The Significance of Urban Hierarchy in Explaining Population Dynamics in the United States

Elizabeth A. Dobis [a*], Michael S. Delgado [a], Raymond J.G.M. Florax [a,b,c], and Peter Mulder [b]
[a] *Department of Agricultural Economics, Purdue University, West Lafayette, IN, United States*
[b] *Department of Spatial Economics, VU University Amsterdam, Amsterdam, The Netherlands*
[c] *Tinbergen Institute, Amsterdam, The Netherlands*

May 27, 2015

Selected Paper prepared for presentation for the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Abstract

In much of the literature focusing on the growth and structure of the urban system, the difference between contagious and hierarchical interrelations across cities comprised in the urban system are obfuscated. In this paper, we clearly distinguish and quantify the effects of both. In other words, we focus on how the structure of the urban system influences population growth by using central place theory as a theoretical basis for addressing the research question: what natural and man-made locational characteristics influence population growth? We make three major contributions to the existing literature. First, we utilize a unique dataset of urban areas with decennial observations from 1990 to 2010 which captures the agglomerated economic activity and built extent of urban locations with at least 2,500 inhabitants, to include all but the smallest rural communities. Second, our analysis includes both the hierarchical relationship among cities of differing sizes and the continuous nature of proximity to other cities. The novel use of a spatially-lagged hierarchical linear model allows us to include both these critical aspects of the urban system in our analysis. Third, we include man-made amenities and characteristics of cities, which have been omitted from previous studies in an effort to avoid endogeneity in the analysis. By focusing on the intercept and lagged population variables in the urban area equation, we use this model to empirically explore the debate on whether there is random or deterministic growth in the distribution of cities in the United States.

JEL Codes: R110, R120, R150
Keywords: population growth, urban hierarchy, spatial lag, hierarchical linear models

Copyright © 2015 by the authors. All rights reserved. DRAFT paper prepared for the 2015 AAEA Meeting. The opinions expressed herein are those of the authors and do not necessarily reflect the views of Purdue University, the VU University Amsterdam, or the Tinbergen Institute Readers may make verbatim copies of this document for non-commercial purposes by any means, provided this copyright notice appears on all such copies. Please do not cite without the permission of the authors.

---

[*] Corresponding author. Address: 403 West State Street, Lafayette, IN 47907; Email: edobis@purdue.edu

1 Introduction

Cities and towns are loci of population and production. In 2010, 80.7 percent of the the United States population resided in urban areas, and 88.2 percent of those individuals lived in an urban area with at least 50,000 inhabitants. These individuals rely on the goods and services available in urban areas, while much of the rural population travel to urban areas for goods and services. As Bosker and Buringh (2011, p. 1-2) stated in their study of the geography and development of the European urban system, cities are "loci for technological innovation, institutional progress, trade, political power and culture." In 2011, 90.1 percent of gross domestic product (GDP) was produced in metropolitan statistical areas (MSAs), emphasizing that urban areas are also centers of production.[1]

The high concentration of people and production in urban areas of the United States makes the structure and growth of the urban system a policy focus for national and local governments due to the economic, environmental, and social implications of urbanization. Governments use policy to alter the incentives of firms and individuals, affecting their choices and impacting the structure of the urban system through the number, location, and sizes of cities, as well as urban form and density. Historically, government policy has influenced urban growth through the alteration of transportation costs. Railroad companies received land grants from the federal government through the Land Grant Act of 1850 to encourage the construction of rail lines across the nation, strengthening flows of people and goods among cities and becoming the impetus for the establishment of new settlements. Eisenhower's Federal Aid Highway Act of 1956 established the Interstate Highway System, which raised and lowered relative transportation costs throughout the United States, altering the land use patterns and growth rates of settlements and promoting urban sprawl. Recently, The Obama Administration's creation of the Partnership for Sustainable Communities, an inter-agency collaboration that focuses on environmental protection while increasing access to affordable housing, increasing transportation options, and decreasing transportation costs, has provided grants to cities like Denver and Indianapolis meant to encourage higher population density through housing and transportation infrastructure and influence a community's urban growth rate.

Academic interest in population change can be split into two literature branches. The first branch focuses on the growth and structure of individual cities. The monocentric city model of von Thünen (1826) and the work on bid-rent curves related to Alonso (1964) are examples of this first branch. More recently, Bettencourt (2013) proposed a model that relates the size and scale of a city to its infrastructure and social dependencies.

The second literature branch focuses on the urban system. Some of the literature in this branch

---

[1] The Bureau of Economic Analysis' regional GDP data were not available for urban areas. Urban area population data are calculated using census tracts, which is more relevant to this dissertation than MSAs, which are comprised of counties and can contain rural land. The proportion of the United States population living in MSAs is 83.7 percent. Population figures are from the U.S. Census Bureau's 2010 Decennial Census and GDP data are from the Bureau of Economic Analysis.

assumes interconnectedness between cities, such as Black and Henderson (2003) and Le Gallo and Chasco (2008) who study the rank-size distribution of cities and the mobility of those cities within the statistical distribution. Other authors explicitly include the spatial proximity between cities (e.g., Partridge et al., 2008; Bosker and Buringh, 2011) or network flows among cities (e.g., Neal, 2011). These authors frequently use central place theory to frame their analysis of city growth. Partridge et al. (2008; 2009), rather than focusing on growth within cities themselves, focus on the effect incremental distance from cities of varying sizes has on population growth, earnings, and housing costs. They find that the further a hinterland county is from large cities, the smaller the values of these variables are. To avoid endogeneity issues, the authors only utilize physical features, income prior to the time period being studied, and distances as explanatory variables, despite asserting that man-made amenities also influence population growth.

In our paper, we intend to further work on the impact of the urban hierarchy on city development by exploring how the structure of the urban system influences population growth. We do this by addressing the research question: what natural and man-made locational characteristics influence population growth? We take this basic question a step further by analyzing whether natural amenities, economic composition, or proximity factors have the most influential role in population change.

Our analysis of growth in the United States urban system incorporates two aspects of central place theory. The first aspect is the idea that cities are markets. The second aspect is the urban hierarchy, which arises from economically efficient spatial land allocation. Urban hierarchy is related to the variety and specialization of goods available in a central place. Because more specialized goods and services require a larger market area to meet minimum demand requirements, cities that are high in the urban hierarchy have a wide variety of goods and services with a wide range of specializations, whereas locations lower in the urban hierarchy have only a small variety of basic goods and services.

Previous studies addressing the influence of the urban system on population growth have not fully captured the connections among cities and often omitted man-made amenities and characteristics that influence their growth. Our first contribution to the literature is the use of a novel econometric method in this field, hierarchical linear modeling, to capture the influence of the structure of the urban system on population change by allowing city-level explanatory variables to affect population growth differently, given the unique characteristics of each location's surrounding market area. Second, we also use a dataset with refined geographic agglomerations that more accurately delineate the boundary between urban and rural area in the United States. Third, we deal with the endogeneity associated with including man-made locational characteristics in our analysis.

Section 2 discusses the theory behind our analysis and Section 3 discusses the model we are using for the analysis and the details of its components. Finally, Section 4 discusses the data, it's construction, and descriptive statistics.

2 Central Place Theory, Markets, and the Urban Hierarchy

Much of the published literature on urban hierarchy utilizes a consumer-based approach to analysis, which focuses on the variety of goods and services available in a central place as well as how specialized those goods are (e.g., Partridge et al., 2008; Le Gallo and Chasco, 2008). They rely on the concept that individuals choose to live close to central places that have a large number of the goods and services they wish to consume. Producer-based analyses instead focus on employment and search costs, agglomeration, and competition (e.g., Wensley and Stabler, 1998; Polèse and Shearmur, 2004; Mori et al., 2008; Tabuchi and Thisse, 2011). Many of the theoretical general and partial equilibrium models that explain the evolution of urban hierarchies have relied strongly on this type of inquiry.

In central place theory, the importance of economic distance, remoteness in terms of time, cost, security, and physical distance, to central places results in a urban system characterized by a regular, hexagonal location pattern of market-based cities that are classified into a limited number of levels, or hierarchical tiers. Each of the settlements in the hierarchy has a hinterland (market area) from which it draws consumers. This hexagonal pattern is the most efficient way, given the costs of economic distance, to provide access to the goods and services available at each level of the urban hierarchy. Pioneering works in this area were simple observations (e.g., Christaller, 1933) or economic models (e.g., Lösch, 1943), while more recent works have incorporated this concept into larger econometric models through incremental distances to cities in the next exogenously-determined tier, such as in Partridge et al. (2008, 2009).

Once a city has established itself at a location, it experiences "inertia" in its tier classification and distributional location due to investment in infrastructure and existing agglomeration economies (Polèse and Shearmur, 2004). These established characteristics of cities make them slow to move within the hierarchy despite industry shocks and churning in urban employment. Low-order central places have less inertia and, thus, have more variance up and down the urban hierarchy (Duranton, 2007).

We implement central place theory in our analysis by including three defining characteristics of central place theory: proximity to other places, hierarchy of cities, and heterogeneity of goods.

The basic tenets of central place theory can be explained starting with the heterogeneity of goods. This heterogeneity implies a hierarchy of goods based on their characteristics. Goods that are more specialized or less common are less price elastic than everyday goods (McCann, 2001). They are consumed less frequently but individuals are willing to travel greater distances to purchase these goods. Therefore, these high-tier goods need a larger market area from which to draw their consumers. These goods are located in the largest cities in an urban system. Consumers are not willing to drive as far for low-tier everyday goods, making them more price elastic with smaller market areas than specialized goods. Because consumers are less willing to travel for these low-tier goods, they are available in the large number of small cities in the urban system that have small market areas. Specialized goods are available in large, centralized cities because they require a large market area. Therefore the heterogeneity of goods leads to a hierarchy of cities characterized by size and the diversity of goods available for purchase.

Because consumers wish to minimize their costs, they travel to the nearest city from which they can obtain the goods they wish to purchase. This implies that there is competition among cities of the same tier in the urban hierarchy. Between cities in different hierarchical tiers, there is both competition (because large cities also have all goods available in small cities) and complementarity (due to the ability to move up tiers to purchase goods). Proximity to other cities determines how much competition or complementarity affects population growth. This cost minimization behavior also leads to smaller market areas being nested within larger market areas, since consumers wish to travel the smallest distance to obtain any good.

3 Model

Much of the empirical work on city growth looks at population or employment change in cities, disregarding the effect of proximity or hierarchy. Glaeser et al. (1992) use employment change to determine whether the knowledge spillover theories of Marshall, Porter, or Jacobs influence economic growth, but ignore the role of the proximity of cities in their analysis. And empirical work that does include proximity variables, such as Partridge et al.'s (2008; 2009) cross-sectional analyses that include the incremental distance to cities in higher tiers, has utilized unilevel econometric models that include proximity variables.

