2023

The Impact of Cluster Strength On Wages: An Empirical Analysis

Devan Schaefer

THE IMPACT OF CLUSTER STRENGTH ON WAGES: AN EMPIRICAL ANALYSIS

BY

DEVAN SCHAEFER

A thesis submitted in partial fulfillment of the requirements for the

Masters of Science

Major in Economics

South Dakota State University

2023

THESIS ACCEPTANCE PAGE

Devan Schaefer

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Joseph Santos
Advisor Date

Joseph Santos
Director Date

Nicole Lounsbery, PhD
Director, Graduate School Date

Contents

ABSTRACT

THE IMPACT OF CLUSTER STRENGTH ON WAGES: AN EMPIRICAL ANALYSIS

DEVAN SCHAEFER

2023

In this thesis, I examine the relationship between clusters (i.e., the grouping of competitive, interconnected industries within a geographical area) and wages, building upon the work of Marshall (1890) and Porter (2003) on the importance of clusters for regional economic development. I seek to answer two research questions. First, after accounting for robustness tests, do clusters continue to affect wages positively? Second, is labor force productivity the only channel through which this relationship occurs?

In my analysis, I employ ordinary least squares, two-stage least squares, and fixed effects regression analyses using panel data from 2009 to 2014 for every U.S. county. My main variable of interest in cluster strength, which U.S. Cluster Mapping (2020) defines as the “percentage of trade labor in a strong cluster.” Using my regression model, I find that cluster strength positively and statistically correlates to the average private wage. However, the increase is not as significant as previously documented in the literature.

Furthermore, I perform additional regression analyses on labor force productivity and patents to reveal that, in the short run, there is no correlation between cluster strength and labor force productivity or patents when using a fixed effects regression. Through my conceptual framework, review of the literature, and empirical findings, I suggest that increases in competition between firms, rather than solely labor force productivity, drive the positive relationship between clusters and wages.

1 Introduction

Marshall (1890) introduced the understanding of industrial districts, which would become the foundation for Michael Porter and other economists to construct their understanding of clusters. The term cluster refers to a grouping of competitive, interconnected industries within a geographical area. Researchers seek to understand clusters to exploit the regional economic advantages they ostensibly offer. These advantages include increases in productivity, innovation, exports, and lower inputs costs, among others. Porter (1990) originally set out to explain why certain nations are able to form competitive advantages and constantly innovate. Porter uses the diamond of national advantage model, which comprises four key attributes: factor conditions, demand conditions, related and supporting industries, and firm strategy, structure, and rivalry to explain the discrepancy of economic performance between nations. Specifically, Porter argues that when a nation has all four attributes, it excels in cluster development, thus creating competitive advantages (e.g., increases in specialization and innovation).

Porter (2003) explores another benefit of clusters, namely, wage increases. The author finds that strong clusters—clusters with a location quotient greater than 0.80—and average private wages in economic areas (EAs) positively correlate. Specifically, the author finds that when the share of trade employment (i.e., total regional traded employment divided by total national traded employment) in a strong cluster increases by one, the average private wage for the year 2000 increases by $102.38. While the relationship between clusters and wages provides an additional reason why policymakers should explore policy decisions to promote cluster development, the academic literature has yet to thoroughly explore the effect of clusters on wages. To my knowledge, the only paper that continues to add to Porter (2003) on cluster and wages is Chrisinger et al. (2015), in which the authors seek to answer how clusters impact wages, wage growth, and employment in Washington state.

Porter (2003) argues that wages are higher in strong clusters due to increases in productivity. However, to my knowledge, researchers have yet to explore the notion that productivity is the primary driver of wage increases across clusters. Given this lack of research, I propose two research questions. First, do clusters positively affect wages when introducing robustness tests? Second, are increases in productivity the only driver with which clusters affect wages?

In this thesis, I test two hypotheses that address my research questions. First, controlling for endogeneity, cluster strength is statistically and positively correlated to the average private wage. I am particularly interested in eliminating endogeneity as endogeneity refers to the predictor variable (that is, cluster strength) being correlated to the error term; thus, causing a higher coefficient on cluster strength in my regression. Second, increases in productivity are not the only driver of wage increases that occur in clusters. Furthermore, in this thesis, I propose a channel through which cluster strength affects wages, namely, competition. This is to say, I propose that clusters foster competition among firms; thus, causing firms to increase wages to attract workers.

To test my first hypothesis, I perform an ordinary least squares (OLS) regression to gain a general view of how clusters affect wages. Moreover, I use a more robust two-stage least squares (2SLS) and fixed effects (FE) regression to eliminate any unobservable variable bias. My coefficient of interest in all three regressions is cluster strength, which I collect from U.S. Cluster Mapping (2020), which defines cluster strength as the "percentage of trade labor in a strong cluster." Where trade labor produces goods and services for use outside their region. To construct this variable, U.S. Cluster Mapping (2020) uses a list of 51 U.S. Benchmark Cluster Definitions from Delgado et al. (2016) to identify strong clusters in various geographical areas (e.g., county, state, MSA, and EA), where a strong cluster is any cluster with a location quotient—employment specialization—in the top 25 percent of the U.S. Then, U.S.

Cluster Mapping (2020) divides the employment of trade labor in strong clusters by all trade labor in the given geographical area to create the cluster strength variable.

Furthermore, for my thesis, I define the geographical areas as a county because of their high number on observations and implications for policy decisions. This is to say, given the varying size of clusters, policy decisions at the smallest government level receive the greatest return on investment due to a greater understanding of the local economy rather than, say a state, as an example, adopting clustering policies to impact only three counties. In addition, I use a robust panel data set (data over multiple individuals and time) rather than time series or cross-sectional data, which limit researchers to one individual or one point in time, respectively. Specifically, my data range from 2009 to 2014.

I report that a one percent increase in cluster strength increases the real regional average private wage by \$63.87 when using an OLS regression. Based on more robust FE regression, I report an increase of \$14.42. Furthermore, when using a 2SLS regression, I find no statistically significant correlation between cluster strength and wages.

To test my second hypothesis, I perform additional regressions on labor force productivity and patents to better understand the channels through which cluster strength increases wages. I report that cluster strength statistically correlates with labor force productivity based on an OLS regression, only; the correlation does not remain statistically significant based on an FE regression. Additionally, I find no correlation between cluster strength and the number of patents based on either regression. Therefore, I conclude, in the short run, that increases in competition, rather than productivity, drives wages within clusters.

The remainder of my thesis is structured as follows. In Section 2, I provide a detailed description of the cluster literature. I primarily focus on clusters' relationships to productivity and wages. In Section 3, I explain my conceptual framework for how clusters could positively affect wages; and I propose a model of how competition could drive this effect. In Section 4, I present the empirical model I use to test my hypothesis, provide detailed descriptions of the data, and report my results from my panel regressions. Finally, in Section 5, I summarize my findings and conclude my analysis.

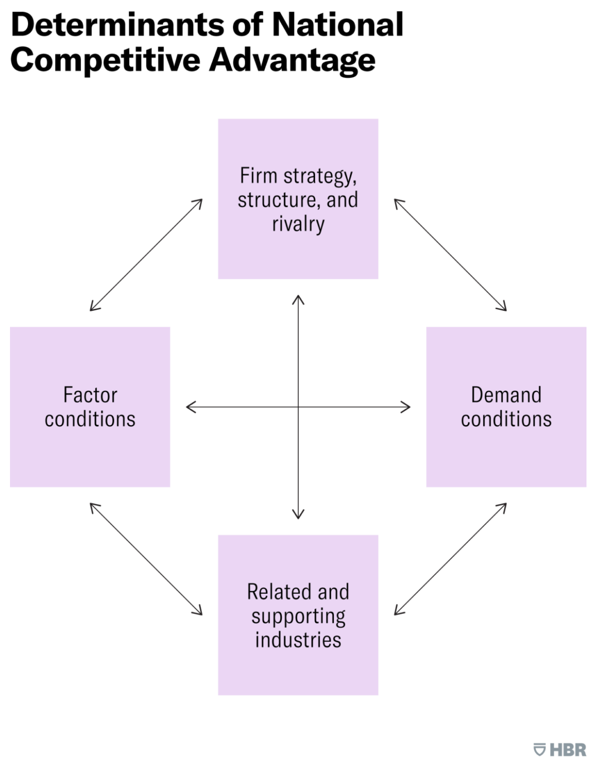
2 Literature Review

2.1 Clusters

Researchers define clusters in different ways. However, the most commonly used definition comes from Michael Porter, who defines clusters by expanding on the topic of specialized industry locations put forth by Marshall (1890) (Porter 2003). A specialized industry location, also referred to as an industrial district, is an observable feature of nations where workers and firms in a location specialize in a certain industry. As understanding on clusters has grown, additional definitions of clusters have emerged.

Porter (1990) introduces the diamond of national advantage, an economic illustration to explain why specific nations excel in particular industries. Specifically, the diamond of national advantage lays out a conceptual framework to understand why certain nations can constantly increase competitive advantages and overcome barriers in innovation. In Figure 1, I show the diamond of national advantage with its four key attributes: factor conditions, demand conditions, related and supporting industries, and firm strategy, structure, and rivalry. Factor conditions are the production variables of a nation (i.e., raw materials and skilled labor). Demand conditions are the understanding of how strong the demand in the home country is for specific products or services; thus, allowing companies to have an earlier understanding of what customers' need. Related and supporting industries are the number of internationally competitive industries in the nation that provide downstream advantages through cost efficiencies and continuous flow of information. Finally, firm strategy, structure, and rivalry is the similarity between firms' performance and management structure in the country, in addition to the strength of local competition that geographic concentration fuels.

Figure 1: The Diamond of National Advantage



Source: Porter (1990)

Porter (1990) hypothesizes that nations with all four attributes of the diamond of national advantage are able to excel in cluster development leading to competitive advantages. These advantages include increases in specialization, innovation, and investments. More specifically, Porter argues when nations have all four attributes, they produce an environment where clusters can flourish through interindustry competition. This competition, as Porter points out, is a way to move away from "static efficiency" to "dynamic improvement". When governments no longer protect and subsidize one company for the sake of economies of scale, they allow domestic rivalry to grow. As companies face domestic rivalry, they compete for everything: market share, top talent, and "bragging rights", which leads companies to innovate and progress. Furthermore, these competitive industries are not haphazard, but rather have observable similarities as Porter points out by stating:

> Competitive industries are not scattered helter-skelter throughout the economy but are usually linked together through vertical (buyer-seller) or horizontal (common customers, technology, channels) relationships. Nor are clusters usually scattered physically; they tend to be concentrated geographically. One competitive industry helps to create another in a mutually reinforcing process. (Porter 1990)

Porter (2000) theorizes how governments should participate in cluster upgrading. Specifically, this cluster upgrading process involves governments eliminating hurdles and restrictions, while making the economy more efficient. These policy decisions could include removing infrastructure constraints or incorporating education policies. However, policymakers should not target specifically one firm or industry, but rather the entire business environment of the cluster to get the highest return on investment. Porter reasons that as governments upgrade clusters, productivity and wages will increase. This is an advantageous decision for the government, whose role involves creating an environment that supports economic growth.

In addition to discussing how governments should upgrade clusters, Porter (2000) expands on competition and cluster's implications in the diamond of national advantage from Porter (1990). Furthermore, Porter develops his definition of clusters, which incorporates two defining features: interconnected industries and geographical proximity. Specifically, Porter states, "A cluster is a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities" (Porter 2000).

Delgado et al. (2014) propose a definition of clusters similar to Porter (2000), again, including the two defining features of clusters: interconnected industries and geographical proximity. The authors define a cluster as "groups of closely related industries co-located in a region". Two years later, in Delgado et al. (2016), the authors propose a new set of U.S. Benchmark Cluster Definitions. U.S. Benchmark Cluster Definitions do not redefine what a cluster is but rather provide a categorization of industries for the purpose of analyzing data and reconciling past findings, like NAICS Sectors (e.g., Finance and Insurance).

To create the set of cluster definitions, Delgado et al. (2016) construct a cluster algorithm that assigns each trade industry—using six-digit NAICS codes—to an exclusive cluster. A trade industry provides goods and services throughout the region and county, excluding industries dependent on location specific natural resources (e.g., mining) (Porter 2003). The authors' cluster algorithm uses a five-step process to assign each trade industry to a unique cluster: (1) capture similarities between industries based on four factors: employment, establishments, inputs & outputs, and occupation; (2) determine parameter choices for the algorithm; (3) create a clustering function, which uses a matrix of similarities between industries and the parameter choices as its inputs; (4) assign validation scores to clusters with the highest relatedness between industries; (5) move any outlier industries to the "next best" cluster.

Through the use of their cluster algorithm, Delgado et al. (2016) create 51 cluster definitions using 778 trade industries.[1] U.S. Cluster Mapping (2020) uses these same 51 cluster definitions to construct my main independent variable of interest for my thesis, cluster strength; a variable that takes a value between 0 and 1, through dividing the sum of trade labor in strong clusters in a geographic area (e.g., MSA, EA, State, or County) by all trade labor in that geographic area. Trade labor comprises labor in trade industries and strong cluster refers to a geographic area whose location

[1] A complete list of each cluster and sub-industries are available at http://clustermapping.us.

quotient—a measure of employment specialization—puts them in the top 25 percent of all geographic areas in the US. In Section 4, I provide a more comprehensive analysis of cluster strength.

Economies in the US and the rest of the world include multiple examples of clusters. Economic researchers analyze these clusters to understand their formation and commonalities. In the US, examples include the Information Technology and Analytical Instruments cluster of Silicon Valley, California and the Automobile cluster of Detroit, Michigan (Klepper 2010; Delgado et al. 2016). Clusters are also prominent within the European Union; examples include the wood-processing and the furniture-manufacturing clusters of Croatia and Slovenia (Stojčić et al. 2019).

One of the most recognized examples of a cluster is the Information Technology and Analytical Instruments cluster of Silicon Valley, located in northern California within the San Francisco Bay Area (Delgado et al. 2016). During its formation in the 1950s, Silicon Valley only comprised about 300,000 people. Over the next 30 years, the population of Silicon Valley increased to 1.3 million people and added roughly 100 semiconductor firms (Klepper 2010). As of 2020, the population of Silicon Valley is over 3 million people, with a GDP of 586 billion dollars (Bureau of Economic Analysis a). Equally, in Detroit, Michigan, 100-plus automobile firms entered the area within 30 years of the start of the industry, and of these, five were industry leaders (Klepper 2010). In 1910, Detroit firms made up 65 percent of the automobile market share and had seven of the top 10 top automobile producers in the US (Klepper 2010).

Silicon Valley, California and Detroit, Michigan are both note-worthy cluster examples as they satisfy both components of the definition of a cluster that Porter (2000) proposes. First, information of consumer demands flows freely and rapidly between firms in the area. This is to say, firms in the area are interconnected. Second, both locations are geographically dense as only three counties make up Silicon Valley: San Mateo, Santa Clara, and Alameda; while, Detroit, Michigan is a single city. It is because of these two components that competition between firms in both areas is fierce and why innovation happens so frequently.

One example of a cluster outside the US is the wood-processing and furniture manufacturing cluster in Croatia and Slovenia. Unlike US clusters that naturally form through firms seeking competitive advantages (bottom-up), European clusters form through organizational financial incentives (top-down). Within Croatia, as of 2018, the wood-processing and furniture manufacturing cluster had 60 members, with 72 percent manufacturing wood and furniture. In Slovakia, the cluster had 94 members, with 79 percent manufacturing wood and furniture (Stojčić et al. 2019). The cluster members that were not involved in manufacturing provided support to the cluster through related services. Additionally, the cluster in both countries have strong interconnectedness between firms and satisfy the condensed geographical location condition (Stojčić et al. 2019).

Several economic advantages seem to accompany clusters. These advantages include increases in innovation, competition, exports, productivity, and employment. Researchers robustly test the causation between the presence of a cluster and these advantages. In doing so, researchers better inform policymakers on the benefits of pursuing cluster upgrading policies.

Researchers demonstrate that clusters increase firm innovation. For example, Huang et al. (2012) examine clustered firms' innovative performance as measured by the number of patents. The authors survey 415 Taiwanese manufacturing firms within the information technology and communication sector. Out of 169 firm responses, the authors use 165 in their regression analysis. The authors find that firms with little research and development capability innovate more from being located within a research park or cluster. Additionally, research has shown a positive relationship between the share of traded employment in a cluster and the number of patents (Porter 2003).

In addition to innovation, research shows that clusters increase competition between firms (Porter 1990, 2000, 2003; Delgado et al. 2014, 2016). This increase is caused by clusters attracting additional firms seeking the same advantages as their competitors (e.g., decreased input costs). Once multiple firms integrate into the same supporting industries, they seek out additional advantages to improve, further increasing competition. This competition creates or enhances all other cluster economic advantages, refuting the notion that competition is "wasteful" (Porter 2000).

