA level)

	Mean	Std. Dev.	Min.	Max.
Number of Workers with CS & Engineering Degree[a]	14215.18	33781.08	131	307363
Number of Workers with History & Arts Degree	7716.82	22350.93	82	281661
Number of Workers with Natural Sciences Degree[b]	9135.32	20959.22	50	219769
Number of Workers with Health Services Degree	6909.09	14488.45	88	173302
Number of Workers with Social Sciences Degree[c]	12202.06	30648.10	339	351832
Number of Workers with Economics & Business Degree	26354.43	64651.42	1023	739337

[a] CS & Engineering Degree: Computer and Information Sciences degree and Engineering degree.
[b] Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics.
[c] Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

Table 2-2: Summary Statistics for Hourly Wage and Total Personal Income

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Hourly wage[a]	42.95	31.65	44.58	36.34	33.95	38.97
	(28.55)	(29.47)	(38.27)	(23.62)	(29.95)	(34.04)
Total personal income	103229.1	76969.52	111668.7	81378.81	82236.65	96986.95
	(77927.8)	(82945.7)	(106671.1)	(62234.2)	(83527.5)	(95254.7)

Note. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology. Means and standard deviations are reported, with standard deviations in parentheses.

[a] Hourly wage is calculated by dividing worker's annual wage income by the usual number of hours worked per week and the number of weeks worked during the last year.

Table 2-3: Urban Wage Premium Regressions

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Log total population in the MSA	0.0312	0.0588	0.0330	0.0321	0.0561	0.0625
	(8.30)	(12.84)	(7.02)	(7.99)	(17.57)	(20.77)
Observations	40,547	21,416	26,872	19,665	34,512	73,142
State FE	48	48	48	48	48	48
Occupation FE	293	297	280	218	287	316
R-squared	0.308	0.334	0.395	0.315	0.371	0.304
Root MSE	0.517	0.599	.594	0.481	0.547	0.591

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

Table 2-4: OLS Elasticity Regressions

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Log No. of full-time workers in own fields in the MSA	0.1320	0.1038	0.0125	0.1089	0.0691	0.1828
	(14.11)	(5.38)	(0.92)	(3.62)	(4.59)	(11.00)
Log total population in the MSA	-0.1170	-0.0643	0.0211	-0.0780	-0.0234	-0.1577
	(-10.73)	(-2.84)	(1.40)	(-2.57)	(-1.38)	(-7.79)
Observations	40,547	21,416	26,872	19,665	34,512	73,142
State FE	48	48	48	48	48	48
Occupation FE	293	297	280	218	287	316
R-squared	0.327	0.336	0.399	0.327	0.371	0.313
Root MSE	0.510	0.592	0.590	0.481	0.543	0.582

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

Table 2-5: GMM Elasticity Regressions[a]

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Log No. of full-time workers in own fields in the MSA	0.1028	0.0484	0.0810	0.3289	0.0591	0.2212
	(5.54)	(0.98)	(1.76)	(3.55)	(1.69)	(4.24)
Log total population in the MSA	-0.0923	-0.0019	-0.0804	-0.3103	-0.0183	-0.2042
	(-4.08)	(-0.03)	(-1.48)	(-3.19)	(-0.45)	(-3.20)
Hansen-J over ID test statistic[b]	21.060	7.322	5.944	5.035	9.273	44.123
	[0.00]	[0.12]	[0.20]	[0.28]	[0.06]	[0.00]
Kleibergen-Paap rk weak ID F-stat.[b]	677.336	272.861	150.775	118.857	564.282	875.417
	-	-	-	-	-	-
Kleibergen-Paap rk under ID stat.[b]	2069.81	877.82	984.12	528.30	2376.88	3546.72
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1^{st} stage F-stat. on inst. for # of workers[b]	385.13	343.01	345.75	140.99	820.24	748.93
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1^{st} stage F-stat. on inst. for total pop.[b]	301.48	335.76	289.44	138.11	671.44	718.27
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	39,121	20,612	25,673	18,736	33,072	70,117
State FE	47	47	47	47	47	47
Occupation FE	293	296	278	217	286	315
Root MSE	0.507	0.590	0.5867	0.477	0.541	0.581

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] GMM instruments include MSA level measures of seismic hazard, landslide hazard, percent of area underlain by sedimentary rock, and city-field specific number of faculty in year 1980.

[b] Test statistics are cluster-robust; P-values are in square brackets.

