Spatial Clustering of the U.S. Biotech Industry

by

Pramod R. Sambidi and R. Wes Harrison[1]

May 30, 2006

Selected Paper
Prepared for Presentation at the American Agricultural Economics Association
2006 Annual Meeting
Long Beach, California
July 23-26, 2006

[1]Pramod R. Sambidi and R. Wes Harrison are graduate research assistant and associate professor, respectively. Department of Agricultural Economics and Agribusiness, Louisiana State University Agricultural Center, Baton Rouge LA.

Abstract:

The study develops a county-level spatial tobit model that analyzes factors affecting site-selection of the U.S. biotech industry. The hypothesis of spatial agglomeration economies is confirmed for the spatial structure of the biotech industry, indicating that biotech firms are positively correlated across counties, resulting in clustering of biotech production.

Spatial Clustering of the U.S. Biotech Industry

Background:

The biotech industry is one of the fastest growing industries in the U.S., increasing sales from $7.7 billion in 1994 to $33.3 billion in 2004 (Ernst and Young, 2005). According to the Biotechnology Industry Organization (BIO, 2005), in 2003 there were 1,473 biotechnology companies in the U.S., employing 198,300 people and spending $17.9 billion on research and development. The top five states in terms of number of biotech companies are: California (420), Massachusetts (193), North Carolina (88), Maryland (84), and New Jersey (77) (figure 1; Ernst and Young, 2004). The biotech industry is mainly concentrated in nine cities/regions (San Francisco Bay Area, Boston/Cambridge, San Diego, Los Angeles, New York, Philadelphia, Raleigh-Durham, Seattle and Washington, DC), accounting for three-fourths of the nation's largest biotechnology firms and for three-fourths of the biotech firms formed in the past decade (Cortright and Mayer, 2002).

As a result of the increasing success of the biotech industry, several state and local economic development agencies are designing and implementing strategies to attract new biotech firms, resulting in stiff competition among and within states. For example, in 2004, 40 states have adopted strategies to stimulate the growth of biotechnology and 50 states have technology based economic development initiatives for biotech firms, compared to merely 14 states in 2001 (Battelle and State Science and Technology Institute (SSTI), 2004). Moreover, a survey of 77 local and 36 state economic development departments indicated that 83% have listed biotechnology as one of the top two targets for industrial development (Grudkova, 2001; Cortright and Mayer, 2002).

As a result of increased competition in attracting biotech firms, the U.S. biotech industry is undergoing changes in its geographical distribution. For example, during 1991-2001, while the concentration of biotech firms in the top states increased slightly, the relative ordering of states changed with New Jersey dropping below Maryland and North Carolina in terms of the number of establishments (Feldman, 2003). Moreover, a study conducted by McCandless (2005) indicated that Dallas, Houston, San Antonio, Memphis, Richmond, and Miami-Dade are making significant progress in becoming important biotech regions of the future. Therefore, it is critical to analyze the economic factors and resource endowments that may affect the location of biotech firms. Additionally, it is also important to analyze statistical implications of the spatial clustering of biotech firms, which was ignored by earlier studies (except Goetz and Rupasingha, 2002).

The primary objective of this paper is to identify county level determinants of the spatial distribution of the U.S. biotech industry. Specifically, this study analyzes the extent to which numerous firm-specific, location-specific, and inter- and intra-industry spatial agglomeration factors affect the location, movement and concentration of the U.S. biotech industry. The study utilizes a Bayesian spatial tobit model that captures the spatial organization of the biotech industry utilizing county-level data for the U.S. Analyzing these factors will aid state and local economic development agencies in designing strategies to better retain and attract biotech firms, which in turn will boost state and local economies and provide employment opportunities for their citizens.

Literature Review:

Biotechnology is defined as "the application of biological knowledge and techniques to develop products and services" (BIO, 2005). Goetz and Morgan (1995) defined it as "any technique that uses living organisms to make/modify products, improve plants or animals, or develop microorganisms for a specific use." Biotech firms are mainly research and development (R&D) oriented and operate in collaboration with research-oriented universities, biomedical research centers, and other diversified companies that aid in the production and distribution of biotech products. Biotech products may be characterized as drugs and pharmaceuticals, agricultural, and environmental, which aid in improving the quality of health, increasing the production of agricultural goods, improving food quality, minimizing environmental hazards and providing a cleaner environment.

Several studies have empirically examined the location aspects of the U.S. biotech industry (Goetz and Morgan, 1995; Darby and Zucker, 1996; Gray and Parker, 1998; Prevezer, 1998; Lerner and Merges, 1998;; Zucker, Darby, and Brewer, 1998; Brennan, Pray and Courtmanche, 1999; Zucker, Darby, and Armstrong, 2003; Xia and Buccola, 2005). The rationale for concentration of the U.S. biotech industry in California and the Northeast has been attributed to proximity to highly research-oriented universities, research parks and laboratories, and well developed infrastructure. Gray and Parker (1998) examined the theoretical arguments surrounding the location and organization of biotech firms and analyzed the prospects for industrial renewal and regional transformation. The authors segregate the U.S. biotech industry into three different categories/regions based on the functions performed by biotech firms in those regions.

The first category includes mature drug producing regions, such as New York, New Jersey, Pennsylvania, Delaware, Illinois, and Indiana. These regions include mature pharmaceutical firms that were established prior to the commercialization of biotechnology (before 1970s), and are now primarily involved in manufacturing (53 %) and marketing (72%) new drugs. Another category includes emerging drug-producing regions, such as San Francisco, San Diego, Los Angeles, Seattle, and Boston. Firms in these regions were established mainly during and after the commercialization of biotechnology, and are primarily involved in R&D activities (82%) for new drugs. A third category includes low-cost periphery regions, such as Puerto Rico, the Southern states of the U.S., and other scattered isolated rural areas. Biotech firms in these regions undertake the production of drug products that have achieved commercial scale and other intermediate products (Gray and Parker, 1998).

Munroe, Craft, and Hutton (2002) conducted a survey of biotech companies in three California counties (Alameda, Contra Costa, and Solano). The survey respondents indicated proximity to leading research centers (i.e. ready supply of skilled labor, access to ongoing research activities, new technology, etc.) as the main reason for their current location. Access to venture capital, a well-trained workforce, space for expansion and access to new technology, were considered to be the most critical requirements for the growth and prosperity of respondents' business. The survey results indicated that, in order to attract biotech firms, a state and local economic development agency should provide financial assistance packages (i.e. subsidies, tax advantages, loan guarantees, etc.), biotechnology incubators/research parks (with appropriate zoning, infrastructure and public transportation), promote public awareness and training programs for the workforce. The results also indicated some of the respondent's willingness to locate in regions associated with lower costs (housing, space, wages, etc.), less congestion/commuting, and good incentives such as subsidies and tax credits (Munroe, Craft, and Hutton, 2002).

Goetz and Rupasingha (2002) analyzed the site-specific determinants of the U.S. high-tech industry, which includes firms that are involved in biotech activities, such as drug and pharmaceutical manufacturing firms and R&D services. Their results indicated that the availability of an existing high-tech firm, number of college graduates, local property taxes, population (urbanization economies), total county income, highway access, and county amenity scale have a positive and significant impact on the location of high-tech firms. Conversely, a county's unemployment and unionization rate, per capita pollution, and the percentage of an African-American population were found to have a negative and significant impact on the high-tech firm's location. The present article differs from previous literature in that we examine the economies of scale associated with county-level spatial agglomeration factors, exclusively for firms involved in the biotech activities. Moreover, our study includes several biotech related sub-industries (Agricultural Feedstock and Chemicals) that are of potential interest to agricultural economists and university agricultural experimental stations.

Data:

The biotech industry is a composition of numerous manufacturing, R&D, and services industries. Consequently, it does not have a separate NAICS or SIC code since different subsectors are involved in the production of biotech products. However, Battelle Technology Partnership Practice and State Science and Technology Institute (SSTI) (2004) classified the bioscience[1] into five major subsectors as follows: 1. Agricultural

[1] "The biosciences are not just biotechnology but rather a range of industry sectors relying on insights into the way living organisms function."(Battelle and SSTI, 2004).

Feedstock and Chemicals (NAICS: 311221, 311222, 311223, 325193, 325199, 325221, 325222, 325311, 325312, 325314, 325320, 424910), 2. Drugs and Pharmaceuticals (NAICS: 325411, 325412, 325413, 325414), 3. Medicinal Devices and Equipment (NAICS: 339111, 339112, 339113, 339114, 334510, 334516, 334517), 4. Research and Testing (NAICS: 541380, 541710), and 5. Academic Health Centers, Research Hospitals, and Research Institutes. The first four subsectors include twenty five industries that are involved in biotechnology activities, with total employment of 885,368 jobs across 17,207 establishments (Battelle and SSTI, 2004). Figure 2 illustrates the U.S. employment distribution across the bioscience subsectors.