Furthermore, research has shown that clusters increase exports. Becchetti et al. (2007) study 25,000 Italian corporations designated as limited liability firms or included in Italy's textile and machinery and equipment industry. The authors separate these firms by inclusion in the industrial district (ID). These IDs compose of both small- and medium-sized firms. The authors find that firms in the ID, while smaller, have a higher number of exports per worker than firms outside of the ID. Additionally, the authors find that firms in the ID have a higher value-added per employee than those outside the ID. Other research has shown that firms in a cluster have a greater chance of becoming an exporter than do non-clustered firms (Stojčić et al. 2019).

Moreover, research has shown that clusters decrease unemployment. Lambert et al. (2017) use data from the Indiana Business Research Center, Purdue Center for Regional Development, BLS, and the US Office of Management and Budget to determine the effect of clusters on the unemployment rate. The authors elect to use data from 2011, the year with the highest number of clusters.[2] The authors use a double-log model to empirically test the effect of clusters on unemployment; they find that clusters are negatively related to the unemployment rate in U.S. metro areas. Specifically, the authors find that a one percent increase in a cluster's establishments and employment causes a 0.65 and 0.17 percent decrease in the unemployment rate, respectively.

[2]The authors only select one year as the Purdue data set does not use the same clusters every year. Therefore, the authors cannot look at the same cluster over time.

Abdesslem and Chiappini (2019) report that clusters increase the productivity of firms. The authors examine how clusters affect productivity in the French optic/photonic industry; they find that total factor productivity (TFP) is 12 percent higher in clustered firms, while labor productivity is 11 percent higher. Stojčić et al. (2019) report a similar result when researching the wood manufacturing industry in Croatia and Slovenia for the period 2013 to 2016. Specifically, the authors find that firms in a cluster have two to three percent higher productivity than those in the control group.

While clusters have multiple economic advantages, they also have disadvantages. These disadvantages predominantly occur because of challenges in establishing a cluster. First, typical US clusters form through a highly innovative and successful first-mover firm entering an area. Examples of first-mover firms include Olds Motor Works in the Automotive cluster of Detroit, Michigan and Fairchild in the Information Technology and Analytical Instruments cluster of Silicon Valley, California (Klepper 2010; Delgado et al. 2016). A first-mover firm pulls in competitors who seek the same advantages, creating a cycle of cluster growth. However, this natural process (i.e., not incentivized by the government) is dependent on both the success of the first-mover firm and competitors moving to replicate success.

Likewise, the establishment of a successful first-mover firm and moving competitors takes time (Porter 1990). This delay makes clusters unattractive to policymakers who rely on short-term economic growth to strengthen reelection chances. Without buy-in from policymakers, the formation and growth of clusters diminish as hurdles and restrictions remain in place (Porter 2000). Finally, latecomer clusters may not be competitive. Barkley and Henry (1997) provide, as an example, Myrtle Beach attempting to become a country music cluster like Branson, Missouri. Myrtle Beach has invested millions of dollars in its country music theatre to attempt to grow the cluster. However, Myrtle Beach has struggled as a latecomer and has not been able to come close to the cluster size of Branson.

2.2 Productivity

As I mention earlier, in Subsection 2.1, research has shown clusters increase firm and region productivity. Researchers use multiple metrics and models to calculate this increase. One of these metrics, put forth by Robert Solow, is total factor productivity. Total factor productivity, A, captures the growth in output, Y (e.g., GDP), that occurs outside the growth of traditional input measures, labor, L, and capital, K. By regressing total factor productivity on clustered and non-clustered firms, researchers determine whether clusters increase productivity (Cingano and Schivardi 2004; Martin et al. 2011; Howard et al. 2014; Abdesslem and Chiappini 2019). In Equation 1, I show the standard equation involving total factor productivity to calculate output in Cobb-Douglas form, where α is the share of capital, and β is the share of labor.

$$Y = A \times K^{\alpha} \times L^{\beta} \tag{1}$$

In Equation 2, I rewrite Equation 1 to solve for total factor productivity.

$$A = \frac{Y}{K^{\alpha} \times L^{\beta}} \tag{2}$$

Another metric researchers use is Data Envelopment Analysis, which calculates the efficiency of decision-making units. Charnes et al. (1978) put forth the standard equation for Data Envelopment Analysis, which I show in Equation 3, where y_r and x_i are the known s outputs (e.g., number of goods) and m inputs (e.g., materials and hours of labor) of the decision-making unit; u_r and v_i are variable weights for each output and input, respectively; and j is the efficiency of one decision-making unit. Solving Equation 3 yields a result less than or equal to one, where one represents a highly efficient decision-making unit. Like total factor productivity, Data Envelopment Analysis provides an explanation for why certain decision-making units have greater output compared to others with the same inputs. Researchers use this result to determine whether clusters affect the efficiency of decision-making units (Kim et al. 2009).

$$max\ h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \tag{3}$$

which is subject to the constraints:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1;\ v_r, v_i \geq 0;\ r = 1, \ldots, s;\ i =, \ldots, m.$$

Researchers also use the count of patents to measure productivity as increases in patenting signifies product and production process advancements (Porter 2003). Researchers elect to use patents when studying the impact of clusters on productivity because patents can track geographical location (Jaffe et al. 1993; Porter 2003; Moretti 2021). Specifically, this benefit allows researchers to easily examine clusters effect on productivity for multiple geographical sizes (e.g., MSA, EA, State, County).

Economic researchers study how clusters affect productivity in multiple contexts. For example, Kim et al. (2009) seek to answer two economic research questions. First, how do clusters enhance the productivity of the biotech industry in the US? Second, what variables cause biotech clusters to form in their location? The authors use Data Development Analysis to determine a decision-making unit's efficiency in the biotech industry. Furthermore, the authors use Directed Acyclic Graphs—an illustration of causation using chosen variables—to establish causality between biotech firms and clusters. The authors find that firms usually form biotech clusters to increase market competition. Additionally, the authors' research shows that clusters positively affect biotech firms' efficiency.

Howard et al. (2014) address whether knowledge and productivity spillovers from clustering occur in Vietnamese firms. The authors study data from 2002 to 2007 from the Vietnamese Enterprise Survey. The data include registered manufacturing enterprises with more than 30 employees and a random sample of 15 percent of manufacturing companies with fewer than 30 employees. However, for the analysis, the authors drop the small firms from the data. The authors use an Olley and Pakes (1996) approach which addresses two sources of biases that occur when using an OLS regression on a production function: simultaneity bias and selection bias. Additionally, the authors add to this approach to allow them to account for firms' selecting their own locations. The authors find a robust productivity spillover within clusters in Vietnam. Further, the authors find that foreign-owned firms benefit the most through this productivity spillover, while domestic firms benefit to a lesser extent.

Lin and Sai (2022) analyze the agglomeration of mining firms in African countries' overall effect on energy productivity. The authors' research fills a gap in the literature as other researchers overlook the African mining cluster entirely or do not provide an empirical analysis. The authors use data from 2009 to 2017 to study of 21 African countries' mining sectors. The authors use a Shephard energy distance function—a combination of a distance function, and the relation of two points and a production function, the relation of physical inputs and outputs—to answer their economic research question. The authors represent the mining industry in 21 African countries using the model's decision-making units. Furthermore, the authors calculate a location quotient which measures the industrial agglomeration of each country's mining sector. Using both a panel regression and threshold regression, the authors find that industrial mining agglomeration positively affects energy productivity.

Cingano and Schivardi (2004) use the growth rate of total factor productivity to measure changes in productivity from sectoral specialization rather than the growth in employment. The authors examine Italian local labor system firms, which the National Statistical Institute separates into self-contained labor markets. The local labor system includes 784 labor markets; of those, the authors use 539. The authors merge employment data from the National Statistical Institute and capital stock data from the Company Accounts Data Service. The authors use this data set to construct total factor productivity for the period 1986 to 1998.

The authors regress the average total factor productivity growth rate on productive variety (i.e., the products the city produces outside the sector), competition, firm size, specialization, city size, human capital at the city level, years of schooling for working age population in 1981, initial city-sector total factor productivity, and geographical and sector dummy variables. Additionally, the authors perform multiple robustness tests, including adding more spatial and selection controls and including all information by adding all city sectors that only had a few years of data. The authors report that increasing sectoral employment causes a 0.4 percent growth in total factor productivity when the concentration index shifts from the 25th to the 75th percentile.

Moretti (2021) uses over four million patents to measure the relationship between inventors' productivity and cluster size.[3] The number of inventors in the data set with non-missing information and an assigned employer is 834,375 over the years of observation. The author uses economic areas for geographical indicators from the Bureau of Economic Analysis. The author defines 895 clusters by multiplying 179 economic areas by five different patent research fields from 1971 to 2007.

To measure the relationship between investor productivity and cluster size, the author uses OLS to regress the natural log of patents (as specified by year, inventor, firm, research field, technology class, and city) on the size of the cluster, city X field,

[3]Private firm patents make up 90.9 percent while universities, government, and nonprofits make up the remaining 9.9 percent.

city X class, field X year, technology class X year, city X year, investor effect, and firm effect. Additionally, the author uses a 2SLS and IV regression for robustness tests. The author's instrumental variable is the firm's spatial network (i.e., the geographical structure of firms with more than one location). The author assumes that firms that employ inventors outside of the central inventor's firm but have a presence in the same location will employ more inventors as the cluster increases. However, shocks in the central inventor's productivity will not affect these outside firms. On balance, the author concludes that there is a positive correlation between inventor productivity and cluster size.

As I mention in Subsection 2.1, Stojčić et al. (2019) use the average treatment effect to determine how clusters impact the productivity of wood-processing and furniture manufacturing firms in Slovenia and Croatia. The authors' data set comprises of 652 and 666 firms in Slovenia and Croatia, respectively, from 2013 to 2016. Of these firms, 39 and 62 are cluster members in Croatia and Slovenia, respectively. The authors use three treatment estimation techniques to remove any hidden selection bias. These methods include inverse probability weighted regression adjustments, nearest neighbor matching procedure, and propensity score matching. The authors regress sales revenue and number of employees on a cluster dummy variable, unit labor cost, unit material cost, subsidies share, market concentration, urbanization economies, and localization economies. The authors showcase each technique's results and a placebo estimation for an extra robustness check. The authors report that clusters positively affect sales revenue and number of employees.

While most economic researchers find a positive correlation between clusters and productivity, a few do not. Martin et al. (2011) find a negative correlation between clusters and productivity when examining a French policy in 1998. The policy sought to fund collaboration projects between firms in the same industry and location by providing an average subsidy of roughly 37,500 Euros. Roughly one hundred projects were funded before the policy was abandoned around 2005. However, I hypothesize that this opposing result stems from failures in the policy rather than failures in clusters. Specifically, this policy incentivizes firms to collaborate, rather than compete. For clusters to increase productivity, competition must be present to push firms to innovate (Porter 1990, 2000, 2003; Delgado et al. 2014). While collaboration between firms is important, it is competition that creates "dynamic improvement" in the cluster (Porter 1990).

2.3 Wages

While multiple economic researchers establish the correlation between clusters and productivity, few investigate the correlation between clusters and wages. Porter (2003) explores multiple research questions regarding clusters. Through using a more innovative and comprehensive data set of 41 clusters within the US, the author aims to better understand regional performance, regional composition, and the role of clusters on performance. The author's core data set includes annual County Business Patterns data from 1990 to 2000. In the data set, the author prefers to use economic areas (EAs) instead of metropolitan statistical areas (MSAs) as EAs cover the whole US and have a more comprehensive historical data set. One theory the author tests is the positive relationship between cluster strength and average wages. The author finds a positive and primarily significant relationship between the share of traded employment in strong clusters and average wages through this analysis. Specifically, the authors reports that a one percent increase in cluster strength increases the average wage by $102.38. However, the author's research has two limitations: no control variables (e.g., education, innovation, population, and productivity) and only one year of data. In Subsection 4.3, I discuss key differences between Porter (2003) and my research as we both examine how clusters affect the average wage.

Fowler and Kleit (2014) examine the relationship between industrial clusters and the poverty rate. Furthermore, the authors aim to understand the effect clusters have on counties or individuals rather than on an entire region. The authors use a Maximum Likelihood Estimation linear regression and a spatial regression model, using census data for the year 2000 to answer their economic research question. The authors regress the percentage of individuals in the county below the poverty rate on demographic, geographic, economic, and cluster variables to answer their economic question. The authors report that industrial clusters have a finite effect on poverty rates. Within the authors' analysis, they find that clusters could either have a positive or negative correlation with the poverty rate—depending on the cluster type. For example, the Textiles and Apparel cluster is linked to a higher poverty rate while the Chemical-based Products cluster is linked to a lover poverty rate.

Chrisinger et al. (2015) test how clusters impact wages, wage growth, and employment in Washington state. The authors add to previous research by providing a comprehensive overview of how clusters affect employment measures. Instead of electing to examine only one employment measure (e.g., wages, wage growth, and employment), the authors interpret how clusters impact workers by looking at multiple variables simultaneously. The authors use wage and employment information from the Washington State's Employment Security Department, trade information from the Bureau of Economic Analysis, and cluster information from the County Business Patterns data from the US Census Bureau. The authors' data set comprises 19 million records from 2003 to 2010. Using the data set, the authors use a methodology from Feser and Isserman (2005) to determine clusters in the data using North America Industry Classification System (NAICS) codes. Through this methodology, the authors find 27 clusters. The authors report that employees in clusters have, on average, higher wages. However, the authors also report that non-clusters have slightly higher wage growth than clusters.

Although some researchers examine the affect clusters have on wages, the method through which clusters affect wages has not been tested. In Porter (2003), the author puts forth the theory that his result of a positive relationship between cluster strength and the average wage stems from increases of productivity in clusters. While Porter does not test this theory, that is, that clusters increase wages through increases in productivity, other researchers do test how productivity affects wages. For example, Judzik and Sala (2013) examine how productivity, de-unionization, and international trade affect real wage growth. The authors use a panel of seven Organization for Economic Co-operation and Development (OECD) countries from 1980 to 2010. The authors use an OLS and IV econometric models to report that productivity growth positively affects real wage growth. Furthermore, if productivity growth is set to zero from 1980 to 2010, real wages decline by a range of eight to 27 percent, depending on the country. Specifically, the United States would have had a nearly eight percent decline in real wages from 1980 to 2010.

Montuenga-Gómez et al. (2007) study the effect of productivity on wages in the Spanish industrial sector through analyzing different wage-setting mechanisms between Spanish sectors. The authors examine 14 different sectors in Spain from 1964 to 1992. Using data from Garcia et al. (1994) and the Industrial Survey, the authors sort their data into two distinct periods. Subsample one, which represents the period of dictatorship in Spain from 1964 to 1997, and subsample two, which represents a period of democracy in Spain from 1998 to 1992. The authors use an OLS econometric model to regress real wages on the productivity of each sector. The authors find in industries with slower productivity growth, productivity is the primary variable that affects the real wage. However, sectors with faster productivity growth set their real wages based on the alternative wage, that is, the wage that a worker could expect to earn in any other sector.

Feldstein (2008) examines whether productivity growth affect wages in the US.

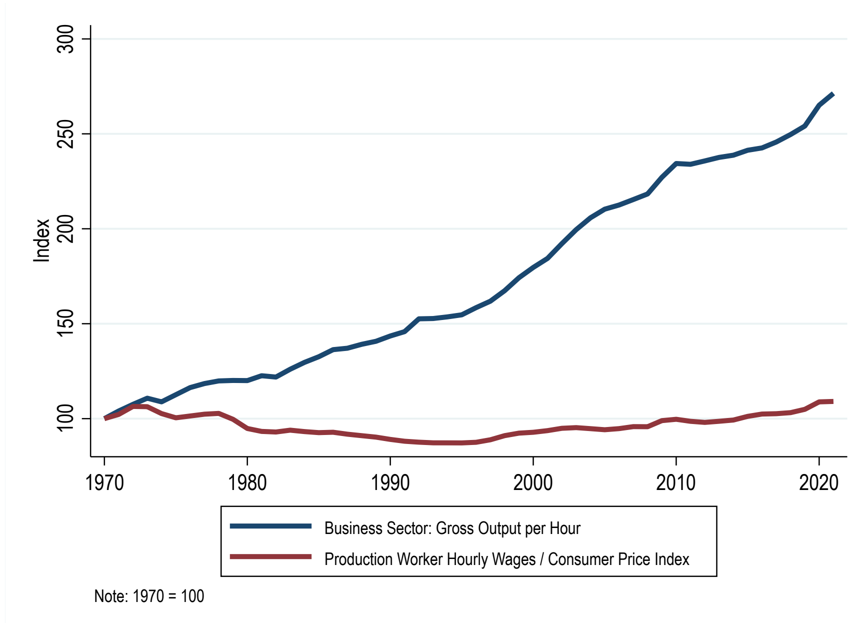
Specifically, the author refutes research that finds that productivity has grown disproportionally to wages through misleading calculations. These misleading calculations include researchers using wages and salaries instead of measures of total compensation and the Consumer Price Index as a deflator instead of deflating by the product price. Using the "correct" calculations, the author regresses wage on productivity for the period 1947 to 2006. The author finds a one percent increase in productivity results in a 0.94 percent increase in compensation when using the "correct" calculations and a two-year lag to account for the change in log productivity and log nominal wages.

