The Determinants of Agglomeration in Health Care Sector Employment in US Metropolitan areas.

Paper Presented on 41st Annual and Conference Mid-Continent
Regional Science Association (MCRSA), June 3-5, 2010
St. Louis, MO

Bhawani P. Mishra
University of Missouri Columbia

The Determinants of Agglomeration in Health Care Sector Employment in US Metropolitan areas.

Bhawani P. Mishra[1]

Abstract: *There is a tendency of industries to co-locate together in order to reap the benefits of externalities. The basic objectives of this study to identify the factors affecting agglomeration of health care sector in US cities. Due to differences in the nature of industries factors affecting agglomeration are likely to be different for different industries. Therefore this study has considered investigating on single health care sector. Moreover, this study used the panel data model in order to capture both cross-section and temporal dimension of the agglomeration. This study found that factors like local competition, population, input availability, and state research expenditure on health care are significant for the agglomeration of health care sector. This study has also identified the fixed effect panel data model for identifying the factors affecting agglomeration in health care sector.*

1. Introduction:

Cities are the home of millions of people and their concentrated economic activities. City formation itself describes the process of agglomeration[2] where concentration of economic activities built-up along with the time in a relatively small area. City helps to reduce the transaction costs and facilitates to knowledge spillover (Glaeser, 2010). Reduction on transaction costs and facilitation of knowledge spillover have positive impact due to externalities (Puga 2009; Ellion and Glaeser, 1999). Agglomeration economies can be localization economies i.e. economies arise from the factors within the industry and urbanization economies i.e. economies arise from the factors outside of industry. There have been efforts in the past to determine the factors affecting agglomeration. Most of those studies concluded that the labor, transportation cost, proximity to inputs or output and specific characteristics of cities are the major factors contributing to agglomeration (Marshall, 1920, Ellion and Glaeser, 1999 and Koo, 2005). Additionally, there have also been efforts to figure out the appropriate model for identifying factors that affect agglomeration. Most of those studies were done by taking all of the sectors of the economy. Since the nature of industries are different from each other i.e. some industries are manufacturing some industries are service oriented. Therefore, there is a need of study to identify

[1] Department of Agriculture and Applied Economics, University of Missouri, Columbia, e-mail: bpm42c@mail.missouri.edu

[2] In this paper agglomeration and clustering is used interchangeably.

the factors for the particular sector of an economy. Visualizing that gap, this study has tried to contribute on identifying factors that affect to health care sectors' agglomeration with an appropriate model. In this study, health care sector is defined as broadly that includes four major sub-sectors Ambulatory Health Care Services, Hospitals, Nursing and Residential Care Facilities and Social Assistance. The definition of health care sector is defined as in North American Industry Classification System (NAICS) that included above four subsectors.

There are three rationales of doing this paper. Firstly, majority of studies done so far are based on the cross sectional data, there is no inclusion of temporal dimension of the agglomeration in the model. In this study panel data are used in order to capture both dimensions i.e. temporal and cross-sectional dimensions of agglomeration. Secondly, previous studies were done based on Standard Industrial Classification (SIC) codes, which is older version of industrial classification system of industries. But, County Business Pattern (CBP) introduces data set under the NAICS from 2001. These two SIC and NAICS coding systems do not match exactly. Besides, there has been lot of technological transformation going on in the health care industry over the years; therefore there is a need of updating the knowledge on agglomeration under NAICS regime specifically within the specific sector. Lastly, since the natures of industry are different from one another some are manufacturing oriented while others are service oriented. Factors affecting to agglomeration would likely to be different based on the nature of the industry. In some industries inputs occupies major of portion of the total cost whereas in other industries supply transaction cost occupies major portion of the total cost. Based on those natures, some industries agglomerate near to the input markets and others industries agglomerate to the output markets; therefore there is a need to indentify drivers of agglomeration for health care sector. The reason of taking health care sector is because health care sector is the most important sector of the US economy.

The major objectives of the study are to investigate the factors affecting the health care industry agglomeration in the Metropolitan Statistical Area (MSA[3]) of lower 48 states of United States. Specifically, following objectives are analyzed in this study.

- Identify the factors affecting health care sector agglomeration in the US cities.
- Examine pattern of agglomeration of the health care sector in the US cities.

[3] Metropolitan Statistical Area includes regions with large city at the core. But, in this study MSA and city are used interchangeably.

- Identifying the appropriate model for determining factors affecting health care sector agglomeration.

2. Conceptual Framework

According to Beckmann (1999) following simple illustration with very simple assumptions is enough to visualize the general tendency of firms to concentrate in particular place (figure -1). Let's take two firms X and Y and their markets are M_1 and M_2 equal size. For the sake of simplicity, let's say markets arranged in a linear fashion and there is only transportation cost involved and other costs are holding constant for both firms. Firm can earn more profit if it can sell to both of the markets and assumed both firms are operated at same technology. But, transportation cost increases with the distance increase. In the current arrangement, due to the proximity to the respective markets X firm will enjoy market M_1 and Y will enjoy market M_2. If the firm X move slightly toward M_2 market still it is can capture M_1 and able to reduce the distance M_2 market thereby lowering the transportation cost. Similarly, firm Y has also incentive to move toward the M_1 market in order to lower the transportation cost without losing the market M_2. Ultimately, X will move toward the M_2 market and Y will move toward the M_1 until they meet together. Once firm X and Y in same place there is no further incentive to move away from each other.