The present study analyzed several categories of variables that are considered to affect the location of biotech firms, such as agglomeration factors, infrastructure factors, and local economic and socioeconomic factors. County-level data was obtained from the 2003 county business patterns (U.S. Census Bureau), Economic Research Service, National Agricultural Statistics Service, Battelle Technology Partnership Practice and SSTI (state-level data), and U.S. Dept. of Labor. The dependent variable considered in the model is the county level number of establishments of firms belonging to the aforementioned NAICS codes.

Economies of scale associated with agglomeration factors are believed to be one of the driving forces in the geographical distribution of the biotech industry (Pisano, Shan, and Teece, 1988; Gray and Parker, 1998; Goetz and Rupasingha, 2002; Munroe, Craft, and Hutton, 2002; McCandless, 2005). Agglomeration economies indicate that performance of one biotech firm is influenced by the other biotech firm located nearby. The resulting spillovers may be due to an already existing industry specific infrastructure, which is associated with lower transaction costs, proximity to research institutions and specialized intermediate industries, good transportation facilities, and availability of skilled labor pool and financial resources. The research oriented biotech firms generate externalities and spill-overs, which tend to be spatially close to where they were created, resulting in a positive economies of scale for firms located in that region (Jaffe, Trajtenberg, and Henderson, 1993; Dahlander and McKelvey, 2003). Zeller (2001) analyzed the spatial clustering of biotech firms in Germany and argued that, even though, knowledge and technology transfer often happens on a global scale, the exchange of tacit knowledge, however, is facilitated by spatial proximity. In this study, we include a spatial lag variable as a proxy for agglomeration economies that accounts for the biotech establishment counts in neighboring counties. The variable is hypothesized to have a positive and significant effect on the location of biotech firms.

One of the factors that are considered to be a prerequisite for locating a biotech firm is the proximity to research institutions. Several studies have analyzed the role of research institutes in the development and commercialization of biotechnology (Powell and Brantley, 1992; Darby and Zucker, 1996; Zucker, Darby, and Brewer, 1998; Prevezer, 1998; Zucker, Darby, and Armstrong, 2003; Dahlander and McKelvey, 2003; Xia and Buccola, 2005). Industry funded university research increased from $630 million in 1985 to $2.1 billion in 2004 (National Science Foundation, 2006), indicating an increasing affiliation between university and industry in technology advancement. Some of the primary reasons for this collaboration are: access to complementary research activity and human capital, increasing commercial opportunities, stringency of patent law and federal policies, and the relative decline of public research funding (Santoro and

Alok, 1999; Yang and Buccola, 2003). This study includes county-level number of colleges, universities, and professional schools (*colleges*) as a proxy for the proximity to research institutions and assumes it will have a positive and significant effect on the location of biotech industry.

Ongoing research intensity in life sciences at research institutions of a particular state is also considered to be a critical factor in the location decision of a biotech firm (Zucker, Darby, and Brewer, 1998; Munroe, Craft, and Hutton, 2002; Xia and Buccola, 2005). State level university life science R&D expenditures (*R&D*), National Institute of Health support for institutions (*NIH*), higher education degrees in biological science (*Biological Degrees*), and average biological scientists in the workforce 2000-2002 (*Biological Scientists*), are included as proxies for ongoing research intensity. All these variables are assumed to have favorable impact on the site-selection of a biotech firm.

Business that provide venture capital are considered to be an important source of capital, especially, for new and small firms (Powell, Koput, Bowie, and Smith-Doerr, 2002). For a small biotech firm, availability of venture capital in a particular region is as important as the strong research capacity of that region. During 2004, venture funding accounted for approximately 23.5% of the total biotech industry finance, as indicated in figure 3 (BIO, 2005; BioWorld, 2005). Most of the biotech firms are small and operate at a loss, spending large amount of money on research and development for several years, before earning a profit (Cortright and Mayer, 2002). For example, only 1 in 5,000 potential new medicines reach the pharmacy shelf, and that is after 12 to 15 years of R&D with an average expenditure of $500 million (California Trade and Commerce Agency, 2001). As a result, most of the small biotech firms depend on venture capital funds, on research contracts and equity investment from large biotech firms, and on sales of their company stock in public markets (Cortright and Mayer, 2002). Therefore, the availability of local venture capital firms (*Venture Capital*) is hypothesized to have a positive and significant impact on the location of a biotech firm in that region.

Agriculture is one of the important components of the biotech industry, along with drugs and pharmaceuticals. Some studies have analyzed issues related exclusively to agricultural biotech firm location and its relationship with research institutions (Kalaitzandonakes and Bjornson, 1997; Graff, 1997; Begemann 1997; Brennan, Pray, and Courtmanche, 1999; Sporleder, Moss, and Nickles, 2002; Yang and Buccola, 2003; Sporleder and Moss, 2004; Xia and Buccola, 2005). Around 13 percent of firms in biotechnology are primarily involved in agriculture (Dibner 1995; Graff 1997). As illustrated in figure 2, agriculture related biotech activities account for 17% of total U.S. employment across bioscience subsectors. According to Ernst and Young (2000), in 1999, agricultural biotech firms employed 21,900 workers, generated $2.3 billion in revenues and $1.4 billion in personal income for employees and owners. The primary goal of agricultural biotech firms is to develop genetically modified high yielding varieties with improved resistance to natural enemies (e.g. pest, diseases, weeds, and adverse growing conditions), better quality and longer shelf life for fruits and vegetables. Since some of the biotech firms seek applications directed toward agricultural production, it is hypothesized that, in order to gain positive economies of scale (low transaction costs), biotech firms prefer to locate in regions with significant agricultural production (*Farmland*). Similarly, since the biotech industry involves drugs and pharmaceutical firms, and medicinal devices and equipment firms, we hypothesize that a county with more hospitals will favor the location of the biotech industry.

In terms of conventional location theory, local property taxes (*Property Tax*) may discourage new investment by increasing the costs of production. However, in case of high-tech firms (such as the biotech firms), high property taxes are consider to be proxies for greater availability or higher quality of local public goods (Goetz and Rupasingha), which in turn reflects high standard of living of the local community. Therefore, we assume that property taxes are positively correlated with the location of biotech firms. Similarly, counties with high unemployment rate (*unemployment*) and poverty rate (*poverty*), which reflect low standard of living of the local community, are considered to have a detrimental effect on the location-decision of biotech firms. Similarly, counties with higher crime rates (*Crime Index*) are also considered to have a negative impact on the location of biotech firms. Urbanization economies (*Rural-Urban*) are measured using the rural-urban continuum codes for U.S. counties, which range from 1 to 9, where 1 represents extremely urban and 9 represents extremely rural. Since biotech firms may wish to locate close to research institutions and regions with well developed infrastructures, we posit a negative relationship between firm location and the circumstance of a county being rural.

The impact of labor quality on the location decision of biotech industries is measured by county-level average wage per job (*Wage*) and percentage of persons with a college degree (*Education*) (Zucker, Darby, and Brewer, 1998). Both variables are considered to have a positive relationship with the site-selection of the biotech industry. Biotech firms prefer to locate in highly populated centers (*Population*) as it provides appropriate services such as contracting for site building, major equipment, and availability of housing (McCandless, 2005). Counties with high per capita incomes (*Income*), which represents a high standard of living, are considered to have a favorable impact on the biotech firms' site-selection. Similarly, median housing values (*Housing Value*) are used as a proxy for the quality of housing in a give county. It is expected to have a positive impact on the location-decision. Table 1 presents the descriptive statistics of all the variables included in the model

Spatial Exploratory Analysis:

The spatial distribution of biotech firms based on 2003 county business pattern data is presented in figure 4. The figure illustrates standard deviations of biotech establishments with the mean of biotech establishments equal to 11.49. A high concentration of firms is seen in the Northeast and West, as well as in major metropolitan cities, which involve 519 counties, accounting for 16.68 percent of the total observations. Most of the counties that are without and not adjacent to a major metropolitan city, indicate a lack of biotech firms. These regions involve the rest of 2592 counties, accounting for 83.32 percent of the total observations. Figure 5 presents the top thirty U.S. counties in terms of the number of biotech establishments, where each of the top thirty counties included at least one major city. This implies that, the U.S. biotech industry exhibit's a spatial pattern, and it is not independently distributed over space.