In unilevel models, all variables, whether they capture aspects of cities or market areas, directly and independently influence population change. When a weights matrix is included in the model, the result is a single interconnected system of cities in which all cities influence all other cities, with the strength of that influence dependent on the distance separating each location. In reality, the effect of city-level explanatory variables on population change may vary depending on the unique characteristics of each market area. These unique characteristics arise because, unlike the assumptions of central place theory, the world is not a flat, featureless plain.

To include this possibility in the analysis, we use a multilevel model, also known as a hierarchical linear model (HLM). This will allow the city-level coefficients to vary based on characteristics of the market areas. Due to the nested structure of the model, characteristics of the top level of the hierarchy will influence the lowest level of the model through their effect on intermediate levels of the model.

The number of levels in the model is the same as the number of tiers in the urban hierarchy because the hierarchical classification of cities is directly related to the goods hierarchy. At each level of the goods hierarchy, goods and services that are more specialized and have a higher price elasticity are available, in addition to the goods available at lower tiers. The larger market area required to sustain these more specialized goods subsumes the smaller market areas for the lower tiers, creating a nested structure. Therefore, each higher tier has a larger market area, resulting in an equal number of model level as city tiers.

A basic three-level conditional hierarchical linear model is specified as:

$$
\begin{aligned}
y_{ijk} &= \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk} a_{pijk} + e_{ijk} \\
\pi_{pjk} &= \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk},\ \forall\, p = 1, \ldots, P \\
\beta_{pqk} &= \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} Z_{sk} + u_{pqk},\ \forall\, p = 1, \ldots, P \text{ and } q = 1, \ldots, Q_p
\end{aligned} \tag{1}
$$

where i is an urban area from any tier, j is a small market area, and k is a large market area. Y is the dependent variable, population change in urban area i. The number of independent variables a in level 1 is p, the number of independent variables X in level 2 is q, and the number of independent variables Z in level 3 is s. The errors are e, r, and u for each level, respectively.

3.1 Defining the Urban Hierarchy

Research on central place theory has been conducted since its introduction by Walter Christaller (1933) and August Lösch (1943), but has experienced a resurgence of interest in recent years (Mulligan et al., 2012). In empirical studies, the urban hierarchy has been approached demographically using population size to classify cities or functionally using a more systematic classification scheme (Beaverstock et al., 1999).

Classifying cities using population size is the most common method, and is utilized in studies such as Borchert (1967), Partridge et al. (2008, 2009), Lorenzen and Andersen (2009), and Maliszewski and Ó hUallacháin (2012). The authors use population as a proxy for the centrality of a city's markets (Neal, 2011). These studies exogenously determine the number of tiers in the urban hierarchy and their population intervals, and each author has his or her own critical points. Others, like Lorenzen and Andersen (2009) and Maliszewski and Ó hUallacháin (2012), completely avoid tier classifications by using the rank-size distribution of cities to represent the urban hierarchy.

The demographic classification of hierarchical tiers is also frequently used in general equilibrium research that studies the evolution of urban hierarchy, as in Fujita et al. (1999) and Hsu et al. (2013). In his newest book, *The New Science of Cities*, Michael Batty devotes two chapters, one theoretical and one empirical, to various methods of creating and defining urban hierarchies using primarily population size.

Because population size is a proxy for centrality, it does not capture the theoretical interurban relationships in central place theory (Neal, 2011). Functional approaches to classifying cities move away from the arbitrary aspects of demographic city classification and incorporate the economic switch from size-based hierarchy to network-based hierarchy that Neal (2011) finds in his work on the evolution of the urban hierarchy in the United States.

Functional approaches to classifying cities in the urban hierarchy are less common in the literature. Beaverstock et al. (1999) assigns points to cities based on the prevalence of global firms in accounting, advertising, banking, and law to determine the cities at the top of the world urban hierarchy. Neal (2011) uses airline passenger information to create a network-based urban hierarchy and capture the flows among cities. And in their study of spatial interactions among American MSAs, Dobkins and Ioannides (2001) utilize an urban hierarchy classification scheme created by Noyelle and Stanback (1984) and modified by Knox and McCarthy (2005) that emphasizes the nodal characteristics of cities. Additional methods, such as the spring and block system Kovács et al. (2013) apply from physics using population and economic weights, can also be used to create functional hierarchical classifications.

The economic and theoretical logic behind functional city classification makes it a more attractive option for our analysis. We use a modified version of the nodal classification system used by Overman and Ioannides (2001) and Dobkins and Ioannides (2001) for our hierarchy. This system has four distinct tiers. The highest tier is the nodal tier, which is characterized by what Knox and McCarthy (2005) call primary and secondary world cities. These cities are international hubs, of finance, trade and government. This is followed by regional nodes and subregional nodes that are considered to be regional control centers (Knox and McCarthy, 2005). These cities are important national hubs of industry, banking, services, education, medicine, and public institutions. Finally, all other cities are considered members of the final tier. A full listing of cities and their tier classification can be found in Table 1.

The tiers, as defined by Overman and Ioannides (2001) and Dobkins and Ioannides (2001), were not properly nested for a HLM analysis. To address this problem, we combined tier 2 and tier 3 and reassigned San Francisco to tier 2. This results in three-tier hierarchy with a physically nested structure that can be used for in a hierarchical level analysis. We refer to tier 1 cities as central place nodes, tier 2 and tier 3 cities as regional nodes, and tier 4 cities as urban areas.

3.2 Our Model

Our dependent variable is the change in the population of an urban area, calculated as the difference of natural logarithms. In the literature urban growth and urbanization are often explained by groups of explanatory variables. Physical geography, amenity, market, economic, demographic, 1^{st} nature, and 2^{nd} nature are a few of the classifications used (Nzaku and Bukenya, 2005; Partridge et al., 2008; Bosker and Buringh, 2011; Olfert et al., 2012). We explain changes in the population of urban areas via hierarchy, heterogeneity, and proximity. Hierarchy is incorporated through our econometric setup, while our explanatory variables address heterogeneity and proximity. This results in the following model:

Urban Area:

$$\begin{aligned}
&\Delta Pop_t = \pi_0 + \pi_1 Pop_{t-1} + \pi_2 GoodsIndex_{t-1} + \pi_3 DistanceRegionalNode + \\
&\qquad \pi_4 GL/OceanProximity + \pi_5 Ruggedness + \\
&\qquad \pi_6 TemperateClimate + e
\end{aligned}$$

Regional Market Area:

$$\begin{aligned}
&\pi = \beta_0 + \beta_1 AggregateIncome_{t-1} + \beta_2 DistanceCPNode + \\
&\qquad \beta_3 AgEmploymentShare_{t-1} + \beta_4 MfgEmploymentShare_{t-1} + \\
&\qquad \beta_4 RegionalRuralLandProportion + r
\end{aligned} \tag{2}$$

Central Place Market Area:

$$\gamma = \omega_0 + \omega_1 CPRuralLandProportion + v$$

In the urban area equation, the variable *GoodsIndex* refers to a goods centrality index calculated to include the variety and balance of products in each urban area. This is an indication of heterogeneity and includes the goods hierarchy in the equation. More about the construction of this index can be found in section 3.4. Other heterogeneity variables are *GL/OceanProximity*, which measures the distance to the nearest large body of water, and *Ruggedness*, which is measured on an ordered categorical scale from 1 to 9 such that the lowest values indicate flat plains and the highest values rugged mountains. *TemperateClimate*, following Glaeser et al. (2001), is measured as the inverse of the average annual temperature minus 70 degrees Fahrenheit, and indicates how mild temperatures are in an urban area. These three variables measure the heterogeneity of natural amenities among urban areas. *DistanceRegionalNode* is the travel distance to the nearest regional node. and is a proximity variable.

The regional market area equation has two main parts. First, *DistanceCPNode* is the travel distance from the regional node to the nearest central place node, and is a proximity variable. Second, *AgEmploymentShare* and *MfgEmploymentShare* indicate the proportion of the regional market area's economy devoted to agriculture and manufacturing. These variables will include how the region's predominant industry influences population growth in urban areas. Economic heterogeneity of the regional market area is also measured by *AggregateIncome*, which represents purchasing power, in similar way to the income bands of market power in Partridge et al. (2008). *RegionalRuralLandProportion* accounts for the construction of regional market area variables by summing urban area values. The central place market area has only one variable, *CPRuralLandProportion*, a control variable constructed in the same way as its regional market area equivalent.

We currently estimate this as a cross-sectional HLM model for urban areas and a panel HLM model for urbanized areas.

3.3 Including Spatial Proximity of Urban Areas

The basic hierarchical linear model discussed above captures the hierarchical structure of urban areas and their markets by allowing city-level explanatory variables to affect population change differently, given the unique characteristics of each location's higher-level regional and central place market areas. However, this model structure only includes the effect of hierarchy in the analysis, and the inclusion of the distance to the nearest regional and central place market nodes only captures proximity to the nearest higher-tier cities. The effect of urban areas on their neighbors is excluded.

Spatially, urban areas both compete with and compliment each other. Looking at where cities formed in Europe over the last 2,200 years,Bosker and Buringh (2011) found that early in European history sites located next to other existing cities were less likely to have a evolved into an urban area, and that, since the 17^{th} century, sites a moderate distance away from existing cities have expanded due to co-location benefits and decreasing transportation costs.

Urban areas within the same tier have similar types of goods available, creating competition among them for population growth. Urban ares located close to each other also tend to have similar natural amenities, such as temperature or a mountainous landscape, intensifying the competition for population among them. They are attracting the same type of residents. Alternatively, urban areas in differing tiers have a different variety of goods and services available, as well as differing characteristics, such as population size or density. This combination of differing characteristics may act in a complementary way, with each urban area attracting different types of residents that may enjoy the different amenities available in nearby urban areas of differing tiers.

The competition and complementarity of urban areas emphasizes the importance of proximity, which we include in our model through a spatial lag. This allows us to include spatial spillovers due to population change, the effect of population change in nearby urban areas on a given urban area. We follow the model and estimation of Baltagi et al. (2014) to include these spatial spillovers in our model. This allows us to concurrently account for both the hierarchical relationship among cities of differing sizes and contiguous effects through the inclusion of neighbors.

The spatial lag three-level conditional hierarchical linear model is specified as:

$$\begin{aligned} y_{ijk} &= \rho \sum_{f=1}^{F_{gh}} \sum_{g=1}^{G_h} \sum_{h=1}^{H} w_{ijk,fgh} yfgh + \sum_{p=0}^{P} \pi_{pjk} a_{pijk} + e_{ijk} \\ \pi_{pjk} &= \sum_{q=0}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk},\ \forall\, p = 0, \ldots, P \\ \beta_{pqk} &= \sum_{s=0}^{S_{pq}} \gamma_{pqs} Z_{sk} + u_{pqk},\ \forall\, p = 0, \ldots, P \text{ and } q = 0, \ldots, Q_p \end{aligned} \tag{3}$$

where $i = 1, \ldots, I_{jk}$ and $f = 1, \ldots, F_{gh}$ represent an urban area from any tier, $j = 1, \ldots, J_k$ and

$g = 1, \ldots, G_h$ represent a small market area, and $k = 1, \ldots, K$ and $h = 1, \ldots, H$ represent a large market area.