Table 2-6: OLS Elasticity Regressions
(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Computer Sciences & Engineering[a]	0.1096	0.0363	0.0103	0.0266	0.0001	0.0115
	(7.10)	(1.38)	(0.45)	(1.00)	(0.01)	(0.76)
History & Arts[a]	0.0037	0.0776	-0.0270	0.0495	-0.0026	0.0287
	(0.21)	(2.72)	(-0.99)	(2.28)	(-0.14)	(1.80)
Natural Sciences[a]	0.0308	-0.0339	0.0215	-0.0252	0.0150	0.0599
	(1.68)	(-1.15)	(0.79)	(-1.03)	(0.72)	(3.69)
Health Services[a]	-0.0421	-0.0228	-0.0328	0.0870	-0.0188	-0.0825
	(-2.23)	(-0.75)	(-1.36)	(2.77)	(-0.96)	(-4.62)
Social Sciences[a]	-0.0901	0.0032	-0.0756	-0.0133	0.0235	-0.0714
	(-4.21)	(0.08)	(-2.24)	(-0.45)	(0.92)	(-3.54)
Economics & Business[a]	0.0949	0.0716	0.1250	0.0287	0.0692	0.1646
	(3.43)	(1.83)	(3.31)	(0.81)	(2.31)	(6.27)
Log total population in the MSA	-0.0974	-0.1051	-0.0002	-0.1384	-0.0501	-0.0830
	(-3.56)	(-2.55)	(-0.01)	(-3.08)	(-1.85)	(-3.75)
Observations	40,547	21,416	26,872	19,665	34,512	73,142
State FE	48	48	48	48	48	48
Occupation FE	293	297	280	218	287	316
R-squared	0.328	0.337	0.400	0.328	0.371	0.313
Root MSE	0.510	0.592	0.589	0.481	0.543	0.582

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] Log number of full-time workers with corresponding Bachelor's degree in the MSA

Table 2-7: GMM Elasticity Regressions[a]

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Computer Sciences & Engineering[b]	0.0932	0.1924	0.0492	-0.0012	-0.0107	0.1752
	(1.81)	(1.46)	(0.50)	(-0.01)	(-0.13)	(2.94)
History & Arts[b]	0.0666	0.0731	-0.2731	-0.2253	-0.1331	0.1076
	(1.02)	(0.69)	(-2.44)	(-1.95)	(-1.49)	(1.32)
Natural Sciences[b]	0.1164	-0.1027	0.2089	0.0485	0.0948	-0.0615
	(1.48)	(-0.64)	(1.58)	(0.40)	(1.00)	(-0.64)
Health Services[b]	0.1603	0.0814	-0.0161	0.5112	0.1413	0.0193
	(1.70)	(0.56)	(-0.13)	(3.46)	(1.32)	(0.21)
Social Sciences[b]	-0.2701	-0.1542	-0.0328	0.1329	-0.0282	-0.3581
	(-3.29)	(-1.34)	(-0.27)	(0.81)	(-0.26)	(-4.46)
Economics & Business[b]	-0.0295	-0.0085	-0.0340	-0.1678	0.0614	0.2782
	(-0.16)	(-0.02)	(-0.16)	(-0.57)	(0.27)	(1.79)
Log total population in the MSA	-0.1132	-0.0469	0.1509	-0.2270	-0.0774	-0.1526
	(-1.36)	(-0.33)	(1.33)	(-1.56)	(-0.79)	(-2.10)
Hansen-J over ID test statistic[c]	3.780	1.313	3.547	4.405	9.092	12.371
	[0.44]	[0.86]	[0.47]	[0.35]	[0.06]	[0.02]
Kleibergen-Paap rk weak ID F-stat.[c]	13.501	5.129	18.534	7.895	12.161	36.000
	-	-	-	-	-	-
Kleibergen-Paap rk under ID stat.[c]	142.29	52.75	181.34	77.75	116.81	383.67
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. CS & Engineering[c]	661.37	743.29	661.98	242.36	1020.90	1214.53
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. History & Arts[c]	520.10	494.17	578.40	189.93	780.31	934.28
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. Natural Sciences[c]	638.85	675.59	668.73	261.41	996.50	1191.01
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. Health Services[c]	505.77	631.11	614.77	176.14	845.24	872.33
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. Social Sciences[c]	569.25	565.39	626.52	217.87	848.50	1005.96
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. Econ. & Business[c]	529.57	586.12	558.42	216.23	858.86	960.03
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
1st stage F-stat. Total pop.[c]	417.45	481.96	480.48	156.91	677.95	751.42
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	39,121	20,612	25,673	18,736	33,072	70,117
State FE	47	47	47	47	47	47
Occupation FE	293	296	278	217	286	315
Root MSE	0.508	0.591	0.589	0.483	0.542	0.583

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] GMM instruments include MSA measures of seismic hazard, landslide hazard, percent of area underlain by sedimentary rock, and city-field specific number of faculty in year 1980.