The spatial association of biotech firms is tested using a global Moran's I, which measures similarities and dissimilarities in biotech establishments across neighboring counties (Anselin, 1995). For the number of biotech establishments, *y*, Morans'I is:

$$I = \left(\frac{n}{\sum_i \sum_j w_{ij}} \right) \times \left(\frac{\sum_i \sum_j (y_i - \mu)(y_j - \mu)}{\sum_i (y_i - \mu)^2} \right)$$

where w_{ij} indicates elements of the spatial weight matrix, $\boldsymbol{W}$ (Rook contiguity weight matrix), μ the mean of all y observations, and i, j=1,…,n. A positive and significant value for Moran's I indicate positive spatial correlation, showing that counties have a high or low number of establishments similar to their neighboring counties. Conversely, a negative and significant value for Moran's I indicates negative spatial correlation, showing that counties have high or low number of establishments unlike neighboring counties (Pacheco and Tyrrell, 2002). We calculate Moran's I for the 2003 number of biotech establishments across all contiguous U.S. counties, employing GeoDa, a spatial data analysis software. The Moran's I statistic is equal to 0.3058, indicating a significant strong positive spatial relationship. However, in the case of uneven spatial clustering, global spatial indicators such as Moran's I are found to be less useful. This resulted in a new general class of local spatial indicators such as Local Indicators of Spatial Association (LISA, also known as Local Moran), which measures the contribution of individual counties to the global Moran's I statistic (Anselin, 1995). The LISA statistic is calculated for the i^{th} county as:

$$I_i = z_i \sum_j w_{ij} z_j$$

where w_{ij} indicates elements of the spatial weight matrix, $\boldsymbol{W}$ (Rook contiguity weight matrix), and z_i and z_j indicates the standardized number of establishments for county i and j, respectively. The sum of LISAs ($\sum_i I_i$) for all observations is proportional to global

Moran's I, implying that LISA statistic can be interpreted as indicators of local spatial clusters and as diagnostics for local instability (spatial outliers) (Anselin, 1995).

Figure 6 illustrates the biotech industry clusters produced by LISA. It indicates the locations with a significant Local Moran statistic classified by type of spatial correlation: (a) high-high association (HH), a county with many biotech firms has neighboring counties with many biotech firms; (b) low-low association (LL), a county with few biotech firms has neighboring counties with few biotech firms; (c) low-high association (LH), a county with few biotech firms has neighboring counties with many biotech firms; and (d) high-low association (HL) , a county with many biotech firms has neighboring counties with few biotech firms. The HH and LL locations suggests clustering of similar values (positive spatial correlation), whereas the HL and LH locations indicate spatial outliers (negative spatial correlation) (Anselin, 1995). A positive and high autocorrelation is found in California, the Northeast, as well as in major cities such as Seattle, Portland, Salt Lake City, Phoenix, Denver, Houston, Dallas, Minneapolis, Chicago, Detroit, Cleveland, Cincinnati, Atlanta, Miami, Orlando, Tampa, and Raleigh. Figure 7 indicates locations with a significant Local Moran statistic. It illustrates that Imperial County (CA), Orange County (CA), Riverside County (CA), San Mateo County (CA), Lee County (KY), and Grant County (WV) have the most significant biotech clusters among all U.S. counties, which is indicated by a p-value of 0.0001.

In addition to the spatial autocorrelation, multivariate spatial correlation is also analyzed employing a multivariate Moran's I statistic. The multivariate spatial correlation "centers on the extent to which values of one variable (z_k) observed at a given location show a systematic (more than likely under spatial randomness) association with another variable (z_l) observed at the neighboring locations." (Anselin, Syabri, and Smirnov, 2002). The multivariate Moran's I is as follows:

$$I_{kl} = \frac{z_k^{'} W z_l}{n}$$

Where n indicates the number of observations, ***W*** indicates rook contiguity weight matrix, and z_k and z_l indicate standardized variables with mean zero and standard deviation equal to one (Anselin, Syabri, and Smirnov, 2002). Using a similar rationale as in the development of LISA, a Multivariate Local Moran Statistic (MLMS) was developed by Anselin, Syabri, and Smirnov (2002). This is defined as follows:

$$I_{kl}^{i} = z_k^{i} \sum_j w_{ij} z_l^{j}$$

where w_{ij} indicates elements of the spatial weight matrix, ***W*** (Rook contiguity weight matrix), and z_k^i and z_l^j indicates the standardized variables for county *i* and *j*, respectively. The MLMS "gives an indication of the degree of linear association (positive or negative) between the value for one variable at a given location *i* and the average of *another* variable at neighboring locations." (Anselin, Syabri, and Smirnov, 2002). Similar to LISA, MLMS suggests two classes of positive spatial correlation, or spatial clusters (HH and LL), and two classes of negative spatial correlation, or spatial outliers (HL and LH) (Anselin, Syabri, and Smirnov, 2002).

Since the affect of agricultural production on the location of the biotech industry is of potential interest to the agricultural economists, this study analyzes spatial correlation between *Farmland* and *spatial lag* of the dependent variable (number of biotech establishments). The Multivariate Global Moran's I statistic is equal to -0.0316, indicating a significant negative spatial relationship between agricultural production and the location of the biotech industry. The MLMSs cluster map indicates a negative spatial correlation mainly in the Northeast and Central U.S., whereas a positive spatial correlation is indicated in parts of Southeast, East North Central and Western regions of the U.S. (figure 8). The significance of the MLMSs spatial clusters is illustrated in figure 9. This indicates that globally, agricultural production and the biotech industry have a negative relation; however, locally there are certain counties where the location of the biotech industry is associated with agriculture rich neighboring counties. We further analyze this relationship along with the other factors discussed earlier, employing a spatial econometric model.

Econometric Model:

Most of the previous empirical studies on industry location have employed non-spatial econometric models, such as the Ordinary Least Squares (OLS), Poisson, Negative Binomial, and Tobit (for recent exception see Roe, Irwin, and Sharp,2002; Goetz and Rupasingha, 2002; Isik, 2004; and Sambidi and Harrison, 2005). Since the data is collected with reference to points in space, employing OLS (and models mentioned above) as an econometric tool will produce spatially autocorrelated residuals, resulting in biased estimates and all inferences based on the model may be incorrect (Anselin, 1988; LeSage, 1999).

The spatial correlation of OLS residuals (without spatial component) is formally tested employing different spatial correlation indices (Morans-I, Wald, Lagrange multiplier and the Likelihood ratio test statistic), as suggested by LeSage (1999). All the tests indicated the presence of spatially correlated residuals in the regression model

(Table 2). To overcome this problem of spatial autocorrelation, different approaches were undertaken, which involved spatial weights matrices (for example, spatial autoregressive model (SAR) and Spatial Error Model (SEM)). However, spatial correlation of data involving discrete dependent variables, have received little attention in the literature. In contrast to general spatial models, estimation of spatial discrete models yields a non-spherical variance covariance matrix, resulting in a heteroskedastic error term (Anselin 2002; Fiva and Rattsø, 2005). To solve this problem, McMillen (1992) employed an error model (EM) algorithm approach to estimate the SAR and SEM probit models containing spatial heteroskedasticity. However, McMillen's EM estimator is associated with certain drawbacks (LeSage, 2000), which were overcome by LeSage (2000), who developed a Gibbs sampling approach to estimate heteroskedastic spatial autoregressive and spatial error probit and tobit models.

The study employs a spatial tobit model, since the dependent variable contains 561 observations with zeros. The spatial tobit model with a spatial lag variable is as follows:

$$Y = \rho WY + \beta X + \varepsilon,$$

where $\boldsymbol{y}$ is a n x 1 vector of endogenous number of biotech establishments variable in each of the n counties for a given time period, $\boldsymbol{\rho}$ is the scalar for spatial lag coefficient, $\boldsymbol{W}$ is the n x n spatial weigh matrix, $\boldsymbol{\beta}$ is the k x 1 parameter vector, $\boldsymbol{X}$ is the n x k matrix of exogenous explanatory variables, $\boldsymbol{\mu}$ is an n x 1 vector of normally distributed error terms with zero mean and variance σ^2. The model is estimated employing a Bayesian estimation method provided by LeSage's econometric toolbox (2005).

Results:

Results of the spatial and standard tobit model along with the marginal effects are presented in table 3. The spatial tobit model resulted in a Pseudo R-Square value of 0.54. Relatively small standard errors for the spatial tobit model, compared to the standard tobit, indicated that the former is a better fit. Moreover, the highly significant spatial lag parameter (ρ) suggests that inference based on the standard tobit specification without a spatial correction, is not valid for the data under consideration.

