Undergraduate Economic Review

Volume 11 | Issue 1 Article 9

2015

Top of the Order: Modeling the Optimal Locations of Minor League Baseball Teams

W. Coleman Conley
Emory & Henry College, wcolemanconley@gmail.com

Recommended Citation
Conley, W. Coleman (2015) "Top of the Order: Modeling the Optimal Locations of Minor League Baseball Teams," *Undergraduate Economic Review*: Vol. 11 : Iss. 1 , Article 9.
Available at: https://digitalcommons.iwu.edu/uer/vol11/iss1/9

Top of the Order: Modeling the Optimal Locations of Minor League Baseball Teams

Abstract

Over the last twenty-five years, minor league baseball franchises have defined firm mobility. Revisiting the work of Michael C. Davis (2006), I construct a logistic regression model to predict which cities house minor league baseball teams. Six variables are tested for inclusion in the model, including population, income level, the number of major-league professional sports teams in a city, five-year population change, and distance from the closest professional team. Based on the model's predicted probabilities, cities are ranked in order of highest probability of having a team at each of the different levels from Class A to Class AAA.

Keywords

Baseball, Econometrics, Spatial Economics, Location Theory

Cover Page Footnote

A special thank you to Dr. Linda Dobkins, the only professor I know whose love of the game is equal to my own. I would also like to thank Dr. Joseph Lane, Dr. Christina Carroll, and Dr. Julia Wilson for serving as my thesis advisors. Complete results can be obtained by emailing wcolemanconley@gmail.com.

INTRODUCTION

Minor League Baseball's relationship to the major leagues makes its structure both complex and unique. Every minor league team is affiliated with a Major League Baseball (MLB) team which is responsible for hiring its players and coaches. There are also seven different levels of domestic minor league teams ascending from Short-Season Rookie to AAA. These classifications serve as preparation and extended tryouts for players to become major leaguers (Ozanian 2008). While my research focuses on the A, A-Advanced, AA, and AAA levels, the 30 MLB teams will, in most cases, have one affiliated team for at least six levels. The MLB's annual fifty-round player draft helps fill out all of these roster spots, where players will begin at some lower level and work their way up toward the major leagues. Because of the complexity of this system, it is not uncommon for the top draft picks to spend several years in the minor leagues before reaching their dream in the MLB. Although a player at any level can be called up to the major leagues, this usually happens at the AA and AAA levels.

All of this does not even touch on the fact that there are leagues outside of the United States, such as the AAA Mexican League, as well as completely unaffiliated, independent professional baseball leagues. Still, if all professional baseball teams are a special subset of economic firms, minor league affiliates are even more unique. Although the Atlanta Braves own nearly all of their affiliates, most minor league teams are privately owned by entities other than their affiliate. Despite this independence, minor league teams accrue huge benefits from having major league affiliates: the players and coaches are compensated by another firm, and their relationship to the MLB comes with a certain amount of brand recognition (Smith 2013). Some other typical assumptions of firms do not apply at all to minor league teams (Bruggink and Zamparelli 1999). For example, while a manufacturing firm might look for a location already high in skilled labor, this is not much of a consideration for a team whose labor effectively comes to them.

Still, the fact remains that a minor league baseball team is an economic firm. In this case, the firm cares about its location largely because it relies on the market area around it to draw spectators to games. For this reason, choosing an optimal location is just as important for a baseball team as it is in any business venture. Unlike the MLB, which has experienced only one major relocation since 1998, minor league team locations change at a more rapid pace. The ultimate purpose of this research is twofold: first, to review the model created by Michael C. Davis (2006) and learn how accurately it predicted the changes in the minor league landscape from 2006-2012, and secondly, to use Davis' work as a starting point to model the ideal locations for minor league baseball teams using demographic, economic, and geographic data from several different sources.

The model will be similar to that of Davis: population, population change, per capita income, distance from the nearest professional baseball team, distance from the nearest major-league baseball team, and the number of a city's major-league professional teams in other sports will be tested as independent variables for their relationship to whether or not the city in question houses a minor league team at a given level. This research will focus on the most prominent levels of minor league baseball (AAA, AA, A-Advanced, and A); short-season and unaffiliated leagues will not be analyzed. The modification of some assumptions, use of Multinomial Logistic Regression instead of a Generalized Ordered Logit Model, and addition of new variables differentiate this research from that of Davis.

LITERATURE REVIEW

There are several theories in spatial and urban economics which help us understand this specific case of team location. One of the simplest ways that a baseball team is related to economics is through spatial monopoly power. Due to league rules which respect the territory of previously established teams, a new ball club must stay at least eighteen miles away from the nearest team at any level (Siegfried 1981). Given these restrictions, each team is effectively granted a baseball monopoly for all consumers in the closest eighteen miles, and even that is the bare minimum. On the other hand, just as a manufacturing firm might reduce transportation costs by locating near a supplier, some teams are located geographically close to its affiliate. For example, the AAA affiliate of the Atlanta Braves is located just north of the city in Lawrenceville, GA. Although classical economic theory often assumes that distance is not a major factor in the cost of doing business, baseball franchises are examples of why such an assumption is unrealistic (Krugman 1991). Not only might a team look for spatial monopoly power or proximity to its affiliate, but the cost of relocating may be more than any benefits of a much better location (McCann 2001).

Since a professional baseball team is a firm in an economic sense, it is easy to assume that it exists solely in order to make profit. However, this is not necessarily true, particularly in the case of past minor league teams. New behavioral theories of firms have helped us understand the alternative goals a firm might choose, as there may be competing objectives for each level of employees in the team ownership hierarchy (McCann 2001). While the owner may be most interested in making a profit, sales directors may be evaluated based on gate revenues, making that their personal goal. And, of course, each minor league team is affiliated with a major league club, whose main goal might be to prepare their future players. Hence, the assumption of pure profit maximization is misleading, particularly since in smaller markets the level of profits is hardly competitive with teams in larger areas (Scully 1995). That said, there is a growing emphasis on finance in the minor leagues since the "moneyball revolution" in the early 2000s, and some well-run teams have proven to be quite profitable (Smith 2013).

One interesting point is that a minor league team may be interested most in maximizing revenue, which does not always happen at the point of profit maximization. In carrying out this strategy, the team would attain their maximum short run market share (McCann 1995). Since for those 176 teams that charge admission, over 90 minor league baseball teams relocated between 1991 and 2009, the short run may be the time period that matters most for some teams (Van Den Berg 2009). While gate receipts are not the only revenue source for these teams, all other sources are driven by bringing people in the gates. The real question, then, for the first scholars in this field, was modeling baseball demand by whether or not a location would draw enough fan attendance at games.

The research of the late 20th and early 21st centuries was instrumental in demonstrating the strong positive relationship between population and game attendance

(Brandvold & Pan 1997; Siegfried & Eisenberg 1980; Bruggink & Zamparelli 1999). Today we can authoritatively assert that a heavily populated area is more likely to result in more people coming through the gates. However, whether minor or major league teams were analyzed, there seemed to be inconsistent results for what else drives the demand for baseball games. Winning percentage (Brandvold & Pan 1997, Bradbury 2007), racial makeup of the area (Siegfried & Eisenberg 1980), and income level (Bruggink & Zamparelli 1999, Racher 2008) were all considered important factors in some studies, while others did not find a significant relationship with attendance.

While most attendance studies had focused on the major leagues, Rodney J. Paul's two 2008 studies analyzed many different factors to find which were the most and least important in causing fans to attend minor league games. More than twenty variables, including day of the week, population, per capita income, win percentage, free tickets, and various promotions, were measured to create an econometric regression model. What is striking about his research is that different leagues seem to have different factors driving attendance numbers. For example, his work on the Northwest League suggests that people are most likely to come to games when the team is winning and scoring runs at will. The only important promotion which significantly affected attendance was firework night. Attendance for teams of the South Atlantic League is also positively related to high-powered offensive baseball, and some of the best attended games in the SAL are accompanied by concerts or fireworks. However, in the South Atlantic League, minor league game attendance is not affected significantly by win percentage, and there was a surprisingly different negative relationship between attendance and per capita income. Paul concluded that the total entertainment package mattered more than baseball, and that these were simply different tastes for baseball consumers in different league areas. The question of winning percentage's effect on attendance was compounded further when a fifteen-year data set showed a significant positive relationship (Gitter and Rhoads 2008). This study also went on to discuss how minor league teams might act as substitutes for nearby major league affiliates. This was particularly true if the two teams were consistent winners, making their main conclusion even stronger.

Finding cities with a higher baseball demand seemed to be a key reason for the constant relocation of minor league baseball, but even assuming that there is a significant demand for baseball in a given city leaves out other important complications. Most of these stem from the fact that baseball teams require a stadium, and it is often the case that the public must want baseball in their city so much that they are able to convince city administration to finance a new stadium (Davis 2008). If done correctly, both the city and the team can enjoy an upward spiral effect of stadium quality, as “increased revenues from a new arena put a franchise in a better position to bid for quality players, resulting in a better team, which, in turn, draws more fans, resulting in more revenues, and so on” (Rascher 2004, 275). Still, the millions of dollars required to build a new stadium is not often an easy sell, adding a hefty cost barrier to an otherwise attractive location.

Even to this day, most sports economics research focuses on the effects that housing a sports team has on the local area. It is often assumed, for example, that having a professional sports team might be the cause behind an increase in population or income level. Instead of taking this approach, Michael C. Davis flipped it on its head to analyze the reasoning behind minor league franchise location in the first place (2006). Could it be that the assumption is one of reverse causality, whereby population and income level make a city desirable as a minor league team's home? Instead of using attendance as his dependent variable, Davis examined the actual placement of minor league teams through a Generalized Ordered Logit Model and found that those cities with higher population, higher income level, and a greater distance from a major league team are more likely to house a minor league team (2006). He also used his model to predict which cities would become future homes of minor league teams, and which cities would not house a minor league team much longer.

Davis later built on that research by studying which other factors affect the location of any team at any level of professional baseball (Davis 2008). To say the least, there is a lot more to team location than choosing a city with a high population. For example, one other consideration is the previously mentioned league rule preventing new teams from locating within close proximity to another team. This factor looms large in New York, whose population houses two major league teams but could possibly support even more, particularly at a minor league level. Another important factor is the distance from a minor league team to its major league affiliate, although this seems to only be a factor at the higher levels. Beyond that, factors can include team winning percentage and the general disposition of the population to different entertainment options. In this way, Davis used his 2008 research to expand into a broader lens which included the ideas of scholars like Paul and Siegfried to again ask what brought people to games in the first place. Needless to say, Davis' work is the inspiration for this research, whereby the same variables will be tested for the creation of a new model.

EVALUATING DAVIS

At the end of his 2006 work, Davis suggested that future research could be dedicated to the movement of teams over the course of time to see whether or not there is conformity to his model. Included in his work were two helpful lists: one of the top ten cities without a team at a given level that the model predicts would house a team, and the bottom ten cities with a team at a given level that the model predicts would not house a team. For the purpose of comparison, these tables have been recreated and are included as Tables A1 and A2 in the Appendix, with the overall city ranking in parentheses. I was able to obtain a more extensive listing of Davis' final rankings with his permission to determine whether his model correctly predicted the franchise moves made in the past six years. After further analysis, I learned that his methodologies required the predicted

Table 1.

2006-2012 MiLB Franchise Moves and Viability Rankings (Davis 2006)						
Class	Original City	Rank	New City	Rank	Change	Year
LoA	Battle Creek, MI	136	Midland, MI	236	-100	2007
AAA	Ottawa, CAN	N/A	Allentown, PA	32	N/A	2008
AA	Wichita, KS	3	Springdale, AR	144	-141	2008
AAA	Richmond, VA	23	Atlanta, GA (Lawrenceville)	N/A	N/A	2009
AAA	Tucson, AZ	35	Reno, NV	69	-34	2009
HiA	Vero Beach, FL	148	Bradenton, FL (Port Charlotte)	11	137	2009
LoA	Columbus, GA	1	Bowling Green, KY	314	-313	2009
AA	Norwich, CT	102	Richmond, VA	22	80	2010
HiA	Sarasota, FL	11	Bradenton, FL	11	0	2010
AAA	Portland, OR	1	Tucson, AZ	35	-34	2011
AA	Raleigh, NC (Zebulon)	55	Pensacola, FL	43	12	2012
HiA	Kinston, NC	413	Raleigh, NC (Zebulon)	140	273	2012
				OVERALL SUM	**-120**	

Note: The viability rankings were obtained from Michael C. Davis and are based on his model's predicted probabilities of housing a team at each given level.

probabilities of all cities as well, rather than just those without a team. With this data, I compiled Table 1, which lists all cities involved in franchise moves during those years and their viability ranking at the relevant level.[1] Change refers to the net change in ranking due to the move, and was calculated by subtracting the new ranking from that of the first city. While I was able to move forward with this data, even having full disclosure did not allow for a perfect analysis due to methodological issues. One stems from the fact that some of the observed franchise moves are simply outside the scope of his research. For example, we do not know how to interpret the relocation of a AAA team in Ottawa to Allentown, PA simply because Canadian cities were not analyzed. Also, cities with MLB teams were not analyzed by Davis for a minor league presence, and the relocation of a Braves affiliate from Richmond, VA, to a part of the Atlanta area posed an issue here as well.

Despite these limitations, we do see that there is some variability when it comes to consistency with the model. Indeed, some of his "Bottom Ten" cities, which housed teams despite inconsistency with his model, saw their franchise move to a different location at some point from 2006-2012: Tucson, AZ; Vero Beach, FL; Norwich, CT; and Kinston, NC. In these cases, three moved to a city with a higher ranking, although none were a "Top Ten" city. However, while the move from Tucson, AZ, was predicted by the model, its new location of Reno, NV was actually rated as even less-consistent with the model. Two years later, when Tucson received a new team, its previous home of Portland, OR, had actually been the highest-ranked city in Davis' AAA-model. The other franchise

[1] While there were 90 relocations from 1991-2009, this includes all levels of Minor League Baseball. Our study is limited to the top four levels, where there is a lesser, albeit considerate, degree of change.

moves, with only a couple of exceptions, were often overwhelmingly in contrast with Davis' predicted probabilities. Cities like Columbus, GA and Wichita, KS echoed the story of Portland, with a very highly ranked city losing its team to another city with a much lower ranking.

One can also see that many of the moves were to cities with a ranking in the triple-digits, raising questions about the effectiveness of the model's predictive power. While Davis was effective in predicting that some locations would not house teams much longer, there is some very obvious inconsistency with his model. To be clear, this is not because Davis' research was poorly conducted or fundamentally flawed in its approach: it simply points to the fact that minor league baseball team location is an enigma that even the best scholars cannot predict with a great amount of accuracy. However, any new knowledge that we can gain from our own analysis can be helpful in making strides toward better predictions. Based on this conclusion, I attempt to construct a similar logit model which correlates independent variables, such as population, five-year population change, per capita income, and distance from a professional baseball team, to the presence of a minor league team in a city. Those cities which best fit the model, according to their model-predicted probabilities, will be designated as optimal team locations.