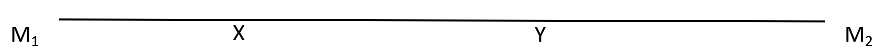


Figure 1: Relative position of firms in linear markets arrangement.

If any one of two firms moves from that equilibrium places, one will lose the market of opposite direction of the movement. Therefore, they do not have incentive to move away from each other. These firms do not have to be agglomerated exactly at the middle of the line. The positions of X and Y are determined by the market size and price of two markets, transportation costs and other specific characteristics and facilities of particular city. This illustration explains the simple supply side story of agglomeration; however this intuition can be applied to explain the demand side as well. In the demand side, firms try to minimize the cost when firm is not a price taker in its output market (competitive market). As long as the firm is a price taker cost minimization would be the strategy of firm to get higher profit.

Let's take other example to explain the agglomeration phenomenon with two inputs and one output center. This process of location decision can be explained by the location triangle which is commonly-known Weber-Moses approach. In fact Weber-Moses approach is used to explain the location decision of a firm. But the intuition can be extended for many firms. This is another simple version of model; however once this process conceptualized more complex situation can be explained by this conceptualization Weber-Moses approach. In Weber-Moses theory of location, two input sources and one output market are located at the vertices of a triangle (figure-2).

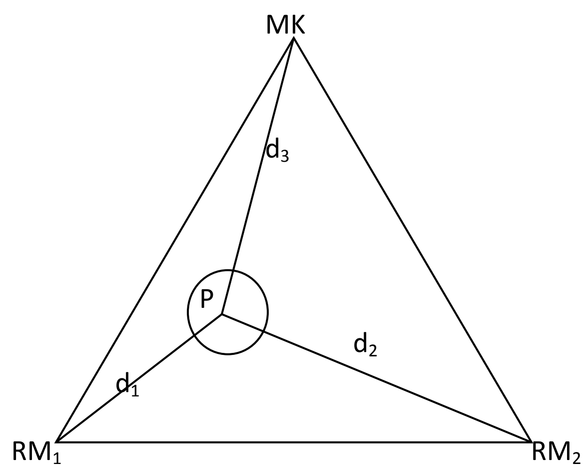


Figure 2: Optimal firm location in location triangle

In the figure above, there two input sites i.e. RM_1 and RM_2 whereas MK is the point of consumption market and d_1, d_2 and d_3 is the distance to input and output markets. P is the strategic location for the firms in which transportation costs for inputs and outputs are minimized. The strategic position of P is determined by the distance and transportation costs of inputs and output markets. In the above drawn picture, transportation costs of inputs from RM_1 site are relatively higher than those of input site RM_2 and output site MK. Therefore, P is located more toward the RM_1 than other markets. Let's say other firm tries to locate in that area. If that new firm has backward or forward linkage with current firm P then it will come near to the firm in order to minimize the cost of transaction. If the new firm is same firm like earlier in location P it has to also locate near the firm P. The mechanism of this process is discussed below with Hotelling spatial game model.

It is easier to understand the agglomeration of interdependent industries (backward and forward linkage) but the question arise how same kind of firms with homogenous product wanted to be co-locate together. Hotelling spatial game model answer that question. Let us assume there is a beach, which is linear as shown below from [-1, 1] (figure-3). Two ice cream sellers, with homogenous quality of ice-cream, wanted to sell their product along side of beach. For the simplicity let's assume that consumers are distributed evenly along the line of beach.

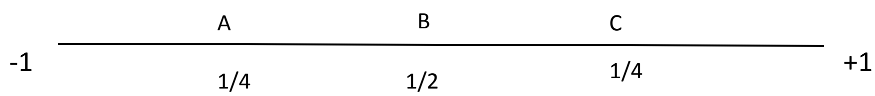


Figure 3: Ice-cream Sellers Location Decision in Linear Market.

If the first seller allowed choosing the place, he/she may choose any place along the line. Everybody has to come to that seller and buy the ice-cream. Let's assume he /she chose the position A and next seller will choose slightly right to the A so that second seller will capture almost 75 percent of the market. If first seller is allowed to move again he/she will choose right of the second seller ultimately these two sellers will end up at the position B. This is the equilibrium position for them. However, this position is not the socially optimal position. Social optimal position is achieved if first seller is on position A and second seller is on position C. In these positions, all consumers have to walk just one-fourth of the total distance. This is how two same firms co-locate together when there is competition occurs.