The spatial lag coefficient (ρ) is positive and significant at the 1% level, indicating the presence of spatial agglomeration economies for the spatial structure of the biotech industry. The positive sign indicates that the spatial distribution of biotech firms is positively correlated across counties. As mentioned earlier, agglomeration factors result in economies of scale, which create a favorable infrastructure for new and existing biotech firms. Hence, counties producing biotech products tend to be concentrated across regions in order to utilize positive externalities associated with agglomeration economies.

Most of the county-level variables in the spatial tobit model have the expected signs and are significantly different from zero. A county's population had an expected positive sign, and was found to be significant at the 1% level. This is in accordance with the present spatial distribution of biotech firms, which are located mainly in major metropolitan cities that are highly populated. The marginal effects for the population variable indicate that, as population in a given county increases by thousand, the number of biotech firms in that county increases by 0.014 units. The median housing value had a negative and significant impact on the location of biotech firms, indicating that biotech firms avoid locating in a county with high housing values. This result is some what surprising since most of the biotech firms are located in urban areas, where housing costs are considered to be high. However, this result is in accordance with the findings of Goetz and Rupasingha (2002) who found that the median housing values have a negative and significant impact on the growth of U.S. high-tech firms.

The unemployment rate had an expected negative sign and was found to be significant at the 10% level. This result is in accordance with the fact that high unemployment rate reflect a lower local quality of life or a weak economy, which is not generally preferred by biotech firm (Goetz and Rupasingha, 2002). The results indicate that, as the unemployment rate in a particular county increases by 1%, possibility of locating a biotech firm in that county decreases by 0.025 units. Conversely, the variable was found to be insignificant in the standard tobit model, which failed to account for spatial autocorrelation.

A county's median household income, which reflects the local standard of living was found to have a positive and significant impact on the location of biotech firms. This result is consistent with previous literature, which indicates biotech firms' preference for locating in regions with a high standard of living and well developed infrastructure. Availability of local venture capital was also found to be positive and significant, indicating biotech firms' dependency on local financial sources. As the number of venture capital firms in a given county increases by one unit, the chances of locating a biotech firm in that county increases by 0.215 units. Similarly, number of colleges and hospitals in a given county were also found to have a positive and significant impact on the location of the biotech industry. As the number of colleges and hospitals in a particular county increase by one unit, the number of biotech firms in that county increases by 0.261 and 0.235 units, respectively. Also noteworthy is that coefficient of hospitals in spatial tobit model is high compared to the one in the standard tobit model, which indicated a downward bias of the estimate.

The variables average wage per job and education, which reflect the labor quality in a given county were positive, as expected, and were found to be significant at the 5% and 10% level, respectively. This result indicates biotech companies' preference for counties with a skilled labor pool. The property tax variable, which was used as a proxy for high standard of living, was found to have an expected positive sign and was significant at the 1% level. The urbanization economies (*Rural-Urban*) had an expected negative sign, but was found to be insignificant in the location of biotech industry. Conversely, Goetz and Rupasingha (2002) indicated that rural counties have a negative and significant impact on the location of high-tech firms, which includes drugs and pharmaceuticals and R&D services. The reason for this may be attributed to the fact that, biotech establishments in our model include firms that are related to agriculture (agricultural feedstock and chemicals), and are assumed to be located in close proximity to rural areas, where farm production is high. Therefore, the significance of urbanization economies variable may be tapered by the location of agricultural biotechnology firms. A county's poverty rate was found to have a negative, but insignificant impact on the location of the biotech industry. Conversely, it was found to be positive and significant in the standard tobit model, implying that models which do not account for spatial correction result in biased and inconsistent estimates. A county's crime index was also found to be insignificant in the site-selection of biotech firms, which is consistent with Goetz and Rupasingha's (2002) findings. However, the variable was found to be negative and highly significant in the standard tobit model, indicating a upward bias of the estimate that failed to correct for spatial autocorrelation.

The amount of farmland in a given county was found to be positive and significant at the 10% level, unlike the MLMS, which indicated a negative spatial association. However, while estimating the bivariate Moran's I statistics, other variables affecting the location of the biotech industry were not considered. The significance of farmland land in the location of biotech firms is attributed to the involvement of agricultural feedstock (which utilizes agricultural products as inputs) and chemical firms (which is utilized in agricultural production as pesticides, insecticides and herbicides) in the biotech industry. The state level variables (*R&D, NIH, Biological Degree, and Biological Scientists*) were found to be insignificant, except for *NIH*, which was found to be positive and significant at the 5% level. The reason for the insignificance of the rest of the variables may be attributed to the fact that they are all measured at the state level and were not able to capture the county level effects.

Conclusions:

Over the past two decades the U.S. biotech industry has experienced significant growth, resulting in an increase in size and number of establishments. Currently, several state and local economic development agencies are designing and implementing strategies to attract new biotech firms, resulting in stiff competition among and within states. As a result of this increasing competition, the U.S. biotech industry is experiencing some changes in its geographical distribution. However, only some new state/regions are likely to attract biotech firms, as most biotech firms are tending to cluster along existing biotech regions. Several studies have analyzed the location aspects of the biotech industry, however, our understanding of the spatial influence on the regional distribution of biotech establishments, is anecdotal. This study employs a Bayesian spatial tobit model that analyzes factors affecting site-selection of the U.S. biotech industry taking the spatial affect into consideration. The study examines the impacts of agglomeration factors, infrastructure factors, and local economic and socioeconomic factors on the county-level biotech establishments. A total of twenty five biotech related industries were analyzed in the study.

The hypothesis of spatial agglomeration economies is confirmed for the spatial structure of the biotech industry, indicating that biotech firms are positively correlated across counties, resulting in clustering of biotech production. Availability of venture capital firms, research institutions, and hospitals were found to have the most significant impact on the location of biotech firms. This indicates that the biotech firms prefers to locate in regions where they have a source for financing their business , access to research institutes to collaborate with skilled labor and obtain new technology, and access to hospitals for research, testing and marketing of new biotech products.

Biotech companies also prefer to locate in counties with a well developed infrastructure. This is indicated by the positive and significant estimates of median household income, average wage per job, education, population, and property tax, and a negative and significant estimate of the unemployment rate. In terms of the theory of industry location, firms should prefer counties with low wages, low property taxes and high unemployment rate; however, the preference of biotech firms seen here is different. In the case of biotech industry location, these three variables are assumed to proxy the standard of living of a given county, thus, indicating there preference to locate in counties with a high standard of living.

The above findings may hinder rural areas hopes of attracting biotech firms; however, they are capable of attracting at least one category of the biotech industry (Agricultural Feedstock and Chemicals), which is involved in the agricultural and biotech activities. Moreover, a county being “rural” was found to be insignificant in the location of biotech firms, whereas, agricultural production was found to be significant, indicating the importance of agriculture for agricultural biotech activities. The rural areas may also want to target the biotech firms that are involved in the manufacturing of intermediate products and drugs that have achieved commercial scale. These type of biotech firms are found to operate in locations that are associated with low costs of production, availability of space for expansion, low median housing values and good incentives (Gray and Parker, 1998).Thus, the state and local economic development agencies should design strategies based on the type of biotech firm they want to attract.

Future research is directed toward a separate analysis of factors affecting the location of agricultural and non-agricultural biotech firms. It will also be interesting to develop a separate model for each census region, in order to analyze any regional differences among the factors affecting the location of biotech industry. Including county-level variables related to the state and local economic development incentives, R&D expenditures, and environmental constraints may further enlighten our understanding of the biotech industry location.

References:

Anselin, L. *Spatial Econometrics, Methods and Models*. Dordrecht: Kluwer Academic, 1988.

Anselin, L., 1995. Local indicators of spatial association – LISA. Geographical Analysis, 27: 93-115.

Anselin, L., I. Syabri, and O. Smirnov. 2002. "Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows." In: Proceedings of the SCISS Specialist Meeting "New Tools for Spatial Data Analysis". Santa Barbara, California, USA. May 10-11, 2002.

Battelle Technology Partnership Practice and State Science and Technology Institute. 2004. " Laboratories of Innovation: State Bioscience Initiatives 2004." Biotechnology Industry Organization, Washington D.C.

Begemann BD. 1997. "Competitive strategies of biotechnology firms: Implication for US Agriculture." Journal of Agricultural and Applied Economics 29:117-22.

Brennan, M.F., Pray, C.E., & Courtmanche, A. 1999. "Impact of industry concentration on innovation in the U.S. plant biotech industry." Paper presented at the Transitions in Agbiotech: Economics of Strategy and Policy conference, Washington, DC.