Y is the dependent variable, population change in urban area i. The number of independent variables a in level 1 is p, the number of independent variables X in level 2 is q, and the number of independent variables Z in level 3 is s. The weight matrix element for urban area i in regional market area j in central place area k is $w_{ijk,fgh}$, while the scalar ρ is the spatial effect. The errors for each level, e_{ijk}, r_{pjk}, and u_{pqk}, are assumed to be independent and identically distributed (i.i.d.), with means and variances of $(0, \sigma_e)$, $(0, \sigma_r)$, and $(0, \sigma_u)$, respectively.

3.4 Incorporating the Goods Hierarchy

In central place theory, urban areas act as central marketplaces for the sale and purchase of goods and services. As Fujita et al. (1999) showed in their dynamic general equilibrium model, product variety can influence the evolution of an urban hierarchy. The differentiation and specialization of products, represented as differing elasticities in both the Fujita et al. (1999) model and in Lösch (1943)'s central place theory model, creates a hierarchy of goods that influences the size and growth of an urban area, making it an essential part of our analysis. We include the goods hierarchy in our analysis through a diversity index.

Diversity indices are widely used throughout social and natural sciences, particularly in ecology, but are conceptualized and applied differently (Stirling, 1998; Maignan et al., 2003). In regional science, diversity is usually calculated relative to a benchmark (Wagner, 2000). Ellison and Glaeser (1997) took the basic Herfindahl Index (known as the Simpson Index in ecology), $H = \sum_i p_i^2$, where p_i is the proportion of individuals in classification i, and replaced the simple proportion with the difference between the regional proportion of national employment in sector i and the total proportion of regional to national employment, which was written as $HHI_i = \sum_{r=1}^{R} (\frac{E_{ir}}{E_{in}} - \frac{E_r}{E_n})^2$ by Black and Henderson (1999). This modification changes the interpretation of the index from a measure of similarity within the community to an indication of concentration within an urban system. Other prominent diversity indexes in regional science, such as the relative diversity index and the location quotient use similar benchmarks.

Using benchmark-based measures for our analysis would give the diversity of an urban area relative to its market area. This is an informative metric, but does not accurately capture the spirit of product diversity related to a hierarchy of goods. Therefore, we use the simpler ecology-based indices that measure diversity within in a community to represent the goods hierarchy in our analysis.

Many of the existing diversity indices focus on either variety ("richness"; e.g., species count and numerical richness), balance ("evenness"; e.g., Shannon and McIntosh Evenness), or a combination of both (e.g., Shannon-Wiener and Simpson Indices) (Stirling, 1998; Maignan et al., 2003). Applying these terms to our analysis, variety is the number of types of establishments in an urban area, as defined by their industry classification, and balance describes how establishments are distributed among these types.

Stirling (1998) acknowledges a third aspect to diversity, disparity, which he defines as the dissimilarity between establishment types. He suggests including all three aspects of diversity in a single measure he calls the integrated multi-diversity index. This is the same as the quadratic entropy index introduced by Rao (1982) and will be referred to as the Rao-Stirling Index henceforth.

Indices including disparity have not been widely used in the literature, compared to variety and balance indices. Desmet et al. (2009) find that including disparity in their analysis results in diversity having a statistic and economic significance on transfers and subsidies, whereas excluding disparity from the diversity measure leads to insignificant results. Leydesdorff and Rafols (2011), on the other hand, indicate that interpretation of the Rao-Stirling Index in analysis can be difficult. They also find that it is extremely sensitive to the distance metric used in its calculation, which, according to Stirling (1998), is not straight-forward. Additionally, Ranaivoson (2005) argues that information can be lost when using the Rao-Stirling Index and suggests using separate measures of diversity to compensate for this as well as to include the demand and supply forces affecting diversity.

Because of the difficulties and drawbacks associated with the Rao-Stirling Index, we initially follow traditional approaches to diversity indices and Ranaivoson (2005)'s suggestion of utilizing separate measures for the differing components of diversity. To capture the variety and balance of establishments in US urban areas, we use the Shannon-Wiener Index, defined as:

$$SW_r = -\sum_{i=1}^{I} \frac{E_{ir}}{E_r} \ln\left(\frac{E_{ir}}{E_r}\right) \tag{4}$$

where E_{ir} is the number of establishments in sector i of urban area r and E_r is the total number of establishments in urban area r. This measure is preferred over the Herfindahl Index because it places more weight on less frequent varieties than the Herfindahl Index does. Because we found no studies that utilized a diversity index that only included disparity, it is not included in this initial model.

The goods hierarchy contains rare products that require a large market area to sustain their profitability. While the Shannon-Wiener index is built from a definition of establishment rarity, it focuses on the rarity of goods within the urban area being studied, such that each urban area has a continuum of rare and common products (Maignan et al., 2003). This does not account for the availability of rare products at higher hierarchical levels, so we modify the Shannon Index to include the range of goods available within the central place market area.

Rarity is most frequently defined as a proportion. We define the rarity of sector i in the central place market area m as $R_i = 1 - (E_{im}/E_m)$. This produces a continuous measure of industry rarity at the central place market level bounded between 0 and 1 that we use to weight the industries included in the Shannon-Wiener Index. This results in a goods centrality index that is defined as:

$$C_r = -\sum_{i=1}^{I} \frac{E_{ir}}{E_r} * R_i \ln\left(\frac{E_{ir}}{E_r}\right) \tag{5}$$

that we use to indicate how centralized an urban area is based on the diversity and rarity of the goods and services available to consumers.

4 Data

4.1 Unit of Anaysis

Utilizing the correct level of spatial aggregation for an urban system analysis can be difficult with existing Census geographies. Metropolitan and micropolitan statistical areas are often chosen as the unit of analysis (e.g., Dobkins and Ioannides, 2001; Neal, 2011). These areas incorporate commuter flows beyond city limits and are created by consolidating county-level information. This construction process often includes rural hinterlands that are not part of the urban area in the MSA, which prevents the clear separation of cities and hinterlands assumed in central place theory. Some authors use incorporated places, the government-defined legal boundaries of settlements when doing their analysis (e.g., González-Val, 2010). This breaks up the agglomerated economic activity that comprises urban locations and is essential to the logic behind central place theory.

Our unit of analysis is urban areas. These are geographic regions created by the United States Census Bureau to capture the built extent of urban locations. Urban areas capture agglomerated economic activity through employment flows while excluding rural hinterlands that are not part of the urban area, in accordance with the clear separation of cities and hinterlands assumed in and essential to central place theory.

The geographic unit "urban area" was introduced by the Census Bureau in 2000 as an improvement on how they measured urban and rural areas. Urban areas are constructed from census blocks and tracts that meet population size and density requirements to create an "urban footprint" that includes residential, commercial, and non-residential urban land uses (US Census Bureau, 2011, 2013; Bureau, 2012a).

Urban areas consist of two subgroups, "urbanized areas" and "urban clusters." These two subgroups are created using the same criteria but are differentiated by total population. If an urban area has at least 50,000 inhabitants, it is classified as an urbanized area. If an urban area has at least 2,500 inhabitants and fewer than 50,000 inhabitants it is classified as an urban cluster. This allows areas that meet population and population density requirements to be considered urban areas, regardless of their legal municipal status.

There are two main issues associated with using urban areas as the base unit of observation. The first issue is that that no urban cluster information exists for 1990. Therefore, there is no

Census-created urban cluster information available for lagged independent variables, such as initial population in the urban area. The second issue is that the Census Bureau changes its urban area criteria every decennial census in order to adjust to current ideas of what is considered urban as well as to take advantage of better spatial information. Therefore, the urbanized areas for 1990, 2000, and 2010 were created using different criteria.

These two issues can be solved with one course of action, creating consistent urban areas using a single decennial censuss urban area criteria. Constructing urban areas is a time and labor intensive process that is difficult to replicate. It requires obtaining boundary files for census blocks, block groups, and tracts as well as lakes and rivers, roads, airports, and other built features, coding an algorithm of the criteria, and checking the results.

This was not a feasible option for us, so we are using the Census-defined urbanized areas for 1990 and urban areas for 2000 and 2010. We address the issue of missing urban cluster information for 1990 by exploring two separate analyses with the data, an an analysis of urbanized areas from 1990 to 2010 (matched urbanized areas, MUZA) and an analysis of urban areas from 2000 to 2010 (matched urban areas, MUA). Urban clusters are 86 percent of the urban areas in the United States, so the 2000 to 2010 analysis allows us to observe interactions throughout the entire urban system. The 1990 to 2010 analysis provides a longer time-period of analysis that focuses on the 88 percent of the population that lives in urbanized areas. See Figure 1 for a visualization of these two datasets.

We next address the issue of changing urban area criteria. To our knowledge, there are no reports, working papers, or published papers about whether this makes a statistically significant difference in variables that are calculated at the aggregate urban area level. The Census Bureau was not able to provide any documentation regarding this issue or any referrals to individuals that may have studied this. This will be accounted for using econometric methods and in statistical error.

4.2 Variables

Data for the urban area and urbanized area variables are from a variety of sources. Population and income data are from United States Census Bureau decennial census data at the urban area geographic level. The data and corresponding boundary files were obtained from the National Historic Geographic Information System (NHGIS).

We utilized the National Establishment Time Series (NETS) database from 1990 and 2000 to determine product diversity in each urban area. Walls and Associates processes yearly data collected by Dun and Bradstreet on the population of establishments in the United States to create the NETS database (Walls & Associates, 2013). The location (both county and lat-long coordinates), employees, sales, and North American Industrial Classification System (NAICS) classification of each establishment is tracked from birth to death, in addition to further details about ownership and headquarters. We use information on the location and NAICS classification to construct our diversity variables. This dataset was also used to calculate agricultural and manufacturing industry employment shares.

In our initial analysis, diversity indices are constructed for 3-digit (subsector), 4-digit (industry group), and 6-digit (national industry) classifications in sectors 48 (transportation and warehousing) through 81 (other services). This sectoral restriction aligns with the 12 NAICS service sectors that were used to develop the North American Product Classification System (NAPCS) (Bureau, 2012b). Developers chose these sectors because they produce the majority of products and include the most dynamic industries in Canada, Mexico, and the United States. We focus on these 12 sectors because many of the products in these sectors are produced and purchased in the same urban area. This is less likely to occur with sectors such as utilities (22) or manufacturing (31-33). We primarily concentrate on the 4-digit classifications in these sectors because the 6-digit classification does not contain enough variation for informative diversity indices.

