The Causal Effects of Active Living-Oriented Zoning on Adult Leisure Time Activities

BY

XUAN YIN
Bachelor, Shanghai University of Finance and Economics, China 2006
Master, Shanghai University of Finance and Economics, China 2010
M.S., University of Illinois at Chicago, Chicago, 2013
M.A., University of Illinois at Chicago, Chicago, 2014

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Chicago, 2017

Chicago, Illinois

Defense Committee:

Frank Chaloupka, Chair and Advisor
Marcus Casey
Darren Lubotsky
Jamie Chriqui, Health Policy and Administration
Lisa Powell, Health Policy and Administration

Abba Anthony said, “I saw all the traps of the enemy spread over the earth, and groaning, said, ‘What can get through these?’ Then I heard a voice saying to me, ‘Humility.’ ”

ACKNOWLEDGMENTS

I acknowledge the professors in the committee - Frank Chaloupka, Marcus Casey, Jamie Chriqui, Darren Lubotsky, and Lisa Powell - for their comments.

I appreciate Prof. Frank Chaloupka for his advice.

I want to thank the research team in the Institute for Health Research and Policy of UIC - Prof. Jamie Chriqui, Julien Leider, and Emily Thrun - for providing the zoning data and related materials to me and helping me understand the zoning data.

I am grateful to my dear friends - Chenjie Tang, Shaolong (Shalong) Jiang, Yanchee (Yanqi) Yang, Duoyi Hu, and Juliane Zhang - for their encouragement and comfort.

I owe special thanks to my mother, Huili Ding, and my grandmother, Meiyu Qian. Without them, it would never have been accomplished.

TABLE OF CONTENTS

1 BACKGROUND AND INTRODUCTION 1

2 LITERATURE REVIEW 10

2.1 Association between Zoning and Active Living 10

2.2 Association between Built Environment and Active Living 12

2.3 Contribution to the Literature 21

3 DATA AND VARIABLE CONSTRUCTION 26

3.1 Zoning Data 26

3.2 NAVTEQ 2011 35

3.3 American Time Use Survey (ATUS) 36

3.4 American Community Survey (ACS) 45

3.5 1900 Census: Population, Housing, Agriculture, and Manufacturing Data 45

3.6 Summary Statistics 46

4 MEDIATIONAL ANALYSIS 60

4.1 Conceptual Framework 60

4.2 Basic Model: Two-Part Model 63

4.3 Three Steps for Mediational Analysis 64

4.4 Sobel Test 65

4.5 Discussions and Results 66

5 INSTRUMENTAL VARIABLE IDENTIFICATION AND ESTIMATION 75

5.1 Tiebout Sorting, Tiebout Equilibrium, and Market Equilibrium 75

5.2 Traveling Behavior and Sorting 78

5.3 Aerobics Regression and the Analysis of Active Living Preference 79

5.4 Instrumental Variable Identification 82

5.5 The Exclusion Restriction 88

5.5.1 Assumption 1: IVs Have No Direct Associations with Individual Active Living Preference 88

TABLE OF CONTENTS (CONTINUED)

5.5.2 Assumption 2: The Causal Chain Goes From IVs to Endogenous Variables 89

5.5.3 Assumption 3: The Sources of Variation of IVs Have No Associations with Individual Active Living Preference 89

5.5.4 Assumption 4: The Uncontrolled Channels from IVs to Active Living Preference Are Weak Enough to Be Negligible 92

5.6 Covariate Balance Test 94

5.7 IV Asymptotical Bias and the Link between Selection on Observables and Unobservables 99

5.8 IV First Stage 105

5.9 IV Probit Model: A Simultaneous Equation System 112

5.10 General Method of Moments 117

5.11 Testing for Endogeneity: Hausman Specification Test and C Test 121

5.12 Overidentification Test 124

5.13 Discussions and Results 125

6 POLICY SIMULATION 143

6.1 Methods 143

6.2 Results 144

7 CONCLUSION 152

REFERENCES 157

APPENDICES 175

Appendix 1 175

Appendix 2 177

Appendix 3 178

Appendix 4 180

Appendix 5 182

Appendix 6 184

TABLE OF CONTENTS (CONTINUED)