In doing so, I will also challenge a few of Davis' assumptions. Most notably, Davis chose to exclude the cities with major-league teams from consideration in his analysis in order to use distance from a major league team as a variable (2006, 257). While I understand this logic, especially when compounded by the population outliers that New York and Los Angeles represent, doing so leaves out a considerable number of minor league teams. For example, the Los Angeles area houses five of the ten teams in the California League. Excluding these cities leaves us unable to say anything at all in situations like these, so I will include all such cities in my research. After all, if we know that some cities house both major and minor league teams, it would be interesting to know which cities seem most capable of doing so. Distance to a major league team will still be included as well, coded as "0" for those cities which have an MLB team. Other challenges will be brought up at a more appropriate time in the methodology and results sections.

DATA & MEASURES

Defining the meaning of the term "optimal" is crucial to understanding the conclusions of this research. Like Davis, we are interested in looking into the factors that may or may not lie behind the choice of a particular city for a baseball team. These are examined through an inductive lens which takes the locations of teams in 2006-2012 as given. We assume that, by and large, many of the best opportunities for team locations are already housing a team, and, therefore, are looking for the data that can best predict similar locations. "Optimal," then, does not imply that these are truly perfect locations for baseball teams, but rather describes cities that are consistent with the values of variables used to model the current landscape. Although there are clear drawbacks to this approach, it should allow us to pick up on any similar cities that do not currently have a minor league team.

Population and per capita income data was obtained for all Metropolitan Statistical Areas (MSAs) in the United States from the Bureau of Economic Analysis. Five-year population change was calculated using this data and additionally converted into percentage change. As these all represent numerical values which begin at zero, they are classified as ratio variables. All data from 2006-2012 were collected into a spreadsheet organized by individual cities. Before moving on, it is important to remember that "city" for the purpose of this research, refers to an MSA as defined by the Office of Budget & Management. Just as Davis classified cities in his work, in the case that the MSA was part of a larger Combined Statistical Area (CSA), the Combined Statistical Area numbers were used. This is because a CSA is composed of at least two geographically adjacent Metropolitan Statistical Areas whose local economies have been judged to be interconnected.

Sometimes this resulted in several large cities being analyzed as one (ie. San Jose-Oakland-San Francisco, CA), but the relationship between the cities justifies the CSA as the better unit of analysis for economic data. In several rare instances, teams were located in smaller Micropolitan Statistical Areas that were not part of a larger CSA. For these special cases, data was obtained for the respective Micropolitan Statistical Areas. Originally, the model was to include two other demographic variables for each city: racial makeup and age makeup. In short, these would have measured the percentage of each city's population identifying with a particular race or age grouping. Although this information is available through the U.S. Census Bureau for Metropolitan and

Micropolitan Statistical Areas, it is not easily convertible into usable data for CSAs.[2] Therefore, these variables were not used in the creation of this study's model.

Non-demographic variables were also used. The number of “major-league” professional sports teams in each city was obtained through the National Football League (NFL), National Basketball Association (NBA), National Hockey League (NHL), and Major League Soccer (MLS) websites. The summation of the number of teams in each city was then tested for use in the final model. Geographic variables were developed through Google MapsEngine to measure the shortest distance from each city to the closest city with a professional baseball team. The distance was measured in miles from the center of one city to the center of the other, and in the case of CSAs, the city listed first in the CSA name was used as the starting point. There were two different variables based on this method: the distance to the closest MLB team and the closest minor league team outside of the city being analyzed. For help in visualization, a map of the Minor League Baseball Landscape is provided as a figure at the beginning of the Appendix.

The individual locations of minor league baseball teams were obtained from the 2006-2012 Baseball America Directories, annual publications which contain all contact information, addresses, affiliates, and historical data for each team in Minor League Baseball (MiLB) that year. This data was first collected into a spreadsheet organized once again by individual cities. The number of teams at each level of the minor leagues (AAA, AA, A-Advanced, A) were coded for each city. A final column added up the number of

2 One might assume that we could simply sum or average the numbers for all of the smaller areas that are part of a CSA, but this is not necessarily true. Especially in the case of Per Capita Income, values would need to be carefully weighted by the population of each area to generate accurate numbers.

teams at each level to determine the total number of minor league teams in each city. Collecting data for the locations of minor league teams meant that sometimes a city had more than one team within its Metropolitan or Combined Statistical Area. For this reason, these dependent variables were recoded into an ordinal measure of whether or not a city had at least one minor league team (0 = No team; 1 = At least one team). These ordinal measures were used as my dependent variable in Binomial Logistic Regression. Another, perhaps more important, dependent variable recorded the Highest Level at which a city had a minor league team (0 = No team; 1 = A; 2 = A-Advanced; 3 = AA; 4 = AAA). Thus, both Multinomial and Ordinal Regression used Highest Level as the dependent variable.

A separate data set was collected based on data almost entirely from the Baseball America Directories. Organized by team name, this data set recorded the classification, league, affiliate, address, and average attendance in the past year for each team. Google MapsEngine was used for one variable in this data set as well, using the same convention to measure the distance from the team's city to their MLB affiliate. In this case, the “team's city” was not the first major city listed in the MSA or CSA name, but the actual town where the team plays. For example, the New Orleans Voodoo's stadium is in Metairie, Louisiana, so distances were calculated from the center of Metairie rather than New Orleans.

The final variables were reported in different units: population was transformed into the natural logarithm of population because of the large amount of skew in the data from cities like New York and Los Angeles. The transformed variable, Log Population, gave a much more normal and helpful representation of the data. Per Capita Income was transformed as well into the more normally distributed Log Per Capita Income. Numerical Population Change was reported in hundred-thousands and rounded to five decimal places, while the Percentage Population Change was coded as a decimal rounded to 3 decimal places. MLB Distance and Baseball Distance were reported in hundreds of miles and, therefore, rounded to two decimal places. The difference between MLB Distance and Baseball Distance is that MLB Distance gives the distance from the city to the closest major league team, while Baseball Distance gives the distance from the city to the closest minor league baseball team outside of the city being analyzed.[3] For this reason, the values of these two variables could be very different for any city. The number of major-league professional sports teams in the city was simply coded as a whole number due to summation.

HYPOTHESES

Based on initial suspicions and the work of previous scholars, I developed the following hypotheses. I believe that the model will predict that:

> *H1: A higher city population, in this specific case the natural logarithm of population, will be associated with a higher probability of having a team at all levels.*
>
> *H2: A higher per capita income for a city, when transformed by the natural logarithm, will be associated with a higher probability of having a team at all*

3 Although cities in Alaska and Hawaii were included in this analysis, their values for these variables were capped at 1,000 miles. This allows us to capture the fact that they are basically limited to plane travel without having outliers so massive (greater than 2,000 miles) that they nullify our results.

levels.

H3: A higher five-year population change, whether by percentage or raw number, will be associated with a higher probability of having a team at all levels. Only one of these measures will be used in the final model to avoid multicollinearity.

H4: The presence of major-league professional sports teams in a city will be associated with a higher probability of having a team at all levels.

H5: A greater distance from the closest professional baseball team will be associated with a higher probability of having a team at all levels.

H6: A greater distance from an MLB team will be associated with a higher probability of having a team at all levels except AAA. For AAA, the relationship will be inversely proportional.

H7: All variables will show significance and be moderate in strength, but there will not be multicollinearity between any variables other than population and raw population change.

The justification for Hypothesis 1 relies on the assumption that population is the best measure of a city's market area. Like any other business, a baseball team will choose to locate in a high population area in hopes of bringing more fans to each game. Hypothesis 2 assumes that cities with a higher per capita income will have the disposable income necessary to attend a minor league baseball game. Although tickets to these games hardly command a high price compared to a major league game, minor league baseball is a recreational activity. Despite their affordability, baseball is not a necessity and should show an elastic demand among the general public. The third hypothesis is one of my own additions to the Davis model, the theory being that population change over the last five years helps show the trend of whether a city is growing or deteriorating. Thus, a minor league team might locate in a smaller population area if there is enough positive population change to show evidence of a promising future. Although these two population change variables will be measuring the same concept relevant to the model, I think it is important to test both to see which is the better indicator of population change. Hypothesis 4 is in regards to major-league sports teams other than baseball, that is, teams in the NFL, NBA, NHL, and MLS to round out the top five American sports. The theory behind this is that seeing another professional team locate in a city may serve a purpose in signaling it as a location which already has some sports popularity.

Concerning the distance to the closest professional baseball team, my fifth hypothesis assumes that there is some spatial monopoly power that a professional baseball team can gain by locating farther away from the next-closest team. It is important to note again that only teams from the AAA to A classifications are being analyzed. Thus, while there may be some teams at lower levels that are technically the “closest professional team,” they will not be considered. Similarly, the sixth hypothesis, for the most part, assumes that there is some spatial monopoly power to be gained. This would reduce their competition for the highest level of baseball, resulting in potential for a greater market share. The AAA-level is a special case, however, in that it is only one level below the major leagues. Therefore, locating close to a major league team, potentially its MLB affiliate, would create an avenue for positive externalities. For example, injuries are quite common in baseball, and often require the injured player to have a temporary rehabilitation assignment to a lower level. The team will pick up a replacement player from a level below to fill out their roster as well, so by locating near each other, the AAA and MLB teams would be able to do this more easily and efficiently.

All of this leads me to my seventh and final hypothesis regarding the relationships between the independent variables. If scholars like Davis used per capita income, population, and distance variables together to build their past models, I expect to see a moderate relationship between them. Although the presence of other professional teams is my addition, I do not expect the relationship to resemble population too closely, and expect its bivariate relationships to be moderate as well. As for the two measures of population change, I believe that the numerical measure will too closely resemble the population variable, resulting in a very strong, collinear relationship. On the other hand, I predict that the percentage change in population will create a level playing field for cities of all sizes, making it the better measure of the two. As it should be measuring something different from population and per capita income, I expect their bivariate relationships to be moderate. As a reminder, it has already been stated that we hypothesize that using both measures of population change would result in multicollinearity.

METHODS & ANALYSIS

Using SPSS, the independent variables were tested using several different methods. Initially, Pearson's R was used to test the relationships between pairs of variables. This was done for two reasons: first, to confirm that the measures were at least somewhat related, and secondly, to test for multicollinearity between variables. Ideally, if these variables are compatible for use in creating a viability index, the relationships between the variables should be significant. However, if the relationship is too strong, the possibility of multicollinearity means that both probably should not be included in the index.

Davis used a Generalized Ordered Logit Model (GOLM), which is a form of logistic regression to produce predicted probabilities of a city having a team at a given level. Whereas the more common linear regression model is helpful for predicting scale dependent variables with a large range, logistic regression models categorical data and finds its predictive power in the logarithm. Linear regression is more powerful when used with continuous dependent variables, such as population. However, our analysis has only a few different dependent variable categories: even when we are analyzing all levels together, there are only five levels from AAA to having no team at all. In this case, it seems clear that logistic regression, also known as a logit model, is the better choice. Since there is some rank-order behind the levels of minor league baseball classifications, we begin with the assumption that our dependent variable is ordered. This is in contrast to a dependent variable such as hair color, which clearly has no order. Generalization is more complex to explain, but basically allows for more flexibility in modeling through the relaxation of some assumptions that an ordinal logistic regression requires, such as the proportional odds assumption we will discuss later.

For several reasons, this analysis will not use a GOLM. Perhaps the most important reason is that I am personally much less familiar with a GOLM than I am other techniques, and do not believe that I am in a position to replicate Davis' methodology. At the same time, the new data set that I am analyzing has undergone several changes in makeup, including updated geographic CSA definitions from the U.S. Office of the Budget & Management. Also, the years from 2006-2012 are being analyzed together, making this study longitudinal in contrast to Davis, whose analysis was cross-sectional. Furthermore, given that the changes in minor league team location that have occurred in the last several years were only sometimes consistent with Davis' model, as we have already seen, it may be that his GOLM is not the best way to model the locations of minor league baseball teams after all. Therefore, our goal is not as much about replication as it is about broadening the usefulness of our results.

All of these factors suggest that it might be smarter to use several other regression methods to find out which best suits our data. Over the course of the analysis, three different forms of logistic regression were used: Binary, Ordinal, and Multinomial. While we will explain the reasons for the different models later, it makes sense to first cover the similarities of all three of these logit models. Again, unlike a simple regression, the line-of-best-fit will not be linear, but based on the shape of the logarithm function. The model produces an equation based on an intercept and coefficients for each variable, corresponding to the log-odds of being in a category based on the exponential function. In reading the model output, the sign of our model coefficients, B, represent a direct or inverse relationship as intuition would tell us. For each one-unit change in a variable, the log-odds will change by the value of B. Because the variables are measured in different units, we cannot say anything about the relative strengths of different values for B until they are converted into actual odds ratios, denoted as Exp(B).

It is important to note that a logit model produces odds ratios here, which are different from probabilities. For this reason, a higher value for Exp(B) represents more strength, but an odds ratio that is two-times the size of another is not twice as strong. As in other regression models, variable significance is measured by our p-value ($p>.05$ is considered insignificant) based on the Wald chi-square test. Finally, although linear regression uses r-square values as a measure of the model's goodness of fit, logistic regression has no perfect r-square measure. Several "pseudo r-squares" have been developed in response to this, and we will report the Cox & Snell and Nagelkerke values here. Low values for these statistics are common, and even though Nagelkerke is built to give a value from 0 to 1 (unlike Cox & Snell), it is still not very comparable in interpretation to conventional r-square.

The mathematics behind this can be seen if we let $\boldsymbol{\alpha}$ represent the intercept of the model, $\boldsymbol{\beta_i}$ represent the coefficients for each variable, and $\mathbf{x_i}$ represent our variables for some integer, i. If **p** is the probability for x, then the model's log-odds for **n** variables can be represented by:

$$\log \frac{p(x)}{1-p(x)} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

We would then calculate the actual odds ratios for a city having a team at the level in question, represented by the expression:

$$e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n}$$

However, because odds ratios can cause confusion in interpretation, all odds ratios were then converted to probabilities which predict the chance of a city having a team. The calculation of the predicted probabilities in some forms of logistic regression is complex and will be discussed shortly, but for a Binomial Logistic Regression, predicted probabilities are simply expressed as follows:

$$\frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}{1+e^{\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}$$

Furthermore, the three types of regression used in our model are all different despite their parental relationship. The Binary Logistic Regression model uses a dependent variable which has only two possibilities. In our case, a value of “0” was used to represent the absence of a team, and a value of “1” was coded if the city has at least one team at the level being analyzed, so our independent variables were used to create a non-linear line of best fit between the values of 0 and 1. Naturally, this was used for our analysis of whether a city had a team at any minor league level. Ordinal Logit modeling is based on several assumptions, including the premise that there is order to the categories of our dependent variable. As Davis suggests, this can be assumed because of the inherent hierarchy of minor league baseball where players are promoted from A up the ladder toward AAA, and then to the major leagues. This leads to the proportional odds assumption that all independent variables should have the same effect in terms of its log-odds at all levels. Finally, Multinomial Logistic Regression expands upon the binomial case in allowing more than 2 possibilities for the dependent variable. In contrast to an ordinal model, multinomial categories are based on the assumption that each should be treated separately. Perhaps, like hair color, no category is better or worse than any other.