[b] Log number of full-time workers with corresponding Bachelor's degree in the MSA

[c] Test statistics are cluster-robust; P-values are in square brackets.

Table 2-8: The Effect of MSA Attributes on Growth of Faculty
(Dependent variable: log(No. of Faculty in 1990 / No. of Faculty in 1980) in the field of degree; t-ratios in parentheses)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
% of African Americans in 1980	-2.4570	2.4519	-0.8376	1.8595	2.2268	1.1629
	(-1.83)	(1.74)	(-0.58)	(1.27)	(1.36)	(0.76)
% of Asians in 1980	-2.3257	0.6323	-2.8424	0.2045	-0.2483	-0.9629
	(-0.82)	(0.23)	(-0.92)	(0.07)	(-0.07)	(-0.29)
% of Other races in 1980	-10.7808	3.0754	-19.6021	-11.0933	23.9214	4.9695
	(-0.94)	(0.27)	(-1.55)	(-0.88)	(1.78)	(0.37)
Unemployment rate in 1980	-7.6589	-0.2042	1.0396	3.9177	-0.2513	-0.3688
	(-1.30)	(-0.03)	(0.16)	(0.61)	(-0.03)	(-0.06)
Log total population in 1980	0.1503	-0.4227	-0.2959	0.3139	-0.6586	-0.1615
	(0.36)	(-0.67)	(-0.56)	(0.65)	(-0.88)	(-0.25)
Average age in 1980	-0.0276	0.0181	-0.0325	-0.0177	0.0872	0.0598
	(-0.56)	(0.37)	(-0.59)	(-0.33)	(1.51)	(1.02)
% of workers with college degree in 1980	-2.0128	-4.1816	0.2625	-0.4658	-5.4283	-2.2683
	(-0.57)	(-1.19)	(0.06)	(-0.13)	(-1.25)	(-0.44)
Log MSA-Field specific median	-0.0783	-0.1339	0.3647	0.3465	0.7948	1.7129
income in 1980	(-0.09)	(-0.11)	(0.61)	(0.67)	(0.53)	(1.68)
Log MSA-Field specific number	-0.0773	0.4614	0.2170	-0.2651	0.4714	0.0655
of workers in 1980	(-0.23)	(0.78)	(0.47)	(-0.61)	(0.67)	(0.11)
Observations	226	226	226	226	226	226
R-squared	0.025	0.028	0.021	0.025	0.041	0.027
Root MSE	1.7408	1.7064	1.8923	1.9191	2.033	1.9989