Physical and environmental features data are collected from the United States Geological Survey Global Ecosystems data (land forms/ruggedness), the PRISM Climate Group (temperature), and the GLOBE Project (elevation). Because urban areas are constructed from small Census geographies (blocks, block groups, and tracts), there are no natural feature data available at that level of geographic aggregation. Instead, we use a geographic information system, ArcGIS, to create aggregated urban area values grid cell data for these variables. Spatial proximity to environmental features such as the Great Lakes/oceans or higher tier urban areas, as well as urban and rural land area, is also calculated with ArcGIS.

Regional market area and central place market area variables are the sum of all urban areas within each market area. To create the regional market areas, urban areas are assigned to the closest regional node (tier 2/3 urban area) by network distance (roads). The regional market areas are then assigned to the central place node (tier 1 urban area) closest to its regional node. Hinterlands were created for each urban area using Thiessen polygons and were aggregated to create each market area polygon. The regional and central place market areas are shown in Figure 3 and Figure 4.

4.3 Descriptive Statistics

The number of urban areas in each central place market area by tier is listed in Table 2. In both the MUZA and MUA datasets, the Chicago market area contains the largest number of urban areas (MUZA - 87; MUA - 1028). In the MUA dataset, this is over twice the number of urban areas than in the second most population market area, Houston (MUA - 507). Houston has only the third largest number of urban areas in the MUZA dataset. This indicates that a large proportion of the urban areas in the Houston market area are under 50,000 inhabitants. This observation is also true for the Denver and Seattle market areas.

The maps of the regional market areas in Figure 3a and Figure 4a show an interesting trend of smaller market areas in the eastern United States and larger market areas in the western United States. This suggests that driving distance may be perceived differently in these parts of the country.

Table 3 contains standard descriptive statistics for the MUZA dataset. Of particular interest is the range of the population change variables from 1990-2000 and 2000-2010. There is much less variation in the population change of urban areas between 2000 and 2010 than there was between 1990 and 2000. In particular the maximum population growth in the first period of the time seriew (1990-2000)was much larger than in the second period (2000-2010). This pattern may be related to the tech boom in the late 1990s and the Great Recession in the late 2000s.

Population change in the datasets was also visualized in Figures 5a, 5b, and 6. There seems to be a visual cluster of high population growth urbanized areas (MUZA dataset) in the Miami central place market between 1990 and 2000. And between 2000 and 2010 a visual cluster of urbanized areas (MUZA dataset) in the Rust Belt had a modest decline in population, as is indicated by their change from green to orange between years.

More interesting descriptive statistics are available for:

- MUZA descriptive statistics by tier (Table 4)
- MUZA descriptive statistics by central place market area (Table 5)
- MUA standard descriptive statistics (Table 6)
- MUA descriptive statistics by tier (Table 7)
- MUA descriptive statistics by central place market area (Table 8)

5 Results

Results are forthcoming for the 2015 AAEA Annual Meeting.

6 Conclusion

Conclusions are forthcoming for the 2015 AAEA Annual Meeting.

References

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent.* Harvard University Press.

Baltagi, B. H., B. Fingleton, and A. Pirotte (2014). Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices. *Journal of Urban Economics 80*, 76–86.

Beaverstock, J. V., R. G. Smith, and P. J. Taylor (1999). A roster of world cities. *Cities 16*(6), 445–458.

Bettencourt, L. M. (2013). The origins of scaling in cities. *Science 340*(6139), 1438–1441.

Black, D. and V. Henderson (1999, May). Spatial evolution of population and industry in the United States. *American Economic Review 89*(2), 321–327.

Black, D. and V. Henderson (2003). Urban evolution in the USA. *Journal of Economic Geography 3*(4), 343–372.

Borchert, J. R. (1967). American metropolitan evolution. *Geographical Review 57*(3), 301–332.

Bosker, M. and E. Buringh (2011). City seeds: Geography and the origins of European cities. Technical report, CEPR Discussion Paper 8066.

Bureau, U. C. (2012a). 2010 urban area FAQs.

Bureau, U. C. (2012b, July). North american product classification system, overview and progress report.

Christaller, W. (1933). *Die Zentralen Orte in Süddeutschland* (1966 Translation ed.). Englewood Cliffs: Prentice Hall.

Desmet, K., S. Weber, and I. Ortuño-Ortín (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association 7*(6), 1291–1318.

Dobkins, L. and Y. M. Ioannides (2001). Spatial interactions among US cities: 1900–1990. *Regional Science and Urban Economics 31*(6), 701–731.

Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. *The American Economic Review*, 197–221.

Ellison, G. and E. L. Glaeser (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy 105*(5), 889–927.

Fujita, M., P. Krugman, and T. Mori (1999). On the evolution of hierarchical urban systems. *European Economic Review 43*(2), 209–251.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of Political Economy 100*(6), 1126–52.

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

González-Val, R. (2010). The evolution of U.S. city size distribution from a long-term perspectve (1900–2000). *Journal of Regional Science 50*(5), 952–972.

Hsu, W.-T., T. J. Holmes, and F. Morgan (2013). Optimal city hierarchy: A dynamic programming approach to central place theory. Technical report.

Knox, P. L. and L. M. McCarthy (2005). *Urbanization: An Introduction to Urban Geography.* Pearson Prentice Hall.

Kovács, A., G. Máté, and Z. Néda (2013). Hierarchical settlement networks. *Regional Statistics* (01), 30–40.

Le Gallo, J. and C. Chasco (2008). Spatial analysis of urban growth in Spain, 1900–2001. *Empirical Economics 34*(1), 59–80.

Leydesdorff, L. and I. Rafols (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics 5*(1), 87–100.

Lorenzen, M. and K. V. Andersen (2009). Centrality and creativity: Does Richard Florida's creative class offer new insights into urban hierarchy? *Economic Geography 85*(4), 363–390.

Lösch, A. (1943). *Die räumliche Ordnung der Wirtshcaft* (2nd Revised Edition ed.), Volume 1967. New York: John Wiley & Sons, Inc.

Maignan, C., G. Ottaviano, D. Pinelli, and F. Rullani (2003). Bio-ecological diversity vs. socio-economic diversity: A comparison of existing measures. Working Paper 13, Fondazione Eni Enrico Mattei (FEEM).

Maliszewski, P. J. and B. Ó hUallacháin (2012). Hierarchy and concentration in the American urban system of technological advance. *Papers in Regional Science 91*(4), 743–758.

McCann, P. (2001). *Urban and Regional Economics*, Volume 15. Oxford: Oxford University Press.

Mori, T., K. Nishikimi, and T. E. Smith (2008). The number-average size rule: A new empirical relationship between industrial location and city size. *Journal of Regional Science 48*(1), 165–211.

Mulligan, G. F., M. D. Partridge, and J. I. Carruthers (2012). Central place theory and its reemergence in regional science. *The Annals of Regional Science 48*, 1–27.

Neal, Z. P. (2011). From central places to network bases: A transition in the US urban hierarchy, 1900–2000. *City & Community 10*(1), 49–75.

Noyelle, T. J. and T. M. Stanback (1984). *The Economic Transformation of American Cities*. Totowa, NJ: Rowman & Allanheld Totowa, NJ.

Nzaku, K. and J. O. Bukenya (2005). Examining the relationship between quality of life amenities and economic development in the southeast USA. *Review of Urban & Regional Development Studies 17*(2), 89–103.

Olfert, M. R., M. Jelinski, D. Zikos, and J. Campbell (2012). Human capital drift up the urban hierarchy: Veterinarians in western Canada. *The Annals of Regional Science 49*(2), 551–570.

Overman, H. G. and Y. M. Ioannides (2001). Cross-sectional evolution of the U.S. city size distribution. *Journal of Urban Economics 49*(3), 543–566.

Partridge, M. D., D. S. Rickman, K. Ali, and M. R. Olfert (2008). Lost in space: Population growth in the American hinterlands and small cities. *Journal of Economic Geography 8*(6), 727–757.

Partridge, M. D., D. S. Rickman, K. Ali, and M. R. Olfert (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. *Journal of International Economics 78*(1), 126–140.

Polèse, M. and R. Shearmur (2004). Culture, language, and the location of high-order service functions: The case of Montreal and Toronto. *Economic Geography 80*(4), 329–350.

Ranaivoson, H. (2005). The economic analysis of product diversity. Technical report.

Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā: The Indian Journal of Statistics, Series A 44*(1), 1–22.

Stirling, A. (1998). On the economics and analysis of diversity. Electronic Working Papers Series 28, Science Policy Research Unit (SPRU).

Tabuchi, T. and J. F. Thisse (2011). A new economic geography model of central places. *Journal of Urban Economics 69*(2), 240–252.

US Census Bureau, C. H. S. (2013, January). Urban and rural areas - history. A description of how the Census Bureau defines the geographic terms urban and rural.

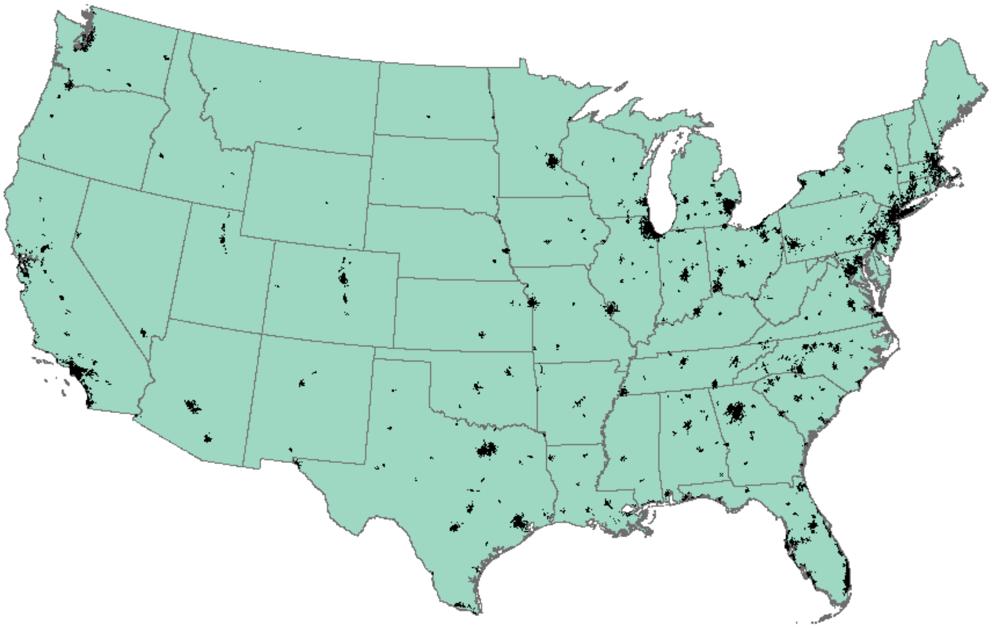
US Census Bureau, D. o. C. (2011, August). Urban area criteria for the 2010 census. *Federal Register Notice 76*(164), 53030–53043.

von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Nationalökonomie und Landwirtschaft.* Gustav Fischer, Stuttgart (reprinted 1966).

