MEASURING SPRAWL IN THE UNITED STATES—A COMPARATIVE ANALYSIS OF PROCEDURES AND RESULTS

by

YUE WU

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Planning and Public Policy

written under the direction of

Robert W. Burchell

and approved by

____

____

____

____

New Brunswick, New Jersey

May, 2013

ABSTRACT OF THE DISSERTATION

MEASURING SPRAWL IN THE UNITED STATES—A COMPARATIVE ANALYSIS OF PROCEDURES AND RESULTS

By YUE WU

Dissertation Director:
Robert W. Burchell

Sprawl is significant, low-density development taking place at the periphery wherein there exists limited infrastructure and public services. It has been the subject of much research, due to its widespread occurrence. Previous empirical studies of sprawl measurement had three significant weaknesses: 1) incomplete coverage of the geography being measured; 2) an absence of regional differences in the density variable, and 3) a lack of exclusion of undevelopable lands when calculating the density of an area. In an effort to overcome these shortcomings, this study: 1) measures sprawl for all 3091 counties in the US using economic areas (EA) to group counties; 2) uses variable densities (locally-determined cultural densities) in the sprawl calculation; and 3) calculates "refined densities" for all states and counties in the US by excluding undevelopable lands from the density calculation. Further, the research results included here are compared with those of Rolf Pendall (1999), Robert W. Burchell (2002), and Reid Ewing (2003b), in order to ascertain the impact of their more comprehensive measurement methods on sprawl measurement results.

Based upon accurate variable densities, one of six county land use types (urban center, urban, suburban, rural center, rural, undeveloped) is assigned to each US county; a sprawl/non-sprawl is then determined for each county. As a result, out of all 3091 US counties, 492 experience sprawl development during the 2000 to 2020 time period. Over 80 percent (or 396 counties) of these 492 sprawling counties are rural or undeveloped counties; the remaining one-fifth are developing suburban and rural center counties. With no exception, the “refined density” of an individual state is greater than its “gross density.” About 65 percent (or 32) of US states have “refined densities” that are at least 1.2 times their original gross densities.

Several conclusions can be drawn using the comparative analyses included here. First, sprawl research must focus on the nation as a whole, or on select component regions. Second, variable density is crucial to the accurate calculation urban versus rural counties nationally. Third, developable land must be employed when calculating variable density (i.e. undevelopable lands must be excluded from the density calculation).

Acknowledgement

First and foremost, I would like to express my deepest gratitude and appreciation to my advisor, Prof. Robert W. Burchell, for the guidance, support, and immense knowledge he has provided me, from when I first began my studies at The Bloustein School, through remote work on my thesis, to completion of my degree. I deeply appreciate his patience when my progress was slowed by the pressures of parenting, and I am ever in his debt for his advocacy during difficult times. I am also deeply grateful to him for carefully reading and revising my thesis. This dissertation would not have been possible without his excellent advice and counsel.

I am extremely grateful to my committee members, Prof. Michael J. Camasso Jr., Prof. Radha Jagannathan, and Prof. David C. Listokin, for their encouragement, their insightful comments, and their constructive criticisms. It is an honor and pleasure to work with Professors Jagannathan and Camasso, from whom I learned much, not the least of which is having both good data management skills and a good work ethic. Special thanks go to Professors David C. Listokin and Robert W. Lake, the former and current directors, respectively, of the doctoral program, for their patience and facilitation throughout the course of my study. I also would like to thank Professors Richard Brail and Reid Ewing for their astute comments on my thesis proposal.

I would like to express my sincerest gratitude to Paul Suino for his excellent job of editing multiple drafts of my thesis. Paul was there whenever I needed his assistance, always fine-tuning my work with quality, and in a timely fashion. I am also deeply grateful to him for his warm encouragement and support.

I would like to acknowledge the financial and academic backing of the Edward J. Bloustein School of Planning and Public Policy, making my graduate study and thesis research possible.

I am indebted to William R. Dolphin for his valuable help in data collection. I would also like to acknowledge the professionals in the United States Department of Agriculture (USDA) Forest Service who generously provided me with the GIS data "Land Ownership Boundary within Federal Forests," information which added great value to my project.

My sincere thanks also go to Lynn Astorga for her enthusiasm and support during my graduate study. Lynn was always there whenever I needed her help. I am also thankful to Barbara Sirman and Amy Cobb whose kind aid significantly improved my process of graduation.

I am also grateful to my friends for standing by me throughout all these years. Among them, I would like to especially acknowledge Yan Shen, Xuejin Ruan, Shuling Tang, Xuanxuan Xu, and Liang Wei, whose unending words of optimism have helped me overcome any number of setbacks to stay focused on my thesis. I am also thankful to my doctoral classmates, Matt, Heather, Pam, John, Sandra, and Jesse, for their spirited and valuable comments.

Most importantly, none of this would have been possible without the love and inspiration of my family. I thank my parents Yutang Wu and Ziyi Han for their belief in me, for their affection, and for their hand, which they have extended throughout my life. I thank my two sons, Eric and Evan, who have brought so much joy to my life. Finally, I want to give special thanks to my husband, Jianye Chen, for his constant encouragement, his valuable help with this project, and his great patience at all times.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Meaning and Importance of this Dissertation: Using Accurate Local Density Definitions to Specify and Locate Sprawl Development

The dissertation which follows is particularly important because it attempts to measure sprawl in the United States using accurate and locally-determined cultural densities. What is meant by this is that location within the United States determines a hierarchy of densities wherein each locality identifies with "place" in terms of the local use of that place. Urban is defined as urban at a different density in low density states than it is in high density states. In both places they function as urban areas. As an example, both urban and rural are one-fifth the densities in North and South Dakota than they are in New Jersey and Rhode Island. What one means is that North and South Dakota have urban areas that function as intense places at much less density than do New Jersey and Rhode Island. Also, New Jersey and Rhode Island have very rural places that function as such at much greater densities than do North/South Dakota.

It is critical to include cultural density (labeled as variable density in this dissertation) as a key measurement of sprawl. Thus, sprawl growth would be defined as: "significant growth taking place in rural areas where the infrastructure is not prepared to take such growth." This would be true in both North/South Dakota and New Jersey/Rhode Island where different densities define these places similarly and they function as such. This is important because one of the leading proponents of anti-sprawl land use strategies (Reid Ewing) defines low density (rural) as less than 1,500 per square mile and high density (urban) as greater than 12,500 people per square mile. He uses this definition as a constant across all locations of the United States. This was presented just recently at the Rocky Mountain Land Use Institute as the best and most current way to define sprawl (Ewing, 2013). In Robert W. Burchell's 2002 classification (Burchell et.al., 2002), urban for high density states was defined as greater than 10,500 people per square mile and rural as less than 875 people per square mile (at an average of 3.5 persons per household). For low density states, Burchell's classifications were greater than 2,100 people per square mile for urban and less than 263 people per square mile for rural counties. This variable density measurement reflective of the culture of an area (as measured in density levels by type of place) allows sprawl to be measured in every county of the United States relative to the way those places function.

This brings up an additional point. One must measure sprawl using EAs (Economic Areas- 172 in U.S.) as opposed to MSAs (Metropolitan Statistical Areas – 400 in U.S.) because the former aggregation attaches rural counties to more urban counties and thereby accounts for all counties in the United States. By using all counties, one is able to identify sprawl in the most rural counties in the United States. This is important because this measurement does not involve just a trickle of development in these locations; it is upper-quartile absolute growth or multiples of average state growth in these areas. Thus, completeness, i.e. measuring the entire country, is important in the discovery of sprawl.

One final point involves the actual measurement of sprawl in the calculation of host-state density, and as well, the densities of urban, suburban, and rural counties. These densities must be calculated using “developable area density”. This means that the area measured for development purposes can be actually converted to residential and nonresidential land uses. One must net out undevelopable water bodies, mountain ranges or severe steep slope areas, deserts or barren soil areas, parklands or established trails, and protected forests or nature preserves. These currently all appear on GIS layers and must be eliminated before developable density is calculated.

The above are the points made in this dissertation. The dissertation calls for a clear and simple definition of sprawl that is applicable in all areas and is calculated precisely. This definition of sprawl is: “significant development taking place in areas which do not have the infrastructure to support this development. These are developing-suburban, developing-rural center, rural, or undeveloped counties in the United States as defined by their local or cultural densities.”

1.2 Research Background

Definition of Sprawl

Sprawl has been the subject of much research, due to its costly nature and widespread occurrence nationwide. Most scholars agree that sprawl is a type of growth pattern that expends more overall resources than the alternative compact development pattern (Burchell et al. 1999; Burchell et al. 2002; Ewing 1997). However, in spite of the large volume of literature on sprawl, no consensus has been reached on a universal definition of sprawl (Ewing 2003a; Burchell et al. 1999, 2002; Galster et al. 2001).

This research has employed the theoretical and empirical definitions of sprawl that were developed and presented by Burchell et al. in 2002. Theoretically, sprawl is defined as “low-density, leapfrog development that is characterized by unlimited outward extension[1].” (Burchell et al. 2002, p. 2) Empirically, sprawl is defined as rapid and significant growth in rural and undeveloped counties, i.e. significant development in a location that might not be able to accommodate this growth (Burchell et al. 2002, p. 3). By definition, urban and urban center counties are not categorically involved in this definition of sprawl. However, some developing suburban and rural center counties are indeed included in the definition of sprawl, in order to “ensure that sprawl has not been overlooked in relatively developed places.” (Burchell et al. 2002, p. 3) The detailed operational definition of sprawl will be explained in Chapter 3 Data and Methods.

Sprawl is more costly than the alternative compact growth scenario. It is the rapid and significant growth occurring in both rural and undeveloped counties. In other words, this significant spread of uncontrolled development happens in locations that do not have existing or sufficient infrastructure and public services components to support it. New infrastructure such as highways, water and sewer systems, power lines, municipal buildings, and public schools must be built to support this rapid or considerable growth.

It would be highly beneficial if one were able to identify where the sprawl growth will happen. If policymakers had the necessary information at hand in order to foresee which counties (and communities) would be experiencing sprawl growth in the future, then they could more readily and carefully plan ahead and control that growth by

---

[1] “In other words, sprawl is significant residential or nonresidential development in a relatively pristine setting. In nearly every instance, this development is low-density, it has leapt over other development to become established in an outlying area, and its very location indicates that it is unbounded.” (Burchell et al. 2002 p. 2)

redirecting the projected sprawl growth from a particular county to one or more locations that were already developed within the EA (Burchell et al. 2002). The existing infrastructure in these more developed locations could partially or even fully contain the redirected future growth, thus saving the unnecessary costs for building new public infrastructure in less developed, sprawl locations.

Measuring Sprawl at County Level

Sprawl has been measured for different geographic units, including metropolitan areas (USA Today 2001; Sierra Club 1998; Fulton et al. 2001; Ewing 2002, 2003a), Extended Urbanized Areas (EUA) (Cutingser et al. 2005), Urbanized Areas (UA) defined by the Census Bureau (Downs 1998; Galster et al. 2001; Glaeser et al. 2001), and counties (Pendall 1999, Burchell et al. 2002, Ewing et al. 2003b).

This study selects counties as the geographic units of analysis, largely due to the following three reasons. First, as this study aims at measuring sprawl nationwide, the units of analysis must consistently cover the country as whole in order to make the comparisons possible, consistent, and accurate among the units. The county as a measurable geographic unit satisfies this requirement perfectly (Burchell et al. 2002). Conversely, none of the metropolitan area, UA, EUA, or places (towns, boroughs, and cities) measurements is able to satisfy this requirement because they only cover a part of the nation and therefore comparisons would be incomplete. Second, the counties or county equivalents are the smallest nationwide geographic units that offer non-educational local government, including land-use and political power (Burchell et al. 2002). This distinction is important because any study findings could significantly affect policy. For example, a county government could develop and implement specific anti-sprawl policies according to the degree of sprawl growth in their particular county. Finally, a variety of historical and projection data is available at county level (Burchell et al. 2002). For most (if not all) of the national socioeconomic projection database, the county is the smallest geographic level because any projection at a sub-county level would be at a risk of incompleteness as well as much bigger projection error than that at the county level. Therefore, the county as a measurable geographic unit is selected as the unit of analysis in this study.

Four Characteristics of the Existing Literature on Measuring Sprawl

Despite the popularity of sprawl in the literature, empirical measurements of sprawl are far from fully developed and could still be improved in many respects. There are four notable characteristics for the existing literature that empirically measures sprawl. First, most of these sprawl measurements are limited to metropolitan areas; second, as noted by many scholars, one important measuring variable of sprawl—residential density--- has not been accurately measured in most studies (Ewing 2002, 2003a; Galster et al. 2001; Cutsinger et al. 2005); third, regional differences in the studied geographic units are somewhat neglected or altogether overlooked in sprawl density measurements (Burchell et al. 2002). Finally, the analyses comparing research results of existing studies on sprawl measurements have not yet been done. Such comparison analyses might shed some light on the possible improvements of sprawl measurement methods and enable further refinement in future studies. These four as-yet-to-be-improved aspects of the current literature on sprawl measurements will be explained below in detail.

Many studies attempt to empirically measure sprawl (Sierra Club 1998; Downs 1998; Pendall 1999; USA Today 2001; Fulton et al. 2001; Galster et al. 2001; Glaeser et al. 2001; Burchell et al. 2002; Ewing 2002, 2003a, 2003b; Cutsinger et al. 2005). Among them, a notable feature is a common focus on large metropolitan areas and the resulting small sample of areas to study (with the only exception being the work of Burchell et al. in 2002). This feature is mainly due to the lack of data, especially if the sprawl measurements are multi-dimensional or if undevelopable lands are excluded from the density calculation. Almost all of these studies limit their sprawl measurements to the most populous metropolitan areas. In these studies, the geographic units for which the sprawl is measured are either the metropolitan area as a whole or some sub-level geographic units within metropolitan area boundaries, such as the “Urbanized Areas” defined by the Census, or the “Extended Urbanized Areas” defined by particular researchers, or counties.

Due to a propensity to focus on the most populous metropolitan areas, these empirical studies usually have small and incomplete samples to consider. Among the studies measuring sprawl within metropolitan areas[2], the largest sample to date is the sample conducted by Ewing et al. (2003b) which includes the 448 counties that make up the 101 most populous metropolitan areas. However, even this “more inclusive” sample covers only 14 percent of the 3141 counties (and county equivalents) in the United States.

These empirical studies assume that sprawl developments mainly take place in the most populous metropolitan areas and thus a study focusing on these areas could capture most of the total sprawl developments (Ewing et al. 2002, 2003a, 2003b). However, what

[2] Burchell et al.’s work (2002) is the only research that measures sprawl across the whole country; all other studies limit their sprawl measurements to metropolitan areas.

if this assumption is wrong? What if a significant amount of sprawl development also happens in small- and non- metropolitan areas? If the abovementioned assumption is misleading or inaccurate, then a study which focuses solely on the most populous metropolitan areas would neglect significant amounts of sprawl development happening in small- or non- metropolitan areas.

Therefore, a study measuring sprawl on a nationwide scale is needed as a necessary addition to the literature of sprawl measurement because it could indeed reveal the actual share of total sprawl development happening in the most populous metropolitan areas as well as any sprawl that may be occurring in previously unmeasured or undetected areas, and thus be used to test whether the abovementioned assumption is a valid one or not. If this assumption is shown to be invalid, then future research should consider measuring sprawl across the whole country (not just focusing on the most populous metropolitan areas). This will enable an examination of all sprawl development, including the significant proportion of sprawl development that happens in small- or non-metropolitan areas.

The second notable characteristic of previous studies measuring sprawl is that when density is used as a measuring variable, undevelopable lands have not been excluded from the density calculation. In order to calculate accurate densities, undevelopable lands, including naturally undevelopable lands like open water, wetlands, and so on as well as legally undevelopable lands such as federal lands and state forests, should be excluded from the density calculation (Galster et al. 2001; Cutsinger et al. 2005). However, few empirical studies have refined their density calculations in such a way. Only three studies have excluded undevelopable lands from their density calculations. Following the argument of Galster et al. that undevelopable lands should be excluded from any density calculation (Galster et al. 2001), Cutsinger et al. (2005) clearly recognized the importance of this idea and excluded naturally undevelopable lands from their study, using the National Land Cover Database (NLCD)[3]. Although not specifically embracing this argument, the work of Pendall and Ewing et al. (Pendall 1999; Ewing et al. 2002, 2003a, 2003b) excluded legally undevelopable lands from the density calculation by using the National Resources Inventory (NRI) dataset[4]. That being said, except for these three studies, virtually all other empirical studies have only excluded bodies of water (only one kind of naturally undevelopable land) from the density calculation, while they have included all other kinds of naturally undevelopable lands and all legally undevelopable lands in their density calculation.

The third characteristic of the existing research measuring sprawl is that most of these studies have not addressed the considerably important effects of regional differences in sprawl density measurements (with the exception of the work of Burchell et al. in 2002). Most of these studies calculate a sprawl indicator (or sprawl index) for each examined geographic unit, then sort and rank these units by their sprawl index score, and finally, analyze the most sprawled geographic units and the least sprawled units. Such a method assumes that one single density variant is applicable to every type of jurisdiction across the country. In these studies, a geographic unit with a higher residential density is usually defined as more compact than one with a lower density. This density cutoff is universal to the nation as a whole regardless of the different level of

---

[3] National Land Cover Database (NLCD) 2001 (data source: United States Geological Survey (USGS)).

[4] National Resources Inventory 1997 (data source: United States Department of Agriculture (USDA)). Pendall (1999) and Ewing et al. (2002, 2003a, 2003b) used the variable "urban areas" from NRI 1997 dataset. The "urban areas" in NRI does not include the federal lands.

density of the state that this jurisdiction is a part of. To wit, a density of 250 households per square mile may represent urban development in a low-density state like Montana, but would very probably represent rural development in a high-density state such as New Jersey. Therefore, it is inappropriate to rank or compare geographic units within different state densities solely on the basis of a single density value. In addition to density values, regional context of the studied geographic units must be taken into account when sprawl is measured.

Finally, although quite a few empirical studies have been conducted to measure sprawl, no research has been done yet to compare the results of these existing studies. Because these studies employ different sprawl definitions, datasets, and methods to measure sprawl, the research results of each might be very different when contrasted with another, and sometimes may even be contradictory. For example, as Ewing et al. (Ewing et al. 2003a) noticed, Los Angeles was claimed as one of the most sprawl metropolitan areas in one study but was noted as a compact metropolitan area in another work.

A study comparing and contrasting the results of the existing empirical studies on sprawl measurements could provide useful information about how to improve measurement methods. The discrepancies among the sprawl score rankings of these studies would reveal the effects of different research methods on sprawl measurement results. With this information drawn from comparative analysis, scholars would then be able to more carefully choose those research methods leading to the most accurate sprawl measurement results—these would be results that would draw heavily upon the differences in development patterns of different geographic units.

In an effort to address the above four problems that exist in the current sprawl measurement literature (incomplete coverage, lack of exclusion of undevelopable lands, regional differences in the density variable, and a lack of comparative analysis), this research has two goals: first, to accurately measure sprawl for all the counties in the United States (answering problems one, two, and three); and second, to compare the research results of three existing sprawl measurement studies with these research results to discover possible improvements in sprawl measurement methods (answering problem four). These two goals are discussed below in detail.

1.3 Research Questions

With regard to the aforementioned four features that exist in the empirical literature on sprawl measurements, this research takes the following steps to counteract these problems, which should greatly aid in improving sprawl measurement methods. First, in response to incomplete coverage, this study measures sprawl for any new development over the period 2000 to 2020 for all the counties in the United States, not just for metropolitan counties. Second, in response to lack of exclusion of undevelopable lands, this analysis excludes all undevelopable lands from the household density calculation. Household densities are calculated for all the counties nationwide, as a criterion for classifying the existing county development pattern types (not as a direct measuring variable of sprawl). Third, in response to regional differences in the density variable, this study takes into account the specific density classification of a county in the process of sprawl measurement. And finally, in response to the fourth feature, the lack of comparative analysis, the research results of this study are compared with the research results of Burchell et al. (2002), Ewing (2003b) and Pendall (1999). The discrepancies and overlaps of these research results are analyzed in order to discover the impacts of different measurement methods on sprawl measurement results. The following section details how this research will resolve these four problems and lists the specific research questions to be answered in this investigation.

Incomplete Coverage

Most existing studies that empirically measure sprawl select their study subjects from the most populous metropolitan areas as a convenience sample, with the implicit assumption that sprawl is unimportant or unmeasurable in any other type of location (Ewing et al. 2003a). However, this assumption has not yet been tested to be true, nor has it been substantiated with any significant empirical data. Differing from these extant studies (with the exception of Burchell et al.'s study in 2002), this study measures sprawl for all of the counties in the United States. One major implication of such a nationwide sprawl analysis is that it provides a platform from which to observe information on sprawl for every county in the country. Another implication is that it provides potentially useful empirical data with which to test the abovementioned assumption that sprawl mainly happens in the most populous metropolitan areas. With sprawl being measured for all the counties nationwide, this study will clarify the following:

- Of the 3091 counties in the United States, which will experience sprawl growth during the time period 2000 to 2020?
- What is the total amount of household growth over the period 2000 to 2020 for each sprawl county?

- Of all defined sprawl counties during the period from 2000 to 2020, how many sprawl counties are in large metropolitan areas[5], versus small metropolitan areas, and non-metropolitan areas? In other words, is it appropriate for sprawl measurement studies to focus solely on large metropolitan areas?
- What is the total household growth from 2000 to 2020 for the sprawl counties in large, versus small, and non- metropolitan areas?

Lack of Exclusion of Undevelopable Lands

Regarding the second notable feature of the existing literature on measuring sprawl, this study will exclude both naturally and legally undevelopable lands from the density calculation. In this research, density in 2000 is not used as the primary measurement variable for sprawl, because sprawl is measured for new development, not for existing development. Usually, existing density is used as the sprawl measurement variable when one measures sprawl for existing land use patterns. In this study, density is used as a criterion to classify the counties' existing land use types (i.e. urban center, urban, suburban, rural center, rural, undeveloped), upon which the sprawl determination for new development is based. By using household density in this way, this research will answer the following questions:

---

[5] Here, the term "large" and "small" refer to the population, not the area, of a metropolitan area. The largest metropolitan area is the most populous metropolitan area. This study uses Ewing et al.'s population threshold for defining the "most populous metropolitan areas". Therefore, same as Ewing et al.'s research (2003b), there are also 101 large (i.e., the most populous) metropolitan areas in this research; and the metropolitan areas other than these 101 metropolitan areas are defined as "small metropolitan areas".

- What is the current household density[6] (in 2000) for each state? How should states be categorized into differing density groups?
- What is the current household density for each county? What is the current land-use development pattern type (urban center, urban, suburban, rural center, rural, or undeveloped) for a county?
- Does the exclusion of undevelopable lands from the density calculation make a significant difference in household density for a state or a county? Answering this question will also provide useful information for researchers who use density as the sole sprawl measurement variable.
- How will such density differences, resulting from the exclusion of undevelopable lands, affect the sprawl classification of a county?

Regional Differences in the Density Variable

In addressing the third notable feature of previous studies, that regional context of studied geographic units are usually neglected when measuring sprawl, this research employs the method developed by Burchell et al. (2002) which speaks to the impacts of regional density differences of counties in sprawl classification. In this method, regional differences of counties are determined by the residential density of the host state. The density of the host state affects the classification of the existing county land-use development type and then, further, affects the sprawl definition for a county. Sprawl is significant development in undeveloped, rural, and developing suburban counties. Regional density differences affect the labeling of a county and whether it is sprawling or

---

[6] The “current household density” refers to the base-year household density in 2000; 2000 is the base year of the projected time period from 2000 to 2020. This meaning of “current” also applies to the term “current land-use type”. Sometimes the “current county land-use type” is called the “existing county land-use type” in this research--both refer to county land-use development types in 2000.

not. Taking into account the regional density difference of a county in sprawl measurement, this research will answer the following questions:

- Which counties will experience sprawl growth over the study period 2000 to 2020?
- Which states, Census regions, EAs, and counties will experience the largest amounts of sprawl growth?

Lack of Comparative Analysis

In response to the fourth feature consideration of the existing literature on measuring sprawl, the research results of three previous studies (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b) will be compared with my research results. These three studies were the only available existing studies that measured sprawl at a county level when the data for this study was gathered. The discrepancies and overlaps of the research results of these studies will be highlighted and the impacts of different measurement methods on sprawl measurement results will be characterized. The comparison analyses may well provide additional empirical data and examples to further answer or clarify any issues regarding the aforementioned first three features/questions in the existing literature on sprawl measurements. In the section on comparison analysis, the following questions will be answered:

- What are the discrepancies and overlaps of the research results between the three previous studies and this study?
- What differences in measurement methods would result in these discrepancies?

- What conclusion or conclusions should be drawn from the comparison analyses in order to improve sprawl measurement methods? Specifically, the following questions could be answered:
    - Is a sample of metropolitan counties appropriate for measuring sprawl nationwide?
    - Is it necessary to calculate a more accurate density when measuring sprawl?
    - Is it necessary to address the regional density definitions of a county when measuring sprawl for that county?

This last question, which includes a set of three sub-questions, actually answers the aforementioned first three characteristic features of the literature on sprawl measurement by employing empirical data and examples derived from the comparative analyses.

The research questions, derived from the two objectives of this research, are summarized in this section. What data and methods are used to achieve these two objectives and answer these research questions? This is explained in the next section.

1.4 Research Methodology

As mentioned above, the objective of this analysis is twofold. First, for 3091 counties nationwide, it will determine whether a specific county will experience sprawl during the period from 2000 to 2020. Second, it will compare the research results of three existing studies measuring sprawl with the research results found here. In the following section, the data and methods employed to achieve these two goals are explained.

The first goal of this research is to measure sprawl for new developments (both residential and employment developments) between 2000 and 2020 for all 3091 counties in the United States. Two steps are undertaken to achieve this goal. The first step is to categorize a county into one of the six existing county development patterns in 2000. These are: undeveloped, rural, rural center, suburban, urban, or urban center counties. The second step is to classify counties as sprawl or non-sprawl: if an undeveloped or rural county (presumably without infrastructure) will experience significant growth over the period 2000 to 2020, then it is classified as a sprawl county for this time period. The data and methods used to achieve this first classification will be described below.

In order to determine the existing development pattern for a county, refined household densities are calculated for the county's host state and the county itself. This is the most time-consuming task undertaken in this research. This kind of density is called "refined" because the undevelopable lands are excluded from the density calculation. The denominator of the household density calculation is the developable lands within a county/state, calculated by excluding the undevelopable lands from the total land areas within that particular county/state. Specifically, in this study "Undevelopable Land" is defined as land that is not appropriate for residential, industrial, or commercial development. It includes naturally undevelopable land and legally undevelopable land; the former refers to any land not suitable for development and the latter refers to land which is reserved for public-purpose use and thus not available for private development. In detail, naturally undevelopable lands include open water, perennial ice/snow areas, barren land (rock/sand/clay), unconsolidated shore areas, and wetlands. Legally undevelopable lands contain federal lands[7], Indian reservations[8], state parks and forests, and local parks and recreational areas.

A series of GIS data layers are employed to calculate the remaining area of developable lands. The input GIS datasets include two components. The first component involves boundary files (in the format of vector polygon) for national parks and forests, Indian reservations, state and local parks and forests, state and county boundaries. The second component is a raster data file for naturally undevelopable land which covers categories such as water, wetlands, deserts, and barren lands.

A series of GIS processing steps are employed to calculate the areas of developable lands within a county/state. First, all the GIS input data files are projected into the Albers Equal-Area Conic projection, ensuring that the areas of polygons on digital maps are not distorted. Second, all the GIS input data files in raster format are converted into vector format, because the main spatial tool employed later, the "union analysis" tool, requires that all the input data files be in this format. Third, all the vector boundary files (including the ones just converted from the raster files) are clipped for each state. This step is conducted because the major GIS analyses are processed by state, due to the large size of the input GIS datasets. Fourth, the GIS spatial analysis tool, the "union analysis", is used to find out which lands are developable. Finally, the areas in acres are calculated for the developable lands within each county and state. All of these

---

[7] Federal lands refer to the lands owned or administered by the Federal government. Some examples of federal lands include national parks, national wildlife refuges, military reservations, Federal prisons, and public-domain land. (U.S. National Atlas, http://nationalatlas.gov/printable/fedlands.html)

[8] "An Indian Reservation is land reserved for a tribe when it relinquished its other land areas to the United States through treaties. There are approximately 275 Indian land areas in the U.S. administered as Indian Reservations (reservations, pueblos, rancherias, communities, etc.)." (U.S. National Atlas, http://nationalatlas.gov/printable/fedlands.html)

processes are conducted in GIS software ArcGIS 9.x[9]. This is the general process for calculating the areas of developable lands. Some special cases make this process more complicated and require special methods to calculate the areas of developable lands, and those exceptions are explained below.

There are mainly three kinds of special cases: 1) some state forests and parks contain large areas of privately-owned lands; 2) National Land Cover Database (NLCD) data files have overlapped regions; 3) Non-Indian households may live on Indian reservations; and 4) NLCD datasets are not available for Alaska and Hawaii. One of these problems, the privately-owned lands existing within federal/state forests boundaries, is summarized below. The other three problems, as well as the special methods used to resolve them, will be discussed in detail in Chapter 3, Data and Methods.

Privately-Owned Lands

Not all of the lands within a federal forest/park boundary are owned by the federal government. In other words, there are some privately-owned lands within federal forest/park boundaries, and these privately-owned lands are developable lands on which people usually reside. Sometimes this is true for state forests as well. Additional datasets were acquired to identify the privately-owned lands within any federal or state forest. For federal forests and parks, this problem is solved by using the newly-acquired GIS data "Land Ownership Boundary within Federal Forests" to identify as developable any privately-owned lands within a federal forest/park boundary. For state forests and parks, the acreage of state-owned lands within particular state forests/parks is obtained from the official websites of these state forests/parks, and used to calculate the acreage of

[9] ArcGIS software is developed by ESRI (www.esri.com).

developable privately-owned lands therein. Finally, these developable lands within federal/state forest boundaries, as well as other developable lands, are used as the denominator of the density calculation.

Some mistakes may occur during GIS processing and during the process of transferring the resulting GIS data from GIS software to SAS software[10]. This is mainly due to the large size of the GIS data files and unstable software performances. For example, for some states, the input NLCD dataset has a size of two gigabytes and contains more than 2 million records. Therefore, ongoing checks of the resulting SAS datasets must be done to ensure that no mistake has occurred during data processing.

The method used to check for any mistakes is to compare the input GIS datasets with the initial SAS datasets. These initial SAS datasets are the attribute tables of the resulting GIS files (the GIS result data files after a series of GIS processing). These attribute tables are exported from the GIS software to SAS software and thus become the initial SAS datasets. The checking has shown that these two datasets (the input GIS datasets and the initial SAS datasets) match each other, which means that no mistakes occurred during the GIS processing and the process of transferring the resulting GIS data from GIS software to SAS. In other words, the initial SAS datasets, which contain the acreage for each developable land polygon, are found to be correct and are ready for the various data analyses in the SAS environment.

One of the more significant data analyses is to calculate the household density for counties and states. The first step of this density calculation is to use these initial SAS

[10] SAS (Statistical Analysis System) is statistical software developed by SAS Institute Inc.

datasets to calculate the total acreage of developable lands for each county and state, which will be used as the denominator in the density calculation.

The second step of the density calculation is to calculate the numerator of the county/state household density. The numerator is the number of the non-Indian households (not the number of the total households) for the county/state in 2000. Thi is a comparable procedure to the exclusion of Indian reservation lands from the denominator of the household density calculation. With the denominator and numerator of the density calculation calculated, the household density is calculated for each county and state of the United States. Unlike some of the existing literature on sprawl measurement, this research does not use the county residential density as an indicator of sprawl. Rather, it is used as a criterion for the classification of the county type—reflecting its existing density of development. Based on the state and county household density in 2000, one of six "county types" (undeveloped, rural, rural center, suburban, urban, and urban center) is assigned to each county.

After each county's' existing type is assigned, Woods & Poole projection data in households and employment over the period from 2000 to 2020 is used to determine whether a county is likely to experience sprawl growth during the above time period. The aforementioned sprawl definition is used to define the sprawl/non-sprawl classification for a county. According to this definition, if an undeveloped or rural county will experience significant growth[11] during the period from 2000 to 2020, then it is defined as a sprawl county.

[11] "Significant growth" is defined in Chapter 3. In short, it is the upper quartile percent of county growth rates in the EA or large amount of numerical growth.

Once the sprawl/non-sprawl classification is defined for all the counties in the United States, it is relatively easy to identify where sprawl will occur in the country over the period from 2000 to 2020. Sprawl growth is then analyzed by state, Census region, EA, and finally by county. With these analyses being conducted, the first goal of this research, measuring sprawl for all counties in the United States, is accomplished.

Subsequent to the achievement of this first goal, the second goal of the study can be addressed. The second goal is to compare the research results of three previous studies (Pendall 1999; Ewing et al. 2003b; Burchell et al. 2002) and this study. In order to conduct the comparative analyses, certain pre-treatments of Ewing et al. and Pendall's research results need to be done as both Ewing et al. and Pendall do not define their counties as sprawl or non-sprawl. Rather, they calculate sprawl scores for counties and then rank counties in ascending order by these sprawl scores. This research does not calculate sprawl scores for counties. Instead, it categorizes a county into a binary sprawl/non-sprawl category. Therefore, in order to compare Pendall and Ewing et al.'s results with these research results, a method to convert their research into a binary sprawl/non-sprawl classification for counties is needed.

The comparative analyses of the previous three studies and this study are conducted by means of the following steps. First, one identifies the overlaps and differences of the sprawl/non-sprawl categorization of counties between the studies. Second, the reasons for these differences in the research results are explored. Specifically, one addresses the impacts of any differences in measurement methods on the differences in any research results from these studies. Third, one addresses some of the research questions raised in the previous section, with information from the empirical data and examples developed in the comparative analyses.

1.5 Organization of the Study

This thesis consists of six chapters.

In Chapter 2, is reviewed the theoretical definitions and empirical measurements of sprawl. First, the theoretical definitions of sprawl are reviewed. Borrowed is Burchell et al.'s definition of sprawl (Burchell et al. 2002) as the theoretical definition used in this research. Next is reviewed the existing studies that empirically measure sprawl. Compared with the review of theoretical definitions, the review of empirical measurements of sprawl is given more credence or emphasis in this research, since a major focus of this study is to empirically measure sprawl for counties nationwide. Specifically, three studies measuring sprawl at the county level are explored (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b).

After reviewing the empirical studies measuring sprawl, I would suggest the following points about sprawl measurement. First, the county level is an appropriate geographic level upon which to measure sprawl; it is better suited overall, for example, than the more generic metropolitan level. Second, a sampling which solely focuses on the most populous metropolitan areas might not be enough to measure sprawl nationwide, if that is the intent, because significant sprawl might well occur in small- or non-metropolitan areas. Third, it is absolutely necessary to calculate a refined sprawl measurement, with undevelopable lands excluded from the density calculation. Fourth, regional density differences should be taken into account when measuring sprawl across regions, especially if these regions have different development patterns for a land use type. Finally, no study has been done that compares research results of the literature on measuring sprawl. This study will compare research results of three previous studies (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b) with the findings of this RESEARCH. A comparative analysis will then shed much needed light on how to improve existing methods of measuring sprawl.

In Chapter 3, are described the datasets and methods used to both calculate existing county land-use development patterns in 2000 and to define a sprawl/non-sprawl classification for counties nationwide. First, the datasets and method used to calculate refined household densities are presented at length. Also explained in detail are the aforementioned four special cases/problems that arise in the process of calculating refined densities, as well as the methods used to avoid these problems. Second, a county is defined as sprawl or non-sprawl based on its existing type in 2000 as well as its projected growth over the period 2000 to 2020. County types and resulting sprawl classification for a small number of counties are adjusted in order to allow research results to reflect these counties' real-world functions in the region.

In Chapter 4, are presented the following empirical findings obtained during the process of measuring sprawl for all counties nationwide: the classification of state density groups in 2000; the classification of county existing development patterns in 2000, and the sprawl/non-sprawl classification for all 3091 counties nationwide for the period 2000 to 2020. Four state density groups (very low, low, moderate, and high density) are presented that are derived using refined state household densities. In this research, the state density groups are used as means to determine county types. In order to answer the research question “whether the difference of refined household density and gross household density is significant”, I compare the two different classifications of state density groups, which are based on the two different density calculation methods, respectively. Preliminarily, it is suggested that a refined density calculation is preferable to the gross density calculation[12], which does not include eliminating undevelopable land.

Second, the existing county type (urban center, urban, suburban, rural center, rural, or undeveloped), which is based on refined county household densities and the classification of state density groups, is defined for each county in the United States. In order to reflect regional context, six county development types are defined for each state density group. This method, first developed by Burchell et al. (2002), is called the “variable-density approach” in this study. For example, urban and rural density are ten times as high in Northeastern states versus Prairie states.

Third, state growth in both households and employment over the period from 2000 to 2020 is projected for each county.

Fourth, the research question “where will sprawl happen in the United States?” is answered. Sprawl/non-sprawl classification is defined for each county in the United States. Sprawl growth is analyzed by state, Census region, EA, and county. This research suggests that forthcoming sprawl household growth from 2000 to 2020 will concentrate in a small number of states. The states that will experience most sprawl growth are in the South and the West. In addition, there is a strong linear association between the amounts of sprawl household growth and the overall household growth for states.

---

[12] Here the “gross density calculation” refers to using the Census land areas as the denominator in density calculation. This method, which does not exclude undevelopable lands other than water bodies, is used by most of existing literature that measure residential density.

Finally, sprawl and non-sprawl classifications are presented for individual counties in five selected EAs[13]. These five EAs are recognizable nationwide (or even worldwide) and represent significant components of national sprawl growth.

In Chapter 5, is compared the research results between the three previous studies (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b) and this study. On the basis of this comparison, several conclusions are drawn. First, a sample solely focused on metropolitan counties is not large enough for measuring sprawl nationwide. Of the 492 sprawl counties in this study, only 55 percent are within metropolitan areas. Second, it is necessary to calculate a more accurate, rather than the gross density, in sprawl measurement. Third, the role that density plays in a regional context must be taken into account when sprawl is measured across the country. This is reflected in the fact that a variable-density approach is preferable to a fixed-density approach (or a static density approach)[14] for measuring sprawl nationwide.

In Chapter 6, I discuss the implications of the aforementioned research results in the context of both public policy and the sprawl measurement literature. First, the implications of the research results developed in the process of measuring sprawl are described. Implications of the comparative analysis between the three existing research works (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b) and this research are then

---

[13] The five selected EAs include Atlanta, GA-AL-NC; Los Angeles-Riverside-Orange, CA-AZ; New York-North New Jersey-Long Island, NY-NJ-CT-PA-MA; San Francisco-Oakland-San Jose, CA; Washington-Baltimore, DC-MD-VA-WV-PA.

[14] The fixed-density approach (or a static density approach) is the opposite of the variable-density approach. It is generally used in the existing literature on measuring sprawl. In this study, the "fixed-density approach" refers to the method that compares geographic units solely on the basis of their density values, although these geographic units may have very different regional contexts (that is, may be located in regions with very different densities). The fix-density approach assumes that a particular density value would represent the same land-use development pattern across the country. Pendall (1999) and Ewing et al. (2003b) use this approach in their studies.

presented. Finally, the last section of Chapter 6 discusses the shortcomings of this study and what could improve for future research.

Chapter 2

Literature Review

The existing studies that defined or quantified sprawl are reviewed in this chapter. Also introduced in this chapter are the theoretical definitions of sprawl employed in this study, as well as the improvement in sprawl measurement fulfilled by this study. Because this analysis measures sprawl at the county level, more emphasis is placed on the empirical studies measuring sprawl in the review, especially those studies that also measured sprawl at the county level (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b).

In the following sections, first the studies theoretically defining sprawl are reviewed; then the studies that operationally measured sprawl are introduced (but not fully criticized yet); and finally, the critiques of these empirical studies of sprawl measurement are addressed.

2.1 Theoretical Definition of Sprawl

Although much literature is related to sprawl, there is no consensus of sprawl's theoretical definition; further, few analyses adequately define it (Burchell et al. 1998; Galster et al. 2001; Ewing et al. 2002; Lopez and Hynes 2003). As Galster (2001) summarized, sprawl has been defined from the aesthetic perspective, as certain land-use patterns, as the reasons that caused such land-use patterns, or as the results that have been caused by such land-use conditions. In the following section, these categories developed by Galster are employed to organize the theoretical definitions of sprawl developed by existing studies.

2.1.1 Defining Sprawl from an Aesthetic Perspective

In early studies, sprawl was defined as an inefficient, wasteful, and unpleasant looking development form. In his book *The Language of Cities,* Charles Abrams defined "sprawl" as "[t]he awkward spreading out of the limbs of either a man or a community. The first is the product of bad manners, the second of bad planning" (Abrams 1971, P293-94). As criticized by Ewing (2002), such aesthetic definitions fail to describe the detailed characteristics of sprawl.

2.1.2 Defining Sprawl as Its Own Type of Growth

In these studies, sprawl was defined as a particular land-use pattern at a given point in time. The characteristics of sprawl defined in such literature are summarized below:

- Low density (especially residential density) land-use development (primarily represented by large-lot single-family housing development) (Popenoe 1979; Downs 1992, 1994; Altshuler and Gomez-Ibanez 1993; Moe 1995; Richmond 1995; Black 1996; Ewing 1997; Orfield 1997; Burchell et al. 1998, 2002; Sierra Club 1998; Galster et al. 2001; Lopez and Hynes 2003; Cutsinger et al. 2005, 2006)
- Unlimited outward expansion of new development (Richmond 1995; Black 1996; Burchell et al. 1998, 2002)

- Spatial segregation of land uses, i.e., lack of mixed land uses (Richmond 1995; Ewing 1997; Burchell et al. 1998; Sierra Club 1998; Galster et al. 2001; Cutsinger et al. 2005, 2006; Yin and Sun 2007; Weitz and Crawford 2012)
- Leapfrog and scattered development (Clawson 1962; Harvey and Clark 1965; Mills 1981; Altshuler and Gomez-Ibanez 1993; Moe 1996; Richmond 1995; Ewing 1997; Burchell et al. 1998, 2002; Galster et al. 2001; Cutsinger et al. 2005, 2006; Burchfield et al. 2006)
- Widespread commercial strip development (Popenoe 1979; Altshuler and Gomez-Ibanez 1993; Moe 1995; Black 1996; Ewing 1997; Burchell et al. 1998)
- Low street accessibility (Ewing 1997)
- Lack of public open space (Popenoe 1979; Ewing 1997)
- Lack of community center (Popenoe 1979; Galster et al. 2001)

2.1.3 Defining Sprawl as the End Result of Certain Types of Growth

In this type of literature, sprawl was defined as the negative result of specific types of land-use development. Results that are used to define sprawl are listed below:

- Heavy reliance on private vehicles as means of transportation (Popenoe 1979; Downs 1992, 1994; Moe 1995; Richmond 1995; Burchell et al. 1998; Sierra Club 1998);
- Great variance in the fiscal capacity of local governments (Richmond 1995; Burchell et al. 1998);

- Reliance on filtering or the "trickle-down" process to provide housing to low-income households (Downs 1992, 1994; Burchell et al. 1998)
- Lack of community focus in the social sense (Popenoe 1979)

2.1.4 Defining Sprawl as the Causative Agent of Certain Types of Growth

Some literature includes the causal elements of sprawl in its definition, because "[a]n adequate definition of sprawl must include the "forces" that underlie sprawl's alleged negative impacts in order for subsequent analysis to respond to those impacts effectively." (Burchell et al. 1998) Some of the causes that are included in the theoretical definitions of sprawl are listed below:

- Fragmentation of governmental land-use authorities among many small localities (Richmond 1995, Burchell et al. 1998, Downs 1998)
- Lack of centralized land ownership or development planning (Richmond 1995; Downs 1998; Burchell et al. 1998)

2.1.5 Defining Sprawl as a Process of Development

In these studies, sprawl is used as a verb, describing it as a sequence of development with definite endings (Galster et al. 2001). As Hayden stated, sprawl is "a process of large-scale real estate development resulting in low-density, scattered, discontinuous car-dependent construction, usually on the periphery of declining older suburbs and shrinking city centers." (Hayden 2004 p.7-8)

The sprawl definition of "unlimited outward extension" (Richmond 1995; Black 1996; Burchell et al. 1998, 2002), to some degree, also describes sprawl as a process of development.

2.1.6 Where Does Sprawl Happen?

As summarized by Burchell et al. (1998), sprawl appears on the metropolitan fringes (Altshuler and Gomez-Ibanez 1993), on the edge of existing communities, declining older suburbs and city centers (Moe 1996; Sierra Club 1998; Hayden 2004;), along major suburban highways or at highway interchanges (Altshuler and Gomez-Ibanez 1993; Moe 1996), or leap-frogging into a “relatively pristine setting”, i.e., rural or undeveloped areas where there is not enough facility infrastructure to support the rapid growth (Moe 1996; Burchell et al. 2002).

2.1.7 Summary of the Literature on the Theoretical Definition of Sprawl

There is no consensus on a theoretical definition of sprawl. At times it has been defined from an aesthetic perspective; most often it has been defined as certain land-use patterns; further, the impacts and causes of such land-use patterns have also been, at times, included in the definition of sprawl.

A basic decision to be made in developing a theoretical sprawl definition is whether to limit the definition to land-use patterns or to include causes and impacts as well (Ewing et al. 2002). On this issue, Galster et al. (2001) argued that “[c]onceptually, a thing cannot simultaneously be what it is and what causes it or what it causes. If sprawl is to be a useful concept for describing something important that occurs in urban areas, it must first be reduced to some objective conditions or traits.” (P685) This study agrees with Galster on this point; therefore, the theoretical definition of sprawl in this study is limited to “what it is”.

This research employs the theoretical and empirical definitions of sprawl that were developed and presented by Burchell et al. in 2002. Theoretically, sprawl is defined as “low-density, leapfrog development that is characterized by unlimited outward extension.” (p. 2). This definition describes sprawl as both a certain land-use pattern and a process of development.

The literature of theoretical definitions of sprawl is introduced in this section. Since the major task of this study is to measure sprawl at the county level, the review of sprawl measurement literature, especially that focused on measuring sprawl at the county level, is emphasized. These studies of sprawl measurement are presented in the following section.

2.2 Measurement of Sprawl

There are several descriptive studies on sprawl but far fewer empirical or quantitative studies. Up to now, only a limited number of studies have empirically measured sprawl, and even fewer works have been completed which measure sprawl from multiple dimensions (Ewing et al. 2002, 2003a; Cutsinger et al. 2005, 2006; Yin and Sun 2007). This is mainly due to the elusive concept of sprawl and more importantly, due to the limited data available.

There is a trade-off between the multi-dimension measurement and the extent of the study area, again due to the limitation of data. Single-dimension measurement, which requires less data than multi-dimension measurement, makes it possible to measure sprawl for a much larger study area than the multi-dimension measurement. For example, the single-dimension sprawl measurement developed by Burchell (2002) has been applied to all counties in the US. In contrast, using multi-dimension measurement, Ewing (2002, 2003a) had to reduce their studied objects from the original 101 metropolitan areas to 83 metropolitan areas, because not all of the required data was available for the eliminated 18 metros to be included.