Note. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

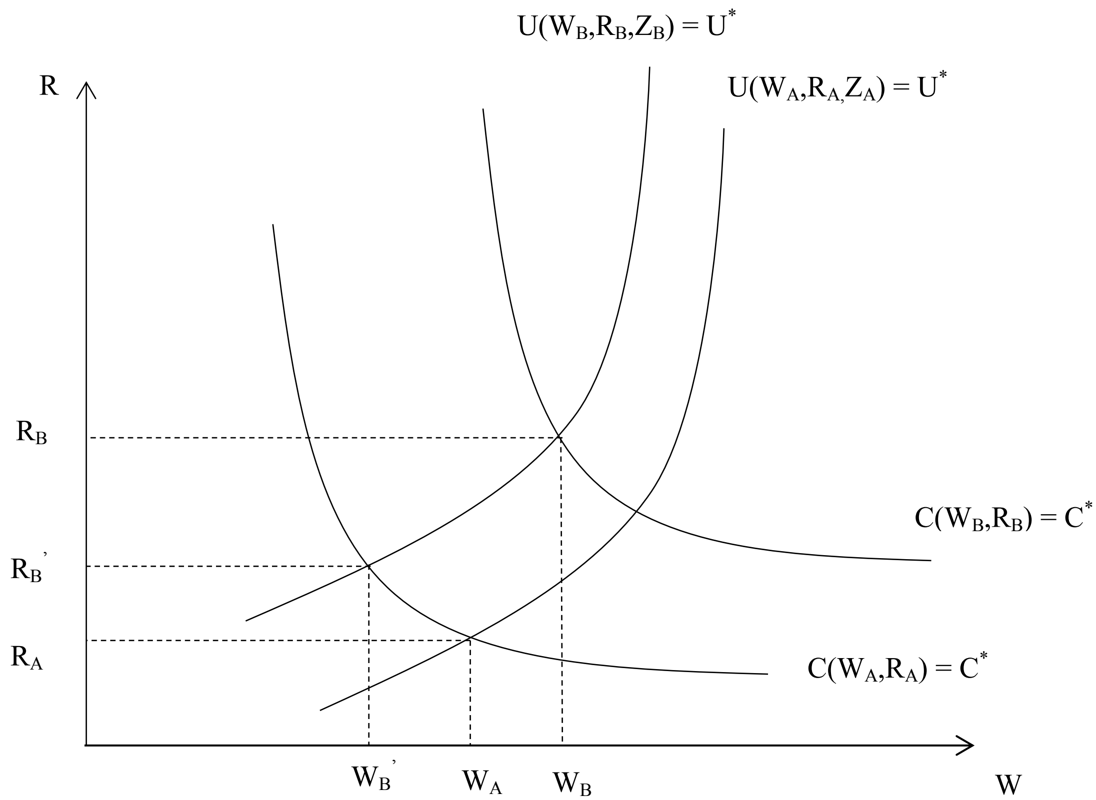


Figure 2-1: Local attributes, Equilibrium wages and Land rents. (W_A, R_A) is the equilibrium in city A. ($W_B{}'$, $R_B{}'$) is the equilibrium in city B without agglomeration economies. (W_B, R_B) is the equilibrium in city B with agglomeration economies.

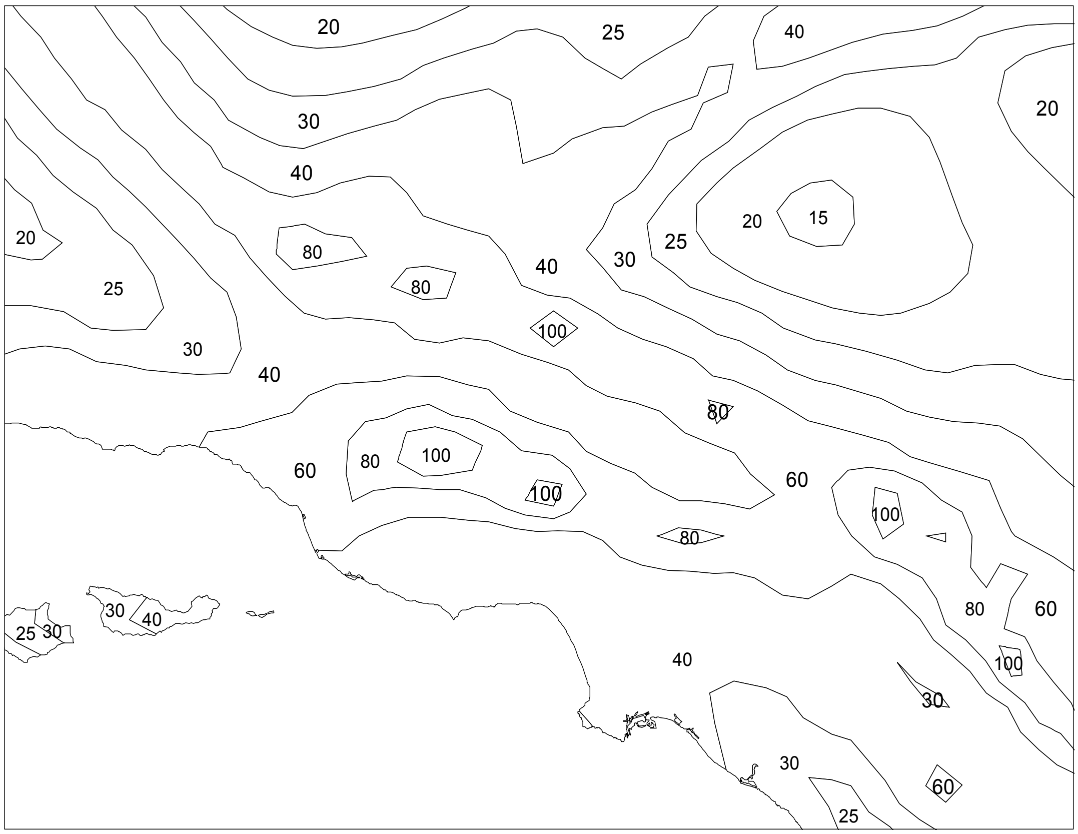


Figure 2-2: Seismic Hazard in Los Angeles
(Scale is from 0 to 100 with 100 as maximum)