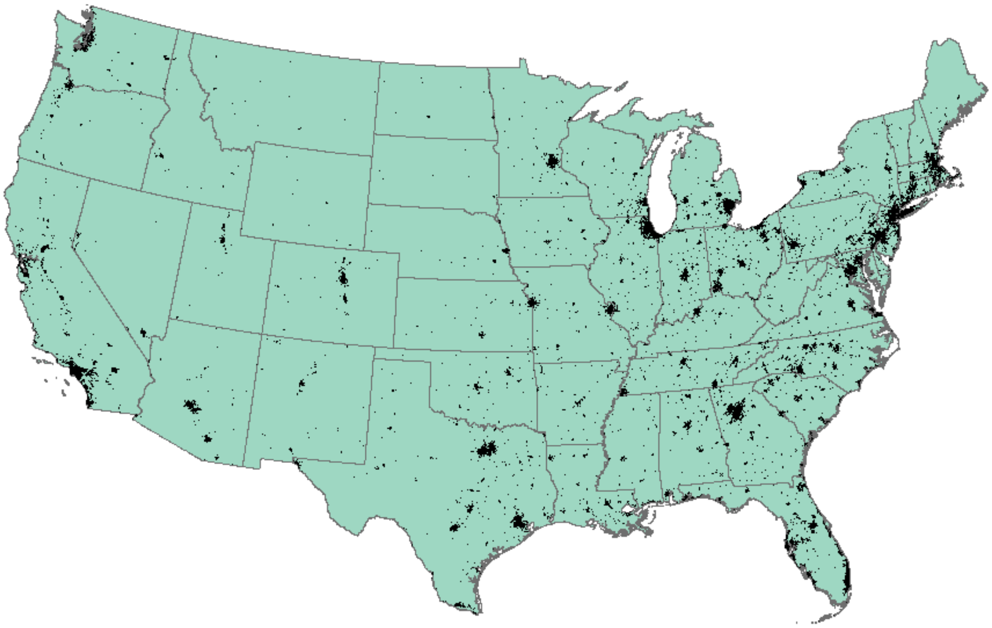
Wagner, J. E. (2000). Regional economic diversity: Action, concept, or state of confusion. *Journal of Regional Analysis and Policy 30*(2), 1–22.

Walls & Associates (2013). National establishment time-series (NETS) database: 2012 database description.

Wensley, M. R. D. and J. C. Stabler (1998). Demand-threshold estimation for business activities in rural saskatchewan. *Journal of Regional Science 38*(1), 155–177.

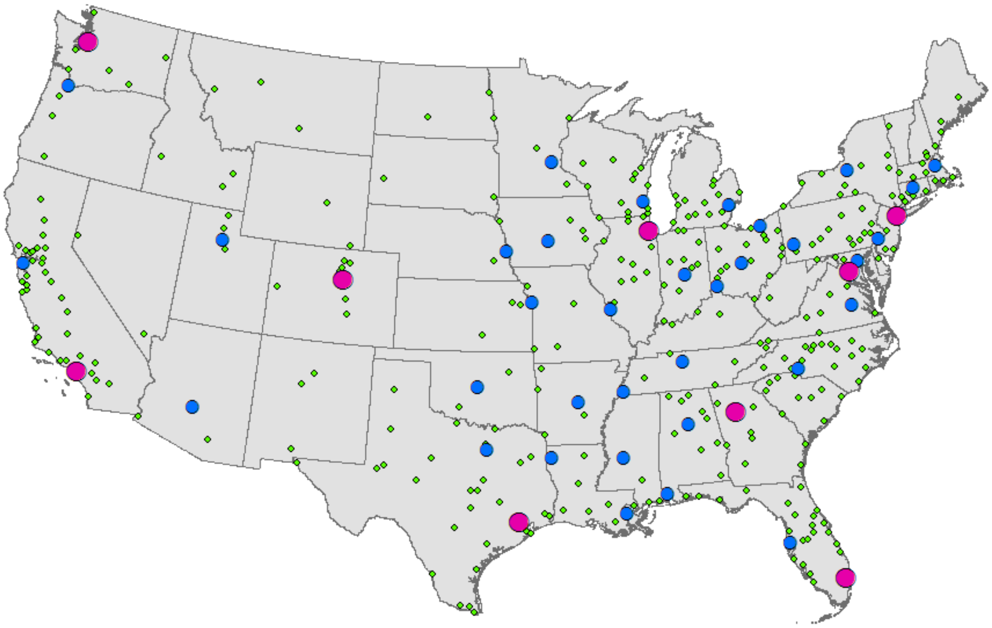


(a) MUZA dataset

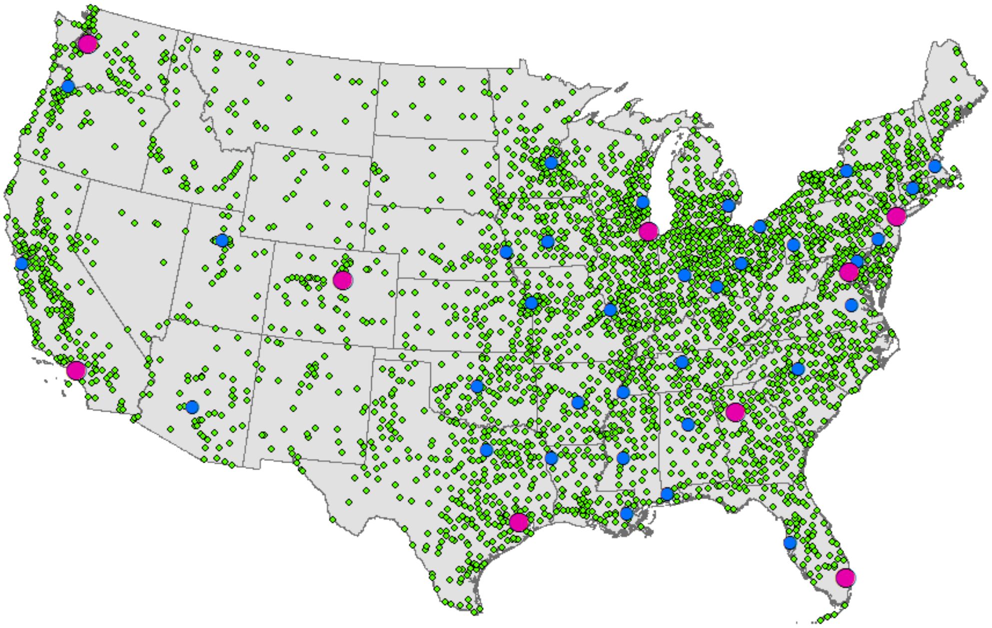


(b) MUA dataset

Figure 1: Maps of urban area in the MUZA and MUA datasets.

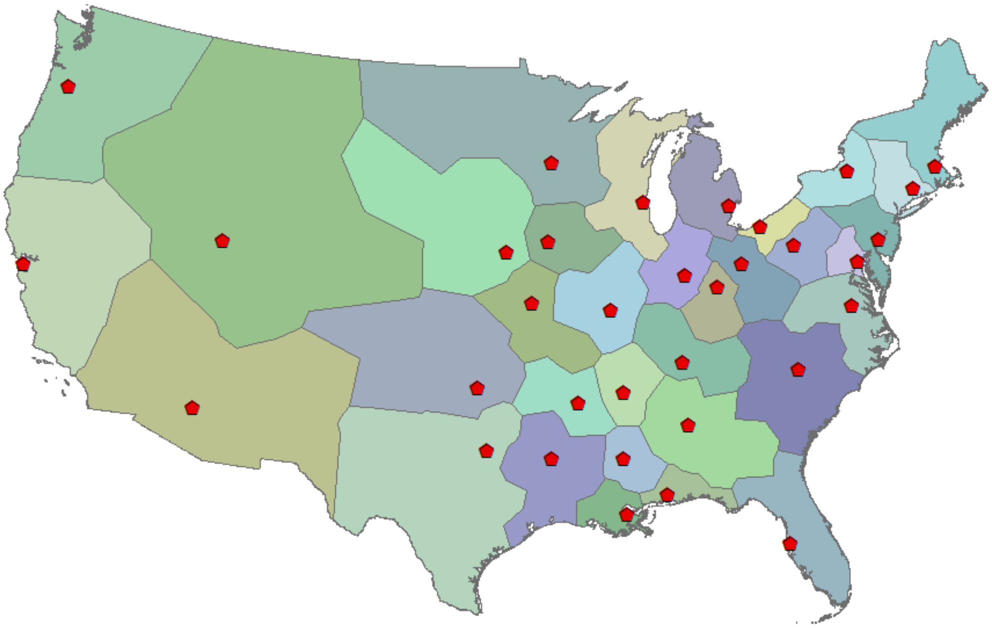


(a) MUZA dataset



(b) MUA dataset

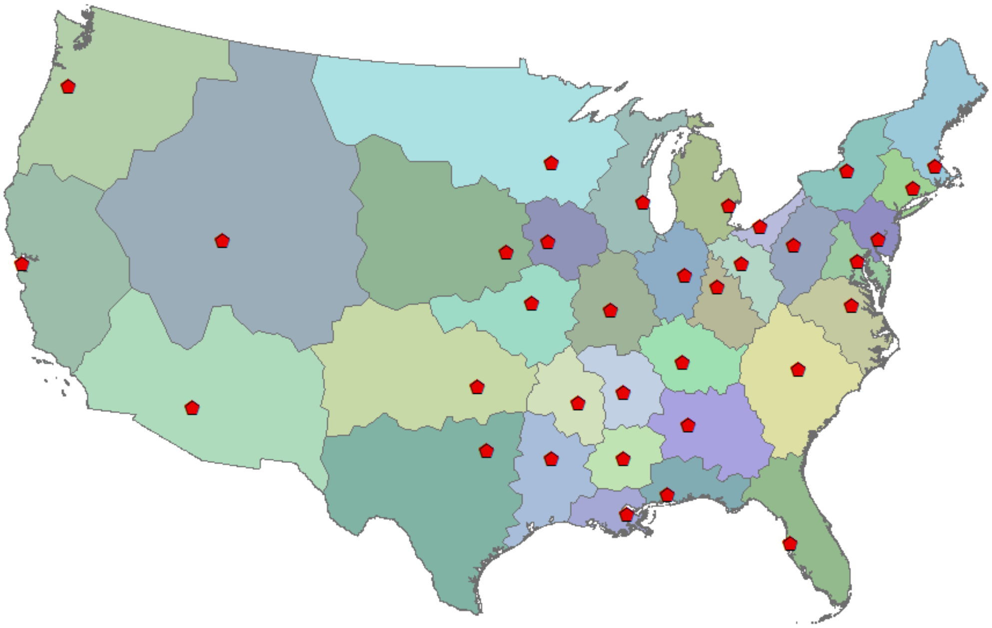
Figure 2: Maps of the assigned tier of each urbanized area in the MUZA and MUA datasets. Tier 1 cities are represented by large purple circles, tier 2/3 cities are represented by the medium blue circles, and tier 4 cities are represented by small green dots.