For this reason, each category of the model is compared to a reference category. In our case, category 0 (No Team) is our reference category.

Multinomial predicted probabilities are a bit more complicated, but are based off of the same logic as binomial. Again, the basic odds ratio for a binomial regression is below.

$$e^{\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}$$

We will call this particular equation e^A. Suppose that we have four different values for coefficients of each variable based on the different categories, denoted e^A, e^B, e^C, and e^D. Then the predicted probability of a city having a team in category A, out of all the other options, would then be:

$$\frac{e^A}{1+e^A+e^B+e^C+e^D}$$

These models will provide output for each individual level of Minor League Baseball as well as the overall presence of a minor league team at any level. The theory behind this is that, as we have seen from previous research, the effects of a variable at one level may not be the same either in significance or direction for another. The predicted probabilities given through these models will serve as our viability indexes, and cities will then be ranked at each level in descending magnitude of probability.

RESULTS

Descriptives & Initial Variable Testing

The univariate analyses for our dependent variables are provided in Tables 2, 3, and 4. Again, all years from 2006-2012 were analyzed, so while there are a total of 2030 cities for all years combined, 290 distinct cities were included in this study. Table 2 categorizes all cities by the number of teams at each level. The MLB is included as well for comparison purposes, as well as a breakdown of cities which have a team at any level. We should not be surprised that the vast majority of cities do not have a minor league team, especially considering there are only 120 MiLB teams being analyzed for each year. Also, the only classification which has no cities housing more than one team is AAA. That said, the only level where there is a very pronounced effect of cities with more than one team is A-Advanced, where Greater Los Angeles houses as many as five teams in a given year.

Table 3 resolves the issue of multiple teams locating in the same city by creating a dummy variable of whether or not a city has at least one team at a given level. Once again, numbers for the MLB and Any Level are included. Most of the conclusions we can draw from it are similar to those previously mentioned, such as the low number of cities with a team. The Binomial regression will use this dichotomous measure of whether or not a city has at least one team as its dependent variable. In Table 4, the issue of multiple

Table 2.

Number of Cities with a Given number of teams at a Given Level												
Teams	**MLB**		**Any Level**		**AAA**		**AA**		**A-Advanced**		**A**	
0	1862	91.7%	1343	66.2%	1822	89.8%	1825	89.9%	1895	93.3%	1823	89.8%
1	133	6.6%	595	29.3%	208	10.2%	201	9.9%	97	4.8%	206	10.1%
2	35	1.7%	59	2.9%			4	0.2%	17	0.8%	1	0.0%
3			17	0.8%					12	0.6%		
4			9	0.4%					2	0.1%		
5			7	0.3%					7	0.3%		
TOTAL	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%

Table 3.

Number of Cities with at Least One Team at a Given Level												
Teams	**MLB**		**Any MiLB**		**AAA**		**AA**		**A-Advanced**		**A**	
0	1862	91.7%	1343	66.2%	1822	89.8%	1825	89.9%	1895	93.3%	1823	89.8%
At Least 1	168	8.3%	687	33.8%	208	10.2%	205	10.1%	135	6.7%	207	10.2%
TOTAL	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%	2030	100.0%

Table 4.

Frequencies for Highest Level			
Highest Level	**Frequency**	**Percent**	**Cumulative Percent**
No Team	1343	66.2	66.2
A	174	8.6	74.7
A-Advanced	118	5.8	80.5
AA	187	9.2	89.8
AAA	208	10.2	100.0
Total	2030	100.0	

teams is resolved in a different way: by coding only the highest level at which a city has a team. In this case, only the levels from A to AAA are included. This Highest Level data will be our dependent variable for Multinomial and Ordinal Logistic Regression models. The percentage of cities without a team is the same as that for Any MiLB level in Table 3. For the other levels, we see the highest percentage of cities have AAA as their highest level, with A and AA in the same ballpark. The relatively low number for A-Advanced is attributed to the large number of cities with more than one team at that level, just as we saw in Table 2. Another finding is that a higher classification does not always have more individual cities with teams than the classification below it.

Descriptive statistics and frequencies for our independent variables are provided in Tables 5 & 6. Only one of our covariates, the number of major-league professional teams in a city (ProTeams), was discrete. Table 5 provides its frequencies. Here we see that while most cities have no major league teams in any of these sports, most that do have only one or two major league teams in all of these sports combined. We also see that there are fewer cities with four major league teams than three, still fewer with five major

Table 5.

Number of Professional Sports Teams in a City		
Teams	**Number of Cities**	**Percentage**
0	1751	86.3%
1	90	4.4%
2	86	4.2%
3	51	2.5%
4	26	1.3%
5	12	0.6%
6	7	0.3%
8	7	0.3%
TOTAL	2030	100.0%

Note: This refers to the teams in the highest professional level in sports other than baseball. NFL, NBA, NHL, and MLS teams are all included.

league teams, and so on. This pattern continues down until only the seven Los Angeles and New York City entries remain as having six and eight teams, respectively, at the highest level.[4] As the rest of my independent variables are continuous, the descriptive statistics of the variables tested in the model are in Table 6.

As I have mentioned, some variables were transformed to better reflect a normal distribution and account for outliers. A full listing of descriptive statistics, including variables before transformations, can be found in the Appendix as Table A3. Population Change is reported in hundred-thousands, with MLB Distance and Baseball Distance being reported in hundreds of miles. Even with the transformations, we can still see some skew in the data, and those variables which were not transformed have mean values that are often much greater than their medians.[5] The mean value for Log Population corresponds to roughly 375,000 people, and the mean for Log Per Capita Income to

4 Since 2006-2012 was analyzed for each of the 230 cities, there will be seven entries for each city.

5 One might wonder why these variables would not be transformed to more normal distributions as well. While this is a valid question, the simplest answer is that transformations are only possible when the all values are positive for a variable. Although this condition is met for Baseball Distance, the transformation showed no significance at all.

roughly $36,000. Skewness and Kurtosis are two measures of the level of skew in the data, which helps explain why our question is so hard for researchers to answer. In a perfectly normal distribution, Skewness should be close to 0 and Kurtosis close to 3.0, but it is not surprising to see that some of our variables are not normally distributed. However, by using the Quintile values, we are able to see how large the spread is for some variables. For example, although the eightieth percentile for MLB Distance and Baseball Distance have values less than 3.0, the maximum value in both cases is 10.00 (or 1,000 miles). This helps us see the effect of outliers in our data set, usually in Alaska or Hawaii.

Pearson's r was used to test the relationships between pairs of variables, with the results in Table 7. Only in three cases is the bivariate relationship insignificant, and they involve Population Change, Percent Population Change, ProTeams, and Baseball Distance. We might not expect some of these to be related anyway, particularly Baseball Distance and ProTeams ($r=-.031$, $p>.05$) or Baseball Distance and Population Change ($r=.033$, $p>.05$). On the other hand, the lack of a significant relationship between Percent Population Change and ProTeams is a bit surprising ($r=.017$, $p>.05$). One would expect that the percentage of population growth would at least have some effect on the locations of the highest level of professional sports teams, particularly given the high correlation between ProTeams and numerical Population Change ($r=.717$, $p<.001$). For the most part, all other variables show significance even if their relationship is too weak to be meaningful ($r<.20$).There appears to be a strong relationship between Log Population and Population Change ($r=.660$, $p<.05$) and a moderate one between Log Population and Log

Table 6.

Descriptive Statistics

		Log Population	Population Change	Percent Population Change	Log Per Capita Income	MLB Distance	Baseball Distance
N		2030	2030	2030	2030	2030	2030
Mean		12.8389	0.45048	5.0385	10.4899	1.9327	1.2475
Std. Error of Mean		0.0264	0.02286	0.1064	0.0033	0.0331	0.0294
Median		12.5378	0.11918	4.5251	10.4808	1.6400	0.8900
Std. Deviation		1.1901	1.03009	4.7958	0.1483	1.4935	1.3233
Skewness		0.8506	3.86224	0.7644	0.1759	2.386	4.562
Kurtosis		0.3797	17.76095	2.5127	0.8211	9.840	25.287
Range		6.5043	10.84017	52.6515	1.0885	10.00	9.82
Minimum		10.4624	-2.84653	-18.4928	9.9683	0.00	0.18
Maximum		16.9666	7.99364	34.1586	11.0568	10.00	10.00
Percentiles	20	11.7897	0.02228	1.1329	10.3710	0.8940	0.6000
	40	12.3140	0.08161	3.4040	10.4489	1.3700	0.7700
	60	12.9249	0.17341	5.5373	10.5196	2.0100	1.0200
	80	13.8320	0.52068	8.6092	10.6103	2.7400	1.5300

Per Capita Income (r=.419, p<.05). This is good news because, according to previous studies, population should be the major factor in choosing a team location. Also, the relationship between Log Population and numerical Population Change does not appear to result in multicollinearity, so we are able to reject the last stipulation of our seventh hypothesis.

Last, we know that a value greater than 0.8 shows a very strong relationship and likely multicollinearity. Although some relationships are close, nothing immediately strikes us as higher than that threshold.[6] However, intuition tells us that we should not use Percentage Population Change and numerical Population Change in the same model since our hypothesis calls for them to be used as a measure of the same aspect of our model. Reviewing the strength of the association factors shows that Population Change has

Table 7.

Pearson's r Correlations for Independent Variables

	Log Population	Population Change	Percent Population Change	Log Per Capita Income	MLB Distance	Baseball Distance	ProTeams
Log Population	1	.660***	.152***	.419***	-.310***	-.058**	.710***
Population Change	.660***	1	.381***	.296***	-.215***	.033	.717***
Percent Population Change	.152***	.381***	1	-.051*	.197***	.162***	.017
Log Per Capita Income	.419***	.296***	-.051*	1	-.150***	.153***	.431***
MLB Distance	-.310***	-.215***	.197***	-.150***	1	.722***	-.314***
Baseball Distance	-.058**	.033	.162***	.153***	.722***	1	-.019
ProTeams	.710***	.717***	.017	.431***	-.314***	-.019	1

***. Correlation is significant at the 0.001 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

6 Relationships with a value of 0.6 or greater were tested further using the Variance Inflation Factor (VIF). All VIF values were lower than the standard multicollinearity threshold of 5.0, and in fact lower than 4.0.

stronger relationships than Percentage Change in almost every case. Also, remember that we were not sure what to make of the insignificant relationship between Percentage Change and ProTeams. Therefore, it seems more rational to use Population Change in the case that both show significance in our final model.

The reverse situation is also true: MLB Distance and Baseball Distance have the highest Pearson's r value ($r=.722$, $p<.001$), suggesting a higher likelihood of multicollinearity. However, these variables were designed with different dimensions of the model in mind: MLB Distance is hypothesized to reflect distance to a potential affiliate, while Baseball Distance might explain a city's level of spatial monopoly power in the baseball industry. This fact, combined with the difference in strengths or relationships with other variables and the r-value below our threshold, justifies the use of both variables in our final model.

Logistic Regression

Binomial logistic regression was used to create a model which predicts whether a city has at least one team at any MiLB level. At the beginning, each variable was tested stepwise, being kept in the model if it showed significance, and being discarded if there was none ($p>.05$). Our base model predicts that no cities have a minor league team, and gives us only an intercept ($\alpha = -.670$). While this seems justifiably crude, it is important to note that the null model correctly predicts that 66.2% of cities have no team, so this is the baseline we are trying to improve upon. All variables will be added to this model stepwise, changing the intercept based on the line-of-best-fit. In all cases, the intercept was significant based on the Wald chi square test ($p<.05$), and there was one degree of freedom for all variables. At this point, we will first report the Wald significance for each variable, as the values of B will change when the final model is produced.

Log Population showed significance (Wald=438.483, $p<.001$) as we would expect, with a positive coefficient (B=1.277) that shows a direct relationship. ProTeams was then added to the model, again showing significance (Wald=111.724, $p<.001$). The coefficient was negative, however, suggesting that having a major league sports team in a city is associated with a lower likelihood of having a MiLB team for all levels combined (B=-.818). With Log Population and ProTeams included in our model, Log Per Capita Income was insignificant when added (Wald=1.930, $p>.05$), and is therefore discarded for this model. Similarly, adding Population Change was clearly insignificant according to our Wald test (Wald=.001, $p>.05$), with $p=.973$. In its place, Percentage Population Change was tested instead. Although the p-value for this variable was much lower ($p=.084$), it was still insignificant according to our Wald test (Wald=2.986, $p>.05$).

After all of this, we are left with only 2 variables in our model so far: Log Population and ProTeams. Adding MLB Distance and Baseball Distance into the model showed significance at the highest level for both variables ($p<.001$). The final binomial logistic regression model for a city having at least one team at any level is provided in Table 8. The positive coefficient of MLB Distance (B=.285) means that a greater distance from an MLB team corresponds to a higher likelihood of having a minor league team, and we will try to explain this later. Baseball Distance had a negative parameter coefficient (B = -.638), with the inverse relationship meaning that a greater distance from

Table 8.

Binomial Logistic Regression: Any Level						
	B	S.E.	Wald	df	Sig.	Exp(B)
Log Population	1.914***	.092	436.024	1	.000	6.777
ProTeams	-0.767***	.081	90.283	1	.000	.464
MLB Distance	0.285***	.066	18.887	1	.000	1.330
Baseball Distance	-0.638***	.093	46.860	1	.000	.528
Constant	-25.177***	1.197	442.512	1	.000	.000

***. All coefficients are significant at the 0.001 level (2-tailed).

any professional baseball team leads to a lower likelihood of having a professional baseball team.

The opposite effects of MLB Distance and Baseball Distance support the idea that they are representing different dimensions of the model, even though we hypothesized that greater distance would correspond to a higher likelihood because of spatial monopoly power. Pseduo r square values are not typically large, and in our case the Cox & Snell r square was .344 and the Nagelkerke r square was .476. Table 9 provides a classification table breaking down the accuracy of our model for having a team at any level. The cut point is set at 0.5, meaning that a predicted probability of 0.5 or higher will be predicted to have a team at some level. Using this model, we correctly predict the status of 82.4% of the cities analyzed, much better than the base null model at 66.2%.