3. Relevant Literatures

The earlier contribution on agglomeration was made by *Weber, von Thünen, Christaller, Isard* through the location theory. Marshall (1920) specifically described the determinants of agglomeration economies that arise from the concentration of economic activities. He suggested that there are three causes of localization of economic activities. These three causes are input sharing, labor pooling and knowledge spillover. These three factors can also be summarized functionally as sharing, matching and learning (Duranton and Puga, 2004). In a city upstream and downstream firms co-locate together in order to reduce the transaction cost between these two cities. Besides, there is a sharing of indivisible goods and facilities which is non-tradable between cities, which ultimately triggers the process of agglomeration in particular location

(Marshall, 1920). Higher population creates the pool of labor in which there is a better match between an employer's needs and a worker's skills (Ellion and Glaeser, 1999). It also reduces risk for both employer and employee. Spillovers of knowledge also enhance the economies of scale which ultimately triggers the process of localization. This allows workers to learn from each other. Urban economist also viewed that cities formation itself is the most significant manifestation of the agglomeration economies i.e. of concentrated human settlements, intended to share social overhead (Ellion and Glaeser, 1999).

Krugman (1999) also attempted to explain uneven distribution of the economic activities in the due to tug of wars between centripetal and centrifugal forces. He has given the list of centripetal and centrifugal forces as following (table-1). Centripetal forces positively contribute to agglomeration whereas centrifugal forces negatively contribute to the agglomeration. Krugman admits that the list of these variables is not comprehensive; it is merely a selection of some forces that may be important in practice.

Table 1: Forces Affecting Geographic Concentration

Centripetal force	**Centrifugal force**
Market size effect (linkage)	Immobile factors
Thick labor market	Land rents
Pure external economies	Pure external diseconomies

Source: Krugman, 1999. (http://irx.sagepub.com/cgi/reprint/22/2/142)

The centripetal force mentioned in the first column represents the Marshallian sources of external economies. According to his explanation a large market size creates the backward and forward linkages and thick labor market is supported by concentration economic activities. It is because workers easily find employers and employers easily find the workers. Moreover, Krugman further explained that a local concentration of economic activity may create more or less pure external economies through information spillovers. In the centrifugal forces lists, immobile factors (negative) contribute to the dispersion of activities, similarly concentration of economic activity increases land rent thereby discouraging the further concentration and pure external diseconomies such as congestion can be created by the concentrations of activity.

According to Glaeser (2010) due to recent advancement of transportation and communication transportation costs has significantly reduced. Therefore transportation costs has little role (but significant) in determining the agglomeration of the industries. However, in the health care industry distance in term of cost may not be important but distance in terms of time is crucial for health care industry. In summary, Glaeser wanted to show that location decision of the firm is basically determined by weighing the factors like backward/forward linkage, local competition and local non-tradable inputs and conditions together rather than just distance.

According to McDonald and McMillen (2007) cities are centers of diversified services, production and specialized services. Due to diversified and specialized services of cities agglomeration of economies realized in cities, firms try to concentrate their activities around cities. These authors also categorized the agglomeration of economies into three categories i.e. urbanization economies, industrialization economies and localization economies. In urbanization of economies, the benefits derived from the agglomeration of population, common infrastructures, availability of labor and market size. But, in industrialization economies the benefits derives from the agglomeration of industrial activities, such as being suppliers or customers and activities near a specific facility such as university, transport terminals, or government institutions respectively. Localization economies are external to the firm but internal to its industry. Localization of economies is limited to the geographic extent and should not extend from the central city locations to the suburbs or vice versa, whereas urbanization economies can extend beyond the boundary of metropolitan area (Rosenthal and Strange, 2003).

Health care sector is one of major sectors of the US economy in which advanced technologies have been introduced in order to improve the quality of services. People might suspect that distance play little role on determining the agglomeration in this sector due to introduction of advance technologies. But, even in a situation of electronic transmission of much information, physical location matters for knowledge flows because electronic contacts have been found to complement rather than substitute for face-to-face encounters (Gaspar and Glaeser, 1998). For example close physical contact is more important for the knowledge flows in the hospital services industry because many medical procedures require to have visually demonstrated. There have been examples of joint ventures and strategic alliance between the hospitals. According to Bates and Santerre (2005) Day Kimball Hospital and Backus Hospital, both located in eastern Connecticut, agreed to share the cost of mobile MRI unit. Not always agglomeration contributes positively to the productivity of industries. Some empirical researchers have found that an evidence of negative impact on productivity once an increased number of hospitals in the same area (Bates and Santerre, 2005). Too many hospitals in same area unnecessarily compete on cosmetic quality items rather than actual productivity (Bates and Santerre, 2005). Such competition makes hospital to engage on a " medical arms race" thereby spend unnecessarily on items such as cosmetic quality improvements, cost-enhancing technologies, and duplicate facilities as a way of attracting more physicians and patients (Robinson and Luft, 1985).