Biotechnology Industry Organization. 2005-2006. "Guide to Biotechnology." Biotechnology Industry Facts. BIO, Washington D.C.

BioWorld Publishing Group. 2005. BioWorld Today.

California Trade and Commerce Agency, Office of Economic Research. 2001. Biotechnology. May 2001.

Cortright, J., & Mayer, H. (2002). "Signs of Life: The Growth of Biotechnology Centers in the U.S." Portland: The Brookings Institution Center on Urban and Metropolitan Policy.

Dahlander, L. and McKelvey, M. (2003): Revisting Frequency and Spatial Distribution:Innovation Collaboration for Biotech Firms. Paper presented at DRUID's Summer 2003 Conference, Helsingore, June 12-14.

Darby M.R. and L.G. Zucker, 1996. "Star Scientists, Institutions, and the Entry of Japanese Biotechnology Enterprises," NBER Working Papers 5795, National Bureau of Economic Research, Inc.

Dibner, Mark D. 1995. *Biotechnology Guide U.S.A.*, 3rd ed., Research Triangle Park, NC: Institute for Biotechnology Information (IBI).

Ernst & Young. 2000. "The Economic Contributions of the Biotechnology Industry to the U.S. Economy." Prepared for Biotechnology Industry Organization. Ernst & Young Economics Consulting and Quantitative Analysis.

Ernst & Young. 2004. "Resurgence: The Americas Perspective–Global Biotechnology Report 2004 Ernst & Young LLP.

Ernst & Young. 2005. “Beyond Borders: The Global Biotechnology Report.” Ernst & Young LLP.

Feldman, M., 2003. “The locational dynamics of the US biotech industry: Knowledge externalities and the anchor hypothesis.” Industry & Innovation, 10(3), 311-328.

Fiva J.H. and J. Rattsø. 2005. “Decentralization with Property Taxation to Improve Incentives: Evidence from Local Governments’ Discrete Choice.” Working Paper Series No. 6/2005. Department of Economics, Norwegian University of Science and Technology.

Goetz, S. J. and A. Rupasingha. 2002. “High-Tech Industry Clustering: Implications for Rural Areas.” American Journal of Agricultural Economics 84(5).

Goetz, S. and S. Morgan. 1995. “State-Level Locational Determinants of Biotechnology Firms”. Economic Development Quarterly, (9):174-184.

Graff D. G. 1997. “The Agricultural Biotechnology Industry in Overview.” Agricultural and Resource Economics, University of California, Berkeley.

Gray M. and Parket E. 1998. “Industrial Change and Regional Development: The Case of the US Biotechnology and Pharmaceutical Industries.” ESRC Centre for Business Research, University of Cambridge, Working Paper No. 95.

Grudkova, V. (2001). The Technology Economy: Why do Tech Companies Go Where They Go? EDA National Forum, Washington, DC (May 30, 2001).

Isik, M.2004. “Environmental Regulation and the Spatial Structure of the U.S. Dairy Sector.” *American Journal of Agricultural Economics* 86: 949-962.

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

Kalaitzandonakes, N. and Bjornson, B. (1997). "Vertical and Horizontal Coordination in the Agro-biotechnology Industry: Evidence and Implications" *Journal of Agricultural and Applied Economics 29*: 129-139.

Lerner, J. and Merges, R. 1998. “The Control of Strategic Alliances: An Empirical Analysis of the Biotechnology Industry.” *Journal of Industrial Economics*, Vol. 46. pp. 125–156.

LeSage, J.P. 1999. “*The Theory and Practice of Spatial Econometrics.”* Unpublished, Dept. of Econ., University of Toledo.

LeSage, J. P. 2000. “Bayesian Estimation of Limited Dependent variable Spatial Autoregressive Models”, *Geographical Analysis*, Volume 32, number 1, p. 19-35.

LeSage, J. P. (2005), "Econometrics Toolbox", MatLab programs. Downloaded from http://www.spatial-econometrics.com.

McCandless M.E. 2005. “Biotechnology Locations: Hitting the Mark.” Business Facilities. http://www.businessfacilities.com/bf_05_05_cover.asp (Date last visited 14rth May 2006).

McMillen, D. P. (1992), ‘Probit with spatial autocorrelation’, Journal of Regional Science 32-3, 335–348.

Munroe T., G. Craft, and D. Hutton. 2002. “A Critical Analysis of the Local biotechnology Industry Cluster — Counties of Alameda, Contra Costa, & Solano in California.” Volume III, Appendix. Munroe Consulting Inc., Craft Consulting and Hutton Associates. Research Monograph prepared for a Consortium of Bay Area Organizations, May 2002.

National Science Foundation. 2006. Industrial Funding of Academic R&D Continues to Decline in FY 2004. NSF 06-315.

Pacheco, Andrada I., and Timothy J. Tyrrell. 2002. “Testing Spatial Patterns and Growth Spillover Effects in Clusters of Cities.” *Journal of Geographical Systems* 4:275-285.

Pisano, G. P., W. Shan, and D. J. Teece. 1988. Joint ventures and collaboration in the biotechnology industry. In *International Collaborative Ventures in US Manufacturing*, edited by D. C. Mowery. Cambridge: Ballinger.

Prevezer, M. 1998. Clustering in biotechnology in the USA, pp. 124–193 in Swan, G. M. P., Prevezer, M. and Stout, D. K. (eds), The Dynamics of Industrial Clustering. International Comparisons in Computing and Biotechnology, Oxford, Oxford University Press.

Powell, W. W., and P. Brantley. 1992. Competitive cooperation in biotechnology:Learning through networks? In *Networks and Organizations. Structure, Form andAction*, edited by R. Eccles. Boston: Harvard Business School Press.

Powell W. W., K.W. Koput, J. Bowie,. and L. Smith-Doerr. 2002. “The Spatial Clustering of Science and Capital. Journal of Regional Studies, 36 (3).

Roe, B., E. Irwin, and J. S. Sharp. 2002. “Pigs in Space: Modeling the Spatial Structure of Hog Production in Traditional and Nontraditional Production Regions.” *American Journal of Agricultural Economics* 84: 259-278.

Sambidi P.R. and W. Harrison. 2005. “Spatial Dependency of the Geographically Concentrated U.S. Broiler Industry.” Paper Presented at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island.

Sporleder, T.L., L.E. Moss, and L.A. Nickles. 2002. “Agricultural Biotechnology Start-Ups: Does Venture Capital Matter?” Prepared for presentation at the Third Annual Agricultural Biotechnology Conference, Ohio’s Future in Functional Foods, Reynoldsburg, OH.

Sporleder, T.L. and L.E. Moss. 2004. “Knowledge Capital, Intangible Assets, and Leverage: Evidence from U.S. Agricultural Biotechnology Firms.” Paper Presented at the 2004 Symposium of the International Food and Agribusiness Management Association Meetings, Switzerland, June 12-13, 2004.

U.S. Census Bureau. 2003. County Business Pattern. Data Available at http://censtats.census.gov/cbpnaic/cbpnaic.shtml

U.S. Department of Agriculture. 2003. Economic Research Service. Data Available at http://www.ers.usda.gov/Data/

___. National Agricultural Statistics Service. *Historical Data.* Available at http://www.nass.usda.gov/Data_and_Statistics/index.asp

U.S. Department of Labor. Bureau of Labor Statistics. Available at http://www.bls.gov/home.htm

Xia, Y., and Buccola, S. 2005. University life science programs and agricultural biotechnology. *American Journal of Agricultural Economics*, *87*(1), 229-243.