There are some other interesting findings in the final model. Just as our literature review showed, Log Population has a huge effect on the probability of a city having a minor league team, as its odds ratio is given as 6.777. The coefficient, B, is positive and the variable still shows significance (Wald = 436.024, $p<.001$). The coefficient for ProTeams is still significantly negative (Wald = 90.283, $p<.001$), with a much lower odds ratio of 0.464 in comparison to Log Population. MLB Distance has a positive coefficient

Table 9.

Classification Table: Binomial Logistic Regression					
Cox & Snell	0.344		Predicted		
Nagelkerke	0.476		Any Level		Percentage Correct
N	2030		No Team	At Least 1	
Observed	Any Level	No Team	1180	163	87.9
		At Least 1	195	492	71.6
	Overall Percentage				82.4

Variables included in the model are Log Population, ProTeams, Baseball Distance, and MLB Distance. The cut value is .500.

and its odds ratio is actually the second largest (Exp(B)=1.330), while Baseball Distance has a negative coefficient that fails to support our hypothesis. The constant, or intercept, is significant at the highest level as well (B = -25.177, Wald = 442.512, p<.001). To go beyond the classification table, predicted probabilities for each city in the data set were calculated for use in ranking the cities, which will be used in the Discussion section.

Now that we have the probabilities for having a minor league team at any level, we will begin using an Ordinal Regression model. This method can be used to create a model for all the different levels simultaneously. Our dependent variable here is Highest Level, which reports the highest classification from A to AAA where a given city has a team. Again, the variables were tested stepwise, being included in the model only if they showed significance in at least one category. We begin with the model for Log Population only, represented in Table 10, which should have the strongest effect on our dependent variable. The thresholds, when negated, represent the constants for each of the models, and they rise as expected for each of the levels. All intercepts are significant at the highest level according to their respective Wald tests, and so is Log Population (p<.001). With just this amount of information, we would typically begin analyzing the coefficients

Table 10.

Ordinal Logistic Regression								
							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold: Highest Level	No Team	15.475***	.629	605.665	1	.000	14.242	16.707
	A	16.071***	.638	635.163	1	.000	14.821	17.321
	A-Advanced	16.537***	.645	657.299	1	.000	15.273	17.802
	AA	17.49***	.661	700.754	1	.000	16.195	18.785
Location	Log Population	1.142***	.048	572.314	1	.000	1.049	1.236

Link function: Logit.

Table 11.

Test of Parallel Lines				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	3747.325			
General	3733.735	13.590	3	.004

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

and explaining what exactly they mean. However, it is more important that we feel secure in our choice of modeling technique, so we will put this off for a moment.

An ordinal regression assumes that there is order behind the categorization, and we have already established why we believe this to be true. Next we need to test the proportional odds assumption. Ordinal Logit Modeling gives us only one value for the slope coefficient of Log Population, assuming that the difference between levels lies only in our threshold intercepts. This would then mean that the slope coefficients would show no difference between levels. If that is true, then the Test for Parallel Lines (TPL) should support this null hypothesis. As we can see in Table 11, the significance level of .004 shows that there is significant difference between the slope coefficients (χ^2=13.590, p<.01). While previously a significant p-value led us to include a variable in the model, this time that calls into question one of our assumptions that allows us to use Ordinal

Regression. We reject the null hypothesis of the TPL, that the slope coefficients are the same, and are left thinking that a higher level of baseball is not quite as ordered as we thought in terms of the factors behind team location. Therefore, despite its significance in the Ordinal Logit model, Log Population violates one of the principal assumptions of the model and cannot be used for drawing any meaningful conclusions.[7] For this reason, we do not need to know any more about the coefficients, pseudo r-squares, or inner-workings of the model, as our variables do not meet the assumptions necessary to use it.

Even if it would have initially gone against our intuition, we now know that we must treat the classifications as completely separate and unordered through Multinomial Regression.[8] The proportional odds violation makes sense with a bit more thought, though, as having a team at a higher level does not always rule out the possibility of having another team at a different level. That said, Multinomial Regression requires the categories to be mutually exclusive, which we have taken care of by coding only the highest level at which a city has a team. This time, the separate models will be calculated for the different calculations while allowing for different values of the slope coefficients within each category. The category 0, representing cities which have no team, served as our reference category to which all other category models are compared. Variables were once again tested stepwise based on significance for at least one category.

The stepwise process yielded significance for at least one classification for all variables, so we will not need to discuss as many details as we did in the binomial model.

---

7 Although they are not reported, other variables were tested as well with the same overall TPL results.

8 Again, a GOLM is another possible way to deal with significance during the TPL. Davis himself says in an endnote that a multinomial model was created along with a GOLM in his research because Multinomial is even more flexible (Davis 2006, 264). For this reason, greater flexibility in our longitudinal data set may be exactly what we need to improve on his results.

Before we skip to the final model, though, it is important to note two things. Recall that, because the testing for multicollinearity supported our idea that Baseball Distance and MLB Distance seem to be measuring different dimensions of the model, both were used in the final model. Also, despite both Population Change and Percentage Population Change showing significance, we know that we cannot use both in the model for the reasons stated earlier. Therefore, because Population Change has stronger, more reliable relationships with our other independent variables, it will be chosen over Percentage Population Change for inclusion in the model. This model is reproduced in Table 12. Even though not all variables are significant in every classification, they are kept because they are significant in at least one classification. Because it is a Multinomial model, which is more complex, we should not worry if the model for one classification is using an insignificant variable or two. After all, insignificant variables should have no noticeable effect on our results

In the final model, all of the intercepts were significant, negative and seemingly substantial with one exception: the A-Advanced model's intercept was not only smaller (B = -12.3565), but also insignificant (Wald=1.561, $p>.05$). Log Population was significant at the highest level ($p<.001$) with a positive coefficient for all classifications. This is as expected, supporting the claim that higher population is associated with a higher likelihood of having a minor league team at any level. Although most values for the coefficient (B) of Log Population were between 1.5 and 2.0, the coefficient was much larger at the AAA level (B=3.233). This also corresponded to a jump in the odds ratio of Log Population from 4.5 to 6.2 to a much larger value for AAA (Exp(B)=25.345).

ProTeams was also significant at the highest level (p<.001), but with negative B coefficients hovering around -1.00 for all classifications.

Log Per Capita Income was only significant at the AA level (B=3.032, Wald=17.380, p<.001) and the AAA level (B=3.579, Wald=19.880, p<.001). In both cases the odds ratio for Log Per Capita Income is even higher than that of population, which is somewhat surprising. The positive coefficients suggest a direct relationship in which a higher income is associated with a higher likelihood of housing a team at the AA and AAA levels. Population Change was only significant for the A-Advanced level (B=.555, Wald=17.402, p<.001), with a positive coefficient and an unsurprising direct effect on the dependent variable. Baseball Distance was significant at the highest level (p<.001) for all classifications with a range of coefficients (-3.3 < B < -0.5). The coefficient was consistently negative, meaning that the farther away a city is from any professional baseball team, the less likely it is to house a minor league team. On the other hand, MLB Distance was negative but insignificant at the A level (p>.05), but had a consistent positive coefficient for the other classifications which showed significance. Its strongest effect was at the AAA level (Exp(B)=2.599). For these variables, there may have been times when the sign of the coefficient was different for one level than another, but in those cases the effect was insignificant. The pseudo r squares for the model were higher than the binomial model as well (Cox & Snell=.450, Nagelkerke=.506).

Finally, for the purpose of comparison with a null model, we include a classification table as Table 13. In short, the predicted classifications for each city are based on the classification with the highest predicted probability of the five different

Table 12.

Multinomial Logistic Regression Parameters

Highest Level		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
A	Intercept	-24.187**	7.546	10.275	1	.001			
	Log Population	1.509***	.139	118.600	1	.000	4.522	3.447	5.933
	Log Per Capita Income	0.411	.733	.315	1	.575	1.509	.358	6.349
	Population Change	-0.301	.201	2.252	1	.133	.740	.499	1.097
	Baseball Distance	-0.936***	.216	18.692	1	.000	.392	.257	.600
	MLB Distance	-0.13	.112	1.362	1	.243	.878	.705	1.093
	ProTeams	-1.278***	.201	40.224	1	.000	.279	.188	.414
A-Advanced	Intercept	-12.365	9.897	1.561	1	.212			
	Log Population	1.817***	.179	102.586	1	.000	6.153	4.329	8.746
	Log Per Capita Income	-1.095	.956	1.314	1	.252	.334	.051	2.176
	Population Change	0.555***	.133	17.402	1	.000	1.741	1.342	2.260
	Baseball Distance	-3.334***	.379	77.293	1	.000	.036	.017	.075
	MLB Distance	0.277*	.134	4.281	1	.039	1.319	1.015	1.715
	ProTeams	-0.799***	.156	26.295	1	.000	.450	.331	.610
AA	Intercept	-56.943***	7.734	54.215	1	.000			
	Log Population	1.794***	.141	163.072	1	.000	6.016	4.568	7.924
	Log Per Capita Income	3.032***	.727	17.380	1	.000	20.728	4.984	86.204
	Population Change	0.106	.104	1.027	1	.311	1.111	.906	1.363
	Baseball Distance	-0.591***	.123	23.147	1	.000	.554	.435	.705
	MLB Distance	0.378***	.099	14.682	1	.000	1.460	1.203	1.772
	ProTeams	-0.887***	.126	49.656	1	.000	.412	.322	.527
AAA	Intercept	-82.93***	8.883	87.157	1	.000			
	Log Population	3.233***	.206	247.228	1	.000	25.345	16.939	37.921
	Log Per Capita Income	3.579***	.803	19.880	1	.000	35.828	7.430	172.759
	Population Change	0.031	.097	.103	1	.749	1.031	.853	1.247
	Baseball Distance	-0.985***	.128	58.996	1	.000	.373	.290	.480
	MLB Distance	0.955***	.114	69.819	1	.000	2.599	2.077	3.251
	ProTeams	-1.255***	.127	97.240	1	.000	.285	.222	.366

The reference category is No Team.

*** Coefficient is significant at the 0.001 level (2-tailed), ** Coefficient is significant at the 0.01 level (2-tailed), * Coefficient is significant at the 0.05 level (2-tailed).

Table 13.

Classification Table: Multinomial Logistic Regression

Observed	Predicted					
	No Team	A	A-Advanced	AA	AAA	Percent Correct
No Team	1272	0	9	5	57	94.7%
A	143	7	4	0	20	4.0%
A-Advanced	61	0	25	5	27	21.2%
AA	108	5	0	0	74	0.0%
AAA	75	0	17	0	116	55.8%
Overall %	81.7%	.6%	2.7%	.5%	14.5%	70.0%

Note: Cells give the number of cities predicted to have a team at the given level.

options. Note that the highest probability may be, and in the vast majority of cases is, that the city has no team at any level. The classification with the next most cases is AAA, while the other three levels are few and far between. In fact, by the classification table, no city is predicted to have a team at the AA level. We see that we are correctly predicting only 70.0% of team locations with this method, most of those coming from cities with no team. Recall that if we simply predicted that none of the 2030 city entries had a minor league team, we would accurately predict the classification of 66.2% of the cities analyzed.

At first, this seems to really cast doubt on the predictive power of our model, especially when we seem to only be predicting either AAA or No Team. Some of this is for good reason: modeling MiLB team locations is no doubt a peculiar problem. However, jumping to the conclusion that our work is unhelpful at this point would be invalid for a couple reasons. First, this simplistic table may predict that 14.5% of the 2030 US cities analyzed have a AAA team (that is, 42 per year), but of course we know that only 30 cities in the US have a AAA team. Thus, we expect those other high probabilities to trickle down to the other levels. More importantly, however, our analysis is focused on ranking the cities based on the predictive probabilities at each level independently. Therefore, cities that have unusually high probabilities at more than one level will be more accurately represented and give a better picture of opportunities in minor league team location.

DISCUSSION & CONCLUSION

Our first hypothesis was that a higher population would correspond to a higher likelihood of housing a minor league team at every classification. Since Log Population had a significant positive coefficient for every classification, our hypothesis is supported. Although it is not always the highest effect, the odds ratios show that it has a very prevalent effect across levels. On the other hand, Log Population was the only variable which fully supported its hypothesis. Log Per Capita Income did show a positive relationship at those levels where it was significant, and in those cases, the odds ratio shows that it had a greater effect than Log Population. The lack of significance at the lowest two levels requires some explanation, and it may have to do with the minor league hierarchy more than anything else. Amid the confusing results and different effects of per capita income in previous studies (Paul 2008), perhaps those cities with the highest income levels are being chosen first for the highest classifications, leaving fewer obvious options for lower-level teams looking for an affluent city. This would help explain the large effect of Log Per Capita Income at the higher levels as well.

Population Change is even harder to make sense of because not only are its coefficients significant for only some classifications: the differences are not easily explained by hierarchy either. We see a positive relationship that supports our hypothesis for the AA and A levels, but the lack of significance for AAA and A-Advanced means that it is only partially supported. One possible explanation is structural: that team locations at the A-Advanced level are historically based on geography more than any other factor. By looking at the map of the 2012 A-Advanced landscape, this is easily seen. The California and Florida State leagues are composed entirely of teams within the respective states. Not only that, but several teams are often located together in some of the most populated CSAs in that state, such as Los Angeles and Miami. For this reason, it seems logical to suppose that geography may be a third variable limiting the mobility of teams at this level.[9] Therefore, the dynamic nature of Population Change makes it less useful for predicting the locations of an A-Advanced team.

AAA is hard to explain as well, but ultimately we are drawn back to the hierarchy and the massive effects of Population and Per Capita Income. Since this is just one level lower than the MLB, we would expect it to be the level which houses teams in the highest population areas left over. Combine this with the higher income levels in those cities, according to the odds ratio, and we have the cities with the largest, most affluent market areas. So although these cities may not have the highest growth, they are likely the best locations for the time being. Of course, if a city keeps growing, its population numbers will show that fact, and it would be much easier to relocate at that point.

---

9 One could also argue the opposite: that the importance of geography should make population growth a more important factor since there are fewer high-population areas to choose from. This might make sense in the long-run, but conversely, the growth rate of a city may also change considerably in the long run. In other words, the high costs of stadium construction and relocation outweigh the benefits of moving to a growing city.