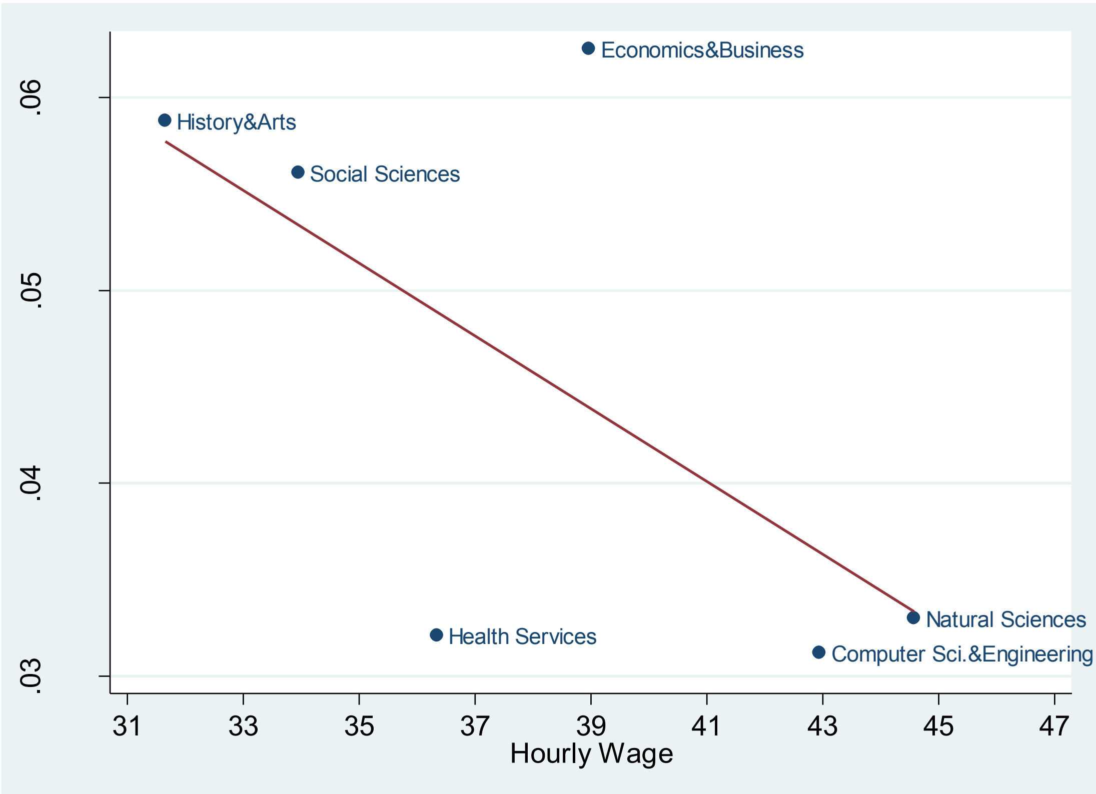


Figure 2-3: Urban Wage Premium and Hourly Wage

Appendix

Table 2-9: Summary Statistics for Instrumental Variables (MSA level)

	Mean	Std. Dev.	Min.	Max.
Number of Faculty Variables				
Number of Faculty in CS & Engineering[a]	107.17	198.77	0.00	2040.00
Number of Faculty in History & Arts	119.91	229.26	0.00	2340.00
Number of Faculty in Natural Sciences[b]	121.33	218.79	0.00	2240.00
Number of Faculty in Health Services	106.28	188.84	0.00	1540.00
Number of Faculty in Social Sciences[c]	43.89	100.96	0.00	900.00
Number of Faculty in Economics & Business	53.81	90.47	0.00	800.00
Geologic Variables				
% of Land Underlain by Sedimentary Rock	0.70	0.37	0.00	1.00
% of Land with Low Landslide Hazard	0.87	0.20	0.06	1.00
% of Land with Medium Landslide Hazard	0.03	0.10	0.00	0.74
% of Land with High Landslide Hazard	0.08	0.17	0.00	0.94
Average Index of Seismic Hazard[d]	5.57	9.18	0.00	56.29

[a] CS & Engineering: Computer and Information Sciences and Engineering.
[b] Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics.
[c] Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.
[d] The scale of the index is from 0 to 100 in the original boundary file downloaded from USGS.