The studies empirically measuring sprawl are reviewed in this section. They will be categorized by the geographic units (i.e. metropolitan areas, urbanized areas (UA)/extended urbanized areas (EUA), and counties) for which sprawl was measured. More emphasis will be given to the studies measuring sprawl at the county level, the same level on which this study is conducted. First, all studies are introduced, and then the common shortcomings of these studies are addressed at the end of this section.

2.2.1 Measuring Sprawl at the Metropolitan Area Level

Several of the empirical studies on sprawl measurement select metropolitan areas as their unit of analysis. Some of them simply measure sprawl (USA Today 2001; Sierra Club 1998; Weitz and Crawford 2012), some relate sprawl to its possible causes (Fulton et al. 2001; Glaeser 2001; Lopez and Hynes 2003; Burchfield et al. 2006; Yin and Sun 2007), and others portray sprawl relative to its impacts (Ewing 2002, 2003a). These studies are listed below.

USA Today, 2001

The USA Today sprawl index (USA Today 2001) was based upon two density measures:

1. Percentage of a metropolitan area’s population living in urbanized areas (defined by the Census Bureau). Since census urbanized areas are density-based, the USA Today sprawl indicators were also density-based.

2. Change in the percentage of metro population living in urbanized areas between the years 1990 and 1999.

The above two sprawl indicators were computed for 271 metropolitan areas and then the metro areas were ranked from 1 to 271 on each sprawl indicator, with a lower ranking number representing lower degree of sprawl. Ultimately, the two ranking numbers were added to create a combined sprawl index score for a metro area.

This study focuses solely on urbanized areas while it neglects the low-density suburbs where sprawl is more likely to take place. Later studies (Ewing et al. 2002, 2003a; Lopez and Hynes 2003, Kurban and Persky 2007; Yin and Sun 2007) corrected this drawback by incorporating low-density suburbs into their sprawl indicator studies.

Sierra Club, 1998

Researchers from the Sierra Club ranked metropolitan areas[15] by series of criteria, which included population shifts from city to suburb, trends in population and land area growth for Census urbanized areas, time wasted in traffic congestion, loss of open-space, and loss of lands involving important habitat or historical importance. Metros were ranked by the above criteria in three size categories: large metros with populations more than one million, medium metros with populations between 500,000 and one million, and small metros with populations between 200,000 and 500,000.

An advantage of this study is that it acknowledges the association between the sizes of metros and their sprawl degrees, which is confirmed by later studies (Lopez and Hynes 2003). However, this study is criticized as being subjective (Ewing 2003a).

---

[15] There are conflicts in the Sierra Club report as to whether the sprawl had been measured for metropolitan areas or for cities. In their report, sometimes the studied objects are referred to as “cities” but sometimes as “metropolitan areas”. According to their resulting tables, it is metropolitan areas for which they measured sprawl. Therefore, here “metropolitan area” is identified as their studied object.

Fulton et al., 2001

This study measures sprawl for new developments (during the time period 1982 to 1997) for 281 out of all 282 metropolitan areas in the US. Fulton et al. (2001) defined sprawl related to pace of development—“If land is being consumed at a faster rate than population growth, then a metropolitan area can be characterized as “sprawling”.” (p. 3)

Their sprawl indicator is borrowed from Pendall’s earlier work (1999). The only difference in the sprawl measurement between these two studies is that Pendall computed it at the county level while Fulton calculated it at the metropolitan area level. In Fulton’s study, sprawl is measured as the change of metropolitan population between 1982 and 1997 divided by the change of urbanized land within that metropolitan area over the same period. The urbanized land16 data was from the National Resources Inventory (NRI) and the population data was from both US Census annual intercensal estimates and their own estimates based on 1990 and 2000 census populations.

This study also examined the association between sprawl and the following assumed causes: 1) population and historical conditions; 2) demographic characteristics; 3) infrastructure endowments and finance; 4) government organization and planning policies; and 5) geographic constraints and agricultural productivity.

---

[16] The United States National Resources Inventory (NRI) defined urbanized land as: “Urban and built-up areas. A *Land cover/use* category consisting of residential, industrial, commercial, and institutional land; construction sites; public administrative sites; railroad yards; cemeteries; airports; golf courses; sanitary landfills; sewage treatment plants; water control structures and spillways; other land used for such purposes; small parks (less than 10 acres) within urban and built-up areas; and highways, *railroads*, and other transportation facilities if they are surrounded by urban areas. Also included are tracts of less than 10 acres that do not meet the above definition but are completely surrounded by urban and built-up land. Two size categories are recognized in the NRI: areas of 0.25 acre to 10 acres, and areas of at least 10 acres.” (Appendix 3. Glossary of Selected Terms. Summary Report, 1997 National Resources Inventory, Revised December 2000)

Conducted by the US Department of Agriculture every five years, the NRI dataset is assembled from data collected at more than 800,000 sample sites nationwide on non-federal lands. Fulton claimed that using NRI urbanized lands is a better choice for density calculation than using Urbanized Areas (UA) as defined by the Census Bureau. This is because the former is an actual measurement of urbanized land, while the latter is just a measurement of population density17.

The drawback of this study is similar to that of Pendall's study, which will be addressed in detail in a later section.

Glaeser et al., 2001

Glaeser and his co-researchers measured sprawl in terms of job decentralization in metropolitan areas. For the 100 largest Primary Metropolitan Statistical Areas (PMSA) and Metropolitan Statistical Areas (MSA) in the US, they calculated the percentage share of overall employment in a metropolitan area that was within a three-mile ring of the Central Business District (CBD), as well as the shares that were within, and beyond a ten-mile ring of the CBD. In their study, the less the share of employment within three-mile ring, or the more the share of employment beyond the ten-mile ring, the higher the degree of sprawl a metropolitan area has.

Based on the percentage share of metro employment that is within the three-mile ring and on the share that is beyond the ten-mile ring, the 100 largest metros are categorized into four groups: dense, centralized, decentralized, and extremely decentralized employment metros. From the dense to the extremely decentralized

---

[17] As explained by Fulton, "Because the Census definition of urban areas includes a density threshold, the Census excludes some areas that would be identified as urban by the NRI. The NRI would also exclude certain areas—especially large parks within urban areas—that the Census incorporates within urban areas. On net, however, the NRI finds more urban acreage than the Census." (Furlton et al. 2001, p.18)

employment metros, the percentage shares of metro employment within the three-mile ring decreased from at least 25 percent to less than 10 percent; while the percentage shares within the ten-mile ring decreased from more than 60 percent to less than 60 percent.

The Glaeser study also explored the factors that cause sprawl. They concluded that job sprawl was not significantly correlated to the age of the primary city in the metropolitan area, but there was a significant relationship between job sprawl and the political fragmentation of an area.

The author of this study agrees with Lopez and Hynes (2003) in that the uniform "three-" or "ten-" mile rings causes Glaeser's sprawl measurement to be affected by the size of metropolitan areas. Large metropolitan areas tend to have higher percentages of employment outside the three- or ten-mile rings than small metropolitan areas. Thus, it is inappropriate to use the uniform three- or ten- mile rings as criteria for sprawl measurement using metropolitan areas of various sizes.

Ewing et al., 2002, 2003a

Ewing (2002, 2003a) measured sprawl for 83 metropolitan areas (MSAs/PMSAs/NECMAs) with populations over 500,000 as of 2000. He defined four sprawl dimensions: residential density, land-use mix, degree of centering, and street accessibility. For each dimension, he used principle component analysis to combine several measurement variables into one sprawl factor.

Except for the mix factor, which was calculated only for the year 1990 due to limited data availability, the three other factors were calculated for both the year 1990 and 2000, revealing the changes of sprawl on each dimension during the involved time period.

Ewing also investigated the associations between sprawl and its hypothesized transportation impacts such as vehicle ownership, commute mode choice, commute time, vehicle miles traveled per capita, traffic delay per capita, traffic fatalities per capita, and ultimate ozone levels. These association analyses were conducted using multiple regression, with the sprawl factors acting as independent variables and the various transportation impacts as dependent variables.

Lopez and Hynes, 2003

The Lopez and Hynes study measured sprawl for 330 out of all 331 US metropolitan areas[18] at two points of time: 1990 and 2000. The changes in degree of sprawl over time and across metropolitan areas were also examined. The authors concluded that metropolitan areas' population sizes are significantly associated with their sprawl degrees. That is, large metropolitan areas are much more likely to have lower degrees of sprawl, while small metropolitan areas are more likely to be characterized by higher degrees of sprawl.

The sprawl indicator in this study was computed as the percentage share of total population in an MA (i.e. metropolitan area) that live in low-density census tracts minus the percentage share of total population living in high-density census tracts. If all of a MA population lives in low-density tracts, then this MA has the highest degree of sprawl; in contrast, if all of a MA population lives in high-density tracts, then this MA has the lowest degree of sprawl.

---

[18] "Metropolitan areas" in their study refers to MSAs and PMSAs (illustrated by the author of this study).

They used uniform criteria to define low- and high- density census tracts across metropolitan areas. Census tracts with densities less than 200 persons per square mile, between 200 and 3,500, and greater than 3,500 persons square mile were classified as rural[19], low-density, and high-density tracts, respectively. This is a kind of static-density approach which is not appropriate for a nationwide study.

Burchfield et al., 2006

The Burchfield study measured the scatteredness dimension of sprawl (i.e., scateredness of residential development) for all 275 metropolitan areas[20] in the contiguous US. Their datasets included the 1992 National Land Cover Data[21] as well as the land cover data for the year 1976 from the US Geological Survey (USGS) and the US Environmental Protection Agency. The datasets were made up by 30×30 meters cells in image format. Each cell represented one of eight land cover categories (i.e. residential development; commercial and industrial development and transportation networks; water; bare rock and sand; forest; range and grassland; agricultural land; and wetlands). The first two categories were defined as developed lands and the other categories were defined as undeveloped lands.

In practice, their sprawl index measured “the percentage of undeveloped land in the square kilometer surrounding an *average* residential development in each metropolitan area” for two points of time, the year 1992 and the year 1976 (p. 605). In order to compute this sprawl index, first the percentage of undeveloped land in the immediate square kilometer was calculated for *each* 30×30 meter cell of residential

---

[19] Rural tracts were excluded from the analysis.
[20] “Metropolitan areas” in their study refers to MSAs and CMSAs (illustrated by the author of this study).
[21] The 1992 National Land Cover Data is based on the Landsat 5 Thematic Mapper satellite imagery.

development. Then the average percentage was computed among all cells of residential development in a metropolitan area, which was used as the sprawl index.

Further they explored the reasons that caused the dramatic variance in the degrees of sprawl across the metropolitan areas. They found that ground water availability, temperate climate, rugged terrain, decentralized employment, early public transport infrastructure, uncertainty about metropolitan growth, and unincorporated land at the urban fringe will increase the degree of sprawl.

Yin and Sun, 2007

They measured sprawl and explored the impacts of state growth management programs on sprawl. Sprawl was measured for the 294 metropolitan areas (MSAs and PMSAs) that have state growth management programs. Two dimensions of sprawl—density and land-use mixture—were calculated for metropolitan areas at two points in time: 1990 and 2000. The set of density variables calculated for each metropolitan area included gross population density, the proportion of population living in low-density census tracts, and the proportion of population living in high-density tracts.

The method to calculate density and land-use mixture variables was borrowed from Ewing et al.'s study (2002) and improved by the authors. One important improvement was that they did not use a uniform density value to define "low-" or "high-" density census tracts as Ewing did[22]. Instead, they used the first and third quartile of census tract densities in a metropolitan area as the threshold densities to identify "low-" or "high-" density census tracts, respectively. A census tract with a density below the first

---

[22] Ewing et al. (2002) defined the census tracts with densities lower than 1,500, between 1,500 and 12,500, and higher than 12,500 persons per square mile as low-, medium-, and high- density census tracts, respectively.

quartile of all tract densities in *its own metropolitan area* was defined as a low-density tract; meanwhile, a tract with a density above the third quartile was identified as a high-density tract in the metropolitan area. This improvement acknowledged that the thresholds for "low-" or "high-" density census tracts varied across metropolitan areas. This shares a common idea with the variable-density approach that was first developed by Burchell et al (2002).

Weitz and Crawford 2012

Weitz and Crawford's study measured job sprawl, using proximity of employment to residential locations, for all of 358 metropolitan areas (MSAs and PMSAs) in the conterminous US. The job sprawl index was defined as "the percentage of change in job accessibility from 2001 to 2006", with a negative value (i.e., decreasing job accessibility) representing job sprawl for a metropolitan area. Sprawl was described as a development process, rather than a static picture, in their study.

In order to calculate the job sprawl index, first zip code points (primarily using the centroid of an areal zip code) and points of populated places[23] are located on digital maps. The attributes for zip code points were total employment for the year 2001 and 2006; the attributes for populated places were total population in 2000[24]. Second, locational job accessibility (i.e., job accessibility for each zip code point) was calculated using a gravity model, based on the populations of surrounding populated places[25] and the straight line distances between the zip code and each of its surrounding populated places. Third, job accessibility at the metropolitan level was calculated by computing the

[23] The datasets of zip code points and points for populated places are from the Census.
[24] Both the employment and population datasets are from the Census.
[25] Only those populated places within a 30-mile distance to the zip code are included in the calculation.

weighted average of locational job accessibilities for all zip codes within the metropolitan area, using the employment at each zip code point as the weight. Finally, job sprawl index was calculated as the percentage of change in job accessibility for a metropolitan area over the time period 2001 to 2006. Their research results showed that 227 metropolitan areas experienced job sprawl (i.e., experienced a decrease in job accessibility) during this time period.

This study explicitly defined job sprawl and job non-sprawl for metropolitan areas nationwide, which is helpful for policy makers to identify sprawl locations and make policies accordingly. However, due to limitations of data availability, their research results were only exploratory. For example, admitted by the two authors, a finer resolution for job locations and population locations (such as census tracts, or block groups) could have made the measurement of job sprawl more accurate. Further, their employment dataset did not include government employment and self-employment, which made their research results less desirable.

Common shortcomings shared by the studies that measured sprawl at the metropolitan area level will be addressed in a later section, right after the studies measuring sprawl at the county level are presented.

2.2.2 Measuring Sprawl at the (Extended) Urbanized Area Level

These studies measured sprawl at the urbanized area[26] (UA) level (Downs 1998; Galster et al. 2001; Kurban and Persky 2007), or at the self-defined extended urbanized

---

[26] The Census Bureau's definition of urbanized areas (1990 census) is presented below: "The Census Bureau delineates urbanized areas (UA's) to provide a better separation of urban and rural territory, population, and housing in the vicinity of large places. A UA comprises one or more places ("central place") and the adjacent densely settled surrounding territory ("urban fringe") that together have a minimum of

area[27] level (Cutsinger et al. 2005). Among them, some simply measured sprawl (Galster et al. 2001); some measured sprawl and explored its various causes (Cutsinger et al. 2005; Kurban and Persky 2007) or its impacts (Downs, 1998b). These studies are introduced below.

Downs, 1998b

Downs measured eight sprawl indicators and their relationship to urban decline for 162 urbanized areas that had at least a population of 150,000 in 1990[28]. Based on data from the 1990 US census, these sprawl indicators included: (1) land area of the Census Bureau defined Urbanized Area; (2) population density of the urbanized area outside the central city/cities; (3) ratio of central city population density to urbanized fringe density; (4) percentage of the total metropolitan area population living outside the urbanized area; (5) percentage of the total metropolitan area population living within the central city/cities; (6) Percentage of urbanized area commuters who drive alone or in carpools; (7) number of separate jurisdictions that control land use per 100,000 metro area residents; and (8) ratio of poor central city residents to poor suburban residents.

50,000 persons. The urban fringe generally consists of contiguous territory having a density of at least 1,000 persons per square mile. The urban fringe also includes outlying territory of such density if it was connected to the core of the contiguous area by road and is within 1 ½ road miles of that core, or within five road miles of the core but separated by water or other undevelopable territory. Other territory with a population density of fewer than 1,000 people per square mile is included in the urban fringe if it eliminates an enclave or closes an indentation in the boundary of the urbanized area. The population density is determined by (1) outside of a place, one or more contiguous census blocks with a population density of at least 1,000 persons per square mile or (2) inclusion of a place containing census blocks that have at least 50 percent of the population of the place and a density of at least 1,000 persons per square mile." (http://www.census.gov/population /censusdata/urdef.txt)

[27] The EUA is defined as "the Census Bureau-defined urbanized area, as well as each additional outlying square-mile cell comprising the metropolitan statistical area that has 60 or more dwelling units and from which at least 30% of its workers commute to the urbanized area." (Cutsinger et al. 2005, p. 237)

[28] Since this article cannot be found online, the author summarizes the content of this article from Ewing et al.'s study (Ewing et al. 2002).

According to his analysis, there was no statistically significant relationship between suburban sprawl and urban decline.

The author of this study agrees with Ewing (2002) in that Downs' sprawl indicator system mixed sprawl with its impacts and causes. The percentage of urbanized area commuters is an indicator of sprawl's transportation impact; the ratio of poor residents is one of sprawl's social impacts; while the number of separate jurisdictions represents a cause of sprawl.

Cutsinger et al., 2005

Cutsinger measured sprawl for the geographic units "Extended Urban Areas (EUA)" (defined by Cutsinger et al.). A EUA is defined as "the Census Bureau-defined urbanized area, as well as each additional outlying square-mile cell comprising the metropolitan statistical area that has 60 or more dwelling units and from which at least 30 percent of its workers commute to the urbanized area." (p. 237) The authors argued that EUAs are better geographic units for sprawl measurement than metropolitan areas or urbanized areas.

Based on Galster et al.'s (2001) conceptual and operational definitions of sprawl, Cutsinger defined and operationalized seven dimensions of land-use patterns: density, continuity, concentration, centrality, proximity, mixed-use, and nuclearity. Fourteen sprawl indicators measuring both residential and employment land use were developed to represent these seven sprawl dimensions. A principal components analysis of the fourteen indicators resulted in seven principal components (factors), which were independent of each other.

These seven factors measured the seven sprawl dimensions for 50 EUAs within the 50 metropolitan areas that were sampled from the 100 largest metropolitan areas in 1990 in the US. Cutsinger's data was primarily from three data sources: the 1990 Census data on housing units, the 1992-93 National Land Cover Data (NLCD) on land-use types, and the 1990 Census Transportation Planning Package (CTPP) data on employment. They ranked the 50 EUAs by scores on each of the seven principal components.

Cutsinger finally employed regression analyses to explore the relationship between the seven principal factors of sprawl measurement and the hypothesized causes of sprawl, which included EUAs' geographic size, age, population, growth, and topographical constraints on development.

A significant contribution of their study was that it measured sprawl on the net available lands in EUAs, with an exclusion of the lands unavailable for development such as open water, perennial ice and snow, woody wetlands, and emergent herbaceous wetlands[29]. This method is borrowed by this study, which will be explained in detail in next chapter.

Kurban and Persky, 2007

Kurban and Persky's study defined sprawl as the rapid growth in low-density suburbs in urbanized areas. Sprawl was measured for urbanized areas[30] in 83 large core-based statistical areas (CBSAs) for two time periods: from 1990 to 2000 and from 2000 to 2004.

---

[29] The data for these undevelopable lands were from the National Land Cover Data (NLCD) that was published by US Geological Survey (USGS).

[30] Each of these large urban areas has a population greater than 250,000 people and 25 or more suburban places.

The data was from the US Department of Housing and Urban Development's HUDUSER–States of Cities data set (2005). All places in urban areas in a CBSA were aggregated into four rings (three suburban rings and the primary, largest city), based on their population densities. The density of a place was calculated as the total population divided by land area for the year 1990 (both datasets are from the1990 Census). The suburban places with densities below the 33rd percentile of all suburban places in *their own CBSA* were grouped as Ring 1 (or low-density suburbs); the suburbs with densities above the 33rd percentile and below the 66th percentile were grouped as Ring 2 (or medium-density suburbs); and the remaining suburbs were grouped as Ring 3 (or high-density suburbs). Then the average annual population growth (for both time periods) for all rings was calculated. Empirically, they defined sprawl as the rapid growth in low-density places in the metropolitan area. This was Burchell's definition (Burchell et al. 2002).

Further, based on regression analyses, the authors claimed that "although metropolitan areas with rich central cities sprawl somewhat less, the pace of suburban sprawl is primarily driven by metropolitan growth" (p. 179).

This study had the advantage that it defined low-, medium-, and high- density suburbs based on their own regional contexts. That is, the thresholds for these suburb classifications vary across metropolitan areas; this method also shares some common ideas with the variable-density approach (which was first developed by Burchell et al. in their 2002 study). However, the authors limit their sprawl identification to urbanized areas, thus neglecting a large portion of overall sprawl development that takes place outside urbanized areas.

2.2.3 Measuring Sprawl at the County Level

So far, three studies measured sprawl at the county level (Pendall 1999, Burchell et al. 2002, Ewing et al. 2003b). Extra emphasis will be placed on the review of these studies, because this study also measures sprawl at the county level. In this section, these three studies are compared from the aspects of study areas, definitions of sprawl, and analyses of sprawl's causes or impacts. For easier reading, Burchell et al.'s study (2002) and Ewing et al.'s study (2003b) will be referred to as "Burchell's study" and "Ewing's study", respectively, in the following sections.

Study Areas

Burchell's study (2002) has the largest study area thus far among all studies measuring sprawl in the US. It measured sprawl for all 3091 counties (and county equivalents) in the US. In contrast, both Pendall (1999) and Ewing (2003b) focused their studies on the counties in large metropolitan areas. Ewing's sample contained all the 448 counties in the 101 most populous metropolitan areas as of 1990 Census; while Pendall's sample covered 159 counties in the 25 largest metropolitan areas.

Theoretical and Operational Definitions

Pendall (1999) and Burchell (2002) measured sprawl for new development that happened during certain time periods. Different from those studies that measured sprawl for existing land use patterns at a particular point in time, this type of study empirically defined sprawl as a process of development. Pendall measured sprawl for new development that happened between 1982 and 1992. Burchell measured sprawl for two time periods: 1980 to 2000 and 2000 to 2025.

Pendall did not develop a theoretical definition of sprawl. Rather, he developed an operational sprawl measurement which he admitted to be actually “a measure of density”. This measurement was “the change in county population between 1982 and 1992, divided by the change in urbanized acres of land over the same period.” (p. 558)

Among the 180 counties in the 25 largest US metropolitan areas (as of 1990), Pendall measured sprawl for the 159 counties that gained population between 1982 and 1992, with an exclusion from his sample of the 21 counties that lost population but gained urbanized area during this time period. However, the author of this study thinks that this exclusion is misleading, because these counties, which lost population but gained urbanized areas, may suggest an even higher degree of sprawl than the counties that gained both population and urbanized areas, as admitted by Pendall (1999).

Burchell (2002) developed a comprehensive theoretical and operational definition for sprawl. Theoretically, sprawl was defined as “low-density, leapfrog development that is characterized by unlimited outward extension.[31]” (p.2) Empirically, sprawl was defined as rapid and significant growth occurring in rural and undeveloped counties[32], where there is insufficient infrastructure to contain the significant growth. Sprawl or non-sprawl was explicitly defined for each of the 3091 counties in the US. Burchell’s method of sprawl measurement included the following steps.

First, based on 1995 household densities of states, the 50 US states were categorized into four density groups (very low; low; moderate, and high density states).

---

[31] “Sprawl is low-density, leapfrog development that is characterized by unlimited outward extension. In other words, sprawl is significant residential or nonresidential development in a relatively pristine setting. In nearly every instance, this development is low density, it has leapt over other development to become established in an outlying area, and its very location indicates that it is unbounded.” (Burchell et al. 2002 p. 2)

[32] A small number of suburban and rural center counties were also classified as sprawl.

By doing so, the regional differences in densities for counties (represented by the densities of their host states) are subsumed into the classification of county land use types.

Second, based on 1995 household densities for both the counties and their host states, one of six land use types (urban center, urban, suburban, rural center, rural, or undeveloped) was assigned to each of the 3091 US counties. “These development patterns reflect a county’s existing land use, infrastructure capacity, and regional function.” (p. 54)

A notable feature of the classifications of county land use types is the application of the variable-density approach. First developed by Burchell (2002), this approach correctly acknowledges that the same density value may not be used for the same land-use pattern in different parts of the country. As a result, the threshold for a given county land use type *varies* across the United States. A rural county in Wyoming may be one tenth the density of a rural county in California or Florida; an urban county in New Jersey may be ten times the density of an urban county in South Dakota.

Finally, sprawl was defined as significant *residential or nonresidential* development that happens in rural or undeveloped counties. A rural or undeveloped county was identified as sprawl if either of the following sets of criteria was met:

1. (a) The county’s annual growth rate (of households, employment, or both) was in the upper quartile of the EA’s[33] annual county growth rates (households, employment, or both); (b) the county’s annual growth rate exceeded the average annual national county

[33] Totally there are 172 EAs (Economic Areas) in the U.S. as defined by BEA (Bureau of Economic Anlysis).

growth rate; and (c) the county's absolute level of growth exceeded 40 percent of the average annual absolute national county growth.

or

2. The county's absolute level of growth exceeded 160 percent of the average annual absolute national county growth." (Burchell et al. 2002, p. 3, p. 572)

The first criterion primarily identifies the relative pace of growth in a county, while the second criterion provides the threshold for "significant" absolute growth in counties. Based on these criteria, all 3091 US counties were classified as sprawl or non-sprawl for two time periods: 1980-2000 and 2000-2025.

Subsequently, for the counties that experienced sprawl during the time period 2000 to 2025, Burchell calculated the relative costs and benefits under two development scenarios: an uncontrolled-growth (i.e. sprawl) scenario and a controlled-growth scenario. The study concluded that sprawl had more costs than benefits even though there were significant instances of each.

Burchell's study provided important information for policy makers and researchers. First, it explicitly identified "which specific state and county locations were the sites of significant sprawl growth nationally". Second, it quantified impact differences between the uncontrolled- (i.e. sprawl) and the controlled- growth scenarios. Third, this study identified county candidates that could act as the receiver of future growth which is redirected from sprawl counties. These research results provided useful, ready-to-use information for policy makers to implement policies to control sprawl.

The simple but elegant sprawl measurement (i.e. density and growth variables) employed in the Burchell study makes it possible to measure sprawl for all US counties.

However, while enjoying the attributes of a large study area, Burchell's study also suffers from the inability to capture the impacts of sprawl development at a local level. This type of local sprawl development includes strip commercial development, lack of mixed land use, and low-density development that happens in places smaller than a county. There is a trade-off between developing a multiple-dimension measurement for sprawl at a larger scale and measuring sprawl at the property or parcel level. Ewing et al. (2003b) measured sprawl for existing land-use patterns at one time point (the year 2000[34]); their study is like taking a "snap-shot" picture of the existing land-use patterns. Their sample contains all 448 counties in the 101 most populous US metropolitan areas.

Ewing's operational definition of sprawl characterized sprawl in four dimensions: low residential density, low level of land-use mix, low degree of centering, and poor street accessibility. However, due to limited data availability, only two dimensions (i.e. low residential density and poor street accessibility) were measured at the county level.

The variables that made up the "low residential density" dimension included: (1) gross population density; (2) percentage of the county population living at low suburban densities; (3) percentage of the county population living at moderate to high urban densities; and (4) the net density in urban areas (county population divided by the amount of NRI urbanized land areas). Meanwhile, the "street accessibility" dimension was made up of two variables: (1) average block size and (2) percentage share of all blocks that had amounts of land less than 0.01 square miles. With principal components analysis, these six variables were then combined into one factor (i.e. sprawl index), which represented the overall degree of sprawl within a county.

---

[34] Five of the six sprawl variables were for the year 2000 and one variable (urban areas) came from the 1997 NRI (national resource inventory) dataset.

Ewing used the amount of land data from the Census to calculate his first three density variables. Census lands include both developable lands and undevelopable lands[35]. Based on this land data, the computed residential density may be underestimated when compared to the land use pattern that exists in the real world. For example, a census tract which has a *low* density in Ewing's study may actually contain a *dense* residential development and a large area of undevelopable lands (say, wetlands). In other words, with undevelopable lands being included in the density calculation, the computed densities do not reflect the realities of residential density at the development level.

Regional Differences in the Density Variable

Neither Pendall (1999) nor Ewing (2003b) explicitly identified a county as sprawl or non-sprawl. Instead, they sorted counties by sprawl index score; the lower the sprawl index score, the higher degree of sprawl a county had. Both of these sprawl index scores were primarily determined by county densities. In general, in these two studies the lower the population density[36], the higher degree of sprawl a county has.

When comparing counties located in different regions by county densities, they applied the "fixed-density" (or "static") approach in their studies. This approach assumed that the same density value could be used to represent the same land use type across the country. It used solely the densities of the studied geographic units as the measuring variable for land-use patterns, while neglecting regional differences in densities of these geographic units. It would be fine if this approach was applied only to the counties in the

---

[35] These undevelopable lands include wetlands, deserts, small areas of inland water bodies, federal lands, state parks and forests, and local parks.

[36] In Ewing's study (2003b), the population density was measured for existing land use patterns at a given point in time; while in Pendall's study (1999), it was computed for the new development that happened during a period of time.

same region, say, the counties within one state; but problems arise when it is employed in a national context, where the sample contains counties from the states with very different average densities. A same density of 140 households per square mile, for example, may represent both an urban county in Montana and a rural county in Massachusetts.

In Ewing's and Pendall's study, counties located in the states with relatively high densities were likely to have greater sprawl index scores and thus ranked toward the "compact" end of the sample list. This is because their densities were relatively high compared with the counties in the same sample that were located in low-density states. However, some of these "compact" counties actually were sprawl counties in the region[37], due to their relatively low densities (or low sprawl index score) compared with their surrounding counties.

On the contrary, Burchell (2002) developed the "variable-density" approach, which took into account the regional differences in densities of counties (determined by the density of a county's host-state) in the classification of county land use types. Burchell argued, "[c]reating a uniform density to categorize the prevailing density in a county could easily ignore regional cultural differences and would employ either an "average" or "best practices" development density to classify the land-use pattern of a region in a particular state." (Burchell et al. 2002, p. 56-57) One of the six land use types (urban center, urban, suburban, rural center, rural, and undeveloped county), which was based on densities of *both* the county *and* its host-state, was assigned to each county. Based on this approach, the density threshold of a given county type *varied* with the host-state density. For example, the density threshold for suburban counties in a very-low

---

[37] Some examples are presented in Chapter 5.

density state ranged from 27 to 116 households per square mile while in a high density state it ranged from 267 to 658. In other words, the same density value cannot represent the same county land use type in states that have very different underlying densities.

With each county being assigned a land-use type, now it was possible to compare land-use types among the counties from different states. In other words, it made less sense to compare density values among counties nationwide (i.e., using the “fixed-density approach”), because in different states the same density value could not represent similar land use types. In contrast, the county type classification, which is based on the variable-density approach, can be compared among counties across the country, because it takes into account the average density in the host state.

Relationships between Sprawl and Its Reasons/Impacts

In addition to measuring sprawl at the county level, these three studies also investigated relationships between sprawl and its possible causes (Pendall 1999), or assessed various impacts of sprawl (Burchell et al. 2002; Ewing et al. 2003b)

Pendall (1999) estimated the impacts of land-use controls on sprawl. He primarily investigated the relationships between sprawl and the following causal variables: percentage of land area under formal (legal) control, farm characteristics, metropolitan fragmentation, housing values, local government spending, transportation infrastructure, and minority population.

Burchell (2002) assessed impacts of sprawl both on resources (such as water and sewer infrastructure, local road infrastructure, local public-service costs, and real estate development costs) and on personal costs (such as travel miles and costs, quality of life, and urban decline). The impacts of sprawl were evaluated by comparing the costs between two scenarios: the uncontrolled-growth (sprawl) scenario and the controlled-growth (compact development) scenario of the new development between 2000 and 2025.

Ewing (2003b) employed hierarchical linear and nonlinear models to estimate the impacts of sprawl on public health which included: leisure time physical activity levels, BMI (body mass index) and obesity, hypertension, diabetes, and CHD (coronary heart disease).

The three studies that measured sprawl at the county level have been summarized in this section. These were the classic sprawl measurement studies in the 1990's and early 2000's. The critique of existing literature measuring sprawl at various geographic levels is addressed in the following section.

2.3 Critique of the Sprawl Measurement Literature

Studies measuring sprawl at various geographic levels are summarized in the previous section, where a few shortcomings specific to some of these studies have also been presented. In this section the common problems shared by most of these studies are addressed.

2.3.1 County Level versus Other Geographic Levels

Which geographic level is appropriate for measuring sprawl? Is it metropolitan areas, urbanized areas (defined by the Census Bureau), extended urbanized areas (Cutsinger et al. 2005), or counties?

A major portion of previous empirical studies measured sprawl at the level of "metropolitan areas", which refer to metropolitan statistical areas (MSA), consolidated metropolitan statistical areas (CMSAs), primary metropolitan statistical areas (PMSAs)[38], and New England county metropolitan areas (NECMAs) (USA Today 2001; Sierra Club 1998; Fulton et al. 2001; Ewing et al. 2002, 2003a; Lopez and Hynes 2003; Burchfield et al. 2006; Yin and Sun 2007; Weitz and Crawford 2012). As correctly criticized by some researchers, the metropolitan area may not be an appropriate unit of analysis for sprawl measurement, because land use patterns usually vary substantially within a metropolitan area (Burchell et al. 1998; Cutsinger et al. 2005; Wolman et al. 2005). As argued by Burchell (1998), studies relating an *average* density of a metropolitan area with its certain characteristics like travel behavior may not be meaningful at all, because travel choices tend to be affected by the density of the *specific* places where people live and work, which may be significantly different from the metro average.

The author of this study agrees with these scholars on the idea that the metropolitan area is too big to be an appropriate geographic unit for measuring sprawl. Two metropolitan areas with very different development patterns might have similar gross densities: one may contain high-density urbanized areas surrounded by low-density rural areas (a compact development pattern) and the other may be characterized by low-density, spread-out development in multiple counties (a sprawl development pattern).

Having realized that the metropolitan area is a coarse analysis unit for sprawl analysis, some scholars employed relatively smaller geographic units as the unit of analysis, which include urbanized areas (UA)[39] (Downs 1998; Galster et al. 2001; Glaeser

---

[38] As illustrated in the foregoing section, for some studies, the "metropolitan areas" represents MSAs and CMSAs; while for other studies, it represents MSAs and PMSAs.
[39] UA is defined by the Census Bureau.

et al. 2001; Kurban and Persky 2007) or extended urbanized areas (EUA)[40] (Cutsinger et al. 2005). Another reason for employing UA/EUA as a unit of analysis, as declared by these scholars, is that it avoided including the outlying rural areas in sprawl measurement, thus avoiding the exaggeration (i.e. overestimation) of sprawl degree for an area.

A primary shortcoming for using UA/EUA as the unit of analysis for sprawl measurement is that they are not legally incorporated places and thus have no governmental management authority responsible for the growth planning within the territory. Therefore, the research results of these studies may offer less policy significance.

Further, a disadvantage specific to studies using *UA* as the analysis unit is that they cannot capture the development occurring at and beyond the urban fringe which might be the essence of sprawl (Galster et al. 2001; Cutsinger et al. 2005). This type of growth, which occurs at urban fringes or in rural areas, often develops at low densities that are far below the density thresholds for defining a UA and thus this type of growth is excluded from the analysis, although in reality it is indeed the essence of “sprawl” development.

In order to solve the aforementioned problems resulting from using metropolitan areas or UA/EUA as a unit of analysis, the county is selected as the unit of analysis for this study. There are several benefits of employing the county as the unit of analysis. First, because this study is aimed at measuring sprawl nationwide, the unit of analysis must consistently cover the whole country. Counties satisfy this requirement perfectly. In

[40] EUA is defined by Cutsinger et al. (2005).

contrast, none of the aforementioned areas (i.e. metropolitan area, UA, or EUA) can satisfy this requirement, because they only cover particular parts of the nation.

Second, a county is the smallest nationwide geographic unit that has non-educational government. Actually, only 12 of the 50 states have continuous local government entities subordinate to the county level (i.e. at the place level) (Burchell et al. 2002). It is important for the studies measuring sprawl that the unit of analysis has both land-use and political significance for the study findings to have a meaningful impact. County governments can make and implement anti-sprawl policies in their counties. As declared by Burchell (2002), "land-use issues such as sprawl are increasingly seeking solutions at this geographic level (i.e., county level- added by the author of this study) because most of these jurisdictions have both land-use and political power and can be assembled into larger units to form a region." (p. 43)

Third, a variety of historical and projection data is available at the county level (Burchell et al. 2002). For specific reasons such as confidentiality or costs, the smaller the geographic areas, the more limited the data for those areas. For most (if not all) of the national socioeconomic projection datasets, a county is the smallest geographic level. Specifically, for projection databases, the smaller the area, the higher risk of projection errors for these areas. In the Woods & Poole database (a major demographic projection dataset employed in this study), the county level is the lowest geographic level for projection.

Due to the above advantages relative to other geographic levels, the county is employed as the basic unit of analysis in this study.

2.3.2 Incomplete Coverage

To date, most of the prior empirical studies focused their study area on (the most populous) metropolitan areas, no matter if their studied objects were the metros themselves or UAs/EUAs/Counties within the largest metros. There are two primary reasons that the largest metros attract the most attention from empirical sprawl measurement studies. One is that sprawl is commonly thought to happen in suburban or urban fringes of metropolitan areas; the other is that metropolitan areas are more populous and thus thought to be of more importance than rural areas (Ewing et al. 2003b).

However, the assumption that sprawl primarily occurs in metropolitan areas should be tested. In order to do so, sprawl must be measured for both metropolitan and *non-metropolitan* areas. Further, it is interesting to investigate the extent and degree at which sprawl takes place outside metropolitan areas. To answer these questions, sprawl must be measured in *non-metropolitan* areas as well. In his earlier work, Ewing (1997) recognized that sprawl is the “spread-out, skipped-over development that characterizes the non-central city metropolitan areas and non-metropolitan areas of the United States” (Burchell 1998, p. 1). However, in his later empirical study measuring sprawl practically, he focused solely on the largest metropolitan areas because “two thirds of the American people live there” (Ewing et al. 2003b); non-metropolitan areas were excluded from his sample.

In response to the incomplete coverage of sampling, this study measures sprawl for all counties in the US, making it possible to discover the extent and degree to which sprawl takes place in non-metropolitan areas.

2.3.3 Lack of Explicit Classifications of Sprawl or Non-sprawl

Most of the prior studies quantifying sprawl did not explicitly identify their studied objects as sprawl or non-sprawl[41]. Rather, they calculated sprawl scores for areas and sorted these areas by their scores. The areas at one end of the list have the greatest degree of sprawl in the sample and those at the other end have the least degree of sprawl (i.e. the highest degree of compactness). A shortcoming of these studies is that it is hard for policy makers to tell whether the areas with midsized sprawl scores are sprawl or not. This study explicitly classifies counties as sprawl or non-sprawl, making the results more informative to policy makers and researchers.

2.3.4 Lack of Exclusion of Undevelopable Lands

Residential density is the most frequently employed sprawl measurement (Galster et al. 2001). By strict definition, it refers to the density of *residential lands*. However, in most (if not all) empirical studies measuring sprawl, it has been defined as the population (or households[42]) divided by all *urban lands* (which include residential, commercial, industrial, transportation, and other urbanized lands[43]). This enlargement of the denominator of density calculation from residential lands to overall urban lands might be due to the lack of data on the amount of residential lands that actually exist.

Two problems need to be addressed when measuring residential densities. The first is how to constrain the measurement of residential density solely to urban lands (i.e.

---

[41] Only two studies (Burchell et al. 2002; Weitz and Crawford 2012) explicitly defined an area as sprawl or not.

[42] Some researchers argued that residential units are a better indicator than population to measure density, because the former better represent the physical land use conditions (Galster et al. 2001).

[43] Some scholars argued that only developable lands within an urban area should be used as the denominator of the density calculation (Galster et al. 2001; Cutsinger et al. 2005; Wolman et al. 2005)

how to exclude rural and undeveloped lands from the density calculation). This problem is especially significant when sprawl is measured as a type of land use pattern. The other problem is related to the exclusion of undevelopable land from the calculation of residential density. The methods that prior studies employed to deal with these two problems are addressed below.

The first problem (i.e. calculate residential density on the basis of urban lands) is especially relevant to studies that measured sprawl as a type of land use patterns. In these studies (Downs 1998; Pendall 1999[44]; Fulton et al. 2001; Ewing et al. 2002, 2003a, 2003b; Galster et al. 2001; Cutsingser et al. 2005), the calculation of residential density for an area (such as metropolitan areas or counties) was constrained to urban lands. That is, rural areas and large areas of continuous undeveloped lands were excluded from the density calculation. If these lands are included in the density calculation, the resulting residential density of the area will be underestimated, and the sprawl calculation for that area will be exaggerated (i.e. overestimated).

These studies primarily employed three methods to exclude the rural or undeveloped lands from the density calculation. The first method was to use the urbanized area (UA, defined by the Census) (Downs 1998; Galster et al. 2001; Glaeser et al. 2001) or extended urbanized area (EUA) (Cutsinger et al. 2005) as the denominator for density calculation. The second method employed urbanized lands[45] from the National Resources Inventory (NRI) as the denominator for density calculation (Pendall 1999; Fulton et al. 2001; Ewing et al. 2002, 2003a, 2003b). A shortcoming of the NRI

---

[44] Among these studies, Pendall's and Fulton's studies analyzed the change of land use patterns during a period of time; while other studies measured existing land use patterns at a given point in time.
[45] The NRI urban and built up areas include residential, industrial, commercial, institutional land, roads and highways, small parks (less than ten acres), and so forth (Fulton et al. 2001).

data is that the amount of urbanized land is an estimate from surveying and thus is subject to all the typical errors of sampling. As recognized by the NRI website, the NRI data is statistically reliable for analysis at the national, state, and multi-county level, but may not be reliable for a county-level analysis. Finally, the third technique was to exclude low-density census tracts (for example, fewer than 100 residents per square mile) within the area, with an assumption that these census tracts represent rural, desert, and other undeveloped areas (Ewing et al. 2002, 2003a, 2003b; Lopez and Hynes 2003). The author of this study agrees with Yin and Sun (2007) that the threshold of low-density census tracts should vary across the country, rather than to apply a uniform threshold to all types of areas.

Density calculation does not need to be constrained to urban lands for the studies where density is not used as a *direct* sprawl indicator. In these kind of studies (Burchell et al. 2002), density is employed as a criterion for classifying the county land use types (i.e. urban center, urban, suburban, rural center, rural, or undeveloped); thus, it should not be constrained to urban lands but should be based on the overall developable lands within the area (i.e. rural and undeveloped lands should not be excluded from the density calculation). This study belongs to this school of study, and thus the density calculation is based on both urban lands and rural lands.

The second problem (i.e. the need to exclude undevelopable lands from the density calculation) is important to both aforementioned types of studies. No matter if residential density is employed as a direct measure variable for sprawl or as the criterion to classify county land use types, undevelopable lands should be excluded from the density calculation. If undevelopable lands are not excluded, the resulting density may be underestimated (Galster et al. 2001; Cutsinger et al. 2005; Wolman et al. 2005).

Two methods have been developed to fulfill this task. The first method (Pendall 1999; Fulton et al. 2001; Ewing et al. 2002, 2003a, 2003b) used NRI urbanized lands as the denominator of density calculation. NRI urbanized land does not include federal lands[46] and large parks (larger than 10 acres), both of which are part of legally undevelopable lands. The second technique (Cutsinger et al. 2005) used USGS data to exclude naturally undevelopable lands such as open water, perennial ice and snow, woody wetlands, and emergent herbaceous wetlands from the density calculation. In order to calculate accurate densities, this study borrows and improves Cutsinger's method by excluding both naturally and legally undevelopable lands (i.e. national, state, and local forests and parks) from the density calculation.

2.3.5 Regional Differences in the Density Variable

Most of the existing studies measuring sprawl did not address the significant effects of regional differences in sprawl density measurement[47]. In these studies, areas across the country were sorted by their densities; the areas with higher densities were defined as more compact (i.e. less sprawl) than those with lower densities. Another example of this problem was that, a uniform density value was often used to represent low-density locations across the country (Ewing 2002, 2003a, 2003b; Lopez and Hynes 2003). By doing so, these studies assumed that a static density value represented the same

[46] However, Indian tribal and individual Indian trust lands are not defined as federal lands in NRI and thus may be included in urbanized lands.
[47] Three exceptions are the studies of Burchell et al. (2002), Yin and Sun (2007), and Kurban and Persky (2007).

land use type across the country. However, this assumption is not true, due to its neglecting the regional differences in densities for the studied geographic areas. As explained in chapter one, a density of 250 households per square mile may represent urban development in a low-density state like Montana, but a figure 10 times this amount would represent urban development in a high-density state such as New Jersey.

This study employs the method developed by Burchell et al. (2002) which addressed the impacts of regional density differences of counties in sprawl classification. In this method, regional differences of counties are determined by the residential density of the host state. As a result, the same density value cannot represent the same land use type in counties across the country; the resulting classifications of county types better reflect the actual land use patterns of the host state.

2.3.6 Lack of Comparative Analysis

Although quite a few empirical studies have been conducted to measure sprawl, no study has yet been done to compare their research results. Based upon different sprawl definitions, datasets, and methods, the research results of these studies sometimes are very different and even contradictory when compared with each other. The discrepancies among these results reveal the effects of different research methods on sprawl measurement results. Therefore, a study comparing the research results of these studies can provide useful information about how to improve methods for sprawl measurement. This study compares its results with the research results of three studies that measured sprawl at the county level (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b); further, the impacts of different sprawl measurement methods on research results is characterized.

2.4 Summary of the Sprawl Measurement Literature and Research Questions Restated

Since the 1990s, multiple studies have empirically measured sprawl. Sprawl has been measured at the metropolitan level (USA Today 2001; Sierra Club 1998; Fulton et al. 2001; Ewing et al. 2002, 2003a; Lopez and Hynes 2003; Burchfield et al. 2006; Yin and Sun 2007; Weitz and Crawford 2012), at the urbanized area (UA) level (Downs 1998; Galster et al. 2001; Glaeser et al. 2001; Kurban and Persky 2007), at the self-defined extended urbanized area (EUA) level (Cutingser et al. 2005), and at the county level (Pendall 1999, Burchell et al. 2002, Ewing et al. 2003b).

These empirical studies have five notable features: incomplete coverage, lack of explicit classifications of sprawl/non-sprawl, lack of exclusion of undevelopable lands, neglecting regional differences in the density variable, and a lack of comparative analysis.

With regard to these, this study takes the following steps to counteract these classic shortcomings. First, in response to the problem of incomplete coverage, this study measures sprawl for all the 3091 counties in the United States (i.e. not just for metropolitan counties). This study agrees with Burchell (2002) in that for a national study of sprawl measurement, the county is a better unit of analysis than a metropolitan area, an urbanized area, an extended urbanized area, or a census tract; thus the county is selected to be the unit of analysis in this study.