Most of the previous studies have found that Marshallian three factors i.e. input sharing, labor pooling and knowledge spillover were the most important factors to determine the agglomeration. Moreover, most of these papers have used cross-sectional model in order to determine the factors. But, later Koo (2005) found that there is endogenity problem in the cross sectional model and he used the three-stage least squares (3SLS) in order to resolve that problem. He introduced the following simultaneous system of two equations model to determine the factors affecting agglomeration.

$$\boldsymbol{FA_{ij} = \beta_0 + \beta_1 KS_{ij} + \beta_2 LP_{ij} + \beta_3 INP_{ij} + \beta_4 P_j + \beta_5 D_i + \varepsilon_{ij}} \dots\dots\dots\dots..(1)$$

Where, FA_{ij} is agglomeration of industry i in region j, KS_{ij} is knowledge spillover created by industry i in region j, LP_{ij} is labor pooling for industry i in region j, and INP_{ij} is input availability for industry i in region j, P_j is population in region j, and D_i is an industry dummy variable.

$$\boldsymbol{KS_{ij} = \alpha_0 + \alpha_1 FA_{ij} + \alpha_2 SE_{ij} + \alpha_3 SP_{ij} + \alpha_4 DV_j + \alpha_5 LC_{ij} + \alpha_6 D_i + \mu_{ij}} \dots (2)$$

Where, SE_{ij} is the percentage of small establishments in industry i in region j, where KS_{ij} is knowledge spillover created by industry i in region ji, FA_{ij} is agglomeration of industry i in region j, SE_{ij} is the percentage of small establishments in industry i in region j, SP_{ij} is specialization of industry i in region j, DV_j is economic diversity of region j, LC_{ij} is the level of local competition of industry i in region j, and D_i is a dummy variable included to capture industry-specific effects.

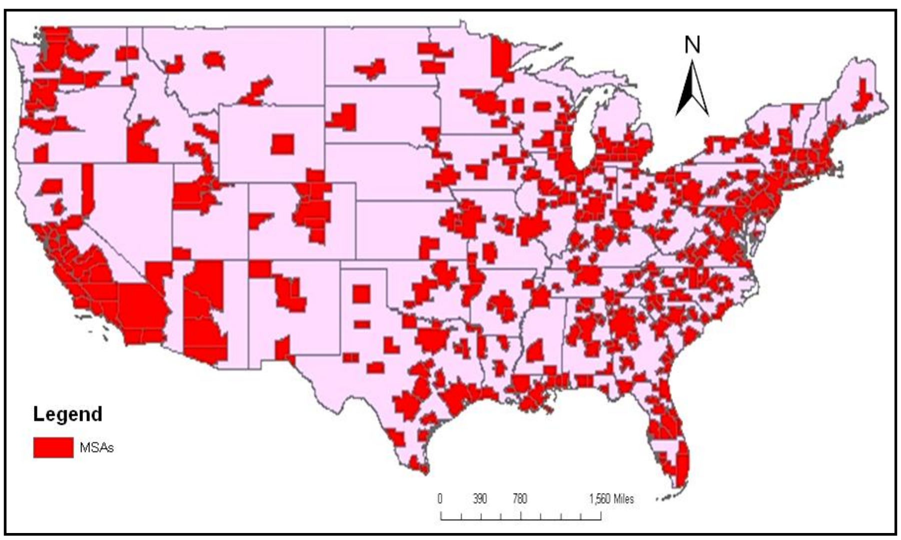
Simultaneous model does not explain the temporal dimension of agglomeration. Therefore, in order to capture the dynamics of agglomeration along with cross-sectional dimensions panel data model is used in this study. There are also several advantages of using panel data. According to Wooldridge, (2002) panel data is used to solve the omitted variable problem. There are other more advantages of using panel data model than cross-sectional or time series model. According to Baltagi (2001), there are following advantages to use panel data.

a) Large number of data points.
b) Increase degrees of freedom & reduce collinearity.
c) Improve efficiency of estimates and
d) Broaden the scope of inference

4. Methods and Procedure

4.1) Data Source

The health care sector[4] employment data of Metropolitan Statistical Area (MSA) is used for the analysis. Since, metropolitan area has geographically compact development pattern. MSAs in this study are only taken for the lower 48 states of US. Employment data of under the heading health care and social assistance are taken from County Business Pattern (CBS) of US Census Bureau. The NAICS code for health care and social assistance is 62. This is broad sector rather than specific industry. This sector includes four major industries i.e. ambulatory health care services, hospitals, nursing and residential care facilities and social assistance. According to Census Bureau, the health care and social assistance sector is arranged on a continuum starting with those establishments providing medical care exclusively, continuing with those providing health care and social assistance, and finally finishing with those providing only social assistance. The selected MSAs were presented in the map with red color (figure-1).



[4] Health sector here as defined by County Business Pattern whose NAICS code is 62.

Figure 1: MSAs included in the data points.

Besides, the employment data, population, physical size of MSAs, other demographic variables are also used and are collected from Census Bureau. Moreover, average health care expenditure by the state is collected from National Center for Health Statistics, labor force from Bureau of Labor Statistics and national input-output coefficient from Bureau of Economic Analysis. Since this study has used panel data, the data are arranged across the MSAs and of five years time period (from 2003 to 2007).

4.2) Analysis

In this paper panel data model is used to determine the factors affecting agglomeration. Since most of the variables are not directly measurable therefore, numbers of indices are calculated. Some of them are proxy indices. The agglomeration of industry i in region j at time t can be measured in the form of relative density of industry employment to the nation employment of that industry. In this study, health care sector is only investigated therefore subscript i is not needed therefore it is nor used here. This agglomeration index, which is dependent variable, is calculated as following.

$$\boldsymbol{HA_{jt} = \frac{E_{jt}}{E_{ust}L_j}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{3})$$

Where,

E_{jt}= health care sector employment number in MSA j at time t.