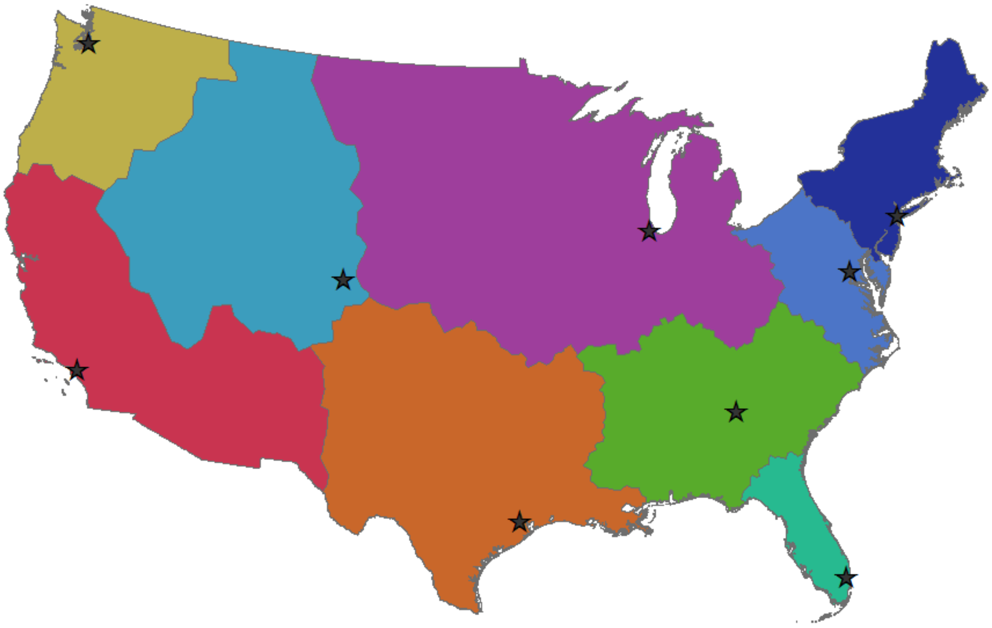
(a) Regional (tier 2/3) MUZA market areas

(b) Central Place (tier 1) MUZA market areas

Figure 3: Maps of the market areas for the MUZA dataset attributed to each tier node in the urban hierarchy. Regional nodes (e.g., Indianapolis, San Francisco) are represented by red pentagons, and central place nodes (e.g., New York City, Denver) are represented by black stars. Regional Market areas are attributed, in their entirety, to a central place market area.

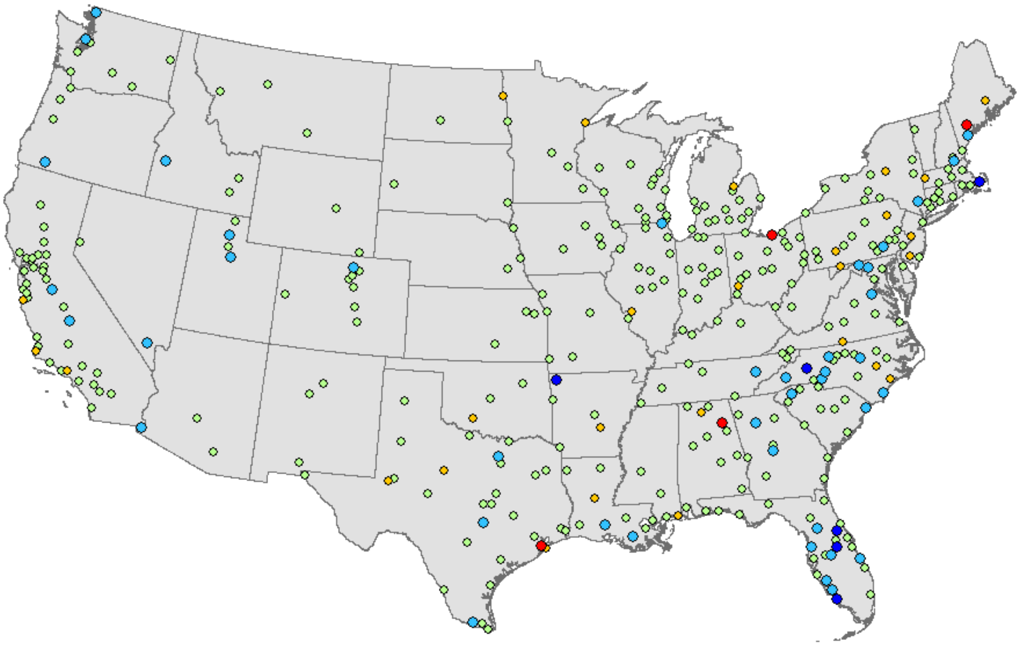


(a) Regional (tier 2/3) MUA market areas

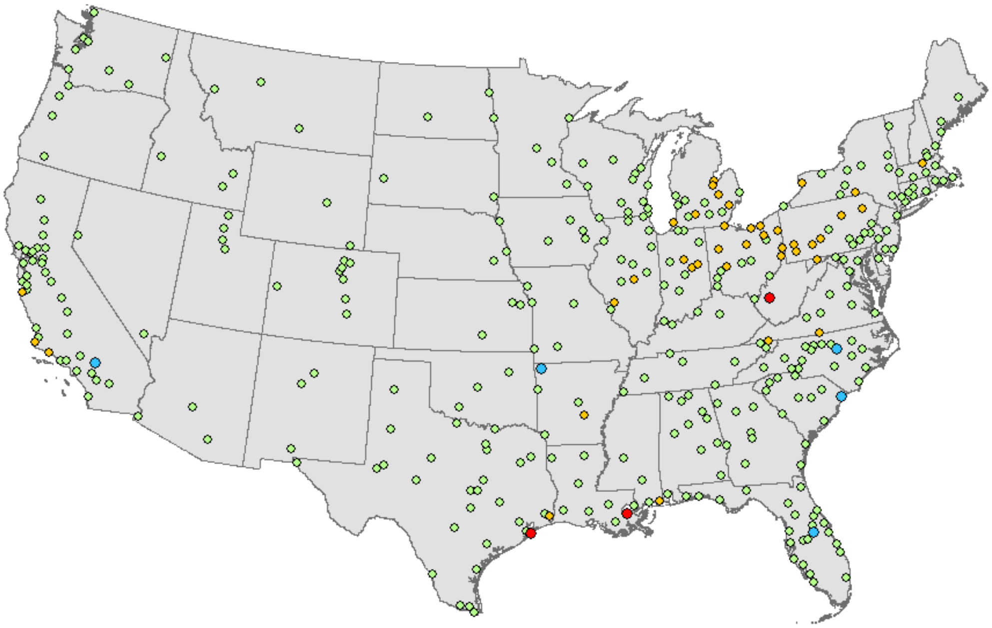


(b) Central Place (tier 1) MUA market areas

Figure 4: Maps of the market areas for the MUA dataset attributed to each tier node in the urban hierarchy. Regional nodes (e.g., Indianapolis, San Francisco) are represented by red pentagons, and central place nodes (e.g., New York City, Denver) are represented by black stars. Regional Market areas are attributed, in their entirety, to a central place market area.



(a) Population change in urbanized areas between 1990 and 2000



(b) Population change in urbanized areas between 2000 and 2010

Figure 5: Maps of the population change of urbanized areas in the MUZA dataset from 1990 to 2010. The same color scheme was used for both maps, where red indicates the largest negative population changes and dark blue indicates the largest positive population changes. Red and orange are negative population change, and green, light blue, and dark blue are positive population change.

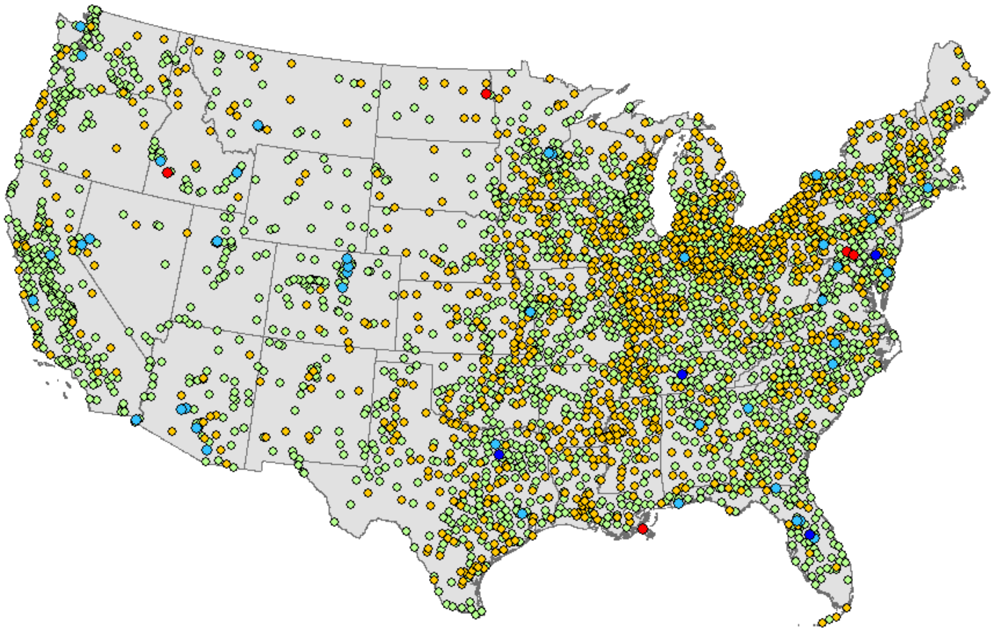


Figure 6: Population change of urban areas in the MUA dataset from 2000 to 2010. The same color scheme was used for both maps, where red indicates the largest negative population changes and dark blue indicates the largest positive population changes. Red and orange are negative population change, and green, light blue, and dark blue are positive population change.