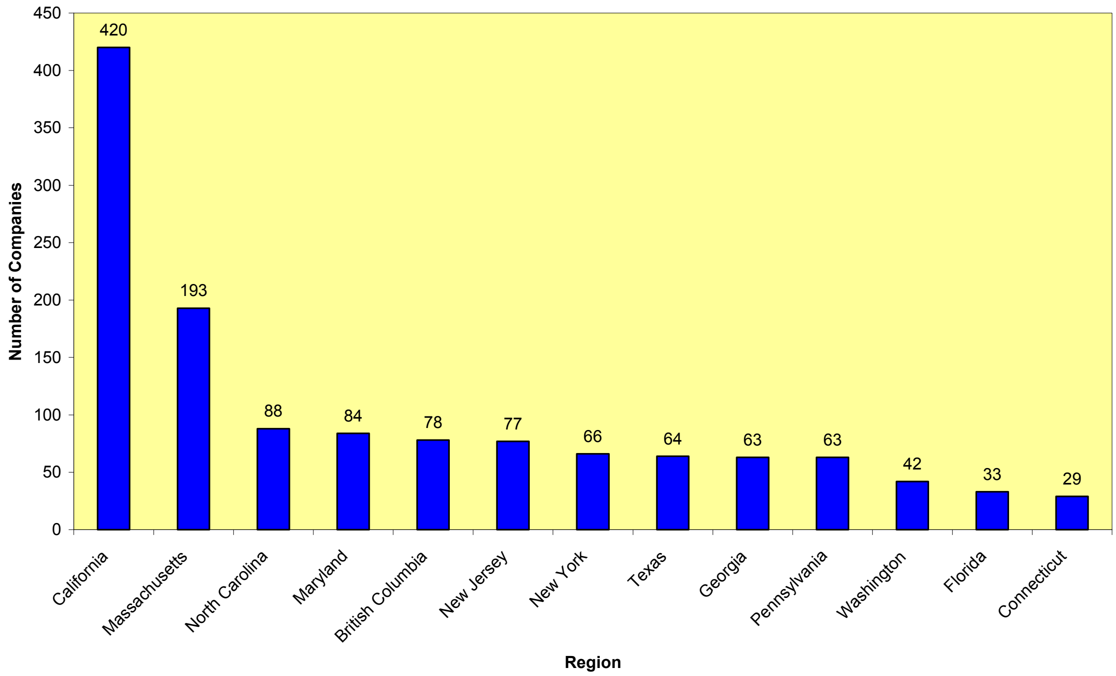
Yang H. and S. Buccola. 2003. "University-Industry Relationships and the Design of Biotechnology Research." Paper Presented at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, July 2003.

Zeller, C., 2001, Clustering biotech: A recipe for success? Spatial patterns of growth of biotechnology in Munich, Rhineland and Hamburg, Small Business Economics, 17, 123-141.

Zucker, L. G., M. R. Darby, and M. B. Brewer. 1998. Intellectual human capital and the birth of U.S. biotechnology enterprises. *The American Economic Review* 88:290-306.

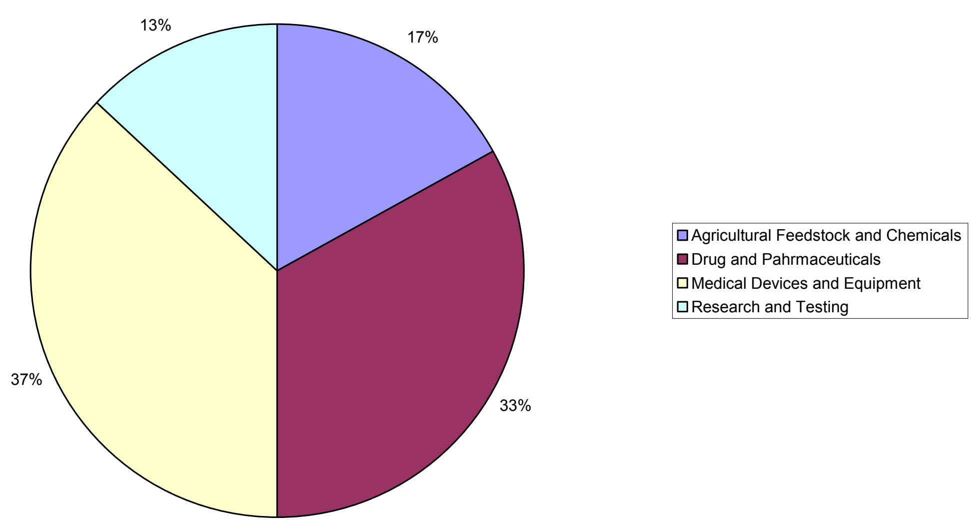
Zucker L. G. & M.R. Darby & J.S. Armstrong, 2003. "Commercializing knowledge: university science, knowledge capture and firm performance in biotechnology," Proceedings, Federal Reserve Bank of Dallas, issue Sep, pages 149-170.

Firgure 1. The U.S. Biotech Companies by State and Province



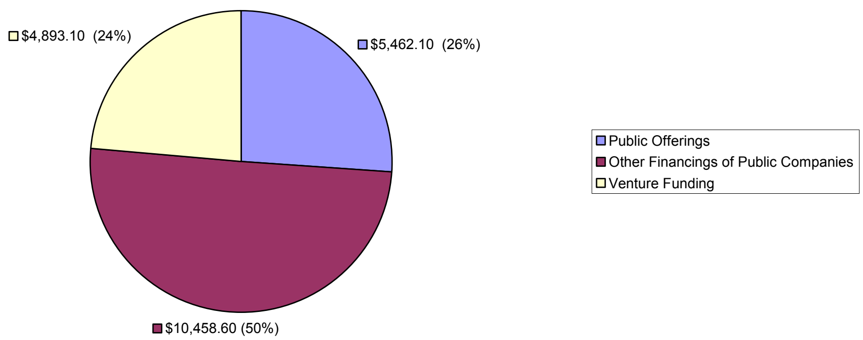
Source: Ernst & Young LLP, America's Biotechnology Report: Resurgence, 2004.

Figure 2: U.S. Employment Distribution across the Bioscience Subsectors



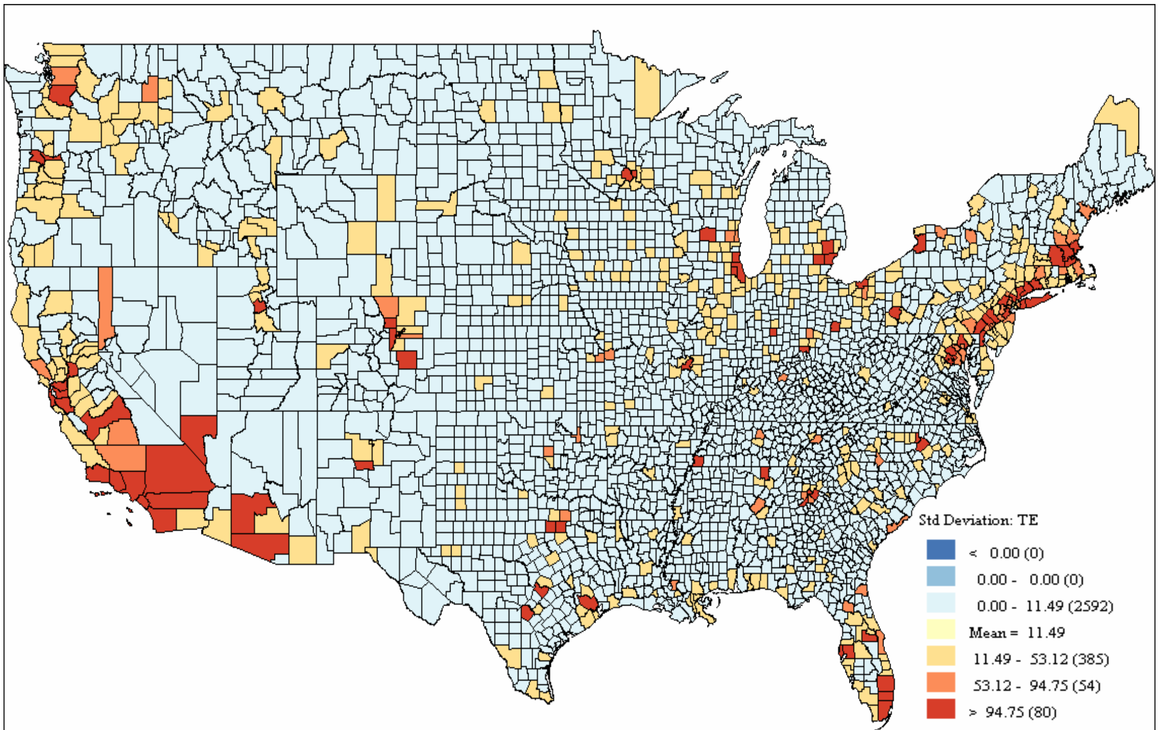
Source: Battelle Technology Partnership Practice and State Science and Technology Institute. 2004.

Figure 3. Biotech Industry Financing, 2004



Source: Biotechnology Industry Organization, 2005; BioWorld, 2005.