E_{ust}= US health care sector employment number at time t (in million).

L_j= Physical size of MSA j (sq. mile of in the year 2000)

HA_{jt}=Proxy for agglomeration of health care sector MSA j at time t

Most of the explanatory variables indexes, which is already been used, are taken from Koo (2005)'s work. Based on above index agglomeration values are calculated. After examining the calculated value of agglomeration, New York, Los Angeles, Trenton-Ewing, New Haven and San Francisco are found most agglomerated MSAs with rank first to fifth respectively (Table-2). But, from the year 2005 Boston gained the fifth rank and San Francisco is on the sixth place.

Table 2: Agglomeration Index for Top Five MSAs.

S.N	MSA name	2003	2004	2005	2006	2007
1	New York-Northern New Jersey-Long Island	11.84	11.58	12.61	11.91	11.07
2	Los Angeles-Long Beach-Santa Ana	7.15	7.00	7.06	7.11	7.20
3	Trenton-Ewing	6.96	6.81	6.96	6.91	7.10
4	New Haven-Milford	6.84	6.69	6.72	6.68	6.79
5	San Francisco-Oakland-Fremont	6.42	6.28			
6	Boston-Cambridge-Quincy			6.53	6.44	5.99

Note: Agglomeration value for San Francisco-Oakland-Fremont is not given because Boston-Cambridge-Quincy MSA is on the fifth position for years 2005 to 2006. Similarly, agglomeration value of Boston-Cambridge-Quincy was not given for year 2003 to 2004 because San Francisco-Oakland-Fremont was at the fifth position.

Similarly, Prescott, Casper, Wenatchee and El Centro and flagstaff are found at the bottom out of 356 MSAs (Table-3). These five MSAs are consistently found at the bottom of the agglomeration index.

Table 3: Agglomeration Index for Least Five MSAs.

S.N	MSA name	2003	2004	2005	2006	2007
1	Prescott	0.08	0.08	0.07	0.08	0.08
2	Casper	0.07	0.07	0.08	0.07	0.07
3	Wenatchee	0.07	0.06	0.07	0.07	0.07
4	El Centro	0.06	0.05	0.06	0.06	0.06
5	Flagstaff	0.03	0.03	0.03	0.02	0.02

Note: Figure in the parenthesis is the rank of MSA.

The first explanatory variable is average state expenditure on health care research in which MSAj is located. This is a proxy variables used to capture the knowledge spillover on health care industry. It is expected to have positive effect on agglomeration.

The second indicator is input availability. According to Koo (2005) this measure evaluates how strong the presence of supplier industries for the health care sector is in MSAj. It is expected to have positive effect on agglomeration.

$$\boldsymbol{INP_{jt}} = \sum_{\boldsymbol{ki=1}}^{\boldsymbol{n}} \boldsymbol{\omega_{kt} LQ_{jt}} \quad \dots\dots\dots\dots\dots\dots (\boldsymbol{4})$$

Where,

INP_{jt}= input availability in health care sector MSA j at time t

ω_{kt}=input-output coefficient from industry k to the health care industry.

LQ_{jt}= Location Quotient for health care industry of MSAj at time t

The third indicator is location quotient that captures the specialization of the health care sector in MSAj. This indicator also provides the information about the input distribution and strength of input industry presence. It was expected to have positive effect on agglomeration.

$$\boldsymbol{LQ_{jt}} = \frac{\frac{\boldsymbol{E_{jt}}}{\boldsymbol{E_{jtot}}}}{\frac{\boldsymbol{E_{ust}}}{\boldsymbol{E_{ustot}}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{5})$$

Where,

LQ_{jt}= Location Quotient of health care sector MSA j at time t

E_{jt}= Employment in health care sector i of MSA j at time t

E_{jtot}= total employment of MSAj at time t

E_{ust}=employment in health care sector in US at time t

E_{ustot}=total employment of US in time t

The fourth indicator is local competition that is developed by Glaeser et al. (1992). This indicator is the ratio per employment establishment number of MSAj to ratio of national level. It is expected to have negative effect on agglomeration.

$$\boldsymbol{LC_{jt}} = \frac{\frac{\boldsymbol{ETS_{jt}}}{\boldsymbol{E_{jt}}}}{\frac{\boldsymbol{ETS_{ust}}}{\boldsymbol{E_{ust}}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{6})$$

Where,

LC_{jt}= Local competition in health care sector in MSA j at time t.

ETS_{jt}= Business establishment number of health care sector in MSAj at time t

E_{jt}=Employment of health care sector of MSA j at time t

ETS_{ust}= Number of business establishment on health care in US at time t

E_{ust}=Employment in health care sector in US at time t

The fifth explanatory variable is proportion of small establishment number of MSAj to the total establishment number in the health care sector. In this study, small industry is defined as industry with less than 100 employments. This is just an arbitrarily chosen. It is expected to have positive effect on agglomeration because of interconnectedness of the many small industries.