The negative coefficient of ProTeams presents us with our first complete rejection of a hypothesis. It was significantly negative at every level, which goes against our intuition. ProTeams was designed as a measure of sports popularity in a given city, which should be positively related to the presence of a team. However, this seems to be showing a completely different phenomenon. Perhaps the most compelling reason for a negative relationship is that ProTeams is only measuring the highest level of professional sports teams. Therefore, it is not surprising that this overlaps with the MLB cities. Although some levels, notably A-Advanced, have teams in the same city as an MLB team, that is not true in most cases. In fact, only one MLB city (Atlanta, GA) also houses a AAA team. So although it is not a perfect answer, this does help explain why the relationship would be negative. In order to better reflect the aspect we were attempting to measure, it might be better to look at the locations of minor league teams in other sports. That said, baseball's hierarchy is far more elaborate than most other sports, making it that much harder to know which leagues to choose for inclusion in such an index. In any case, the relationship does not appear to be large according to the odds ratios despite its statistical significance.

For Baseball Distance, we hypothesized that spatial monopoly power would cause us to see a direct relationship between it and Highest Level. The variable is significant for all classifications, but the relationship is in fact inversely proportional. That is, the farther a city is from the closest professional baseball team, the less likely that city is to have a minor league team regardless of the classification. As a consumer of minor league baseball, I still believe that spatial monopoly power is very much an important factor. In reviewing the data, I concluded that the importance of spatial monopoly power is overshadowed by a lack of professional baseball teams in some isolated areas. For example, cities in the Rocky Mountain region were often closest to either the Colorado Rockies or the Seattle Mariners despite being hundreds of miles away. To the extent that this is prevalent, the lack of professional baseball in that region would then be associated with a great distance from the closest team at any level. So although we capped distance at 1,000 miles for two main outlier states, Alaska and Hawaii, perhaps distance would need to be limited even further to be able to see the effect of a spatial monopoly. Based on these findings, we reject the hypothesis that a greater distance from a professional baseball team, as we defined it, corresponds to a higher likelihood of housing a team.

Finally, MLB Distance was hypothesized to be significant for all levels, but have different effects depending on the level being considered. We predicted that a greater distance from an MLB team would lead to a higher probability of having a team at the lower three levels, but a lower probability of having a AAA team. Despite insignificance at the A level, the positive coefficients in the A-Advanced and AA models support our hypothesis for those levels. However, the AAA coefficient is positive as well, suggesting the opposite of our hypothesis for that level. Note that this is the opposite of the effect of Baseball Distance, where greater distance from a professional baseball team corresponded to a lower likelihood of housing a minor league team in that city. This supports our original thought that these variables were measuring different aspects of the question.

The original theory was that AAA teams would want to locate closer to an MLB team, if it is their affiliate, to allow for quicker transfer of players at the highest levels due to injuries and the like. Once again, the data seem to be measuring a different aspect than we intended, albeit an important one: AAA games may be seen as substitutes for MLB games. In fact, all levels other than A appear to be substitutes for the MLB to some degree. This is especially true in areas like the Southeast, where MLB games are only played in Georgia and Florida. Perhaps consumers are not as concerned about the level of baseball as we had originally supposed.

Granted, this oversimplifies the clear differences in quality of play, but the basic point is that consumers of baseball appear to be, to some degree, consumers of all levels of baseball. To the extent that minor league teams are competitors with MLB firms, consumers would be likely to choose the option closest to them geographically. Furthermore, because there is a higher quality of play in the MLB, it makes even more sense that a minor league firm would need to locate farther away in order to compete effectively. Therefore, we reject the hypothesis that the greater a city's distance from an MLB team, the higher its likelihood of housing a AAA team. Here we conclude that although Baseball Distance and MLB Distance do seem to be measuring different aspects of team location, they do not appear to be measuring the aspects we intended for them. MLB Distance is effectively describing the distance to a competitor rather than a potential affiliate, while Baseball Distance is a measure of complete geographic isolation rather than spatial monopoly power. Even though these outcomes are different from our hypotheses, the models results are not invalidated. Instead, we use them to learn how to evaluate these variables based on what they actually appear to be measuring.

Earlier we learned that Classification Table 13 shows the predicted classifications of the 2030 cities analyzed for the seven years combined. It was clear to see that the model was predicting more cases correct at some levels than others, most notably the fact that not a single city was correctly predicted to have a team at the AA level. However, this analysis relies on choosing the classification with the highest probability of the five choices, including No Team. Because a limited number of teams play at each level, this is not the best way to analyze our results. Simply put, the city ranked thirty-first at AAA would still not have a team in our choice of optimal locations because the level is limited to thirty teams. However, it is quite possible that a team that just misses one classification would be a good candidate for another, lower level.

For this reason, all cities were ranked according to their predicted probabilities at each level. Since there were seven entries for each city due to analyzing seven years of data together, the highest probability for each city was used in creating the rankings. Using the actual team locations from 2012, cities were also coded as to whether or not they housed a team at the given level during the final year of our analysis. This allowed us to produce “Top 10” and “Bottom 10” lists just as Davis did. Table 14 contains the cities predicted most likely to have a team that do not for each level, while Table 15 ranks the cities least likely to have a team at a given level that had a team in 2012. No city was excluded from being ranked at any level, so there are times when the same city is shown as an optimal choice for more than one level. This must be interpreted with caution, as it is typically unusual for a city to have more than one (or certainly two) minor league teams.

In comparing our list to Davis' Tables A1 and A2 in the Appendix, we see both similarities and differences.[10] One thing that we notice immediately is that many major league cities are listed as high probability areas for minor league teams as well. Once again, we must interpret this with caution. On one hand, there are clear instances where a city can house both a major and minor league team, such as Los Angeles. However, we also know that these cases are not common, so our model seems to be a bit overconfident in predicting minor league locations in cities with MLB teams. For ease in picking out these cities, the Top 10 lists denote which cities already house MLB teams. Another potential issue is showing that a city does not have a team at one level when it actually has a team at a higher level. The table also makes these inconsistencies easy to pick out.

In Table 14, four of our top ten AAA cities were also predicted by Davis, albeit in a different order. The similarities seem to increase when we see that our top ten list is composed of six non-MLB cities. Personally, I am a bit surprised to see that Greensboro, NC is ranked second overall for AAA team locations. As someone who has spent some time in the area, I know very well that it is not the largest, richest, or most well-known city even within its state. That said, a review of the data shows that despite these facts, Greensboro consistently has "enough" in every category to make it an attractive location according to the model: a population that is higher than 1,000,000; a middle-of-the pack per capita income level, a couple hundred miles of separation from the nearest MLB team, and 100 miles of proximity to several other AAA professional teams to compete

10 Remember that some of the cities listed as likely locations for Davis (ie. Allentown, PA) have since been added as part of a larger CSA. This issue means that it would be inaccurate to immediately assume that certain cities have "dropped out" of the Top 10, as they may have simply been reclassified by the government.

Table 14.

Cities Most Likely to Have a Team That Do Not	
Rank	**AAA**
1	Chicago-Naperville, IL-IN-WI (1)[m]
2	Greensboro--Winston-Salem--High Point, NC (4)
3	Portland-Vancouver-Salem, OR-WA (6)
4	San Jose-San Francisco-Oakland, CA (7)[m]
5	San Antonio-New Braunfels, TX (10)
6	Tulsa-Muskogee-Bartlesville, OK (11)
7	Little Rock-North Little Rock, AR (15)
8	Hartford-West Hartford, CT (19)
9	San Diego-Carlsbad, CA (20)[m]
10	Philadelphia-Reading-Camden, PA-NJ-DE-MD (21)[m]
Rank	**AA**
1	Houston-The Woodlands, TX (1)[m]
2	Seattle-Tacoma, WA (2)*[m]
3	Austin-Round Rock, TX (3)*
4	Cape Coral-Fort Myers-Naples, FL (5)
5	Omaha-Council Bluffs-Fremont, NE-IA (6)*
6	Miami-Fort Lauderdale-Port St. Lucie, FL (7)[m]
7	St. Louis-St. Charles-Farmington, MO-IL (8)[m]
8	Des Moines-Ames-West Des Moines, IA (9)*
9	Albany-Schenectady, NY (10)
10	North Port-Sarasota, FL (13)
Rank	**A-Advanced**
1	Atlanta--Athens-Clarke County--Sandy Springs, GA (1)*[m]
2	Phoenix-Mesa-Scottsdale, AZ (5)[m]
3	Charlotte-Concord, NC-SC (6)*
4	Fresno-Madera, CA (7)*
5	Sacramento-Roseville, CA (9)*
6	New York-Newark, NY-NJ-CT-PA (11)*[m]
7	Rockford-Freeport-Rochelle, IL (15)
8	Cincinnati-Wilmington-Maysville, OH-KY-IN (19)[m]
9	Nashville-Davidson--Murfreesboro, TN (20)*
10	Columbus-Marion-Zanesville, OH (22)*

*. City has a team at a higher level

[m]. City has an MLB team

Table 14.

Cities Most Likely to Have a Team that Do Not (cont.)	
Rank	**A**
1	Youngstown-Warren, OH-PA (2)
2	Toledo-Port Clinton, OH (3)*
3	Lakeland-Winter Haven, FL (4)*
4	Springfield-Greenfield Town, MA (5)
5	Modesto-Merced, CA (7)*
6	Chattanooga-Cleveland-Dalton, TN-GA-AL (8)*
7	Lancaster, PA (10)
8	Bakersfield, CA (13)*
9	Rockford-Freeport-Rochelle, IL (14)
10	North Port-Sarasota, FL (15)*
Rank	**Any Level**
1	San Diego-Carlsbad, CA (5)[m]
2	St. Louis-St. Charles-Farmington, MO-IL (6)[m]
3	Portland-Vancouver-Salem, OR-WA (10)
4	Houston-The Woodlands, TX (13)[m]
5	Cincinnati-Wilmington-Maysville, OH-KY-IN (19)[m]
6	Detroit-Warren-Ann Arbor, MI (20)[m]
7	Milwaukee-Racine-Waukesha, WI (23)[m]
8	Albany-Schenectady, NY (32)
9	Phoenix-Mesa-Scottsdale, AZ (38)[m]
10	Pittsburgh-New Castle-Weirton, PA-OH-WV (39)[m]

*. City has a team at a higher level
[m]. City has an MLB team

against on the field.

For AA, only Albany, NY is predicted in the top ten for both models, and our lists for A-Advanced are completely different. The Class A rankings have nothing for comparison, while the rankings for any MiLB level are virtually unhelpful because they predict so many MLB cities. One aspect that I think is a strength of our model is the lack of Honolulu as an optimal city. Although Davis predicts it as a highly ranked potential city, our model accounts for its distance to the closest team and effectively penalizes it for

Table 15.

Cities Least Likely to Have a Team that Do	
Rank	**AAA**
1	Toledo-Port Clinton, OH (114)
2	Scranton--Wilkes-Barre--Hazleton, PA (99)
3	Colorado Springs, CO (96)
4	Buffalo-Cheektowaga, NY (91)
5	Reno-Carson City-Fernley, NV (75)
6	Tucson-Nogales, AZ (71)
7	Fresno-Madera, CA (64)
8	Indianapolis-Carmel-Muncie, IN (54)
9	Syracuse-Auburn, NY (50)
10	Columbus-Marion-Zanesville, OH (49)
Rank	**AA**
1	Altoona, PA (257)
2	Jackson, TN (209)
3	Binghamton, NY (155)
4	Erie-Meadville, PA (154)
5	Springfield-Branson, MO (118)
6	Fayetteville-Springdale-Rogers, AR-MO (94)
7	Chattanooga-Cleveland-Dalton, TN-GA-AL (84)
8	New York-Newark, NY-NJ-CT-PA (81)[m]
9	Midland-Odessa, TX (73)
10	Corpus Christi-Kingsville-Alice, TX (69)
Rank	**A-Advanced**
1	Montgomery, AL (125)
2	Lynchburg, VA (101)
3	Roanoke, VA (86)
4	North Port-Sarasota, FL (76)
5	Myrtle Beach-Conway, SC-NC (72)
6	Miami-Fort Lauderdale-Port St. Lucie, FL (67)[m]
7	Palm Bay-Melbourne-Titusville, FL (45)
8	Raleigh-Durham-Chapel Hill, NC (25)
9	San Jose-San Francisco-Oakland, CA (21)[m]
10	Bakersfield, CA (18)

[m]. City has an MLB team

Table 15.

Cities Least Likely to Have a Team that Do (cont.)	
Rank	**A**
1	Washington-Baltimore-Arlington, DC-MD-VA-WV-PA (277)[m]
2	Burlington, IA-IL (265)
3	Charlotte-Concord, NC-SC (256)
4	Clinton, IA (255)
5	Denver-Aurora, CO (253)[m]
6	Chicago-Naperville, IL-IN-WI (233)[m]
7	Rome-Summerville, GA (186)
8	Greensboro--Winston-Salem--High Point, NC (163)
9	Bowling Green-Glasgow, KY (128)
10	Charleston-North Charleston, SC (92)
Rank	**Any Level**
1	Burlington, IA-IL (288)
2	Clinton, IA (287)
3	Jackson, TN (242)
4	Rome-Summerville, GA (241)
5	Altoona, PA (235)
6	Bowling Green-Glasgow, KY (185)
7	Binghamton, NY (156)
8	Lynchburg, VA (154)
9	Midland-Odessa, TX (143)
10	Roanoke, VA (139)

[m]. City has an MLB team

being so isolated.[11] Granted, it is still ranked in the top 30 cities for a AAA team, but not nearly as high as in the Davis model. Although it seems clear that our model does not always orient itself toward the real-world, in this case it seems to do that very nicely.

Table 15 seems to be more similar to the original work done by Davis, although still quite different. While six of Davis' bottom ten are in ours as well, there seem to be some notable differences in ranking. For instance, Toledo, OH ranks as a bottom ten city for both models. Although Davis ranks them 38th, our model ranks them at 114th.

[11] When a model was run without capping the distances from MLB teams and the closest professional baseball team, Honolulu was ranked very highly just as it was for Davis.

Similarly, six cities are found to be unlikely to have teams at the AA level in both models, but more major differences in city ranking is noticed. The addition of three variables appears to have a significantly different effect on predicted probabilities. In the coming years we will be able to tell whether or not this has a positive effect on our power to predict where MiLB teams will locate in the future.

Again, the full rankings, Top 10 lists, and predicted probabilities of every city and every level analyzed can be obtained by email. With the predicted probabilities in plain sight, we learn that there is usually very little difference in predicted probabilities among the top five or ten. In the case of the AA level, even as this is expanded to the top fifty, the difference in these rankings is not always very large. However, in comparing a top twenty city to one with a triple digit ranking, the probability may be many times greater for the higher ranked city.