Figure 4. Spatial Distribution of the U.S. Biotech Establishments, 2003



Note: TE represents total number of biotech establishments in each county

Figure 5. U.S. Biotech Establishments by County, 2003

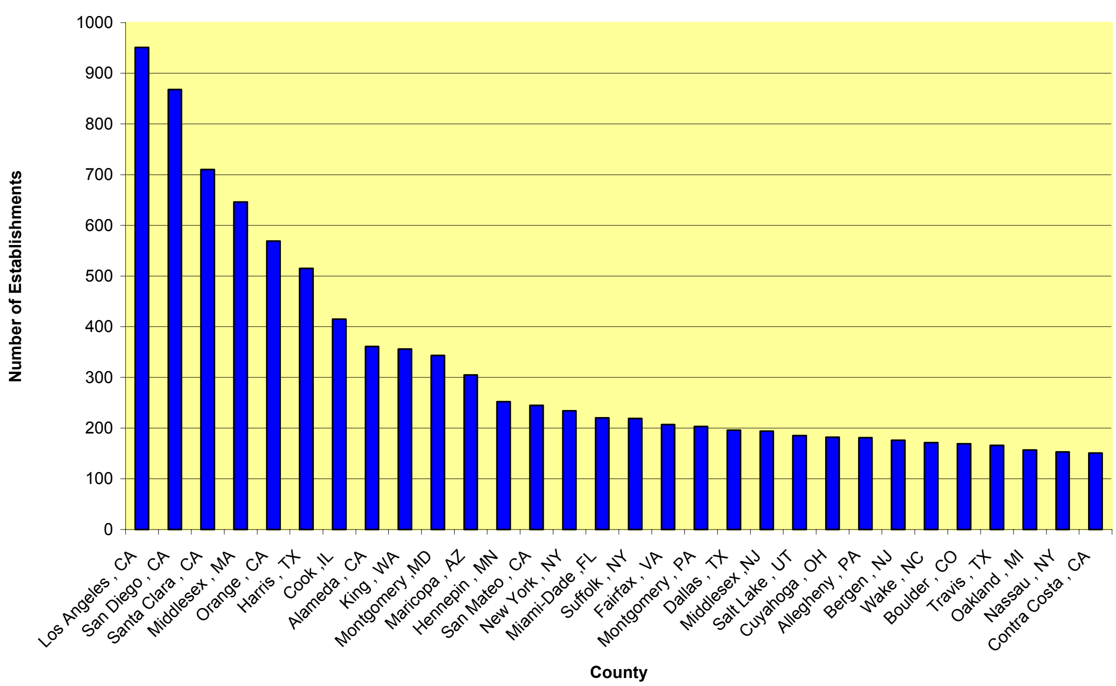


Figure 6. Local Indicator of Spatial Association (LISA) Cluster Map for Biotech Establishments

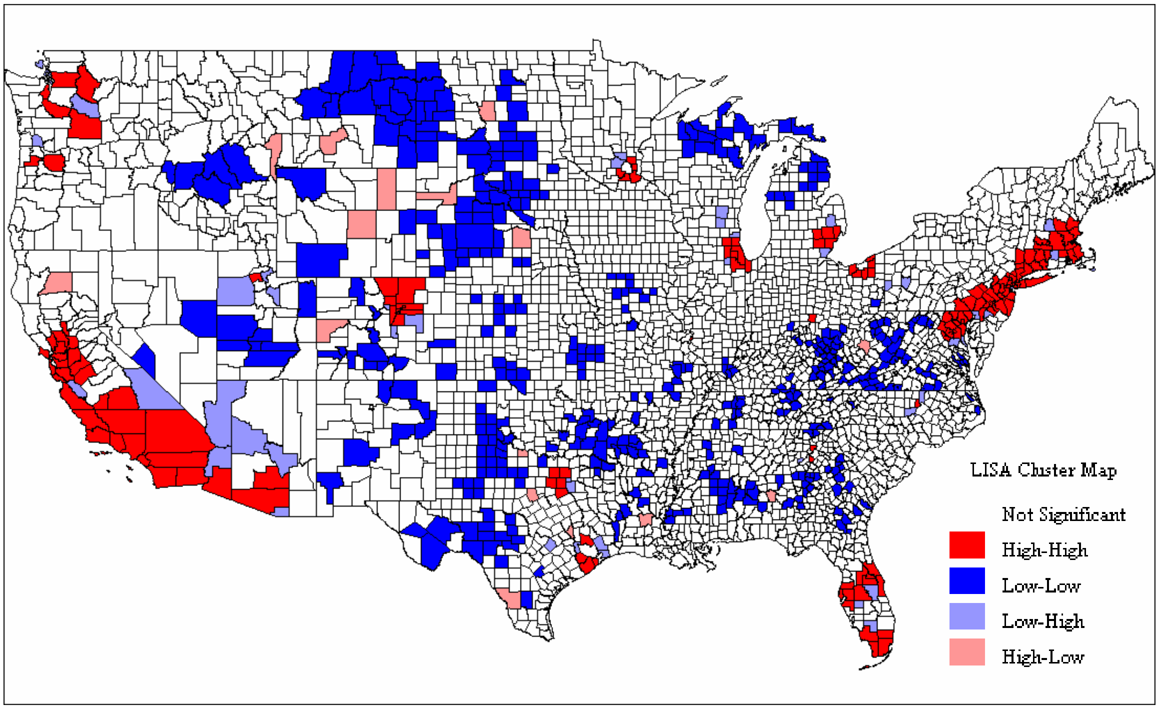


Figure 7. Local Indicator of Spatial Association (LISA) Significance Map for Biotech Establishments

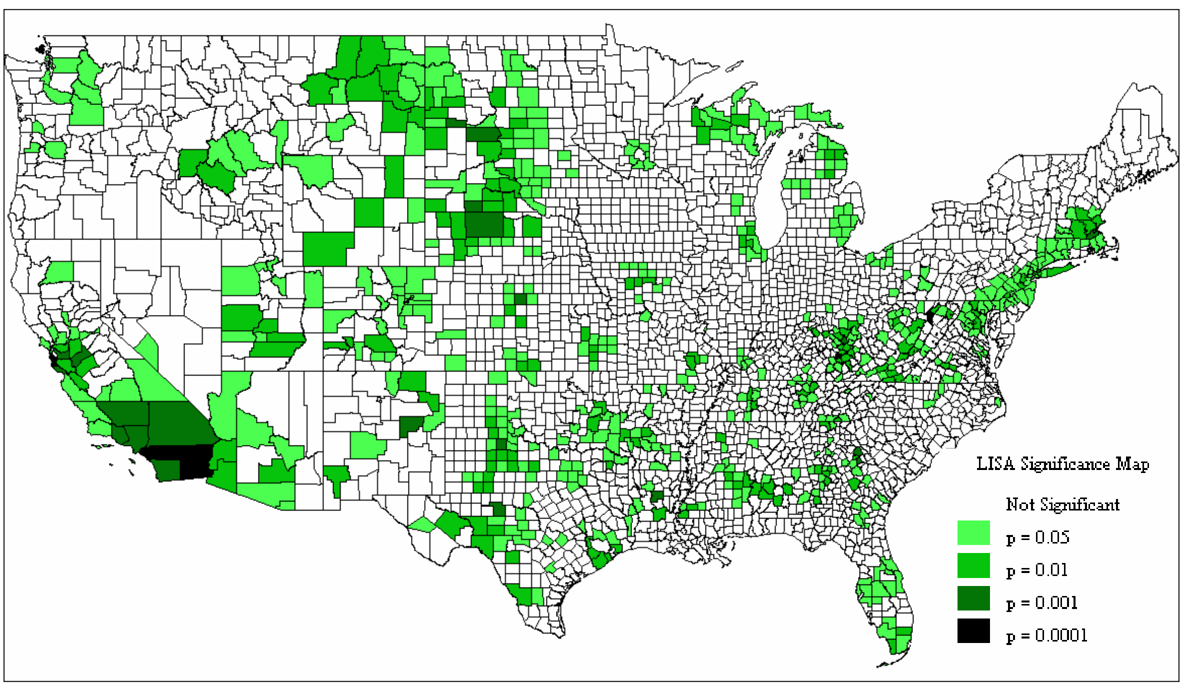


Figure 8. Bivariate LISA Cluster Map for Farmland and Spatial Lag of Biotech Establishments