Second, in response to the lack of explicit classifications of sprawl/non-sprawl, each one of the 3091 US counties is explicitly defined as sprawl or non-sprawl.

Third, in response to the lack of exclusion of undevelopable lands, this analysis excludes all undevelopable lands from the household density calculation. Household densities are calculated for all counties and states nationwide, as criteria for classifying the existing county land use types. The NRI urbanized land area data is not used in this study, since it solely covers urbanized lands. A variety of GIS datasets are employed to exclude both naturally and legally undevelopable lands from the density calculation.

Fourth, in response to regional differences in the density variable, this analysis represents the regional context of a county with the density of the county's host state. Based on densities of both states and counties, the county land use types classified in this study better reflect the real world than the studies that use "fixed-density" (or "static-density") approach[48].

Finally, in response to the lack of comparative analysis, the research results of this study are compared with the research results of Pendall (1999), Burchell et al. (2002), and Ewing (2003b). The discrepancies and overlaps of these research results are analyzed in order to discover the impacts of different measurement methods on sprawl measurement results.

The existing literature on sprawl definition and measurement is introduced and criticized in this chapter. The following major research questions will be addressed in the following chapters:

1) Why is the variable-density approach better than the fixed or static-density approach?

---

[48] The fixed-density approach is defined in Chapter 1.

2) Based on the variable-density approach, what is the existing land use type (urban center, urban, suburban, rural center, rural, or undeveloped) for a county?

3) Of the 3091 counties in the US, which will experience sprawl growth during the time period 2000 to 2020?

4) Does the exclusion of undevelopable lands from the density calculation make a significant difference in household density for a state or county?

Chapter 3

Data and Methods

All counties experiencing sprawl growth during the period from 2000 to 2020 are identified in this study. Based on a review of the literature , this study will improve existing sprawl measurement by using more accurate densities, by embracing a superior variable-density approach in the process of measuring sprawl, and by measuring sprawl for a much larger area—all US counties.

Datasets and methods which are employed in order to fulfill this task are explained in this chapter. Of special note, however, is that most emphasis is put on the measurement of accurate density, because it is one of the most important tasks of this study. GIS and non-GIS datasets are processed to calculate densities, using both GIS and SAS software[49].

Based on calculating the accurate densities of states and counties, one of six existing land-use types (urban center, urban, suburban, rural center, rural, and undeveloped) is then assigned to each county by applying the variable-density approach. Finally, based on county land use types and the projected growth during the 2000 to 2020 period, counties are then carefully classified as sprawl or non-sprawl.

3.1 Geographic Levels

Major geographic levels that are involved in the analysis are explained in this section.

[49] SAS (Statistical Analysis System) is statistical software developed by SAS Institute Inc.

County and County Equivalents

The reason for selecting the county as the basic geographic unit of analysis has been addressed in chapter 2. This section explains why 3091 counties and county groups are used as the basic unit of analysis in this study.

According to the Census Bureau, counties are "the primary legal division of every state except Alaska and Louisiana" (Census glossary documents[50]). County equivalents[51] are "geographic entities that are not legally designated as a county, but are treated by the BEA[52] and the Census Bureau as equivalent to a county for data presentation purposes" (Census glossary documents).

There are 3141 counties and county-equivalents in 2000. The estimated households and employment in the year 2000 and the projected households and employment in 2020 used in this study are from the Woods & Poole database. This database contains demographic and economic data (past and projected) for every US county for every year from 1970 to 2025. Because boundaries between counties (or between counties and independent cities) may have changed since 1969, in order to make historical data consistent over time, the Woods & Poole data has grouped counties which have experienced boundary changes into new county groups that have consistent boundaries from 1970 to 2025. These 3141 counties and county equivalents are combined to create the 3091 counties and county groups that have consistent boundaries over time (Woods & Poole 2003). These 3091 counties and county groups are the basic, constant geographic

---

[50] This definition is available on Census website http://www.census.gov/dmd/www/glossary.html#C (accessed September 2012).
[51] These county equivalents include the boroughs, city and boroughs, municipality, and census areas in Alaska; parishes in Louisiana; and independent cities (independent of any county) in Maryland, Missouri, Nevada, and Virginia; and the District of Columbia (Census Glossary).
[52] The Bureau of Economic Analysis (BEA)

units of analysis in this study and are hereafter referred to as 'counties' for the purposes of this study.

Further, in order to ensure that these geographic units of analysis are consistent among all the datasets, all datasets from any other data sources based on 3141 counties have been processed to fit into the same county groups as defined by the Woods & Poole database. This creates the ability for a more uniform and stable analysis over time.

Economic Area (EA)

EAs are relatively independent economic units. According to the Bureau of Economic Analysis' (BEA) 1995 redefinition of economic areas[53] (EA), "(e)ach economic area consists of one or more economic nodes—metropolitan areas or similar areas that serve as centers of economic activity—and the surrounding counties that are economically related to the nodes." In total, there are 172 EAs identified for the year 2000. Their boundaries are determined[54] so that "(1) each economic area includes, as far as possible, the place of work and the place of residence of its labor force; and (2) each economic area is economically large enough to be part of BEA's local area economic projections program." (Johnson, 1995) The EA links rural and undeveloped counties to metropolitan areas.

---

[53] "BEA first defined economic areas in 1969, and then redefined them in 1974, 1977, 1983, 1995, and most recently, 2004. The major changes made by the 2004 redefinition include: (1) 179 economic areas in 2004 redefinition; (2) the redefinition used the OMB's (U.S. Office of Management and Budget) 2004 revised standards for MSAs and new standards that recognize, for the first time, micropolitan statistical areas." (Johnson, Kenneth P. and John R. Kort, 2004)

[54] The boundary of an EA is mainly determined by commuting patterns. In less populous EAs, the boundary is determined by newspaper circulation pattern.

Following Burchell et al.'s method (2002), this analysis uses Economic Area (EA), developed by the Bureau of Economic Analysis (BEA), as the host region[55] in the operational definition of sprawl. This is done for two reasons. First, in Burchell et al.'s study (2002), both the operational definition of sprawl and the concept of redirecting future sprawl growth require a relatively independent economic area as the host region of counties; that is, economic relationships are strong within the EA, but relatively weak across EAs. As relatively independent economic units, EAs function as a commuter shed or region.

Using EAs as host regions in the definition of sprawl, a county is defined as sprawl if it experiences significant growth relative to the growth rates of other counties in the EA and that growth takes place in an area likely to not have infrastructure, i.e. a rural or undeveloped county. Further, any possible redirection of future sprawl growth should be conducted within an EA (Burchell et al. 2002).

The second reason for using EAs as the host region of a county is that this study is aimed at measuring sprawl for all counties nationwide and thus EAs, which cover all counties nationwide, are able to be used for a complete analysis. The more generally used geographic level—the metropolitan area—cannot fulfill this task, because metropolitan areas only cover 836 of the 3091 US counties. Therefore, based on these two reasons, the economic area (EA) is used to delineate a host region of a county for the operational definition of sprawl.

3.2 Dataset for Growth Projection

[55] The other kind of host region used in the sprawl classification analysis is the host state of a county.

The major dataset used in this project to measure the amount of growth is the "Woods & Poole 2003 Regional Projections and Database," which utilizes data sources of the 2000 US census and at the time, the most current datasets from the US Department of Commerce. It contains more than 900 demographic and economic variables for every county in the United States for each year, extending from 1970 to 2025.

From this database, two variables capturing two time points (total households and employment at county levels for the years 2000 and 2020, respectively) are used in for sprawl measurement. Specifically, they are used to measure both household and employment growth for each county during the 2000 to 2020 time period; in addition, the household figure in 2000 is used to calculate residential densities for counties and states.

The household data[56] in the Woods & Poole database is from both Census Bureau counts in 1970, 1980, 1990, and 2000[57] and projections from 2000 to 2030. In the Woods & Poole database, total employment data includes both full- and part-time jobs by place of work. Historical employment data for 1969-2000 is from the US Department of Commerce, Bureau of Economic Analysis. The projected households and employment data for the year 2020 in the Woods & Poole database is based on the projection model developed by Woods & Poole.

In addition to the households and employment data provided by the Woods & Poole database, the other primary datasets employed in this project are GIS datasets, which are

---

[56] Households are defined as occupied housing units. "A housing unit is a house, an apartment, a group of rooms, or a single room occupied as separate living quarters. The occupants of a housing unit may be a single family, one person living alone, two or more families living together, or any group of related or unrelated persons who share living quarters. All people are part of a household except those who reside in group quarters, which include living arrangements such as prisons, homes for the aged, rooming houses, college dormitories, and military barracks." (Woods & Poole 2003)

[57] The household data from the decennial censuses are adjusted from April 1st to July 1st (Woods & Poole 2003).

used in the calculation of residential densities for counties and states. The detailed information about these GIS datasets is discussed in the following section.

3.3 Calculating Household Density

In order to measure sprawl for new developments over the period 2000 to 2020 in a given county, the first step is to measure household density at the state level and the county level for the base year 2000. Based on these densities, one of the six land use types (i.e., urban center, urban, suburban, rural center, rural, and undeveloped) is then assigned to each county. If a rural or undeveloped county[58] experiences significant growth during the 2000 to 2020 period, the county is classified as sprawl (i.e., the county will experience sprawl growth[59]).

The following section will describe the datasets and methods employed to calculate the household densities in 2000 for states and counties.

3.3.1 Datasets for Calculating Developable Land Areas; Other Ancillary Datasets

As addressed in chapters 1 and 2, the household density should be based on the amount of "developable" land instead of "total" land area. In this study, a major task in calculating household density is to calculate its denominator—the amount of developable land. The datasets employed to calculate developable lands are described in this section.

---

[58] For some developing suburban and rural center counties, their new developments may also be classified as sprawl. However, the number of such sprawl suburban and rural center counties is small.
[59] This concept could be presented in various ways: "the county's new developments will be classified as sprawl", "the county is classified as sprawl", "the county is identified as the location of sprawl growth", or "the county will experience sprawl growth". These narrations represent the same situation.

Definition of Developable vs. Undevelopable Lands

In order to calculate more accurate household densities, lands that are undevelopable—that is, not appropriate for development—are excluded from the land area (in measurements of square miles) that is used as the denominator of the density calculation. If this step is not taken, the household density of a county (or state) will be underestimated and the comparison of densities among counties (or states) will not be as accurate. For example, lakes and state parks within a county should be excluded from the calculation of the county's household density. Otherwise, such a county may be calculated as having the same household density as a county that has the same total area (but no lakes nor state parks within it) when in actuality, relative to actual developable lands, the former county has a much higher household density than the latter.

In this study, "undevelopable area" is defined as the area that is not appropriate for residential or nonresidential development. There are two kinds of "undevelopable lands": "naturally undevelopable lands" and "legally undevelopable lands". The former refers to lands that are realistically unsuitable for development; the latter refers to lands that are reserved for public-purposes and thus unavailable for private development. More specifically, "naturally undevelopable lands" include open water, perennial ice/snow, barren land (rock/sand/clay), unconsolidated shore lands, and wetlands; while "legally undevelopable lands" refer to federal lands[60], American Indian reservations[61], state parks and forests, and local parks and recreational lands.

---

[60] Federal lands refer to the lands owned or administered by the Federal government. Some examples of federal lands include national parks, national wildlife refuges, military reservations, Federal prisons, and public-domain land. (U.S. National Atlas, http://nationalatlas.gov/printable/fedlands.html)
[61] "An Indian Reservation is land reserved for a tribe when it relinquished its other land areas to the United States through treaties. There are approximately 275 Indian land areas in the U.S. administered as Indian

Conversely, "developable land[62]" is land that is *both* naturally *and* legally developable. In practice, it refers to land that is not characterized in any way as undevelopable. The amount of developable land within a geographic unit is used as the denominator in the calculation of its residential density.

The exclusion of the aforementioned "undevelopable lands" will be determined via GIS (Geographic Information Science) processing. The datasets describing these undevelopable lands, which are mainly GIS overlays obtained from various data sources, will be described in the following section.

Datasets of Naturally Undevelopable Lands

The data of "naturally undevelopable lands" for the year 2000[63] is taken from the National Land Cover Database (NLCD) 2001 (data source: U.S. Geological Survey (USGS)). The definition and classification of "naturally undevelopable lands" in this study is mainly based upon the Cutsinger et al. study (2005), which classified as undevelopable areas the NLCD categories of open water, perennial ice and snow, woody wetlands and emergent herbaceous wetlands. In addition to Cutsinger et al.'s classification of undevelopable areas (all of the undevelopable lands defined by therein are actually *naturally* undevelopable lands), this study also defines the NLCD barren lands (rock/sand/clay) and unconsolidated shore areas as naturally undevelopable lands. Legally undevelopable lands are not addressed in Cutsinger et al.'s study but are analyzed

---

Reservations (reservations, pueblos, rancherias, communities, etc.)." (U.S. National Atlas, http://nationalatlas.gov/printable/fedlands.html)

[62] It can be named as "Developable Area" or "Developable Land". Since "developable area" must be land (i.e. not be water), here in this study the "Developable Land" is used more often.

[63] The NLCD 2001 dataset are primarily for the year 2000, although individual dates may range from 1999 to 2006 (http://www.mrlc.gov/mrlc2k_product_desc.asp).

in this study. The NLCD land cover classes, defined as naturally undevelopable lands in this study, accompanied by their USGS definitions, are listed in the following table.

Table 3.1 NLCD Land Covers Defined as Naturally Undevelopable Lands

NLCD Land Covers	**Definitions**
Open Water	All lands underneath open water, generally with less than 25% cover of vegetation or soil.
Perennial Ice/Snow	All lands characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
Barren Land (Rock/Sand/Clay)	Barren lands of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
Unconsolidated Shore	Unconsolidated lands such as silt, sand, or gravel that is subject to inundation and redistribution due to the action of water. Characterized by substrates lacking vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable. Erosion and deposition by waves and currents produce a number of landforms representing this class.
Woody Wetlands	Lands where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
Emergent Herbaceous	Lands where perennial herbaceous vegetation accounts for

NLCD Land Covers	Definitions
Wetlands	greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Datasets of Legally Undevelopable Lands

The datasets of “legally undevelopable areas” are primarily the compilation of GIS boundary data from various data sources. These datasets depict boundaries for federal and Indian lands, state forests and parks, and local parks and recreation lands. Usually, every feature of the dataset is a polygon that depicts the boundary of a legally undevelopable geographic unit such as a federal park, an Indian reservation, and so on. In addition to the GIS boundary data, an ancillary dataset for state forests and parks—the acreage of state-owned lands within a state forest or park—is obtained from the official websites of these state forests and parks. The datasets of these legally undevelopable lands are described in the following section.

The data of federal lands and Indian reservations for the year 2000[64] is from the GIS digital map “U.S. National Atlas Federal and Indian Land Areas” (data source: National Atlas, available from the U.S. ESRI Data & Maps 2003 CDs). Only areas equal to or greater than 640 acres (one square mile) are included in the dataset. At a scale of 1:2,000,000, this dataset is very appropriate for spatial analysis at a national or regional level, and also appropriate for this level of analysis, due to its unique nature. This is the only publically available spatial data of federal lands and Indian reservations and is used

[64] The dataset of “U.S. parks” is for the year 1997.

in this study to exclude federal lands and Indian reservations from the calculation of household density.

The datasets of state parks and forests, and local parks and recreation areas for both 1997 and 2002 come from the digital map "U.S. Parks" (1997) and "U.S. GDT Park Landmarks" (2002) (data source: U.S. ESRI Data & Maps 2003 CDs). The "U.S. Parks" mainly focuses on state forests and parks; while the "U.S. GDT Park landmarks" focuses on local parks and recreation areas. At a large scale of 1:500,000[65] and 1:100,000, respectively, these two datasets meet any accuracy requirements for spatial analysis at a county level.

For the aforementioned data of federal lands ("U.S. National Atlas Federal and Indian Land Areas"), the boundary data of national forests turns out to be somewhat less than desired. The data itself only depicts the boundaries of national forests, but does not delineate the boundaries of privately-owned lands that are located within these national forests. The total area within national forest boundaries is about 232 million acres[66]; of these 232 million acres, about 8 percent (or 18 million acres[67]) are privately owned lands. The ESRI boundary data of national forests was originally used to calculate household densities for counties but problems arise for a few counties which are fully contained by one or several national forests. For these counties, developable lands are calculated as zero because they are fully covered by national forests. Meanwhile, these counties show

---

[65] The source data of the regional and local parks and forests in the "U.S. Parks" is at a scale of 1:100,000, but the overall scale of this dataset is reduced to 1:500,000 by other less accurate data sources (i.e. the data sources for national parks and forests). Therefore, the real scale of the regional and local park and forest boundary polygons (features) in the dataset "U.S. Parks" may be at an even better 1:100,000, instead of 1:500,000.
[66] This number is calculated from these GIS boundary data in U.S. ESRI Data & Maps 2003 CDs.
[67] This number is calculated from the "National Forest Land Ownership" GIS data from the United States Department of Agriculture (USDA) Forest Service, which will be explained in the following section.

reasonably large populations. Where do these people live? The fact is that these people live on those privately-owned lands that are located within the federal forest boundaries. However, these privately-owned lands, which are indeed developable lands, cannot be identified using the ESRI datasets.

In order to solve this problem, a GIS dataset that depicts the boundaries of both the national forest and the privately-owned lands located within the national forest is needed. Fortunately, the dataset that fulfills this requirement, the "National Forest Land Ownership" GIS data, is available from the United States Department of Agriculture (USDA) Forest Service. This data (in the form of a digital map) contains the basic ownership parcels of the surface estate within national forests in year 2000 for the 48 continental states, as well as for Alaska and Puerto Rico; polygons (every feature of the digital map is a polygon) in the map depict ownership parcels by organization type. The basic ownership types include USDA Forest Service, private ownership, and state agencies. By using this GIS dataset, privately-owned lands within national forest boundaries can now be readily identified.

Rather than using both this dataset and the ESRI "federal and Indian lands data" as the data input of national forests, this land ownership data is the only data input of national forests for the subsequent GIS analyses. This is due to three reasons. First, this land ownership data covers almost all (96.4%) of the national forest lands owned and managed by USDA Forest Service[68]. Secondly, this land ownership data includes all national forests and grasslands in the "federal and Indian lands data". The exception is a

---

[68] The total acreage of national forest lands owned and managed by USDA Forest Service is 192,383,077 acres (This number is obtained from the official USDA Forest Service website.). This "National Forest Land Ownership" dataset covers 185,410,000.00 acres (96.4%) of lands owned and managed by USDA Forest Services.

national forest in the state of Utah, which has been manually added for later GIS data input processing. Finally, at a scale of 1:24000, this ownership data allows a much more accurate spatial analysis than the “federal and Indian lands” dataset which is at the 1:2,000,000 scale. Thus, only the land ownership data has been used as the data input of national forests for subsequent GIS analyses.

Table 3.1 presents the amount of lands within national forest boundaries by ownership types, summarized from the “National Forest Land Ownership” data. As shown in the table, eight percent (around 18 million acres) of the total area within national forest boundaries is privately owned, which is the second largest ownership type (to the USDA) for lands within national forest boundaries. The naturally developable portion of these privately-owned lands is used as developable lands in the household density calculations. The lands owned by all other agencies in this dataset, including the 185 million acres owned by USDA forest service, are legally undevelopable and therefore excluded from the calculation.

Table 3.2 Amount of Lands within National Forests by Ownership in 2000

Land Ownership	Area (in 1000 Acres)	Percentage Share (%)
USDA Forest Service	185,410	83.3
Private	18,129	8.1
State	16,642	7.5
Other	2,313	1.0
Total	222,495	100

Source: summarized from the GIS “National Forest Land Ownership” data by United States Department of Agriculture (USDA) Forest Service.

Other Datasets

In addition to the aforementioned major datasets, various supplemental datasets from multiple sources are also used for sprawl measurement. These datasets include: the acreage of state forests and parks (obtained from the official websites of these state forests and parks); the GIS boundary data for census tracts, counties and states in 2000 (obtained from U.S. ESRI Data & Maps 2003 CDs); land areas of census tracts in 2000 (obtained from the Census 2000 website); and the number of American Indian households[69] by census tract in 2000 (obtained from the Census 2000 website). These datasets are described in detail in the following section.

State Forests and Parks

The data regarding the acreage of state forests and parks, obtained from the official websites of these state forests and parks, serves as supplementary data for calculating developable lands within state forest and park boundaries. This is the total acreage of lands within state forest/park boundaries that are "publicly owned", i.e., owned and managed by governmental agencies. The aforementioned problem of privately-owned lands being located within national forest boundaries also exists for some state forests and parks. Not all state forests and parks have such a problem, but for some, this is a particularly significant issue. Unfortunately, the GIS data of land ownership within forest (park) boundaries are unavailable for most state forests and parks. Therefore, the new dataset, the acreage of state forests and parks, is collected in order to calculate the areas of developable lands within the state forest (park) boundaries.

[69] In this study, American Indian households are defined as the households with an American Indian householder.

Census Tracts, Counties and States

The GIS boundary data for census tracts, counties and states in 2000, extracted from the Census Tiger/line files, is available from the U.S. ESRI Data & Maps 2003 CDs. In this dataset, state and county boundaries have been created by merging census tract boundaries, which ensures the consistency of boundaries through these different levels of geographic units. The polygon features of this dataset are intersected (through GIS "union" analysis) with the polygons of the abovementioned various undevelopable areas, resulting in the polygons of developable lands with the census tract identification. Areas of these resulting polygons are summarized for census tracts, counties and states using this identification. In addition, this GIS boundary data of census tracts has been checked to ensure that it contains all of the census tracts that have developable lands[70]; thus it can be used as the basic GIS dataset in this study.

Land Areas of Census Tracts

The third ancillary dataset, the acreage of land areas for a census tract, obtained from the Census 2000 website, is employed to calculate the "gross household density" by using the gross density calculation[71]. This gross density is then compared with the new, refined density calculated in this study (which is based on the new density calculation),

---

[70] Except for one census tract with a small area of developable lands, all census tracts with a nonzero developable lands can be found in this ESRI census tract dataset. In detail, 106 census tracts are not included in the ESRI census tract dataset. However, of these 106 census tracts, 105 are totally covered by water body and have neither developable lands nor population; the only one that has a population only contain 3.1 square mile lands, which is occupied by five persons in one housing unit, according to Census 2000 data.

[71] As mentioned in chapter 1, this method uses the area of lands (with the water bodies being excluded from the total area) as the denominator of density calculation. In other words, only one kind of naturally undevelopable areas (the water body) is excluded from the density calculation, while other kinds of naturally undevelopable areas and all legally undevelopable areas are not excluded.

highlighting the advantage of the new density calculation method over the gross density calculation.

American Indian Households

The data of total American Indian households[72] in a census tract, obtained from the Census 2000 website, is used to calculate the area of developable lands within an Indian reservation. It is also used to calculate the number of non- American Indian households for a county, the numerator of the county household density calculation.

The datasets employed to calculate the acreage of developable lands have been described in this section. The methods employed to calculate the amount of developable lands are addressed in the following section.

3.3.2 Methods for Calculating Developable Land Areas

As defined in the foregoing section, “developable lands” are those lands that are *both* naturally *and* legally developable. In practice, these lands refer to those that do not belong to any of the abovementioned “undevelopable areas”. By definition, developable and undevelopable lands are mutually exclusive. The sum of the developable areas and undevelopable areas within a geographic area is equal to its total. For a geographic area, this study uses the amount (measured in square miles) of its developable lands as the denominator in the calculation of its density.

This section describes both the general method used to calculate the amount of developable lands and the specific methods used to solve problems caused by three kinds

[72] In this study, Indian households are defined as the households with an Indian householder.

of special cases. Obviously, checking is conducted to ensure that all GIS processing results are correct.

Data processing

As the first step in data processing, the Albers Equal-Area Conic Projection is used to project all the GIS datasets. A major goal of the GIS analysis in this study is to create polygons of developable lands in a digital map that have undistorted areas compared with their corresponding developable land parcels in the real world. Then the areas of these polygons will be summed to create the amount of developable lands in a county, which will then be used in the density calculation. Thus, one must ensure that the areas (in square miles) of these polygons in the digital map are exactly the same as the areas of their corresponding land parcels in the real world. The Albers Equal-Area Projection fulfills this requirement perfectly, and is selected as the projection method for this study.

For the lower 48 states, the USA contiguous Albers Equal-Area Conic Projection is applied to all input GIS datasets. For the states of Alaska and Hawaii, the Albers Alaska and Hawaii Projections are applied, respectively. The Geographic Coordinate System for all 50 states is GCS_North_American_1983.

Once the input GIS data files have been projected into the Albers Equal-Area Projection, the second step of data processing is to clip out all the projected GIS data files by state, by using the GIS "clip" tool. All the above-mentioned GIS boundary data files encompass the whole country[73]; they are used as the input data files of the "clip" procedure. The "clip" procedure segregates boundary data for each individual state.

[73] The NLCD 2001 data is the only exception. It contains 15 subordinate datasets, each of which covers several states. The details of processing NLCD 2001 datasets will be addressed in a later section "NLCD

The major reason for clipping out the GIS boundary data by state is to speed GIS processing. Some of these original datasets are of such a large size that it is highly impractical to conduct the subsequent “union” analysis (which is the most complicated and time-consuming GIS analysis of this study) for the country as a whole[74]. The more pragmatic and efficient method is to first clip out the original boundary data (by state) and then to conduct the “union” analysis (by state), by using the resulting data as the input data files for “union” analysis. Separating the components in this way, the GIS processing time can be dramatically reduced. The GIS “clip” processing is illustrated by the following figure.

Figure 3.1 GIS “Clip” Analysis

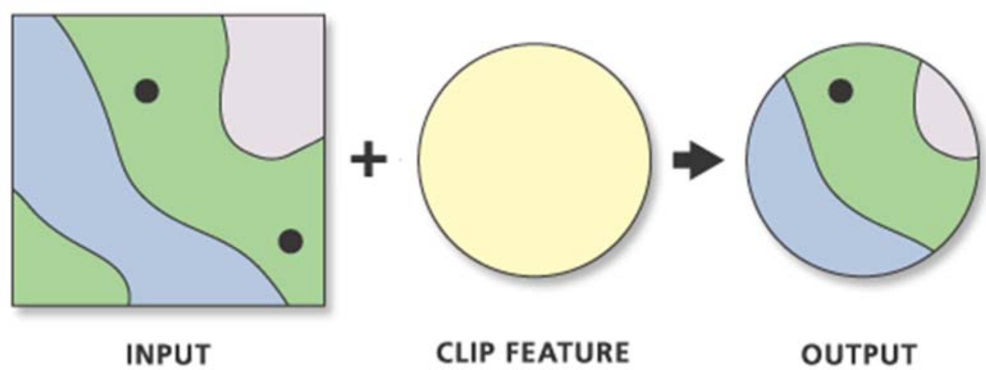


Figure source: ArcGIS Desktop Help file.

During this GIS “clip” processing, the national GIS boundary data files, i.e., the NLCD 2001 data file, the federal land and Indian reservation boundary file, various park boundary files, the national forest ownership boundary file, and the census tract boundary file, are used as the “input feature” data file. The state boundary file is used as the “clip feature” data file. For one “clip” procedure, only one input file can be used; thus, “clip”

Data Processing” in this chapter.
[74] The compute will run the “union” analysis for a long time and then crash.

procedures have been conducted for each of the input data files for each state. For any input data file, the “clip” analysis is conducted 50 times; each time for one state (that is, using the boundary file of that state as the “clip feature” data file), thus resulting in 50 output GIS files.

The state of California is used as an example to explain “clip” analysis. The NLCD 2001 data file (for the zone that contains CA), federal and Indian land boundary file for the US, two park boundary files for the US[75], the national forest ownership boundary file for the US, and the census tract boundary file for the US are all used as separate input features (i.e., each input file for one clip procedure). The state boundary file of California is used as the clip feature. When processed, the GIS “clip” analysis creates an output boundary file for California for each of the abovementioned input data files, producing six output boundary files, which include the NLCD 2001 data for California, federal and Indian land boundary file for California, two park boundary files for California, the national forest ownership boundary file for California, and the census tract boundary file for California. These six output boundary data files are used as the input data files for the GIS “union” analysis for California. These “clip” analyses are then repeated for all US states.

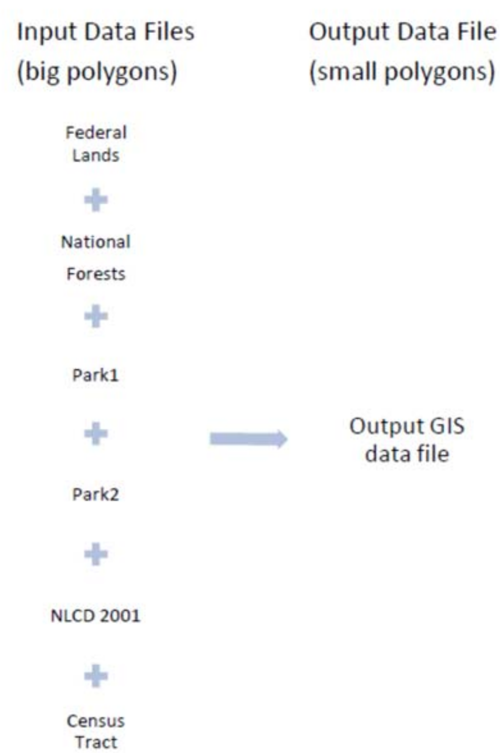
For the third step of data processing, the GIS spatial analysis tool “union”[76] is used to combine together the properties from all input datasets. The NLCD 2001 data, federal and Indian land boundary file, two park boundary files, national forest ownership

[75] These two data files are “U.S. Parks” (1997) and “U.S. GDT Park landmarks” (2002) from U.S. ESRI Data & Maps 2003 CDs.

[76] “The “union” tool computes a geometric intersection of the Input Features. All features will be written to the Output Feature Class with the attributes from the Input Features, which it overlaps.” (ArcGIS 9.2 Desktop Help File)

boundary file, and census tract boundary file in combination constitute the input of the "union" tool. The resulting GIS data (one data file) contains many small polygons resulting from the geometric intersections of all input boundary polygons. The polygon features in the output data file are "small" relative to those polygons in the input data files (which are called "big" polygons). Every resulting small polygon can carry all variables from all its input features[77] (i.e. the input features that resulted in the small polygon through intersection). Figure 3.2 presents the input data files and output data file of the "union" spatial analysis.

Figure 3.2 Input and Output Data Files of Union Analysis



[77] The "Input features" refer to the features that are contained in the input data files. Each feature is a polygon with attributes (variables) that describe the characteristics of the polygon.

The following figure illustrates the "union" processing. In this example, three big input features (polygons) result in five relatively small output features (polygons).

Figure 3.3 GIS Union Analysis

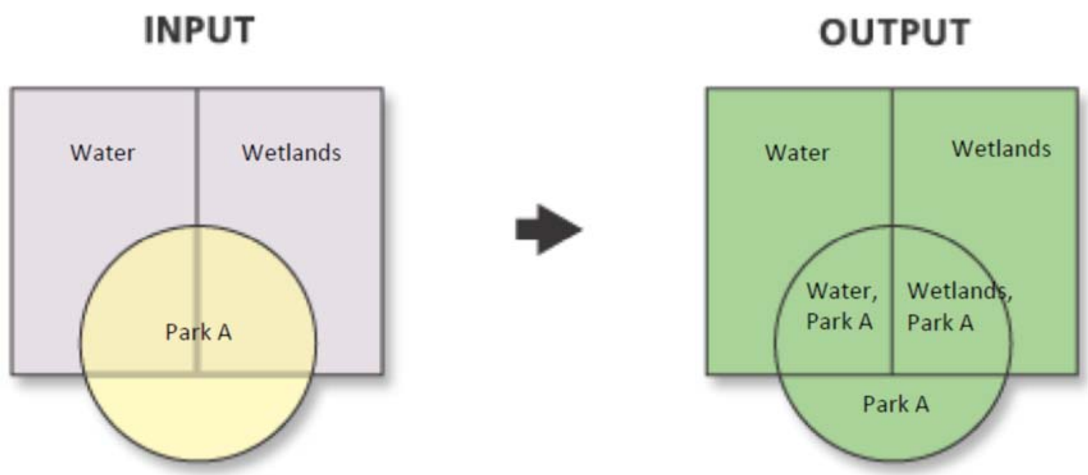


Figure source: ArcGIS 9.2 Desktop Help File. The original figure has been modified by the author.

A simplified example is used to explain the "union" analysis. This is shown in Figure 3.3. In this example, there are only two input GIS data files. One file, the NLCD 2001 file, contains only two polygon features, which have "water" and "wetlands" as the attributes[78], respectively. The other input data, the state park file, contains only one polygon feature, with "Park A" as its value for the attribute variable "name". The polygon of "Park A" overlaps partially with the two polygons from the first data—the "water" and "wetlands" polygons. The "union" analysis computes "a geometric intersection of the input features" (ArcGIS 9.2 Desktop Help File) and thus produces an

---

[78] In ArcGIS 9.2 Desktop software, one digital map has only one attribute table, which contains all variables that describe the characteristics of the spatial features in the map. In an attribute table, a column (or a field) is also called a variable or an attribute. In this case, the values of the variable "type" for these two features are "water" and "wetlands", respectively.

output file containing five relatively small polygon features (compared to the relatively big polygons in the input data) that result from the intersection. The attributes of these output features are shown in Figure 3.3. Since the output features carry all the attributes of the input features, the two output features that represent the overlapped areas of input polygons have attributes "Water, Park A"[79] and "Wetlands, Park A", respectively.

This "union" analysis is conducted for each of the 50 states, resulting in 50 output GIS data files (i.e. with one output data file for each state). For each output data file, amount of area (measured in square miles) is calculated for every polygon feature by using the GIS "calculate geometry" tool. These calculated land areas (measured in square miles), together with other attribute variables of the features, are stored in the attribute table of this output data file. This is the end of the GIS processing, with an output data file that contains relatively small polygons, the result of intersection (compared with the relatively big polygons in the input data files). These polygons represent corresponding land parcels in the real world. The attribute variables of this output data file describe the characteristic features of every polygon, such as whether or not a polygon is truly developable land; its area; the census tract to which it belongs; and so on. All portions of this resultant information are useful when calculating the amount of developable lands for counties.

The next step of data processing is to export the attribute tables of the aforementioned 50 GIS output data files to SAS software for area summarization and subsequent statistical analyses. These 50 attribute tables are first exported from ArcGIS

[79] In detail, for this polygon, its attribute variable "Type" has a value of "Water" and the variable "name" has a value of "Park A".

9.2 software into ".dbf" tables; and then these ".dbf" tables are imported into SAS software and stored as SAS datasets[80].

In the SAS environment, the imported attribute tables are processed to calculate the area of developable lands for counties and states. First, the polygons characterized as "developable lands" are queried out for each state data file. These polygons represent the lands that do not belong to any of the following kinds of areas: naturally undevelopable areas, federal lands, American Indian reservations, state parks and forests, and local parks and recreation areas. Second, the areas of these developable land polygons are summed up by county for each state. Third, the 50 attribute tables for states, containing the total area of developable lands for each county, are combined into one large table that contains the area of developable lands, as well as other attribute variables, for all 3091 US counties, with each record (or row) representing one county. Based on this table, the area of developable lands for all states is summarized.

In this section an explanation of the general, basic data processing steps to calculate areas of developable lands for counties/states has been undertaken. In addition to these basic steps, some specific data processing methods have been developed and employed to resolve problems due to special cases. These methods, as well as the special cases themselves, are addressed in the following section.

Special Problems in Data Processing

The previous steps describe the general process used in this study to calculate amounts of developable land. Some special cases make this process more complicated

---

[80] All SAS datasets have a name with a ".dat" file extension.

and require special methods to be developed. There are mainly three kinds of special cases where this applies: 1) where NLCD data files have overlapping regions; 2) where some state forests and parks contain large areas of privately-owned lands; and 3) when non- American Indian households live on Indian reservations. These special cases, as well as the special methods used to resolve these problems, are addressed below.

a. NLCD Data Processing

The first special case is when the NLCD data files have regions that are overlapped, which makes the GIS data processing more complicated. As mentioned earlier, the vector data of NLCD undevelopable areas needs to be clipped out by state boundaries, because the subsequent GIS "union" spatial analysis is conducted by state. However, during this clip-out step, two problems arise due to the way that the NLCD datasets are organized. The NLCD datasets comprise 15 subordinate data files—the whole country territory[81] is divided into 15 geographic zones, with each zone being contained in a subordinate data file. The boundary of a zone may not be consistent with state boundaries. In other words, a zone may consist of several whole states and only portions of some other states. That is, a state may cross over two or more zones.

The problem occurs when a state crosses multiple NLCD data zones; it is compounded if these multiple zones overlap. For example, the State of Louisiana falls in two zones, zone 10 and 12; furthermore, these two zones (zone 10 and zone 12) overlap with each other. In order to solve this problem, first every portion of the state is clipped out from all NLCD zones that contain this state (in this case, zones 10 and 12), and then

---

[81] During the time of my data processing, the NLCD land cover dataset for Hawaii is not yet available on their website.

the multiple clipped-out parts are merged together. After merging, the duplicate NLCD polygons, resulting from the overlapped parts between the two NLCD zones, are eliminated. If this is not done, the summed lands of all the NLCD polygons within the state will be larger than its total lands, due to the inclusion of duplicate NLCD polygons. In order to fulfill this task, the GIS spatial analysis "union" tool, instead of the "merge" tool[82], is used to combine together all NLCD data portions of a state that are located in multiple NLCD data zones. The output data of the "union" analysis carries all the features and attributes from the input data files, including the identification of NLCD zones (in this case, the identification "zone 10" and "zone 12"); and thus the duplicate polygons can be identified by zone and eliminated.

After this special processing, one complete NLCD data file for a state is used as one of the input data files for the GIS spatial analysis "union" to create the final dataset.

b. State Forests/Parks

Similar to the ESRI national forest boundary data, the ESRI state forest/park boundary data also does not any depict privately-owned lands located within a forest/park boundary, which makes it impossible to identify these privately-owned lands using this data. This problem is especially significant for certain states, Minnesota and New York being examples. In the state of Minnesota, The Richard J. Dorer Memorial Hardwood State Forest covers nearly two million acres of land, of which only 2.3 percent is state-owned and 97.7 percent is privately-owned. In the state of New York, nearly 52 percent of the 6 million acres that comprise Adirondack Park is privately owned. For these states,

[82] In "merge" tool, the output data does not carry the attributes from the input data; thus the duplicate polygons cannot be eliminated.

if these whole state forests/parks are identified as undevelopable areas, then the privately-owned areas within the forest/park boundaries, which are actually developable lands, would be incorrectly excluded from the density calculation.

In order to correctly calculate the household density for a county that contains state forests/parks, the amount of privately-owned lands within the state forests/parks needs to be calculated, and further, be categorized as developable lands in the density calculation. In order to fulfill this task, the amount (in acres) of state-owned lands within a state forest/park, obtained from the official website of these state forests/parks, is used. The method of using this supplemental data to calculate the developable lands within a state forest/park is described in Appendix 1.

After this special processing, the amount of developable lands within state forests/parks is calculated and further used as a component of the denominator for the density calculation.

c. American Indian Reservations

Indian reservations must be treated cautiously when calculating household density for states and counties because their residential densities are usually much lower than the densities of non-reservation regions. Therefore, in this study, Indian reservations are classified as “legally undevelopable lands” and excluded from household density calculation.

In addition, corresponding to the exclusion of Indian reservation areas from the denominator of density calculation, American Indian households living on these reservations are simultaneously excluded from the numerator of density calculation. That is, the Indian reservation lands, as well as the American Indian households living on them, are excluded from the density calculation. In this study, the "American Indian household" is defined as a household with an American Indian householder; the "non-American Indian household" is defined as a household with a non- American Indian householder[83].

In order to calculate accurate densities, both the denominator (developable lands) and the numerator (non- American Indian households) are calculated first at the census tract level (rather than at the county level), and are summed for the whole county. The county density is then calculated as the sum of non- American Indian households of the county divided by the total undevelopable lands of the county.

In this study, there are two types of census tracts: the first does not contain Indian reservations. For these, the numerator of the density calculation is the total number of households living in the census tract; and the denominator is the amount of "developable lands" within the census tract. In other words, since no Indian reservation lands are excluded from the denominator, the corresponding American Indian households are also not excluded from the numerator.

The other type of census tract does contain Indian reservations. Corresponding to the exclusion of Indian reservation lands (as a type of legally undevelopable land) from the denominator of the density calculation, the number of "non-American Indian households" living in the census tract (calculated as total households minus the number of American Indian households) is used as the numerator of density calculation.

Using this method, all land within the Indian reservation boundaries were originally excluded from the denominator of density calculation, with an assumption that

---

[83] Such a household may have Indian American as a non-householder member.

all non- American Indian households would live outside Indian reservations. However, this assumption is not true in the real world. Therefore, the denominator “developable lands” in the density calculation is adjusted to include a portion of lands within Indian reservations. This adjustment is explained below.

Since some non- American Indian households live on Indian reservations, the exclusion of all areas within Indian reservation boundaries will result in an upward biased household density for some counties. For example, for Osage County, Oklahoma, 97.7 percent of its total county lands (2304 square miles) are occupied by within Indian reservations; there were 14,691 non- American Indian households and 2,028 American Indian households recorded living in the county in the year 2000. If all Indian reservation lands are excluded from the density calculation, then the resultant household density of Osage County is 9,508 households per square mile[84], even denser than that of San Francisco County, CA (8,464 households per square mile). Clearly, the real household density in Osage County is not as high as 9,508 households per square mile. This density value is upwardly biased, due to the wrong assumption that all non- American Indian households live outside Indian reservations.

In order to solve this problem, not all Indian reservation lands are excluded from the denominator of the density calculation. Rather, because the numerator of the density calculation includes those non- American Indian households living on reservations, the denominator should accordingly include the amount of lands within Indian reservation boundaries on which these non- American Indian households reside. These lands are named as the “adjusted developable lands within Indian reservations” in this study and

[84] This density is calculated as 14,691 households divided by the area of developable lands.

used as a component of developable lands in the density calculation. In the case of Osage County, for instance, after this adjustment is taken, it has a much lower density of 7.7 households per square mile; and therefore, based on this low density, its existing land use type in 2000 classifies it as an "undeveloped" county. The detailed steps to calculate these "adjusted developable lands within Indian reservations" are described in Appendix 1.

To simplify, in the following sections household density is calculated as "non- American Indian households" divided by "developable lands". As explained in the foregoing section, "non- American Indian households" actually include two components: one is the non- American Indian households living in census tracts on Indian reservations; the other is total households living in census tracts off Indian reservations. In addition, "developable lands" include the abovementioned "adjusted developable lands within Indian Reservations".

Checking Process

Some mistakes may occur during the GIS processing and the steps of transferring the output GIS data files from GIS software to SAS software. This is mainly due to the large size of the GIS data files and unstable software performance. For some states, the NLCD dataset has a size of two gigabytes, with more than 2 million records. Therefore, ongoing checks of the resulting SAS datasets are routine to ensure that no mistake occurs during the data processing.

The method used to check for any mistakes is to compare the input GIS data files of the GIS "union" analysis with the initial SAS datasets. If these two data files match each other, then it is determined that no mistakes occurred during the data processing. As explained before, the "initial SAS datasets" are attribute tables of the output GIS data files of the "union" spatial analysis. They are exported from GIS software to SAS software and become the "initial SAS datasets". Because all GIS processing is conducted by state, checking is also conducted by state for every input data file.

As illustrated in Figure 3.2, the features in the input GIS data files of "union" analysis are relatively big polygons, and the features in the output GIS data file are relatively small polygons carrying the attributes of all input features. In the checking process, the total land of the initial big polygons[85] in an input GIS data file is compared with the total land of the resulting, small polygons in the initial SAS dataset. If the values of these two areas match each other, no mistakes have been produced during data processing.

As an example, for the state of New Jersey, first the Indian reservation polygons are checked. The total area of Indian reservation polygons is calculated for the input data file, the "Federal and Indian Lands" GIS boundary data file. Then for the initial SAS table for the state of New Jersey, the lands are summed for all the records[86] (i.e., rows) that are characterized as Indian reservation. The checking shows that these two summarized areas are equal to each other—therefore it is concluded that, for Indian reservation polygons, no mistakes have been created during data processing.

Subsequently for the state of New Jersey, the same checking process is performed for all other input datasets, including federal lands, state forests, state and local parks, and

---

[85] The polygons in an input dataset of "union" analysis are called "big" polygons; they are "big" relative to the polygons in the resulting output GIS datasets (which are called "small" polygons). Therefore, here "big polygons" and "small polygons" are used to describe the input polygons and the resulting output polygons of the "union" analysis, respectively.

[86] Each record represents an abovementioned "small" polygon that carries all attributes of input features.

census tract boundary files, the national forest ownership file, and NLCD 2001 naturally undevelopable area polygons. Then the same checking process is conducted for all 50 US states.

For all GIS input datasets, cross checking shows that they match the initial SAS datasets. This means that no mistakes have occurred either during the GIS processing or the process of transferring the output GIS data from GIS software to SAS. Therefore, the initial SAS datasets, which contain the area (in square miles) for every polygon of developable lands, are correct and ready for subsequent data analyses.

3.3.3 Household Density Calculations

Once the checking process ensures that the initial SAS datasets, which contain areas of polygons for developable lands, are correct, then the subsequent analyses can be conducted in an SAS environment. These analyses include the calculation of household density for states/counties, classification of the existing county land use types, and categorization of sprawl/non-sprawl for a county. This section describes the first substantive analysis, the calculation of household densities in 2000 for a county/state. The density is calculated for the year 2000 because this is the base year of the time period 2000 to 2020, for which sprawl is measured.

As the first step of the density calculation, the denominator for the density, i.e. the area (in square miles) of developable lands for a county/state, is calculated. This is achieved by summarizing the areas (in square miles) of all developable land polygons for a county/state. In sequence, first the areas of developable lands are summarized for each census tract[87], further summarized for a county, and then lastly, for a state. This aggregated area (in square miles) of developable lands for a county/state is used as the denominator of the county's/state's household density calculation.

Second, the numerator of the household density calculation, the number of Non-American Indian households living in a county/state, is calculated. The calculation is performed first at the census tract level. In sequence, for the census tracts that contain Indian reservations, the number of non- American Indian households (calculated as total households minus the number of American Indian households) is calculated, in response to the exclusion of Indian reservation areas[88] from density calculation. For census tracts that do not contain Indian reservations, total households are used as a component of the numerator. At this point, the numbers of households at the census tract level are summed for a county or state. For census tracts that contain Indian reservations, non-American Indian households are summed; for census tracts that do not contain Indian reservations, total households are summed for the county or state.

This aggregated number of households is used as the numerator of the county/state household density calculation. It is referred to as "non- American Indian households", with an emphasis on the exclusion of American Indian households from the numerator, although it is total households for most census tracts.

---

[87] As explained in the foregoing section, for the census tracts that contain Indian Reservations, the "Adjusted Developable Lands", instead of the original developable lands, are used in this process of aggregation.