Appendix 7 187
Appendix 8 189

VITA 192

LIST OF TABLES

Table 3.1.1 Active Recreation Markers 32

Table 3.1.2 Active Recreation Markers Not Included 33

Table 3.1.3 Passive Recreation Markers 33

Table 3.1.4 Passive Recreation Markers Not Included 33

Table 3.1.5 Population-Weighted County-Aggregated Zoning Variable (Sidewalks) Construction Instruction 35

Table 3.3.1 ATUS Coding Structure: First-Tier Activity Groups 39

Table 3.3.2 Dependent Variable Construction 42

Table 3.3.3 Activity Examples for Physical Activity 2010 - 2015 42

Table 3.3.4 Activity Examples for Sedentary Behavior 2010 - 2015 43

Table 3.6.1 Characteristics of All Counties from the Data of Zoning, NAVTEQ 2011, and ACS 2011-2015 5-Year Estimates (Before Merged with ATUS and 1900 Census) 51

Table 3.6.2 Characteristics of Counties in the Analysis 52

Table 3.6.3 Characteristics of Counties Excluded from the Analysis 53

Table 3.6.4 Characteristic Comparison between Counties Included in the Analysis and Counties Excluded from the Analysis 54

Table 3.6.5 Characteristic Comparison between Counties Included in the Analysis and All Counties (Counties in the Zoning Data) 55

Table 3.6.6 Physical Activity Summary Statistics 56

Table 3.6.7 Sedentary Behavior Summary Statistics 58

Table 4.5.1 The Mediational Analysis, Step 1: Physical Activity Regression 69

Table 4.5.2 The Mediational Analysis, Step 1: Sedentary Behavior Regression 70

Table 4.5.3 The Mediational Analysis, Step 2: Walkability Regression 71

Table 4.5.4 The Mediational Analysis, Step 3: Physical Activity Regression 72

Table 4.5.5 The Mediational Analysis, Step 3: Sedentary Behavior Regression 73

LIST OF TABLES (CONTINUED)

Table 5.3.1 The Mediational Analysis, Aerobics Regression 81

Table 5.6.1 Correlation Analysis on Manufacturing Establishment Density 1900 and Other Covariates 97

Table 5.6.2 Correlation Analysis on Farmland Proportion 1900 and Other Covariates 98

Table 5.7.1 2SLS Estimates of Zoning Effects and Estimates of Potential Bias Using Altonji, Elder, and Taber (2005b) Methodology 104

Table 5.8.1 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 1: Physical Activity Regression 108

Table 5.8.2 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 1: Sedentary Behavior Regression 109

Table 5.8.3 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 3: Physical Activity Regression 110

Table 5.8.4 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 3: Sedentary Behavior Regression 111

Table 5.13.1 The Mediational Analysis with Instrumental Variables, Step 1: Physical Activity Regression 138

Table 5.13.2 The Mediational Analysis with Instrumental Variables, Step 1: Sedentary Behavior Regression 139

Table 5.13.3 The Mediational Analysis with Instrumental Variables, Step 2: Walkability Regression 140

Table 5.13.4 The Mediational Analysis with Instrumental Variables, Step 3: Physical Activity Regression 141

Table 5.13.5 The Mediational Analysis with Instrumental Variables, Step 3: Sedentary Behavior Regression 142

Table 6.1.1 The Coefficients of Active Living-Oriented Zoning from Step 1 of the Mediational Analysis 143

Table 6.2.1 The Simulation Results on Physical Activity 146

Table 6.2.2 The Simulation Results on Sedentary Behavior 147

LIST OF FIGURES

Figure 2.1.1 Euclidean Zoning vs. Transect-Based Zoning 1 8

Figure 2.1.2 Euclidean Zoning vs. Transect-Based Zoning 2 9

Figure 4.1.1 Conceptual Framework 1 62

Figure 4.1.2 Conceptual Framework 2 62

Figure 4.5.1 Physical Activity, Probit Marginal Effects 74

Figure 4.5.2 Sedentary Behavior, Probit Marginal Effects 74

Figure 5.1.1 Tiebout Sorting Illustration 77

Figure 5.4.1 The Illustration of IV Identification of Zhao and Kaestner (2010) 86