Similar to Feldstein (2008), Strain (2019) provides an intuitive example of the relationship between wages and productivity. Strain's example is that if a worker can only produce \$15 per hour of revenue, it would not make sense for the employer to pay her more. Conversely, the worker should not accept a wage less than \$15 per hour. This relationship between the productivity (output per hour) of the worker and their wage (dollars per hour), as the author describes, is the "marginal revenue product." However, as Feldstein (2008) mentions, some researchers do not find this same result, i.e., that workers' wages follow productivity.

Lawrence (2016) examines a figure comparing wage and worker productivity growth that wage linkage critics commonly show. In Figure 2, I recreate this figure and expand its period to 2021, keeping the base year the same. As I show in Figure 2, worker productivity has grown tremendously throughout the years, while worker wages are stagnant. At first glance, one could conclude that wages have not kept up with worker productivity, leading the intuitive thought exercise by Strain (2019) to be incorrect. However, as Strain (2019) points out, Lawrence (2016) makes several adjustments more suitable to show the relationship between workers' wages and productivity.

First, Lawrence corrects the output metric. As Strain (2019) mentions, Lawrence uses the total output from the economy rather than exclusively the business sector. Second, he uses net output rather than gross output, which does not include de-

Figure 2: Hourly Wages and Output per Hour, 1970-2021



Source: Lawrence (2016) and St.Louis Federal Reserve (FRED), Series: OPHPBS, AHETPI, & CPIAUCSL

preciation in its calculation. Third, Lawrence includes part-time employees instead of only examining full-time employees. Fourth, Lawrence adjusts the wage variable. This adjustment includes using compensation instead of only wages. This change allows for a better representation as non-wage compensation is 17.93 percent of all compensation (Bureau of Economic Analysis b,c).[4] Additionally, Lawrence allows for professional workers and deflates wages using an output price index. In Figure 3, I replicate the changes made by Lawrence (2016) using the personal consumption expenditure (PCE) price index as the price deflator for net domestic product.[5]

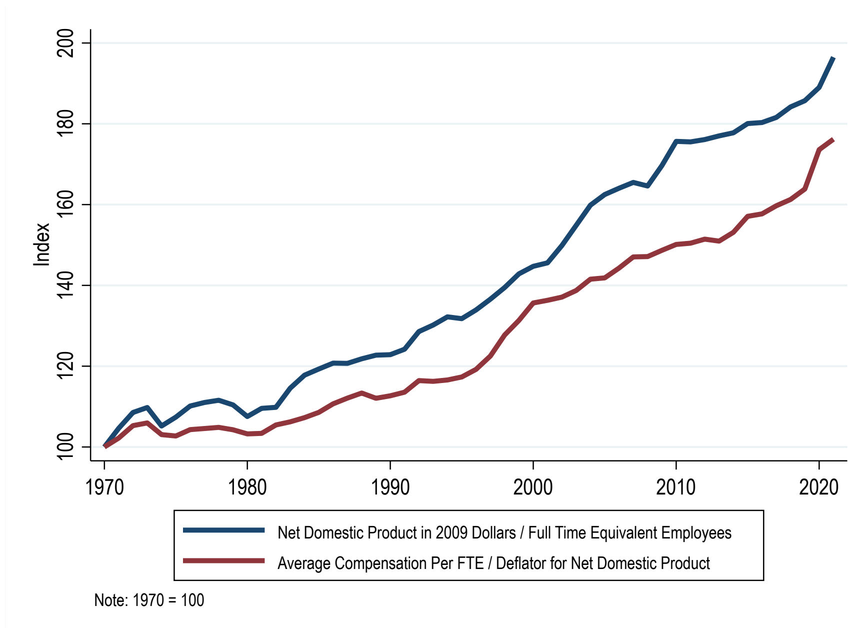
As I show in Figure 3, the correlation between productivity and labor income remains strong even after the slight divergence of 2001 (Strain 2019). Stansbury and Summers (2017) also find a strong correlation between worker compensation and

[4]Feldstein (2008) provides an overview of the benefits of using compensation rather than solely just wage as a measurement between income and productivity.

[5]I choose to use PCE rather than Lawrence's deflator based on its conservative estimate and reliability.

productivity. The authors examine the relationship between productivity growth and compensation growth for production/nonsupervisory workers over the period 1973 to 2015. The authors find that a one percent increase in productivity growth, on average, increases production/nonsupervisory real compensation growth by 40 to 70 basis points.[6] Furthermore, the authors suggest that some other factor might have caused this divergence in productivity and pay (see **Figure 2**). The authors test whether technological progress could have caused this divergence but do not find strong evidence of the relationship.

Figure 3: Net Domestic Product and Real Product Compensation per Full-Time Equivalent Employee



Source: Lawrence (2016) and St.Louis Federal Reserve (FRED), Series: PCEPI, A362RC1A027NBEA, A4301C0A173NBEA, & A4401C0A052NBEA

[6]Stansbury and Summers (2017) highlight the period 1973 over 1948 as this is when the divergence between productivity and compensation occurred.

3 Conceptual Framework

Through my review of the academic literature on clusters, I hypothesize that clusters will positively affect wages. Before expanding on the data and model I will use to test my two hypotheses empirically, I explain the concepts that support and disagree with my hypotheses. I further explain the main control variables I include in my empirical model and how they could affect wages. Likewise, I also discuss the channels through which this relationship occurs to address my second hypothesis that wage increases come from other channels than solely increases in labor force productivity, namely competition.

3.1 Supporting Theory for Hypotheses

There are four ways that I anticipate clusters could increase wages: increasing labor force productivity, raising competition, boosting innovation, and decreasing input costs. These four channels, as I discuss in Section 2, are found to be benefits of clusters. In this subsection, I discuss how each variable has been shown to or could increase wages.

3.1.1 Labor Force Productivity

In Porter (2003), the author examines the relationship between strong clusters and wages, stating, "The proportion of strong clusters in the economy should be positively related to productivity and hence average wages." This relationship, that Porter mentions, can be shown algebraically. To demonstrate this, I follow Nicholson (1995) derivation of the Lagrangian expression associated with a cost minimization problem. In Equation 4, I include the original expression from Nicholson (1995), where K is capital; L is labor; $f(K, L)$ is the production function of the firm; q_0 is the level of output; v is the per-unit hiring cost of capital; and w is the per-unit hiring cost of labor.

$$\mathcal{L} = vK + wL + \lambda[q_0 - f(K, L)] \tag{4}$$

Next, assuming that the firm's input choices do not affect the price of inputs, that is v and w, I solve for the first-order condition for a minimum in Equation 5 through 7.

$$\frac{\partial \mathcal{L}}{\partial K} = v - \lambda \frac{\partial f}{\partial K} = 0 \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial L} = w - \lambda \frac{\partial f}{\partial L} = 0 \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = q_0 - f(K, L) = 0 \tag{7}$$

Where, focusing only on labor, Equation 6 can be written as Equation 8.

$$\lambda \frac{\partial f}{\partial L} = \lambda MP_L = w \tag{8}$$

However, as Nicholson (1995) points out, the Lagrangian multiplier, λ, can be seen as the marginal cost, MC, as it captures the change in objective total costs for a one-unit change in the constraint output, that is, output minus q_0. Given this interpretation, I construct Equation 9.

$$MC \times MP_L = w \tag{9}$$

Next, incorporating output choices and following the general rule of marginal cost, MC, equals marginal revenue, MR, for profit maximization, I construct Equation 10.

$$MR \times MP_L = w \tag{10}$$

Furthermore, assuming the firm is a price taker for its outputs; thus, indicating that marginal revenue is the same as the market price, I rewrite Equation 10 as Equation 11.

$$P \times MP_L = w \tag{11}$$

Next, assuming that firm utilizes a Cobb-Douglas production function, I rewrite Equation 11 as Equation 12, where P is the market price; $(1-\alpha)$ is the labor share; and Y is output.

$$P \times (1-\alpha)\frac{Y}{L} = w \tag{12}$$

Thus, since $(1-\alpha)$ and P are a constant, w is proportional to $\frac{Y}{L}$; thus, w and $\frac{Y}{L}$ grow at the same rate as I show in Equation 13. Put differently, as labor force productivity increases, so too should the wage at an equal rate.

$$\frac{Y}{L} \propto w \tag{13}$$

In summary, given this theoretical thought experiment and past research, it stands to reason that clusters could increase the wage through advancing labor force productivity.

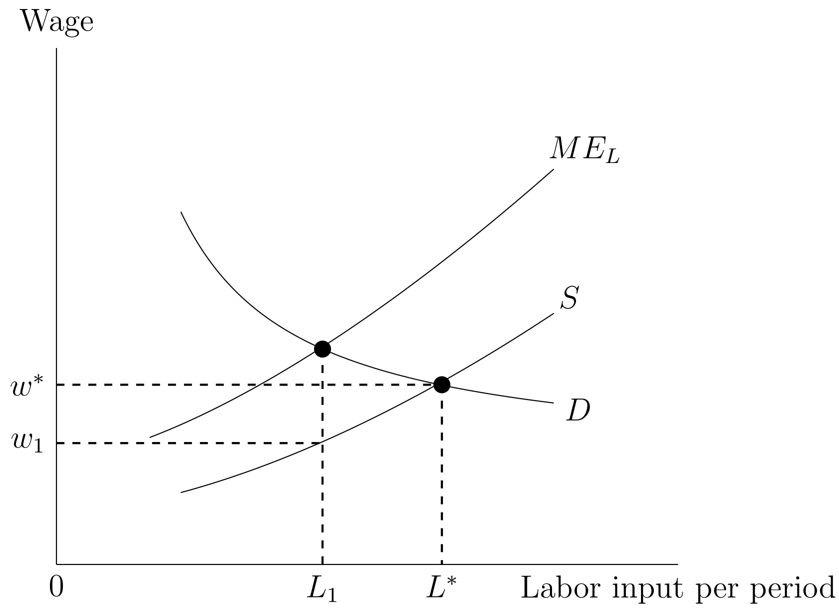
3.1.2 Monopsonistic Labor Market

In addition to wage increases from advancements in labor force productivity, clusters could mitigate monopsonistic labor markets (i.e., labor markets with a single buyer), thus, increasing wages. The profit-maximizing firm will hire to the point where marginal input expenses, ME, equals the marginal revenue product, MRP. However, as Nicholson (1995) states, since there is only one buyer, the firm faces the entire market supply curve. Therefore, the firm does not only pay a higher wage to the new worker but also to all other workers. Thus, the marginal expense for hiring the new worker, ME_L, will surpass the wage rate, w (Nicholson 1995).

In Figure 4, I recreate the Nicholson (1995) illustration of the pricing in a monopsonistic labor market, where the ME_L curve lies above the positively sloped supply curve as each additional worker raises the wage for all workers; L_1 is the level of labor for profit maximization for the firm; w_1 is the wage rate at L_1; L^* is the labor level for a perfectively competitive labor market; and w^* is the wage rate at L^*. Therefore, as more firms enter the labor market—which cluster require to form—the labor market moves towards a perfectively competitive labor market from a monopsonistic labor market, thus increasing wages. Muehlemann et al. (2013) empirically tests this theory by researching how monopsonies affect wages by examining one standard deviation increase in the number of establishments. The authors' data set comprises of 3,562 Swiss firms in 2004. Through their research, the authors find that when the number of establishments increases by one standard deviation, skilled workers' wages increase by 1.8 percent. However, the increase in wages for unskilled workers is not statistically significant.

Furthermore, as the labor market moves closer to being perfectively competitive, firms may need to provide a wage above the current wage rate to attract more workers (Nicholson 1995). Since my data examines the county level, it is more logical to suggest that the rise in competition among firms, which is more present in clusters, rather than a change in labor markets will increase the wage (Porter 2000).

Figure 4: Pricing in a Monopsonistic Labor Market



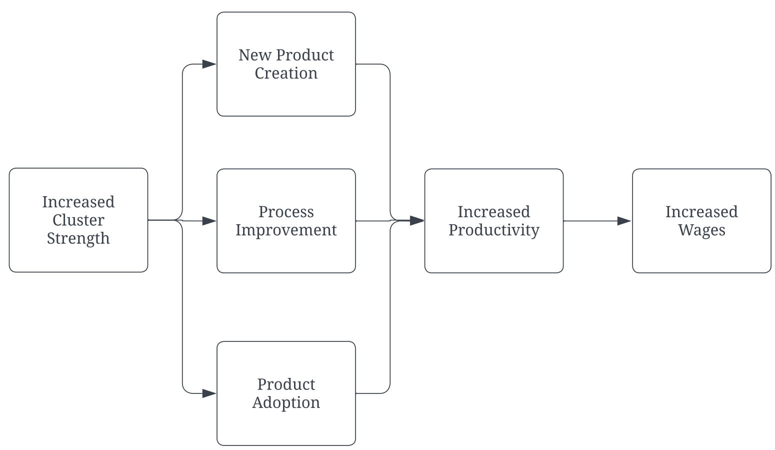
Source: Nicholson (1995)

3.1.3 Innovation

As I mention in Subsection 2.3, there is a strong linkage between productivity and worker wages (Stojčić et al. 2019; Abdesslem and Chiappini 2019; Kim et al. 2009; Howard et al. 2014; Lin and Sai 2022). This linkage, however, can be driven by multiple factors. One of these drivers of productivity that researchers thoroughly explore is innovation. In Mohnen and Hall (2013), the authors seek to better understand how innovation impacts productivity through a review of previous literature. The authors propose three ways that innovation could increase productivity: (1) the creation of a new product, which, under assumption, firms could create with less inputs due to technological advancements; (2) the improvement of processes that allow firms to minimize production costs; (3) the adoption of product to a firm that is not new the market. In addition, the authors touch on that innovation is dependent on both *competition* and *complementarity products*, precisely, the driving factors of clusters that Porter mentions in Porter (1990).

Furthermore, researchers demonstrate a robust positive correlation between innovation—primarily measured through patents—and clusters (Huang et al. 2012; Porter 2003). Therefore, increases in innovation that are set upon by the presence of a cluster could be one channel through which clusters increase productivity, thus increasing wages. In Figure 5, I illustrate how clusters could positively affect wages when using solely productivity as the driver. While clusters could affect other variables that impact productivity, I focus solely on innovation in Figure 5 because of its ease and measurement reliability through patents. This reasoning furthers my understanding of why I expect wages and cluster strength to be positively correlated.

Figure 5: Simplified Cluster Impact on Productivity



3.1.4 Decreased Input Costs

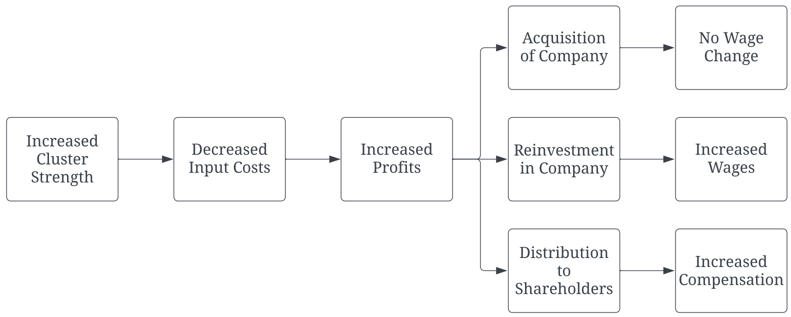
One traditional advantage of clusters from agglomeration theory is decreased input costs. These inputs range from required services, components, machinery, and transportation (Porter 2000; Krugman 1991). In Glaeser and Gottlieb (2009), the authors look closer into one of these decreased costs by comparing the relationship between average shipment length and the logarithm of value per ton. The authors find that the firms ship heaviest goods the least distance, indicating that transport costs are still necessary to firm location choices. Therefore, for a firm with high transport costs, locating within clusters is advantageous as it decreases the input costs relative to the firm locating outside the cluster.

If the firm can reduce costs by locating within the cluster and assuming the firm is profitable, the profits of the firm should increase. The firm can distribute these profits in three different ways: reinvesting in the business, disbursing to shareholders, or acquiring other firms. Of these three, reinvestment into the firm is only option that could affect wages as shareholder disbursement affect compensation and research has shown that the acquisition of other firms does not positively impact average worker wages (Ouimet and Zarutskie 2011).[7] Regarding reinvestment in the firm, whether the firm chooses to increase R&D, buy new capital, or directly increase wages, all channels lead to higher wages, in theory, either through the productivity channel or directly through wage increases.