[88] As stated in the foregoing section, only part of the Indian Reservation area is excluded from the denominator of the household density calculation; in other words, a portion of the Indian Reservation lands is included in the denominator, because it often happens that some non- American Indian households live within Indian Reservations.

Household density for a county is calculated as the number of non- American Indian households[89] divided by the total area (in square miles) of developable lands; the same calculation method is applied to a state for state household density calculation. In the subsequent analysis, one of six county land use types is assigned to each county, based on the county and state densities calculated in this section. The methods necessary to complete this task are described in the following section.

3.4 Classifying County Land Use Type

Apart from most existing studies of sprawl measurement, residential density is not used as a direct measurement of sprawl. Rather, it is used as the criterion for classifying the existing county land use types, upon which the sprawl classifications are based. The method employed to classify the county land use types is described in this section. For some counties, the original county land use type is adjusted according to certain criteria in order to better reflect the land-use pattern of the county in the real world.

3.4.1 County Land Use Type Based on Density

Based on the household densities of states and counties, one of six existing land use types (urban center, urban, suburban, rural center, rural, or undeveloped) is assigned to each county. This task is undertaken in several steps.

First, all 50 US states are categorized into four density groups (very low, low, moderate, and high density states) based on their density. All states are sorted by their

[89] As defined in the foregoing section, for the census tracts that contain no Indian reservations, the "non-American Indian households" actually refer to the total households of the census tract. This concept is the same for all "non- American Indian households" used in the following sections.

residential density in 2000 in ascending order; then three obvious density gaps (i.e. density breaks) are observed for the sorted state densities. Based on these density gaps, the 50 states are classified into four density groupings, with significant density gaps between each density group. The names of states in each density group are presented in chapter 4.

The following table presents the density threshold ranges for the four density groups, as well as the number of states in each group. Column A and B present the observed minimum and maximum state density in each density group, respectively. Together, these two columns show the density range in each density group, as well as the density gap between each density group. Based on these two columns, Column C lists the designated density range for each density group.

Table 3.3 Density Thresholds for State Density Groups (in Households/Sq. Mile)

Density Group of States	Minimum Density (A)	Maximum Density (B)	Designated Density Range (C)	Number of States (D)	Examples
Very Low	1.2	15.9	1-18	9	AK, WY, MO
Low	19.2	56.8	19-59	19	OR, CO, AZ
Moderate	60.3	121	60-149	13	GA, NH, TE
High	152.9	539.9	>=150	9	CA, FL, NJ

Second, within each state density group, all counties that make up those states are sorted in ascending order by their residential densities in 2000; then they are classified into six county land use categories by using density gaps among counties and known counties as references. The six county land use types are urban center, urban, suburban, rural center, rural, and undeveloped counties, which are in a descending order of development density (that is, urban center counties represent the highest density of development in the region, while undeveloped counties are the lowest density of development in the region).

Here, the aforementioned variable-density approach[90] is applied to the classification of county land use types. According to this approach, the same density value may represent different land-use types in states with very different densities; and regional context must be taken into account when establishing land-use types for counties based on their densities. In this study, regional context is represented by the host-state density. Under this approach, the density range of a particular county land use type varies with the host-state density. Thus four sets of density thresholds (rather than one set of density thresholds) are developed for the six existing county land use types according to the density of their state. This is shown in table 3.4.

Table 3.4 presents the density thresholds of six county land use types for each state density group. Each column presents the six density ranges for counties within one of the four state density groups. Each row shows that the density range for a particular county land use type varies with the state density, based on the variable-density approach. For example, the density range for a rural county in a high-density state (from 34 to 158 households per square mile) is six times higher than that for a rural county in a very low-density state (from 6 to 20 households per square mile). For an urban county in a high density state, the density range (658 – 1240) may also be six times higher than the density range in a very low density state (116 – 300).

[90] This approach has been defined in Chapter 2 Literature Review.

Table 3.4 Density Thresholds for County Land use Types in 2000 by State Density Groups (in Households/Sq. Mile)

County Land-Use Type	Very Low-Density States	Low-Density States	Moderate-Density States	High-Density States
Undeveloped	< 6	< 25.5	< 30.5	< 34
Rural	6 – 20[a]	25 - 70	30.5 - 115	34 -158
Rural Center	20 - 27	70 - 90	115 - 169	158 - 267
Suburban	27 - 116	90 - 363	169 - 455	267 - 658
Urban	116 - 300	363 - 773	455 - 1199	658 - 1240
Urban Center	>= 300	>= 773	>= 1199	>= 1240

Note:

a. The range "6 – 20" represents "equal or greater than 6 and less than 20" households per square mile. This denotation is the same for other ranges in this table.

3.4.2 County Land Use Type Adjustment

Thus far, six county existing land-use types have been assigned to counties based on county and state densities. This is the "original" county land use type, the one that exists before adjustment. In addition to residential densities, two other factors are taken into account when assigning a particular land use type to a county. According to these two factors, the original county land use types of some (not all) counties are adjusted to different land use types[91].

The first situation that requires adjustment of the land use type is the case where an EA only contains rural and/or undeveloped counties (that is, it has no urban features). Given this situation, then one or more counties in this EA that meet certain criteria will

---

[91] For those counties to which the adjustments are not applied, their original county land use types are just the final ones.

be adjusted to rural center counties. To simplify, this kind of economic area is entitled as "Rural EA" in this study. Why must some rural counties in a "Rural EA" be adjusted to a rural center county? This study agrees with Burchell et al. (2002) that there must be at least one center county (either economic or residential) within each EA to which future growth in the EA could be directed. Future growth can be encouraged to take place in center-orientated counties, where there are already sufficient facilities to accommodate it. According to the definition of sprawl (Burchell et al. 2002), the rural center county is the lowest level of county type where the growth could be encouraged or reallocated to[92]. Therefore, at least one rural center county must be designated in each "Rural EA".

While this first factor mainly affects "Rural EAs", the second factor may affect all EAs. According to the definition of an economic area (by BEA), each EA must have one or more economic nodes (metropolitan areas or similar areas) that act as the economic activity centers of the EA. In this study, the counties that contain the center city (or cities) of these economic nodes are called "economic center counties". Some economic center counties may not have high residential densities—this could result in an under-classified county type. These economic center counties are adjusted to the highest level of land use types existing in that EA[93].

Aimed at resolving the problems caused by these two factors, three general rules are employed to adjust original to final county types. If a county meets either of the following two criteria and its original county type is not the highest level of development

---

[92] I agree with Burchell et al. (2002) that all of the urban center and urban counties, and some of the suburban and rural center counties are the counties to which future growth should be directed.

[93] Here, the highest level of land use types refers to the particular highest level of original county land use types within that EA. For example, the original county land use types for counties in one particular EA are suburban, rural center, and rural counties; then the highest level of county types in this EA is "suburban".

type within the EA, then its original county type is adjusted to the highest level of the county type in the EA[94]: 1) its density is above the county density gap[95] within the EA; and 2) the county is the "economic center county" of the EA. 3) In addition to these two rules, the third criterion is that the work team's practical knowledge of the counties validates the adjustment. A county may be adjusted to a higher or lower level of county land use type (relative to its original type) by using the third rule. An example explaining application of these rules is given in Appendix 1.

In total, the land use type of 75 counties in 53 EAs was adjusted and the adjusted types were used as the final land use types. For the other 3016 counties, their original types, which are density-based, are used as their final land use types. After this adjustment, of 3091 U.S. counties in the year 2000, 84 percent (2583 counties) are undeveloped and rural counties, 13 percent (395 counties) suburban and rural center counties, and less than 4 percent (113 counties) urban and urban center counties. The result of classification of county land use types is presented in detail in the following chapter.

With county land use type assigned to counties, the last step in sprawl measurement, sprawl/non-sprawl classification, can now be made. This procedure is addressed below.

3.5 Classifying Sprawl/Non-Sprawl for Counties

---

[94] The exception is for the counties in the 33 "Rural EAs". For these EAs, the highest level of original county land use type is rural or undeveloped county. Therefore, for these EAs, the rural center county, instead of the highest level of original county land use types, is assigned to counties that meet either of the first two criteria.

[95] All the counties within an EA are sorted in descending order by density. The first density gap/break is called the "county density gap". The counties with a density higher than this density gap are the top-density counties in the EA, and usually, they are the most developed counties in the EA.

With counties classified according to their development levels and growth during the 2000 to 2020 period calculated[96], nationwide locations of sprawl at the county level can be determined. As mentioned in the previous section, this study borrows Burchell et al.'s empirical definition of sprawl and reinforces the statement that sprawl will predominantly happen in "nonurban" locations (Burchell et al. 2002, p. 3.). By definition, this study acknowledges that all urban center and urban counties in 2000 will not be subject to sprawl growth during the 2000 to 2020 period.

Sprawl development may only happen in sparsely developed areas without significant infrastructure. These are rural and undeveloped counties, and some developing suburban and developing rural center counties. In this study, the criteria for sprawl classification are somewhat different between these two groups of counties: the first group contains rural and undeveloped counties; and the second group, suburban and rural center counties. The classification criteria for the latter are "looser" than those for the former group. While all rural and undeveloped counties are potential locations of sprawl growth, only a portion of the suburban and rural center counties are candidates. Specifically, these looser criteria are used to screen out from sprawl categorization the developed suburban and developed rural center counties that have sufficient infrastructure to contain future growth. These counties are not candidates for sprawl designation; given their infrastructure, they are classified as non-sprawl even if they experience significant growth in the future, because these are the locations to which future growth is directed. Only those developing suburban and rural center counties, which are less established and have less infrastructure, are possible locations for future

[96] As mentioned in the foregoing section, the households and employment growth of a county is calculated using the Woods and Poole dataset (the 2003 edition).

sprawl growth. In the following section, the "tighter" criteria of sprawl classification for rural and undeveloped counties are addressed first, and then the "looser" criteria for suburban and rural center counties.

3.5.1 Classifying Sprawl/Non-Sprawl for Rural and Undeveloped Counties

In this study, a rural or undeveloped county will experience sprawl during the 2000 to 2020 time period if *either* of the following sets of criteria is met:

1. (a) The county's annual growth rate (of households, employment, or both) is in the upper quartile of the EA's[97] annual county growth rates (households, employment, or both)[98]; (b) the county's annual growth rate exceeds the average annual national county growth rate[99]; and (c) the county's absolute level of growth exceeds 40 percent of the average annual absolute national county growth[100].

or

2. The county's absolute level of growth exceeds 160 percent of the average annual absolute national county growth. (Burchell et al. 2002, p. 3, p. 572)

In this definition, EAs, rather than metropolitan areas, are used as the host regions of counties because the 172 EAs bind urban, suburban, and rural counties and include all

---

[97] Totally there are 172 EAs (Economic Areas) in the U.S. as defined by BEA (Bureau of Economic Analysis).

[98] Here, the annual growth rate is calculated by dividing a county's growth rate during the period 2000 to 2020 by 20 years. And the upper quartile value of the EA's annual county growth rates is calculated based on all the counties within the EA.

[99] The average annual national county growth rate is calculated as the mean of all 3091 counties' annual growth rates, with both the positive and negative county annual growth rates being included in the mean calculation.

[100] The average annual absolute national county growth is calculated as the mean of all 3091 counties' annual absolute growth, with both positive and negative county growth being included in the mean calculation. The annual absolute county growth is calculated by dividing the absolute growth of a county during the 2000 to 2020 period by 20 years.

counties in the country while metropolitan areas include no rural counties and only 836 of the 3091 counties.

The thresholds in this definition are borrowed from Burchell's work (Burchell 2002). Numerous analyses of the population growth rate define the 75 percentile as a separator between significantly rapid and average growth. The '40 percent of average absolute national county growth' acts to eliminate those counties whose rapid growth rates are due to large increases to a small base. The '160 percent of average absolute national county growth' acts to include those suburban and rural center counties where the growth rates would never exceed the thresholds due to their large base population. (Burchell 2002) Burchell's work team established these thresholds based on their considerable and exacting study of theoretical and practical knowledge.

A rural or undeveloped county that will experience significant household or employment growth in the future is defined as sprawling. The first criterion primarily captures the relative pace of growth in a county, while the second criterion captures the occurrence of "significant" absolute growth in counties. These two criteria are also used to define "significant growth". A county meeting either of these two criteria is said to experience "significant growth". If projected "significant growth" will take place in a *rural or undeveloped* county, then this county is defined as sprawling. If "significant growth" will take place in a *suburban or rural center* county (assuming some infrastructure), then this county *may or may not* be defined as sprawl, which is explained in the next section.

3.5.2 Classifying Sprawl/Non-Sprawl for Suburban and Rural Center Counties

A suburban or rural center county is classified as sprawl if *both* of the following criteria are met:

1. It will experience "significant growth" during the 2000 to 2020 period (i.e., it meets either of the abovementioned two criteria of "significant growth"—upper quartile growth rate and above average absolute growth);

and

2. Its urban population percentage share[101] is below 85% (applied to suburban counties) or 80% (applied to rural center counties) in 2000.

The second criterion makes the set of criteria "looser" when compared with the criteria for rural and undeveloped counties. In fact, here the high urban population percentage share is used as a criterion to identify the abovementioned developed suburban and rural center counties (with an urban percentage share above 85% and 80%, respectively), to which we want to direct future growth. Developed suburban or rural center counties, which actually have land-use types close to those of urban counties, are

---

[101] Urban population percentage is defined as the urban population of a county divided by its total population. The data of urban population and total population is from the Census 2000 data. "The U.S. Census Bureau classifies as urban all territory, population, and housing units located within urbanized areas (UAs) and urban clusters (UCs). It delineates UA and UC boundaries to encompass densely settled territory, which generally consists of:

• A cluster of one or more block groups or census blocks each of which has a population density of at least 1,000 people per square mile at the time, and

• Surrounding block groups and census blocks each of which has a population density of at least 500 people per square mile at the time, and

• Less densely settled blocks that form enclaves or indentations, or are used to connect discontiguous areas with qualifying densities.

Rural consists of all territory, population, and housing units located outside of UAs and UCs." (Summary File 1 2000 Census of Population and Housing Technical Documentation pp. A-22 – A-23)

Due to the limited time and resource, this study did not define urban and rural areas using the more accurate density calculation developed in this project. Instead, the Census 2000 definitions of urban and rural, which are based on less accurate density calculation, are used by necessity.

classified as non-sprawl, even if they will experience significant growth during the 2000 to 2020 time period. Therefore, only developing suburban and rural center counties (with an urban percentage share below 85% and 80%, respectively) that will experience significant growth are classified as sprawl counties.

3.5.3 Sprawl Adjustments

Finally, for some particular suburban and rural center counties that have been classified as sprawl by the listed criteria, the sprawl classification is further negated by applying the following rules of sprawl adjustment. Such a county will be adjusted from sprawl to non-sprawl if *any* of the following criteria[102] is met:

(1) It is one of the EA's highest-density counties (with a density higher than the abovementioned "county density gap" in the EA); *or*

(2) it is the county that contains the largest economic node in the EA[103]; *or*

(3) according to practical knowledge of the counties in the EA, the county should be categorized as non-sprawl.

The reason to adjust these counties from sprawl to non-sprawl is due to the requirement that for each EA, there must be some appropriate location(s) where projected future growth is encouraged to take place. These locations are the most developed counties that have sufficient infrastructure to contain future growth. In this analysis, if the counties with the highest county land use type in the EA are urban center or urban

---

[102] Actually, as mentioned in the foregoing section of classification of county land use types, the counties that meet either of the first two criteria have been classified as/adjusted to the highest level of the original county types in the EA.

[103] Here, the "largest economic node" actually refers to the largest center city among economic nodes in the EA. For all the EAs in the sample, the suburban and rural center counties that contain the second largest node are originally classified as non-sprawl by the aforementioned two criteria and thus need not to be adjusted.

counties, or, are suburban or rural center counties that were originally classified as non-sprawl, then the requirement is fulfilled (that is, these counties are defined as appropriate locations to which future growth can be directed) and there is no need to adjust their original “non-sprawl” classification. This is the case for most of the EAs in this study.

However, for some EAs, wherein the highest county type is the suburban or rural center county designation and these counties were originally classified as sprawl, then a problem exists. That is, being classified as sprawl means that these counties, which are the most developed counties in the EA, are identified as inappropriate locations to contain significant future growth. The question is then, to which location(s) should future growth for this EA be directed? In order to solve this problem, in these EAs, the most developed counties that have been originally classified as “sprawl” are adjusted to the classification of “non-sprawl” by using the abovementioned criteria of sprawl adjustments[104]. Being adjusted to non-sprawl defines these counties (in the real world) as appropriate locations to which future growth in the EA can be directed. The examples that explain the sprawl adjustments are given in Appendix 1.

Sprawl adjustments are only applied to the suburban or rural center counties; that is, rural or undeveloped counties are not adjusted for sprawl classification, because they are never the most developed counties in the EA[105]. In total, 47 counties in 41 EAs have been adjusted from their original classification of “sprawl” to a final classification of “non-sprawl”. Among these 47 counties, 26 meet both of the first two criteria; seven meet the

---

[104] The counties that meet the third criterion may not be the most developed county in the EA (i.e., the county with the highest development type in the EA), but should be the locations to which future growth be directed.

[105] The highest level of county types in an EA can only be rural center or above levels, according to the rules of county type adjustments.

first criterion only; 12 meet the second criterion only; and the remaining two counties satisfy the third criterion.

After processing this adjusted data, all 3091 U.S. counties are classified into two categories: sprawl and non-sprawl counties.16 percent (492 counties) of US counties will experience sprawl during the 2000 to 2020 time period. The results created during the sprawl measurements will be presented in the next chapter.

3.6 Summary

The datasets and methods employed to classify counties as sprawl or non-sprawl are described in this chapter. Three major tasks have been fulfilled during the process of measuring sprawl for counties.

First, refined residential densities are calculated for all US counties and states. Various GIS boundary data files are used to calculate areas of developable lands for counties and states, through a series of GIS spatial analyses. These areas of developable lands are used as the denominator of density calculation. Then the numbers of total households (or non- American Indian households for tracts that contain Indian reservations) are summed by census tracts for counties and states; this becomes the numerator of the density calculation. Based on these denominators and numerators, refined household densities are calculated for counties and states.

Second, based on these refined densities, one of six existing land-use types (urban center, urban, suburban, rural center, rural, and undeveloped) is assigned to every county. The variable-density approach is the basis of this process, in that 50 US states are first categorized into four density groups (very low, low, moderate, and high density states) and then one set of density thresholds of six county land use types is specified for each of the four state groups. In other words, the density range of county land use type varies with and parallels host-state densities. The original classifications of county land use types, which are density-based, have been adjusted for 75 counties. As a result, of the 3091 US counties: 84 percent (2583 counties) are undeveloped and rural counties; 13 percent (395 counties) are suburban and rural center counties; and less than 4 percent (113 counties) are urban and urban center counties.

Finally, the locations of sprawl are identified at the county level nationwide. Based on the county land use types and the projected growth of counties during the 2000 to 2020 time period, all US counties are classified as sprawl or non-sprawl. A rural or undeveloped county is identified as sprawl if it will experience significant household or employment growth during this time period. Some developing suburban or rural center counties that satisfy certain criteria have also been classified as sprawl. In order to ensure that the most developed counties in an EA are classified appropriately, 47 counties have been adjusted from their original classification of “sprawl” to “non-sprawl”. As a result, 492 of the total 3091 US counties will experience sprawl development during the 2000 to 2020 time period.

The results created during the process of sprawl measurements are presented in the next chapter.

Chapter 4

Empirical Findings—Results of the Analysis

The results collected and identified during the process of sprawl measurements are presented in this chapter. They have been categorized and placed into four main groupings: refined densities for states and counties; land use types (urban center, urban, suburban, rural center, rural, and undeveloped) assigned to each county; projected growth over the 2000 to 2020 period; and finally, sprawl growth at various geographic levels. These results are addressed in detail in this chapter.

4.1 State Density Groups

According to the variable-density approach, regional density differences of counties must be taken into account when defining land use types for specific counties. In this study, regional differences of counties are represented by the residential density of the host state, which reflects the social, economical, and historical development of the state. States are classified into four density groups (very-low, low, moderate, and high), which represent the four kinds of regional density and are the basis for classifying county land use types.

Further, as addressed in Chapter 3, the densities calculated in this study are more accurate than the densities in previous studies, in that naturally and legally undevelopable areas have been excluded from the density calculation.

State Density Groups

Overall, the density of the United States as a whole is 46.3 households per square mile[106] as determined for the year 2000. State densities vary considerably, from a minimum of 1.2 households per square mile in Alaska to a maximum of 540 households per square mile in New Jersey. The top five states, with a density higher than 200 households per square mile, are New Jersey, Rhode Island, Massachusetts, and Connecticut, and Maryland (in descending order). The median state density is about 48 households per square mile (the average densities of Wisconsin and Arizona), very close to the average density of the United States (46.3 households per square mile). The densities of the 50 US states are presented in Table 4.1.

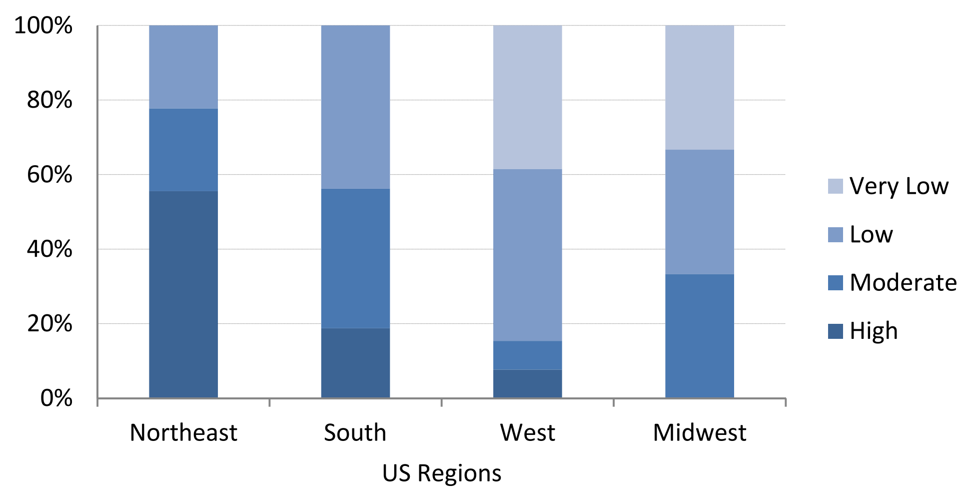
As mentioned in chapter 3, the 50 US states are categorized into four natural density groups: very-low, low, moderate, and high. The density thresholds of these groups are shown in Table 3.3. Column D of Table 3.3 presents the number of states within each density group. Corresponding to the generally low density of the United States, 40 percent of US states (19 states) are in the low-density group, followed by one-quarter (13 states) in the moderate range.

The distribution of states among the four density groups for each census region is shown in Figure 4.1. Each 100% stack column represents the four state density classes' percentage share of all states in that region. At the regional level, the South has the most states (16), followed by the West (13), the Midwest (12), and the Northeast (9). Not surprisingly, as shown in Figure 4.1, the Northeast and the South have no "very low-

[106] This household density for the U.S. as a whole is calculated as the total households (with American Indian households living in the Indian reservations excluded) divided by the total developable lands within the country.

density" states; in contrast, the Midwest has no high-density states. Most (nearly 80 percent) of the Northeastern states have high- or moderate- densities. In contrast, the West and Midwest are overall characterized by very low- and low- density states: nearly 85 percent of Western states and 70 percent of Midwestern states belong to these two density groups. For the South, most states (over 80 percent) have a moderate- or low-density.

Figure 4.1 Distributions of States among Density Groups by Region



Differences between Refined and Gross Densities

Is it really necessary to take a cautionary approach to calculate more refined densities, rather than simply calculating gross densities based on the ready-to-use land areas? This question is addressed in two parts. First, are the values of refined densities significantly different from those of gross densities? Second, will the differences between these two kinds of densities affect the assignment of county land use types? If the answer is "Yes" to both questions, then it is definitely necessary to calculate specific densities in a study whose task is determining accurate land-use types.

The densities calculated in this study will be referred to as “refined density” or “developable lands density” (because its denominator is the developable land area in a geographic unit). Conversely, the denominator of gross density is the area of non-water lands (i.e., the total area minus the area of any water body) within a geographic unit; this gross density is the one generally used in most existing sprawl studies. Therefore, for the same geographic unit, the denominator of its refined density is always less than or equal to that of its gross density, because developable lands, as defined in this study, amount to a share of non-water lands. The answers to the two questions are presented below, supported by the empirical results developed in this study.

The analysis begins with why care is needed in determining densities: are the values of refined densities significantly different from those of gross densities? The answer is yes. With no exception, the “developable lands density” of an individual state is larger than its “gross density”, because the former has a much smaller denominator than the latter. A substantial proportion of US states have “developable lands (or refined) densities” that are significantly larger than their gross densities. As shown in Table 4.1, about 65 percent (32 states) have “developable lands densities” (column D) that are at least 1.2 times their gross densities (column A). Further, nearly one-third (15 states) have refined densities that are at least 1.5 times their gross densities, due to the denominator of the former being only about 65 percent of the denominator of the latter (shown in column I), which means that the areas of developable lands are only around 65 percent of the non-water lands in these states. Table 4.1 presents the two kinds of densities for all 50 US states.

Table 4.1 Comparison between Developable Lands Densities and Gross Densities for States, 2000 (in households per square mile)

Name	Old Dens. (A)	Rank of Old Dens. (B)	Old Dens. Group (C)	New Dens. (D)	New Dens. Group (E)	Rank of New Dens. (F)	Develop -able Lands (sq. mi.) (G)	Area of non-water Lands (sq. mi.) (H)	Percent-age of non-water Lands that is develop-able (%) (I)
Alaska	0.4	1	Very Low	1.2	Very Low	1	179,786	571,951	31.4
Wyoming	2.0	2	Very Low	4.1	Very Low	3	47,346	97,100	48.8
Montana	2.5	3	Very Low	3.7	Very Low	2	95,518	145,552	65.6
North Dakota	3.7	4	Very Low	4.1	Very Low	4	60,847	68,976	88.2
South Dakota	3.8	5	Very Low	4.6	Very Low	5	60,871	75,885	80.2
New Mexico	5.6	6	Very Low	8.9	Very Low	6	72,880	121,356	60.1
Idaho	5.7	7	Very Low	15.9	Very Low	9	29,517	82,747	35.7
Nevada	6.9	8	Very Low	56.8	Low	28	13,320	109,826	12.1
Utah	8.6	9	Very Low	30.4	Low	18	23,111	82,144	28.1
Nebraska	8.7	10	Very Low	9.1	Very Low	7	73,479	76,872	95.6
Kansas	12.7	11	Low	13.0	Very Low	8	79,987	81,815	97.8
Oregon	14.0	12	Low	31.0	Low	19	43,192	95,997	45.0
Colorado	16.1	13	Low	26.1	Low	14	64,023	103,718	61.7
Maine	16.8	14	Low	19.2	Low	10	27,035	30,862	87.6
Arizona	16.9	15	Low	48.4	Low	26	38,689	113,635	34.0
Oklahoma	19.6	16	Low	20.4	Low	11	65,912	68,667	96.0
Arkansas	20.1	17	Low	24.5	Low	13	42,691	52,068	82.0
Iowa	20.6	18	Low	21.3	Low	12	54,152	55,869	96.9
Mississippi	22.4	19	Low	27.6	Low	15	37,971	46,907	80.9
Minnesota	23.9	20	Low	32.4	Low	20	58,546	79,610	73.5
Vermont	26.1	21	Low	29.9	Low	16	8,087	9,250	87.4
Texas	28.4	22	Low	30.3	Low	17	244,957	261,797	93.6
West Virginia	30.6	23	Low	34.0	Low	21	21,711	24,078	90.2
Missouri	32.0	24	Low	34.3	Low	22	64,131	68,886	93.1
Washington	34.3	25	Low	52.7	Low	27	43,121	66,544	64.8
Alabama	34.3	26	Low	38.8	Low	23	44,905	50,744	88.5
Louisiana	38.1	27	Low	62.4	Mod.	30	26,611	43,562	61.1
Wisconsin	38.5	28	Low	47.4	Low	25	44,009	54,310	81.0
Kentucky	40.2	29	Low	42.8	Low	24	37,260	39,728	93.8
South Carolina	51.2	30	Mod.	67.7	Mod.	33	22,750	30,109	75.6

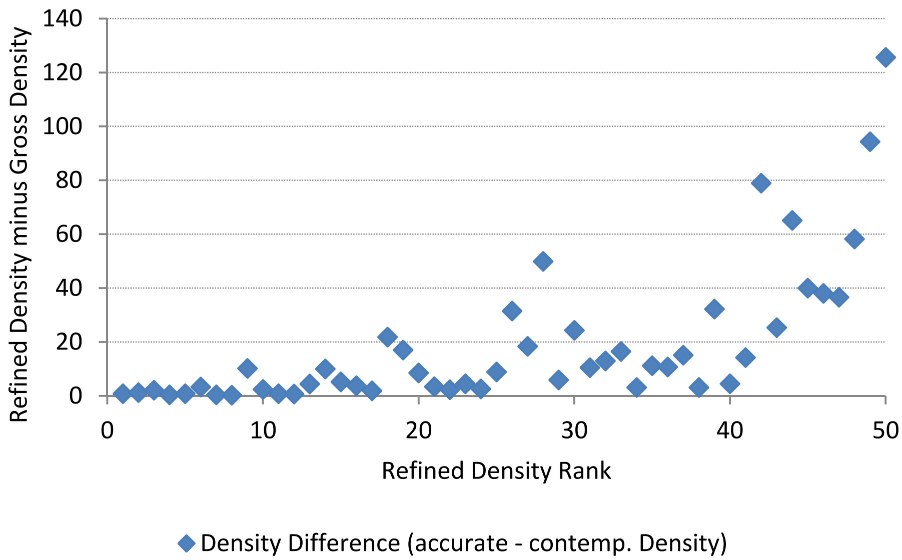
Name	Old Dens. (A)	Rank of Old Dens. (B)	Old Dens. Group (C)	New Dens. (D)	New Dens. Group (E)	Rank of New Dens. (F)	Develop -able Lands (sq. mi.) (G)	Area of non-water Lands (sq. mi.) (H)	Percent-age of non-water Lands that is develop-able (%) (I)
Georgia	52.2	31	Mod.	62.7	Mod.	31	48,230	57,906	83.3
New Hampshire	53.1	32	Mod.	66.1	Mod.	32	7,207	8,968	80.4
Tennessee	54.4	33	Mod.	60.3	Mod.	29	37,222	41,217	90.3
Hawaii	63.0	34	Mod.	74.2	Mod.	35	5,451	6,420	84.9
North Carolina	64.7	35	Mod.	79.8	Mod.	37	39,443	48,711	81.0
Indiana	65.3	36	Mod.	68.4	Mod.	34	34,245	35,867	95.5
Michigan	66.8	37	Mod.	99.0	Mod.	39	38,318	56,804	67.5
Virginia	68.4	38	Mod.	79.1	Mod.	36	34,238	39,594	86.5
California	74.0	39	Mod.	152.9	High	42	75,389	155,959	48.3
Illinois	82.8	40	Mod.	85.9	Mod.	38	53,584	55,584	96.4
Pennsylvania	106.8	41	High	121.0	Mod.	41	39,553	44,817	88.3
Ohio	108.8	42	High	113.3	Mod.	40	39,332	40,948	96.1
Florida	118.1	43	High	183.2	High	44	34,744	53,927	64.4
New York	149.8	44	High	189.8	High	45	37,244	47,214	78.9
Delaware	153.6	45	High	178.9	High	43	1,678	1,954	85.9
Maryland	203.3	46	High	241.3	High	46	8,235	9,774	84.3
Connecticut	269.2	47	High	305.8	High	47	4,264	4,845	88.0
Massachusetts	312.5	48	High	370.7	High	48	6,610	7,840	84.3
Rhode Island	391.8	49	High	486.1	High	49	842	1,045	80.6
New Jersey	414.3	50	High	539.9	High	50	5,691	7,417	76.7

Note:

1. The “Dens.” in column heading represents “Density”. The “Old Dens.” (in column A, B, and C) refer to the gross densities; while the “New Dens.” (in column D, E, and F) refers to the refined densities (i.e. developable lands densities) calculated in this study.
2. The “Mod.” In column C and E represents “Moderate”.

Figure 4.2 is a scatter plot presenting the differences between these two kinds of densities. The horizontal “x” axis represents the rank in refined densities for states, with the “1st” representing the lowest state density and the “50th” the highest density. The vertical “y” axis represents the value of refined density minus gross density for a state. Each box in this figure represents one state. The biggest differences between these two types of densities occurs in several of the states in both the high-density (42nd to the 50th) and low-density (10th to the 28th) groups[107].

Figure 4.2 Differences between Two Types of Densities (in households/square mile)



For example, Nevada (the 28th state in Figure 4.2), a state in the low-density group, has a “developable lands or refined density” of 56.8 households per square mile, more than 8 times its gross density (6.9 households per square mile). This significant difference is caused by large areas of federal lands in Nevada: 83 percent of its total area is in federal lands[108] and only 12 percent of its non-water lands are defined as developable in this study.

---

[107] However, there is a strong positive linear correlation between the refined and the grow densities, with a correlation coefficient of 0.99. The correlation coefficient between the *ranks* in these two densities has a slightly smaller value of 0.95.

[108] The total state area of Nevada (including open water) is 110,561 square miles, calculated from the ESRI GIS state boundary data. The federal lands cover more than 92,198 square miles—this number does not include the 1547 square miles of Indian Reservation lands, because part of the Indian Reservation lands had been adjusted as developable. If the other portion of Indian reservation lands that are defined as undevelopable is counted in the federal lands, then the federal lands will cover more than 83 percent of the total state area.

Another example, New Jersey (the 50th state in Figure 4.2), a state in the high-density group, has a noteworthy difference of 126 households per square mile between the two kinds of densities, mainly due to the significantly different denominators between the two densities (resulting from the large amount of naturally undevelopable lands—such as wetlands—in the state) and to the large number of households in the state.

Now to the second part of the question: how will the differences in values between the refined densities and the gross densities affect the assignment of county land use types? As will be shown, the differences affect the assignment of county types in a compounding fashion. If gross densities were used to classify state density groups, seven states (Nevada, Utah, Kansas, Louisiana, California, Pennsylvania, and Ohio) would be placed into different density groups outside their current assignation (the one based on refined densities). This would result in different compositions of state density groups (that is, some of the counties in each state density group would be different) between these two kinds of density systems, which in turn would affect the classification of land use types for inclusive counties in two ways. First, density thresholds within county land use types of each state density group could be affected, due to the different composition of counties in that density group. Second, the counties in these seven states may be classified as very different land use types and further, in different sprawl/non-sprawl categories, in one or the other density system.

For example, Ohio is classified as a high-density state based on the gross density system, but as a moderate-density state based on the refined density system; many counties in Ohio have higher levels of county types in the refined density system than they would have in the gross density system. For example, Summit County, Ohio would be classified as a suburban county based on the gross density system[109], but as an urban county based on the refined density system[110]. As an urban county, it is classified as non-sprawl over the 2000 to 2020 period in this study; however, it could have been classified as a sprawl suburban county under the gross density system if it experienced significant growth during this time period. Therefore, different county land use type classifications based on the two different levels of densities could result in different sprawl/non-sprawl classifications.

The reasons that these seven states have different state density classifications under the two different density systems are explained below.

Four of the abovementioned seven states, Nevada, Utah, Louisiana, and California, belong to higher levels of density groups in the refined density system than they do in the gross density system. Containing large areas of undevelopable, non-water lands, these states have "developable lands or refined densities" that are much higher than gross densities. As a result, they increase in their levels of state density classification.

The situation which caused the great fluctuation in ranking for the state of Nevada was explained earlier. It is the state that changes most in ranking due to these two density systems, moving from the eighth lowest density state (in the very-low density group) as determined by the gross density system, to the 28th lowest density state (in the low density group) as adjusted in the more refined density system.

---

[109] This classification is borrowed from Burchell et al.'s study (2002).

[110] The gross and "developable lands or refined density" for Summit County are 529 and 614 households per square, respectively.

In Utah, as in Nevada, nearly 60 percent of its total lands are federal lands[111]; thus only 28 percent of the non-water lands are developable (as shown in column I in table 4.1). Therefore, based on the gross density system, it would be classified as a "very-low" density state (with a gross density of 8.6 households per square mile); while based on the refined density system, it is defined as a "low" density state (with a refined density of 30.4 households per square mile, which is 3.5 times its gross density).

In Louisiana, 40 percent of the total area is naturally undevelopable, encompassing features such as open water and wetlands; only 60 percent of its non-water lands are developable. Based on the gross density system, it would be classified as a "low" density state (with a gross density of 38 households per square mile); while based on the refined density system, it is classified as a "moderate" density state (with a refined density of 62 households per square mile).

California contains large areas of naturally undevelopable lands, national forests and other federal lands, as well as state and local forests and parks. As a result, only 48 percent of its non-water lands are developable lands. In the gross density system, California would be classified as a "moderate" density state (with a gross density of 74 households per square mile); while in the refined density system, it is classified as a "high" density state (with a refined density of 153 households per square mile, twice its gross density).

The refined density calculation, with undevelopable land areas being excluded from the computation, better represent actual densities in the real world. This conclusion

---

[111] Indian Reservation lands are not counted in this area of federal lands. The federal lands (with Indian reservation lands excluded) and total area (including water) in Utah is 48,599 and 84,897 square miles, respectively.

can be supported by the cases of California and Louisiana. In this study, California and Louisiana are classified as high- and moderate-density states, respectively, based on the refined density calculation. However, based on gross densities, they would be classified as moderate- and low-density states, respectively. Based on realistic knowledge of these two states, Burchell et al. (2002) correctly adjusted California from a moderate- to a high-density state, and also adjusted Louisiana from a low- to a moderate- density state. Their adjustments match the results developed in this study, which implies that the refined densities fittingly reflect the realities of these two states.

On the other hand, two of the abovementioned seven states, Pennsylvania and Ohio, experience a decrease in their level of state density classification (from high-density states to moderate-density states), when changing from the gross density system to the refined density system. As shown in Table 4.1, based on gross densities, these two states would be classified as high density states, and California would be classified as a moderate density state. The "developable lands densities" of these two states are a bit higher than their gross densities. However, the significant increase in California's densities (from gross density to refined density) results in a big gap in refined densities between California and these two states; thus based on the refined density scheme, these two states are classified as moderate density states while California is classified as a high density state.

Based on these discussions, it is concluded that it is necessary to calculate refined densities in a study that measures land-use types, for two reasons. First, the refined density may be significantly different from the gross density for the same geographic unit. More importantly, the former better reflects the density of the real world. Second, the different compositions of state density groups between the refined and gross density systems will result in different classifications of land use types for some counties, which will further affect the sprawl classification for these counties.

4.2 County Land Use Type Classifications

As described in Chapter 3, based on the densities of both counties and their host states, as well as the author's practical real world knowledge of counties in the United States, one of six existing land use types is assigned to each county. The classification results are described here.

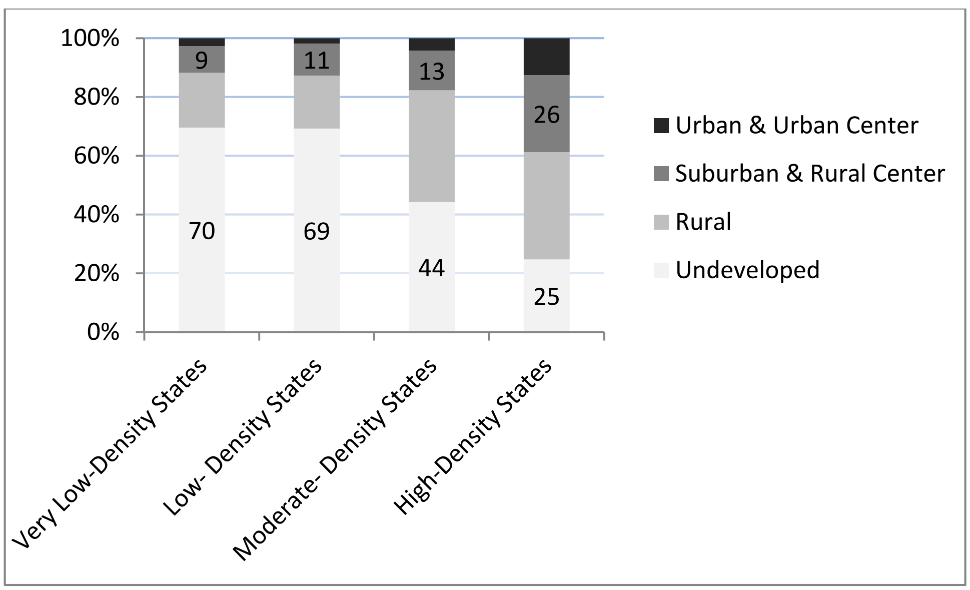
Generally, the lower the level of county land use types, the more counties found in that classification group. As shown in Table 4.2, the largest county land use type is the undeveloped county (the lowest level of the land use types), containing nearly 60 percent of all US counties. This group contains many more counties than any of the other five classification groups. The second largest group is the rural classification, with approximately one-quarter of all counties. The suburban county category follows as the third largest group, consisting of about seven percent of all counties. Together, urban and urban center counties account for less than four percent of all US counties.

Table 4.2 Number of Counties by Land use Type Classification

Land use Type	Number of Counties	Percentage Share (%)
Undeveloped	1771	57.3
Rural	812	26.3
Rural Center	173	5.6
Suburban	222	7.2
Urban	76	2.5
Urban Center	37	1.2
Total	3091	100

Figure 4.3 shows the distribution of county land use types among the counties in each state density group. The corresponding table for this figure (Table 4.4) can be found in Appendix 2. Each 100% stacked column represents the six county land use types' percentage shares of all counties in an individual state density group. For example, the stacked column for the low-density states shows that of all 1330 counties in low-density states, 69 percent are undeveloped, and less than 2 percent are urban or urban center counties. Generally speaking, in the four state density groups, the largest percentage share is undeveloped counties and the smallest share is urban and urban center counties.

Figure 4.3 Distribution of County Land Use Types in Each State Density Classification



Certain patterns can also be observed when a given land use type's percentage shares are compared across state density groups (from very-low density states at one end of the spectrum to high-density states at the other end). Overall, the higher the density in a state group, the higher the percentage of all counties in that group categorized as developed (i.e. urban and urban center, or suburban and rural center). As shown in Figure 4.3, the suburban and rural center counties' percentage shares increase from the lowest at nine percent of all counties in very-low density states to the highest, at 26 percent in high-density states. A similar pattern can be observed for urban and urban center counties. In contrast, the undeveloped counties' percentage share of all counties in density groups decreases from 70 percent in very-low density states to 25 percent in high-density states.

4.3 U.S. Growth, 2000-2020

Projected growth in households and employment over the 2000 to 2020 period is employed to develop a series of criteria for classifying counties as sprawl/non-sprawl. These criteria have been presented in the basic definition of sprawl contained in Chapter 3. Overall growth at the national, state, and regional level is discussed in this section. Growth at the EA and county level will be addressed with sprawl growth in a later section.

From 2000 to 2020, households in the United States will increase by 25.3 million from 106 million to 131.2 million, a growth rate of about 24 percent over 20 years. Meanwhile, employment will grow by about 46.5 million to nearly 214 million jobs, a growth rate of nearly 28 percent.

Household and employment growth will not be evenly distributed across the country (Burchell et al. 2002). About one-third of projected national household growth from 2000 to 2020 will take place in only three states; more than one-half of all household growth will occur in ten states.

Table 4.3 lists the variables describing household growth during the 2000 to 2020 period for the top 20 household growth states. The contributions of these 20 states to overall national household growth vary considerably, ranging from the highest, 13 percent, for California, to the lowest, or about 2 percent, for South Carolina. The top 3 growth states—California, Texas, and Florida (in descending order of absolute growth)—account for one-third of national household growth, with each of them individually representing more than one-tenth of all national growth. The growth rates of these 20 states over the 20-year study period vary from the highest, 63 percent for Nevada, to the lowest, or close to 10 percent, for New York. Of these 20 states, eight are in the South; six are in the West, four are in the Midwest; and only two are in the Northeast.

Table 4.3 Household Growth for Top 20 States (in thousands), 2000-2020

State	Rank	Households In 2000	Households In 2020	Household Growth	% Of National Household Growth	Growth Rate (%)
California	1	11534	14800	3267	12.9	28.3
Texas	2	7430	10159	2730	10.8	36.7
Florida	3	6367	8936	2569	10.2	40.4
Arizona	4	1916	2951	1035	4.1	54.1
North Carolina	5	3149	4162	1013	4	32.2
Georgia	6	3024	3975	951	3.8	31.5
Virginia	7	2710	3468	758	3	28
Washington	8	2283	3018	735	2.9	32.2
New York	9	7071	7759	688	2.7	9.7
Colorado	10	1669	2322	654	2.6	39.2
Illinois	11	4603	5243	640	2.5	13.9
Tennessee	12	2243	2873	630	2.5	28.1
Michigan	13	3796	4354	558	2.2	14.7
Maryland	14	1987	2522	535	2.1	26.9
New Jersey	15	3073	3596	523	2.1	17
Ohio	16	4456	4955	499	2	11.2
Nevada	17	758	1239	481	1.9	63.4
Minnesota	18	1902	2377	475	1.9	24.9
Oregon	19	1340	1790	450	1.8	33.6
South Carolina	20	1541	1989	448	1.8	29.1

Like household growth, projected employment growth over the 2000 to 2020 period is also not evenly distributed across the US. Table 4.4 presents the absolute growth and the growth rate of employment for the top 20 growth states. As shown in Table 4.5, three states will comprise about 30 percent of the projected national employment growth; ten states will account for more than one-half of national employment growth; and twenty states will comprise three-quarters of national employment growth over the 2000 to 2020 period. The top three employment growth states—California, Texas, and Florida (in descending order of absolute growth) are also the top three household growth states.

Table 4.4 Employment Growth for Top 20 States (in thousands), 2000-2020

State	Rank	Employment in 2000	Employment in 2020	Growth	% of National Employment Growth	Growth Rate (%)
California	1	19655	25267	5612	12.1	28.6
Texas	2	12314	16878	4564	9.8	37.1
Florida	3	8951	12667	3716	8.0	41.5
Illinois	4	7442	9108	1666	3.6	22.4
North Carolina	5	4943	6585	1642	3.5	33.2
Georgia	6	4906	6512	1606	3.5	32.7
Arizona	7	2822	4359	1537	3.3	54.5
Ohio	8	6878	8411	1534	3.3	22.3
Virginia	9	4432	5791	1360	2.9	30.7
Michigan	10	5655	6964	1310	2.8	23.2
Pennsylvania	11	7003	8273	1270	2.7	18.1
New York	12	10548	11771	1223	2.6	11.6
Washington	13	3560	4703	1143	2.5	32.1
Colorado	14	2961	4089	1128	2.4	38.1
Tennessee	15	3507	4604	1097	2.4	31.3
Minnesota	16	3357	4291	934	2.0	27.8
Indiana	17	3692	4614	922	2.0	25.0
New Jersey	18	4771	5692	921	2.0	19.3
Wisconsin	19	3444	4309	865	1.9	25.1
Maryland	20	3111	3941	831	1.8	26.7

Overall, states with the largest absolute employment growth are concentrated in the South and Midwest regions. Of the top twenty employment growth states, seven are in the South, six are in the Midwest, four are in the west, and three are in the northeast (the region for each state is listed in Table 4.4 in Appendix 2).