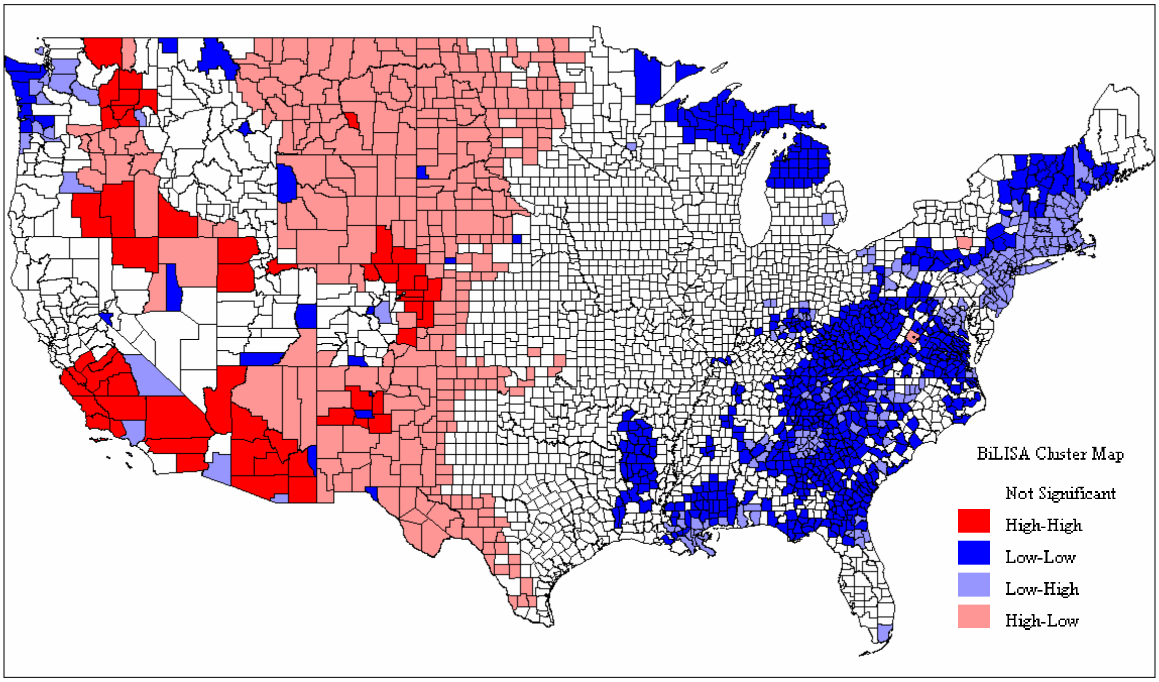


Figure 9. Bivariate LISA Significance Map for Farmland and Spatial Lag of Biotech Establishments

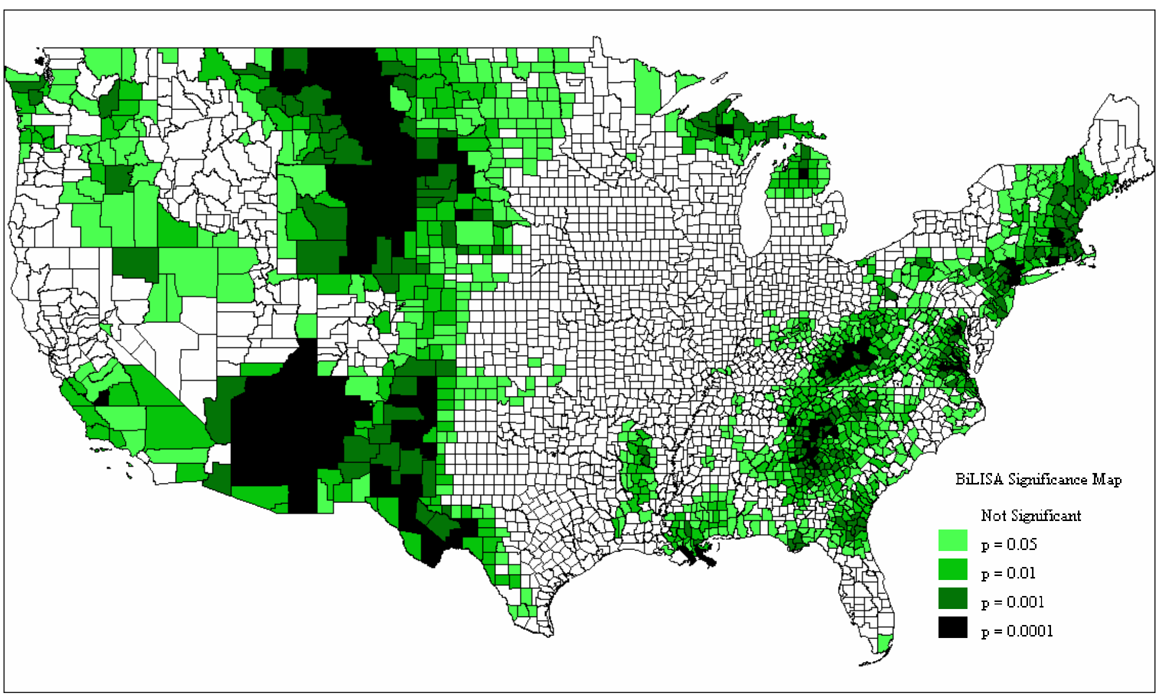


Table 1. Summary Statistics of Variables

Variable	Mean	Std. Dev.
Total Number of Biotech Establishments	11.50	41.64
Rural-Urban Continuum Code	5.11	2.68
Poverty Rate	13.36	4.89
Population (in 1000s)	92.92	303.47
Owner-Occupied Housing Units: Median Value (in $1000)	80.93	41.94
Unemployment Rate	5.97	1.96
Median Household Income (in $1000)	36.73	9.28
Total Number of Venture Capital Firms	1.89	11.45
Number of Colleges, Universities and Professional Schools	1.08	4.86
Number of General Medical and Surgical Hospitals	1.75	3.70
Average Wage per Job (in $1000)	27.05	5.79
Percent of Persons with a College Degree	16.51	7.80
Property Tax (in $1000)	22.74	23.41
University Life Science R&D Expenditures (in $1000)	563.61	573.18
NIH Support for Institutions (in $1000)	427.71	516.33
Higher Education Degrees in Biological Science (in 100)	27.26	21.44
Biological Scientists in Workforce 2000-2002 Avg (in 100)	118.38	106.01
Crime Index (in 100)	35.95	139.73
Land in Farm Acres (in 1000)	301.14	385.21

Table 2. Tests for Spatial Correlation in Residuals of a Regression Model

Test Statistics	Value
Moran's I Statistic	9.099^{***}
Lagrange Multiplier	77.110^{***}
Likelihood Ratio	69.594^{***}
Wald	812.299^{***}

Note: *** indicates significance at the 1% level

Table 3. Estimates of Factors Affecting the Location of U.S. Biotech Industry

Variable	Coefficients: Spatial Tobit	Coefficients: Tobit	Marginal Effects: Spatial Tobit	Marginal Effects: Tobit
Constant	-1.9758**	-42.5502***	-0.9702	-27.2834
	(0.9230)	(5.9758)		
Rural-Urban	-0.0109	0.3104	-0.0054	0.1991
	(0.0345)	(0.2250)		
Poverty	-0.0033	0.3548**	-0.0016	0.2275
	(0.0233)	(0.1680)		
Population	0.0294***	0.0968***	0.0144	0.0620
	(0.0086)	(0.0056)		
Median Housing Value	-0.0162***	-0.0021	-0.0080	-0.0013
	(0.0037)	(0.0197)		
Unemployment	-0.0508*	-0.2555	-0.0250	-0.1638
	(0.0365)	(0.2871)		
Median Household Income	0.0695***	0.2875**	0.0341	0.1844
	(0.0213)	(0.1220)		
Venture Capital	0.4369***	0.4663***	0.2146	0.2990
	(0.0842)	(0.0573)		
Colleges	0.5311***	1.5823***	0.2608	1.0146
	(0.1024)	(0.2166)		
Hospitals	0.4785***	0.3099	0.2350	0.1987
	(0.0807)	(0.3255)		
Average Wage Per job	0.0265**	0.5949***	0.0130	0.3815
	(0.0157)	(0.1117)		
Education	0.0541***	0.4186***	0.0266	0.2684
	(0.0140)	(0.0883)		
Property Tax	0.0213***	0.0087	0.0104	0.0056
	(0.0067)	(0.0227)		
Life Science R&D	-0.0010	-0.0193***	-0.0005	-0.0123
	(0.0008)	(0.0047)		
NIH Support for Institutions	0.0009**	0.0156***	0.0005	0.0100
	(0.0005)	(0.003)		
Higher Education Degrees in Biological Science	0.0061	0.038	0.0030	0.0247
	(0.0122)	(0.0869)		
Biological Scientists in Workforce	-0.0008	0.0385*	-0.0004	0.0245
	(0.0031)	(0.0201)		
Crime Index	0.0051	-0.0657***	0.0025	-0.0422
	(0.0093)	(0.0095)		
Farm Land	0.0002*	0.0032***	0.0001	0.0021
	(0.0002)	(0.0011)		
rho	0.0235***			
	(0.0071)			

Note: *,**,*** Statistical significance at the 10, 5, and 1% levels, respectively.