Figure 5.4.2 The Illustration of IV Identification 86

Figure 5.13.1 Physical Activity, IV Probit Marginal Effects 135

Figure 5.13.2 Sedentary Behavior, IV Probit Marginal Effects 135

Figure 6.2.1 The Predicted Probability of Physical Activity in the Full Analytical Sample .. 148

Figure 6.2.2 The Predicted Minutes of Physical Activity within the Analytical Sample of Physical Activity Participation 149

Figure 6.2.3 The Predicted Probability of Sedentary Behavior in the Full Analytical Sample 150

Figure 6.2.4 The Predicted Minutes of Sedentary Behavior within the Analytical Sample of Sedentary Behavior Participation 151

LIST OF ABBREVIATIONS

ACS	American Community Survey
ATE	Average Treatment Effect
ATUS	American Time Use Survey
ATV	All-Terrain Vehicle
COEFF	Coefficients
CPS	Current Population Survey
DAG	Directed Acyclic Graph
GIS	Global Information System
GMM	General Method of Moments
IV	Instrumental Variable
LATE	Local Average Treatment Effect
MLE	Maximum Likelihood Estimation
NCI	National Cancer Institute
OLS	Ordinary Least Squares
RV	Recreational Vehicle
SUR	Seemingly Unrelated Regression
SW	Sanderson-Windmeijer (Sanderson and Windmeijer 2016)
TND	Traditional Neighborhood Development
YMCA	Young Men's Christian Association
2SLS	Two-Stage Least Squares

SUMMARY

The author of this dissertation investigates the causal effects of zoning on adults' leisure-time physical activity and sedentary behavior. This is the first study on the causal effects of zoning on active living outcomes, and the first study to quantitatively verify its underlying mechanisms (i.e., the mediation of the built environment for the effect of zoning on active living outcomes).

The first chapter provides an introduction to physical activity trends and zoning. By 2015, a certain percentage of the adult population still did not reach the standards of physical activity guidelines for Americans set by U.S. Department of Health and Human Services. It is important to understand the causal factors that drive this fact. Numerous empirical studies revealed a positive association between active living-friendly built environment and walking/bicycling. Researchers also found that, among built environment elements, community- and street-scale urban design and land use policies and practices were effective at promoting physical activity. To implement community- and street-scale urban design and land use policies and practices, zoning is the most commonly used and useful tool. Zoning, at the county and municipal levels, is governed by local governments through zoning codes. The birth and development of zoning has a close relationship with public health. Traditional Euclidean zoning classifies residential, commercial, and industrial land uses into distinct zones without intersection. It prevented the spread of infectious disease in early ages when sanitation was inadequate, but it also created communities of auto reliance and a lack of physical activity. The Smart Growth and New Urbanism movements of the 1990s spurred zoning code reform to promote active living.

Chapter 2 provides reviews of literature on the association between zoning and active living and the association between built environment and active living. Researchers found a significant positive association between active living-oriented zoning/built environment and physical activity. A summary of the contributions of this dissertation is presented at the end of this chapter.

In Chapter 3, the author provides descriptions of data sources and variable construction. The data used in this project comes from five sources. Zoning data originate from the research team at the Institute for Health Research and Policy at the University of Illinois at Chicago and are used to construct the independent variable—active living-oriented zoning. NAVTEQ 2011 third quarter GIS data and the *American Community Survey* (ACS) 2011 1-year estimates were combined to build the county-level independent variable—walkability—which is used to measure built environment. The *American Time Use Survey* 2010–2015 is used to build the outcome variables of time usage and individual-level control variables. The ACS 2011–2015 five-year estimates are used to build the county-level control variables. The 1900 census is used to build the two instrumental variables, which are manufacturing establishment density and farmland proportion. The final analytical data cover 2,453 municipalities and unincorporated areas in 251 counties of 37 states and represent 39.08% of the U.S. population. Only 8.05 percent of the survey respondents in the sample participated in physical activity (running, walking, jogging, and bicycling). Conditional on participation, people spent an average of 1 hour per day engaging in physical activity. About 82 percent of the survey respondents in the sample engage in sedentary behavior. Conditional on participation, people spent nearly 4 hours per day engaging in sedentary behavior on average. In the final analytical data about physical activity analysis, the average county-level walkability scale was 1.55 in the full sample and 1.59