However, with respect to absolute growth, the South and West are still the regions where the largest amounts of growth, in both households and employment, will take place, as shown in Table 4.5.

Table 4.5 Household and Employment Growth by Regions (in thousands), 2000 to 2020

Region	Rank	Household Growth	Employment Growth
South	1	11370	19110
West	2	7628	12758
Midwest	3	3974	9440
Northeast	4	2338	5184

The growth in households will also be analyzed in comparison with sprawl growth at the state, region, EA, and county levels.

4.4 Sprawl Growth in the US, 2000-2020

According to the definition of sprawl, a rural or undeveloped county[112] is classified as sprawl due to significant growth in either households, employment, or both. Overall, 492, or 16 percent, of US counties will experience sprawl growth during the 2000 to 2020 period. Of these 492 sprawl counties, 86 percent (424 counties) will experience sprawl

[112] As defined in Chapter 3, some developing suburban and rural center counties are also classified as sprawl.

household growth or sprawl household and employment growth[113]; while only 14 percent (68) will experience sprawl employment growth.

Although significant sprawl will be found in 16 percent of all US counties, it will affect about one-third of new development nationally over the 2000 to 2020 period. Table 4.6 presents the distribution of national numeric growth that will take place in sprawl and non-sprawl counties. As shown in Table 4.6, of all national household growth (25 million households), nearly 30 percent will take place in *sprawling* counties; about 60 percent will occur in already developed urban and suburban, *non-sprawling* counties; and 10 percent will be found in very low-growth rural and undeveloped counties (these are *non-sprawling* counties, because the definition of a sprawling county is significant or fast growth in rural or undeveloped locations).

Table 4.6 Numeric Growth in Sprawl and Non-sprawl Counties (in thousands), 2000-2020

Growth	Sprawl Counties	Non-sprawl Counties		Total
		Urban & Suburban[1]	Rural and Undeveloped	
Households	7039	15605	2668	25312
Employment	9562	31538	5394	46494

Note:

[1]This column presents the growth in urban center, urban, suburban, and rural center counties. To save space, the column head is written as “Urban & Suburban”.

Most of the counties that are sprawling in the United States are rural or undeveloped counties. Table 4.7 shows the number of sprawling counties aggregated by county land use type. As shown in Table 4.7, of the 492 counties that are sprawling, over 80 percent

[113] 302 counties will experience sprawl growth in both households and employment.

(396 counties) are rural or undeveloped counties; the remainder are developing suburban or rural center counties. A more detailed table showing the number of sprawl counties by the six county land use types is found in Table 4.5 in Appendix 2.

Table 4.7 Number of Sprawl Counties by County Land Use Types

Sprawl Classification	Rural & Undeveloped	Suburban & Rural Center	Urban Center & Urban	Total
Sprawl	**396**	**96**	**0**	**492**
Non-sprawl	2187	299	113	2599
Total	2583	395	113	3091

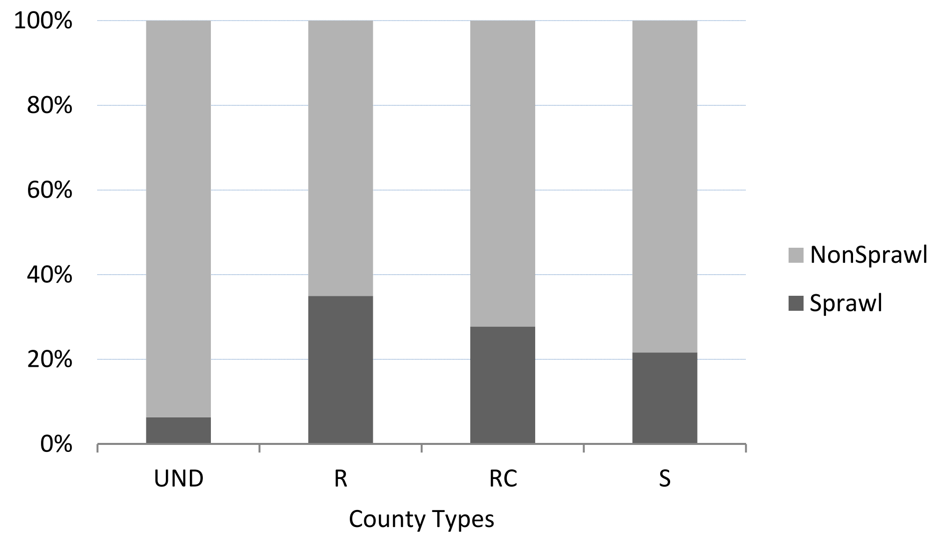
Note:

The numbers of sprawl counties are shown in bold.

Rural counties are more likely to experience sprawling growth than any other county type[114]. As shown in Table 4.5 (in Appendix 2) and Figure 4.4, 35 percent of the county type experiencing sprawl, rural counties will be most impacted by sprawl, followed by rural center counties (28 percent), and suburban counties (22 percent). Due to their large numbers (1771 undeveloped counties), only six percent of undeveloped counties will experience significant sprawl growth.

Figure 4.4 Sprawl Counties' Percentage Shares of all Counties by County Types

[114] Of the overall six current county land use types, four types (undeveloped, rural, rural center, suburban) may experience sprawl; while the other two types (urban and urban center) will not experience sprawl according to the sprawl definition employed in this study.



Note:

UND = “Undeveloped”; R = “Rural”; RC=”Rural Center”; and S = “Suburban”.

Sprawl growth will be analyzed at different geographic levels (states, census regions, EAs, and counties) in the following sections.

4.4.1 Sprawl at the State Level

Sprawling counties are primarily concentrated in a few states. As shown in Table 4.8, over half of the US counties experiencing sprawl are concentrated in twelve, or roughly one-quarter, of US states. Further, one-third of US counties experiencing sprawl are found in six states: Texas, Georgia, California, North Carolina, Tennessee, and Florida, in descending order of the number of sprawl counties in each state.

Conversely, three states will not experience much growth in sprawl form over this period. North Dakota and Vermont are relatively low growth, low density states. Massachusetts on the other hand, is a high density state that will experience a moderate amount of growth (27th of the 50 states in volume of growth over the period). The counties in MA that will experience significant growth are the state's developed suburban and rural center counties, as well as urban and urban center counties. In addition, small amount of growth are happening in low-growth rural counties. Therefore, growth in all 14 counties in MA is happening in non-sprawling counties.

Table 4.8 Top 12 States with the Greatest Number of Sprawl Counties, 2000-2020

State	Rank	Sprawl Counties	All Counties
Texas	1	39	254
Georgia	2	26	159
California	3	23	58
North Carolina	4	23	100
Tennessee	5	21	95
Florida	6	20	67
Indiana	7	19	92
Ohio	8	18	88
Michigan	9	17	83
Missouri	10	17	115
Pennsylvania	11	16	67
Oklahoma	12	16	77

Sprawl results for states are presented in Table 4.9. The table presents sprawl household growth and overall household growth[115] in states for the period 2000 to 2020. In this table, states are listed in descending order of their contribution to national sprawl household growth (i.e. column E[116]). Column A is a numerical ranking of column E. The state with the largest amount of sprawl household growth (i.e., with the largest contribution to national sprawl household growth) ranks 1st; while the state with the least

[115] The total household growth of a state is the "net growth", which is the sum of both the households increase and households decrease for all counties in the state. The sprawl household growth for a state is the sum of household growth between 2000 and 2020 for all the sprawl counties in the state. Generally, sprawl household growth is positive for a sprawl county (with only one exception—the county is an employment-sprawl county that will experience a household decrease).

[116] Column E is calculated as (column C*100/7038865). The national sprawl household growth during the 2000 to 2020 period is 7038865 households.

amount of sprawl growth ranks 48th[117]. Column F[118] scores an individual state's contribution to national overall household growth. Column B is a numerical ranking of Column F.

Table 4.9 Sprawl Growth Compared with Overall Growth in US States, 2000-2020 (in Thousands of Households)

State	Sprawl Growth Rank (A)	Overall Growth Rank (B)	House-hold Growth in Sprawl Counties (C)	House-hold Growth in All Counties (D)	Sprawl House-hold Growth's Perc. Share of all US Sprawl Growth (%) (E)	Overall Growth's Perc. Share of all US House-hold Growth (%) (F)	Perc. of State Growth Desig. as Sprawl (%) (G)
California	1	1	950.7	3266.5	13.5	12.9	29.1
Texas	2	2	537.6	2729.8	7.6	10.8	19.7
Florida	3	3	528.2	2569.1	7.5	10.2	20.6
Georgia	4	6	332.8	951.4	4.7	3.8	35.0
Tennessee	5	12	284.6	630.1	4.0	2.5	45.2
North Carolina	6	5	269.8	1012.8	3.8	4.0	26.6
Pennsylvania	7	24	258.2	400.0	3.7	1.6	64.6
South Carolina	8	20	229.8	447.9	3.3	1.8	51.3
Washington	9	8	229.6	735.1	3.3	2.9	31.2
Ohio	10	16	226.9	498.6	3.2	2.0	45.5
Michigan	11	13	200.3	557.7	2.8	2.2	35.9
Minnesota	12	18	186.4	474.6	2.6	1.9	39.3
Maryland	13	14	181.8	534.6	2.6	2.1	34.0
Oregon	14	19	180.5	450.0	2.6	1.8	40.1
Alabama	15	25	171.4	342.7	2.4	1.4	50.0
Virginia	16	7	160.8	758.3	2.3	3.0	21.2
Indiana	17	21	159.9	446.5	2.3	1.8	35.8
Missouri	18	22	153.7	437.0	2.2	1.7	35.2
Arizona	19	4	149.6	1035.5	2.1	4.1	14.4
Oklahoma	20	29	120.7	269.9	1.7	1.1	44.7
Wisconsin	21	23	119.1	406.5	1.7	1.6	29.3
New Hampshire	22	37	114.3	130.2	1.6	0.5	87.8
Colorado	23	10	109.2	653.6	1.6	2.6	16.7

[117] The three states (North Dakota, Vermont, and Massachusetts) that will not experience significant sprawl growth over the 2000 to 2020 period have the same ranking order of 48th.
[118] Column F is calculated as (Column D*100/25311676). The overall household growth in the US during the 2000 to 2020 period is 25311676 households.

State	Sprawl Growth Rank (A)	Overall Growth Rank (B)	House-hold Growth in Sprawl Counties (C)	House-hold Growth in All Counties (D)	Sprawl House-hold Growth's Perc. Share of all US Sprawl Growth (%) (E)	Overall Growth's Perc. Share of all US House-hold Growth (%) (F)	Perc. of State Growth Desig. as Sprawl (%) (G)
Mississippi	24	33	108.6	196.9	1.5	0.8	55.2
Kentucky	25	28	106.6	301.4	1.5	1.2	35.4
New York	26	9	105.2	687.8	1.5	2.7	15.3
Arkansas	27	30	102.4	260.8	1.5	1.0	39.3
Louisiana	28	31	95.2	241.3	1.4	1.0	39.4
New Mexico	29	32	80.1	236.5	1.1	0.9	33.9
Connecticut	30	38	61.0	125.3	0.9	0.5	48.7
Utah	31	26	59.2	331.1	0.8	1.3	17.9
Idaho	32	34	58.4	181.4	0.8	0.7	32.2
Illinois	33	11	51.3	640.0	0.7	2.5	8.0
New Jersey	34	15	44.8	522.7	0.6	2.1	8.6
Montana	35	40	41.4	91.1	0.6	0.4	45.4
Iowa	36	36	39.3	132.6	0.6	0.5	29.6
Delaware	37	41	38.2	77.6	0.5	0.3	49.2
Hawaii	38	43	37.4	74.3	0.5	0.3	50.3
Nevada	39	17	34.2	481.1	0.5	1.9	7.1
Maine	40	42	32.0	74.8	0.5	0.3	42.8
Alaska	41	44	23.9	60.7	0.3	0.2	39.3
Kansas	42	35	14.3	180.3	0.2	0.7	7.9
West Virginia	43	48	13.5	44.9	0.2	0.2	30.1
South Dakota	44	45	13.4	53.3	0.2	0.2	25.2
Rhode Island	45	46	12.7	51.1	0.2	0.2	24.9
Nebraska	46	39	6.3	120.7	0.1	0.5	5.2
Wyoming	47	49	3.7	31.3	0.1	0.1	11.8
Massachusetts	48	27	0.0	301.7	0.0	1.2	0.0
Vermont	48	47	0.0	45.2	0.0	0.2	0.0
North Dakota	48	50	0.0	26.5	0.0	0.1	0.0

Note:

1. In the column headings, “Perc.” and “Desig.” represents “Percentage” and “Designated”, respectively.

As shown in Table 4.9, household growth in sprawl form is concentrated in a small number of states. The top three states in which sprawl growth found are California, Texas, and Florida; they account for nearly 30 percent of all national household growth in sprawl form. The top ten states having a sprawl development pattern contain more than one-half of national sprawl household growth; and the top 20 sprawl growth states contain nearly 80 percent of all national sprawl household growth.

As shown in Column E in Table 4.9, for the top ten sprawl growth states, their contributions to national sprawl household growth vary from a high of 13.5 percent for California to a low of 3.2 percent for Ohio, with the highest percentage share (13.5% for California) being much larger than the second largest share (7.6% for Texas). That is, nearly one-seventh of national sprawl household growth likely to take place over the 2000 to 2020 period is taking place in California.

States with the largest amount of sprawl growth are concentrated in the South and the West. Among the top 10 sprawl growth states, six are in the South and two are in the West. The remaining two are one each in the Northeast (Pennsylvania) and the Midwest (Ohio). For the top 20 sprawl growth states, ten are in the South, five are in the Midwest, four are in the West, and only one is in the Northeast.

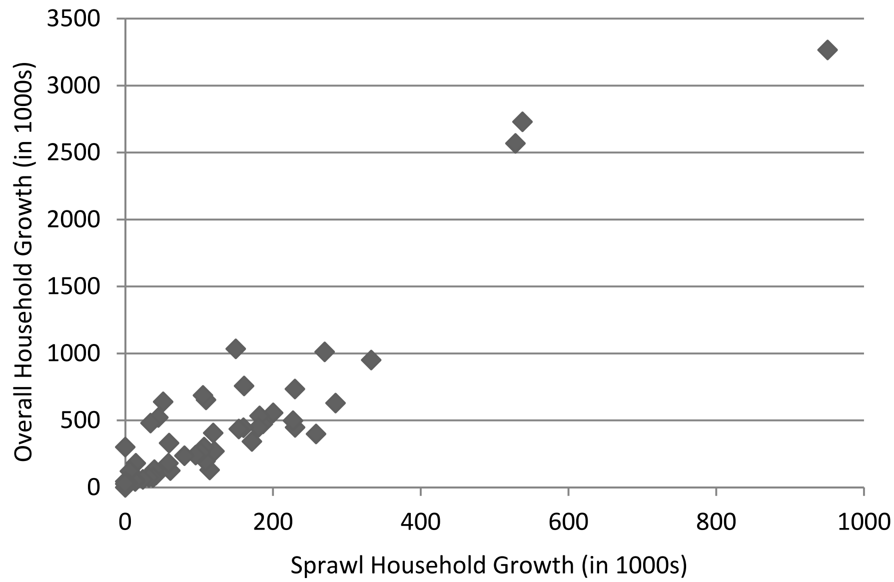
Column G in Table 4.9 lists the percentage share of overall state growth designated as sprawl, which is calculated as “column C (i.e. sprawl growth in state)*100/ column D (i.e. overall growth in state)”. Among the top 20 sprawl growth states, the percentage shares of all growth designated as sprawl vary considerably, from a high of 65 percent for Pennsylvania to a low of 14 percent for Arizona. That is, 65 percent of all household growth during the 2000 to 2020 period in Pennsylvania will take place in sprawl counties and thus be designated as sprawl growth; in contrast, only 14 percent of future household growth in Arizona will occur in sprawl counties and so be designated as sprawl growth.

This variable can explain to some extent the difference between an individual state's contribution to national *sprawl* growth and its contribution to national *overall* growth, which will be addressed in a later section.

As argued by Burchell el al. (2002), sprawl trends follow growth trends. States with the largest amount of overall growth are also those likely to have the largest amount of sprawl growth. Figure 4.5 is a scatter plot for the 50 US states that presents overall household growth (the vertical "y" axis) by sprawl household growth (the horizontal "x" axis). In this figure, each point represents one individual state. As shown in Figure 4.5, the points basically fall around a straight line with a positive slope, clearly demonstrating that there is a strong positive linear relationship between the amount of overall household growth in states and the amount of sprawl household growth in states. This is also confirmed by a high Pearson correlation coefficient value of 0.91976[119] between these two variables. Further, the correlation coefficient between a state's contributions to national sprawl growth (Column E in Table 4.9) and to national overall growth (Column F in Table 4.9) is also 0.91976. This is because the former is an exact linear function of a state's sprawl growth, while the latter is an exact linear function of a state's overall growth. Therefore, the correlation coefficients between these two pairs of variables have the same value of 0.91976.

Figure 4.5 All Growth by Sprawl Growth in Households for US States, 2000 to 2020 (in Thousands of Households)

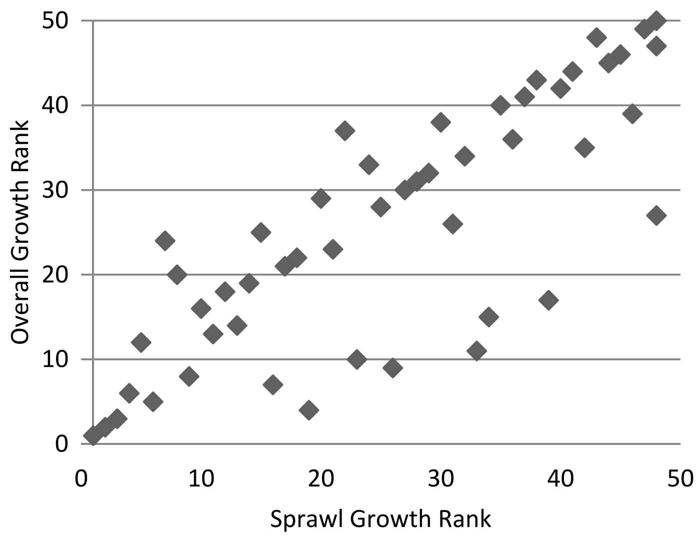
[119] The p-value of the test of null hypothesis "there is no linear relationship between the sprawl household growth and the overall household growth" is less than 0.0001), and thus the null hypothesis will be rejected.



Although there is a strong linear relationship between the amount of sprawl growth and overall growth for states, very often the ranking orders of these two kinds of growth (i.e. column A and B in Table 4.9) for an individual state are different. The Spearman rank-order correlation coefficient between these two kinds of ranking orders is 0.81634[120], implying a weaker positive linear association relative to the one between the amounts of these two kinds of growth (the abovementioned coefficient of 0.91976). The scatter plot Figure 4.6 displays the significant linear association between these two kinds of ranks (the "y" axis represents the overall growth rank-Column B in Table 4.9; the "x" axis represents the sprawl growth rank-Column A in Table 4.9), with each point representing one state.

Figure 4.6 Sprawl Growth Rank by Overall Growth Rank for US States

[120] The p-value of the test for the null hypothesis "there is no linear relationship between the two rank-orders" is less than 0.0001", and thus the null hypothesis will be rejected.



As mentioned above, the sprawl growth rank for a state may not be exactly the same as its overall growth rank. I agree with Burchell et al. (2002) that sprawl growth's percentage share of all growth in a state provides some explanation for this discrepancy. For states with a small sprawl growth percentage share of all growth (that is, a significant proportion of all growth in the state will take place in non-sprawl counties), their sprawl growth rank tends to be less significant than their overall growth rank; in other words, their contribution to national sprawl growth is likely to be less significant than their contribution to national overall growth. For example, Virginia and Arizona have low percentage shares of overall growth in their states that are categorized as sprawl (21 percent for Virginia and 14 percent for Arizona); as a result, they are only the 16th and 19th most significant states relative to sprawl growth, but are the 7th and 4th most significant overall growth states, respectively.

In contrast, for states with a large percentage share of overall growth designated as sprawl, their sprawl growth rank tends to be more significant than their overall growth rank. For example, among the top 10 sprawl growth states, Tennessee, Pennsylvania, South Carolina, and Ohio have sprawl growth ranks of 5th, 7th, 8th, and 10th, respectively, yet have overall growth ranks of 12th, 24th, 20th, and 16th, respectively, mainly due to their high percentage shares of all growth designated as sprawl growth (which range from 45 to 65 percent for these four states).

4.4.2 Sprawl at the Region Level

Not surprisingly, at the regional level, the South is home to the most sprawl counties and the Northeast is party to the least. As shown in Table 4.10, of the 492 sprawl counties nationwide, nearly one half (239 counties) are in the South; one quarter (124 counties) are in the Midwest; 18 percent (86 counties) are in the west; and only 9 percent (43 counties) are in the Northeast.

Table 4.10 Numbers of Sprawl Counties by Region, 2000-2020

Region	Sprawl Counties	All Counties
South	239	1394
Midwest	124	1054
West	86	426
Northeast	43	217
Total	492	3091

Again, at the regional level, the greatest amounts of sprawl household growth are in those regions with the greatest amount of overall growth. As shown in Table 4.11, sprawl growth rank is the same as total growth rank for the four regions. The South is experiencing much more sprawl growth than any other region over the 2000 to 2020 period. Nearly half of all national sprawl household growth is taking place in the South, followed by the West (28 percent), the Midwest (17 percent), and the Northeast (9 percent). The amount of sprawl growth found in the South is nearly twice the amount as in the West, the second most significant sprawl region.

Table 4.11 Sprawl and Overall Household Growth by Region, 2000-2020 (in thousands of households)

Region	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	Perc. of US House-hold Growth Designated as Sprawl (%)	Perc. of All US Household Growth (%)	Perc. of All Regional Growth Desig. as Sprawl (%)
South	1	1	3282	11370	46.6	44.9	28.9
West	2	2	1958	7628	27.8	30.1	25.7
Midwest	3	3	1171	3974	16.6	15.7	29.5
Northeast	4	4	628	2339	8.9	9.2	26.9

Note:

In column headings, "Perc." and "Desig." represent "Percentage" and "Designated", respectively.

In short, among the four regions, sprawl growth is most significant in the South, which contains half the future sprawl counties and half the projected sprawl household growth during the 2000 to 2020 period. The West has the second largest amount of sprawl growth, while the Midwest contains the second largest number of sprawl counties. An interesting note concerns the Northeast. Although the Northeast contributes the least to national sprawl household growth (only 9 percent), this is more than one-quarter of household growth in the Northeast region as a whole. In fact, the percentage share of sprawl growth in relation to all growth in a region is quite similar among the four regions, varying modestly from 26 percent to 30 percent.

4.4.3 Sprawl at the EA Level

Eighty percent, or 136, of the 172 EAs are experiencing sprawl growth over the 2000 to 2020 period. The top three EAs with the largest number of sprawl counties are the Dallas-Fort Worth, TX-AR-OK EA, the New York-Northern New Jersey-Long Island NY-NJ-CT-PA-MA EA, and the Minneapolis-St. Paul, MN-WI-IA EA, which contain 19, 17, and 16 sprawl counties, respectively.

At the other end of the spectrum, one-fifth, or 36, of the 172 EAs have no counties experiencing significant sprawl growth during this time period. Most of these non-sprawl EAs are very-low growth "Rural EAs" that contain only rural and undeveloped counties before county type adjustments. They do not meet the threshold that defines sprawl growth. Two examples of such non-sprawl "Rural EAs" are the Amarillo (TX-NM) EA and Springfield (IL-MO) EA. The EAs that will not experience significant sprawl growth are listed in Table 4.12 in Appendix 2.

As at the state and regional levels, sprawl growth tends to concentrate in a few EAs. As shown in Table 4.12, the top 10 sprawl growth EAs together account for more than one-third (36 percent) of national sprawl household growth; further, the top 20 sprawl growth EAs represent more than one-half (54 percent) of national sprawl household growth. Table 4.12 in Appendix 2 lists the same variables for all EAs.

Table 4.12 Sprawl Growth Compared with Overall Growth in EAs (Top 30 EAs), 2000-2020

EA Names	Sprawl Growth Rank (A)	Overall Growth Rank (B)	Household Growth in Sprawl Counties (C)	Household Growth in All Counties (D)	Perc. of U.S. Household Growth Designated as Sprawl (%) (E)	Perc. of All US Household Growth (%) (F)	Perc. of County Growth Designated as Sprawl (%) (G)

EA Names	Sprawl Growth Rank (A)	Overall Growth Rank (B)	House-hold Growth in Sprawl Counties (C)	House-hold Growth in All Counties (D)	Perc. of U.S. Household Growth Designated as Sprawl (%) (E)	Perc. of All US House-hold Growth (%) (F)	Perc. of County Growth Designated as Sprawl (%) (G)
Los Angeles-Riverside-Orange County, CA-AZ	1	1	404005	1436989	5.7	5.7	28.1
San Francisco-Oakland-San Jose, CA	2	5	327204	907694	4.6	3.6	36.0
Atlanta, GA-AL-NC	3	7	252694	764981	3.6	3.0	33.0
Washington-Baltimore, DC-MD-VA-WV-PA	4	4	236559	937491	3.4	3.7	25.2
Orlando, FL	5	12	232560	619015	3.3	2.4	37.6
New York-Northern NJ-Long Island, NY-NJ-CT-PA-MA	6	2	232398	1204837	3.3	4.8	19.3
Dallas-Fort Worth, TX-AR-OK	7	3	224305	1052869	3.2	4.2	21.3
Portland-Salem, OR-WA	8	17	219368	418783	3.1	1.7	52.4
Nashville, TN-KY	9	23	212588	326668	3.0	1.3	65.1
Minneapolis-St. Paul, MN-WI-IA	10	14	206589	482443	2.9	1.9	42.8
Sacramento-Yolo, CA	11	18	184328	381686	2.6	1.5	48.3
Houston-Galveston-Brazoria, TX	12	9	168511	719581	2.4	2.8	23.4
Philadelphia-Wilmington-Atl. City, PA-NJ-DE-MD	13	21	135209	349099	1.9	1.4	38.7
Jacksonville, FL-GA	14	30	132608	257093	1.9	1.0	51.6
Boston-Worcester-Lawrence-Lowell-Brockton, MA-NH-RI	15	15	127030	478513	1.8	1.9	26.5
Seattle-Tacoma-Bremerton, WA	16	13	117505	518454	1.7	2.0	22.7
Charlotte-Gastonia-Rock Hill, NC-SC	17	24	115245	300699	1.6	1.2	38.3
Detroit-Ann Arbor-Flint, MI	18	27	112403	290172	1.6	1.1	38.7
Denver-Boulder-Greeley, CO-KS-NE	19	11	95306	621111	1.4	2.5	15.3
Indianapolis, IN-IL	20	29	94937	258180	1.3	1.0	36.8
Little Rock-North Little Rock, AR	21	41	92219	148657	1.3	0.6	62.0

EA Names	Sprawl Growth Rank (A)	Overall Growth Rank (B)	House-hold Growth in Sprawl Counties (C)	House-hold Growth in All Counties (D)	Perc. of U.S. Household Growth Designated as Sprawl (%) (E)	Perc. of All US House-hold Growth (%) (F)	Perc. of County Growth Designated as Sprawl (%) (G)
Cincinnati-Hamilton, OH-KY-IN	22	36	88484	175740	1.3	0.7	50.3
Birmingham, AL	23	54	87570	120265	1.2	0.5	72.8
Columbus, OH	24	31	84075	238227	1.2	0.9	35.3
Wilmington, NC-SC	25	52	83270	123194	1.2	0.5	67.6
Las Vegas, NV-AZ-UT	26	16	77349	459379	1.1	1.8	16.8
St. Louis, MO-IL	27	34	74376	181956	1.1	0.7	40.9
San Antonio, TX	28	25	71961	297843	1.0	1.2	24.2
Grand Rapids-Muskegon-Holland, MI	29	35	70695	175984	1.0	0.7	40.2
Springfield, MO	30	50	65705	127547	0.9	0.5	51.5

Note:

"Perc." In column headings represents "Percentage".

The top 20 sprawl-growth EAs represent 54 percent of national sprawl household growth and 49 percent of national overall household growth. Their contributions to national sprawl growth range from a high of 5.7 percent for the Los Angeles-Riverside-Orange County EA to a low of 1.3 percent for the Indianapolis EA. For these twenty EAs, three quarters (15 EAs) are in the South or West: nearly one-half (9 EAs) are in the South, followed by those located in the West (6 EAs).

Sprawl growth as a percentage of overall growth in EAs varies widely, from zero to 92 percent among the 172 EAs. Among the top 20 sprawl-growth EAs, this variable ranges from a high of 65 percent for the Nashville, TN-KY EA to a low of 15 percent for the Denver-Boulder-Greeley, CO-KS-NE EA. That is, about 65 percent of overall household growth in the Nashville EA is taking place in sprawl counties.

As it occurs at the state level, there is a strong positive linear relationship between an EA's sprawl household growth and its overall household growth. Generally, large amounts of sprawl household growth take place in EAs with large amounts of overall household growth. The Pearson correlation coefficient between these two variables is 0.85474[121] for EAs, implying a strong positive linear association between these two variables.

Similarly, there is a strong positive linear association between an EA's rank on sprawl growth and its rank on overall growth, with a Spearman rank-order correlation coefficient of 0.86513[122]. Figure 4.7 presents the scatter plots of these two rank variables for EAs (with the "Y" axis representing the overall growth rank-Column B in Table 4.9 and the "X" axis representing the sprawl growth rank-Column A in Table 4.9). However, there are still significant differences between these two kinds of ranks for particular EAs. Again, if sprawl growth's percentage share of overall growth in an EA (Column G in Table 4.9) is high, the EA's sprawl growth rank (Column A) tends to be more significant than its overall growth rank (Column B). For example, the Orlando, FL EA, the Portland-Salem, OR-WA EA, and the Nashville, TN-KY EA rank as high as the 5th, 8th, and 9th most significant sprawl EAs, yet only the 12th, 17th, and 23rd most significant overall-growth EAs, respectively, mainly due to their relatively high sprawl growth percentage shares in comparison to overall growth in EA (38 percent, 52 percent, and 65 percent, respectively). Conversely, the New York-Northern NJ-Long Island EA, the Dallas-Fort Worth EA, and the Denver-Boulder-Greeley EA rank low, as the 6th, 7th, and 19th most

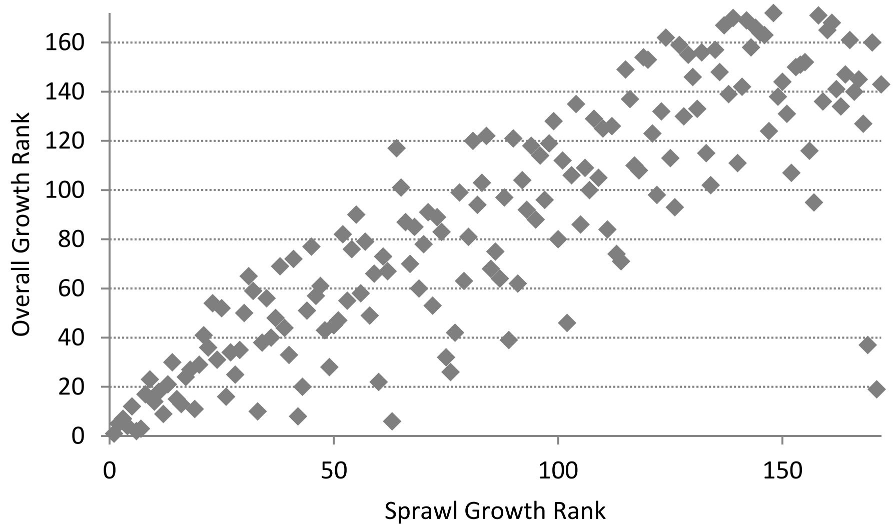
---

[121] The p-value for testing the null hypothesis "there is no linear relationship between these two variables" is less than 0.0001, and thus the null hypothesis will be rejected.

[122] The p-value for testing the null hypothesis "there is no linear relationship between these two rank variables" is less than 0.0001, and thus the null hypothesis will be rejected.

significant sprawl growth EAs, but rank high as the 2nd, 3rd, and 11th most significant overall-growth EAs, mainly due to their relatively low percentage share of overall growth in EA that is sprawl (19 percent, 21 percent, and 15 percent, respectively).

Figure 4.7 Sprawl Growth Ranks by Overall Growth Ranks for EAs



4.4.4 Sprawl at the County Level

Again, just as is seen at the state and EA levels, sprawl growth is also a phenomenon of concentration at the county level. The top 20 sprawl-growth counties (out of 3091 counties) represent nearly one-fifth of national sprawl household growth; the top 100 sprawl counties account for more than one-half (53 percent) of national sprawl growth; the top 200 sprawl counties contribute to three-quarters of national sprawl growth; and finally, the top 300 sprawl counties, or one-tenth of all US counties, comprise 88 percent of national sprawl growth.

Table 4.13 presents the top 30 sprawl growth counties. These 30 counties represent one-quarter of national sprawl growth, but only 7 percent of national overall growth. As shown in Table 4.13, their contributions to national sprawl household growth range from a high of 3.3 percent for San Bernardino County, CA to a low 0.5 percent for Pinal County, AZ. Among these 30 counties, 17 are rural and undeveloped counties and 13 are developing suburban and rural center counties.

Most of the top 30 sprawl growth counties are in the South and the West; 14 are in the South; 13 are in the West; only 2 are in the Northeast and one is in the Midwest. Nearly one-half of these 30 counties are concentrated in two states: eight counties are in California; and four are in Florida.

Table 4.13 Top 30 Sprawl Growth Counties, 2000-2020

Name	Sprawl Growth Rank	Overall Growth Rank	Household Growth	Percentage of US Household Growth Designated as Sprawl (%)	Percentage of All US Household Growth (%)	Land use Type
San Bernardino, CA	1	11	235262	3.3	0.9	Rural
Montgomery, TX	2	47	88134	1.3	0.3	Suburban
Clark, WA	3	57	78109	1.1	0.3	Suburban
Placer, CA	4	59	77734	1.1	0.3	Rural
Kern, CA	5	64	72582	1.0	0.3	Rural
Sonoma, CA	6	67	71954	1.0	0.3	Rural
San Joaquin, CA	7	68	71676	1.0	0.3	Rural
Clackamas, OR	8	73	67070	1.0	0.3	Suburban
Stanislaus, CA	9	76	64501	0.9	0.3	Rural
Polk, FL	10	87	60367	0.9	0.2	Rural Center
Washington, MN	11	96	56182	0.8	0.2	Suburban
Chester, PA	12	103	51467	0.7	0.2	Suburban
Charlotte, FL	13	106	49908	0.7	0.2	Rural
Osceola, FL	14	112	47528	0.7	0.2	Rural
Shelby, AL	15	114	46896	0.7	0.2	Rural Center
Rutherford, TN	16	115	46515	0.7	0.2	Rural
Lexington, SC	17	123	43832	0.6	0.2	Rural Center
Frederick, MD	18	125	43617	0.6	0.2	Rural

Name	Sprawl Growth Rank	Overall Growth Rank	Household Growth	Percentage of US Household Growth Designated as Sprawl (%)	Percentage of All US Household Growth (%)	Land use Type
San Luis Obispo, CA	19	129	43000	0.6	0.2	Rural
Monterey, CA	20	131	42774	0.6	0.2	Rural
Horry, SC	21	133	42382	0.6	0.2	Rural Center
Harford, MD	22	138	41010	0.6	0.2	Rural Center
Cherokee, GA	23	139	40959	0.6	0.2	Rural Center
Lancaster, PA	24	145	39755	0.6	0.2	Suburban
Mohave, AZ	25	146	39731	0.6	0.2	Undeveloped
Lake, FL	26	147	39461	0.6	0.2	Rural Center
Baldwin, AL	27	148	39384	0.6	0.2	Rural
Jackson, OR	28	150	39249	0.6	0.2	Rural
Williamson, TN	29	152	38337	0.5	0.2	Rural
Pinal, AZ	30	153	38283	0.5	0.2	Undeveloped

Since the county is the basic geographic unit in this study, the amount of overall growth in a sprawl county is used as the amount of its sprawl growth; thus the ranks of both sprawl-growth and overall-growth are based on the amounts of overall growth in counties.

While sprawl growth ranks and overall growth ranks are close to each other for states and EAs, they are significantly different for counties[123] (as shown in Table 4.13). Actually, all 492 sprawl counties have much more significant sprawl growth ranks than their overall growth ranks. This is because 84 percent of counties are non-sprawl and a substantial number of the non-sprawl, urban and suburban counties have greater amounts of growth than the sprawl, rural, and undeveloped counties do[124]. With these non-sprawl, big-growth developed counties getting involved in the overall growth ranking, the sprawl

[123] Therefore, the ranks of the sprawl growth counties are based on the 492 sprawl counties; while the ranks for the overall growth counties are based on all 3091 counties. For both ranks, the rank order of 1st represents the largest amount of growth.

[124] In contrast, only 6 percent of all states and 21 percent of all EAs are non-sprawl; further, most of these non-sprawl states or EAs have much lower growth than sprawl states or EAs do.

counties' overall growth ranks are much less significant than their sprawl growth ranks. For example, San Bernardino, CA ranks as the 1st most significant sprawl growth county, but only the 11th most significant overall growth county. Montgomery, TX, is the 2nd most significant sprawl growth county, but only the 47th most significant overall growth county. In other words, among the top 40 overall growth counties, San Bernardino is the only one that will experience significant sprawl; the other 39 counties are urban center, urban, and developed suburban counties, with large amounts of growth that are defined as non-sprawl.

Therefore, the top 30 sprawl growth counties are very different from the top 30 overall growth counties. As a comparison to the top 30 sprawl growth counties, Table 4.14 in Appendix 2 lists the top 30 overall-growth counties. These 30 counties (out of 3091 counties) account for one-quarter of national overall household growth, a much higher percentage than the contribution of the top 30 sprawl growth counties (7 percent). As shown in Table 4.14 (in Appendix 4), their contributions to national overall household growth vary from a high of 2.6 percent for Maricopa County, AZ, to a low of 0.5 percent for Denton County, TX. Of these 30 counties, thirteen, or nearly half, are urban counties; ten are developed suburban counties; five are urban center counties; one is rural center and one is a purely rural county. The only rural county among these 30 is the abovementioned San Bernardino County.

4.5 Sprawl Classifications in Five Selected EAs

The purpose of this case analysis is to present in detail the sprawl and non-sprawl classifications for individual counties in the selected EAs. Do the sprawl classifications make sense, given the situations in the real world? If the answer to this question is yes, then it implies that the data and method employed in this study are appropriate for classifying counties as sprawl or non-sprawl.

Five example EAs are selected for this analysis, based on two major selection criteria—their general recognizability and their contribution to national sprawl household growth. These five EAs are recognizable nationwide (or even worldwide) and represent significant components of national sprawl growth. The five selected EAs are listed below:

- Atlanta, GA-AL-NC
- Los Angeles-Riverside-Orange, CA-AZ
- New York-North New Jersey-Long Island, NY-NJ-CT-PA-MA
- San Francisco-Oakland-San Jose, CA
- Washington-Baltimore, DC-MD-VA-WV-PA

In order to control sprawl growth in an EA, part of future growth in the sprawl counties (i.e. "sending" locations) should be redirected to non-sprawl urban and suburban counties[125] (i.e. receiving locations) within the same EA (Burchell et al. 2002). Further, the non-sprawl, low-growth rural or undeveloped counties are not appropriate candidates for the "receiving" locations. These concepts are borrowed and employed in this study. In the following analysis, the "sending" and "receiving" locations are indicated for individual counties.

In each of the following case analyses, first both the *overall* and the *sprawl* household growth in EA are presented; then the sprawl counties as well as the non-sprawl

[125] In detail, the receiving locations include urban center, urban, suburban, and some rural center counties.

urban and suburban counties are listed. Due to their insignificant sprawl growth, the non-sprawling, low-growth rural and undeveloped counties are not listed in these analyses.

ATLANTA, GA-AL-NC EA (EA 40)

The Atlanta, GA-AL-NC EA will experience the seventh greatest *overall* household growth and the third largest *sprawl* household growth nationwide during the 2000 to 2020 time period. Households in this area will increase by 38 percent (or 764,981 households) to 2,795,619 in 2020; jobs will grow by 37 percent (or 1,257,383 jobs) to 4,702,282 in 2020. One third (or 252,694) of the overall household growth is sprawl household growth.

This South Region EA contains 67 counties. Of these 67 counties, 15 are sprawling rural, suburban, or rural center counties[126], representing sending locations; seven are non-sprawling urban and suburban counties, representing receiving locations[127]; and the remaining 45 are non-sprawling, low-growth rural or undeveloped counties. Under the sprawl or uncontrolled-growth scenario, sprawl counties increase by an average of 16,846 households; while non-sprawl urban and suburban counties increase by an average of 53,844 households.

The 15 sprawl counties[128] include Cherokee, Henry, Forsyth, Paulding, Fayette, Coweta, Douglas, Hall, Rockdale, Oconee, Dawson, White, Madison, Union, and Lumpkin County, in order of descending sprawl household growth. Among these 15 counties, Cherokee County will experience the largest sprawl household growth, or

[126] Of these 15 sprawling counties, eight are rural counties, four are suburban counties, and the remaining three are rural center counties.
[127] Of these seven counties, five are urban counties; one is a suburban county, and the remaining county is a rural center.
[128] All of these 15 counties are located in GA.

40,959 households, followed by Henry County, which will increase by 33,373 households. The four sprawl suburban counties are Forsyth, Fayette, Douglas, and Rockdale. These developing suburban counties will experience significant growth during this time period; meanwhile, their urban population percentage shares of total county population are less than 85 percent[129]. Therefore, they are classified as sprawl counties. The three sprawling rural center counties are Cherokee, Henry, and Hall. With significant growth between 2000 and 2020, as well as urban population percentage shares that are below 80 percent, they are classified as sprawl in this study.

The seven non-sprawling urban and suburban counties[130] consist of Gwinnett, Cobb, Fulton, De Kalb, Clayton, Clarke, and Spalding County, in order of descending sprawl household growth. Among these seven counties, Gwinnett County will experience the largest sprawl growth of 128,431 households, followed by Cobb County, which will increase by 127,612 households.

LOS ANGELES-RIVERSIDE-ORANGE COUNTY, CA-AZ (EA 160)

The Los Angeles-Riverside-Orange, CA-AZ EA will experience both the largest *overall* household growth and the largest *sprawl* household growth among all EAs during the 2000 to 2020 time period. Households in the EA will increase by 24 percent (or 1,436,989 households) to 7,339,730 households in 2020; jobs will increase by 26 percent (or 2,573,292 jobs) to 12,583,868 in 2020. Twenty-eight percent (or 404,005) of the overall household growth in this EA is sprawl household growth[131].

---

[129] As defined in Chapter 3, the 85 percent of county population as urban population is one of the thresholds defining developed counties.

[130] All of these seven counties are located in GA.

[131] As defined in this study, the 404,005 households is the total household growth that will take place in the sprawl counties.

This West Region EA comprises ten counties. Of these ten counties, five are sprawling rural counties, representing sending locations; four are non-sprawling urban-center and suburban counties, representing receiving locations; and the remaining one is a low-growth undeveloped county[132]. Under the sprawl or uncontrolled-growth scenario, sprawl counties increase by an average of 80,801 households; urban center and suburban counties increase by an average of 255,154 households.

The five sprawling counties[133] include San Bernardino, Kern, San Luis Obispo, Santa Barbara, and Yuma and La Paz (AZ) County, in order of descending sprawl household growth. Among these five counties, San Bernardino boasts the greatest sprawl growth of 235,262 households, followed by Kern with a growth of 72,582 households.

The four non-sprawling urban-center and suburban counties consist of Los Angeles (urban center), Riverside (rural center), Orange (urban center), and Ventura (suburban), in the descending order of sprawl household growth. Among these four counties, Los Angeles will experience the largest growth of 334,478 households, followed by Riverside with a growth of 307,587 households.

NEW YORK-NO. NEW JER.-LONG ISLAND, NY-NJ-CT-PA-MA (EA 10)

The New York-North New Jersey-Long Island, NY-NJ-CT-PA-MA EA will experience the second largest *overall* household growth and the sixth largest *sprawl* household growth nationwide between 2000 and 2020. The households in this EA will increase by 13 percent (or 1,204,837 households) during this time period, reaching a total of 10,715,476 households in 2020; jobs will increase by 15 percent (or 2,117,470 jobs)

---

[132] This undeveloped county is Imperial County.
[133] Except for Yuma and La Paz County which are located in AZ, the other four counties sit in CA.

for a total of 16,667,585 in 2020. Around 19 percent (or 232,398 households) of this EA's overall household growth is sprawl growth.

This Northeast Region EA comprises 58 counties. Of these 58 counties, 17 counties are sprawling rural and rural center counties[134], representing sending locations; 26 are non-sprawling urban/urban center and suburban counties, representing receiving locations[135]; and the remaining 15 are non-sprawling, slow- or no-growth rural and undeveloped counties. Under the sprawl or uncontrolled-growth scenario, sprawl counties increase by an average of 13,670 households; non-sprawl urban center, urban, and suburban counties increase by an average of 35,886 households.

The 17 sprawling counties include Orange, NY; Monroe, PA; Dutchess, NY; Northampton, PA; Hunterdon, NJ; Sussex, NJ; Middlesex, CT; Litchfield, CT; Tolland, CT; Putnam, NY; Warren, NJ; New London, CT; Pike, PA; Windham, CT; Wayne, PA; Union, PA; and Luzerne, PA[136] (in order of descending sprawl household growth). Orange, NY will experience the greatest sprawl growth of 37,152 households, followed by Monroe, PA with a growth of 22,366 households.

The 26 non-sprawling urban and suburban areas consist of the counties of Queens, NY; Suffolk, NY; Kings, NY; Ocean, NJ; Middlesex, NJ; Monmouth, NJ; Bronx, NY; Richmond, NY; Somerset, NJ; Morris, NJ; Westchester, NY; Fairfield, CT; Nassau, NY; New Haven, CT; New York, NY; Lehigh, PA; Mercer, NJ; Rockland, NY; Bergen, NJ; Hartford, CT; Passaic, NJ; Hudson, NJ; Union, NJ; Essex, NJ; Hampden, MA; and

---

[134] Of these 17 sprawling counties, ten are rural counties; five are rural center counties; one is a suburban county, and the final is an undeveloped county.

[135] Of these 26 non-sprawling urban and suburban counties, 12 are suburban counties; seven are urban counties; and the remaining seven are urban center counties.