MSA population is the last variable that is considered for this model for MSAj at time t. It is expected to have positive effect on agglomeration. Population can be viewed as the labor pooling source as well as pool of consumers who consume the health care sectors services.

Instead of looking at the individual intercepts of 356 MSAs, MSAs were clustered into four groups. In order to cluster them average cluster linkage method is used. Four categories are found to be appropriate to categorize all MSAs based on the agglomeration value. Cluster one has the lowest value of HA_{jt} (i.e. lowest agglomerated MSAs) whereas cluster 4 has the highest value of HA_{jt} (i.e. highest agglomerated MSAs). But, there are not very many MSAs in fourth cluster. After dividing them into four clusters least square dummy variable (LSDV) model is calculated in order to see the fixed effect model of panel data. After LSDV, random effect model is also calculated. Later Hausman's specification test is calculated in order to choose appropriate model between fixed-effect model or random effect model. Detail description of the fixed and random effect models are presented below.

a) Fixed effect model.

A fixed effect model assumes differences in intercepts across groups or time periods. Following model used to estimate the parameters.

$$\boldsymbol{y_{it} = (\alpha + u_i) + X'_{it}\beta + \varepsilon_{it}}$$

$$\boldsymbol{\varepsilon_{it} \sim (0, \sigma_\varepsilon^2)}$$

Where, α is usual intercept and u_t is intercept for the individual intercept for the MSA. X'_{it} is the vector of explanatory variables and β is usually parameter to be estimated. u_t is calculated by creating the dummies for each MSA. As earlier explained in this paper cluster of MSAs is used rather than dummy for each MSA. Before calculating model, descriptive statistics is calculated and presented below (table-4).

Table 4: Descriptive Statistics of Variables.

Variable	Abbreviation	N	Mean	Std Dev	Minimum	Maximum
agglomeration	FA	1760	1.096077	1.238713	0.021373	12.60626
Location Quotient	LQ	1760	1.196833	0.28179	0.489426	2.342737
Local competition	LC	1760	0.877256	0.218629	0.394344	1.728426
Population	pop	1760	691578.8	1570532	54724	18922571
Input availability	INP	1760	3.871999	0.627109	2.202	18.5673
Labor Pooling	LP	1760	15743.81	35468.39	766.84	516323.9
Proportion small establishments	small	1760	0.103673	0.044513	0.06072	1.6791
State average health care research expenditure	State_Res_exp	1760	4670.2	316.4083	4221	5155

Descriptive statistics values presented above do not have usual straight forward meaning because they are obtained after stacking over five-year period of each variable; therefore they are average over the five-year period of each variable.

Since most of the variables in the model contain the density concept, therefore it is suspected to have high correlation among them. Therefore, there is a need to test the multicollinearity. If there a multicollinearity, there might be chance of false conclusion of no linear relationship between an independent and a dependent variable (Green, 1993). Moreover, coefficients will have the wrong sign or implausible magnitude (Green, 1993). In order to check correlation among the explanatory variables, variance inflation faction (VIF) is also calculated through Ordinary Least Square (OLS). The result of the OLS is presented below (table-5). This OLS model is overall significant and has 57 percent r-square.

Table 5: OLS without dummy variables.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	1.48931	0.33441	4.45	<.0001	0
LQ	1	0.16533	0.08765	1.89	0.0594	1.6168
LC	1	-0.22298	0.11962	-1.86	0.0625	1.81243
pop	1	-3.25E-07	6.34E-08	-5.12	<.0001	26.27505
INP	1	0.0828	0.03475	2.38	0.0173	1.2586
state_res_exp	1	-6.9E-05	6.16E-05	-1.12	0.2622	1.00761
LP	1	3.97E-05	2.77E-06	14.33	<.0001	25.5649
small	1	0.0242	0.44744	0.05	0.9569	1.05124

As expected, there is a high VIF value for population and labor pooling that indicates that these two variables likely to be highly correlated. Therefore, later population is kept as independent variable and labor pooling is removed from the model in order to get rid of the multicollinearity problem. Since the labor pooling is poorly calculated by using different time period than 2003 to 2005, therefore, labor pooling removed instead of population. The result has also shown the opposite sign then the study done by Koo (2005) especially highly correlated variables. But, once labor pooling (LP) removed the sign is population turned out to be positive which is as expected.
After OLS calculation, LSDV is calculated with adding dummy variables for clusters of MSAs and time periods. The result of LSDV model presented below (table-6). This model is overall significant and has r-square almost 90 percent. However, usually r-square is unreliable in panel data model.