in the participation sample. On average, 56.2 percent of the people in the full sample and 59.4 percent of the people in the participation sample were exposed to active living-oriented zoning. In the final analytical data about sedentary behavior analysis, the average county-level walkability scale was 1.55 in the full sample and 1.52 in the participation sample. On average, 56.2 percent of the population in the full sample and 55.8 percent of the population in the participation sample were exposed to active living-oriented zoning. In the physical activity analysis and sedentary behavior analysis, both the full and participation samples featured large proportions of non-Hispanic White adults who were married and had college degrees. Both samples included slightly more females than males.

In Chapter 4, the author introduces and discusses the mediational analysis. A conceptual framework based on a directed acyclic graph is introduced. A two-part model is employed as a basic model. A Probit model is used to characterize whether adults participated in physical activity or sedentary behavior. Conditional on participation, the linear model is used to characterize how many minutes adults spent on physical activity or sedentary behavior. For the linear model, a logarithmic transformation of the outcome variable—the minutes—is also utilized to address the highly right-skewed distribution of the outcome variable. A one percent increase in the population that is exposed to the promotion of active living-oriented zoning is found to be associated with a 2.8 percentage point increase (Probit model) in the likelihood that an adult engaged in physical activity and a 4.6 percentage point decrease (Probit model) in the likelihood that an adult engaged in sedentary behavior when other factors remained unchanged. The effect of zoning are statistically significant at the 1% level. Nonetheless, the effects of zoning on the time people spent on physical activity or sedentary behavior is statistically insignificant. The Sobel test is conducted on the indirect effects of zoning. It reveals that zoning has an insignificant indirect effect on physical activity and a significant indirect effect on sedentary behavior. The indirect effect of zoning on sedentary behavior is established. The insignificant indirect effect of zoning on physical activity that was revealed by the Sobel test may be due to the fact that the Sobel test has insufficient power to correctly reject the false null hypothesis when the indirect effect is positive and zoning has a positive indirect effect on physical activity. Therefore, rather than being based on the Sobel test, the indirect effect of zoning on physical activity is established based on the changes in the magnitude of marginal effects.

In Chapter 5, the author introduces and discusses instrumental variable identification and estimation. Previous literature did not address the self-selection problem, which makes the literature fail to establish a causal relationship between zoning and active living outcomes. Because people choose where they live, the observed zoning and built environment can be the choice results. Self-selection makes the estimates in the mediational analysis upward biased and inconsistent. Self-selection is modeled as unobserved confounder problem. The unobserved confounder, here, is active living preference. The use of instrument variable identification and estimation are proposed to solve the problem. Two instrumental variables are manufacturing establishment density 1900 and farmland proportion 1900. The assumption for the two instrumental variables is they affect active living outcomes only through the zoning and the built environment, which is also known as exclusion restriction. Although the assumption of exclusion restriction of instrumental variables cannot be rigorously verified using statistical analysis, an exploration of the history and the institution of American zoning and American manufacturing make the assumption of exclusion restriction convincible. The two instrumental variables pass the tests of weak identification and underidentification that are performed along with the first-stage regression. An IV Probit model and general method of moments (GMM) are employed for estimation. Hansen J statistic, which is based on the general method of moments, is utilized to test the endogeneity of zoning and walkability, and to test the overidentification in the first step of mediational analysis when there is only one endogenous variable—zoning—but two instrumental variables.