[136] Luzerne, PA, a rural center county, will lose 3311 households between 2000 and 2020. It is classified as a sprawl county due to its significant employment growth. Jobs in this EA will increase by 17 percent (or 29,218 jobs) during this time period.

Lackawanna, PA (in the order of descending sprawl household growth). Among these 26 counties, Queens, NY will experience the largest growth of 94,693 households, followed by Suffolk, NY with a growth of 89,453 households.

SAN FRANCISCO-OAKLAND-SAN JOSE, CA (EA 163)

The San Francisco-Oakland-San Jose, CA EA will experience the fifth greatest *overall* household growth and the second largest *sprawl* household growth nationwide during this time period. The households in this EA will increase by 28 percent (or 907,694 households) to 4,157,174 in 2020; jobs will increase by 27 percent (or 1,568,180 jobs) to 7,348,983 in 2020. Thirty-six percent (or 327,204 households) of this EA's overall household growth is sprawl household growth.

This West Region EA consists of 22 counties. Of these 22 counties, nine are sprawling rural and undeveloped counties, representing sending locations; eight are non-sprawling urban or suburban counties[137], representing receiving locations; and the remaining five are non-sprawling, slow-growth undeveloped and rural counties. Under the sprawl or uncontrolled-growth scenario, sprawl counties increase by an average of 36,356 households; non-sprawl urban and suburban counties increase by an average of 69,318 households.

The nine sprawling counties[138] include Sonoma, San Joaquin, Stanislaus, Monterey, Merced, Napa, Lake, Calaveras, and San Benito, in order of descending sprawl household growth. Among these nine counties, Sonoma County will experience the largest sprawl

---

[137] Of these eight counties, three are urban/urban center counties; three are suburban counties; and two are rural center counties.

[138] All of these nine counties are located in CA.

growth of 71,954 households, followed by San Joaquin County with a growth of 71,676 households.

The eight non-sprawling urban and suburban counties[139] comprises Contra Costa, Alameda, Santa Clara, Solano, San Mateo, Santa Cruz, San Francisco, and Marin, in order of descending sprawl household growth. Among these eight counties, the suburban county Contra Costa will experience the largest growth of 137,986 households, followed by the urban county Alameda which will increase by 132,219 households.

WASHINGTON-BALTIMORE, DC-MD-VA-WV-PA (EA 13)

The Washington-Baltimore, DC-MD-VA-WV-PA EA will experience the fourth greatest *overall* household growth and also the fourth greatest *sprawl* household growth nationwide. Households in this EA will increase by 29 percent (or 937,491 households) to 4,133,000 in 2020; jobs will increase by 28 percent (1,564,919 jobs) to 7,116,892 in 2020. One quarter (or 236,559) of the overall household growth of this EA is sprawl household growth.

This South Region EA comprises 52 counties. Of these 52 counties, ten are sprawling rural or rural center counties[140], representing sending locations; 14 are non-sprawling urban and suburban counties[141], representing receiving locations; and the remaining 28 are non-sprawling, low-growth rural and undeveloped counties. Under the sprawl or uncontrolled-growth scenario, sprawl counties will increase by an average of 23,656 households; non-sprawl urban and suburban counties will increase by an average of 44,721 households.

---

[139] All of these eight counties are located in CA.
[140] Of these ten sprawl counties, eight are rural counties; and two are rural center counties.
[141] Of these 14 non-sprawling core counties, four are urban center counties; three are urban counties; five are suburban counties; and the remaining two are rural center counties.

The ten sprawling counties, in order of descending sprawl household growth, are Frederick, MD; Harford, MD; Spotsylvania, VA[142]; Carroll, MD; Charles, MD; Stafford, VA; Calvert, MD; Frederick, VA[143]; St. Marys, MD; And Fauquier, VA. Among these ten counties, Frederick, MD will experience the greatest sprawl growth of 43,617 households, followed by Harford, MD, which will increase by 41,010 households.

The 14 non-sprawling urban and suburban counties, in order of descending sprawl household growth, are Fairfax, VA[144]; Montgomery, MD; Prince William, VA[145]; Anne Arundel, MD; Howard, MD; Baltimore, MD; Prince Georges, MD; Loudoun, VA; Berkeley, WV; Alexandria (Independent City), VA; Jefferson, WV; Arlington, VA; District Of Columbia; and Baltimore (Independent City), MD. Among these 14 counties, Fairfax County, VA will experience the largest sprawl household growth of 146,031 households, followed by Montgomery, MD with a growth of 84,885 households.

According to the author's practical knowledge of the counties in these EAs, the sprawl/non-sprawl classifications for these counties conform to their situations in the real world. This conformability practically supports that the method of sprawl measurements employed in this study is appropriate.

4.6 Summary

This chapter presents five groupings of results developed during the process of sprawl measurements, which are summarized below.

---

[142] It is the county group that consists of Spotsylvania and Fredericksburg, VA.
[143] It is the county group that comprises Frederick and Winchester, VA.
[144] It is the county group that consists of Fairfax, Fairfax City, and Falls Church, VA.
[145] It is the county group that comprises Prince William, Manassas, and Manssas Park, VA.

First, the four state density groups are presented and analyzed. Refined densities are calculated for each of the 50 US states. These states are then categorized into four density groups: very-low, low, moderate, and high. Corresponding to the overall low density of the country[146], 40 percent of US states (or 19 states) are in the low-density group, followed by one-quarter (13 states) in the moderate range. Of the remaining 18 states, nine are in each of the very low- and high- density groups.

Further, the differences between refined densities and gross densities for states are investigated. Both the *refined* and the *gross* densities are presented for each of the 50 states. With no exception, the "refined density" (i.e. the "developable lands density") of an individual state is larger than its "gross density". About 65 percent (or 32) of the US states have "refined densities" that are at least 1.2 times their gross densities.

The analysis of the differences in values between these two kinds of densities shows that the differences considerably affect the assignment of county development types. If gross densities were used to classify state density groups, seven states (Nevada, Utah, Kansas, Louisiana, California, Pennsylvania, and Ohio) would be placed into different density groups outside their current assignation (the one based on refined densities). This would affect the classifications of land use types and further affect the sprawl/non-sprawl classifications for some particular counties.

Second, the land use types of US counties are analyzed. Based on the variable-density approach and refined densities, one of six land use types (urban center, urban, suburban, rural center, rural, or undeveloped) is assigned to each county. Generally, the lower the level of county type, the more counties are in that classification group. Of the

---

[146] Overall, the density of the country as a whole is 46.3 households per square mile in the year 2000.

3091 US counties, nearly 60 percent are undeveloped counties; approximately one-quarter are rural counties; seven percent are suburban counties; and the remaining four percent are urban and urban center counties.

Third, projected household growth during the 2000 to 2020 time period is analyzed at the national, state, regional, EA, and county level[147]. Household and employment growth will not be evenly distributed across the country (Burchell et al. 2002). More than one-half of the projected national household growth from 2000 to 2020 will occur in only ten states.

Fourth, incidences of sprawl are analyzed at the state, regional, EA, and county level. Sixteen percent, or 492 of all US counties, will experience sprawl growth during the 2000 to 2020 time period. Although significant sprawl will be found in only16 percent of all US counties, it will affect about one-third of new development nationally. Of the 492 sprawl counties, over 80 percent (or 396 counties) are rural or undeveloped counties; only one-fifth are developing suburban and rural center counties.

At the state level, national sprawl household growth will concentrate in a small number of states. The top three sprawl growth states, California, Texas, and Florida, account for nearly 30 percent of all national sprawl household growth; the top ten sprawl growth states represent more than one-half of national sprawl growth.

Sprawl trends follow growth trends (Burchell et al. 2002). The states with the largest amount of overall growth are also those with the largest amount of sprawl growth. Although there is a strong linear relationship between the amount of sprawl growth and

---

[147] Growth at the EA and county levels are addressed together with sprawl growth in section 4.4.

overall growth for states, very often the ranking orders of these two kinds of growth for an individual state are different.

At the regional level, sprawl growth will be most significant in the South Region, which will contain half the future sprawl counties and half the projected sprawl household growth during the 2000 to 2020 time period. The West will have the second largest amount of sprawl growth; the Midwest will contain the second largest number of sprawl counties.

At the EA level, eighty percent (or 136) of the 172 EAs will experience sprawl growth over this time period. As at the state and regional levels, sprawl growth tends to concentrate in a few EAs. The top 10 sprawl growth EAs together account for more than one-third (36 percent) of national sprawl household growth; further, the top 20 sprawl growth EAs represent more than one-half (54 percent) of national sprawl household growth.

At the county level, sprawl growth is also a phenomenon of concentration. The top 20 sprawl-growth counties represent nearly one-fifth of national sprawl household growth; the top 100 sprawl counties account for more than one-half (53 percent) of national sprawl growth; and the top 200 sprawl counties contribute to three-quarters of national sprawl growth. Most of the top 30 sprawl growth counties are in the South and the West. Nearly one-half of these 30 counties are concentrated in only two states: California and Florida.

While sprawl growth ranks and overall growth ranks are close to each other for states and EAs, they are significantly different for counties. In fact, all 492 sprawl counties have much more significant sprawl growth ranks than their overall growth ranks. This is because a substantial number of the non-sprawl, urban and suburban counties have greater amounts of growth than the sprawl, rural, and undeveloped counties do. With these non-sprawl, big-growth developed counties getting involved in the overall growth ranking, the sprawl counties' overall growth ranks are much less significant than their sprawl growth ranks.

Finally, the sprawl/non-sprawl classifications for individual counties are investigated for five selected EAs. These example EAs are nationally recognizable and contribute significantly to national sprawl household growth. The conformability of sprawl/non-sprawl classifications to the real world situation for the counties in these EAs implies that the method employed in this study is appropriate.

Chapter 5

Comparing Three Research Results with the Research Result of this Study

The research results contained here will be compared with the research results of three existing studies (Pendall 1999; Ewing et al. 2003b; Burchell et al. 2002). As described in Chapter 2, the basic definition of sprawl may be more or less different in these studies. Ewing et al. (2003b) defined sprawl as places with low population densities at a given time point (the year 2000); Pendall (1999) defined sprawl as low population densities encompassing newly developed urban lands over the time period 1982 to 1992. If a rural/undeveloped county (i.e. a place likely to not have infrastructure) will experience significant growth[148] (i.e. significant demands on infrastructure) over the time period 2000 to 2025 or over the period 2000 to 2020, it is defined as sprawl by Burchell et al. (2002) and this study, respectively.

The sampling methods, land use types (if applicable), and sprawl classifications will be compared between these three studies and the research contained here. Through comparison, the following questions will be answered: 1) Is an accurate density calculation necessary for sprawl measurement? 2) Is a sample solely focused on metropolitan areas large enough for a study measuring sprawl nationwide? 3) What are the overlaps and discrepancies of sprawl classifications between the three studies and this study and why do these discrepancies exist? and 4) Why is a variable-density approach

---

[148] Compared with other counties in the EA or with the average national growth level.

potentially better than a static-density approach? These questions will be addressed in the following sections.

5.1 Comparing Research Results between Burchell et al.'s Study and this Study

The classifications of state density groups are different between Burchell et al.'s study and this study, primarily due to the different density calculation methods employed in these two studies. The differences in classifications of state density groups between these two studies have been addressed in Chapter 4[149], with a conclusion that the accurate densities developed in this study better reflect the true density of any given county in real world analyses. Therefore, in the following sections, emphasis will be placed on the comparisons of county land use types and sprawl classifications between these two studies.

5.1.1 Difference in County Land Use Type Classifications

Several reasons contribute to the differences in classifications of county land use types between these two studies (Burchell's versus this study). These reasons include the different density calculation methods utilized, the varying classifications placed upon state density groups, complex classification criterion for rural centers, and multiple methods for adjusting county land use types between these two studies. For efficiency purpose, in the following sections, the Burchell et al.'s study (2002) will be also referred

---

[149] The classifications of state density groups based on the gross densities employed in Chapter 4 represent the classifications developed in Burchell et al.'s study (2002).

to as the “Report”, an abbreviation for the name of their published work (Costs of Sprawl—2000).

Both Table 5.1 and Figure 5.1 compare the numbers of counties in each county land use type between the Report and this study. There are only slight differences between these two studies in the number of counties for three of the six land use type categories: suburban, urban, and urban center. However, significant differences exist for the other three land use categories: this study has less undeveloped counties but more rural and rural center counties than Burchell et al.’s study. The reasons for these differences are addressed below.

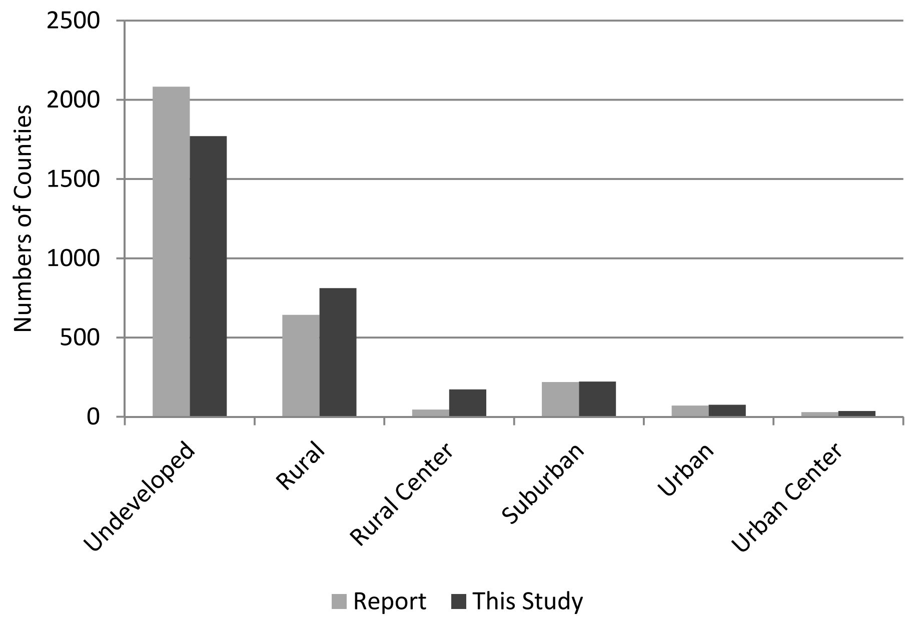
Table 5.1 Numbers of Counties by County Types in the Two Studies

County Type	In the Report (A)	In This Study (B)	Differences (C)
Undeveloped	2083	1771	-312
Rural	643	812	169
Rural Center	46	173	127
Suburban	219	222	3
Urban	71	76	5
Urban Center	29	37	8
Total	3091	3091	

Note:

Difference (Column C) is calculated as Column B minus Column A.

Figure 5.1 Differences in Numbers of Counties by County Types between the two studies



There are more counties classified as undeveloped in the Report than in this study, mainly because 317 of the 2083 undeveloped counties in the Report are classified not as undeveloped, but as rural in this study. Further, gaining these 317 counties, the rural county group in this study (which contains 812 rural counties) is much larger than that in the Report (which contains 643 rural counties).

The difference in the number of rural center counties between these two studies is mainly caused by the different classification criteria for this type of county. In the Report, rural center counties are exclusively defined for the "Rural EAs"[150]; that is, the economic center counties of these "Rural EAs", which contain the largest economic node in EA, are classified as rural centers. Overall, 46 rural center counties are defined for the 45 "Rural EAs" in the Report[151].

---

[150] "Rural EAs" is defined in Chapter 3.
[151] One rural center county is assigned to each of the 44 "Rural EAs"; while two rural centers are defined for the Casper, WY-ID-UT EA.

However, in this study rural centers are defined by two criteria: 1) the county household densities and 2) the economic center counties in an EA. As described in Chapter 3, rural center counties are first defined based on their densities; thus all EAs, including those non-Rural EAs that already have suburban and higher levels of county types, may have rural center counties. Secondly, for the 43 “Rural EAs” in this study, certain rural/undeveloped counties are adjusted upwards to rural center counties if they have the highest densities in, *or* contain the economic nodes of, the EA. Due to these differences in criteria, 105 counties that were previously classified as rural/undeveloped[152] in the Report are now classified as rural centers in this study; thus there are many more rural center counties in this study (in total 173) than in the Report (46). Significant development in a rural center county in the “Rural EAs” does not count as sprawl because it is development in the densest portion of the EA.

The abovementioned variance between the two studies in the number of counties for each land use type is a rather significant difference. A further underlying, deep-level difference that plays into the equation would be the divergent sprawl/non-sprawl classifications between the two studies for the *same* individual county. A closer examination of the reasoning behind any deep-level differences here could reveal the impact of different sprawl measurement methods on sprawl measurement results.

These deep-level differences will be addressed in the following sections. First, different land use type classifications for the same individual county and the impacts of such differences on sprawl classification for the county will be addressed. Then, sprawl classifications of the same individual county will be addressed for both studies (including

---

[152] Of these 105 counties, 91 are classified as rural and 14 are classified as undeveloped in Burchell et al.’s study.

both same and different classifications) and the reasons behind that classification will be investigated in the subsequent section.

The classification of the current land use type for an individual county may affect its sprawl classification. Specifically, two kinds of changes in county type classifications may affect the sprawl classification of an individual county. One is the change from the rural/undeveloped county group to the suburban/rural center county group, or vice versa; the other is the change from the urban/urban center county group to any other county group, or vice versa. The second kind of change will affect the sprawl classification for an individual county because as long as it is classified as an urban/urban center county, it will be classified as non-sprawl.

The first kind of change may affect sprawl classification because the criteria are different between these two development classification groups. That is, as explained in Chapter 3 and noted above, the sprawl classification criteria for suburban/rural center counties are looser than those for rural/undeveloped counties. For example, if two counties (one rural and the other suburban) will experience significant growth during the period 2000 to 2020, then growth in the rural county will be defined as sprawl growth. However, growth in the suburban county may be classified as non-sprawl if the county has a high percentage of all population as urban population, or if the county meets the sprawl adjustment criteria[153].

Therefore, the six county land use types are combined into the abovementioned three land use type groups when analyzing the impact of different land use types between the two studies for sprawl classifications of counties. Table 5.2 compares the number of

---

[153] As explained in Chapter 3, only suburban and rural center counties can be adjusted from sprawl to non-sprawl; that is, sprawl adjustments are not applied to the rural/undeveloped counties in this study.

counties by county land use types between the Report and this study. In Table 5.2, cells in the diagonal line running from the upper left to lower right store the counties that have the same county type in both studies; other cells store the counties that have been assigned different county types by the two studies. For example, 2573 counties are classified as undeveloped/rural counties in both studies; 151 counties are classified as suburban/rural center in this study but as rural/undeveloped in Burchell et al.'s study. The first row in Table 5.2 shows that of the 2583 counties classified as rural/undeveloped in this study, 2573 are also classified as rural/ undeveloped in the Report, but 10 are classified as suburban/ rural center in the Report.

Table 5.2 Number of Counties by Land Use Type Classifications in the Two Studies

County Types in this Study	County Types in the Report		
	Rural & Undeveloped	Suburban & Rural Center	Urban & Urban Center
Rural & Undeveloped	2573	**10**	0
Suburban & Rural Center	**151**	232	**12**
Urban & Urban Center	**2**	**23**	88

Note:

1. The Report represents Burchell et al's study (2002).
2. Counties that fall in different combined land use type groups between the two studies are shown in bold.

How will the abovementioned two kinds of changes in county land use types between the two studies affect sprawl classification? For the first type of change, as shown in Table 5.2, 151 counties are defined as rural/undeveloped in the Report but as suburban/rural center in this study. Of these 151 counties, 38 are classified as sprawl in the Report but as non-sprawl in this study, due to the different county type classifications for these counties. These 38 counties will experience significant growth in both studies. In the Report, they are classified as sprawl because they are rural/undeveloped counties. In this study, they are classified as non-sprawl because they are suburban/rural center counties that either have high percentages of all population as urban population (26 counties) or meet the sprawl adjustment criteria (12 counties).

For the sample used in this study, the impact of the abovementioned second type of change in county land use types between the two studies for sprawl classification is not significant. As shown in Table 5.2, of the 37 counties that are classified as urban/urban center counties in one study but as other county types in the other, only eight have different sprawl classifications between the two studies[154] (these eight counties are different because they have significant growth).

The impacts of two kinds of changes in an individual county's land use type classification between the two studies regarding sprawl classification have been addressed in this section. Overall, at least 46 counties' sprawl classifications are affected by these two changes[155]. In detail, 38 counties are classified as *sprawling* rural/undeveloped counties in Burchell's study but as *non-sprawling* suburban/rural center counties in this study; 8 counties are classified as *sprawling* suburban/rural center counties in the Burchell study but as *non-sprawling* urban/urban center counties in this study.

---

[154] Six of these eight counties are classified as *sprawling* suburban/rural center counties in the Report but as *non-sprawling* urban/urban center counties in this study.

[155] Four counties are classified as *non-sprawling* suburban/urban center counties in the Report but as *sprawling* rural/undeveloped counties in this study. However, the reasons they are classified as non-sprawling in the Report are complicated. Due to limited information, it is hard to decide whether differences in sprawl classifications for these four counties are exclusively due to their different development type classifications between the two studies. Therefore, these four counties are *not* counted in the abovementioned 46 counties.

In addition to the change in a county's land use classification (which results from the different density calculations between these two studies), other factors may also result in different sprawl classifications for the same individual counties between these two studies. These factors will be addressed in the next section.

5.1.2 Difference in Sprawl Classifications

Two kinds of differences in sprawl classifications between the Report and this study will be addressed in this section. First, the difference in the number of sprawl counties between the two studies will be investigated; then the difference in the amounts of sprawl household growth between the two studies will be analyzed.

Difference in Numbers of Sprawl Counties

The differences in sprawl measurement methods between this study and Burchell et al.'s study mainly exist in the following aspects: household density calculation, the classification criteria for rural center counties, adjustments of county types, the sprawl classification criteria for suburban/rural center counties[156], and sprawl adjustments[157]. Further, different versions of the Woods & Poole datasets are employed in these two studies. These differences may affect the sprawl/non-sprawl classifications of counties. In this section, first the differences of sprawl classification results between these two studies will be described, and then the reasons causing these differences will be investigated.

---

[156] Burchell et al. (2002) used GIS datasets and population data to adjust the suburban counties that satisfied certain criteria from sprawl to non-sprawl.

[157] Burchell et al. also adjust sprawl classification for some counties (from sprawl to non-sprawl), but not in as grand a scale as this study does.

Overall, the number of sprawl counties in this study is 73 less than that in the Report; 492 counties will experience sprawl during the 2000 to 2020 period in this study; 565 counties will experience sprawl during the 2000 to 2025 period in Burchell et al.'s study.

Table 5.3 compares sprawl classification results between these two studies. The cells in the diagonal line running from the upper left to lower right store counties that have the same sprawl/non-sprawl classification in both studies; the remaining cells contain counties that have different sprawl classifications between these two studies. For example, 145 counties are classified as sprawl in this study but non-sprawl in Burchell et al.'s study.

Table 5.3 Number of Counties by Sprawl Classifications between the Two Studies

This Study	The Report		Total
	Sprawl	Non-sprawl	
Sprawl	347	**145**	**492**
Non-sprawl	**218**	2381	2599
Total	**565**	2526	3091

Note:

1. The numbers of sprawl counties in both studies are shown in bold (i.e. 565 sprawl counties in the Report and 492 sprawl counties in this study).
2. The numbers of counties that have different sprawl/non-sprawl classifications between the two studies are shown in bold.

As shown in Table 5.3, nearly 90 percent (88 percent or 2728) of the 3091 US counties have the same sprawl/non-sprawl classification in both studies: 2381 counties are classified as non-sprawl in both studies; and 347 counties are classified as sprawl in both studies. However, 363 counties in total have different sprawl/non-sprawl classifications between the two studies: 218 counties are classified as non-sprawl in this study but as sprawl in the Report; 145 counties are classified as sprawl in this study but as non-sprawl in the Report. These overlaps and differences in sprawl results between the two studies, as well as their reasons, are addressed below.

Of the 2381 counties that are classified as non-sprawl in both studies, 88 percent (or 2094 counties) are very-low growth rural/undeveloped counties in both studies; 5 percent are developed suburban/rural center counties in both studies; and nearly 4 percent (or 88 counties) are urban/urban center counties in both studies. Land use types of these 2381 counties largely overlap between the two studies. If these 2381 counties are grouped into the abovementioned three county type groups that will not affect sprawl classifications[158], then nearly 97 percent of the 2381 counties fall in the same county type group in both studies.

All of the 347 counties that are classified as sprawl in both studies will experience significant growth in the future. Of these 347 counties, 76 percent (or 265 counties) are rural/undeveloped counties in both studies; and 8 percent (or 28 counties) are developing suburban/rural center counties in both studies. The other 16 percent (or 54 counties) are assigned different county types by these two studies. 51 of these 54 counties are classified as rural/undeveloped counties in the Report but as suburban/rural center counties in this study.

County land use types of the 2728 counties that have the same sprawl/non-sprawl classification between the two studies have been addressed. The 363 counties with different sprawl classifications will be investigated in the following section.

---

[158] These three county type groups are: rural/undeveloped county group, suburban/rural center group, and urban/urban center group.

218 counties are classified as sprawl in the Report but as non-sprawl in this study. 113 of these 218 counties will not experience significant household or employment growth during the 2000 to 2020 period in this study, but will experience significant growth during the 2000 to 2025 period in the Report, mainly due to the different versions of Woods & Poole datasets employed in these two studies, but also due to the different time periods for which projected growth is analyzed in these two studies. Of these 113 very-low growth counties, 80 percent are rural/undeveloped counties in this study; 14 percent are suburban/ rural center counties; and 6 percent are urban/ urban center counties in this study[159].

In addition, 105 of the abovementioned 218 counties will experience significant growth during the period 2000 to 2020 in this study. All of these 105 counties are suburban/rural center counties in this study[160]. Of these 105 counties, 71 are classified as non-sprawl in this study because they are developed suburban/ rural center counties[161]; the other 34 counties satisfy the sprawl adjustment criteria and thus are adjusted from sprawl to non-sprawl.

In summary, several reasons result in these 218 counties being classified as sprawl in the Report but as non-sprawl in this study. For 113 of these 218 counties, the major reasons include the different datasets used and different growth periods measured in the two studies (this study has better and more current information); for the other 105 counties, the major reasons for differences between the two studies can be attributed to

---

[159] In Burchell et al.'s study, 88 percent of these 113 counties are rural/undeveloped counties; 12 percent are suburban/rural center counties.
[160] In Burchell et al.'s study, of these 105 counties, 64 percent are suburban/rural center counties, and 33 percent are rural counties.
[161] The urban population percentage shares of all population in these counties range from 85.1 percent to 98.3 percent for the suburban counties and from 80.1 percent to 95.8 percent for the rural center counties.

the different classifications of land use types (caused by different density calculations[162]), the different sprawl classification criteria for suburban/ rural center counties, and the different sprawl adjustment methods (again, this study contains improved methods for county designation).

Conversely, the other difference in sprawl classification between these two studies is that 145 counties are classified as non-sprawl in the Report but as sprawl in this study. In the Report, 90 percent of these 145 counties are very-low growth rural/undeveloped counties; and nearly 10 percent are suburban counties. However, in this study, 88 percent are classified as rural/undeveloped counties with significant growth; and the other 12 percent are developing suburban/ rural center counties with significant growth in the future. In both cases they are counted as sprawl locations.

The major reasons for the difference in sprawl classifications of these 145 counties might well be the different datasets employed and the variance in time periods for which sprawl is measured in these two studies. While these counties would not experience significant growth during the 2000 to 2025 period based on the 1995 Woods & Poole dataset (which is employed in Burchell et al.'s study, the Report); they will do so during the 2000 to 2020 period based on the 2003 Woods & Poole dataset (which is used in this study). These counties are developing suburban/rural center counties in this study whose percentage shares of all population in said counties defined as urban population are not high enough to consider them being classified as non-sprawl. Further, they do not satisfy the criteria for any sprawl adjustment. Thus these counties are classified as sprawl locations in this study.

---

[162] The major difference is higher density in this study due to lands classified as undevelopable. This leads to locations, due to their higher density, that are not classified as sprawl.

The difference in county type classifications between these two studies is not a significant reason for the different sprawl classifications of these 145 counties between the two studies. Of these 145 counties, 74.5 percent have the same county type classifications in both studies; 17 percent are classified as rural in this study but as undeveloped in the Report (as explained in the foregoing section, this kind of difference in county type classifications does not affect their sprawl growth definitions).

The difference in the numbers of sprawl counties between the two studies has been addressed in this section; the difference in the amounts of sprawl household growth between the two studies will be discussed in the next section.

Difference in Amounts of Sprawl Household Growth

The difference in the amounts of national sprawl household growth between the two studies is reasonably large. The sprawl household growth is 11,079,960 households in the Report[163] but only 7,038,865[164] households in this study; that is, sprawl household growth would decrease by 36 percent (or 4,041,095 households) by moving from Burchell et al.'s research definitions to the research definitions in this study.

There are two primary reasons for the 36 percent (or 4,041,095) decrease in households in this study when compared to the Report. First, there are simply more counties defined as sprawl in the Report than in this study. There are 565 sprawl counties in the Report but only 492 sprawl counties in this study. With the additional 73 sprawl

---

[163] As stated in the foregoing section, the two studies used different datasets. As a reference, the overall national household growth is 25,311,676 during the *2000 to 2020* time period in the Woods & Poole 2003 dataset (which is used in this study); while it is 23,454,410 during the *2000 to 2025* period in the Woods & Poole 1995 dataset (which is employed in Burchell et al.'s study).

[164] "Sprawl growth households" is defined as the total household growth for the counties that are classified as sprawl for the 2000 to 2020 period (or for the 2000 to 2025 period in the Report). These "sprawl" counties include both household sprawl counties and employment sprawl counties.

counties, the Report would clearly have a larger amount of the total sprawl household growth than found in this study.

The second reason for the decrease plays a more important role than the first reason. Among the counties that are classified as sprawl in the Report but as non-sprawl in this study, are some suburban/rural center counties[165] with large amounts of projected household growth in the future; while in contrast, many of the counties that are classified as sprawl in this study but as non-sprawl in the Report will experience a relatively small amount of household growth in the future (i.e. more sprawl counties and more growth in the Report versus less sprawl counties and less growth in this study).

For example, of the top 12 most significant sprawl counties[166] in the Report, only one (San Bernardino, CA) is also classified as sprawl in this study; the other 11 counties are classified as non-sprawl in this study (due to the fact that they are no longer designated as rural/undeveloped counties). Now classified as urban/urban center, and developed suburban/ rural center counties in this study, these 11 counties will experience large amounts of household growth from 2000 to 2020[167], ranking as the 3rd to 45th most significant household growth counties among all US counties in this study. The total sprawl household growth[168] of these 11 counties (2,009,385 households) represents half of the sprawl household difference between these two studies. Specifically, two of these 11 counties, Clark, NV and Broward, FL, which are classified as *sprawling suburban*

---

[165] With high percentages of all population as urban population, these counties are identified as *non-sprawl, developed* suburban/rural center counties in this study.
[166] As mentioned in the foregoing section, the "Most Significant Sprawl Counties" refers to the sprawl counties that will experience the largest amount of overall growth among all sprawl counties.
[167] Burchell et al. used the overall growth between 2000 and 2025, which has no significant difference with the overall growth between 2000 and 2020.
[168] As mentioned in the foregoing section, for a sprawl county, its sprawl household growth is equal to its overall household growth.

counties in the Report but as *non-sprawling urban center and urban* counties[169], respectively, in this study contribute 16 percent of the sprawl household difference between these two studies.

In contrast to these eleven counties, the 145 counties that are classified as sprawl in this study but as non-sprawl in the Report will experience small amounts of household growth from 2000 to 2020—their overall household growth ranks from 96th to 1780th among all US counties. As a result of the reasons brought to light here, total sprawl household growth in this study is much less than the sprawl growth found in the Burchell's study.

5.2 Comparing Research Results between Ewing's Study and this Study

The research results of Ewing et al.'s study (2003a) and this study will be compared in this section. Through the comparison, four questions will be addressed: 1) Is a sample solely focusing on metropolitan areas large enough for a study measuring sprawl nationwide? 2) What are overlaps and discrepancies of sprawl classifications between these two studies? 3) What are overlaps and discrepancies of county type classifications between these two studies? and 4) Why is the variable-density approach[170] potentially better than the static-density approach? For consistency purposes, Ewing et al.'s study (2003a) will be referred to as "Ewing's study" in the following sections.

[169] Even though there is a lot of growth, it is not happening where there is a lack of infrastructure.
[170] The static-density approach and variable-density approach have been defined in Chapter 3 Data and Method.

5.2.1 Comparing Sampling Method

The sample in this study includes all 3091 counties/county groups in the US; while Ewing's sample contains all 448 counties in the 101 most populous US metropolitan areas (as of the 2000 census). Most existing studies measuring sprawl focus their samples solely on metropolitan areas. Both Ewing's sampling method, which focuses on a part of metropolitan areas, and a hypothesis method, which focuses on all metropolitan areas, will be addressed in this section.

There are 276 metropolitan areas (including 259 MSAs and 17 CMSAs) which contain 831 counties/county groups in the US in 2000[171]. Table 5.4 compares the distribution of county land use types between the metropolitan and non-metropolitan areas. As shown in Table 5.4, the percentage share of all counties as urban/urban centers or suburban/rural centers is much higher in metropolitan areas than in non-metropolitan areas. However, metropolitan areas still contain a significant number of rural/undeveloped counties: 46 percent of the 831 metropolitan counties are classified as rural/undeveloped in this study. From another perspective, 98 percent of the 113 urban/urban center counties nationwide, as well as 85 percent of all 395 suburban/rural center counties in the country are located in metropolitan areas.

Table 5.4 Number of Counties by Land use Types in Metropolitan and Non-Metropolitan Areas

Land use Type	Non-metro. Areas	Metro. Areas
Rural & Undeveloped	2198	385
Suburban & Rural Center	60	335
Urban & Urban Center	2	111

[171] These summary numbers are based on Woods & Poole 2003 dataset. As explained in the foregoing section, Woods & Poole team combined the 3141 US counties into 3091 counties/county groups.

Total	2260	831

Focusing on the most populous metropolitan areas, Ewing's sample[172] has a much lower percentage share labeled as rural/undeveloped counties and a much higher percentage share labeled as urban/urban center counties than all of the metropolitan areas in the United States. Of Ewing's 448 counties, 38 percent are rural/undeveloped counties; 41 percent are suburban/rural center counties; and 20 percent are urban/urban center counties. The percentage shares of these three land use-type groups for metropolitan areas as a whole are 46, 40, and 13 percent, respectively.

More than 70 percent of sprawl counties are rural/undeveloped counties in both the metropolitan and non-metropolitan areas. Of the 216 sprawl counties in this study that are located in non-metropolitan areas, 94 percent (or 204 counties) are rural/undeveloped counties; and less than 6 percent (or 12 counties) are suburban/rural center counties. For the other 276 sprawl counties that are located within metropolitan areas, nearly 70 percent are rural/undeveloped counties; and approximately 30 percent (or 84 counties) are suburban/rural center counties.

Ewing's sample represents 37 percent of the 276 metropolitan areas and 54 percent of all 831 metropolitan counties/county groups. Does this sample include most of the sprawl counties in the US? Further, does a sampling method that focuses on all metropolitan areas contain most of the sprawl counties nationwide? These questions will be addressed below.

[172] As explained in the foregoing section, "Ewing's sample" actually refers to "Ewing et al.'s (2003a) sample".

In order to answer the first question, it is necessary to calculate how many sprawl counties Ewing's sample has missed. Of the 492 sprawl counties in this study, only 168 counties are included in Ewing's sample. That is, with the sprawl classification results from this study being used as a reference, Ewing's sample not only misses 66 percent (or 324 sprawl counties) of all sprawl counties nationwide, but he also misses 40 percent of the 276 sprawl metropolitan counties.

In order to answer the second question, we need to know how many sprawl counties are located outside metropolitan areas. Of the 492 sprawl counties in this study, only 56 percent are located in metropolitan areas; and the other 44 percent (or 216 counties) are located outside metropolitan areas. Thus it may be inappropriate to only select metropolitan areas and counties as research subjects when measuring sprawl nationwide, since this would neglect nearly one-half of the sprawl counties in the country that are found in rural areas.

Based on the answers to the above two questions[173], it can be concluded that a sample solely focusing on the most populous metropolitan areas, or even on all metropolitan areas, indeed misses a significantly large portion of sprawl counties nationwide.

---

[173] As an interesting reference, a metropolitan sample would miss a less significant proportion of national sprawl growth than it would miss for number of sprawl counties. Of the (7,038,865) sprawl household growth nationwide during the 2000 to 2020 time period, 78 percent (or 5,460,346 households) will take place in metropolitan counties, while 22 percent will occur in non-metropolitan areas. Ewing's sample contains sprawl growth of 3,696,368 households and misses 47 percent of national sprawl household growth.

However, in this case the incidences of sprawl are of more policy significance than the amount of sprawl growth. Therefore, it is still concluded that a sample solely focusing on metropolitan areas is not enough for a study measuring sprawl nationwide.

5.2.2 Background to Comparing Sprawl Classifications

Two points must be clarified before any comparison of sprawl results is conducted between these two studies (Ewing versus this study). The first point involves the different time periods for which sprawl is measured in these two studies. As explained in Chapter 2, the sprawl index in Ewing's study is based on county densities in 2000, and thus is a static sprawl measurement at a particular point in time. However, the sprawl measurement in this study is based on county growth during the time period from 2000 to 2020. Therefore, great caution is required in interpreting the comparison results between these two studies. For example, if a county is classified as sprawl in both studies, then this means that: 1) it would be classified as sprawl for the year 2000, based on Ewing's research results; and 2) it will also experience sprawl growth during the time period from 2000 to 2020, based on the research results of this study.

Secondly, Ewing did not explicitly classify the 448 counties into binary sprawl/non-sprawl categories. Rather, these counties were sorted by their sprawl index scores in ascending order. The lower the sprawl index score for a given county, the higher the degree of sprawl for that county. Conversely, this study classifies all US counties into binary sprawl/non-sprawl categories, but does not calculate "sprawl index scores" for counties. Therefore, in order to compare the sprawl results between these two studies, a specific criterion has been developed in this study to first classify Ewing's sample into binary sprawl/non-sprawl categories; and then the sprawl/non-sprawl classifications can be compared between the two studies.

Based on Ewing's concept that "the lower the sprawl index score for a given county, the higher the degree of sprawl for that county", it is reasonable to find a particular sprawl index value which serves as the cut-off value between his sprawl and non-sprawl categories. After investigating the relationship between the sprawl index values in Ewing's study and the sprawl/non-sprawl classifications in this study, the 50th percentile of sprawl index values is used as the threshold cut-off between the sprawl and non-sprawl classification for Ewing's sample. Counties with sprawl index scores lower than the 50th percentile would be defined as sprawl for Ewing's study. To simplify for clarity, these counties will be referred to as "being classified as sprawl *in Ewing's study*" in the following sections. Based on this criterion, all 448 counties in Ewing et al.'s study are classified into binary sprawl/non-sprawl categories.

Why is the 50th percentile of Ewing's sprawl index value selected as the threshold cut-off? Table 5.5 presents the number of counties that are classified as sprawl/non-sprawl for each of the ten percentile groups in Ewing's study. Each percentile group contains 44 or 45 counties, with the 1st group having the lowest sprawl scores (i.e. most sprawl counties) and the 10th group having the highest sprawl scores (i.e. most compact development counties). As shown in Table 5.5, more than 70 percent of the counties in the 2nd, 3rd, and 4th percentile groups are classified as sprawl in this study; while in the 6th through 10th percentile groups, more than 80 percent of the counties are classified as non-sprawl in this study. Because there is an obvious gap for the percentage shares of all counties in groups that are classified as sprawl in this study between the 5th percentile group (50 percent) and the 6th percentile group (20 percent), the counties in the 5th percentile group are ascribed to Ewing's sprawl category. That is, the 50th percentile of sprawl index value is used as the threshold to identify sprawl counties in Ewing's sample.

This threshold ensures the maximum overlap of sprawl classifications between these two studies.

Some interesting facts are presented by Table 5.5. Significant discrepancies in sprawl classifications between the two studies exist in the 1st percentile group, which have the lowest sprawl index scores. The counties in this percentile group are the greatest sprawl counties in Ewing's study (i.e., their degrees of sprawl are higher than other counties). However, nearly one-half of the counties in this percentile group (22 counties) are classified as non-sprawl in this study, because they are very-low growth rural/undeveloped counties during the time period from 2000 to 2020. In other words, these 22 greatest *sprawling* counties in Ewing's study are classified as *non-sprawling* in this study, because there is actually no growth taking place in these counties. The argument "the lower the density for a given county, the higher the degree of sprawl for that county" is true in Ewing's study but it does not mean much because no growth is taking place in these locations.

Table 5.5 Sprawl Classifications in this Study for Counties in Ewing's Percentile

Percentile Group in Ewing's Study	In this Study		Total Counties
	Non-Sprawl Counties	Sprawl Counties	
1st	22	23	45
2nd	9	37	46
3rd	10	34	44
4th	13	32	45
5th	22	22	44
6th	36	9	45
7th	40	5	45
8th	40	5	45
9th	44	1	45
10th	44	0	44
Total	280	168	448

Note:

The percentile groups are based on the percentiles of Ewing's sprawl index values. The counties in the 1st percentile group have sprawl scores (in Ewing's study) lower than the 10th percentile; and those in the 10th percentile groups have sprawl scores higher than the 90th percentile.

5.2.3 Comparing Sprawl Classifications

After employing the abovementioned method to classify Ewing's 448 counties into binary sprawl/ non-sprawl categories, the sprawl classifications can now be more readily compared between these two studies. As shown in Table 5.6, one-half of the 448 counties (i.e. 224 counties) would be classified as sprawl and the other half as non-sprawl in Ewing's study. Conversely, of these 448 counties, 38 percent (or 168 counties) and 62 percent (or 280 counties) are classified in this study as sprawl and non-sprawl, respectively.

Table 5.6 presents the sprawl classifications for the same individual county between the two studies. Again, the cells on the upper left to lower right diagonal line illustrate the number of counties that have the same sprawl classifications between the two studies. Because the method classifying Ewing's sample into sprawl/non-sprawl categories is geared to ensure the maximum overlap between the two sprawl classifications (in the two studies), as a result, the two sprawl classifications indeed have a lot of sprawl/non-sprawl counties in common. As shown in Table 5.6, almost 80 percent (or 352 counties) of the 448 counties have the same sprawl/non-sprawl classifications in both studies: 148 counties and 204 counties are classified in both studies as sprawl and non-sprawl, respectively. The other 21 percent (or 96 counties) of the 448 counties have different sprawl/non-sprawl classifications between the two studies: 76 counties would be classified as sprawl in Ewing's study but as non-sprawl in this study; while in contrast, 20 counties would be classified as non-sprawl in Ewing's study but as sprawl in this study.

Table 5.6 Number of Counties by sprawl classifications in the Two Studies

In Ewing's Study	In This Study		
	sprawl	non-sprawl	total
sprawl	148	76	224
non-sprawl	20	204	224
total	168	280	448

Note:

The sprawl classifications in Ewing's study, which are based on Ewing's sprawl index scores, are developed by the author of this study.

In the following sections, overlaps of sprawl classifications between the two studies will be identified and the reasons for those overlaps will be addressed. Then the counties with discrepant sprawl classifications between the two studies will be investigated.

First, 204 counties are classified as non-sprawl in both studies. In other words, these 204 counties have relatively high densities in 2000 in Ewing's study, and will not experience sprawl growth during the time period 2000 to 2020 in this study. Three major reasons lead to this result. First, 112 of these 204 counties will not experience significant growth in the future. Of these 112 counties, most (83 percent) are urban/urban center counties; 11 percent are suburban/ rural center counties; and only 6 percent are rural/undeveloped counties in this study. Secondly, 88 of these 204 counties that will experience significant growth in the future are *developed suburban/rural center* counties (80 suburban counties and 8 rural center counties) and thus satisfy the criteria of non-sprawl classification (i.e. growth going into a developed area). Finally, 4 suburban counties meet the sprawl adjustment criteria and are adjusted from sprawl to non-sprawl in this study.

Secondly, 148 counties are classified as sprawl in both studies. The 224 counties with relatively low densities in 2000 are classified as sprawl in Ewing's study. Of these 224 counties, 148 will experience sprawl growth during the time period 2000 to 2020 in this study. In this study, all of these 148 counties will experience significant growth in the future. Of them, 68 percent (or 101 counties) are rural/undeveloped counties; and 32 percent (or 47 counties) are developing suburban/rural center counties.

The counties with the same sprawl classification in the two studies have been analyzed. The counties with different sprawl classifications between the two studies will be addressed in the following sections.

On one hand, 76 counties are classified as *sprawling* based on Ewing's sprawl scores but are classified as *non-sprawling* in this study. The different sprawl results of these 76 counties result from the two different sprawl measurement methods: Ewing measures sprawl based on density and this analysis measures sprawl based on growth and density. One (Ewing) has a static density applying to all places; the other (this study) has a variable definition of density denoting type of development according to the density group of the state. The 76 counties are classified as sprawl for the year 2000 in Ewing's study based on their relatively low densities. In this study, however, they are defined as non-sprawl for the time period 2000 to 2020, because most of these counties will not experience significant growth during this time period. In detail, of these 76 counties, 59 will not experience significant growth in the future (90 percent of these 59 counties are low-growth rural/undeveloped counties); 13 counties will experience significant projected growth and are developed suburban/ rural center counties that satisfy non-sprawl classification criteria; and 4 suburban/rural center counties are adjusted from sprawl to non-sprawl because they satisfy the sprawl adjustment criteria.

On the other hand, 20 counties are classified as *non-sprawling* based on Ewing's sprawl scores but are classified as *sprawling* in this study. These counties have relatively high densities in Ewing's sample and thus are defined as non-sprawl based on their sprawl index scores. However, they are classified as sprawling in this study, mainly due to two reasons. First, 9 of these 20 counties are located in high-density states and thus have relatively high sprawl index scores (which represent compact counties) due to Ewing's density-based sprawl measurement[174]. However, these nine counties are classified as rural counties in this study because their densities are relatively low compared with all counties in the high-density state group. In fact, most of these counties are tourist or agricultural communities; that is, they are indeed rural counties in the real world. In other words, these nine counties have relatively "high" densities if compared to some counties located in low- and moderate- density states in Ewing's sample, and thus are classified as *non-sprawl urban/suburban* counties in their study[175]. However, they have a relatively "low" density if compared with their surrounding counties in the region and with the counties in the same high-density state group and thus are classified as *sprawl rural* counties in this study. The same density value represents different land use types for a county between the two studies, mainly due to the different density

---

[174] These nine counties are St. Johns, FL, San Bernardino, CA, Sonoma, CA, Osceola, FL, Yolo, CA, Napa, CA, Stanislaus, CA, San Joaquin, CA, and Santa Barbara, CA (in the order of ascending sprawl index scores).
[175] Of the 9 counties, 3 are in each of the 6th, 7th, and 8th percentile groups of sprawl index scores. The 6th and 7th percentile groups represent suburban/rural center counties; while the 8th percentile group contains suburban and urban counties in Ewing's study.

approaches employed in the two studies: the static-density approach used in Ewing's study and the variable-density approach employed in this study.