Finally, since this is a topic in spatial economics, maps are provided at the beginning of the appendix that compare the actual MiLB landscape of 2012 to the optimal cities chosen by the model. Although it is not a perfect assumption, the locations of the top thirty cities at each level were plotted on the map as distinct, optimal locations. In the case that a city was predicted to have a team at more than one level, only the highest level is shown in the map of all predicted cities combined. While all levels have geographic limitations, those of the A-Advanced level are very noticeable, showing why our optimal A-Advanced landscape would be impossible without a complete overhaul of the league structure. That said, the model does a better than expected job of placing many teams in California and Florida.

We also see that Class A teams are generally limited to geographic locations east of the Mississippi River. While the model does predict many cities in the Midwest would have a Class A team, there are few teams listed that would fit into the actual South Atlantic League. In addition, because the model does not account for transportation costs between competitors, there are a few teams in California or otherwise isolated areas. For AA, the model predicts virtually a complete overhaul of team location from the Southeast to the Midwest and Texas. Although there are fewer teams in the Northeast according to the model's output, there are still a good number of teams in that region. On the other hand, the model seems to predict fewer AAA teams in the Northeast than are there currently, but more in the Southeast instead. By looking at the actual locations of AAA teams, there is no question that this level is the most geographically disbursed. Also, by looking at previous maps created by the model, we notice that places like Tucson and Colorado Springs were predicted to house much lower-level teams than AAA.

The limitations of this study are many to say the least. To begin, because of the complexity of predicting minor league baseball team locations, we are already at a disadvantage. The fact that the two distance variables did not measure the intended aspects of team location gives us pause, and some might call into question the validity of the model for that very reason. However, I contend that since we were able to make sense of the dimensions these statistically significant variables might be measuring, this could be an area of future research. Maybe by capping the distances at an even lower number, or creating a dummy variable representing a distance less than a given number of miles to the nearest team, one could get a better idea of how spatial monopoly power plays a role.

Other major limitations to this research are geographic, as our model does not account for the current spatial setup of leagues at the different levels. This means that everything from transportation costs to overall eligibility for relocation at the beginning stages of that process are not considered to the degree that they would need to be. Additionally, the rule requiring teams to locate at least eighteen miles apart from each other was not taken into account either. Although the inclusion of MLB cities is a step in the direction of more practical application, future research could be dedicated to making the model even more applicable to the real-world.

In that same vein, there are other real-world application questions that are not related to geography. Since we learned that stadium availability can often be the deciding factor in a relocation decision, any measure that accounts for whether a suitable stadium already exists in different cities would be helpful. Differences in stadium quality could also be an important aspect to consider if it could be quantified in some way. Then there is the most practical of dependent variables: attendance at ballgames. As this is the factor that drives revenues for a minor league team, it would be interesting to see whether or not our model's predicted cities were also those with higher attendance and profitability.

Another related question would be to analyze the trend surrounding changes in affiliation as opposed to relocation. Whereas there are usually one or two franchise moves in a given year, affiliations change at an even higher rate. What are the motives behind these decisions? Is it motivated by profit potential, geographic proximity to the major league affiliate, or something completely different? A crude look at the distances from a minor league team to its affiliate shows that the average distance has decreased from 742 to 701 miles from 2006-2012, while the median has consistently hovered around 490 miles. Since this question may be even more complex than franchise location, it appears to be an obvious area for potential research.

Overall, the main question that we are left wondering is whether or not this model will be more effective in predicting franchise moves in the coming years than Davis was in 2006. Although I hope that this work will prove to have strong predictive power, the very fact that a professor with many years of experience was unable to perfectly answer this question prevents me from feeling overly-confident. In either case, I believe that academics are just beginning to skim the surface of topics like these in sports business. The game of baseball is itself based on a series of mind games and adjustments made after every pitch. Sports economists and statisticians should take the same approach to answering complicated questions in our field. Over time, our predictive ability will increase as previous assumptions continue to be challenged and ultimately discarded as models become more realistic. My greatest hope in writing this is that we will stand in the batter’s box together, working the count until we are able to knock one out of the park.

REFERENCES

Bradbury, J.C. *The Baseball Economist: The Real Game Exposed*. New York: Dutton, 2007.

Branvold, S. E., David W. Pan, and T. E. Gabert. "Effects of Winning Percentage and Market Size on Attendance in Minor League Baseball." *Sport Marketing Quarterly* 6, no. 4 (1997): 35-42.

Bruggink, Thomas H., and Justin M. Zamparelli. "From *Emerging Markets in Baseball: an Econometric Model for Predicting the Expansion Teams' New Cities,*" in *Sports Economics: Current Research,* edited by John Fizel, E. Gustafson, and L. Hadley, 49-59. Westport, CT: Praeger, 1999. http://ebooks.abc-clio.com/.

Davis, Michael C. "Called up to the Big Leagues: An Examination of the Factors Affecting the Location of Minor League Baseball Teams." *International Journal of Sport Finance* 1, no. 4 (2006): 253-264.

Davis, Michael C. "Income and the Locations of AAA Minor League Baseball Teams." *Atlantic Economic Journal* 35, no. 3 (2007): 10.1007/s11293-007-9079-5.

Davis, Michael C. "From *New Franchise Location in Major and Minor-League Baseball*." In Vol. 3, *The Business of Sports,* edited by Brad R. Humphreys and Dennis R. Howard, 157-166. Westport, CT: Praeger, 2008.

Gitter, Seth R., and Thomas A. Rhoads. "Determinants of Minor League Baseball Attendance." *Journal of Sports Economics* 11 no. 6, (2010): 10.1177/1527002509359758.

Krugman, Paul. *Geography and Trade.* Cambridge, MA: MIT Press, 1991.

McCann, Philip. *Urban and Regional Economics.* New York: Oxford University Press, 2001.

Ozanian, Michael K. “Minor Leagues, Major Profits.” *Forbes*, Aug. 6, 2008. http://www.forbes.com/2008/08/06/baseball-minors-sacramento-biz-sports-cz_mo_0806minors.html.

Paul, Rodney J., Andrew P. Weinbach, and Michael Toma. "The Minor League Experience: What Drives Attendance at South Atlantic League Baseball Games?" *Coastal Business Journal* 8, no. 1 (2008): 70-84.

Paul, Rodney J., Kristin K. Paul, and Elizabeth Holihan. "Attendance in the Northwest Baseball League: Effects of Promotions, Winning, and Scoring." *Business Research Yearbook: Global Business Perspectives* 15, no. 1 (2008): 807-812.

Rascher, Daniel A. “From *Franchise Relocations, Expansions, and Mergers in Professional Sports Leagues*.” In Vol. 2, *The Business of Sports,* edited by Brad R. Humphreys and Dennis R. Howard, 67-106. Westport, CT: Praeger, 2008.

Rascher, Daniel A., and Heather Rascher. "NBA Expansion and Relocation: a Viability Study of Various Cities." *Journal of Sport Management* 18 no. 3, (2004): 274-295.

Scully, Gerald W. *The Market Structure of Sports.* Chicago: The University of Chicago Press, 1995.

Siegfried, John J. Review of *Baseball Economics and Public Policy,* by Jesse W. Markham and Paul W. Teplitz. *The Antitrust Bulletin* 28, (1981): 783-790.

Siegfried, John J., and Jeff D. Eisenberg. "The Demand for Minor League Baseball." *Atlantic Economic Journal* 8, no. 2 (1980): 59-69.

Smith, Chris. “Minor League Baseball's Most Valuable Teams.” *Forbes*, July 17, 2013. http://www.forbes.com/sites/chrissmith/2013/07/17/minor-league-baseballs-most-valuable-teams/.

Van Den Berg, David. “Ballpark Boom: New Minor League Stadiums Spring Up Across the Region.” *Region Focus,* Winter 2009, 24-27, 46. http://www.richmondfed.org/publications/research/.

APPENDICES

The Appendix to this work contains several different elements. It begins with visual representations of the 2012 Minor League Baseball landscape, organized by level. Interspersed between these are the maps of the Top 30 city locations predicted by the model for each level in order to provide easy comparison between our predictions and the current reality. A map of the major league locations is provided as well. Followed by these maps are some tables referenced in the work, including Davis' model-predicted locations and an extended table of descriptive statistics.