Table 2-10: Selected Complete OLS, 1st and 2nd Stage Regressions
(t-ratios based on standard errors clustered at MSA/occupation level in parentheses)

	Computer Sciences & Engineering				History & Arts			
	OLS Log wage	1st Stage Log # of workers	1st Stage Log Population	GMM Log wage	OLS Log wage	1st Stage Log # of workers	1st Stage Log Population	GMM Log wage
Log No. of full-time workers in	0.1320	-	-	0.1028	0.1038	-	-	0.0484
own field in the MSA	(14.11)	-	-	(5.54)	(5.38)	-	-	(0.98)
Log total population in the MSA	-0.1170	-	-	-0.0923	-0.0643	-	-	-0.0019
	(-10.73)	-	-	(-4.08)	(-2.84)	-	-	(-0.03)
% of sedimentary rock	-	-0.9535	-1.1247	-	-	-1.0668	-1.2234	-
	-	(-11.06)	(-10.89)	-	-	(-13.06)	(-16.36)	-
% of medium landslide hazard	-	0.4505	-0.0615	-	-	-4.7513	-0.7403	-
	-	(0.40)	(-0.05)	-	-	(-5.91)	(-0.96)	-
% of high landslide hazard	-	1.6799	0.0561	-	-	-4.6559	-0.8325	-
	-	(1.67)	(0.04)	-	-	(-6.11)	(-1.11)	-
Ave. index of seismic hazard	-	0.0515	0.0126	-	-	0.0374	0.0273	-
	-	(10.86)	(2.34)	-	-	(9.69)	(8.06)	-
Log # of Faculty in own field	-	0.2701	0.2259	-	-	0.3760	0.3091	-
In the MSA in 1980	-	(37.95)	(37.81)	-	-	(33.37)	(31.51)	-
Male	0.0898	-0.0313	-0.0239	0.0899	0.1046	0.0074	0.0123	0.1045
	(11.11)	(-2.66)	(-2.34)	(11.04)	(10.07)	(0.61)	(1.18)	(9.89)
African American	-0.1374	0.1226	0.1092	-0.1338	-0.0667	0.0951	0.0817	-0.0660
	(-8.95)	(6.31)	(6.47)	(-8.65)	(-3.29)	(4.20)	(4.15)	(-3.22)
Asian	-0.0331	0.1318	0.0886	-0.0335	-0.1085	0.0712	0.0757	-0.1068
	(-3.18)	(9.49)	(6.70)	(-3.19)	(-4.76)	(3.06)	(3.76)	(-4.66)
Other races	-0.1670	0.0602	0.0823	-0.1699	-0.0936	0.0519	0.0511	-0.0903
	(-8.64)	(2.22)	(3.60)	(-8.71)	(-3.51)	(1.85)	(2.03)	(-3.40)
Graduate degree	0.1328	0.0364	0.0049	0.1340	0.1233	0.0056	0.0020	0.1252
	(18.59)	(3.60)	(0.54)	(18.61)	(9.72)	(0.36)	(0.15)	(9.75)
Child under 18	0.0420	-0.0261	-0.0264	0.0396	0.0805	-0.0075	-0.0138	0.0811
	(5.29)	(-2.18)	(-2.38)	(4.93)	(5.44)	(-0.41)	(-0.86)	(5.40)
Married	0.1063	-0.0313	-0.0265	0.1092	0.1086	-0.0324	-0.0274	0.1067
	(13.15)	(-2.79)	(-2.62)	(13.39)	(9.94)	(-2.46)	(-2.38)	(9.61)
Age	0.0652	0.0056	0.0028	0.0666	0.0690	0.0085	0.0073	0.0695
	(25.70)	(1.45)	(0.80)	(25.72)	(18.61)	(2.01)	(2.01)	(18.43)
Age squared	-0.0006	-0.0001	0.0000	-0.0007	-0.0007	-0.0001	-0.0001	-0.0007
	(-21.40)	(-1.56)	(-0.81)	(-21.64)	(-15.44)	(-2.33)	(-2.34)	(-15.31)
No. of years in US 6 to 10	0.0974	0.0112	0.0005	0.0983	-0.0490	0.0419	0.0307	-0.0455
	(5.30)	(0.46)	(0.02)	(5.33)	(-0.65)	(0.54)	(0.51)	(-0.61)
No. of years in US 11 to 15	0.1009	0.0676	0.0422	0.1074	0.0475	0.0187	0.0179	0.0608
	(5.14)	(2.84)	(1.86)	(5.48)	(0.62)	(0.28)	(0.34)	(0.79)
No. of years in US 16 to 20	0.1177	0.0844	0.0791	0.1188	0.0174	0.0445	0.0339	0.0291
	(6.04)	(3.34)	(3.27)	(6.05)	(0.24)	(0.67)	(0.63)	(0.40)
No. of years in US 20 or more	0.1828	0.0026	0.0119	0.1861	0.1064	-0.0340	-0.0233	0.1151
	(11.41)	(0.13)	(0.65)	(11.59)	(1.77)	(-0.55)	(-0.49)	(1.91)
Observations	40,547	39,121	39,121	39,121	21,416	20,612	20,612	20,612
State FE	48	47	47	47	48	47	47	47
Occupation FE	293	294	294	293	297	296	296	296
R-squared	0.327	0.694	0.673	0.134	0.336	0.770	0.756	0.113
Root MSE	0.510	0.753	0.667	0.507	0.592	0.726	0.628	0.590