Table 1: Overman and Ioannides (2001) and Dobkins and Ioannides (2001) Tier Classification

National Nodal (Tier 1)	Regional Nodal (Tier 2)	Subregional Nodal (Tier 3)
Atlanta	Baltimore	Birmingham
Chicago	Boston	Charlotte
Denver	Cincinnati	Des Moines
Houston	Cleveland	Detroit
Los Angeles	Columbus	Hartford
Miami	Dallas	Jackson, MS
New York	Indianapolis	Little Rock
San Francisco	Kansas City	Memphis
Seattle	Minneapolis	Milwaukee
Washington, DC	New Orleans	Mobile
	Philadelphia	Nashville
	Phoenix	Oklahoma City
	Portland	Omaha
	St. Louis	Pittsburgh
		Richmond
		Salt Lake City
		Shreveport
		Syracuse
		Tampa

Table 2: Number of Urban Areas by CP Market Area and Tier

(a) MUZA dataset

CP Market Area	Number of Urbanized Areas
Atlanta, GA	
Tier 1	1
Tier 2/3	6
Tier 4	54
Chicago, IL–IN	
Tier 1	1
Tier 2/3	10
Tier 4	76
Denver–Aurora, CO	
Tier 1	1
Tier 2/3	1
Tier 4	17
Houston, TX	
Tier 1	1
Tier 2/3	5
Tier 4	42
Los Angeles–Long Beach–Santa Ana, CA	
Tier 1	1
Tier 2/3	2
Tier 4	42
Miami, FL	
Tier 1	1
Tier 2/3	1
Tier 4	20
New York–Newark, NY–NJ–CT	
Tier 1	1
Tier 2/3	4
Tier 4	38
Seattle, WA	
Tier 1	1
Tier 2/3	1
Tier 4	10
Washington, DC–VA–MD	
Tier 1	1
Tier 2/3	4
Tier 4	25

(b) MUA dataset

CP Market Area	Number of Urban Areas
Atlanta, GA	
Tier 1	1
Tier 2/3	6
Tier 4	482
Chicago, IL–IN	
Tier 1	1
Tier 2/3	10
Tier 4	1017
Denver–Aurora, CO	
Tier 1	1
Tier 2/3	1
Tier 4	143
Houston, TX	
Tier 1	1
Tier 2/3	5
Tier 4	501
Los Angeles–Long Beach–Santa Ana, CA	
Tier 1	1
Tier 2/3	2
Tier 4	279
Miami, FL	
Tier 1	1
Tier 2/3	1
Tier 4	83
New York–Newark, NY–NJ–CT	
Tier 1	1
Tier 2/3	4
Tier 4	235
Seattle, WA	
Tier 1	1
Tier 2/3	1
Tier 4	140
Washington, DC–VA–MD	
Tier 1	1
Tier 2/3	4
Tier 4	251

Table 3: Descriptive Statistics for the MUZA dataset, 1990–2010

(a) Urbanized Areas

Variable	N	Mean	St. Dev.	Min	Max
Total Pop 1990	367	428,582.80	1,235,902.00	50,066.00	16,044,012.00
Total Pop 2000	367	515,174.70	1,397,374.00	50,902.00	17,832,182.00
Total Pop 2010	367	574,970.80	1,480,237.00	44,022.00	18,388,132.00
ln(Pop 1990)	367	12.008	1.112	10.821	16.591
ln(Pop 2000)	367	12.225	1.115	10.838	16.697
ln(Pop 2010)	367	12.347	1.128	10.692	16.727
Δ Pop 1990-2000	367	0.217	0.213	-0.285	1.295
Δ Pop 2000-2010	367	0.121	0.113	-0.218	0.520
Avg Temp (° C, 1990 area)	367	13.867	4.583	4.408	24.344
Avg Temp (° C, 2000 area)	367	13.852	4.583	4.385	24.307
Temperate Climate (1990 area)	367	-0.143	0.794	-5.524	8.424
Temperate Climate (2000 area)	367	-0.136	0.767	-4.768	7.928
Dist to GL/Ocean (km, 1990 area)	367	232.318	277.499	0.000	1,232.600
Dist to GL/Ocean (km, 2000 area)	367	232.576	277.507	0.000	1,237.063
Ruggedness (category, 1990 area)	367	1.924	1.428	1.000	10.000
Ruggedness (category, 2000 area)	367	1.926	1.315	1.000	10.000
Ruggedness (integer, 1990 area)	367	2.170	1.109	1.000	6.267
Ruggedness (integer, 2000 area)	367	2.168	1.072	1.000	6.057
Elev Diff (m, 1990 area)	367	170.698	195.273	2.000	1,388.000
Elev Diff (m, 2000 area)	367	176.185	186.649	4.000	1,340.000
Dist to Tier 2/3 (km)	358	212.814	190.868	0.000	914.290
SW Index (6-dig, 1990 area)	367	4.623	0.147	4.090	4.914
SW Index (6-dig, 2000 area)	367	4.782	0.112	4.325	5.000
SW Index (4-dig, 1990 area)	367	3.978	0.085	3.698	4.146
SW Index (4-dig, 2000 area)	367	4.061	0.063	3.795	4.185
SW Index (3-dig, 1990 area)	367	2.855	0.052	2.601	2.986
SW Index (3-dig, 2000 area)	367	2.868	0.057	2.570	3.034
Centrality Index (6-dig, 1990 area)	367	4.564	0.148	4.041	4.861
Centrality Index (6-dig, 2000 area)	367	4.730	0.113	4.275	4.953
Centrality Index (4-dig, 1990 area)	367	11.980	1.726	6.592	15.256
Centrality Index (4-dig, 2000 area)	367	13.215	1.365	8.038	15.676
Centrality Index (3-dig, 1990 area)	367	33.639	5.961	14.321	44.128
Centrality Index (3-dig, 2000 area)	367	36.998	4.510	19.380	44.204

Table 3: Descriptive Statistics for the MUZA dataset, 1990–2010 –Continued–

(b) Regional Market Areas

Variable	N	Mean	St. Dev.	Min	Max
Real Agg Income 1989 (millions)	34	105,769.200	117,978.500	7,692.602	585,626.800
Real Agg Income 1999 (millions)	34	288,323.900	365,797.100	29,117.770	1,564,246.000
Dist to Tier 1 (km)	34	448.919	191.687	67.810	780.711
Ag Employment Share 1990	34	0.003	0.002	0.001	0.009
Mfg Employment Share 1990	34	0.157	0.042	0.076	0.248
Ag Employment Share 2000	34	0.003	0.002	0.001	0.008
Mfg Employment Share 2000	34	0.118	0.037	0.054	0.210
Urban Area 1990 (km^2)	34	4,671.188	3,286.176	703.259	12,984.420
Rural Area 1990 (km^2)	34	224,092.700	265,136.700	26,276.360	1,341,098.000
Urban Area 2000 (km^2)	34	5,435.659	4,317.279	587.548	17,064.660
Rural Area 2000 (km^2)	34	223,328.200	265,163.100	25,009.990	1,340,541.000
Total Area (km^2)	34	228,763.900	265,737.600	31,132.750	1,345,350.000
Proportion of Rural Area 1990	34	0.962	0.041	0.812	0.997
Proportion of Rural Area 2000	34	0.955	0.052	0.769	0.997

(c) Central Place Market Areas

Variable	N	Mean	St. Dev.	Min	Max
Urban Area 1990 (km^2)	9	17,646.710	9,894.266	4,251.798	34,050.600
Rural Area 1990 (km^2)	9	846,572.500	599,604.700	141,397.100	1,852,483.000
Urban Area 2000 (km^2)	9	20,534.710	11,612.680	4,808.316	37,156.800
Rural Area 2000 (km^2)	9	843,684.500	600,406.400	138,990.800	1,849,377.000
Total Area (km^2)	9	864,219.200	603,562.100	152,591.300	1,886,533.000
Proportion of Rural Area 1990	9	0.967	0.027	0.927	0.997
Proportion of Rural Area 2000	9	0.961	0.034	0.905	0.996

Table 4: Descriptive Statistics for Urbanized Areas (MUZA dataset) by Tier, 1990–2010

Variable	Mean	St. Dev.	Min	Max
Tier 1				
Total Pop 1990	5,658,645.00	5,007,423.00	1,517,977.00	16,044,012.00
Total Pop 2000	6,642,846.00	5,315,725.00	2,010,212.00	17,832,182.00
Total Pop 2010	7,236,398.00	5,266,982.00	2,374,203.00	18,388,132.00
Δ Pop 1990-2000	0.222	0.125	0.093	0.496
Δ Pop 2000-2010	0.127	0.085	0.031	0.253
Avg Temp (° C, 2000 area)	15.019	5.176	9.889	24.307
Temperate Climate (2000 area)	−0.559	1.408	−4.285	0.313
Dist to GL/Ocean (km, 2000 area)	186.109	394.066	0.276	1,193.023
Ruggedness (integer, 2000 area)	2.201	0.876	1.001	3.047
Elev Diff (m, 2000 area)	419.111	435.460	8.000	1,340.00
SW Index (4-dig, 2000 area)	4.119	0.034	4.032	4.147
Centrality Index (4-dig, 2000 area)	15.317	0.282	14.942	15.611
Tier 2/3				
Total Pop 1990	1,396,655.00	1,137,382.00	256,489.00	4,671,827.00
Total Pop 2000	1,626,009.00	1,301,799.00	275,213.00	5,190,255.00
Total Pop 2010	1,775,598.00	1,393,913.00	298,317	5,441,567
Δ Pop 1990-2000	0.156	0.117	0.034	0.599
Δ Pop 2000-2010	0.099	0.102	−0.183	0.409
Avg Temp (° C, 2000 area)	13.853	4.071	7.586	22.603
Temperate Climate (2000 area)	−0.186	0.515	−2.741	0.847
Dist to GL/Ocean (km, 2000 area)	254.496	253.067	2.069	928.005
Ruggedness (integer, 2000 area)	2.269	0.967	1.000	5.536
Elev Diff (m, 2000 area)	209.941	162.260	10.000	718.000
Dist to Tier 2/3 (km)	0.000	0.000	0.000	0.000
SW Index (4-dig, 2000 area)	4.121	0.029	4.051	4.185
Centrality Index (4-dig, 2000 area)	14.792	0.459	13.784	15.676
Tier 4				
Total Pop 1990	181,715.500	235,888.100	50,066	2,348,417
Total Pop 2000	228,392.500	284,561.400	50,902	2,681,237
Total Pop 2010	263,939.400	334,629.200	44,022	2,956,746
Δ Pop 1990-2000	0.224	0.222	−0.285	1.295
Δ Pop 2000-2010	0.124	0.115	−0.218	0.520
Avg Temp (° C, 2000 area)	13.819	4.627	4.385	23.624
Temperate Climate (2000 area)	−0.119	0.765	−4.768	7.928
Dist to GL/Ocean (km, 2000 area)	231.566	277.059	0.000	1,237.063
Ruggedness (integer, 2000 area)	2.156	1.089	1.000	6.057
Elev Diff (m, 2000 area)	165.895	173.835	4.000	1,149
Dist to Tier 2/3 (km)	235.146	187.077	34.692	914.290
SW Index (4-dig, 2000 area)	4.053	0.062	3.795	4.172
Centrality Index (4-dig, 2000 area)	12.991	1.286	8.038	15.525