Further, 11 of the abovementioned 20 counties are developing suburban/rural center counties that will experience significant growth in the future and thus are classified as sprawl in this study. Because these counties have similar land use type classifications in both studies[176], their different sprawl classifications mainly result from the different sprawl definitions employed in the two studies. In other words, these counties are classified as *non-sprawl* for the year 2000 in Ewing's study based on their relatively high one-type density; however, they are classified as *sprawl* in this study because they will experience significant growth during the time period 2000 to 2020 and this growth will take place in a developing context.

5.2.4 Comparing County Land Use Type Classifications

Ewing's sprawl measurements for counties for the year 2000 are relatively comparable to the county land use types in 2000 classified in this study. However, Ewing did not explicitly classify their counties into 6-category county land use types as this study does. Fortunately, concepts developed in Ewing's study make the comparison possible. As claimed by Ewing, the counties with low sprawl index scores in their study are low-density, rural-like counties; while counties with high sprawl scores are high-density urban counties. Based on this concept, this study can reasonably extrapolate that counties with medium-size sprawl scores in Ewing's study have suburban/rural center land use types.

---

[176] 7 of the 11 counties are in the 6th percentile group of sprawl scores in Ewing et al.'s study; the other 4 are evenly distributed in the 7th and 8th percentile groups.

In order to compare county land use types between these two studies, first the 448 counties are divided into four quartile groups based on their sprawl index scores. According to Ewing's concept, most of the counties in the first quartile group and a large proportion of the 2nd quartile group should be rural/undeveloped counties; most of the counties in the 4th quartile group should be urban/urban center counties; while the middle two quartile groups mainly contain suburban/rural center counties. Then, within each quartile group, the land use types classified by this study are examined to determine whether they are consistent or discrepant with the land use types based on Ewing's concepts.

Table 5.7 presents the number of counties by land use types (which are classified in this study) for each quartile group of sprawl index scores (which are developed in Ewing's study). Each quartile group contains 112 counties. The first quartile group represents the counties with the lowest sprawl scores and densities; while the 4th quartile group represents the counties with the highest sprawl scores and densities. As shown in Table 5.7 and Figure 5.2, there is a high degree of overlap in the land use types for counties between the two studies. Most of the counties in the 1st quartile group are also classified in this study as rural/undeveloped; a major proportion of the 4th quartile group are also classified as urban/urban center counties in this study; a major proportion of the 2nd quartile group and most of the 3rd group are also classified as suburban/rural center counties in this study.

Table 5.7 Number of Counties by Land Use Types in Each Quartile Group of Sprawl Index Values

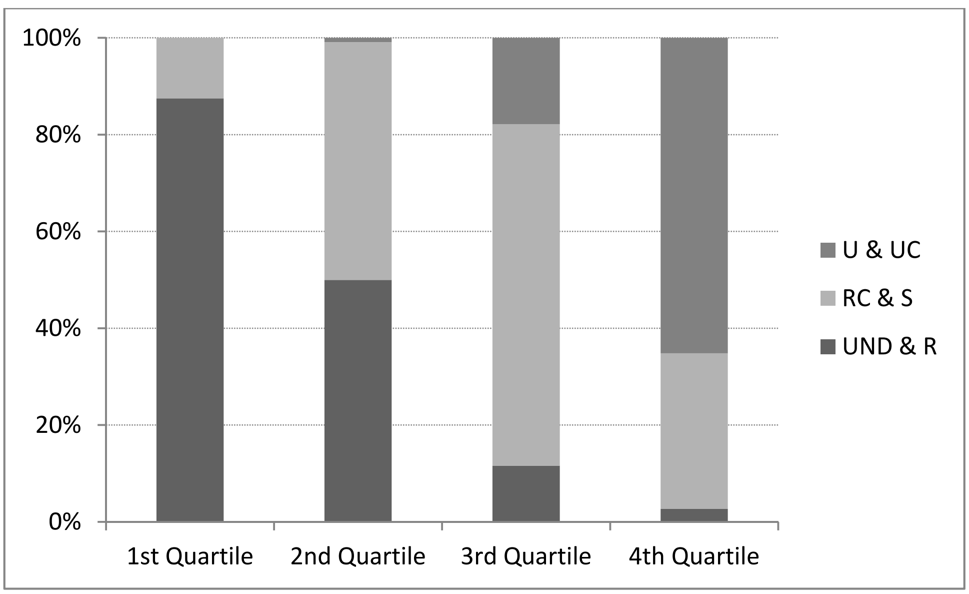
Land use Types in	Quartile Group in Ewing's Study			
	1st	2nd	3rd	4th

this Study				
UND & R[1]	98	56	13	3
RC & S	14	55	79	36
U & UC	0	1	20	73
Total	112	112	112	112

Note:

1. To save space, abbreviations for land use types are used in this table. Specifically, UND = Undeveloped; R = Rural; RC = Rural Center; S = Suburban; U = Urban; and UC = Urban Center.

Figure 5.2 Percentage Shares of County Land use Types for Each Quartile Group of Sprawl Index Values



Note:

The county land use types are developed in this study; while the quartile groups are based on the sprawl index scores developed in Ewing's study.

It is interesting to investigate the differences in county land use types between these two studies. On one hand, five counties in the 4th quartile group are classified in this study as rural or rural center counties. However, based on Ewing's theory, the 4th quartile group should only contain urban/urban center and suburban counties, because the counties in this group have the highest densities in their sample. These five counties are Stanislaus, CA, San Joaquin, CA, Santa Barbara, CA, Solano, CA, and Santa Cruz, CA. The first three counties are classified as rural counties and the latter two are classified as rural center counties in this study. These five counties have relatively high densities in Ewing's study, with the ranking orders of their sprawl index scores ranging from the 344th to 392nd, and thus are assumed to be suburban or urban counties by Ewing. They rank toward the most compact county end in Ewing's study because they are located in the high-density state of California, and thus have relatively high densities compared with most of the counties in Ewing's sample. However, these high-density counties function as rural or rural center counties in the region, due to their relatively low densities compared with surrounding counties. Therefore, they would be classified as urban/suburban in Ewing's study but are classified as rural/rural center counties in this study.

On the other hand, based on Ewing's theory, all of the 1st quartile group and part of the 2nd quartile group should be classified as rural/undeveloped counties, due to their relatively low densities. However, some of the counties in the 1st quartile group are classified as suburban/rural center counties in this study. For example, Washington, WI and De Soto, MS rank as the 61st and 77th greatest sprawl counties (or lowest density counties) in Ewing's study, respectively. They would also be defined as rural/undeveloped counties in Ewing's study, based on their relatively low densities. These two counties are located in low-density states and thus have relatively low densities compared with counties located in high- or moderate- density states in the sample. However, they function as suburban counties in the region, due to their relatively high densities compared with their surrounding counties in the region; thus they are classified as suburban counties in this study.

5.2.5 Summaries

Several conclusions can be drawn from the comparison analysis of research results between Ewing's study and this study. First, for studies aimed at measuring sprawl, a sample solely focusing on the most populous metropolitan areas is not large enough. Actually, even a sample solely focusing on metropolitan areas is not large enough, because it will miss a substantial proportion of locations where significant sprawl growth will take place. For example, 44 percent of the 492 sprawl counties in this study are located in non-metropolitan areas.

Secondly, the same individual county may have different sprawl classifications between the two studies, primarily due to two reasons. The first reason is that different sprawl definitions are employed in these two studies. The second reason is that the static-density approach is used in Ewing's study while the variable-density approach is employed in this study to measure sprawl. Some counties located in high-density states are classified as non-sprawl in Ewing's study, due to their relatively *high* densities in the *sample*. However, they are classified as rural counties in this study, due to their relatively *low* densities in the *region*. Further, these counties would be classified as sprawl in this study if they will experience significant growth in the future.

Thirdly, the differences between the static- and variable- density approaches also result in different classifications of land use types between these two studies for some particular counties. For example, some counties located in low-density states may be classified as rural counties in Ewing's study, based on the static-density approach; but are classified as suburban counties in this study, based on the variable-density approach.

Finally, it can be concluded that the variable-density approach is a better instrument than the static-density approach for measuring sprawl across the country. This is true because the former takes into account the regional density differences of counties, which is vital in a national study of sprawl measurements, while the latter fails to do so. In Ewing's sample, the regional contexts of counties (which are characterized by densities of their host states) vary significantly. Of the 448 counties, 27 percent are from high-density states; 46 percent are from moderate-density states; 24 percent are from low-density states; and 3 percent are from very-low density states. In order to more accurately classify land use types and sprawl/non-sprawl for such a sample, the variable-density approach, rather than the static density approach, should be employed. The comparison analysis between these two studies provides practical examples that support this conclusion. Further, the variable-density approach indeed is a better reflection of the land use types and functions of counties in the real world than is the static-density approach.

5.3 Comparing Research Results between Pendall's Study and this Study

The research results of Pendall's study (1999) and this study are compared in this section. Through the comparison, three questions will be addressed: 1) Is Pendall's sample large enough for a study measuring sprawl nationwide? 2) What are the overlaps and discrepancies of sprawl classifications between these two studies? and 3) Why is the variable-density approach better than the static-density approach?

5.3.1 Comparing Sampling Method

Pendall's sample (1999) includes the 159 counties that gained population between 1982 and 1992 in the 25 most populous metropolitan areas[177]. Based on the description of data and method in his published work, this study replicates his study results. However, even strictly following his method, the sample re-created by this study includes 181 counties, which is 22 counties more than in Pendall's published work. There are two primary reasons for this sample discrepancy. First, the Census website has updated their population estimates for the years 1982 and 1992, which may result in more counties gaining population during this time period than the older version used by Pendall. Second, following Pendall's method, this study selects counties from the 25 most populous CMSAs and MSAs. However, for some CMSAs, Pendall might have actually used the component PMSAs instead of the whole CMSAs to select the counties, although it is claimed that the "metropolitan areas" in his study refer to CMSAs and MSAs[178].

According to the census data, 208 counties constituted the 25 metropolitan areas in Pendall's study. Based on county groups in the Woods & Poole 2003 and the NRI 1997 datasets[179], these 208 counties are combined into 200[180] counties/county groups. With

---

[177] In Pendall's article, totally there are 180 counties in the 25 most populous metropolitan areas (as of the 1990 census); with an exclusion of the 21 counties that had lost population during this time period, eventually his sample contains 159 counties.

[178] For example, the Boston CMSA contains 8 gaining-population counties in this replicated sample but only 4 gaining-population counties in his sample. Another example is the New York-Northern New Jersey-Long Island CMSA: there are 21 gaining-population counties in this study but only 15 gaining-population counties in Pendall's article.

[179] Both Woods & Poole and NRI datasets have their specific county groups. First, a list of county groups is created that contain all county groups from these two datasets; then based on this list, the census counties are combined into county groups.

[180] Washington D.C. is included in this 200-county list. However, it was excluded from the final 181-county sample because it has missing urban acre data in the NRI dataset.

counties that had lost population during the period 1982 to 1992 being excluded, there are a final 181[181] counties in the sample for replicating Pendall's sprawl measurements.

In spite of the abovementioned discrepancy between the two samples, this study has followed Pendall's method in order to most accurately replicate his sprawl measurements. The research results replicated by this study, which reflect both his theory and method, can be used as a substitution for his research results. For simplification, the replicated sample and sprawl measurement results created by this study will be referred to as "Pendall's sample" and "Pendall's research results", respectively, in the following sections.

Containing only one-quarter of the 831 metropolitan counties, Pendall's sample covers even fewer metropolitan counties than does Ewing's sample. As a result, of the 492 counties classified as sprawl in this study, only 65 counties (or 13 percent) are contained by Pendall's sample. That is, his sample misses 76 percent of all *sprawl metropolitan* counties and 87 percent of all *sprawl* counties in the US.

Table 5.8 presents the land use types for the counties in Pendall's sample. As shown in Table 5.8, his sample presents a much higher percentage share as urban counties and a much lower share as undeveloped counties than the nation as a whole. Of the 200 counties/county groups in his sample, 31 percent are urban/ urban center counties; 47 percent are suburban/rural center counties; and only 23 percent are rural/undeveloped counties.

---

[181] The 181 counties include 179 counties that gained both population and urban acres and two counties that only gained population between 1982 and 1992. These two counties, Bronx, NY and Queens, NY, had not gained urban acres but gained population during this time period. The sprawl score of their neighbor county—the county of New York, NY (which has the highest sprawl score)—is used as the scores for these two counties.

Table 5.8 County Land use Types for Pendall's Sample and for all US Counties

	# of Counties in Pendall's Sample	# of Counties in the U.S.	% in Pendall's Sample	% in the U.S.
Rural & Undeveloped	46	2583	23	83.6
Suburban & Rural Center	93	395	46.5	12.8
Urban & Urban Center	61	113	30.5	3.7
Total	200	3091	100	100

Note:

All of the 200 counties in Pendall's sample (which include the counties that lost population between 1982 and 1992) are presented in this table.

5.3.2 Background to Comparing Sprawl Classifications

Two points need to be clarified before any comparisons of sprawl results are to be made between Pendall's study and this study. First, one should be cautious when interpreting any comparison of sprawl results, due to the different sprawl definitions employed in these two studies. Pendall defines his sprawl measurement as the change of population divided by the change of urban lands during the period 1982 to 1992, with an assumption that all the increased population would live on the newly developed urban lands. If a county has experienced low-density developments on the newly developed urban lands, then it would be defined as sprawl in his study. This study does not measure densities for *newly developed* urban lands as Pendall did; rather, it measures densities on *developable* lands in 2000 as well as the growth rate and absolute amount of new growth during the period from 2000 to 2020. Therefore, if a county is defined as sprawl in both studies, it experienced low-density growth on newly developed urban lands during the period 1982 to 1992, based on Pendall's sprawl definition; and also will experience significant household or employment growth during the time period 2000 to 2020, based on the definition in this study.

Second, similar to Ewing's study, Pendall did not explicitly classify his 159 counties into binary sprawl/non-sprawl categories. Rather, he calculated sprawl indicator scores for counties and sorted the counties by their sprawl scores in ascending order; the lower the sprawl score value for a given county, the higher the degree of sprawl for that county. In order to compare the research results between Pendall's study and this study, first a method needs to be adopted to classify the counties in his sample into binary sprawl/non-sprawl categories. The method used is similar to the one that has been applied to Ewing's sprawl results; the 30th percentile of Pendall's sprawl scores is used as the threshold cut-off to distinguish between sprawl and non-sprawl counties in his sample.

Why does the 30th percentile figure as an appropriate cut-off? In order to answer this query appropriately, the research results in this study are used as a reference. The cut-off value should maximize the overlap of sprawl classifications between these two studies. First, based on their sprawl scores, all of the 181 counties in Pendall's sample are classified into ten percentile groups, with the 1st group having sprawl scores below the 10th percentile and the 10th group having sprawl scores above the 90th percentile. Then within each percentile group, the numbers of sprawl and non-sprawl counties, based on the research results of this study, are summarized as shown in Table 5.9. For example, for the 3rd percentile group, 10 of the 17 counties are classified as sprawl in this study; while for the 4th group, 8 of the 18 counties are classified as sprawl in this study.

Table 5.9 Number of Counties by Sprawl Classifications (Developed in this Study) for Counties in Pendall's Percentile Groups

Percentile Group[1]	Number of Counties: Sprawl	Non-sprawl	Total	% Designated as Sprawl
1	5	12	17	29.4
2	10	9	19	52.6
3	10	7	17	58.8
4	8	10	18	44.4
5	7	10	17	41.2
6	1	17	18	5.6
7	6	11	17	35.3
8	9	9	18	50.0
9	8	9	17	47.1
10	1	22	23	4.3
Total	65	116	181	

Note:

1. The 10 percentile groups are based on Pendall's sprawl scores.
2. The numbers of sprawl and non-sprawl counties are based on the sprawl classifications developed in this study.

As shown in Table 5.9, more than one-half of the counties in the 2nd and 3rd groups are defined as sprawl in this study; while less than one-half of the counties in the 4th through 10th groups are sprawl in this study. Therefore, the 30th percentile of Pendall's sprawl scores is used as the cut-off value, which maximizes the overlaps of the sprawl classifications between these two studies. Based on this cut-off value, the 53 counties with a sprawl score below this value are defined as sprawl; and the other 128 counties with sprawl scores above it are defined as non-sprawl for Pendall's study. These sprawl/non-sprawl classifications, although developed by this study, are based on the concepts and associated with the sprawl results developed in Pendall's study. Therefore , these sprawl (or non-sprawl) counties will be referred to as "being classified as sprawl (or non-sprawl) *in Pendall's study*" in the following sections.

Differing from all other percentile groups, a significant discrepancy in sprawl classifications between the two studies exists in the first percentile group. This percentile group contains the most *sprawl* counties in Pendall's sample, but 70 percent of its counties are classified as *non-sprawl* in this study. This discrepancy is mainly due to the different sprawl definitions employed in these two studies. These 17 counties had the lowest densities on their newly developed urban lands during the period 1982 to 1992, compared with other counties in the sample; thus they would be classified as sprawl in Pendall's study. However, 12 of these 17 counties are classified as non-sprawl in this study, primarily due to two reasons. First, nine counties will not experience significant growth during the period 2000 to 2020; second, 3 counties that will experience significant growth in the future are developed suburban counties and thus are classified as non-sprawl.

5.3.3 Comparing Sprawl Classifications

After using the abovementioned method to classify Pendall's sample into binary sprawl/non-sprawl categories, the sprawl classifications can now be compared between the two studies.

Table 5.10 presents the sprawl classifications for the same individual county between the two studies. Again, the cells on the upper left to lower right diagonal line illustrate the number of counties that have the same sprawl classifications between the two studies. As shown in Table 5.10, 62 percent (or 113 counties) of the 181 counties have the same sprawl/non-sprawl classifications in both studies: For instance, 25 and 88 counties are classified in both studies as sprawl and non-sprawl, respectively. The other 38 percent (or 68 counties) of the 181 counties have different sprawl classifications between the two studies: 28 counties are classified as *sprawl* in Pendall's study but as *non-sprawl* in this study; conversely, 40 counties are classified as *non-sprawl* in Pendall's study but as *sprawl* in this study.

Table 5.10 Numbers of Counties by sprawl classifications in the Two Studies

	In This Study		
In Pendall's Study	Sprawl	Non-sprawl	Total
Sprawl	25	28	53
Non-sprawl	40	88	128
Total	65	116	181

Note:

The sprawl classifications in Pendall's study, which are based on Pendall's sprawl indicator scores, are developed by the author of this study.

In the following sections, first the overlap of sprawl classifications between the two studies will be addressed. Then the counties with discrepant sprawl classifications between the two studies will be investigated.

First, 88 counties are classified as *non-sprawl* in both studies. These counties have relatively high densities on newly developed urban lands during the time period 1982 to 1992 (in Pendall's study), mainly because most (72 percent) of these counties are located in high- or moderate- density states[182] (47 percent in high-density states and 25 percent in moderate-density states). These counties will also not experience sprawl growth during the time period 2000 to 2020 (in this study), mainly due to two reasons. First, just over one-half of the 88 counties (or 45 counties) will not experience significant growth in the future. They do not count as sprawl if they are not growing very fast. Of these 45 counties, most (38 counties) are developed urban/urban center counties; only one-tenth

[182] Only 27 percent of these 88 counties are from low-density states and 1 percent is from a very-low density state.

(or 5 counties) are low-growth rural/undeveloped counties. Second, the other one-half of these 88 counties (43 counties) that will experience significant growth in the future are *developed* suburban/ rural center counties (36 suburban counties and 7 rural center counties), thus satisfying the criteria for non-sprawl classification in this study (i.e. significant growth that takes place in already reasonably developed suburban or rural center counties does not count as sprawl growth).

Second, 25 counties are classified as sprawl in both studies. These counties are defined as sprawl in Pendall's study because they experienced relatively low-density new developments during the time period 1982 to 1992. Most of these counties are located in low- or moderate-density states[183] and thus have relatively low densities on newly developed urban lands. In this study, they are rural/ undeveloped counties (13 counties) and developing suburban/ rural center counties (12 counties) that will experience significant growth during the time period 2000 to 2020, and thus are classified as sprawl.

Counties with the same sprawl classification in the two studies have now been presented and analyzed. The counties with different sprawl classifications between the two studies will be addressed in the following sections.

On one hand, it is observed that 28 counties are classified as *sprawl* based on Pendall's sprawl scores, yet are classified as *non-sprawl* in this study. Although a large portion of these 28 counties are located in high- and moderate- density states[184], they experienced low-density growth during the period 1982 to 1992; and thus would be classified as sprawl in Pendall's study. The rationale which leads to the different sprawl

---

[183] Of these 25 counties, 12, 9, and 4 counties are located in low-, moderate-, and high- density states, respectively.

[184] Almost one-half (or 12 counties) of these 28 counties are located in high-density states; 7 are in moderate-density states, 8 are in low-density states, and 1 is in a very-low density state.

classifications between these two studies can be found in the difference in sprawl definitions. These 28 counties are classified as non-sprawl in this study, mainly due to two reasons. First, 17 of these 28 counties will not experience significant growth during the time period 2000 to 2020. These are low-growth rural/ undeveloped counties (9 counties) and developed urban/ suburban counties (8 counties), which are mainly located in low- and moderate- density states[185]. Secondly, the other 11 of these 28 counties that will experience significant growth in the future are *developed suburban counties*, and thus are classified as non-sprawl in this study. Eight of these 11 counties are located in high-density states. It is interesting to note that these 11 counties, which are located in high-density states and classified as suburban counties in 2000, experienced relatively low-density development on the newly developed urban lands during the period 1982 to 1992.

In contrast, 40 counties would be classified as *non-sprawling* in Pendall's study but are classified as *sprawling* in this study. These 40 counties have experienced relatively high-density developments on newly developed urban lands during the period 1982 to 1992. Almost three-quarters (or 29 counties) of these 40 counties are from high- or moderate- density states. The existing high-density development in these counties resulted in the high-density development in the new urban areas.

However, these 40 counties are classified as sprawling in this study, again primarily for two reasons. First, nearly one-half (or 19 counties) of these 40 counties are rural counties that will experience significant growth during the period 2000 to 2020 and thus are classified as sprawl in this study. Fourteen of these 19 counties are located in high- or

---

[185] Of these 17 counties, 6 are from low-density states; 6 are from moderate-density states; 4 are from high density states; and 1 is from a very-low density state.

moderate density states and thus had experienced relatively high-density developments in new urban areas, compared with the new developments of counties in the sample that are located in low- or very low-density states—therefore they would be classified as non-sprawl in Pendall's study[186]. However, the new urban areas in these counties may have been developed at relatively *low* densities compared with the new developments in their surrounding counties. Although this study does not measure densities for new developments, the densities of these counties in 2000, which include the land-use results of developments during the period 1982 to 1992, are relatively low compared with their surrounding counties. These 14 counties are classified as rural/undeveloped in this study. In other words, these 14 counties may have been classified as *sprawling* counties if the variable-density approach, instead of the static density approach, had been employed in Pendall's study. Further, these 14 counties will experience significant growth during the period 2000 to 2020 and thus are classified as sprawl in this study.

The other one-half (or 21 counties) of these 40 counties are developing suburban/rural center counties which, by definition, will experience significant growth in the future and thus are classified as sprawl in this study[187]. The different sprawl classifications between these two studies mainly result from the different sprawl definitions employed in the two studies.

Why would the variable-density approach be better than the static-density approach for Pendall's sample? The regional contexts of counties in Pendall's sample vary

---

[186] Ten of these 14 counties, by descending sprawl scores, are San Bernardino, CA, Sonoma, CA, Napa, CA, Coweta, GA, Carroll, MD, Monroe, IL, Charles, MD, Frederick, MD, Calvert, MD, and Dearborn, IN. They rank from 157th (San Bernardino) to 129th (Dearborn, IN) in Pendall's sample.

[187] Of these 21 counties, four are located in high-density states; eleven are in moderate-density states; five are in low-density sates; and one is in very-low density state.

significantly. Of the 200 counties in his sample, nearly 40 percent (or 76 counties) are located in high-density states, 32 percent (or 64 counties) are in moderate-density states; 28 percent (or 54 counties) are in low-density states; and 2 percent are in very-low density states. In each of these cases Pendall uses the same density to classify sprawl development. In order for such a sample to be a more accurate and logical representation, the variable-density approach, instead of the static-density approach, should be applied, because the former takes into account the different contexts of counties. This argument will be addressed in detail in the next section.

5.4 Summary

Through the comparison of research results between the existing three studies (Burchell et al.'s, Ewing et al.'s, and Pendall's studies) and this study, the following conclusions can be derived.

First, it is necessary to calculate more accurate densities in a study that measures sprawl. Between the Report (i.e. Burchell's study) and this study, different density calculations result in somewhat different classifications of land use types and, further, in different sprawl classifications for counties. At least 46 counties' sprawl classifications are affected by these calculations. Among these 46 counties, 38 counties are classified as *sprawl* rural/undeveloped counties in the Report but as *non-sprawl* suburban/rural center counties in this study; 8 counties are classified as *sprawl* suburban/rural center counties in one study but as *non-sprawl* urban/urban center counties in the other study.

Second, the Report has 73 more sprawl counties than does this study. Nearly ninety percent (88 percent or 2728 counties) of the 3091 US counties have the same sprawl/non-sprawl classification in both studies[188]; while the other 12 percent or 363 counties have different sprawl/non-sprawl classifications. In detail, 218 counties are classified as *non-sprawl* in this study but as *sprawl* in the Report; while the other 145 counties are classified as *sprawl* in this study but as *non-sprawl* in the Report. The major reasons for these different sprawl classifications include: 1) the different versions of Woods & Poole datasets employed in these two studies (i.e. different amounts of growth for same individual counties), 2) the different classifications of county types (caused by different density calculations—better information leading to better classifications), and 3) the different sprawl classification criteria (i.e. an improved definition) for suburban/ rural center counties between these two studies.

Third, the difference in the amount of national sprawl household growth between the Report and this study is considerably large. Sprawl household growth would decrease by 36 percent (or 4,041,095 households) by moving from the Report to this study, chiefly attributed to two reasons. First, more counties are defined as sprawl in the Report than in this study. Second, and more importantly, among the counties that are classified as *sprawl* in the Report but as *non-sprawl* in this study, some are developed suburban/rural center counties with large amounts of projected household growth in the future, which are counted as sprawl growth in the Report; while in contrast, all of the counties that are classified as *sprawl* in this study but as *non-sprawl* in the Report will experience relatively small amounts of household growth during the same time period.

Fourth, for studies aimed at measuring sprawl nationwide, a sample solely focused on the most populous metropolitan areas is not large enough. Of the 492 counties

---

[188] Here, “both studies” refers to the Report and this study.

classified as sprawl in this study, 87 percent (or 427 counties) are missed by Pendall's sample; 66 percent (or 324 sprawl counties) are missed by Ewing's sample; and 44 percent (or 216 counties) would even be missed by a sample that contains all metropolitan counties.

Fifth, 21 percent (or 96 counties) of the 448 counties in Ewing's sample, as well as 38 percent (or 68 counties) of the 181 counties in Pendall's sample, would have different sprawl/non-sprawl classifications from this study. Two major reasons cause these different sprawl classifications: 1) the different sprawl definitions between these studies and this study; and 2) the difference between the static-density approach used in these two studies and the variable-density approach employed in this study.

Further, the difference between the static- and variable- density approaches also results in different classifications of land use types for some particular counties between Ewing's study and this study.

Finally, the variable-density approach is demonstrably better than the static-density approach when measuring sprawl across the country because the former takes into account the regional contexts of counties, which is arguably essential in a national study of sprawl measurements. On one hand, counties in *high-density* states would be defined as *non-sprawl* or "urban" based on the static-density approach, because their densities are relatively high compared with the counties located in *low-density* states. However, these counties may actually act as *rural* counties in the region, based on the variable-density approach, due to their relatively low densities compared with their surrounding counties. If these counties will experience significant growth in the future, they will be classified as *sprawl* in this study. On the other hand, counties in *low-density* states will be defined as *sprawl* or "rural" by the static-density approach, because their densities are relatively low compared with the counties in *high-density* states. But these counties may function as *developed suburban* or *urban* counties in the region, based on the variable-density approach, because of their relatively high densities compared with the surrounding counties. Since they are urban/developed suburban counties, they are classified as *non-sprawl* in this study. As supported by the practical examples in this chapter, the variable-density approach better reflects the land-use types and functions of counties in real world applications than does the static-density approach.

Chapter 6

Implications of Research

As Burchell et al. clarified, "sprawl is low-density, leapfrog development that is characterized by unlimited outward extension" (Burchell et al. 2002). Sprawl has been criticized as a kind of growth pattern that comes at greater costs than the alternative, compact growth scenario (Ewing 1997; Burchell et al. 2002). Sprawl, as defined in this study[189], is characterized as significant growth taking place in undeveloped or rural counties, where there is not enough of an existing infrastructure and not sufficient public services to support such rapid and significant growth. New infrastructure and public schools must be built in order to support this rapid or large amount of growth. Due to its costly nature and widespread episodes of occurrence nationwide, sprawl has attracted much attention from researchers and policymakers. As such, it is important to identify where sprawl will happen in the near future. If policymakers have reliable information on which counties will experience sprawl growth in the foreseeable future, then they could plan accordingly, in an effort to control that sprawl by redirecting the projected sprawl growth from one particular area or county to those already developed locations within the EA (Burchell et al. 2002). The existing infrastructure in these more developed locations could partially or even fully contain the redirected growth, thus saving many unnecessary costs of building support structures and new facilities in less developed, sprawl counties.

Because of the importance of identifying the occurrence of sprawl, several studies empirically measuring sprawl have emerged since late the1990s. These works have four

[189] As mentioned in chapter 1, this definition is first developed by Burchell et al. (2002).

notable characteristics. First, the sprawl measurements used in these studies focus a primary eye on the most populous metropolitan areas (with the only exception being that of Burchell et al.'s research in 2002). Second, regional density differences of the studied geographic units are inappropriately neglected by using the "fixed-density approach" to measure sprawl across the country. Third, land-use densities are not accurately measured due to the inclusion of undevelopable lands in the density calculation. Finally, no research has ever been undertaken that compares and contrasts the research results of existing sprawl-measurement studies. Sometimes the sprawl measurement results of these works might be contradictory due to their different measurement methods. A comparison analysis of the research results of these existing studies would be highly useful because it could shed a direct light on the needed improvements of currently in vogue sprawl-measurement methods.

Aimed at resolving the abovementioned problems, this study first measures sprawl for all counties in the United States using a variable-density approach. Undevelopable lands are excluded from the density calculation. The research results of this study are then compared to three prominent previous studies (Pendall 1999; Burchell et al. 2002; Ewing et al. 2003b), highlighting how to achieve improvements in the methods for measuring sprawl. The major research results of this project include: 1) residential densities in 2000 for all the 3091US counties and the 50 US states; 2) existing land use type in 2000 for these counties; 3) sprawl/non-sprawl classification during the 2000 to 2020 period for these counties; and 4) conclusions drawn from comparison analysis between this study and the three previous studies. These research results have been presented in Chapters 4 and 5.

The purpose of this chapter is to discuss the implications of the research results in both the sprawl measurement literature and in public policies. In section 6.1, implications of the research results that have been developed in the process of measuring sprawl for all US counties are described; in section 6.2, a comparative analytic discussion is made to highlight an implicit understanding and any logical relation between the previously referenced three existing research studies and this study; finally, in section 6.3, some improvements that can be made for future research on measuring sprawl are highlighted.

6.1 Implications of the Sprawl Measurement

In this section a discussion is contained on the implications of research results created in the process of measuring sprawl for the counties nationwide. These results include: 1) accurate densities being calculated for all 3091 counties and all US 50 states; 2) one of the six existing land use types being assigned to each county, based on residential density of the county and its host state, as well as the county's real-world function within the EA[190]; and 3) sprawl/non-sprawl classification being defined for the projected growth during the period 2000 to 2020 for all 3091 counties. The implications of these research results will be detailed in the following section. For each research result, a description of its characteristics is given, followed by any implications which that result may have in terms of both the sprawl-measurement literature and public policy.

As the first research result created during the process of measuring sprawl, an accurate residential density in 2000 is calculated for all 3091 counties and all US 50

[190] As explained in Chapter 3, the economic and social functions of a county within an EA, based on the information obtained from websites and the author's knowledge of the county, are employed to adjust the county land use type for particular counties.

states. As density is a basic and important variable in the field of planning and public policy, it is desirable to calculate an accurate density for the studied geographic units. The density measured in this study is more accurate than the current, gross means of measuring density, because the undevelopable lands have been excluded from the denominator of the calculation. With only the developable lands as the denominator, this residential density better reflects the land use reality in a state/county than does the rough contemporary density. Thus, this more accurate representation provides policymakers and researchers with a clearer description of the existing land use types for their studied geographic units.

In addition, this project is the first one that includes a calculation of accurate density for all counties in the United States. Mainly due to the lack of data, two drawbacks currently exist in the previous literature measuring accurate densities at county and sub-county levels: first, there have been only a small number of studied subjects and second, there has been a trend to focus merely on the most populous metropolitan areas. The biggest sample in the existing literature calculating accurate density at county level is Ewing et al.'s 448 counties[191] that comprise the 101 most populous metropolitan areas (Ewing et al. 2003b). A problem resulting directly from these two drawbacks is that policymakers who are interested in non-metropolitan counties have no data upon which they may rely for accurate density information in order to help them understand the land use types of those specific counties. Aimed at solving this problem, this project would provide policy makers and researchers with accurate

[191] Ewing's team (2003b) uses a method and dataset different from this study.

densities for all metropolitan and non-metropolitan counties, thereby eliminating this gap in useful data.

The second research result developed in the process of measuring sprawl is the classification of existing land use types in 2000 for all counties in the United States. One of the six land use types (urban center, urban, suburban, rural center, rural, and undeveloped) has been assigned to each of the 3091 counties nationwide. As defined in Chapter 3, this county type is density-based and “denotes a county as more or less developed relative to the other counties of its region” (Burchell et al. 2002). In this project, it is assumed that the more developed a county is, the more important its economic and social function is in the region. It is also the county most likely to receive growth without the requirement to develop significant new infrastructure. Therefore, the highest-density county in a region is assumed to be the most developed county and the most important economic and social center in the region, and the most ready to receive development. Of these 3091 counties, most (84 percent) are undeveloped and rural counties; nearly 13 percent are suburban and rural center counties; and only 3.7 percent are urban and urban center counties. The latter two growth of counties can receive development without sprawling.

The county land use types in this study, based on the abovementioned accurate county densities, better describe how developed a county is relative to other counties in the region than would be the case if contemporary rough densities were used.

In addition, based on the variable-density approach, the county existing land use configuration can be compared among counties nationwide. As discussed in previous chapters, it makes less sense to compare density values among counties nationwide (i.e., using the “fixed-density approach”), because in different states the same value of density may represent very different land use types. A same density of 150 households per square mile, for example, might represent an urban county in Montana but only a rural county in New Jersey.

To solve this problem, comparable county land use types, instead of densities, are used as the measuring variable of the existing land use types of counties. The “variable-density approach”, first developed by Burchell et al. (2002), is employed to make the county land use types comparable across the country. This approach takes into account regional context of a county (in this research, they are mainly represented by a county’s host state density) when defining county land use type, thus making it possible and meaningful to compare county types among counties across the country. According to this approach, the density thresholds of county land use types vary with state densities. For example, the density threshold range of suburban counties is from 27 to 116 households per square mile for the states with very low densities (i.e., very low-density state group); but between 267 and 658 households per square mile for the states with high densities (i.e., high-density state group). Such county type classification based on the variable-density approach is especially valuable if policymakers want to conduct any cross-country analysis. The practical proof to the advantage of a variable-density approach over a fixed-density approach is shown in chapter 5, and will be explained in a later section in terms of the implications of comparison analysis.

Finally, the accurate county land use types developed in this study would be helpful to the policymakers when they want to control the sprawl growth of a county. In order to control future sprawl for a particular county, a portion of the growth could be redirected to other, more developed counties (in the same EA) that will not experience sprawl during the same time period[192] (Burchell et al. 2002). With the county land use type information for all the counties within the EA, policymakers could easily identify the nearby, more developed counties (i.e., the urban center, urban, or developed suburban counties) of the sprawl county as potential receivers of any redirected growth.

The third research result developed in the process of measuring sprawl is the classification of sprawl/non-sprawl for all counties nationwide. A rural or undeveloped county is defined as sprawl if it will experience significant growth over the period 2000 to 2020. Some of the developing suburban or rural center counties are also defined as sprawl areas if they meet certain criteria. Overall, 492 counties in the country will experience sprawl over the period 2000 to 2020. Of these 492 sprawl counties, most (80 percent, or 396 counties) are rural and undeveloped counties; one fifth (96 counties) are developing suburban and rural center counties. Sprawl growth during this time period is analyzed by state, EA, and county. This study also presents the states and EAs that will experience the largest amounts of projected sprawl growth.

Based on the abovementioned accurate, comparable county land use types, this sprawl classification for counties would provide the credible identification of sprawl counties nationwide. With this classification, policymakers and researchers can more readily and easily identify which counties will experience sprawl in the future, and

---

[192] . There are two ways to control sprawl growth. The first method, which could be named as "inter-county redirection", redirects a portion of the growth from the sprawl county to the urban center, urban, and developed suburban and rural center counties within the same EA. The second method, named as "intra-county redirection", redirects the sprawl growth to the more developed (urbanized) locations within the same county (Burchell et al. 2002).

therefore would be able to adjust or make policies accordingly in a better, more modified effort to control the projected sprawl growth.

Finally, this study also provides policymakers with highly useful information for controlling sprawl growth. Such information includes the sprawl classifications and the projected growth over the period 2000 to 2020 for all counties nationwide. The latter is calculated from the Woods & Poole projection (Woods & Poole 2003). All this information would be useful when policymakers decide to control the sprawl growth in a county by shifting, or redirecting a portion of its projected growth to other, more developed counties which will not experience sprawl during the same given time period. With this information at hand , they can logically decide 1) how much of a county’s sprawl growth should be redirected to other, more developed locations; and 2) which more developed counties within the EA will be non-sprawl during the period 2000 to 2020, and thus could act as the receivers of the redirected growth without experiencing excessive growth themselves.

Implications of the research results created during the process of measuring sprawl for counties have been the focus of the discussion in this section. In the next section, implications of the comparison analysis of relative research results between the existing three studies and this study are highlighted.

6.2 Implications of the Comparison Analysis

Although many studies have measured sprawl since 1990s, no research has ever been done to compare and contrast results of these studies. As presented in Chapter 5, quite often a geographic unit will be defined as sprawl in one research project but as compact in another, due to the different sprawl definitions, datasets, and measurement methods used in different studies. A comparison analysis would indeed shed a much needed light on what may be causing these discrepancies in sprawl results among the various studies. In addition, it would also contribute to the literature regarding the measurement of sprawl by providing practical proofs to key questions about existing sprawl measurement methods. These questions, raised in Chapter 1, will be answered below, with the practical proofs from the comparison analysis developed in Chapter 5.

The first question is raised in Chapter 1. Is it enough to measure sprawl only in the most populous metropolitan areas? Research results here suggest that it may not be enough to do so. In fact, of the 492 counties that will experience sprawl growth during the period 2000 to 2020, only one third (168 sprawl counties) are contained by the 101 largest metropolitan areas[193]. Further, it may well be that it is not enough to measure sprawl only for the geographic units located in metropolitan areas. To wit, only 56 percent of this study's 492 sprawl counties are located within metropolitan areas; that is, nearly half of the sprawl counties are located outside metropolitan areas. Therefore, a study measuring sprawl only for metropolitan counties will miss nearly half of the sprawl counties nationwide.

This leads us to a second key question. Is a variable-density approach more preferable to a fixed-density approach? These research results suggest that to be the case, that the former is better than the latter, especially for research that measures sprawl across the country. The comparison analyses of the research results between Ewing et al.'s/ Pendall's study and this study provide practical supportive proofs. A major reason

[193] The 101 largest metropolitan areas are from Ewing et al.'s sample (2003b), which is the largest sample among the studies focusing on the most populous metropolitan areas.

for the different sprawl classifications between Ewing et al's / Pendall's research and research provided herein is that the former use the fixed-density approach while the research here uses the variable-density approach. For research whose goal is to accurately provide the framework for measuring sprawl, if the sample includes counties from different density contexts, that is, from states with very different residential densities, it is inappropriate to simply use static county densities as the indicator of sprawl or to compare density values among these counties. In other words, a static density value may not reflect the real function and land use type of a specific county.

On one hand, a low-density, "sprawl" county in Ewing et al.'s study may be defined as non-sprawl in this study, because this low-density county may actually act as urban county in its region, due to its higher density relative to other counties in the region. On the other hand, a high-density, "compact" county in Ewing et al.'s study may be defined as sprawl in this study, because that high-density county may actually act as rural or undeveloped county in its region, due to its actual lower density relative to other counties in its region. If this rural or undeveloped county will experience significant growth during the period 2000 to 2020, then it will be classified as sprawl herein. For example, San Bernardino County, CA, a high density and thus a "compact" county in Ewing et al.'s study (2003b), is classified as a sprawl county in both Burchell et al.'s study (2002) and this study. It is calculated as a compact county in Ewing et al.'s research (2003b), because as a county located in a high-density state (CA), San Bernardino County has a higher density than some counties in the sample[194] that are located in low-density states. However, in Burchell et al.'s research as in this research it is defined as a

[194] Ewing et al.'s sample (2003b) includes 408 counties from the 101 most populous metropolitan areas in 2000.

sprawl county, mainly because it functions as a rural county in its region (i.e., CA)[195] and will experience significant growth over the period 2000 to 2020. It is classified as a rural county in 2000 in this investigation, because its density is much lower than the densities of surrounding counties in the region (i.e., in southern CA). In fact, based on the variable-density approach, the land use type classification of rural county better represents the regional function of San Bernardino in the real world than does the probable classification of urban county, which would be based on the fixed-density approach.

Third, this project suggests that a more accurate density calculation is necessary when measuring sprawl. The comparison analysis of research results between Burchell et al.'s (2002) study and this study suggests that the accurate density calculation, with the undevelopable lands being excluded, would better reflect a county's density in the real world. In addition, the county's existing land use type, if based on the more accurate density value, would better reflect the county's real function in the region.

The above three conclusions, drawn from the comparison analysis, shed some intriguing light on the methods of measuring sprawl for future research. However, due to limited resources and data availability, this project inevitably has drawbacks that could be improved with further, even more detailed future research, which will be discussed in the next section.

6.3 Limitations of this Study and Future Research

[195] If the fixed-density approach, instead of the variable-density approach, is employed in my study, then San Bernardino, CA might be classified as an urban county in 2000 and thus be defined as non-sprawl over the period 2000 to 2020.

Due to several data and resource limitations, this study certainly has its own shortcomings that can be improved with additional research.

Improved Density Calculation

The density calculation could be improved if more accurate GIS data would be available for use in such a study. The GIS data for the boundaries of national parks, state parks and forests, and Indian reservations used in this project are on a small scale, and not accurate enough for a spatial analysis at the county level. However, these were the only publicly available GIS data on the boundaries of these geographic units when this project was initiated and conducted. If more accurate GIS boundary data for these legally undevelopable lands would become available in the future, then even more accurate residential densities for counties and states could be calculated.

Causes of Sprawl in Rural/Undeveloped Counties

The research contained here never intended to nor does it address the actual causes of sprawl found in rural/undeveloped counties. Sprawl's causes are of interest to planners and policy makers because they are useful in defining strategies that contain sprawl development. Currently, in the United States, the primary cause of sprawl development, in locations where it should not take place, is the availability of less expensive land in these locations. Land price is one of the most important underlying factors affecting where development takes place. Sprawl occurs in areas where lower land prices exist because potential buyers can seek out these locations and build larger houses there. A second factor facilitating to sprawl development is that due to reduced social and public services, movers usually experience lower taxes in peripheral locations than they do in locations closer-in. A third important factor affecting sprawl is the subsidizing of road infrastructure by federal, state, and local governments (Ewing 1997; Burchell et al. 1998; Pendall 1999), which assists costly, unlimited outward development in areas lacking most other infrastructure by facilitating automobiles to get there. Additional factors include "white flight" (i.e. one ethnic group moving away from another to suburbs and exurbs) and property value increases which usually take place much faster in the more remote areas. (Pendall 1999).

Non-sprawling Rural/Undeveloped Counties with Significant Growth

As argued by Reid Ewing in his recent presentation at the Rocky Mountain Land Use Institute, if *new growth happens* in a rural county in close proximity to an existing built-up central area (i.e. new development can benefit from the existing infrastructure of adjacent areas, saving the cost of building new infrastructure), then this type of development should *not* be defined as sprawl. This study agrees with Ewing's position, which is similar to Burchell's intra-county redirection of future growth (Burchell et al. 2002). By definition, new growth immediately adjacent to built-up areas, is not sprawl.

Policies Constraining Sprawl

The study included here further does not address how various planning strategies might affect sprawl development, an area of great interest to policy makers. Numerous policies exist to contain sprawl, encourage high density development, provide efficient public services and infrastructure, and preserve land resources/critical habitats. One of these policies, an Urban Growth Boundary (UGB) contains sprawl by establishing a growth boundary around existing communities. Only land inside the boundary can be converted to urban use before a specified date; land outside the boundary is preserved for nonurban use until that land is required and no land remains inside the boundary (Knaap 1985, p.26). Other planning strategies similar to the Urban Growth Boundary are “Greenlines” and urban service areas. Empirical studies (Knaap 1985; Shen 1996) show that this constraint is only effective when accompanied by strict zoning ordinances designed to prevent low-density development outside the boundary, and incentive-laden policies to encourage infill inside the boundary. If zoning ordinances and incentive policies are not present, the UGB has no significant impact on the control of sprawl.

Other means of constraining sprawl rely on a series of financial strategies to position development. These strategies include adequate public facilities ordinances (APFOs), development impact fees, and transfer of development rights (TDR). By making growth pay its own way, if development is intending to go into areas absent infrastructure, there is some likelihood that it will refrain from going there, thereby limiting sprawl and preserving land and other resources (Pendall 1999).

Demographic Projection -- Woods & Poole

The Woods &Poole projections of demographic trends employed in this study would reflect greater accuracy if they were somewhat more recent. 2003 versus 2012 projections were used because this period was closest to the date of available land cover data and because this time period was also close to the time period of the Burchell (2002), Ewing (2000), and Pendall (1999) studies.

The Woods & Poole projection data, as a whole, is a relatively solid source of demographic and economic projections at the county level. As with all projection data, however, it does have its limitations. Woods & Poole use an "export-base" approach[196] to project employment, earnings, and income in each of the counties; it then projects population migration *as a function of employment growth*[197]. Individuals and families are assumed to migrate primarily in response to employment opportunities. Political and social factors that affect migration are generally not taken into account. For example, changing political views, changes in economic and tax policies (e.g., more/less favorable tax system, immigration issues, climate, etc.) could induce migration. Not including these factors potentially increases errors in the Woods & Poole demographic projections, and in turn, affects the accuracy of the overall research results offered here. This is not something that can be changed by users of the Woods & Poole data, their data is what it is. There are certain things which would increase study accuracy as will be noted below.