Table 2-11: GMM Elasticity Regressions for Male[a]

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Computer Sciences & Engineering[b]	0.0875	0.4077	0.0152	0.0864	-0.0350	0.2136
	(1.54)	(2.17)	(0.12)	(0.34)	(-0.25)	(2.56)
History & Arts[b]	0.1164	0.1439	-0.2232	-0.0364	-0.2606	0.1939
	(1.68)	(0.81)	(-1.62)	(-0.14)	(-1.99)	(1.77)
Natural Sciences[b]	0.0876	-0.2726	0.3217	-0.1963	0.2396	-0.1065
	(1.02)	(-1.24)	(1.77)	(-0.70)	(1.53)	(-0.73)
Health Services[b]	0.1341	0.2011	-0.0113	0.2243	0.1914	-0.0296
	(1.26)	(0.99)	(-0.08)	(0.61)	(1.14)	(-0.25)
Social Sciences[b]	-0.2915	-0.0598	-0.0656	-0.1508	0.0708	-0.4944
	(-3.36)	(-0.36)	(-0.44)	(-0.42)	(0.43)	(-4.77)
Economics & Business[b]	0.0274	-0.1705	-0.2040	0.4902	-0.2267	0.4117
	(0.14)	(-0.39)	(-0.82)	(0.81)	(-0.64)	(1.99)
Log total population in the MSA	-0.1510	-0.2502	0.2444	-0.4813	0.1251	-0.2096
	(-1.67)	(-1.24)	(1.74)	(-1.56)	(0.71)	(-2.09)
Observations	32,714	10,298	15,651	3,824	14,363	42,517
State FE	47	47	47	47	47	47
Occupation FE	287	268	267	169	280	313
Root MSE	0.511	0.627	0.620	0.536	0.591	0.626

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] GMM instruments include MSA measures of seismic hazard, landslide hazard, percent of area underlain by sedimentary rock, and city-field specific number of faculty in year 1980.

[b] Log number of full-time workers with corresponding Bachelor's degree in the MSA

Table 2-12: GMM Elasticity Regressions for Female[a]

(Dependent variable: log of individual wage; t-ratios based on standard errors clustered at MSA/occupation level)