Regarding shareholder disbursement, assuming the worker invests in the firm through an employee stock ownership plan (ESOP) or employee stock purchase plan (ESPP), the worker would receive extra compensation from the additional distribution. However, because my primary interest is how cluster strength impacts wages, I do not consider shareholder disbarments as an option. In Figure 6, I illustrate the three channels I mention and how they could affect wages. Given that decreased input costs can only affect wages through one of three channels, I do not foresee this traditional advantage of clusters playing a significant role in understanding how cluster strength could affect wages.

[7] Ouimet and Zarutskie (2011) shows that average workers both from the acquiring and acquired firm receive fewer wage increases. The authors state that workers may already receive a wage premium for the acquiring firm; however, the authors' research shows that the most skilled workers receive a wage increase.

Figure 6: Simplified Cluster Impact on Input Costs



3.2 Opposing Theory for Hypotheses

I reason two ways cluster strength could negatively affect wages, causing my first hypothesis to be incorrect. First, firms could collude to fix wages, lowering wages relative to outside firms. Second, the cluster strength could correlate with an outside variable positively affecting wages. While I attempt to remove endogeneity in my model through statistical methods, which I discuss more in Subsection 4.2, I aim to thoughtfully include variables correlating to cluster strength and wages in my empirical model. I further discuss both ways cluster strength could negatively affect wages below.

3.2.1 Collusion

If firms within a cluster collude, the wage relative to firms outside the cluster would be lower. So far, as Mosur and Posner (2023) mentions, most researchers have focused on mergers or no-poaching agreements to address collusion. However, the authors attempt to fill the gap in the literature by utilizing collusion in the product markets to understand collusion in the labor market. Specifically, the authors argue that although antitrust laws apply to all markets, wage agreements cause more harm than price fixing; thus, collusion in the labor market deserves more attention from policymakers.

The authors point to three key differences between the product and labor markets that support their argument that collusion in the labor markets deserves more attention than the product markets. First, since firms have greater control over their workers than their customers, employees face a higher switching cost. Second, labor markets have pay equity norms, thus, causing a greater risk of stealing workers than customers. Third, labor markets face downward nominal rigidity (i.e., workers oppose nominal wage decreases). Therefore, once an employer raises wages to poach workers, they will face difficulties lowering wages. In summary, all three factors allow collusion in the labor markets to be more sustainable than in the product markets.

While the presence of firms colluding could cause clusters to affect wages negatively, I do not foresee this playing a prominent factor for two reasons: illegality and geographical distance. Regarding the latter, Hoffer and Prewitt (2018) mention the illegality of horizontal wage fixing by stating:

> The DOJ has staked out its position that "naked" no-poach and *wage-fixing agreements* are per se illegal under Section 1 of the Sherman Act, meaning the DOJ may prosecute such agreements without any inquiry into their justifications or competitive effects [emphasis added].

Although Mosur and Posner (2023) point out there are likely firms that collude and do fix the wage, regardless of its illegality, it needs to be more common than currently documented to affect my results substantially.

In addition to the illegality of wage fixing, the geographical distance of non-citied clusters would provide an additional challenge as wage levels differ geographically. For example, Wayne County, Michigan, the home of Detroit, and Oakland County, Michigan, are both a part of the Automotive cluster (Delgado et al. 2016). While the two counties touch geographically, they still have significantly different average private wages. In 2016, the average wage for Wayne County, Michigan, was $73,065, while the annual wage for Oakland Country, Michigan, was $67,992 (U.S. Cluster

Mapping 2020). Thus, those working in Wayne, Michigan, made an additional $5,073 on average.

If firms within these two counties were to collude, they would have an outside force working against them, the cost of living. That is, if Wayne County lowers its average annual wage to that of Oakland Country, a decrease of $5,703, assuming low friction of moving and transferable skills, all of which take place in clusters, a majority of the automotive workers in Wayne County would move to Oakland Country since they no longer receive a premium.[8] Therefore, it would not only be illegal to fix wages, but firms would find it exceptionally challenging to fix wages over counties, which is the geographical unit of measurement for my data set. For these reasons, I do not anticipate the presence of collusion significantly lowering the coefficient between clusters and wages.

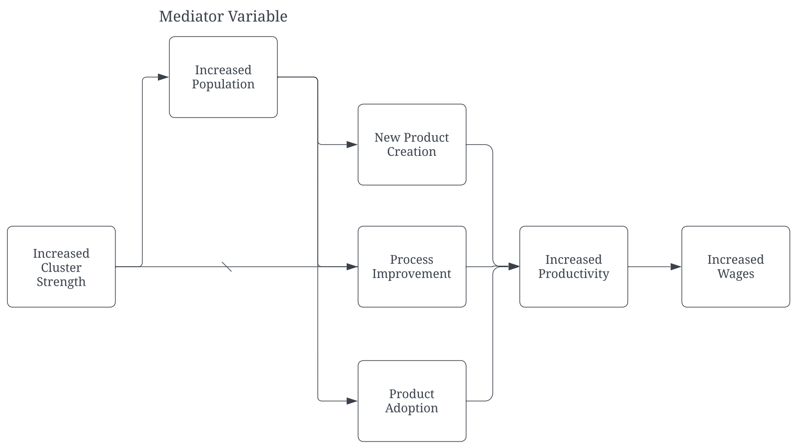
3.2.2 Population

Regarding the latter way that cluster strength could not affect wages, a mediator variable (i.e., a variable that is affected by the independent variable but affects the dependent variable) could make it appear that cluster strength affects wages, when in fact, it is the mediator variable. One mediator variable that I foresee is population as it could closely relate to clusters and could also affects wages. Specifically, an increase in population could, in theory, impact some of the same channels that clusters would go through to increase wages. I foresee two initial examples: a change in innovation and labor force productivity, which I illustrate in Figure 7.

Regarding innovation, in Gössling and Rutten (2007), the authors research whether population impacts innovation by examining European Union countries in the early 2000s. Specifically, the authors attempt to answer whether regional characteristics are essential to innovation in their study. The authors test how six different vari-

[8] It would not stand to reason that Oakland Country would want to increase their average salary by $5,703 as this only increases their expenses and provides no benefit.

Figure 7: Simplified Cluster Impact on Productivity with Mediator Variable



ables influence innovation in a region as a measure of input (R&D expenditure) and output measure (patents per million inhabitants). One of these variables is urbanization, which they measure through the variable population density. The authors find that population density negatively correlates to R&D expenditure and the number of patents. Therefore, population will not likely substantially increase innovation, increasing wages.

However, population could inflate wages another way. In Glaeser and Maré (2001), the authors conclude that workers in dense metropolitan areas make a wage premium of 25 percent after controlling relevant factors. Yankow (2006) expands on Glaeser and Maré (2001) by providing an explanation of this wage premium. The author's calculation shows a 19 percent wage premium, with two-thirds of the premium coming from large urban areas' ability to capture highly skilled workers. Therefore, I determine it is important to control for both education as a measure of skill and population density in the model to ensure that population increase is not driving the wage change.

4 Empirical Results

4.1 Data

I collect a majority of my variables from U.S. Cluster Mapping, which is "led by Harvard Business School's Institute for Strategy and Competitiveness in partnership with the U.S. Department of Commerce and U.S. Economic Development Administration" (U.S. Cluster Mapping 2020). U.S. Cluster Mapping stores multiple open data records, including U.S. Census Bureau County Business Patterns, U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics, STATS America, Regional Innovation Acceleration Network, and U.S. Patent and Trademark Office. Outside of U.S. Cluster Mapping, I use County Business Patterns for data on the business environment (e.g., number of establishments and employment) and demographics (e.g., population and land area). Furthermore, I use the U.S. Bureau of Labor Statistics to collect regional and national consumer price index data to convert annual private wages to real annual private wages.

4.1.1 Cluster Strength

My main variable of interest is cluster strength, which U.S. Cluster Mapping (2020) defines as the "percent of traded employment in strong clusters." To better define this variable, I further explain traded employment and strong cluster. Regarding traded employment, Delgado et al. (2016) classify trade industries—which employ trade labor—as industries "that are more geographically concentrated and produce goods and services that are sold across regions and countries"—think, Hardware Manufacturing, for example. However, researchers exclude industries that rely on location specific natural resources (e.g., mining) from the trade industry classification (Porter 2003). Opposite to trade industries are local industries, which Delgado et al. (2016) define as industries that "serve primarily the local markets"—think, retail or barber shops, for example—and "whose employment is evenly distributed across regions in proportion to regional population."

Prior to defining strong clusters, a further quantitative understanding of clusters is needed. As I mention in Section 2, Delgado et al. (2016) create a set of 51 U.S. Benchmark Cluster Definitions using 778 trade industries. To create these definitions, the authors construct a cluster algorithm that assigns each trade industry—using six-digit NAICS codes—to an exclusive cluster. The authors' algorithm follows a five-step process to create each cluster configuration, C. First, the authors construct a similarity matrix, M_{ij} that captures similarities between industries i and j. Specifically, the authors use a similarity matrix, $LC - IO - Occ_{ij}$, which comprises an average of the four (standardized) matrices: (1) employment locational correlation, which captures co-locational formations in employment for region r; (2) establishment locational correlation, which captures co-locational formations in establishments for region r; (3) input-output links, which capture cross-industry flows between buyers and sellers; and (4) labor occupation links, which captures industries with similar skills. I show each matrix in Equations 14 through 17. Furthermore, in Table 1, I include the first twelve of 51 U.S. Benchmark Cluster Definitions from Delgado et al. (2016), where the complete table is available on page 43. Note DownStream Metal Products, which I use as an example later in this section.

$$LC - Employment_{ij} = Correlation(Employment_{ir}, Employment_{jr}) \tag{14}$$

$$LC - Establishments_{ij} = Correlation(Establishments_{ir}, Establishments_{jr}) \tag{15}$$

Table 1: First Twelve Proposed Set of Benchmark Cluster Definitions

Cluster Name	No. Industries	% Traded Employ	WCR Rank	WCR Score	WCR^{LC-Emp}	WCR^{LC-Est}	WCR^{IO}	WCR^{Occ}
Aerospace Vehicles and Defense	7	1.3%	1	1.93	0.2	0.63	0.15	0.87
Agricultural Inputs and Services	9	0.2%	1	1.14	0.35	0.53	0.1	0.46
Apparel	21	0.4%	1	2.28	0.45	0.74	0.11	1
Automotive	26	1.9%	1	2.27	0.31	0.62	0.22	0.65
Biopharmaceuticals	4	0.6%	1	3.35	0.59	0.76	0.23	1
Business Services	33	24.2%	1	1.17	0.66	0.83	0.04	0.25
Coal Mining	4	0.2%	2	2.3	0.44	0.53	0.22	0.62
Communications Equipment and Services	8	1.3%	1	2.37	0.47	0.79	0.23	0.41
Construction Products and Services	20	1.8%	1	1.81	0.39	0.61	0.21	0.29
Distribution and Electronic Commerce	62	13.0%	1	2.19	0.67	0.82	0.12	0.63
Downstream Chemical Products	13	0.6%	1	1.3	0.39	0.69	0.04	0.71
Downstream Metal Products	16	1.0%	1	1.05	0.28	0.58	0.02	0.82

Note: WCR is the average of the (standardized) WCR^{LC-Emp}, WCR^{LC-Est}, WCR^{IO}, and WCR^{Occ}.

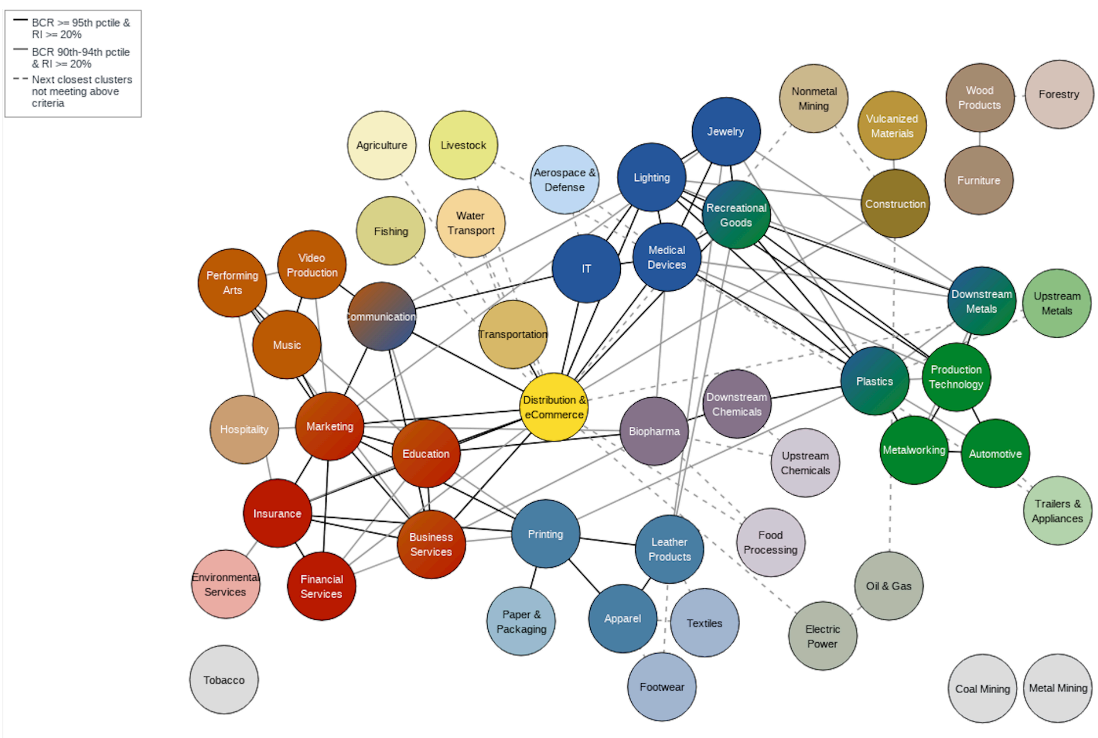
Source: Delgado et al. (2016)

$$IO_{ij} = Max\{input_{i\to j}, input_{i\leftarrow j}, output_{i\to j}, output_{i\leftarrow j}\} \tag{16}$$

$$Occ_{ij} = Correlation(Occupation_i, Occupation_j) \tag{17}$$

In step two, the authors create their broad parameter choices β. These parameters include the initial number of clusters, the normalization of the data, and starting values for the clustering function. Third, through the construction of the similarity matrix, M_{ij}, and broad parameter choices, β, the authors create a clustering function, which takes as inputs the similarity matrix and parameter choices. This clustering function allows the authors to construct the initial cluster configuration, C. Fourth, the authors assign validation scores (VS) that report which clusters and industries have the highest Within Cluster Relatedness (WCR) comparative to Between Cluster Relatedness (BCR) of different clusters. In Figure 8, I provide a list of related clusters from U.S. Cluster Mapping (2020) that better demonstrates Between Cluster Relatedness for each of the 51 clusters. Furthermore, configurations Cs with the highest validation scores receive the label of C^* and move on to step five. In the final step, the authors examine each C^* and correct for any mistakes that might have been made through data limitations. The authors move any outliers to the "next best" cluster, leaving them with the finalized set of cluster definitions, C^{**}.

Figure 8: Full Portfolio View of Related Clusters



Source: U.S. Cluster Mapping (2020)

U.S. Cluster Mapping (2020) uses these 51 cluster definitions from Delgado et al. (2016) to determine strong clusters. Specifically, U.S. Cluster Mapping (2020) defines a strong cluster "as those where the location quotient, i.e. the cluster's relative employment specialization, puts them into the leading 25% of regions across the U.S. in their respective cluster category." In Equation 18, I demonstrate the formula to calculate a cluster's location quotient, where a location quotient greater than one signifies that County y's trade labor specializes more in Cluster x than the U.S. U.S.

Cluster Mapping (2020) calculates the location quotient for every cluster definition to generate a list of strong clusters by county. Through taking the list of strong clusters in each county and summing all trade labor in strong clusters and then dividing it by all trade labor in that county, U.S. Cluster Mapping (2020) creates the cluster strength variable. In Equation 19, I provide a mathematical representation of cluster strength for easier understanding.

$$\text{Location Quotient} = \frac{\frac{\text{Trade labor in cluster } x \text{ in county } y}{\text{All trade labor in county } y}}{\frac{\text{Trade labor in cluster } x \text{ in United States}}{\text{All trade labor in United States}}} \tag{18}$$

$$\text{Cluster Strength} = \frac{\sum_{i=1}^{n}(\text{Trade labor in strong cluster } x_i \text{ in county } y)}{\text{All trade labor in county } y} \tag{19}$$

As an example, I calculate the location quotient and cluster strength for Brookings, SD, my university's location, in 2014. Regarding location quotient, I use equation 18 and information in Table 2—a subsample of Table A1, which I collect from U.S. Cluster Mapping (2020)—to determine the location quotient for one of Brookings, SD's strong clusters, Downstream Metal Products, which I show in Equation 20.