Figure 1. 2012 Locations of MiLB Teams

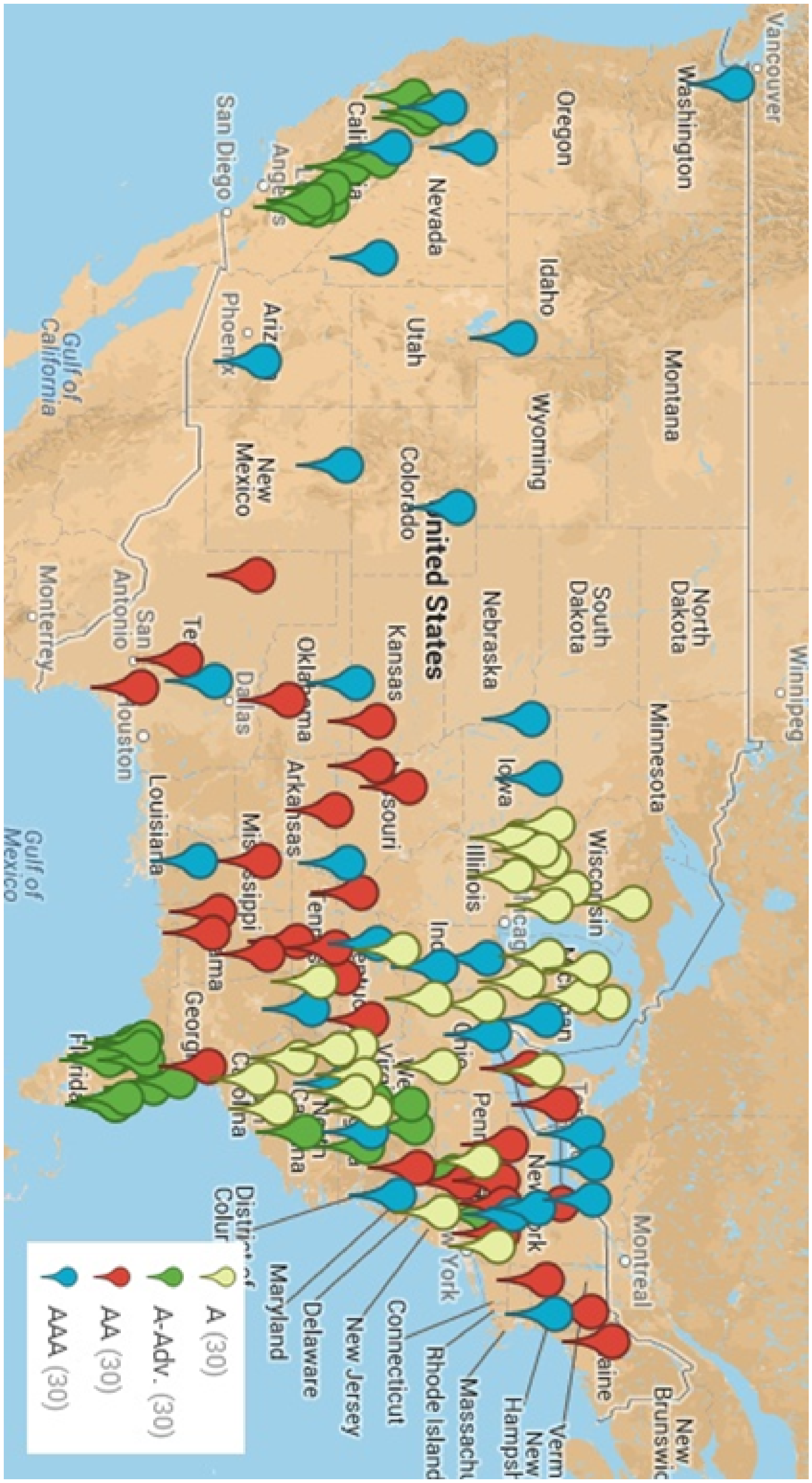

Figure 2. Predicted Team Locations – All Levels

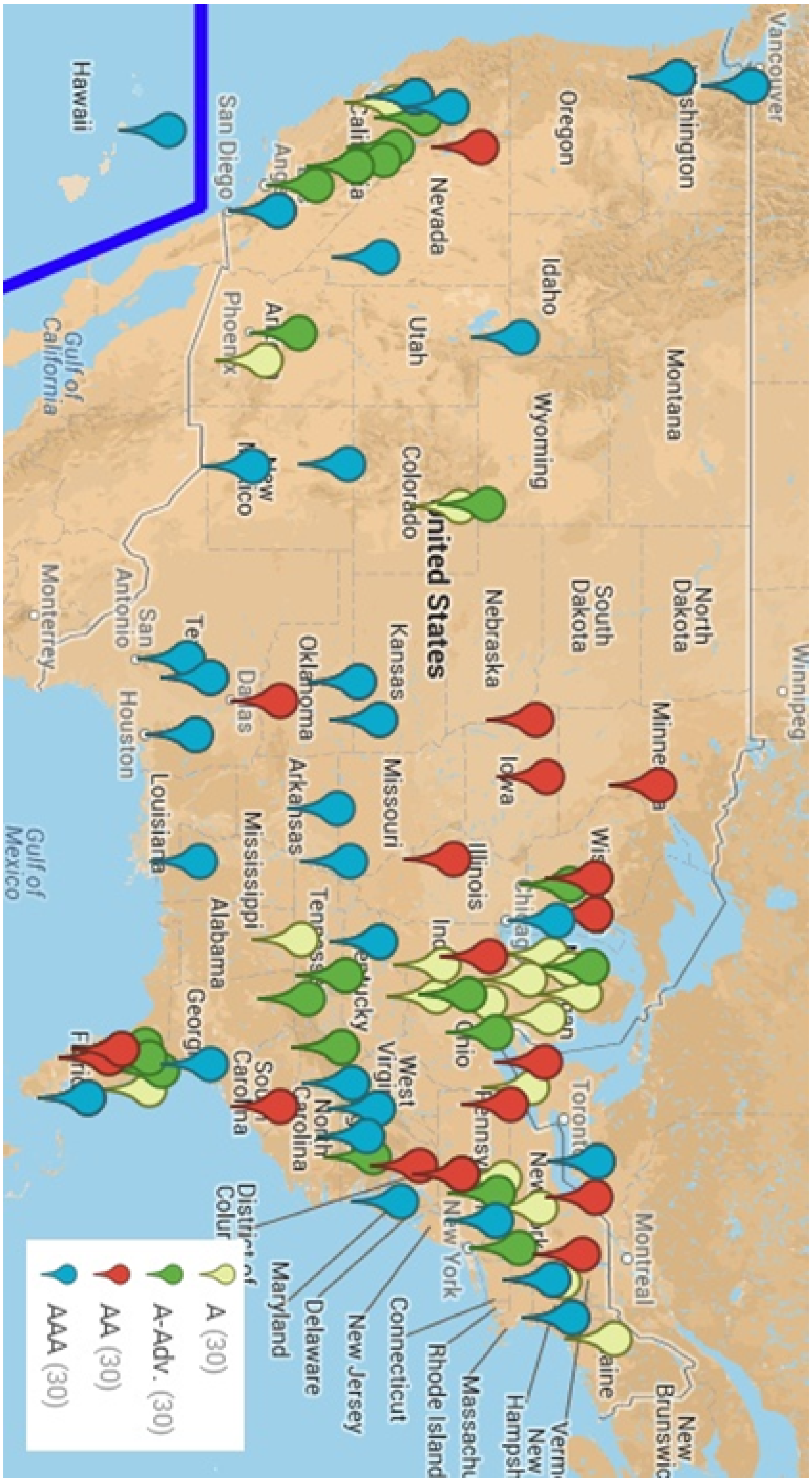

Figure 3. 2012 Class A Team Locations

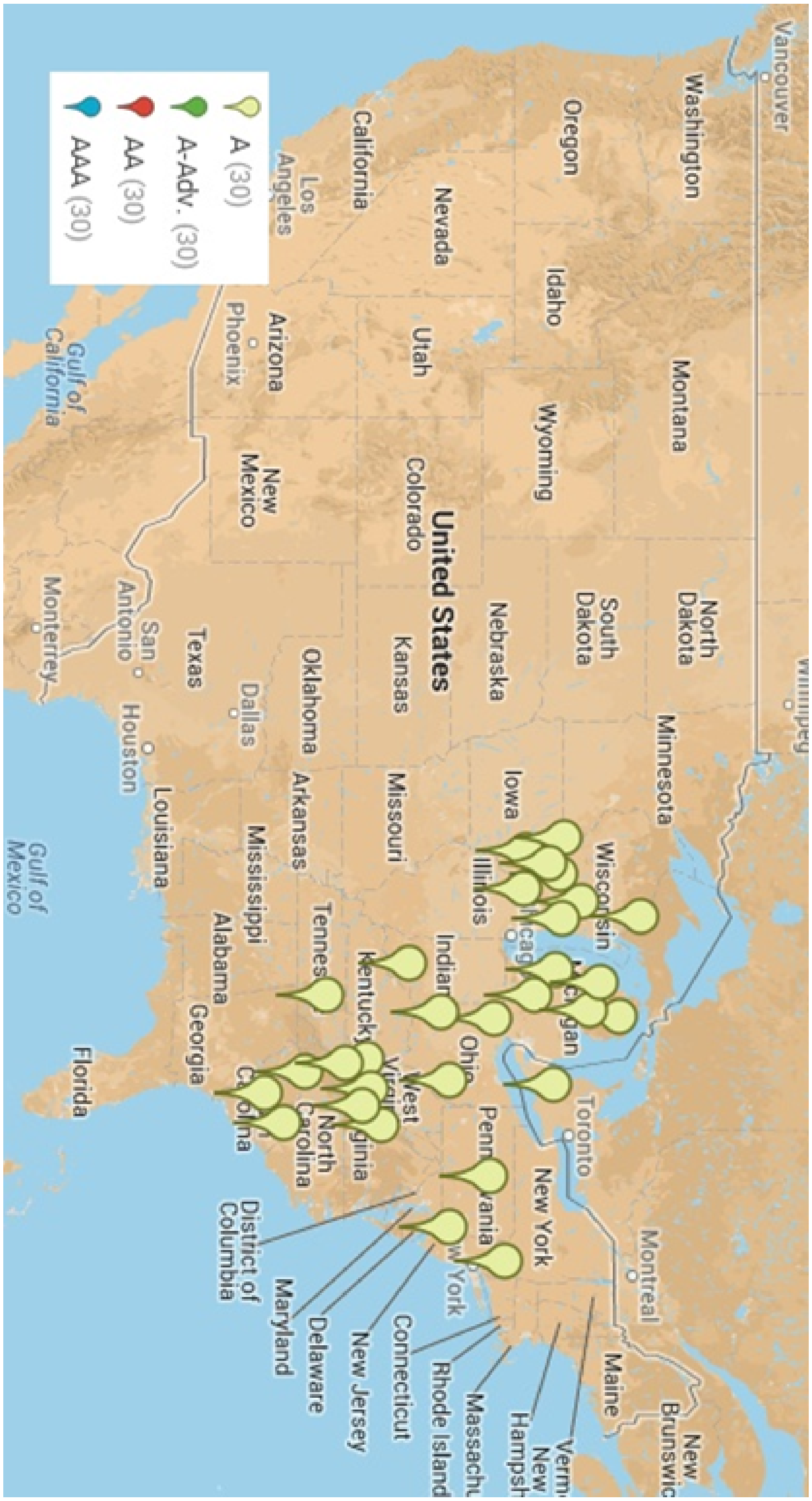

Figure 4. Predicted Class A Team Locations

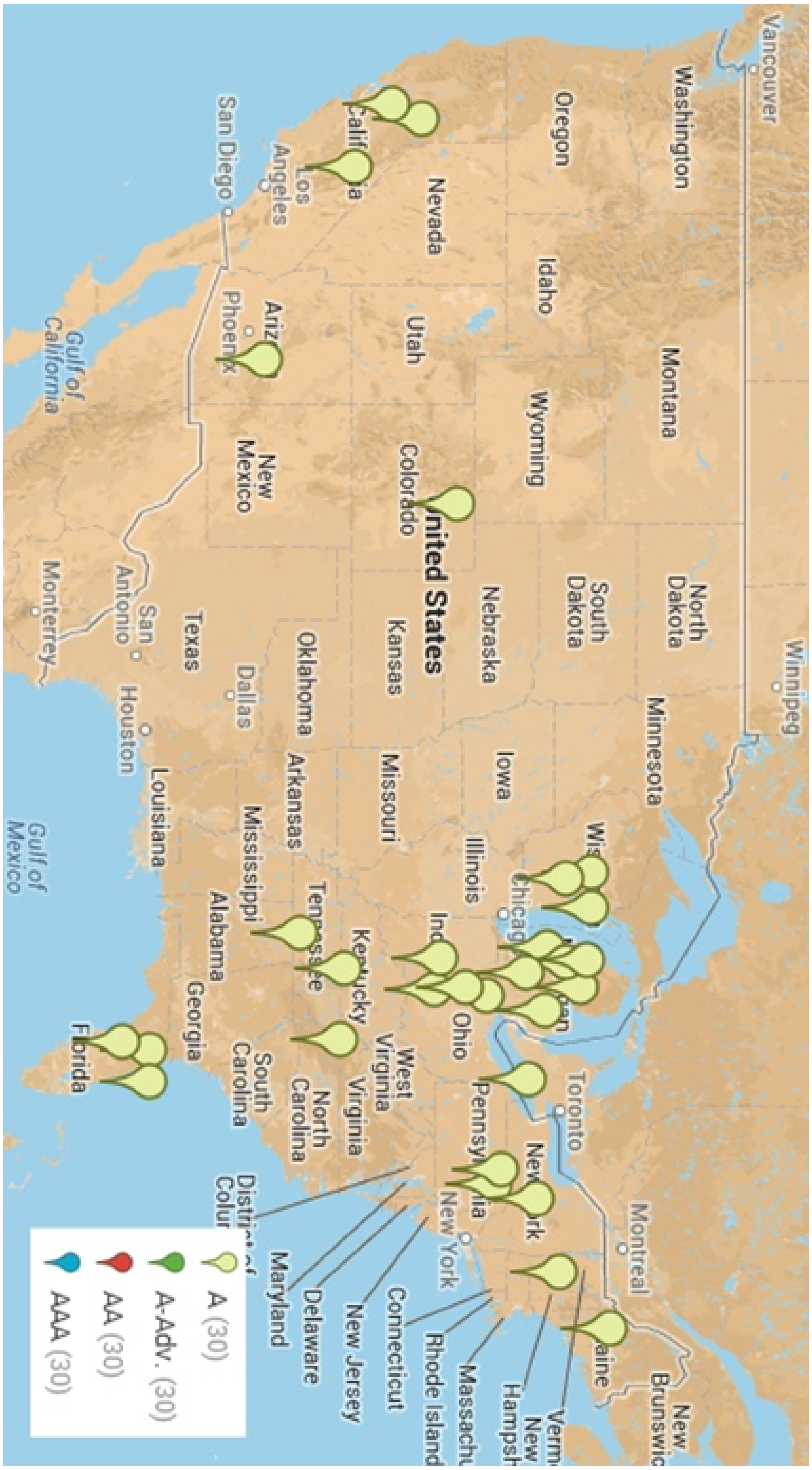

Figure 5. 2012 Class A-Advanced Team Locations

Figure 6. Predicted Class A-Advanced Team Locations

Figure 7. 2012 Class AA Team Locations

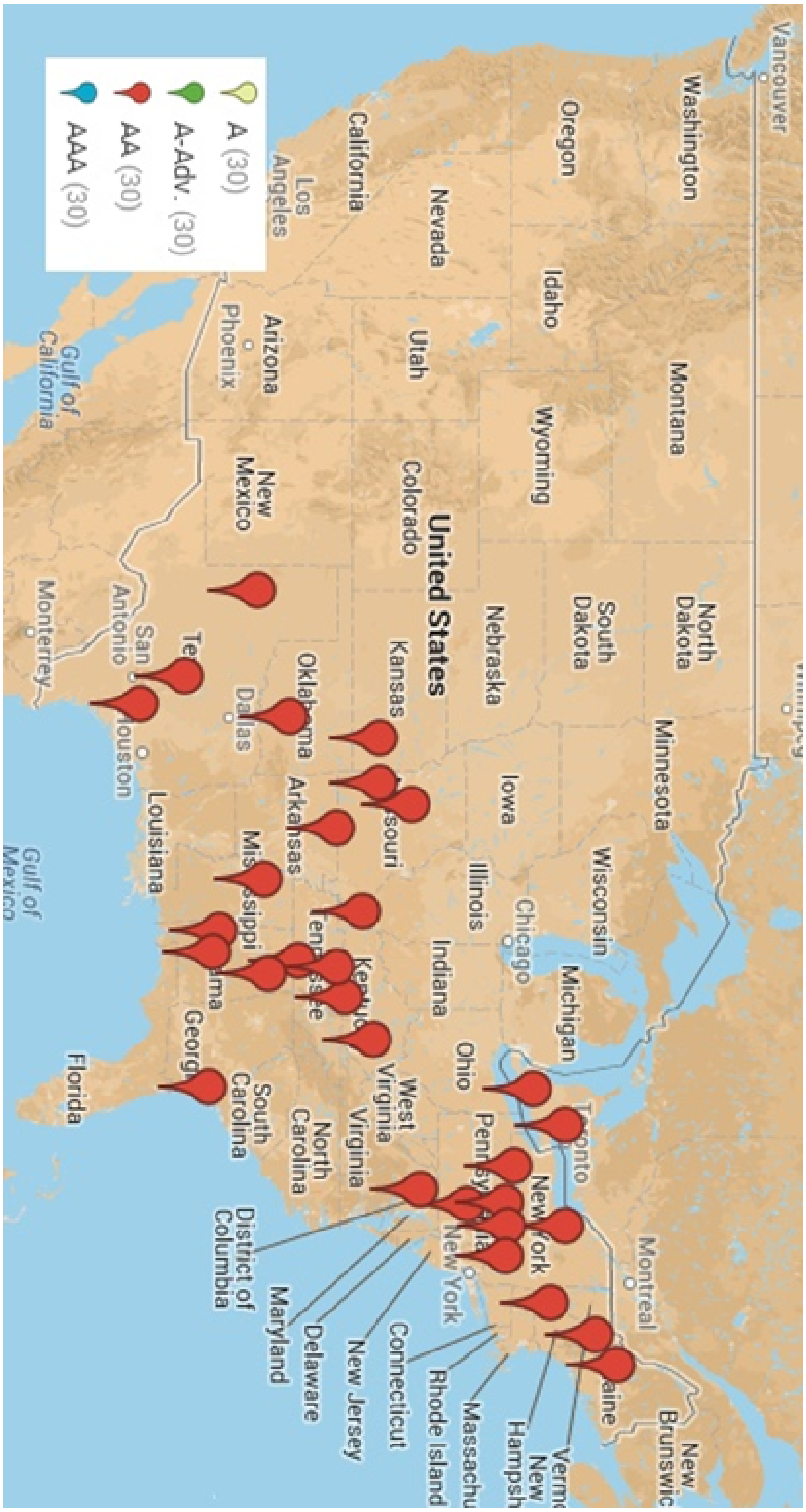

Figure 8. Predicted Class AA Team Locations

Figure 9. 2012 Class AAA Team Locations

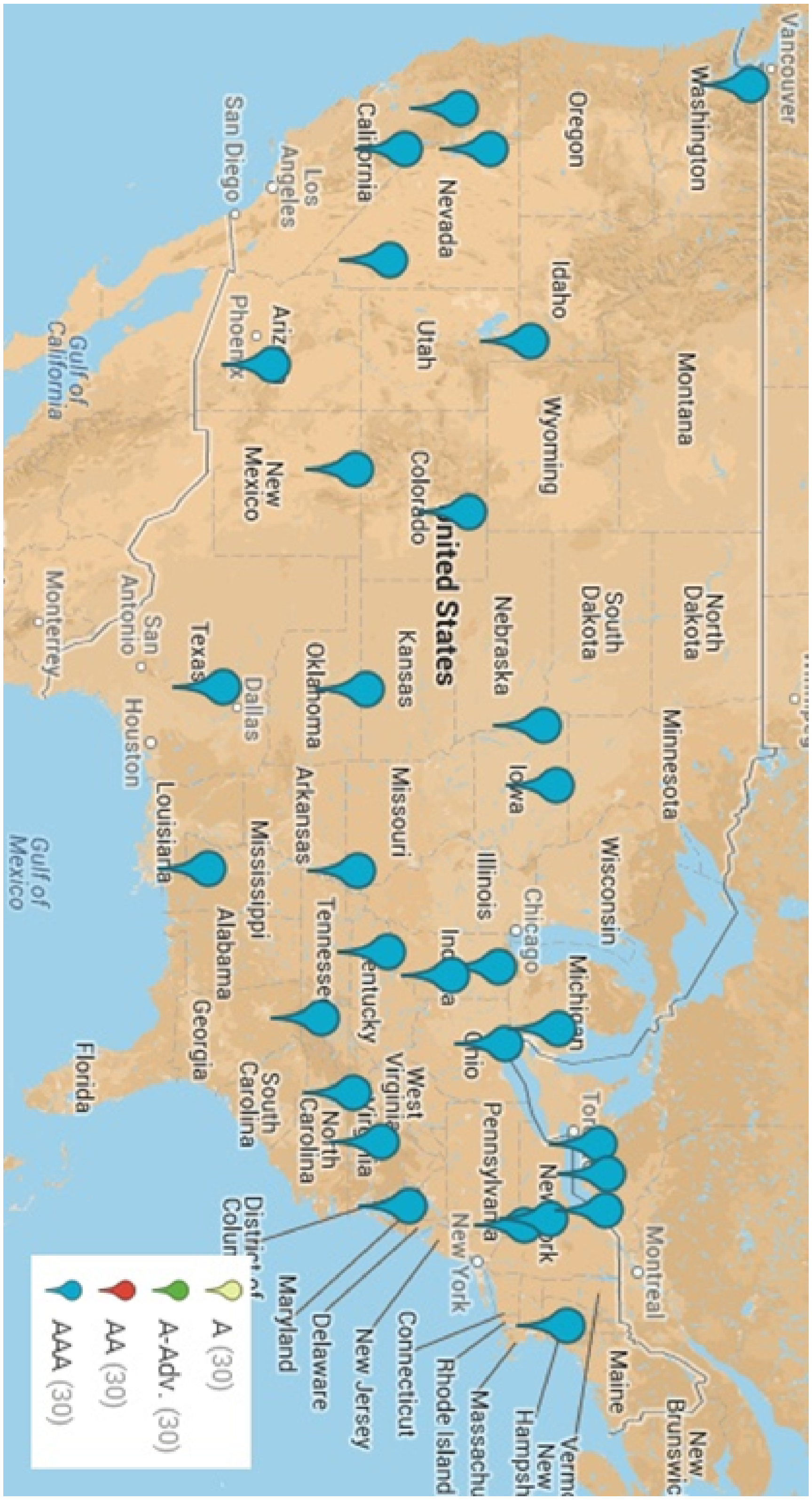

Figure 10. Predicted Class AAA Team Locations

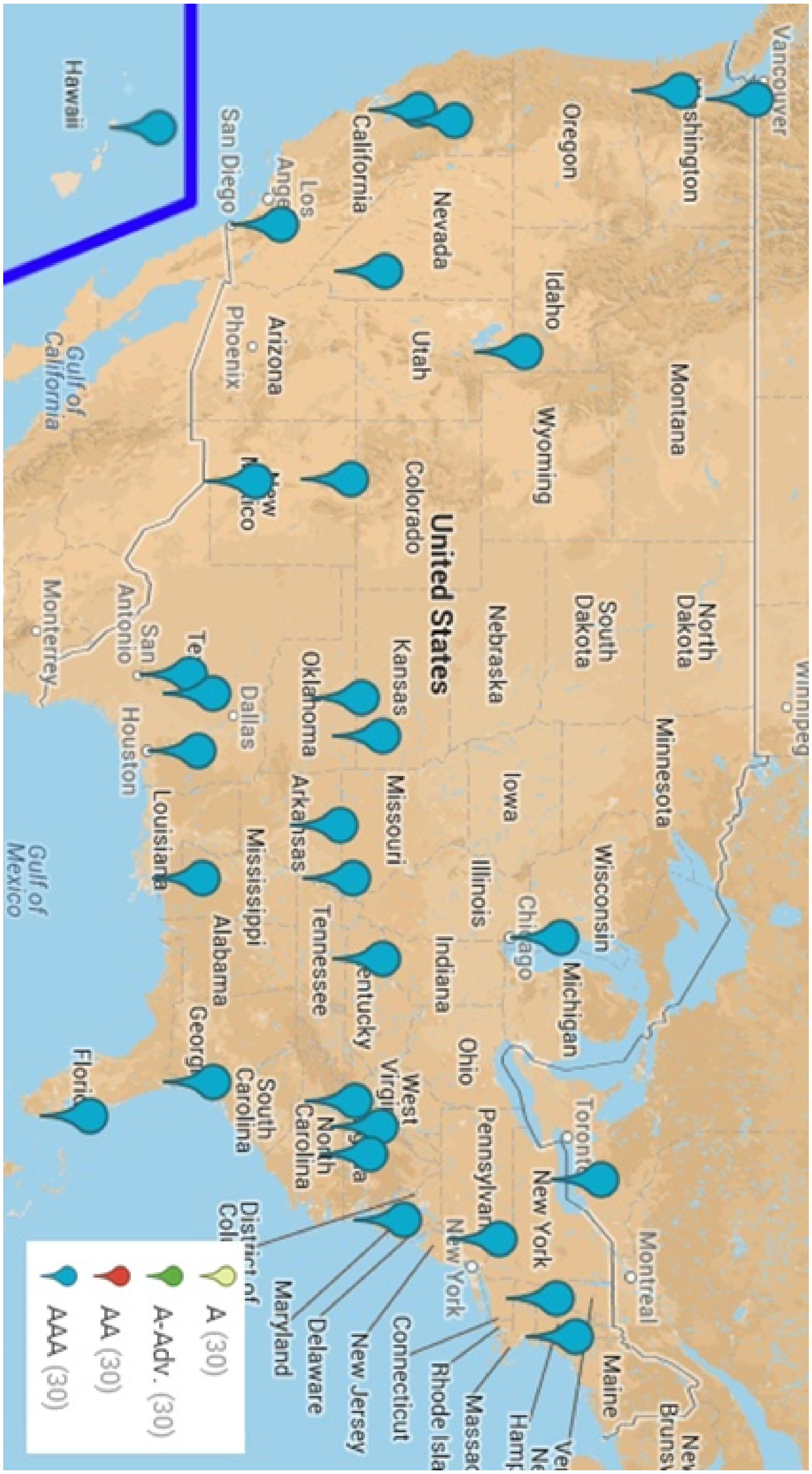

Figure 11. MLB Team Locations 2006-2012

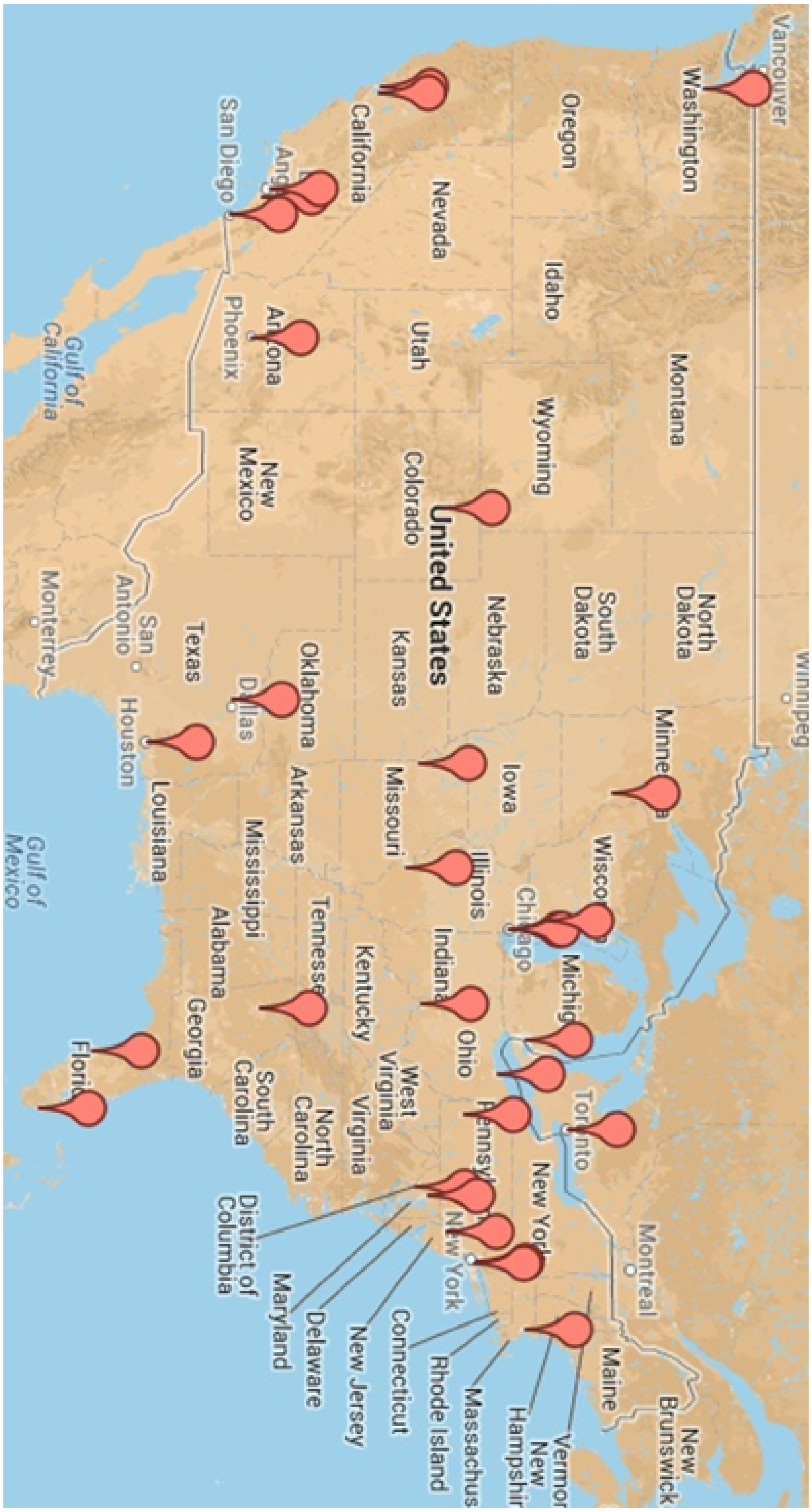

Table A1.

Top 10 Cities with Highest Probabilities to Have Team at Particular Level that do not Have a Team at that Level or Higher				
Rank	**AAA**	**AA**	**A-Advanced**	**Any Level***
1	San Antonio, TX (7)	Greensboro, NC (4)	Honolulu, HI (2)	Honolulu, HI (29)
2	Orlando, FL (8)	Grand Rapids, MI (7)	Grand Rapids, MI (3)	Springfield, MA (47)
3	Hartford, CT (17)	Albany, NY (8)	Albany, NY (4)	Anchorage, AK (49)
4	Jacksonville, FL (18)	Dayton, OH (10)	Dayton, OH (5)	Lafayette, LA (57)
5	Greensboro, NC (20)	Honolulu, HI (12)	Allentown, PA (6)	Gulfport, MS (59)
6	Grand Rapids, MI (21)	Springfield, MA (19)	Springfield, MA (8)	Fayetteville, NC (65)
7	Austin, TX (22)	Charleston, SC (20)	Charleston, SC (9)	Reno, NV (69)
8	Albany, NY (24)	Allentown, PA (21)	Baton Rouge, LA (10)	Madison, WI (77)
9	Birmingham, AL (26)	Baton Rouge, LA (22)	Boise, ID (11)	Lancaster, PA (83)
10	Tulsa, OK (27)	Boise, ID (23)	Jackson, MS (12)	Santa Barbara, CA (87)

Table A2.

Top 10 Cities with Lowest Probabilities to Have Team at Particular Level that do Have a Team at that Level				
Rank	**AAA**	**AA**	**A-Advanced**	**Any Level***
1	Scranton, PA (59)	Altoona, PA (226)	Kinston, NC (336)	Clinton, IA (459)
2	Colorado Springs, CO (53)	Jackson, TN (153)	Vero Beach, FL (114)	Richmond, IN (417)
3	Tucson, AZ (46)	Erie, PA (110)	Lynchburg, VA (87)	Oneonta, NY (416)