Table 5: Descriptive Statistics for Urbanized Areas (MUZA dataset) by Central Place Market Area, 1990–2010

Variables	Atlanta	Chicago	Denver	Houston	Los Angeles
Total Pop 1990	203,351.40	369,393.40	220,654.60	320,698.60	672,130.10
Total Pop 2000	276,421.50	422,069.90	293,505.90	407,817.10	806,920.80
Total Pop 2010	338,370.30	448,501.70	353,189.30	491,377.50	907,278.80
Δ Pop 1990-2000	0.245	0.149	0.271	0.195	0.252
Δ Pop 2000-2010	0.161	0.073	0.178	0.131	0.147
Avg Temp (° C, 2000 area)	16.473	9.830	9.260	18.779	16.502
Temperate Climate (2000 area)	−0.280	−0.092	−0.086	−0.400	−0.213
Dist to GL/Ocean (km, 2000 area)	259.362	239.111	1,015.25	331.503	138.084
Ruggedness (integer, 2000 area)	2.462	1.905	2.222	1.457	2.307
Elev Diff (m, 2000 area)	138.082	91.230	388.368	75.667	368.622
Dist to Tier 2/3 (km)	170.532	157.696	542.040	298.492	307.443
SW Index (4-dig, 2000 area)	4.027	4.071	4.117	4.055	4.053
Centrality Index (4-dig, 2000 area)	12.995	12.884	13.900	13.201	13.705

	Miami	New York	Seattle	Washington, DC
Total Pop 1990	443,542.90	802,524.70	386,233.10	467,170.00
Total Pop 2000	583,936.20	917,549.00	505,718.10	521,807.40
Total Pop 2010	695,976.90	955,948.50	581,755.90	561,048.30
Δ Pop 1990-2000	0.430	0.195	0.316	0.146
Δ Pop 2000-2010	0.231	0.068	0.150	0.077
Avg Temp (° C, 2000 area)	22.176	9.727	10.971	12.192
Temperate Climate (2000 area)	0.652	−0.090	−0.099	−0.122
Dist to GL/Ocean (km, 2000 area)	23.707	61.225	99.751	98.818
Ruggedness (integer, 2000 area)	1.163	2.601	2.682	3.137
Elev Diff (m, 2000 area)	32.182	234.070	312.500	205.933
Dist to Tier 2/3 (km)	192.145	113.012	254.344	120.725
SW Index (4-dig, 2000 area)	4.060	4.080	4.112	4.041
Centrality Index (4-dig, 2000 area)	14.025	13.230	14.343	12.411

Table 6: Descriptive Statistics for the MUA dataset, 2000–2010

(a) Urban Areas

Variable	N	Mean	St. Dev.	Min	Max
Total Pop 2000	3,174	69,358.93	493,711.40	2,501	17,832,182
Total Pop 2010	3,174	77,589.23	525,599.10	2,503	18,388,132
ln(Pop 2000)	3,174	9.313	1.366	7.824	16.697
ln(Pop 2010)	3,174	9.396	1.398	7.825	16.727
Δ Pop 2000-2010	3,174	0.083	0.200	-1.255	2.072
Avg Temp (° C, 2000 area)	3,174	13.004	4.457	1.589	25.442
Temperate Climate (2000 area)	3,174	-0.111	1.430	-16.521	54.734
Dist to GL/Ocean (km, 2000 area)	3,174	308.871	281.172	0.000	1,275.840
Ruggedness (category, 2000 area)	3,174	2.143	1.631	1.000	10.000
Ruggedness (integer, 2000 area)	3,174	2.213	1.296	1.000	8.109
Elev Diff (m, 2000 area)	3,174	86.206	113.442	0.000	1,340
Dist to Tier 2/3 (km)	3,165	240.965	178.843	0.000	1,224.910
SW Index (6-dig, 2000 area)	3,174	4.121	0.506	0.000	5.000
SW Index (4-dig, 2000 area)	3,174	3.695	0.344	0.000	4.185
SW Index (3-dig, 2000 area)	3,174	2.809	0.173	0.000	3.095
Centrality Index (6-dig, 2000 area)	3,174	4.070	0.503	0.000	4.953
Centrality Index (4-dig, 2000 area)	3,174	7.456	2.998	0.000	15.684
Centrality Index (3-dig, 2000 area)	3,174	17.092	10.016	0.000	44.355

(b) Regional Market Areas

Variable	N	Mean	St. Dev.	Min	Max
Real Agg Income 1999 (millions)	34	283,590.100	307,971.100	35,107.440	1,229,517.000
Dist to Tier 1 (km)	34	448.921	192.001	67.810	780.711
Ag Employment Share 2000	34	0.005	0.003	0.002	0.013
Mfg Employment Share 2000	34	0.126	0.036	0.070	0.210
Urban Area 2000 (km^2)	34	7,014.042	4,642.909	1,341.056	17,837.430
Rural Area 2000 (km^2)	34	221,749.900	248,122.500	25,605.740	1,196,363.000
Total Area (km^2)	34	228,763.900	248,975.600	30,436.380	1,202,999.000
Proportion of Rural Area 2000	34	0.942	0.064	0.662	0.996

(c) Central Place Market Areas

Variable	N	Mean	St. Dev.	Min	Max
Urban Area 2000 (km^2)	9	26,493.100	14,813.160	6,636.161	50,781.810
Rural Area 2000 (km^2)	9	837,726.100	616,906.900	132,008.100	2,053,201.000
Total Area (km^2)	9	864,219.200	623,509.500	147,753.500	2,103,983.000
Proportion of Rural Area 2000	9	0.952	0.040	0.888	0.994

Table 7: Descriptive Statistics for Urban Areas (MUA dataset) by Tier, 2000–2010

Variable	Mean	St. Dev.	Min	Max
Tier 1				
Total Pop 2000	6,549,324.00	5,231,462.00	2,010,212.00	17,832,182.00
Total Pop 2010	7,125,998.00	5,172,898.00	2,374,203.00	18,388,132.00
ln(Pop 2000)	15.454	0.712	14.514	16.697
ln(Pop 2010)	15.580	0.651	14.680	16.727
Δ Pop 2000-2010	0.126	0.086	0.030	0.253
Avg Temp (° C, 2000 area)	15.029	5.176	9.889	24.307
Temperate Climate (2000 area)	−0.559	1.407	−4.282	0.313
Dist to GL/Ocean (km, 2000 area)	186.136	394.058	0.276	1,193.020
Ruggedness (integer, 2000 area)	2.174	0.853	1.001	3.090
Elev Diff (m, 2000 area)	420.889	434.964	8.000	1,340
SW Index (4-dig, 2000 area)	4.119	0.034	4.032	4.147
Centrality Index (4-dig, 2000 area)	15.318	0.277	14.949	15.618
Tier 2/3				
Total Pop 2000	1,597,923.00	1,265,224.00	275,213	5,190,255
Total Pop 2010	1,743,944.00	1,358,097.00	298,317	5,441,567
ln(Pop 2000)	13.984	0.809	12.525	15.462
ln(Pop 2010)	14.082	0.797	12.606	15.510
Δ Pop 2000-2010	0.098	0.102	−0.183	0.409
Avg Temp (° C, 2000 area)	13.848	4.070	7.586	22.602
Temperate Climate (2000 area)	−0.186	0.516	−2.748	0.847
Dist to GL/Ocean (km, 2000 area)	254.465	253.362	0.000	928.005
Ruggedness (integer, 2000 area)	2.266	0.962	1.000	5.537
Elev Diff (m, 2000 area)	206.882	157.723	10.000	718.000
Dist to Tier 2/3 (km)	0.000	0.000	0.000	0.000
SW Index (4-dig, 2000 area)	4.121	0.029	4.051	4.185
Centrality Index (4-dig, 2000 area)	14.795	0.458	13.785	15.684
Tier 4				
Total Pop 2000	34,133.49	113,630.60	2,501.00	2,681,237.00
Total Pop 2010	39,233.53	133,138.90	2,503.00	2,956,746.00
ln(Pop 2000)	9.244	1.238	7.824	14.802
ln(Pop 2010)	9.327	1.273	7.825	14.900
Δ Pop 2000-2010	0.083	0.201	−1.255	2.072
Avg Temp (° C, 2000 area)	12.989	4.459	1.589	25.442
Temperate Climate (2000 area)	−0.109	1.437	−16.521	54.734
Dist to GL/Ocean (km, 2000 area)	309.815	281.058	0.000	1,275.840
Ruggedness (integer, 2000 area)	2.212	1.300	1.000	8.109
Elev Diff (m, 2000 area)	83.934	108.695	0.000	1,231
Dist to Tier 2/3 (km)	243.581	178.030	23.808	1,224.910
SW Index (4-dig, 2000 area)	3.689	0.342	0.000	4.172
Centrality Index (4-dig, 2000 area)	7.353	2.887	0.000	15.575

Table 8: Descriptive Statistics for Urban Areas (MUA dataset) by Central Place Market Area, 2000–2010

Variables	Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	Seattle	Washington, DC
Total Pop 2000	44,030.22	44,369.11	47,103.77	47,555.79	141,783.00	162,544.60	174,467.90	54,620.87	72,791.61
Total Pop 2010	52,924.41	47,132.87	57,513.94	56,547.76	161,503.90	195,052.30	181,935.00	63,250.12	78,737.93
Δ Pop 2000-2010	0.084	0.049	0.205	0.058	0.156	0.205	0.050	0.153	0.072
Avg Temp (° C, 2000 area)	16.235	9.941	8.222	17.871	15.871	21.598	8.652	10.422	11.701
Temperate Climate (2000 area)	−0.254	−0.094	−0.085	−0.143	−0.083	0.452	−0.082	−0.095	−0.116
Dist to GL/Ocean (km, 2000 area)	313.770	333.103	936.396	418.049	218.069	46.927	90.261	164.627	102.508
Ruggedness (integer, 2000 area)	2.410	1.912	2.638	1.591	2.527	1.167	3.055	2.709	2.967
Elev Diff (m, 2000 area)	72.928	47.322	190.786	41.854	193.181	22.141	153.450	143.873	104.723
Dist to Tier 2/3 (km)	194.699	198.931	531.323	291.934	329.724	256.531	167.106	334.407	148.307
SW Index (4-dig, 2000 area)	3.664	3.711	3.797	3.656	3.614	3.732	3.776	3.808	3.641
Centrality Index (4-dig, 2000 area)	7.352	7.157	8.188	7.129	8.060	8.759	7.993	8.323	7.002