Furthermore, for understanding, in Table 3, I provide the North American Industry Classification System (NAICS) codes—comprising 16 industries—for the Downstream Metal Products cluster from Delgado et al. (2016). Due to many industries in each cluster, it is unlikely that one firm is causing an increase in cluster strength, but rather an increase of labor in multiple firms within the grouping of similar industries.

$$\text{Location Quotient} = \frac{\frac{810}{4{,}800}}{\frac{402{,}823}{43{,}733{,}043}} = 18.32 \tag{20}$$

A location quotient of 18.32 is significant as it signifies that Brookings, SD trade labor is 18.32 times more specialized in Downstream Metal Products than the U.S for the year 2014. This puts Brookings, SD in the top 25 percent of Downstream

Table 2: Condensed Trade Employment Data for Brookings, SD, 2014

Cluster Name	**2014 Employment**	**Strong Cluster**	**National Rank**	**2014 U.S. Employment**
Downstream Metal Products	810	TRUE	134	402,823
Production Technology and Heavy Machinery	435	TRUE	592	978,399
Electric Power Generation and Transmission	120	TRUE	306	150,379
...	...	...	...	...
Total	**4,800**	**3**		**43,733,043**

Source: U.S. Cluster Mapping (2020)

Table 3: Downstream Metal Products Industries

Cluster Name: **Downstream Metal Products**
Cluster Code: **12**
Description: This cluster contains establishments that manufacture metal containers, prefabricated metal structures, and end user metal products. These end user products include ammunition, kitchenware, hardware, metal bathroom fixtures, and similar metal products used in home finishing such as doors, windows and ornamentation.

Number of Industries 16

NAICS	NAICS Name	Subcluster Name
332211	Cutlery and Flatware (except Precious) Manufacturing	Metal Products
332213	Saw Blade and Handsaw Manufacturing	Metal Products
332214	Kitchen Utensil, Pot, and Pan Manufacturing	Metal Products
332321	Metal Window and Door Manufacturing	Metal Products
332323	Ornamental and Architectural Metal Work Manufacturing	Metal Products
332510	Hardware Manufacturing	Metal Products
332998	Enameled Iron and Metal Sanitary Ware Manufacturing	Metal Products
332999	All Other Miscellaneous Fabricated Metal Product Manufacturing	Metal Products
332992	Small Arms Ammunition Manufacturing	Ammunition
332993	Ammunition (except Small Arms) Manufacturing	Ammunition
332994	Small Arms Manufacturing	Ammunition
332995	Other Ordnance and Accessories Manufacturing	Ammunition
332311	Prefabricated Metal Building and Component Manufacturing	Fabricated Metal Structures
332312	Fabricated Structural Metal Manufacturing	Fabricated Metal Structures
332431	Metal Can Manufacturing	Metal Containers
332439	Other Metal Container Manufacturing	Metal Containers

Source: U.S. Cluster Mapping (2020)

Metal Products cluster in the U.S. Additionally, Brookings, SD has two other strong clusters: Production Technology & Heavy Machinery and Electric Power Generation & Transmission (Delgado et al. 2016). Using Table 2, again, I construct the cluster strength variable for Brookings, SD in Equation 21, using Equation 19.

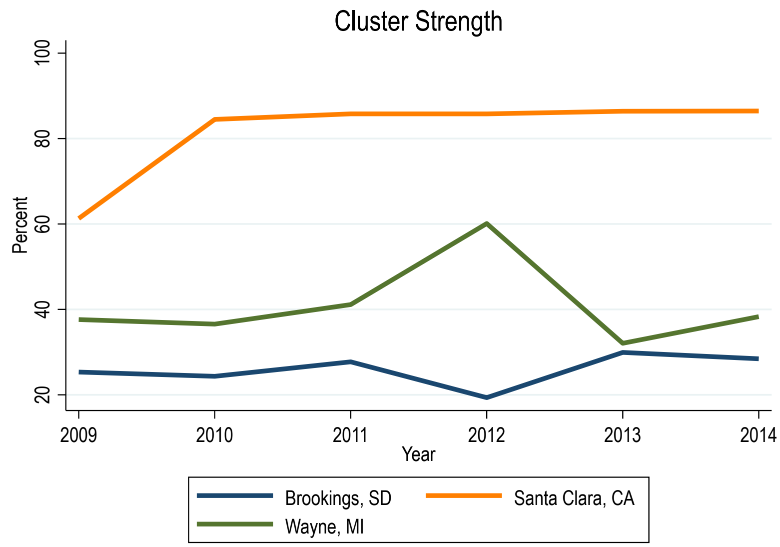
$$\text{Cluster Strength} = \frac{810 + 435 + 120}{4,800} = 0.284 \tag{21}$$

4.1.2 Control Variables

To more robustly understand how cluster strength affects wages, I use a plethora of control variables. The first control variable is labor force productivity (LFP), which is the "real GDP in 2005 dollars per labor force participant" by county (U.S. Cluster Mapping 2020). I divide labor force productivity by 1,000 to allow for a more straightforward interpretation. To capture differences in education throughout the counties over time, I use the percentage of the population over 25 receiving a high school diploma or more, some college or associate's degree or more, and completing a bachelor's degree or more. Furthermore, I use the count of patents to measure the county's innovation. I use each county's population and square miles to create population density. Lastly, I include the number of establishments and the number of employees to help showcase the business environment of the counties over time. I use the number of establishments rather than the number of firms or enterprises because establishments better represent the decision-making and development of the county (Sadeghi et al. 2016).

In Figure 9, I demonstrate the movement of my primary independent variable of interest, cluster strength, over the life of the data set, 2009 to 2014. I include three specific counties to highlight the data: Brookings County in South Dakota, my university's location, Santa Clara County in California, one of the county locations for Silicon Valley, and Wayne County in Michigan, the home of Detroit, which, as I previously state, is the home of the Automotive cluster (Delgado et al. 2016). Furthermore, in Figure 10, I show the movement of all control variables using the same counties.

Figure 9: Cluster Strength



4.1.3 Real Regional Average Private Wage

My dependent variable is the real regional annual private wage by county. I construct this variable by deflating each county's annual private wage by the corresponding regional consumer price index. Specifically, I use four census regions from the Bureau of Labor Statistics: West, Midwest, Northeast, and South.[9] In Figure 11, I show the breakdown of census regions by state.

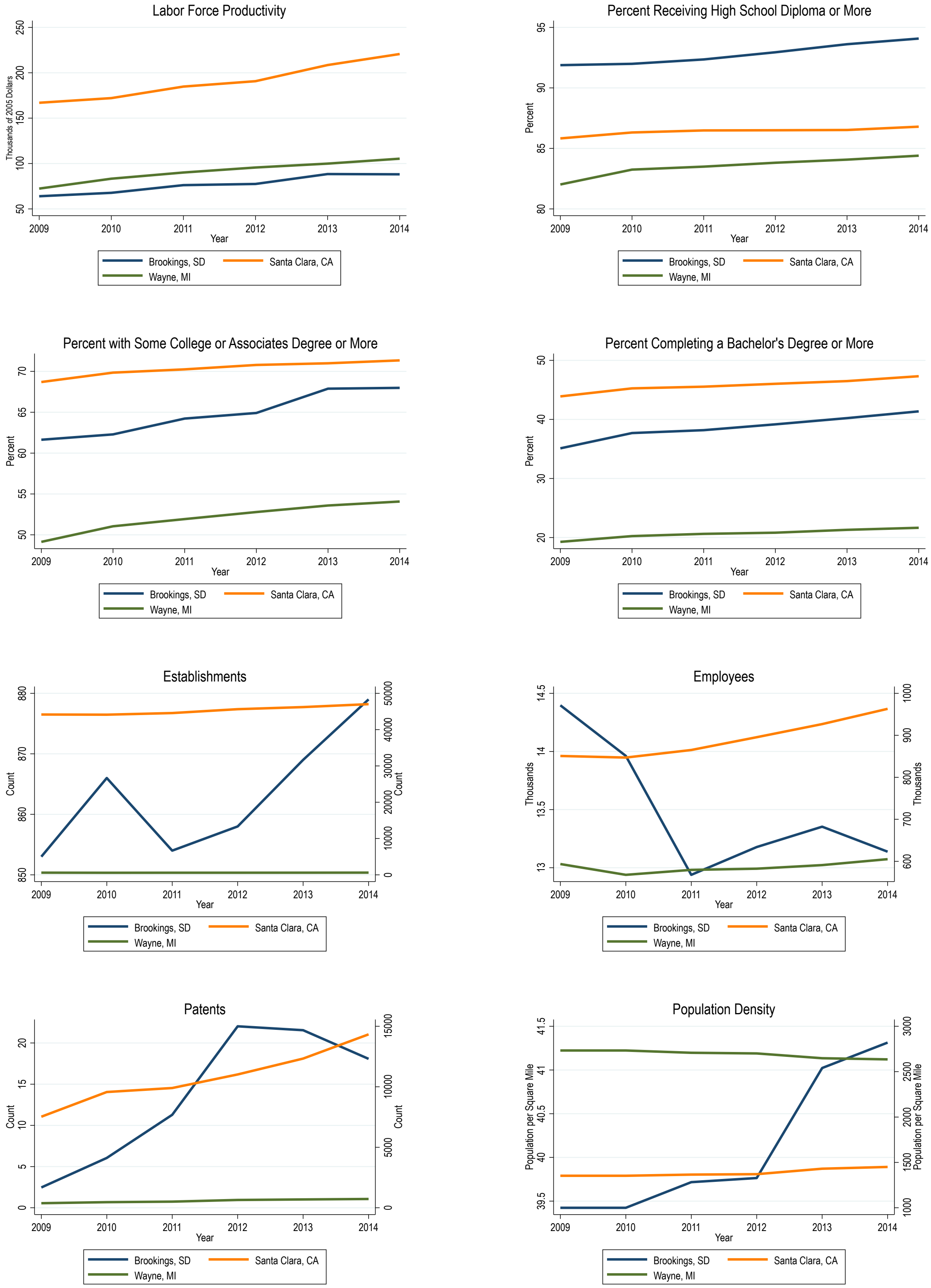
Likewise, I include the regional nominal average private wage and the real national average private wage, which I adjust for the national consumer price index in my results. I illustrate the movement of the real regional annual private wage in Figure 12 over the same three counties. Finally, in Figure 13, I illustrate the bi-variate

[9]In December of 2018, the Bureau of Labor Statistics added nine divisions to their regions. Since my data range is 2009-2014, I could not account for this additional granularity.

relationship of real regional average private wage and cluster strength for my data set's first and last year.[10] As I show in Figure 13, multiple counties experience wage and cluster strength increases over the range of the data set. However, this similar movement does not necessarily imply causality. To conclude, I display the summary statistics for all my data in Table 4.

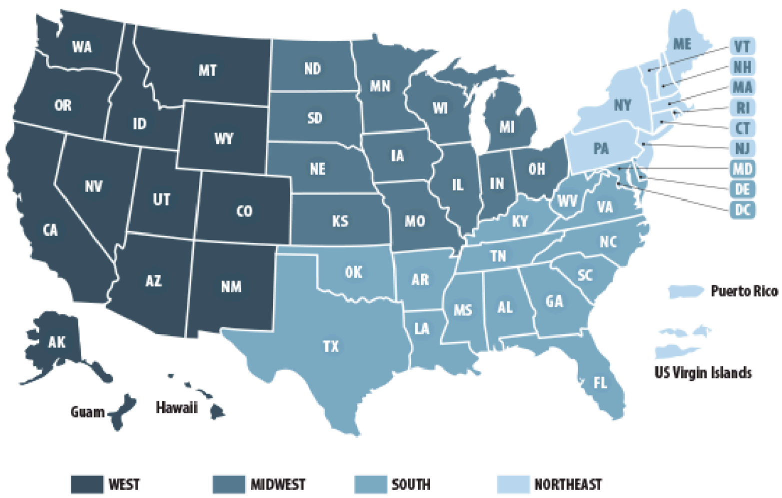
[10]I form the bi-variate matrix using bins created by percentile data distribution instead of hard cutoffs (i.e., 33% and 66%) to better visualize the data.

Figure 10: Control Variables



Note: Wayne, MI, and Santa Clara, CA, are represented on the secondary Y axis for figures with two Y axes.

Figure 11: Census Regions



Source: U.S. Bureau of Labor Statistics

Figure 12: Real Regional Average Private Wage

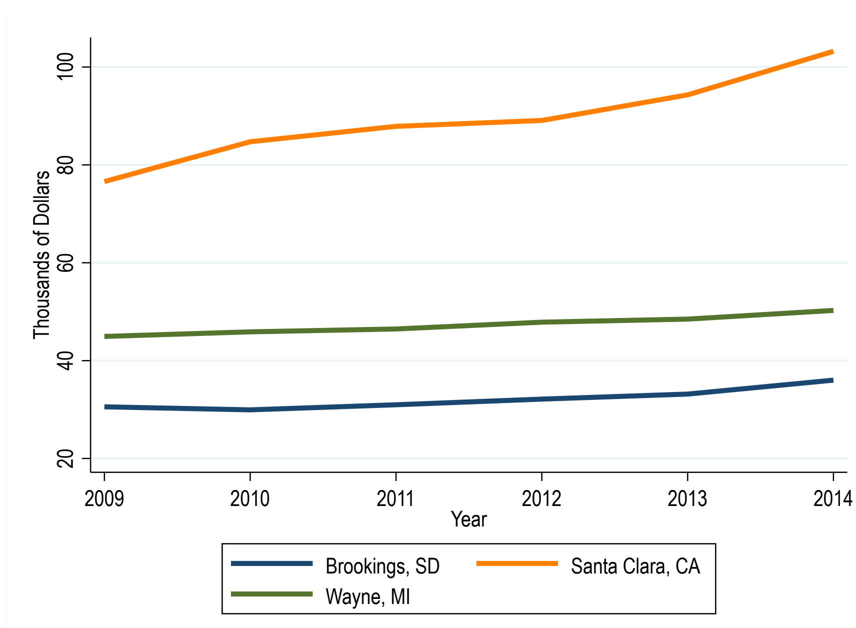
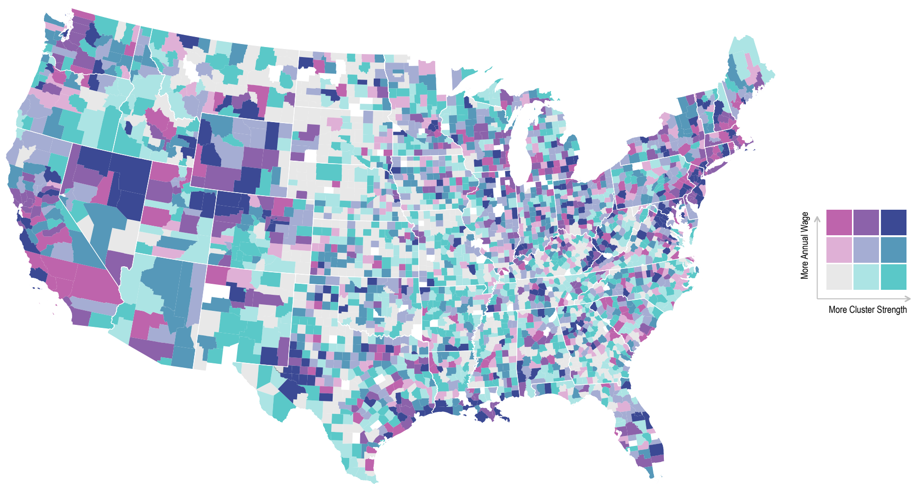
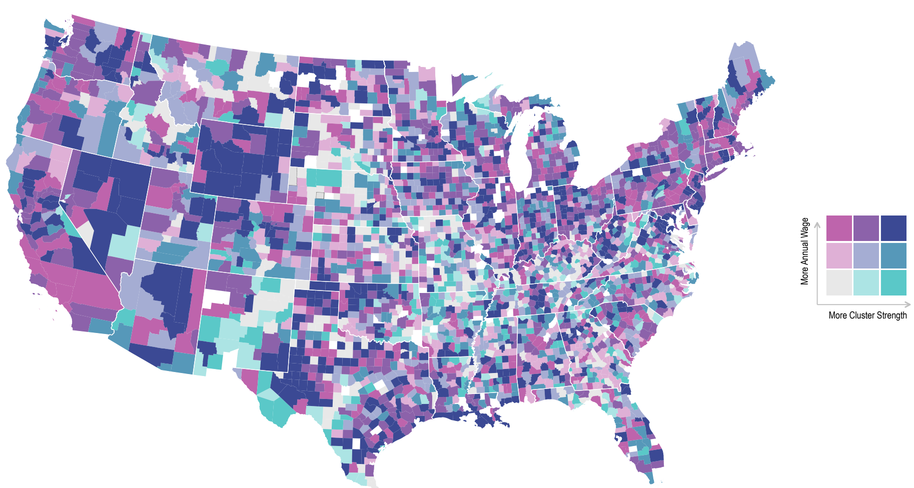


Figure 13: Cluster Strength and Real Regional Wage: Bi-Variate Map

2009



2014



Note: Alaska and Hawaii removed due to graphical limitations.