Table 6: Estimates LSDV model with time and MSA cluster effect.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
LQ	1	0.08307	0.05696	1.46	0.1449
LC	1	-0.55066	0.07486	-7.36	<.0001
pop	1	1.77E-07	1.31E-08	13.59	<.0001
INP	1	0.07088	0.02272	3.12	0.0018
state_res_ex	1	0.000371	2.92E-05	12.71	<.0001
smallest	1	0.00909	0.29002	0.03	0.975
d2 (cluster 2)	1	2.48651	0.08064	30.83	<.0001
d3 (cluster 3)	1	4.73686	0.09431	50.23	<.0001
d4 (cluster 4)	1	7.6136	0.33524	22.71	<.0001
y4 (year 2004)	1	-0.11821	0.04201	-2.81	0.005
y5 (year 2005)	1	-0.1722	0.04428	-3.89	0.0001
y6 (year 2006)	1	-0.23996	0.04613	-5.2	<.0001
y7(year 2007)	1	-0.39958	0.05118	-7.81	<.0001

It can be seen that all MSAs clustered in group 2, 3, and 4 are significantly different from group 1 which have lowest agglomeration. Similarly, over the year agglomeration of health care sector is significantly decreasing as compared with the year 2003. This may be the reason because of increasing trend of suburbanization of cities. It may be due to improvement on infrastructures in an around the cities. Moreover, local competition, population, input availability, and state average expenditure on health care turned out to be significant. The expected sign of significant variable are seen as expected. In addition to estimation of this model joint test also calculated for the MSAs's coefficients. In this joint test, the null hypothesis is all coefficients of the MSAs are zero and alternative hypothesis is at least one different from zero. Null hypothesis is rejected and conclude that they are different from zero.

Fixed effect model cannot estimate effects of variables which vary across individuals but not over time. The use of fixed effects is inefficient if α_i is uncorrelated with x_{it} (i.e., if appropriate model is random effects). Further, the use of fixed effects can exacerbate biases from other types of specification problems, especially measurement error (Green, 1993). Therefore, I also estimate the random effect model.

b) **Random effect model:**

The random effects model examines how group and/or time affect error variances. Additionally, a random effect model is estimated by generalized least squares (GLS) when the variance structure is known and feasible generalized least squares (FGLS) when the variance is unknown (Green, 1993). Here, fixed effect and random effect models were judged in order to identify the better model for agglomeration of health care sector. The general form of random effect model can be presented as below. In this model, individual specific constant term is randomly distributed across cross-sectional units. u_i is random disturbance i^{th} observation and constant over time.

$$\boldsymbol{y_{it} = \alpha + \beta' x_{it} + u_i + \varepsilon_{it}}$$

Where,

y_{it}=dependent variable across i and time t

α = usual intercept

x_{it}= vector of independent variables across the MSA and time t

β= vector of parameters

$$\boldsymbol{u_i \sim iid(0, \sigma_u^2)}$$

The assumption of this model is presented here,

$$\boldsymbol{E[\varepsilon_{it}] = E[u_i] = 0}$$

$$\boldsymbol{E(\varepsilon_{it}^2) = \sigma_\varepsilon^2}$$

$$\boldsymbol{E(u_{it}^2) = \sigma_u^2}$$

$$\boldsymbol{E[\varepsilon_{it} u_j] = 0 \ \forall\ i, t, and\ j}$$

$$\boldsymbol{E[\varepsilon_{it} \varepsilon_{js}] = 0\ if\ t \neq s\ or\ i \neq j}$$

$$\boldsymbol{E[u_i u_j] = 0\ \ if\ i \neq j}$$

Random effect model can be calculated either one-way random effect model or two-way random effect model depending upon the purpose of study. The model presented below is one-way random effect model. The result of the one-way random effect model is presented below (table-7). The result shows that location quotient, local competition, population, input availability, and state average research expenditure on health care turn out to be significant. However, proportion of small firms is not found significant in one-way random effect model. Similarly, state average health care research expenditure is found significant but the sign is contrary to the expectation.

Table 7: Estimates of one-way random effects model.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.353217	0.0793	17.06	<.0001
LQ	1	0.125293	0.0293	4.28	<.0001
LC	1	-0.729	0.0402	-18.15	<.0001
pop	1	4.98E-07	2.68E-08	18.57	<.0001
INP	1	0.007465	0.00352	2.12	0.0341
state_res_exp	1	-0.00004	5.84E-06	-6.44	<.0001
small	1	0.033686	4.68E-02	0.72	0.4716

Since there is no data of average health care research expenditure at the MSA level, here state level data is used. It is likely that average research expenditure is high in less health care access area i.e. rural area than metro area. That may be reason to have negative sign of this variable.

In order to compare the fixed effect model and random effect model Hausman's specification test is also calculated. The Hausman's specification test examines that if the individual effects are uncorrelated with the other regressors in the model (Green, 1993). If correlated a fixed effect model is preferred. The essential result of the Hausman's specification test is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero (Green, 1993). Hausman's specification test process involves two steps. Firstly, obtaining the coefficient estimates of the fixed Effects model and subtracting the coefficient estimates of the random effects model to form a vector of the difference in the coefficient estimates of the two methods. Secondly, obtain the variance-covariance matrix from both fixed and random effect model substrate the variance-covariance matrix of random effect model from of variance-covariance matrix of fixed effect model. It is clearer by looking at the following Hausman's specification test formula.

$$\left[\hat{\beta}_{FE} - \hat{\beta}_{RE}\right]' \left[\mathrm{Var}(\hat{\beta}_{FE}) - \mathrm{Var}(\hat{\beta}_{RE})\right]^{-1} \left[\hat{\beta}_{FE} - \hat{\beta}_{RE}\right] \sim \chi^2_K$$

The result also shows that if it rejects the null hypothesis that means random effect model is not preferred than fixed effect model (table- 8).