Updated Data

The sprawl classification for counties would be more reliable if the updated Woods & Poole growth projections over the period 2000 to 2020 (i.e. which has the benefit of the 2010 Census) were used in the project. This study used the 2003 version of Woods & Poole datasets, created before Hurricane Katrina devastated Louisiana and the Gulf Coast in 2005, and before the Great Recession from December 2007 to June 2009 had left its impact on the nation as a whole. Regular recession factors have been taken into account in the Woods & Poole data projection, but the scale and degree of the 2008-2009

[196] This projection approach is based on differentiating "basic" industrial sectors from "non-basic" sectors. Certain industrial sectors are considered "basic" (such as mining, manufacturing, and the federal government) which produce output that is "exported" out of the region for national or international consumption. In contrast, "non-basic" sectors (such as retail trade, real estate, and construction) produce output that is consumed locally.

[197] Woods & Poole does integrate some demographic factors in their projection model. For the population aged 65 and over, and for the college or military-aged population, migration patterns over the forecast period are based on historical net migration and not employment opportunities.

Recession is obviously not a part of the 2003 version of Woods & Poole datasets. Therefore, a sprawl classification based on updated Woods & Poole growth projection data would more accurately capture these unexpected economic and social changes and obviously be more reliable. This does not take into account comparison time periods as will be noted. Sprawl classifications based on updated Woods & Poole projections were recently calculated by the author and the results are presented below.

As stated previously, a major goal of this study was to investigate the impact of different sprawl measurement methods on results through comparative analysis. For the three previous studies involved in the comparative analysis, the studied time periods were before 2000 (Pendall's ten year study from 1982-1992); in 2000 (Ewing's static study of 2000); or using 2000 as the base year (Burchell's twenty-five year study from 2000-2025). Thus, this study uses 2000 as the base year and selects the 2000-2020 as projected time period in order to be consistent with prior research and to build upon previous sprawl measurement exercises. Selection of 2000 rather than 2010 as the base year was predicated on the fact that during the period of this study, i.e. 2007-2010, updated data for the 2010 Census was not yet available.

However, it would be interesting to apply the methods employed in this study using the updated data of the 2010 Census to measure sprawl for a projected time period, say, 2010 to 2020. If such a study were to be conducted, household densities in 2010 would be larger than those found in this study, because the numerator (i.e. the number of households) would have increased, by an average of 10 percent for these counties from 2000 to 2010, while the denominator (i.e. developable lands) probably would have increased slightly[198]. Without actually doing the analysis, it is difficult to analyze the change in county type classifications between the years 2000 and 2010, or the sprawl/non-sprawl classifications for counties during the period 2010 to 2020 (using the new projection). Any change analysis would be complicated, because a county's type and sprawl classification can only be defined relative to changes in other counties.

Since the Woods & Poole 2012 data is now available, this study actually looks at sprawl/non-sprawl classifications over the period 2000 to 2020[199] for all US counties using this updated version. The 2003 and 2012 versions of Woods & Poole data will be referred to as the *"old"* and *"new"* projections, respectively. The analysis ultimately shows that sprawl classifications are quite similar between the two projections.

First, based on the new projection, 443 counties will experience sprawl development during the 2000 to 2020 period, compared with 492 counties said to experience sprawl under the old projection. The 49 fewer sprawling counties cited under the new projection are mainly the result of lack of growth during the 2008-2009 Recession[200]. Of these 443 *sprawling* counties, 87 percent (or 384 counties) are also defined as *sprawl* in the old projection; only 13 percent (or 59 counties) are classified as non-sprawl in the old projection.

---

[198] The denominator of density calculation (i.e. developable lands) is calculated as total area minus undevelopable lands (such as water bodies, wetlands, desert, steep slope mountains, national, state, and regional parks and forests, etc.). These undevelopable lands may decrease slightly over time (for example, some wetlands or desert may be developed), resulting in slightly increased developable lands. Developable lands include both *developed* lands and *undeveloped* (but developable) lands; thus, even though large amounts of *undeveloped* (but developable) lands have been converted to *developed* lands over the period 2000 to 2010, the amount of total *developable* land does not change due to this conversion.

[199] The density calculation and county land use type classification of this new analysis are the same as those in this study using the Woods & Poole 2003 data, because the base year used is still the year 2000.

[200] The projected household growth in the US increases by 6 percent from 25,311,676 to 26,817,449 households when moving from the old to the new projection. However, employment growth decreases a substantial 35 percent, from 46,493,819 jobs in the old projection to only 30,227,152 jobs in the new projection.

Second, the county type distribution of sprawling counties is almost identical between the two projections. In the new projection, out of 443 sprawling counties, 59 percent are rural; 19 percent are undeveloped; 11 percent are developing rural center; and the remaining 11 percent are developing suburban counties[201].

Third, at the regional level, the four regions rank the same between the two projection systems for sprawl household growth, as shown in Table 6.1. The South Region increases by 1,100,000 in household growth in the new projection.

Table 6.1 Sprawl Household Growth by Region in Two Projection Systems, 2000-2020 (in Thousands of Households)

Region	Old Projection (A)	New Projection (B)	% Change (C)
South	3,282	4,363	32.9
West	1,958	1,968	0.5
Midwest	1,171	1,115	-4.8
Northeast	628	567	-9.7
Total	7,039	8,013	13.8

Note:

1. The “old” and “new” projections refer to the 2003 and 2012 Woods & Poole projections, respectively.
2. Column C is calculated as (column B minus column A)*100/column A.

Fourth, the top ten states with the largest sprawl household growth are ordered similarly in the two projection exercises, although the rank order may change slightly for individual states, as shown in Table 6.2.

---

[201] With the old projection, of the 492 sprawling counties, 57 percent were rural; 23 percent were undeveloped; 10 percent were rural center; and the remaining 10 percent were developing suburban counties.

Table 6.2 Sprawl Growth Compared with Overall Growth in Top Ten Sprawl States, 2000-2020 (in Thousands of Households)

	Sprawl Growth Rank		Household Growth in Sprawl Counties			Household Growth in All Counties		
State	Old Proj.	New Proj.	Old Proj.	New Proj.	% Change	Old Proj.	New Proj.	% Change
California	1	1	951	770	-19.0	3267	2808	-14.0
Florida	3	2	528	719	36.0	2569	2562	-0.3
Texas	2	3	538	714	32.8	2730	3580	31.2
Georgia	4	4	333	479	44.1	951	1227	29.0
North Carolina	6	5	270	398	47.7	1013	1288	27.2
Tennessee	5	6	285	355	24.7	630	671	6.5
South Carolina	8	7	230	340	47.8	448	568	26.7
Washington	9	8	230	278	21.2	735	840	14.3
Pennsylvania	7	9	258	262	1.6	400	534	33.5
Ohio	10	10	227	251	10.6	499	466	-6.6

Note:

"Proj." in the column headings means "Projection".

Finally, sprawl classifications are similar between the two projection exercises for six states substantially affected by the 2008-2009 Recession. These six states (Arizona, California, Florida, Illinois, Michigan, and Ohio) contain 412 counties in total. Based on the new projection, 72 of the 412 counties experience sprawl development over the period 2000 to 2020, which is 19 sprawl counties less than the old projection (91 sprawl counties). All states in this category experience sprawl county reductions using with this new data except Arizona and Florida. Further, 90 percent (or 65) of these 72 counties defined as sprawl in the new projection were defined similarly in the old projection. County type distribution of sprawling counties is also the same between the two projections. In the old projection, 14 percent (or 13) of 91 sprawl counties were rural center and developing suburban counties; using the new projection, 17 percent (or 12) of the 72 sprawl counties are rural center and developing suburban counties.

Table 6.3 compares the number of sprawl counties, the household growth in those sprawl counties, and the overall household growth between the two projection exercises for these six states. California, Michigan, and Ohio lose five to six sprawl counties each compared with the old projection, due to their slow growth rate during the 2008-2009 Recession. Sprawl household growth in California also decreases significantly (by 14 percent) when switching from the old projection to the new projection. In contrast, sprawl growth in Arizona, Florida, Illinois, and Ohio increases, when transitioning from the old to the new projection. Different periods of data analysis do affect results but not significantly.

Table 6.3 Sprawl Growth Compared with Overall Growth in Six Recession States, 2000-2020 (in Thousands of Households)

State	Number of Sprawl Counties		Household Growth in Sprawl Counties			Household Growth in All Counties		
	Old Proj.	New Proj.	Old Proj.	New Proj.	% Change	Old Proj.	New Proj.	% Change
Arizona	6	7	149	198	33.5	1035	1030	-0.5
California	23	18	892	770	-13.6	3267	2808	-14.0
Florida	20	21	539	719	33.4	2569	2562	-0.3
Illinois	7	6	51	67	31.5	640	593	-7.3
Michigan	17	8	140	137	-1.8	558	294	-47.3
Ohio	18	12	194	251	29.1	499	466	-6.6

Note:

“Proj.” in the column headings represents “Projection”.

Improved Comparison Analysis

The comparison analysis in this project could be improved in future research. In this project, the time periods for which Pendall has measured sprawl and for which sprawl is measured here are different: Pendall measures sprawl for the time period 1982 to 1992, while here it is measured for the period 2000 to 2020. Obviously, the comparison analysis would be greatly improved and shed a more pinpointed spotlight on sprawl measurement methods if new developments for which sprawl are measured would be during the same time period for both studies. Further, this discrepancy in the measured time periods is mainly due to the limited data availability. The NRI land cover data used in Pendall's study has been released at 5-year intervals from 1997 through 2007[202], while the GIS boundary data used in this project is only available for the year 2000 and not available for the years around 1982. However, it is possible to improve this comparison analysis. When the NRI 2012 data release is made available, it would be interesting to use both Pendall's method and the method in this study to measure sprawl for new developments over the same time period from 2002 to 2012. Then the comparison analysis of the resulting two sprawl classifications would be more informative for improving sprawl measurement methods than the analysis conducted in this research.

Redirection of Future Growth

Due to both time and resource limitations, this project only measures sprawl for counties under the uncontrolled-growth, or sprawl, scenario, but did not investigate how

---

[202] "Periodic NRIs were conducted in 1977, 1982, 1987, 1992, and 1997. Since 2000, NRI data have been gathered annually; major releases of these data, however, will continue to be reported at 5-year intervals." (http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS//stelprdb1041379.pdf)
(U.S. Department of Agriculture. 2009)

to redirect the projected sprawl growth to more developed locations in the EA under the alternative, or controlled-growth, scenario. Questions currently remain unanswered. For instance, how much household /employment growth of a sprawl county should be redirected to other, more developed counties or to developed locations within a county? To which counties and locations in the EA should these sprawl growth be redirected? These questions are addressed in Burchell et al.'s project (Burchell et al. 2002) but are unfortunately not covered in this project. A project that answers these questions of redirection would be very useful to policy makers while they are creating or refining policies in their efforts to counteract future sprawl growth.

Appendices

Appendix 1 (for Chapter 3)

1. Special Problems in Data Processing

1.1 State Forests/Parks

After investigation of data[203], six states (Minnesota, New York, Michigan, Wisconsin, South Carolina, and Tennessee) are identified as in a need of an adjustment of developable lands for their state forests/parks. In order to calculate the area (in acre) of developable lands within state forests, new data—the acreage of the state-owned lands within state forests (obtained from the websites of state forests)—is employed together with the ESRI state forest boundary data. The method used to fulfill this task is explained below.

First, the area (in acre) of "GIS non-federal state forest lands for the whole state" is calculated. It is the area of non-federal lands within the GIS state forest boundary[204] for the whole state, calculated from the ESRI state forest/park boundary data (therefore, it is named with "GIS").

---

[203] The areas (in acre) of the state-owned lands for large state forests/parks nationwide are obtained from the official website of these forests/parks. Then for every large forest/park, this area is compared to the total area within the forest/park boundary that is calculated from the ESRI GIS boundary data. For 44 states, the values of these two areas are close to each other, which means that most of the areas within the large state forests/parks are state-owned for these states. However, this is not true for six states, which means that there are significant areas of privately-owned lands in their state forests/parks.

[204] Sometimes the federal lands would fall into the state forest/park boundaries.

Second, the area (in acre) of "GIS Potentially Developable lands" within a state forest is calculated for both the states and the counties that contain the state forests (again, it is named with "GIS" because it was calculated from the ESRI GIS state forest boundary data and federal and Indian land boundary data). The "GIS Potentially Developable lands" is defined as the lands within a state forest boundary that do not belong to any of the following kinds of undevelopable areas: federal lands, state or local park lands, or naturally undevelopable lands. They are the lands where private lands could potentially be located (thus it is named as "potentially").

Finally, the total area of privately-owned lands within state forest boundaries for a county is calculated by using the following equation:

PL2 = D2*(A1-B1)/D1

Where

PL2 = the total area of private lands within state forests for a county;

D2 = the total "GIS Potentially Developable lands" within state forests for a county;

D1 = the total "GIS Potentially Developable lands" within state forests for a state;

A1 = acreage of "GIS non-federal state forest lands for the whole state"; and

B1 = acreage of the state-owned forest lands for the whole state

As described in the foregoing section, the B1 "acreage of the state-owned forest lands for the whole state" is obtained from the official websites of state forests. (A1-B1) represents the total acreage of the privately-owned lands within GIS state forest boundaries for a state[205]. (A1-B1)/D1 describes the private land proportion of the total

---

[205] Other public lands within state forest boundaries, such as those owned by the county government, if

"GIS Potentially Developable lands" within GIS state forest boundaries in a state. To simplify the calculation, this proportion is assumed to be uniform in a state. This rate may vary among the abovementioned six states; but within each state, it is the same for all the state forests within that state.

For example, if (A1-B1)/D1 is equal to 0.6 for State Michigan, then 60 percent of all the "GIS Potentially Developable Lands" within state forest boundaries for Michigan are privately owned. If D2 (total "GIS Potentially Developable lands" within state forests in a county) is 1000 acres in a particular county, then the private lands within state forest boundaries in this county would be 600 (1000*0.6=600) acres. Therefore, in this hypothetical example, originally these 1000 acres would be classified as undevelopable lands and thus excluded from density calculation. However, now 600 acres out of these 1000 acres will be identified as developable lands and contribute to the denominator of density calculation.

1.2 Indian Reservations

The "adjusted developable lands within Indian reservations" are calculated through the following steps. First, the acreage of "GIS Potentially Developable Lands within Indian Reservations" is calculated for each census tract that contains Indian reservations. These are the lands within Indian reservations that do not belong to any of the following kinds of undevelopable lands: naturally undevelopable lands, other federal lands[206], or state and local forests and parks.

---

exist, are of small area, and thus neglected in this equation.

[206] "Other federal lands" refer to the federal lands other than Indian reservation lands.

Second, the “GIS Potentially Developable lands within Indian Reservations” are divided into two parts, corresponding to the two components of the households (American Indian and non- American Indian households) living within Indian reservations. One part of the “GIS Potentially Developable Lands” is ascribed to the American Indian households (but this part is not adjusted to developable lands); the other part is assigned to the non- American Indian households—it is this part that is adjusted to developable lands, named as “Adjusted Developable Lands within Indian Reservations”. Its percentage share of the total “GIS Potentially Developable Lands within Indian Reservation” is equal to the non-American Indian households’ percentage share of the total households of that census tract. For example, in a census tract that contains Indian reservations, if 40 percent of its total households are non- American Indian households, then 40 percent of this census tract’s “GIS Potentially Developable Indian Reservation Lands” are “adjusted developable lands”, which will contribute to density calculation of the county.

2. An Example of County Development Type Adjustments

Three general rules are employed to adjust the original county types to final county types. If a county meets any of the following two criteria and its original county type is not the highest level of development type within the EA, then its original county type will be adjusted to the highest level of the county types in the EA[207]: 1) its density is

[207] The exception is for the counties in the 33 “Rural EAs”. For these EAs, the highest level of original county type is rural or undeveloped county. Therefore, for these EAs, the rural center county, instead of the highest level of original county types, is assigned to the counties that meet any of the first two criteria.

above the “natural county density gap”[208] within the EA; and 2) the county is the “economic center county” of the EA. 3) In addition to these two rules, the third criterion is that the work team’s practical knowledge of the counties validates the adjustment. A county may be adjusted to a higher or lower level of county development type (relative to its original type) by using the third rule.

Here is an example to explain applications of these rules. EA “State College, PA” (EA id: 009) is a “Rural EA”—all its counties are originally classified as rural or undeveloped counties, based on their densities. In order to apply the first criterion of adjustment, all counties in this EA are sorted by residential density in descending order. Then an obvious natural density gap is observed between the 2nd (91 households/sq. mi.) and the 3rd (59 households/sq. mi.) highest county density (the top four county densities are 100, 91, 59, and 58 households per square mile[209]). The counties with a density above the density gap, in this case, the 1st and 2nd highest-density county (Blair and Cambria, respectively), are adjusted from the original “rural” county to the “rural center” county. Further, the second criterion is applied to the counties. The 4th highest-density county, Centre, PA, has a residential density lower than the natural density gap cut-off, but still is adjusted from rural to “rural center” county, because it is the economic center county, containing the center city (State College, PA) of the EA’s economic node. However, the 3rd highest density county, Mifflin, PA, is not adjusted to rural center county, because it does not meet any of the abovementioned three criteria.

---

[208] All the counties within an EA are sorted in descending order by density. Then the first natural density gap/break is called the “natural county density gap”. The counties with a density higher than this density gap are the top-density counties in the EA, and usually, they are the most developed counties in the EA.

[209] The top four highest-density counties, from the 1st to the 4th highest-density order, are Blair, Cambria, Mifflin, and Centre, PA. They are originally classified as “Rural” county, based on their residential densities.

3. Sprawl Adjustment

For some particular suburban and rural center counties that have been classified as sprawl by the abovementioned criteria, the sprawl classification are further loosened by applying the following rules of sprawl adjustment. Such a county will be adjusted from sprawl to non-sprawl if *any* of the following criteria[210] is met:

(1) It is the EA's highest-density counties (with a density higher than the abovementioned "natural county density gap" in the EA); or

(2) it is the county that contains the largest economic node in the EA[211]; or

(3) according to the research team's practical knowledge of the counties in the EA, the county should be categorized as non-sprawl.

Here is an examples that explains these rules. For EA "Idaho Falls, ID-WY" (EA id: 148), the first highest density county is Teton, WY, classified as a suburban county. Originally it was classified as "sprawl" because it will experience significant growth over the 2000 to 2020 time period and meanwhile its urban population percentage (55%) is lower than 85% (the threshold of urban population percentage for a suburban county to be classified as non-sprawl). However, because its density is higher than the density gap of its EA[212], it is adjusted to "non-sprawl". Further, the third highest-density county in the

---

[210] Actually, as mentioned in the foregoing section of classification of county development types, the counties that meet either of the first two criteria have been classified as/adjusted to the highest level of the original county types in the EA.

[211] Here, the "largest economic node" actually refers to the largest center city among economic nodes in the EA. For all the EAs in the sample, the suburban and rural center counties that contain the second largest node are originally classified as non-sprawl by the aforementioned two criteria and thus need not to be adjusted.

[212] The top three highest county densities are 68, 37, and 31 households per square mile; thus the natural density gap is between the first (68 households per square mile) and the second (37 households per square mile) highest county density. The top three highest density counties are Teton, WY, Bannock, ID, and Bonneville, ID.

EA, Bonneville, ID (classified as a suburban county), is originally classified as “sprawl”, due to its projected significant growth in the future and its urban population percentage of 81.2%. However, it contains the center city (the Idaho Falls City) of the EA’s economic node; thus is adjusted from the original “sprawl” to “non-sprawl”, no matter its density is below the EA’s “natural county density gap”.

Appendix 2 (for Chapter 4)

Table 4.4 Number of Counties by Development Patterns for State Density Classifications, 2000

Land-Use Development Pattern	Very Low-Density States	Low-Density States	Moderate-Density States	High-Density States
Undeveloped	336	921	449	65
Rural	90	240	386	96
Suburban & Rural Center	44	145	137	69
Urban & Urban Center	13	24	43	33
Total	**483**	**1330**	**1015**	**263**

Table 4.5 Number of Sprawl Counties by County Development Type

	Undeveloped	Rural	Rural Center	Suburban	Urban	Urban Center	Total
Sprawl	**112**	**284**	**48**	**48**	**0**	**0**	**492**
Non-Sprawl	1659	528	125	174	76	37	2599
Total	1771	812	173	222	76	37	3091

Note:

The numbers of sprawl counties are shown in bold.

Table 4.12 Sprawl Growth Compared with Overall Growth in EAs, 2000-2020

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
LOS ANGELES-RIVERSIDE-ORANGE COUNTY, CA-AZ	1	1	404005	1436989	5.7	5.7	28.1
SAN FRANCISCO-OAKLAND-SAN JOSE, CA	2	5	327204	907694	4.6	3.6	36.0
ATLANTA, GA-AL-NC	3	7	252694	764981	3.6	3.0	33.0
WASHINGTON-BALTIMORE, DC-MD-VA-WV-PA	4	4	236559	937491	3.4	3.7	25.2

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
ORLANDO, FL	5	12	232560	619015	3.3	2.4	37.6
NEW YORK-Northern NJ-LONG ISLAND, NY-NJ-CT-PA-MA	6	2	232398	1204837	3.3	4.8	19.3
DALLAS-FORT WORTH,TX-AR-OK	7	3	224305	1052869	3.2	4.2	21.3
PORTLAND-SALEM, OR-WA	8	17	219368	418783	3.1	1.7	52.4
NASHVILLE,TN-KY	9	23	212588	326668	3.0	1.3	65.1
MINNEAPOLIS-ST. PAUL, MN-WI-IA	10	14	206589	482443	2.9	1.9	42.8
SACRAMENTO-YOLO, CA	11	18	184328	381686	2.6	1.5	48.3
HOUSTON-GALVESTON-BRAZORIA, TX	12	9	168511	719581	2.4	2.8	23.4
PHILADELPHIA-WILMINGTON-ATL. CITY, PA-NJ-DE-MD	13	21	135209	349099	1.9	1.4	38.7
JACKSONVILLE, FL-GA	14	30	132608	257093	1.9	1.0	51.6
BOSTON-WORCESTER-LAWRENCE-LOWELL-BROCKTO, MA-NH-RI	15	15	127030	478513	1.8	1.9	26.5
SEATTLE-TACOMA-BREMERTON, WA	16	13	117505	518454	1.7	2.0	22.7
CHARLOTTE-GASTONIA-ROCK HILL, NC-SC	17	24	115245	300699	1.6	1.2	38.3
DETROIT-ANN ARBOR-FLINT, MI	18	27	112403	290172	1.6	1.1	38.7
DENVER-BOULDER-GREELEY, CO-KS-NE	19	11	95306	621111	1.4	2.5	15.3
INDIANAPOLIS, IN-IL	20	29	94937	258180	1.3	1.0	36.8
LITTLE ROCK-NORTH LITTLE ROCK,AR	21	41	92219	148657	1.3	0.6	62.0
CINCINNATI-HAMILTON, OH-KY-IN	22	36	88484	175740	1.3	0.7	50.3
BIRMINGHAM,AL	23	54	87570	120265	1.2	0.5	72.8
COLUMBUS, OH	24	31	84075	238227	1.2	0.9	35.3
WILMINGTON, NC-SC	25	52	83270	123194	1.2	0.5	67.6
LAS VEGAS, NV-AZ-UT	26	16	77349	459379	1.1	1.8	16.8
ST. LOUIS, MO-IL	27	34	74376	181956	1.1	0.7	40.9

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
SAN ANTONIO,TX	28	25	71961	297843	1.0	1.2	24.2
GRAND RAPIDS-MUSKEGON-HOLLAND, MI	29	35	70695	175984	1.0	0.7	40.2
SPRINGFIELD, MO	30	50	65705	127547	0.9	0.5	51.5
HARRISBURG-LEBANON-CARLISLE, PA	31	65	64577	93403	0.9	0.4	69.1
CLEVELAND-AKRON, OH-PA	32	59	64200	104233	0.9	0.4	61.6
CHICAGO-GARY-KENOSHA, IL-IN-WI	33	10	62974	637171	0.9	2.5	9.9
GREENSBORO-WINSTON-SALEM-HIGH POINT, NC-VA	34	38	58539	171647	0.8	0.7	34.1
OKLAHOMA CITY, OK	35	56	56174	114696	0.8	0.5	49.0
MEMPHIS,TN-AR-MS-KY	36	40	55826	155104	0.8	0.6	36.0
KNOXVILLE,TN	37	48	54961	131390	0.8	0.5	41.8
NEW ORLEANS, LA-MS	38	69	54362	84913	0.8	0.3	64.0
FRESNO, CA	39	44	50484	144388	0.7	0.6	35.0
SARASOTA-BRADENTON, FL	40	33	49908	196308	0.7	0.8	25.4
JACKSON, MS-AL-LA	41	72	49692	78594	0.7	0.3	63.2
PHOENIX-MESA,AZ-NM	42	8	49002	720424	0.7	2.8	6.8
AUSTIN-SAN MARCOS,TX	43	20	48723	349783	0.7	1.4	13.9
TULSA, OK-KS	44	51	47356	123461	0.7	0.5	38.4
SAVANNAH, GA-SC	45	77	46355	73675	0.7	0.3	62.9
EUGENE-SPRINGFIELD, OR-CA	46	57	44009	110089	0.6	0.4	40.0
COLUMBIA, SC	47	61	43832	96007	0.6	0.4	45.7
RICHMOND-PETERSBURG,VA	48	43	43234	146031	0.6	0.6	29.6
RALEIGH-DURHAM-CHAPEL HILL, NC	49	28	42763	277453	0.6	1.1	15.4
LEXINGTON, KY-TN-VA-WV	50	45	41635	134446	0.6	0.5	31.0
ALBUQUERQUE, NM-AZ	51	47	39496	133410	0.6	0.5	29.6
MOBILE,AL	52	82	39384	61475	0.6	0.2	64.1
LOUISVILLE, KY-IN	53	55	38336	114768	0.5	0.5	33.4

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
HONOLULU, HI	54	76	37353	74310	0.5	0.3	50.3
AUGUSTA-AIKEN, GA-SC	55	90	34471	47626	0.5	0.2	72.4
RENO, NV-CA	56	58	34228	107893	0.5	0.4	31.7
RICHLAND-KENNEWICK-PASCO, WA	57	79	34014	70799	0.5	0.3	48.0
GREENVILLE-SPARTANBURG-ANDERSON, SC-NC	58	49	33914	129261	0.5	0.5	26.2
SPOKANE,WA-ID	59	66	32982	92284	0.5	0.4	35.7
TAMPA-ST. PETERSBURG-CLEARWATER, FL	60	22	32979	343650	0.5	1.4	9.6
BATON ROUGE, LA-MS	61	73	32750	78344	0.5	0.3	41.8
PENSACOLA, FL	62	67	32674	91600	0.5	0.4	35.7
MIAMI-FORT LAUDERDALE, FL	63	6	31563	834525	0.4	3.3	3.8
PITTSBURGH, PA-WV	64	117	30336	32978	0.4	0.1	92.0
SALISBURY, MD-DE-VA	65	101	30038	39009	0.4	0.2	77.0
ROANOKE,VA-NC-WV	66	87	29386	55172	0.4	0.2	53.3
TALLAHASSEE, FL-GA	67	70	27462	81933	0.4	0.3	33.5
MISSOULA, MT	68	85	27264	56397	0.4	0.2	48.3
EL PASO,TX-NM	69	60	26714	99776	0.4	0.4	26.8
CHARLESTON-NORTH CHARLESTON, SC	70	78	26577	71820	0.4	0.3	37.0
ALBANY-SCHENECTADY-TROY, NY	71	91	26020	46676	0.4	0.2	55.7
MILWAUKEE-RACINE,WI	72	53	25304	121260	0.4	0.5	20.9
PORTLAND, ME	73	89	24253	52889	0.3	0.2	45.9
ANCHORAGE,AK	74	83	23867	60690	0.3	0.2	39.3
KANSAS CITY, MO-KS	75	32	23274	217357	0.3	0.9	10.7
SALT LAKE CITY-OGDEN, UT-ID	76	26	21582	292142	0.3	1.2	7.4
NORFOLK-VIRGINIA BEACH-NEWPORT NEWS,VA-NC	77	42	20659	146234	0.3	0.6	14.1
APPLETON-OSHKOSH-NEENAH,WI	78	99	19984	39748	0.3	0.2	50.3

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
FLAGSTAFF,AZ-UT	79	63	19346	95079	0.3	0.4	20.3
DES MOINES, IA-IL-MO	80	81	18388	65719	0.3	0.3	28.0
COLUMBUS, GA-AL	81	120	18322	29585	0.3	0.1	61.9
FAYETTEVILLE, NC	82	94	17862	44960	0.3	0.2	39.7
FORT SMITH,AR-OK	83	103	17439	38329	0.2	0.2	45.5
ALBANY, GA	84	122	17392	28279	0.2	0.1	61.5
MADISON,WI-IL-IA	85	68	17218	90070	0.2	0.4	19.1
GREENVILLE, NC	86	75	17189	76217	0.2	0.3	22.6
BOISE CITY, ID-OR	87	64	16363	94657	0.2	0.4	17.3
CEDAR RAPIDS, IA	88	97	15854	41495	0.2	0.2	38.2
TUCSON,AZ	89	39	15559	156152	0.2	0.6	10.0
MONTGOMERY,AL	90	121	15332	28740	0.2	0.1	53.3
HUNTSVILLE,AL-TN	91	62	14549	95515	0.2	0.4	15.2
BILLINGS, MT-WY	92	104	14120	38075	0.2	0.2	37.1
SANTA FE, NM	93	92	13907	45646	0.2	0.2	30.5
FARMINGTON, NM-CO	94	118	13857	32731	0.2	0.1	42.3
ELKHART-GOSHEN, IN-MI	95	88	13678	54451	0.2	0.2	25.1
WAUSAU,WI	96	114	13168	33789	0.2	0.1	39.0
FORT WAYNE, IN	97	96	12561	41578	0.2	0.2	30.2
NORTHERN MICHIGAN, MI	98	119	11738	30328	0.2	0.1	38.7
TOLEDO, OH	99	128	11735	22278	0.2	0.1	52.7
CHATTANOOGA,TN-GA	100	80	11320	66738	0.2	0.3	17.0
SIOUX FALLS, SD-IA-MN-NE	101	112	11229	34676	0.2	0.1	32.4
MCALLEN-EDINBURG-MISSION,TX	102	46	10900	134396	0.2	0.5	8.1
JOHNSON CITY-KINGSPORT-BRISTOL, TN-VA	103	106	10189	37814	0.1	0.1	26.9
SHREVEPORT-BOSSIER CITY, LA-AR	104	135	9514	19402	0.1	0.1	49.0
WICHITA, KS-OK	105	86	9403	55395	0.1	0.2	17.0
TUPELO, MS-AL-TN	106	109	8317	36657	0.1	0.1	22.7
GREEN BAY,WI-MI	107	100	7838	39294	0.1	0.2	19.9
BANGOR, ME	108	129	7775	21896	0.1	0.1	35.5
BILOXI-GULFPORT-PASCAGOULA,MS	109	105	7395	37983	0.1	0.2	19.5

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
JOPLIN, MO-KS-OK	110	125	7321	25310	0.1	0.1	28.9
ASHEVILLE, NC	111	84	7128	60566	0.1	0.2	11.8
BEAUMONT-PORT ARTHUR,TX	112	126	7081	24193	0.1	0.1	29.3
FAYETTEVILLE-SPRINGDALE-ROGERS, AR-MO-OK	113	74	6468	78116	0.1	0.3	8.3
OMAHA, NE-IA-MO	114	71	6148	79012	0.1	0.3	7.8
CHARLESTON,WV-KY-OH	115	149	5965	11499	0.1	0.0	51.9
TWIN FALLS, ID	116	137	5897	16398	0.1	0.1	36.0
REDDING, CA-OR	117	110	5881	35619	0.1	0.1	16.5
ROCHESTER, NY-PA	118	108	5755	37034	0.1	0.1	15.5
PEORIA-PEKIN, IL	119	154	5624	9727	0.1	0.0	57.8
SAN ANGELO,TX	120	153	5343	10135	0.1	0.0	52.7
EVANSVILLE-HENDERSON, IN-KY-IL	121	123	4902	26868	0.1	0.1	18.2
HICKORY-MORGANTON, NC-TN	122	98	4791	40938	0.1	0.2	11.7
LAKE CHARLES, LA	123	132	4401	21460	0.1	0.1	20.5
STATE COLLEGE, PA	124	162	4257	5810	0.1	0.0	73.3
CORPUS CHRISTI,TX	125	113	4242	34066	0.1	0.1	12.5
TRAVERSE CITY, MI	126	93	3980	45129	0.1	0.2	8.8
GRAND ISLAND, NE	127	159	3768	6763	0.1	0.0	55.7
CASPER,WY-ID-UT	128	130	3689	21499	0.1	0.1	17.2
SIOUX CITY, IA-NE-SD	129	155	3637	9678	0.1	0.0	37.6
PADUCAH, KY-IL	130	146	3573	13129	0.1	0.1	27.2
FARGO-MOORHEAD, ND-MN	131	133	3515	21399	0.0	0.1	16.4
DAYTON-SPRINGFIELD, OH	132	156	3245	9242	0.0	0.0	35.1
IDAHO FALLS, ID-WY	133	115	3180	33769	0.0	0.1	9.4
BURLINGTON,VT-NY	134	102	2721	38572	0.0	0.2	7.1
DULUTH-SUPERIOR, MN-WI	135	157	2543	8166	0.0	0.0	31.1
CHAMPAIGN-URBANA, IL	136	148	1827	11552	0.0	0.0	15.8
SAN DIEGO, CA	171	19	0	378111	0.0	1.5	0.0
FORT MYERS-CAPE CORAL, FL	169	37	0	174448	0.0	0.7	0.0
MACON, GA	157	95	0	41730	0.0	0.2	0.0
COLUMBIA, MO	152	107	0	37431	0.0	0.1	0.0

EA Names	Sprawl Growth Rank	Overall Growth Rank	House-hold Growth in Sprawl Counties	House-hold Growth in All Counties	% of U.S. House-hold Growth Desig. as Sprawl	% of All U.S. House-hold Growth	% of County Growth Desig. as Sprawl
LINCOLN, NE	140	111	0	35447	0.0	0.1	0.0
LAFAYETTE, LA	156	116	0	33383	0.0	0.1	0.0
ROCHESTER, MN-IA-WI	147	124	0	25477	0.0	0.1	0.0
STAUNTON,VA-WV	168	127	0	23117	0.0	0.1	0.0
SYRACUSE, NY-PA	151	131	0	21491	0.0	0.1	0.0
AMARILLO,TX-NM	163	134	0	19844	0.0	0.1	0.0
DOTHAN, AL-FL-GA	159	136	0	19340	0.0	0.1	0.0
LA CROSSE,WI-MN	149	138	0	15914	0.0	0.1	0.0
TOPEKA, KS	138	139	0	15178	0.0	0.1	0.0
ODESSA-MIDLAND,TX	166	140	0	15173	0.0	0.1	0.0
PUEBLO, CO-NM	162	141	0	15028	0.0	0.1	0.0
RAPID CITY, SD-MT-NE-ND	141	142	0	14275	0.0	0.1	0.0
PENDLETON, OR-WA	172	143	0	14072	0.0	0.1	0.0
DAVENPORT-MOLINE-ROCK ISLAND, IA-IL	150	144	0	13271	0.0	0.1	0.0
BUFFALO-NIAGARA FALLS, NY-PA	167	145	0	13186	0.0	0.1	0.0
LUBBOCK,TX	164	147	0	11970	0.0	0.0	0.0
SPRINGFIELD, IL-MO	153	150	0	10664	0.0	0.0	0.0
JONESBORO,AR-MO	154	151	0	10617	0.0	0.0	0.0
MONROE, LA	155	152	0	10213	0.0	0.0	0.0
BISMARCK, ND-MT-SD	143	158	0	7369	0.0	0.0	0.0
ABILENE,TX	170	160	0	6363	0.0	0.0	0.0
HOBBS, NM-TX	165	161	0	6324	0.0	0.0	0.0
ERIE, PA	146	163	0	5373	0.0	0.0	0.0
GRAND FORKS, ND-MN	145	164	0	4277	0.0	0.0	0.0
GREAT FALLS, MT	160	165	0	2525	0.0	0.0	0.0
MINOT, ND	144	166	0	1905	0.0	0.0	0.0
WESTERN OKLAHOMA, OK	137	167	0	1396	0.0	0.0	0.0
SCOTTSBLUFF, NE-WY	161	168	0	1172	0.0	0.0	0.0
ABERDEEN, SD	142	169	0	1165	0.0	0.0	0.0
NORTH PLATTE, NE-CO	139	170	0	783	0.0	0.0	0.0
GREENVILLE, MS	158	171	0	-4775	0.0	0.0	0.0
WHEELING,WV-OH	148	172	0	-7546	0.0	0.0	0.0

Note:

“Desig.” in column headings represents “Designated”.

Table 4.14 Top 30 Overall Growth Counties, 2000-2020

NAME	Overall Growth Rank	Household Growth	Percentage of All U.S. Household Growth (%)	Land Use Type
MARICOPA, AZ	1	660249	2.6	Suburban
SAN DIEGO, CA	2	378111	1.5	Suburban
CLARK, NV	3	373355	1.5	Urban Center
HARRIS, TX	4	341663	1.3	Urban Center
LOS ANGELES, CA	5	334478	1.3	Urban Center
RIVERSIDE, CA	6	307587	1.2	Rural Center
ORANGE, CA	7	297677	1.2	Urban Center
BROWARD, FL	8	273333	1.1	Urban
PALM BEACH, FL	9	271020	1.1	Suburban
TARRANT, TX	10	264411	1.0	Urban
SAN BERNARDINO, CA	11	235262	0.9	Rural
BEXAR, TX	12	189281	0.7	Urban
DALLAS, TX	13	186156	0.7	Urban Center
TRAVIS, TX	14	183409	0.7	Urban
SACRAMENTO, CA	15	180869	0.7	Suburban
ORANGE, FL	16	178525	0.7	Urban
MIAMI-DADE, FL	17	178064	0.7	Urban
COLLIN, TX	18	168134	0.7	Suburban
KING, WA	19	165002	0.7	Urban
HILLSBOROUGH, FL	20	164548	0.7	Suburban
FAIRFAX + FAIRFAX CIT + FALLS CHURCH, VA	21	146031	0.6	Urban
WAKE, NC	22	139715	0.6	Suburban
CONTRA COSTA, CA	23	137986	0.5	Suburban
MECKLENBURG, NC	24	137832	0.5	Urban
PIMA, AZ	25	136568	0.5	Suburban
ALAMEDA, CA	26	132219	0.5	Urban
GWINNETT, GA	27	128431	0.5	Urban
COBB, GA	28	127612	0.5	Urban
DU PAGE, IL	29	125125	0.5	Urban
DENTON, TX	30	124446	0.5	Suburban

References

1. Abrams, Charles. 1971. *The Language of Cities*. New York: Viking.

2. Altshuler, Alan A., and Jose A. Gomez-Ibanez. 1993. *Regulation for Revenue: The Political Economy of Land Use Exactions*. Washington, DC: Brookings Institution.

3. Black, Thomas J. 1996. "The Economics of Sprawl." *Urban Land* 55, 3: 6-52.

4. Brody, Samuel D., Virginia Carrasco, and Wesley E. Highfield. 2006. "Measuring the Adoption of Local Sprawl". *Journal of Planning Education and Research* 25: 294-310. http://research.arch.tamu.edu/epsru/pdf/JPER_sprawl.pdf (accessed March 2013).

5. Burchell, Robert W., Anthony Downs, Samuel Seskin, Terry Moore, Naveed Shad, David Listokin, Judy S. Davis, David Helton, Michelle Gall, and Hilary Phillips. 1998. *TCRP Report 39: The Costs of Sprawl- Revisited*. Washington, DC: National Academy Press.

6. Burchell, Robert W., George Lowenstein, William R. Dolphin, Catherine C. Galley. 2002. *TCRP Report 74: Costs of Sprawl- 2000*. Washington, DC: National Academy Press.

7. Burchfield, Marcy, Henry G. Overman, Diego Puga, Matthew A. Turner. 2006. "Causes of Sprawl: A Portrait from Space." *The Quarterly Journal of Economics* May 2006: 587-633.

8. Census Glossary, http://www.census.gov/dmd/www/glossary.html#C (accessed April 2009).

9. Cervero, Robert. 1991 "Land Use and Travel at Suburban Activity Centers." *Transportation Quarterly* 45, 4: 479-91.

10. Cutsinger, Jackie, George Galster, Harold Wolman, Royce Hanson, and Douglas Towns. 2005. "Verifying the Multi-dimensional Nature of Metropolitan Land Use: Advancing the Understanding and Measurement of Sprawl." *Journal of Urban Affairs* 27, 3: 235-259.

11. Downs, Anthony. 1994. *New Visions for Metropolitan America*. Washington, DC: Brookings Institution; and Cambridge, MA: Lincoln Institute of Land Policy.

12. Downs, Anthony. 1998a. "How America's Cities Are Growing: The Big Picture." *Brookings Review* 16, 4: 8-12.

13. Downs, Anthony. 1998b. “Relating Suburban Sprawl to Urban Decline.” Unpublished manuscript.

14. Downs, Anthony. 1999. “Some Realities about Sprawl and Urban Decline.” *Housing Policy Debate* 10, 4: 955-974.

15. Ewing, Reid. 1997. “Is Los Angeles-Style Sprawl Desirable?” *Journal of the American Planning Association* 63: 107-126.

16. Ewing, Reid, Rolf Pendall, and Donald Chen. 2002. *Measuring Sprawl and Its Impact* (Technical Report). World Wide Web page < http://www.smartgrowthamerica.org/documents/MeasuringSprawlTechnical.pdf> (accessed September 2012).

17. Ewing, Reid, Rolf Pendall, and Donald Chen. 2003a. “Measuring Sprawl and Its Transportation Impacts.” *Transportation Research Record* 1831: 175-183.

18. Ewing, Reid, Tom Schmid, Richard Killingsworth, Amy Zlot, and Stephen Raudenbush. 2003b. “Relationship between Urban Sprawl and Physical Activity, Obesity, and Morbidity.” *American Journal of Health Promotion* 18, 1: 47-57.

19. Frenkel, Amnon, and Maya Ashkenazi. 2005. “Measuring Urban Sprawl—How Can We Deal With It?” *European Regional Science Association ERSA Conference Papers ersa05p50*. http://ideas.repec.org/p/wiw/wiwrsa/ersa05p50.html (accessed March 2013)

20. Galster, George, Royce Hanson, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage. 2001. “Wrestling Sprawl to The Ground: Defining and Measuring an Elusive Concept.” *Housing Policy Debate* 12, 4: 681-717. Fannie Mae Foundation 2001.

21. Glaeser, E., M. Kahn, and C. Chu. 2001. *Job Sprawl: Employment Location in U.S. Metropolitan Areas*. Center for Urban and metropolitan Policy, The Brookings Institution, Washington, D.C.

22. Gordon, Peter, and Harry W. Richardson. 1997. “Are Compact Cities a Desirable Planning Goal?” *Journal of the American Planning Association* 63: 95-106.

23. Harvey, Robert O., and William A. V. Clark. 1965. “The Nature and Economics of Sprawl.” *Land Economics* 41, 1:1-9.

24. Haydon, D. 2004. *A Field Guide to Sprawl*. W. W. Norton.

25. Knaap, Gerrit J. 1985. “The Price Effects of an Urban Growth Boundary in Metropolitan Portland, Oregon.” *Land Economics* 61: 26-35.

26. Knaap, Gerrit, Yan Song, Reid Ewing, and Kelly Clifton. 2005. "Seeing the Elephant: Multi-disciplinary Measures of Urban Sprawl." http://smartgrowth.umd.edu/multidisciplinarymeasuresurbansprawl.html (Accessed March 2013).

27. Kurban, Haydar and Joseph Persky. 2007. "Do Metropolitan Areas with Rich Central Cities Experience Less Sprawl?" *Economic Development Quarterly* 21, 2: 179-184.

28. Lockwood, Charles. 1999. "Sprawl." *Hemispheres*, September: 82-91.

29. Lopez, Russ and Patricia H. Hynes. 2003. "Sprawl in the 1990s Measurement, Distribution, and Trends." *Urban Affairs Review* 38, 3: 325-355.

30. Mills, David E. 1980. "Growth, Speculation, and Sprawl in a Monocentric City." *Journal of Urban Economics* 10: 201-226.

31. Moe, Richard. 1996. *Growing Smarter: Fighting Sprawl and Restoring Community in America*. Washington, DC: National Trust for Historic Preservation.

32. Orfield, Myron. 1997. *Metropolitics: A Regional Agenda for Community and Stability*. Washington, DC: Brookings Institution, and Cambridge, MA: Lincoln Institute of Land Policy.

33. Pendall, Rolf. 1999. "Do Land-use Controls Cause Sprawl?" *Environment and Planning B: Planning and Design* 26, 4: 555-571.

34. Popenoe, David. 1979. "Urban Sprawl: Some Neglected Sociology." *Sociology and Social Research* 31, 2:181-188.

35. Richmond, Henry R. 1995. *Regionalism: Chicago as an American Region*. Chicago: John D. and Catherine T. MacArthur Foundation.

36. Shen, Q. 1996. "Spatial Impacts of Locally Enacted Growth Controls: the San Francisco Bay Area in the 1980s." *Environment and Planning B: Planning and Design* 23: 61-91.

37. Sierra Club. 1998. *Sprawl: The Dark Side of the American Dream*. World wide web page http://www.sierraclub.org/sprawl/report98/ <Accessed September 2012>.

38. Torrens, Paul and Marina Alberti. 2000. *Measuring sprawl*. Unpublished paper No. 27. University College, London. Center for Advanced Spatial Analysis.

39. USA Today, February 22, 2001.

40. U.S. Department of Agriculture. 2009. Summary Report: 2007 National Resources Inventory, Natural Resources Conservation Service, Washington, DC, and Center for

Survey Statistics and Methodology, Iowa State University, Ames, Iowa. 123 pages. http://www.nrcs.usda.gov/technical/NRI/2007/2007_NRI_Summary.pdf

41. Wolman, Harold, George Galster, Royce Hanson, Michael Ratcliffe, Kimberly Furdell, Andrea Sarzynski. 2005. “The Fundamental Challenge in Measuring Sprawl: Which Land Should Be Considered?” *The Professional Geographer* 57, 1: 94–105.

42. Woods & Poole 2003 Regional Projections and Database

43. Woods & Poole 2012 Regional Projections and Database

44. Weitz, Jerry and Tom Crawford. 2012. “Where the Jobs Are Going: Job Sprawl in U.S. Metropolitan Regions, 2001-2006.” *Journal of the American Planning Association* 78, 1: 53-69.

45. Yin, Ming and Jian Sun. 2007. “The Impacts of State Growth Management Programs on Urban Sprawl in the 1990s.” *Journal of Urban Affairs* 29, 2: 149-179.