Table 4: Summary Statistics

	Observations	Mean	SD	Min	Max
Nominal Regional Wage	17,671	33379.47	8105.10	11,757	106747
Real National Wage	17,671	32836.71	7947.02	11,513	105023
Real Regional Wage	17,671	32830.21	7946.43	11,504	104764
Labor Force Productivity (Thousands)	17,659	74.98	100.67	11	3,692
Cluster Strength	17,669	47.94	20.33	0	97
Percent Receiving High School Diploma +	17,487	83.73	7.17	45	99
Percent with Some College or Associates Degree +	17,487	48.35	10.77	18	86
Percent Completing a Bachelor's Degree +	17,487	19.23	8.57	3	72
Patents	17,671	34.56	265.10	0	14,333
Population Density	17,650	190.22	1216.30	0	48,417
Establishments	17,561	2383.37	8171.31	3	258982
Employment (Thousands)	17,561	36.21	134.58	0	3,933

4.2 Panel Regression and Estimates

I estimate a model that takes the general form of Equation 22, where $w_{i,t}$ is the wage measure for county i at time t; $x_{i,t}$ is a vector of control variables, excluding cluster strength, cs, that vary across counties (that is, i), across time (that is, t), or some combination of both counties and time; in a fixed effects estimation, u_i captures unobserved heterogeneity, i.e., the difference across units, that is, counties; $v_{i,t}$ is an idiosyncratic error, i.e., any unobserved factors that affect $w_{i,t}$. Additionally, I cluster standard errors by state; thus, providing a more robust analysis, as the state where county i resides will likely impact the wage due to different state policies and features. My primary coefficient of interest is γ, the measure of wage movement from cluster strength.

$$w_{i,t} = x_{i,t}\beta + cs_{i,t}\gamma + u_i + v_{i,t} \tag{22}$$

4.2.1 Ordinary Least Squares

In Table 5, I report my initial results using a less robust ordinary least square (OLS) regression.[11] I include the nominal regional wage, real regional wage, and real nation

[11] Ordinary least square regressions are less robust due to multiple limitations, including: severe estimation errors from outliers, sensitivity to correlation in independent variables, and susceptibility

wage in columns one through three, respectively. Furthermore, I report standard errors in parentheses below the coefficients. I indicate the significance of each coefficient by including the p-value, which I denote by the number of stars to the right of the coefficient.

Table 5: OLS Panel Regression Results for Average Private Wage

	Nominal Regional Wage	Real Regional Wage	Real National Wage
	(1)	(2)	(3)
Labor Force Productivity (Thousands)	22.626***	22.147***	22.164***
	(7.321)	(7.169)	(7.164)
Cluster Strength	64.993***	63.867***	63.877***
	(11.775)	(11.510)	(11.523)
Percent Receiving High School Diploma +	33.106	32.810	31.859
	(51.530)	(50.703)	(50.523)
Percent with Some College or Associates Degree +	−1.020	−3.929	−5.151
	(61.619)	(60.287)	(60.361)
Percent Completing a Bachelor's Degree +	249.867***	248.737***	250.205***
	(77.089)	(75.423)	(75.582)
Patents	4.518***	4.423***	4.414***
	(0.632)	(0.612)	(0.609)
Population Density	0.678*	0.664*	0.664*
	(0.364)	(0.358)	(0.358)
Establishments	−0.374***	−0.367***	−0.367***
	(0.123)	(0.121)	(0.121)
Employment (Thousands)	32.658***	32.093***	32.132***
	(6.043)	(5.916)	(5.943)
Constant	20388.836***	20127.173***	20242.800***
	(2799.396)	(2756.050)	(2745.633)
State Fixed Effects	*No*	*No*	*No*
R-Squared	0.415	0.416	0.416
Observations	17349	17349	17349

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

I report a statistically significant relationship (p<0.01) between the regional nominal wage and cluster strength in column one. Specifically, the regional nominal wage increases by \$64.99 when cluster strength increases by one percent. In addition, the regional real wage increases by \$63.87, and the real national wages increase by \$63.88 when increasing cluster strength by one percent, as I report in columns two and three, respectively. This change is a substantial increase as it implies that a county that starts with a zero percent cluster strength and increases it to 100 percent would

to heteroskedasticity (i.e., the standard deviations of the predicted variables are not constant).

increase their regional real wage by an estimate of \$6,387.

Furthermore, the control variables mostly perform as I expect. Labor force productivity and count of patents correlate positively and statistically significantly ($p<0.01$) to all wage measures. Additionally, as the percentage of the population over 25 with a bachelor's degree or more increases, all measures of wages increase ($p<0.01$). However, the same does not hold true when looking at the completion of high school or some college, which is not statistically significant.

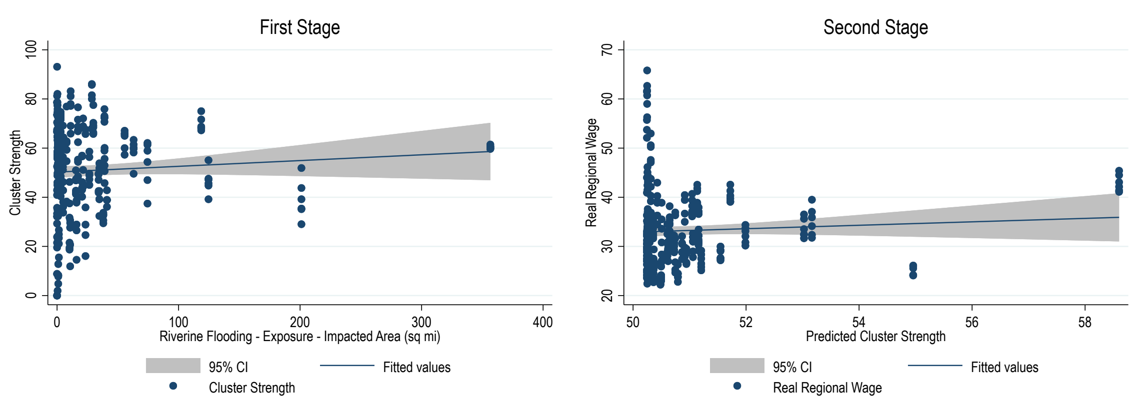
One unexpected result I report is the negative and statistically significant correlation between wages and the number of establishments. This relationship is likely due to the nature of the establishments. That is, if there are more local establishments as opposed to trade establishments, the average wage could decrease as local labor, on average, is paid less than trade labor (Porter 2003).

4.2.2 Two-Stage Least Squares

In attempt to eliminate endogeneity, I perform a two-stage least squares regression. To do this, researchers utilize an instrumental variable, which is correlated with the independent variable of interest, but could not, through *assumption*, be correlated with the error term. In the first stage, the researcher regresses the independent variable of interest, X on the instrumental variable to create estimated linear predictor values (that is, $\hat{X}$). In the second stage, the researcher regresses the dependent variable on the linear predicted values to create a linear regression. For further understanding, in Figure IV, I provide a graphical representation of both stages with my instrumental variable, riverine flooding - exposure - impacted area by square mile, using a sample of 50 counties from my data set. For the remainder of my thesis, I refer to this variable as $River_{Exp}$ for simplicity.

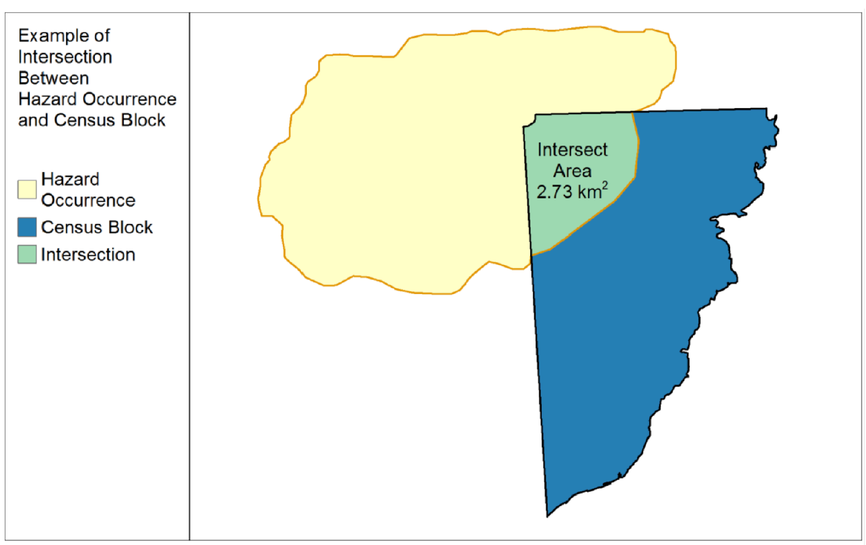
To create a two-stage least squares regression, I use the variable $River_{Exp}$ by county from the Federal Emergency Management Agency (FEMA) National Risk In-

Figure 14: Two-Stage Least Square Example



dex. FEMA constructs this variable by analyzing susceptible areas that have either historical occurrences of a riverine flooding or identifiable risks. Specifically, FEMA creates an annualized frequency of the area of the intersection between the county block and hazard block from 1996 to 2019. In Figure 15, I provide an example of the intersection between the two blocks from National Risk Index Technical Documentation, which, in the example, uses census blocks instead of county blocks.

Figure 15: Intersection Example



Source: National Risk Index Technical Documentation

My assumption is that $River_{Exp}$ is not correlated with my error term, while counties with a higher $River_{Exp}$ are more likely to have a higher cluster strength as early clustered firms presumably chose to locate by rivers to ship their trade goods. As a test, I run the first stage in my two-stage least square regression by regressing clusters strength on $River_{Exp}$, while also keeping my control variables. I report my results in Table 6, where I find that $River_{Exp}$ statistically and positively correlates to my main independent variable of interest, cluster strength. In addition, in Table A2, I provide a correlation matrix with the entirety of my data.

Table 6: First Stage

	(1)
Riverine Exposure	0.035^{***}
	(0.010)
Labor Force Productivity (Thousands)	0.012^{**}
	(0.005)
Percent Receiving High School Diploma +	-0.071
	(0.144)
Percent with Some College or Associates Degree +	-0.064
	(0.151)
Percent Completing a Bachelor's Degree +	0.339^{**}
	(0.138)
Patents	0.003^{***}
	(0.001)
Population Density	0.000^{**}
	(0.000)
Establishments	-0.000
	(0.000)
Employment (Thousands)	-0.007
	(0.008)
Constant	48.843^{***}
	(9.932)
R-Squared	0.019
Observations	17348

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In the second stage, which I report in Table 7, I regress the real regional wage on the linear predicted value of cluster strength, $\hat{cs}$, from the first stage, keeping my control variables. I report that there is no statistical significance between $\hat{cs}$ and the real regional wage. While my result is not statistically significant (p=0.20), cluster strength could still affect wages. My result simply shows that I am unable to eliminate endogeneity in the model by using the instrumental variable $River_{Exp}$. I further determine whether using an instrumental variable is necessary or if a standard OLS regression is sufficient by using Durbin–Wu–Hausman test, which Davidson and MacKinnon (1993) suggest. Through running the test, I report a p-value of 0.265, therefore I don't reject my null hypothesis that my OLS regression is efficient. In addition, I also perform an Augmented Regression Test, where, in the first step, I regress cluster strength on $River_{Exp}$ and my control variables, which I show in Table 8. In step two, I regress the real regional wage on cluster strength, my control variables, and the predicted residual, $zres$ of Table 8. I report, in Table 9, that $zres$ is not statistically significant (p=0.395), therefore, I again don't reject my null hypothesis that my OLS regression is efficient.

4.2.3 Fixed Effects

As an extra robustness test, I attempt to avoid omitted-variable bias in my regression, where omitted-variables bias is when relevant variables are left out of the model. This can be done through either using a fixed effects (FE) or random effects (RE) model. To determine which model to use, I run another Durbin–Wu–Hausman test. I report a p-value of 0.000, therefore, I reject my null hypothesis that random effects is the preferred model. By using an FE model, I control for all variables that differ over the cross-sectional units but remain constant over time; thus, eliminating the omitted-variable bias. In Figure 16 and 17, I provide an example of how FE differs from OLS within my data by using the same three counties above. In Figure 16,

Table 7: Second Stage

	(1)
$\hat{cs}$	178.993
	(138.580)
Labor Force Productivity (Thousands)	20.844**
	(8.004)
Percent Receiving High School Diploma +	44.959
	(55.242)
Percent with Some College or Associates Degree +	0.115
	(66.036)
Percent Completing a Bachelor's Degree +	212.648**
	(88.414)
Patents	4.081***
	(0.900)
Population Density	0.627*
	(0.358)
Establishments	-0.368***
	(0.125)
Employment (Thousands)	33.078***
	(6.480)
Constant	14184.384**
	(7006.999)
R-Squared	0.391
Observations	17350

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

I show a significant relationship between cluster strength and the nominal regional wage as OLS uses *between* county variation. However, this relationship becomes less significant when I use the FE regression, which uses *within* county variation, as I show in Figure 17. This difference allows FE to better eliminate endogeneity in the regression model, providing a more robust result.

In Table 10, I show my results from my FE regression for which I attempt to remove any unobservable variable bias from the heterogeneous performance of each

Table 8: Augmented Regression Test - First Step

	(1)
Riverine Exposure	0.035***
	(0.010)
Labor Force Productivity (Thousands)	0.012**
	(0.005)
Percent Receiving High School Diploma +	-0.071
	(0.144)
Percent with Some College or Associates Degree +	-0.064
	(0.151)
Percent Completing a Bachelor's Degree +	0.339**
	(0.138)
Patents	0.003***
	(0.001)
Population Density	0.000**
	(0.000)
Establishments	-0.000
	(0.000)
Employment (Thousands)	-0.007
	(0.008)
Constant	48.843***
	(9.932)
R-Squared	0.019
Observations	17348

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

state. I report that cluster strength is still positively and statistically significant (p<0.05) for each wage measure. However, the coefficients for all wage measures have decreased ($14.42 for the real regional wage) relevant to Table 5. This decrease in the coefficient is likely due to the removal of any unobservable variables. For example, one of these unobservable variables could be the difference in the cost of living between states, which could explain part of the wage differences.

Similar to Table 5, most coefficients respond as I expect. Labor force productivity

Table 9: Augmented Regression Test - Second Step

	(1)
Cluster Strength	178.739
	(135.936)
Labor Force Productivity (Thousands)	20.848***
	(7.685)
Percent Receiving High School Diploma +	44.933
	(48.114)
Percent with Some College or Associates Degree +	0.242
	(61.363)
Percent Completing a Bachelor's Degree +	212.664**
	(83.574)
Patents	4.082***
	(0.888)
Population Density	0.627*
	(0.350)
Establishments	-0.368***
	(0.121)
Employment (Thousands)	33.081***
	(6.298)
zres	-115.385
	(134.536)
Constant	14192.774**
	(6615.018)
R-Squared	0.417
Observations	17348

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

and patents remain statistically significant (p<0.01) and have a similar coefficient to the OLS regression. However, the coefficient for the total receiving a high school degree and total receiving some college or associate degree is now statistically significant (p<0.01), while the total completing a bachelor's degree is no longer significant. Furthermore, population density is now statistically insignificant, while number of es-

Figure 16: Ordinary Least Squares

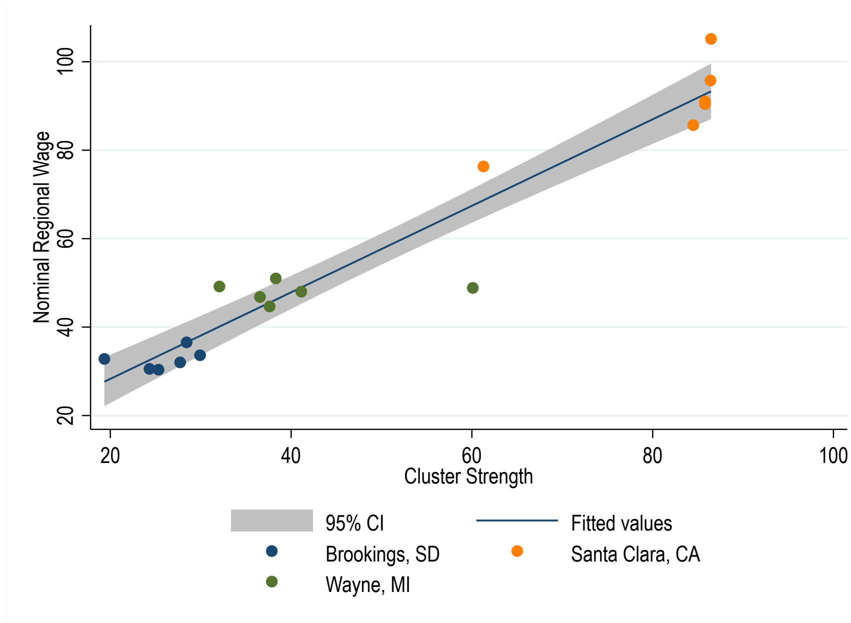
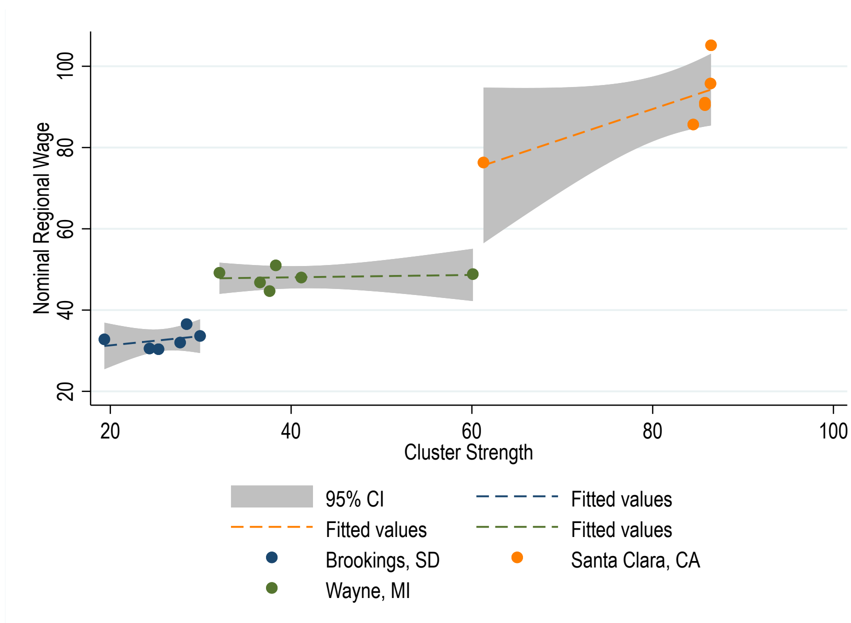


Figure 17: Fixed Effect



tablishments and employment remain statistically significant with similar coefficients to the OLS regression.