Table 8 Results of one-way random effect model Hausman's specification test.

Hausman Test		
DF	m-value	Pr
4	44.95	<0.0001

Based on above one-way random effect model, fixed effect model is preferred than random effect model. Similarly, two-way random effect model is also calculated here. There general form of equation can be presented below.

$$y_{it} = \alpha + \beta' x_{it} + u_i + \gamma_t + \varepsilon_{it}$$

Where, γ_t= random time factor

Other parameters are usual and as defined earlier. Two-way model includes both individual-specific and period-specific effect. Two-way model includes both individual-specific effects u_i and period-specific effects γ_t .The two way random effects model has the null hypothesis that variance components for groups and time are all zero. The result of the two-way random effect model is presented below (table-9)

Table 9: Results of two-way random effect model.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.322265	0.0946	13.98	<.0001
LQ	1	0.13433	0.0294	4.57	<.0001
LC	1	-0.69811	0.0413	-16.89	<.0001
pop	1	4.96E-07	2.68E-08	18.49	<.0001
INP	1	0.007264	0.00355	2.05	0.041
state_res_ex	1	-0.00004	1.20E-05	-3.15	0.0016
small	1	0.027612	4.67E-02	0.59	0.5548

The result shows same results as in the case of one-way random effect model in term of significance of the variables. In this model also proportion of small firm is not significant. Similarly, Hausaman's specification test also found significant that means fixed effect model is preferred over random effect model (table-10).

Table 10: Results of two-way random effect model Hausman's specification test.

Hausman Test		
DF	m-value	Pr
3	42.36	<0.0001

Here also Hausman's specification test is turn out to be significant therefore, it is concluded that fixed effect model is preferred than random effect model.

5. Conclusion

The phenomenon of agglomeration is getting attention for long period of time. Some industries have tendency to co-locate with input markets, whereas others have tendency to co-locate with output markets. However, there are other factors equally affect to the clustering of the health care industries. While analyzing the clustering factors of health care sector, location quotient is not found significant but population and input availability are significantly contributing to the agglomeration of health care industries. This study also shows that Marshallian factors are important for the health care sector as well. Besides, there is also a significant difference between the agglomeration index values between the time periods. Similarly, there is also significant difference between the values of agglomeration index within the group of MSAs. In continuation of effort of finding an appropriate model for identifying the factors affecting agglomeration, this study found the fixed-effect model is appropriate to indentify the factors of agglomeration in health care sector with many advantages over cross-section model.

References:

Baltagi, B.H. 2001. *Econometric Analysis of Panel Data*. Wiley. Chichester

Beckmann, M.J.1999. *Lectures on location theory*, Springer-Verlag Berlin, Germany.

Bates, L.J. and R.E.Santerre.2005. Do Agglomeration Economies exist in the Hospital Services Industry? Eastern Economic Journal.Vol31, No.4, 617-.628.

Ellison, G. and E. L. Glaeser.1999. *The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?*. The American Economic Review, Vol. 89, No. 2, pp. 311-316 [available: http://www.jstor.org/stable/117127]

Ellison, G. D., E. L. Glaeser and W. R. Kerr .2007. *What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns*. NBER Working Paper No. W13068: [Available: http://ssrn.com/abstract=986918]

Gaspar, J., and E. J. Glaeser. 1998. *Information Technology and the Future of Cities*. Journal of Urban Economics 43. 136-156.

Green, W., 1993, *Econometric Analysis*. Macmillan publishing company, Canada

Glaeser, E.L.(ed).2010. Agglomeration Economics. NBER. The university Chicago Press

Glaeser, E.L., H.D.Kallal, J.A.Scheinkman and A. Shleifer.1992.*Growth in Cities*. Journal of political economy100: 1126-1152.

Duranton, G. and D. Puga.2004. *Micro-foundations of Urban Agglomeration economies*. In: J.V.Henderson and J.F. Thisse (ed.) *Handbook of regional and Urban Economics, Volume 4*. Elsevier.

Koo, J. 2005. *Agglomeration and spillovers in a simultaneous framework*. Ann Reg Sci 39(1): 35–47.

Krugman, P.1991. *Geography and trade*. Cambridge, MA: MIT Press.

Krugman, P.1999. *The Role of Geography in Development*, International Regional Science Review.22, 142

Marshall, A. 1920. *Principles of Economics*. London: MacMillan.

McDonald,J.F. and D.P.McMillen. 2007. *Urban Economics and Real Estate Theory and Policy*. Blackwell. Publishing.

Robinson, J. and H. Luft. 1985. *The impact of Hospital Market Structure on Patient Volume Average Length of Stay, and the Cost of Care*. Journal of Health Economics, 333-356.

Rosenthal, S. S and W.C. Strange.2001. *The determinants of Agglomeration*. Journal of Urban Economics 50,191-229.

Wooldridge, J. M. 2002. *Econometrics analysis of cross section and panel data*, MIT.