4	Syracuse, NY (45)	Binghamton, NY (93)	Myrtle Beach, SC (71)	Burlington, IA (410)
5	Des Moines (44)	Norwich, CT (84)	Visalia, CA (68)	Kinston, NC (369)
6	Toledo, OH (38)	Midland, TX (77)	Wilmington, NC (60)	Rome, GA (354)
7	Fresno, CA (35)	Reading, PA (52)	Roanoke, VA (57)	Martinsville, VA (330)
8	Albuquerque, NM (34)	Mobile, AL (40)	Port St. Lucie, FL (52)	Helena, MT (306)
9	Omaha, NE (31)	Huntsville, AL (35)	Modesto, CA (46)	Jamestown, NY (280)
10	Oklahoma City, OK (30)	Chattanooga, TN (33)	Lakeland, FL (36)	Great Falls, MT (279)

Source: Michael C. Davis. "Called Up to the Big Leagues: An Examination of the Factors Affecting the Locations of Minor League
*Includes levels lower than A-Advanced, including Rookie, Short-Season, and Independent Leagues.
Does not include MLB cities.

Table A3.

Descriptive Statistics

		Population[t]	Log Population	Population Change	Percent Population Change	Per Capita Income[t]	Log Per Capita Income	MLB Distance*	MLB Distance	Baseball Distance*	MLB Distance
N		2030	2030	2030	2030	2030	2030	2030	2030	2030	2030
Mean		89.9799	12.8389	0.45048	5.0385	3.6350	10.4899	2.117	1.933	1.431	1.248
Std. Error of Mean		4.6653	0.0264	0.02286	0.1064	0.0122	0.0033	0.061	0.033	0.060	0.029
Median		26.3971	12.5378	0.11918	4.5251	3.5625	10.4808	1.640	1.640	0.890	0.890
Std. Deviation		210.1968	1.1901	1.03009	4.7958	0.5519	0.1483	2.737	1.494	2.720	1.323
Skewness		6.3660	0.8506	3.86224	0.7644	0.7889	0.1759	6.223	2.386	7.278	4.562
Kurtosis		52.0114	0.3797	17.76095	2.5127	1.5190	0.8211	45.057	9.840	55.600	25.287
Range		2266.4631	6.5043	10.84017	52.6515	4.2034	1.0885	24.00	10.00	23.82	9.82
Minimum		3.3275	10.4624	-2.84653	-18.4928	2.1339	9.9683	0.00	0.00	0.18	0.18
Maximum		2269.7906	16.9666	7.99364	34.1586	6.3373	11.0568	24.00	10.00	24.00	10.00
Percentiles	20	12.7030	11.7897	0.02228	1.1329	3.1921	10.3710	0.89	0.89	0.60	0.60
	40	21.1422	12.3140	0.08161	3.4040	3.4507	10.4489	1.37	1.37	0.77	0.77
	60	39.7781	12.9249	0.17341	5.5373	3.7036	10.5196	2.02	2.01	1.02	1.02
	80	92.6867	13.8320	0.52068	8.6092	4.0551	10.6103	2.74	2.74	1.53	1.53

[t]. This is the original variable before it was transformed through the natural logarithm.

*. This is the original distribution of distance before being capped at 1,000 miles.