Table 10: FE Panel Regression Results for Average Private Wage

	Nominal Regional Wage	Real Regional Wage	Real National Wage
	(1)	(2)	(3)
Labor Force Productivity (Thousands)	16.839^{***}	16.088^{***}	16.110^{***}
	(4.003)	(3.651)	(3.639)
Cluster Strength	15.074^{**}	14.420^{**}	14.466^{**}
	(6.701)	(6.553)	(6.539)
Percent Receiving High School Diploma +	282.756^{***}	254.284^{***}	255.985^{***}
	(30.825)	(29.004)	(29.013)
Percent with Some College or Associates Degree +	422.517^{***}	369.842^{***}	370.856^{***}
	(33.150)	(31.725)	(31.797)
Percent Completing a Bachelor's Degree +	-10.480	4.839	3.055
	(57.502)	(56.685)	(56.588)
Patents	4.988^{***}	4.202^{***}	4.253^{***}
	(0.770)	(0.494)	(0.477)
Population Density	0.784	0.643	0.586
	(0.745)	(0.732)	(0.720)
Establishments	-0.737^{**}	-0.637^{*}	-0.597^{*}
	(0.349)	(0.353)	(0.350)
Employment (Thousands)	52.089^{***}	56.053^{***}	54.894^{***}
	(9.722)	(8.620)	(8.557)
Constant	-13058.770^{***}	-9209.544^{***}	-9406.575^{***}
	(3015.777)	(2851.813)	(2837.958)
State Fixed Effects	Yes	Yes	Yes
R-Squared	0.292	0.270	0.271
Observations	17349	17349	17349

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

4.2.4 Regression Results by Quartile

Although I determine that cluster strength positively affects the average private wage, I further wish to understand if counties with different wages, labor force productivity, and patents benefit differently from a one percent increase in cluster strength. To do so, I rerun my OLS and FE regression, this time, however, I separate my data by quartiles for each one of the variables I mention. I include summary statistics for each quartile of wages, labor force productivity, and patents prior to the corresponding regression results due to limitations of including all data in one table.

In Table 11, I provide summary statistics by quartile for the real regional wage. I report, in Table 12, that counties in the top (fourth) quartile receive the largest increase in real regional wage when cluster strength increases by one percent. Furthermore, counties in the bottom (first) quartile receive a smaller increase, or worse,

Table 11: Real Regional Wage Quartile Summary Statistics

	Observations	Mean	SD	Min	Max
1	4,418	25043.32	2312.99	11,504	27,822
2	4,418	29606.82	1004.37	27,822	31,368
3	4,418	33401.59	1288.46	31,368	35,904
4	4,417	43271.49	8013.64	35,905	104764
Total	17,671	32830.21	7946.43	11,504	104764

a decrease in their real regional wage as cluster strength increases when using an FE model. This decrease is likely due to more trade labor entering lower paying trade industries (think, Glove and Mitten Manufacturing, which is a part of the Apparel cluster, as an example) within the county.

I provide summary statistics by quartile for labor force productivity in Table 13. In Table 14, I report that similar to counties in the top quartile of real regional wages, counties in the top quartile of labor force productivity also see a greater increase in their wage when increasing cluster strength by one percent. Furthermore, only quartile four has a statistically significant coefficient when I use an FE regression. When I quartile by patents, I find a different result. I report in Table 16 that only counties in the third quartile experience a positive and statistically significant increase in their real regional wage when cluster strength increases by one percent. This difference, however, likely stems from the large standard deviation for quartile four as I show in Table 15.

Through my examination of Table 11, 13, and 15, it is apparent that counties with greater wages, labor force productivity, and patents see a more significant benefit in real regional wage increases when increasing cluster strength. Incidentally, Chrisinger et al. (2015) find a similar result when looking at clusters in Washington state. Specifically, the authors report a “suggestive rather than definitive” result that some clusters negatively rather than positively affect the wage. While, as Chrisinger et al. (2015) suggest, researches do need to analyze this results further, I propose

Table 12: Quartile Panel Regression Results for Real Regional Wage

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Force Productivity (Thousands)	8.250**	31.329***	1.684	34.020***	8.514**	43.908***	12.217***	9.628***
	(3.426)	(9.955)	(1.473)	(10.019)	(3.244)	(10.841)	(4.476)	(1.910)
Cluster Strength	21.646***	−7.141**	−0.950	−0.919	−0.247	−0.281	74.607***	28.590**
	(3.782)	(3.464)	(1.632)	(2.528)	(1.951)	(2.917)	(11.019)	(13.252)
Percent Receiving High School Diploma +	15.899	146.852***	25.051***	240.186***	20.347*	296.097***	11.871	446.153***
	(24.320)	(18.961)	(8.453)	(36.006)	(12.081)	(43.580)	(55.134)	(80.269)
Percent with Some College or Associates Degree +	14.186	197.897***	39.018***	298.757***	39.430***	370.494***	−119.339	483.480***
	(16.772)	(27.524)	(11.202)	(30.965)	(12.295)	(44.771)	(73.388)	(90.826)
Percent Completing a Bachelor's Degree +	−39.786	−18.341	−46.080***	−1.770	−36.570***	25.027	264.184***	211.370**
	(26.686)	(32.672)	(10.108)	(38.585)	(12.693)	(51.398)	(96.632)	(95.159)
Patents	84.889***	111.574***	1.062	29.350**	8.620***	21.789***	4.251***	3.978***
	(18.053)	(14.910)	(1.075)	(13.934)	(2.428)	(4.859)	(0.497)	(0.309)
Population Density	8.104***	2.377***	−0.051	−1.818***	0.316***	0.352*	0.569***	0.726
	(2.410)	(0.731)	(0.401)	(0.533)	(0.084)	(0.189)	(0.170)	(0.895)
Establishments	1.635*	−2.257	−0.475**	−5.279***	−0.374***	−3.067***	−0.407***	−0.378
	(0.874)	(1.476)	(0.181)	(1.828)	(0.104)	(0.975)	(0.056)	(0.329)
Employment (Thousands)	−115.111*	19.721	49.449***	312.304***	28.725***	214.680***	26.410***	38.370***
	(59.256)	(125.875)	(18.230)	(61.743)	(6.779)	(51.317)	(3.693)	(5.905)
Constant	22164.783***	4244.675***	26446.745***	−4898.714	29557.498***	−13699.356***	35525.749***	−31736.070***
	(1692.216)	(1233.305)	(457.361)	(3110.417)	(842.071)	(4218.092)	(3703.036)	(7476.047)
State Fixed Effects	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
R-Squared	0.110	0.314	0.054	0.477	0.087	0.519	0.391	0.309
Observations	4374	4374	4349	4349	4322	4322	4304	4304

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 13: Labor Force Productivity Quartile Summary Statistics

	Observations	Mean	SD	Min	Max
1	4,415	36.21	6.67	11	46
2	4,415	53.10	4.19	46	60
3	4,415	69.15	5.61	60	80
4	4,414	141.48	184.43	80	3,692
Total	17,659	74.98	100.67	11	3,692

two potential explanations of why counties in the fourth quartile of each variable experience larger wage increases.

First, is a case of arithmetic. If wage increases by companies are done by percentages rather than a fixed value, then firms with higher paid workers will receive a larger wage increase from a five percent increase than those of a lower paying firm. Since counties needs high paying firms to have a high average wage, it makes sense that counties in the fourth quartile receive the greatest benefit from cluster strength as I report wages in levels instead of growth rates. Second, counties with higher wages, labor force productivity, and patents must comprise more skilled workers who are in a lower supply than non-skilled workers. Therefore, companies in the fourth quartile likely compete more for workers in this skilled labor pool, thus causing a higher increase in the average wage as companies bid up labor. I further touch on this increase in competition in the next subsection.

4.3 Implications, Limitations, and Areas for Future Research

While I demonstrate that cluster strength has a robust correlation with wages, I further explore the channel through which this increase occurs. I first test cluster strength's relationship with labor force productivity as I show in Table 17. I report a statistically significant increase ($p<0.05$) of $276 in the labor force productivity when cluster strength increases by one percent. However, this result does not remain when I use a more robust FE regression.

Table 14: Quartile Panel Regression Results for Labor Force Productivity

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Force Productivity (Thousands)	184.419***	198.184***	215.685***	177.480***	172.661***	155.005***	12.093***	12.813***
	(11.694)	(14.762)	(15.208)	(13.900)	(12.248)	(7.816)	(3.102)	(1.567)
Cluster Strength	16.708**	0.741	34.130***	3.812	42.257***	4.581	102.664***	47.561**
	(6.624)	(3.102)	(9.207)	(3.307)	(8.142)	(4.001)	(17.558)	(22.739)
Percent Receiving High School Diploma +	74.046*	58.872*	95.539*	118.022***	30.558	118.591***	29.460	311.552***
	(42.707)	(34.931)	(50.352)	(36.569)	(53.005)	(33.811)	(79.921)	(53.529)
Percent with Some College or Associates Degree +	−19.399	147.994***	19.005	167.998***	−36.310	206.135***	−139.729	389.949***
	(46.078)	(31.805)	(42.859)	(40.971)	(50.337)	(38.948)	(128.094)	(52.091)
Percent Completing a Bachelor's Degree +	24.447	31.863	−2.606	12.997	134.543***	7.563	390.829***	80.353
	(40.437)	(45.077)	(42.321)	(52.670)	(48.607)	(40.369)	(145.327)	(114.912)
Patents	4.995	39.005***	14.601***	12.930**	14.266***	6.535***	4.117***	4.098***
	(5.265)	(13.284)	(3.950)	(5.459)	(1.685)	(1.600)	(0.438)	(0.329)
Population Density	0.415***	0.064	0.307**	−0.228	1.768*	−0.784	0.871***	2.450
	(0.150)	(0.973)	(0.134)	(0.198)	(1.043)	(0.538)	(0.275)	(1.481)
Establishments	1.008	−1.414	−0.529*	−0.777	−0.188	−0.518	−0.253**	−0.370
	(0.985)	(1.424)	(0.286)	(0.616)	(0.256)	(0.433)	(0.095)	(0.446)
Employment (Thousands)	−51.834	194.033*	55.777**	63.675	27.112	81.695***	19.559***	36.145***
	(75.955)	(105.790)	(26.550)	(59.092)	(17.900)	(18.494)	(5.619)	(8.871)
Constant	14490.226***	8255.749***	8253.836**	2801.078	14686.279***	102.428	27181.035***	−15956.250***
	(2277.013)	(2243.210)	(3371.935)	(2555.865)	(3486.653)	(2700.941)	(5590.833)	(5932.211)
Fixed Effects	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
R-Squared	0.198	0.336	0.236	0.415	0.363	0.499	0.415	0.254
Observations	4361	4361	4345	4345	4331	4331	4312	4312

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 15: Patents Quartile Summary Statistics

	Observations	Mean	SD	Min	Max
1	6,376	0.00	0.00	0	0
2	2,460	0.44	0.21	0	1
3	4,419	2.42	1.30	1	6
4	4,416	135.61	517.35	6	14,333
Total	17,671	34.56	265.10	0	14,333

Next, in Table 18, I test the relationship between cluster strength and patents to determine if this drives the wage increase. I report that cluster strength has no statistically significant relationship with patents for either my OLS or FE regression. However, incidentally, I find a similar result to Gössling and Rutten (2007), who, find that population density negatively correlates with patents, which I show in column one.

I attribute the absence of a statistically significant results between cluster strength and labor force productivity and patents to the limited years in my data set. This is to say, it is unlikely that in increase in cluster strength will cause any growth in labor force productivity or patents as both of these variables take multiple years to develop. However, I find that this result—or lack thereof—provides an interesting view into the short- and long-run economic implications of clusters.

Because I report, in the short run, that cluster strength does not affect labor force productivity or patents and that I include population density in my regressions, I assume, based on my literature review and conceptual framework that competition increases are the primary channel for which cluster strength increases wages in the short run. This assumption differs from Porter (2003), who predicts wage increases from cluster strength happen through increases in productivity. Furthermore, unlike Porter (2003), I find a more modest relationship between cluster strength and wages. Specifically, Porter finds that an increasing the share of traded employment in strong clusters by one increases the nominal average private wage by \$102.38. However, as

Table 16: Quartile Panel Regression Results for Patents

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor Force Productivity (Thousands)	17.507***	13.218***	50.405***	51.260**	59.808***	45.436**	144.444***	103.688***
	(4.150)	(2.114)	(14.107)	(20.864)	(14.312)	(18.469)	(15.048)	(19.446)
Cluster Strength	66.287***	17.180	32.624**	7.798	34.466***	21.697**	35.681**	−2.969
	(14.107)	(10.437)	(12.622)	(6.475)	(11.975)	(9.248)	(15.184)	(4.504)
Percent Receiving High School Diploma +	14.978	184.750***	114.808**	243.688***	125.517***	356.700***	−51.516	311.587***
	(45.324)	(30.858)	(46.565)	(66.664)	(35.647)	(52.912)	(101.490)	(89.170)
Percent with Some College or Associates Degree +	119.983**	305.482***	24.965	242.196**	14.804	295.836***	−70.145	400.655***
	(50.091)	(40.115)	(67.542)	(103.149)	(51.547)	(47.969)	(86.538)	(61.381)
Percent Completing a Bachelor's Degree +	−98.547	−50.354	−108.112	−15.256	−42.292	17.247	274.822***	237.117***
	(63.711)	(54.760)	(83.062)	(80.898)	(73.904)	(62.141)	(88.637)	(79.131)
Patents	0.000	0.000	837.529*	1368.684***	541.131***	199.900***	3.582***	3.230***
	(.)	(.)	(468.350)	(192.135)	(105.740)	(31.083)	(0.548)	(0.319)
Population Density	−4.009	2.541***	−5.533**	−0.658	−2.614*	−5.181***	0.053	−0.084
	(4.038)	(0.710)	(2.323)	(1.975)	(1.395)	(0.617)	(0.202)	(0.555)
Establishments	−10.821***	−7.144	−6.006***	−3.591	−2.332**	−1.734	0.061	0.076
	(3.373)	(10.706)	(1.733)	(4.461)	(0.943)	(2.501)	(0.089)	(0.231)
Employment (Thousands)	903.061***	2141.343***	518.615***	943.997***	205.086***	493.826***	−1.298	5.035
	(233.465)	(772.283)	(118.210)	(338.892)	(54.280)	(135.528)	(5.769)	(7.986)
Constant	21159.811***	−4003.781	17067.045***	−8429.309	14251.162***	−21851.665***	25677.354***	−28185.852***
	(2825.954)	(3228.789)	(2894.186)	(6494.821)	(2762.276)	(5886.460)	(6986.527)	(7312.666)
Fixed Effects	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*
R-Squared	0.243	0.240	0.308	0.326	0.317	0.405	0.693	0.593
Observations	6268	6268	2628	2628	4131	4131	4322	4322

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 17: Panel Regression Results for Labor Force Productivity

	Labor Force Productivity	
	(1)	(2)
Cluster Strength	0.276***	0.127
	(0.102)	(0.095)
Percent Receiving High School Diploma +	−2.693	1.747***
	(2.130)	(0.515)
Percent with Some College or Associates Degree +	2.223**	1.774***
	(1.022)	(0.397)
Percent Completing a Bachelor's Degree +	−0.571	1.098
	(0.416)	(1.331)
Patents	0.005**	0.005**
	(0.002)	(0.003)
Population Density	0.004***	0.004
	(0.001)	(0.004)
Establishments	−0.005***	−0.003
	(0.001)	(0.003)
Employment (Thousands)	0.311***	0.277**
	(0.039)	(0.123)
Constant	190.327	−188.424**
	(139.914)	(89.575)
State Fixed Effects	*No*	*Yes*
R-Squared	0.031	0.032
Observations	17349	17349

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

I previously mention, I report an increase in the nominal average private wage of $64.99 or $15.07 when using an OLS and FE regression, respectively.