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
Computer Sciences & Engineering[b]	0.0934	0.0520	0.0549	0.0374	0.0083	0.1165
	(0.84)	(0.31)	(0.44)	(0.25)	(0.09)	(1.45)
History & Arts[b]	-0.0943	0.0391	-0.4045	-0.2619	-0.0219	0.0575
	(-0.58)	(0.32)	(-2.53)	(-2.04)	(-0.18)	(0.56)
Natural Sciences[b]	0.2657	-0.0289	0.1096	0.0944	-0.0190	-0.0475
	(1.32)	(-0.13)	(0.77)	(0.65)	(-0.18)	(-0.48)
Health Services[b]	0.1689	0.0154	0.0310	0.6084	0.0737	0.0841
	(0.91)	(0.08)	(0.19)	(3.66)	(0.56)	(0.64)
Social Sciences[b]	-0.2861	-0.2183	-0.0075	0.2255	-0.1145	-0.1829
	(-1.40)	(-1.40)	(-0.05)	(1.28)	(-0.89)	(-1.69)
Economics & Business[b]	-0.0994	0.1300	0.2437	-0.4310	0.2876	0.1236
	(-0.23)	(0.26)	(0.83)	(-1.36)	(1.09)	(0.59)
Log total population in the MSA	0.0135	0.0675	0.0044	-0.1663	-0.2113	-0.1135
	(0.07)	(0.36)	(0.03)	(-1.00)	(-2.02)	(-1.21)
Observations	6,407	10,314	10,022	14,912	18,709	27,600
State FE	47	47	47	47	47	47
Occupation FE	206	236	220	189	228	246
Root MSE	0.472	0.537	0.519	0.466	0.495	0.499

Note. Each regression includes additional controls for the worker's education (Bachelor degree, and more than a Bachelor's), whether a child is present in the household, whether the worker is married, age and age squared of the worker, gender of the worker, race of the worker (White, African American, Asian, and other), and the number of years the worker has been in the United States (less than 6 years, 6 to 10 years, 11 to 15 years, 16 to 20 years, 20 years or native citizen) and a constant. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] GMM instruments include MSA measures of seismic hazard, landslide hazard, percent of area underlain by sedimentary rock, and city-field specific number of faculty in year 1980.

[b] Log number of full-time workers with corresponding Bachelor's degree in the MSA

Table 2-13: Summary Statistics for MSA Attributes

Panel I: Variables Not Change by Field of Degree

	Mean	Std. Dev.	Min.	Max.
% of African Americans in 1980	0.11	0.10	0.00	0.45
% of Asians in 1980	0.01	0.04	0.00	0.62
% of Other races in 1980	0.01	0.01	0.00	0.08
Unemployment rate in 1980	0.07	0.02	0.02	0.15
Total population in 1980	667985.00	1388916.00	99660.00	1.46E+07
Average age in 1980	32.93	2.60	25.19	45.71
% of workers with college degree in 1980	0.21	0.05	0.10	0.42

Panel II: Variables Change by Field of Degree[a]

	Computer Sci. &Engineering	History &Arts	Natural Sciences	Health Services	Social Sciences	Economics &Business
MSA-Field specific median income in 1980	23498.96	17241.06	26889.31	23464.78	17389.80	21984.20
	(3569.66)	(2173.15)	(6744.90)	(7025.70)	(2116.20)	(3368.78)
MSA-Field specific number of workers in 1980	3591.50	6471.59	3117.52	2374.78	7388.05	6706.02
	(9214.41)	(17166.21)	(7976.39)	(5735.15)	(19253.11)	(18751.65)
Log(No. of Faculty in 1990 / No. of Faculty in 1980)	0.08	-0.43	-0.27	-0.58	-0.52	-0.67
	(1.73)	(1.70)	(1.87)	(1.90)	(2.03)	(1.99)

Note. Natural Sciences include Biology and Life Sciences, Physical Sciences and Mathematics. Social Sciences include General Social Sciences (excludes Economics), Public Affairs, Policy and Social Work, and Psychology.

[a] Means and Standard Deviations are reported.

VITA

NAME OF AUTHOR: Shimeng Liu

PLACE OF BIRTH: Lichuan, China

DATE OF BIRTH: April 26th, 1988

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENED:

Huazhong University of Science & Technology, Wuhan, China
Syracuse University, Syracuse, New York, USA

DEGREES AWARDED:

Master of Arts in Economics, 2011, Syracuse University
Bachelor of Science in Economics, 2009, Huazhong University of Science & Technology

AWARDS AND HONORS:

AREUEA Doctoral Session Travel Grant, January 2014
Maxwell Summer Fellowship , Summer 2009 – 2013
Roscoe-Martin Graduate Award, Fall 2011
Syracuse University Graduate Assistantship, Fall 2009 – Present

PROFESSIONAL EXPERIMENCE:

Teaching Assistant, Department of Economics, Syracuse University, 2009-2014
Research Assistant, Department of Economics, Syracuse University, 2011-2012