I attribute this difference in results to five factors: (1) I analyze cluster strength and wages from 2009 to 2014, while Porter (2003) only looks at the year 2000; (2) I use counties for my observations due to the increase in observations, while Porter uses economic areas; (3) I provide multiple control variables to ensure cluster strength is increasing the wage and not an unobservable variable; (4) my strong cluster's location quotient is in the top 75 percent of regions in the U.S., while Porter's cutoff is an

Table 18: Panel Regression Results for Patents

	Patents (1)	Patents (2)
Labor Force Productivity (Thousands)	0.022	0.010
	(0.022)	(0.016)
Cluster Strength	0.359	-0.007
	(0.270)	(0.027)
Percent Receiving High School Diploma +	-1.180	-0.322
	(1.114)	(0.263)
Percent with Some College or Associates Degree +	-1.217^{**}	0.401^{***}
	(0.556)	(0.086)
Percent Completing a Bachelor's Degree +	5.548^{**}	1.334^{**}
	(2.192)	(0.626)
Population Density	-0.029^{***}	-0.035
	(0.009)	(0.035)
Establishments	-0.007	0.036^{**}
	(0.018)	(0.015)
Employment (Thousands)	1.440	3.780^{**}
	(1.274)	(1.536)
Constant	35.674	-201.554^{***}
	(49.597)	(52.329)
Fixed Effects	*No*	*Yes*
R-Squared	0.300	0.185
Observations	17349	17349

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

location quotient ≥ 0.80; (5) my data include 51 cluster definitions while Porter (2003) uses 41. By adding these differences in my analysis, I provide further credibility to Porter's original results and deliver a more robust relationship between clusters and wages for policymakers.

Regarding policy implications, given my results, I do not recommend sweeping increases in policy instruments to increase cluster strength (e.g., elimination of restrictions on competition) but rather selecting targeted policies to increase the counties wage while pursuing cluster development in the background. As an example, a county could pursue some sort of means tested income policy, such as housing subsidies. Therefore, I encourage counties to understand their position in my model; thus, not using cluster strength as a sweeping policy instrument but rather an aid to other policies.

While my paper aims to provide a more robust result for the relationship between clusters and wages, it still contains limitations. One limitation is that my data do not directly separate strong clusters in the cluster strength variable. Separating by clusters would allow me to have another layer of granularity in the data and determine if a considerable share of the wage increase is arriving from specific clusters. Additionally, my data time range is relatively low due to the limited availability of the data; a more extensive data time range would provide more robust results and allow me to understand long-run implications. Finally, there are certain control variables that would add to the robustness of the paper that I could not include—specifically, a competition index and a cost-of-living index. At the same time, I try to include variables that correlate to these indexes (i.e., number of establishments and population density). However, having an actual index would provide more robust results.

I suggest two specific recommendations for future research based on the limitations and findings of my paper. The first is to incorporate a data set that includes the share of clusters in the cluster strength variable. Future researchers could accomplish this by using the cluster classifications from Delgado et al. (2016) and reconstructing the cluster strength variable while understanding which clusters make up the largest share. This would provide a more robust result as it would better determine which clusters policymakers should pursue to increase wages and which they should avoid. The second is to provide more extensive research on how cluster strength impacts wages. To my knowledge, the literature has yet to demonstrate why cluster strength empirically increases wages. Although I provide a possible explanation, that is, that cluster strength drives competition, thus increasing wages, I am unable to empirically test my explanation.

5 Conclusion

In this thesis, I aim to answer two research questions regarding cluster strength's effects on wages to fill the gap in the literature. First, after introducing robustness tests, does cluster strength still positively affect wages as Porter (2003) finds? Second, are increases in productivity the primary driver through which cluster strength affects wages. Two answer these research questions, I put fourth two hypothesis, which I construct through my review of the academic literature on clusters, productivity, and wages as well as my conceptual framework. Given that multiple researchers show that clusters provide a plethora of economic advantages to the firms that reside in the cluster (e.g., increases in productivity, patents, exports), I anticipate that after including robustness tests, cluster strength will still positively affect wages. However, while clusters may positively affect the wage through increased productivity, I predict through my conceptual framework that clusters also increase competition, thus moving the average wage closer to the marginal product of labor.

I construct a panel data set comprising multiple economic variables for U.S. counties over the period 2009 to 2014. Within this data set is my primary variable of interest, cluster strength, which I collect from U.S. Cluster Mapping (2020), which defines cluster strength as the "percentage of trade labor in a strong cluster." To construct this variable, U.S. Cluster Mapping (2020) identifies every strong cluster by county, where a strong cluster is any cluster with a location quotient—employment specialization—in the top 25 percent of the U.S. U.S. Cluster Mapping (2020) then calculates the percent of trade labor (i.e., labor that produces goods and services for use outside their region) that work within strong clusters for that county. Furthermore, my main dependent variable for my empirical model is the real regional wage, which I construct by deflating the average private wage by the regional consumer price index.

To empirically test my two hypotheses, I exploit the heterogeneous performance of U.S. counties through three panel regressions: OLS, 2SLS, and FE. Using multiple regressions, I attempt to eliminate endogeneity from my model to test whether cluster strength still positively affects wages when including robustness tests. Similar to Porter (2003), I report that cluster strength and wages are statistically significant and positively correlate when using an OLS regression. Specifically, I find that when cluster strength increases by one percent, the real regional wage increases by \$63.87 when using an OLS regression. However, when I use a 2SLS model to eliminate endogeneity, with $River_{Exp}$ as my instrumental variable, I do not find a statistically significant result between cluster strength and wages. As a secondary robustness test, I use an FE regression, where I report a statistically significant increase of \$14.42 in the real regional wage when cluster strength increases by one percent.

Furthermore, I find no correlation between cluster strength and labor force productivity or patents when using an FE regression. Therefore, I conclude that in the *short run*, increases in productivity or innovation are not the primary channels through which cluster strength affects wages. Rather, given my conceptual framework, I determine that wage increase in the short run stem from escalations in competition between firms. This determination differs from Porter (2003), who predicts wage increases stem from clusters increasing productivity.

My findings on how cluster strength affects wages further demonstrates the importance of clusters for economic regional development. Given the positive relationship I find between cluster strength and wages, and the plethora of researchers who document the positive externalities clusters offer, polices that accelerate cluster development seem appropriate. These policies, as Porter (2000) mentions, should focus on eliminating hurdles and restrictions on firms (e.g., infrastructure constraints) while also furthering the education of the benefits of clusters to firms. In doing so, policymakers should aim to enact policy that not only benefits one firm or industry, but rather the entire business environment of the cluster. Therefore, as Porter (2000) mentions, policymakers will be fulfilling their role in establishing an environment that supports economic growth.

6 References

Abdesslem, A. B. and Chiappini, R. (2019). Cluster policy and firm performance: a case study of the french optic/photonic industry. *Regional Studies*, 53(5):692–705.

Barkley, D. L. and Henry, M. S. (1997). Rural industrial development: To cluster or not to cluster? *Review of Agricultural Economics*, 19(2):308–325.

Becchetti, L., De Panizza, A., and Oropallo, F. (2007). Role of industrial district externalities in export and value-added performance: Evidence from the population of italian firms. *Regional studies*, 41(5):601–621.

Bureau of Economic Analysis. Gdp by county, metro, and other areas. Electronic.

Bureau of Economic Analysis. Table 6.2c. compensation of employees by industry. Electronic.

Bureau of Economic Analysis. Table 6.3c. wages and salaries by industry. Electronic.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.

Chrisinger, C. K., Fowler, C. S., and Kleit, R. G. (2015). Industry clusters and employment outcomes in washington state. *Economic Development Quarterly*, 29(3):199–210.

Cingano, F. and Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2(4):720–744.

Davidson, R. and MacKinnon, J. (1993). *Estimation and inference in econometrics*. Oxford University Press, New York.

Delgado, M., Porter, M. E., and Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, 43(10):1785–1799.

Delgado, M., Porter, M. E., and Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, 16(1):1–38.

Feldstein, M. (2008). Did wages reflect growth in productivity? *Journal of Policy Modeling*, 30(4):591–594. Growth, Productivity, and Wages in the U. S. Economy.

Feser, E. and Isserman, A. (2005). Clusters and rural economies in economic and geographic space. *Urbana-Champaign: University of Illinois*.

Fowler, C. S. and Kleit, R. G. (2014). The effects of industrial clusters on the poverty rate. *Economic Geography*, 90(2):129–154.

Glaeser, E. L. and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of economic literature*, 47(4):983–1028.

Glaeser, E. L. and Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.

Gössling, T. and Rutten, R. (2007). Innovation in regions. *European planning studies*, 15(2):253–270.

Hoffer, D. and Prewitt, E. (2018). To hire or not to hire: Us cartel enforcement targeting employment practices. *CONCURRENCES COMPETITION L. REV.*, 3:78–79.

Howard, E., Newman, C., Rand, J., and Tarp, F. (2014). Productivity-enhancing manufacturing clusters: Evidence from vietnam. WIDER Working Paper 2014/071, Helsinki.

Huang, K.-F., Yu, C.-M. J., and Seetoo, D.-H. (2012). Firm innovation in policy-driven parks and spontaneous clusters: the smaller firm the better? *The Journal of Technology Transfer*, 37(5):715–731.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly journal of economics*, 108(3):577–598.

Judzik, D. and Sala, H. (2013). Productivity, deunionization and trade: Wage effects and labour share implications. *International labour review*, 152(2):205–236.

Kim, M.-K., Harris, T. R., and Vusovic, S. (2009). Efficiency analysis of the us biotechnology industry: Clustering enhances productivity. *Agbioforum (Columbia, Mo.)*, 12(3-4):422–436.

Klepper, S. (2010). The origin and growth of industry clusters: The making of silicon valley and detroit. *Journal of Urban Economics*, 67(1):15–32. Special Issue: Cities and Entrepreneurship.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Lambert, T. E., Mattson, G. A., and Dorriere, K. (2017). The impact of growth and innovation clusters on unemployment in us metro regions. *Regional Science Policy & Practice*, 9(1):25–37.

Lawrence, R. Z. (2016). Does productivity still determine worker compensation? domestic and international evidence. *THE US LABOR MARKET-QUESTIONS AND CHALLENGES FOR PUBLIC POLICY*, pages 42–62.

Lin, B. and Sai, R. (2022). Has mining agglomeration affected energy productivity in africa? *Energy*, 244:122652.

Marshall, A. (1890). *Principles of Economics*, volume 1. Macmillan.

Martin, P., Mayer, T., and Mayneris, F. (2011). Public support to clusters: A firm level study of french "local productive systems". *Regional Science and Urban Economics*, 41(2):108–123.

Mohnen, P. and Hall, B. H. (2013). Innovation and productivity: An update. *Eurasian Business Review*, 3(1):47–65.

Montuenga-Gómez, V., Fernández, M., and Romeu, A. (2007). The link between wages and productivity in spain. *International Review of Applied Economics*, 21(2):247 – 272.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–75.

Mosur, J. and Posner, E. A. (2023). Horizontal collusion and parallel wage setting in labor markets. *The University of Chicago law review*, 90(2):545–578.

Muehlemann, S., Ryan, P., and Wolter, S. C. (2013). Monopsony power, pay structure, and training. *ILR Review*, 66(5):1097–1114.

Nicholson, W. (1995). *Microeconomic theory : basic principles and extensions*. The Dryden Press series in economics. Dryden P., 6th ed.. edition.

Olley, S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry.

Ouimet, P. and Zarutskie, R. (2011). Acquiring labor. *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-32*.

Porter, M. (2003). The economic performance of regions. *Regional studies*, 37(6-7):549–578.

Porter, M. E. (1990). The competitive advantage of nations. Magazine.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly*, 14(1):15–34.

Sadeghi, A., Talan, D. M., and Clayton, R. L. (2016). Establishment, firm, or enterprise: does the unit of analysis matter. *Monthly Lab. Rev.*, 139:1.

Stansbury, A. M. and Summers, L. H. (2017). Productivity and pay: Is the link broken? Working Paper 24165, National Bureau of Economic Research.

Stojčić, N., Anić, I.-D., and Aralica, Z. (2019). Do firms in clusters perform better? lessons from wood-processing industries in new eu member states. *Forest Policy and Economics*, 109:102043.

Strain, M. R. (2019). The link between wages and productivity is strong. *Expanding Economic Opportunity for More Americans, The Aspen Institute.*

U.S. Cluster Mapping (2020). Institute for Strategy and Competitiveness, Harvard Business School (http://clustermapping.us). Electronic.

Yankow, J. J. (2006). Why do cities pay more? an empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, 60(2):139–161.

Table A1: Trade Employment Data, 2014

Cluster Name	**Brookings, SD Employment**	**U.S. Employment**	**Strong Cluster**	**National Rank**
Downstream Metal Products	810	402,823	TRUE	134
Medical Devices	750	254,115		87
Production Technology and Heavy Machinery	435	978,399	TRUE	592
Hospitality and Tourism	373	3,106,368		938
Business Services	367	11,572,491		1,133
Distribution and Electronic Commerce	311	5,616,900		1,471
Food Processing and Manufacturing	235	982,745		847
Financial Services	200	1,892,564		723
Recreational and Small Electric Goods	175	152,085		230
Plastics	175	674,103		781
Transportation and Logistics	172	1,617,134		1,051
Education and Knowledge Creation	148	3,046,194		863
Electric Power Generation and Transmission	120	150,379	TRUE	306
Marketing, Design, and Publishing	89	1,322,741		840
Lighting and Electrical Equipment	70	283,062		575
Upstream Chemical Products	60	175,777		612
Biopharmaceuticals	60	236,046		301
Insurance Services	40	1,511,440		825
Vulcanized and Fired Materials	20	241,281		890
Printing Services	20	485,809		1,093
Oil and Gas Production and Transportation	20	736,804		1,297
Information Technology and Analytical Instruments	20	1,057,686		878
Construction Products and Services	20	807,294		2,087
Apparel	20	128,270		573
Wood Products	10	342,947		1,917
Video Production and Distribution	10	245,380		730
Trailers, Motor Homes, and Appliances	10	127,497		492
Performing Arts	10	346,436		1,365
Paper and Packaging	10	351,976		948
Nonmetal Mining	10	80,874		1,347
Livestock Processing	10	473,520		1,234
Communications Equipment and Services	10	423,190		1,199
Agricultural Inputs and Services	10	97,440		1,571
Water Transportation	0	309,444		
Upstream Metal Manufacturing	0	399,936		
Tobacco	0	14,557		
Textile Manufacturing	0	190,070		
Music and Sound Recording	0	27,477		
Metalworking Technology	0	488,833		
Metal Mining	0	45,025		
Leather and Related Products	0	30,654		
Jewelry and Precious Metals	0	25,652		
Furniture	0	328,265		
Forestry	0	64,674		
Footwear	0	14,793		
Fishing and Fishing Products	0	38,157		
Environmental Services	0	85,995		
Downstream Chemical Products	0	239,739		
Coal Mining	0	82,946		
Automotive	0	892,726		
Aerospace Vehicles and Defense	0	532,330		
Total	4,800	43,733,043		

Table A2: Correlation Matrix

	(1)									
	RFLD_EXP_AREA	real_regional_w	cs	lfp	h	c	cm	pat	pop_dens	establishments	employment
RFLD_EXP_AREA	1.00										
real_regional_w	-0.01	1.00									
cs	0.06***	0.22***	1.00	0.06***							
lfp	-0.03***	0.34***	0.06***	1.00							
h	-0.10***	0.21***	0.03***	-0.04***	1.00						
c	-0.05***	0.36***	0.07***	0.06***	0.78***	1.00					
cm	-0.07***	0.44***	0.09***	0.06***	0.61***	0.87***	1.00				
pat	-0.01	0.35***	0.05***	0.05***	0.08***	0.20***	0.28***	1.00			
pop_dens	-0.03***	0.29***	0.03***	0.07***	0.02**	0.13***	0.22***	0.14***	1.00		
establishments	0.03***	0.40***	0.02**	0.06***	0.10***	0.27***	0.37***	0.51***	0.43***	1.00	
employment	0.02*	0.42***	0.02**	0.07***	0.09***	0.26***	0.36***	0.52***	0.45***	0.98***	1.00

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$