Through IV estimation, a one percent increase in the population exposed to the promotion of active living-oriented zoning is found to be associated with a 2.8 percentage point increase (IV Probit) in the likelihood that an adult engaged in physical activity, a 9.957 unit increase (GMM) or a 29.7 percent increase (GMM Log) in the minutes an adult spent on physical activity, a 1.4 percentage point decrease (IV Probit) in the chances that an adult chose sedentary behavior, and an 8.9 percent decrease (GMM Log) in the minutes an adult spent on sedentary behavior. Although the marginal effects in participation probabilities are nearly insignificant due to the inefficiency of IV estimation and the marginal effect calculation procedure, the corresponding estimated coefficients in IV Probit are significant, which means the effects of zoning are significant. For sedentary behavior participation, endogeneity is found to be significant. For other models, endogeneity is found to be insignificant. Nonetheless, IV estimation for linear models corrects the measurement error caused by the impossibility of linking survey respondents to the zoning metrics of their specific areas within their county. Note that the marginal effects of zoning on the participation probabilities are causal effects on the whole population while the marginal effects on the minutes are weighted averages of local average treatment effects (LATE) in the sense of the causal effects of zoning on "compliers" whose behaviors change as the values of IVs change.

In the physical activity participation analysis, walkability is a mediator for zoning. This is concluded from the Probit model because the endogeneity tests show insignificant endogeneity in the physical activity participation analysis. In the sedentary behavior participation analysis, walkability is not a mediator for zoning. This is concluded from IV Probit because the endogeneity tests show significant endogeneity in the sedentary behavior participation analysis. For the duration of physical activity and sedentary behavior, the mediation of walkability for zoning cannot be established because most of the coefficients are insignificant in both ordinary least squares (OLS) and GMM. Note that, the mediational analysis in linear models with IV estimation can be invalid because the compliers of zoning and walkability in Step 3 can be different.

The failure of mediation of walkability in the causal path from zoning to sedentary behavior participation implies that some built environment elements that cannot be captured by the walkability metric play important roles in people's decisions to engage in sedentary behavior. The zoning variable, active living-oriented zoning, includes many more elements of built environment than walkability, active recreation (equipment that supports physical activity), passive recreation (not equipment, opportunities like open space for physical activity), mixed-use development, bike parking, and bike lanes. Many of these zoning provisions cannot be captured by the walkability metric that is constituted by four density metrics: population density, housing unit density, the ratio of four-way intersections to all intersections, and the total number of intersections divided by land area.

Based on the estimated models that have consistent and significant coefficients of zoning, Chapter 6 provides policy simulation to do within-analytical sample prediction. Compared to the real situation, if all places promoted active living-oriented zoning, then they could increase the proportion of the adult population that engaged in physical activity by 1.25 percentage points (from Probit) or 0.95 percentage points (from IV Probit) and the time that an adult spent on physical activity by 4 minutes (from GMM) or 8 minutes (from GMM Log). They could also decrease the proportion of the adult population that engaged in sedentary behavior by 0.6 percentage points (from IV Probit) and the time that an adult spent on sedentary behavior by 9 minutes (from GMM Log).

Chapter 7 concludes the dissertation. In it, the author summarizes the analysis and opens the discussion to policy implications. Overall, the results of this dissertation suggest that active living-oriented zoning, indeed, promotes physical activity, such as walking, running, and bicycling, and discourages sedentary behaviors at home, such as TV watching, radio listening, and music listening, partly through shaping built environment. Nowadays, an increasing number of professionals, researchers, and leaders from public and private organizations advocate that local governments should upgrade their built environments to promote physical activity. All of them provide detailed strategies with which to make active living-friendly built environment, and all of them mention that implementing zoning laws could be an effective way to promote active living-friendly built environment. This dissertation serves as a sound foundation for policies and strategic plans to address pedestrian- and cyclist-friendly zoning provisions to promote active living.

1 BACKGROUND AND INTRODUCTION

Being physically active is helpful to a person's health. It reduces the risk of many chronic diseases. To promote physical activity, the U.S. Department of Health and Human Services released the *2008 Physical Activity Guidelines for Americans*.[1] Some of the key guidelines for adults are the following:

> All adults should avoid inactivity. Some physical activity is better than none, and adults who participate in any amount of physical activity gain some health benefits. For substantial health benefits, adults should do at least 150 minutes (2 hours and 30 minutes) a week of moderate-intensity, or 75 minutes (1 hour and 15 minutes) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous intensity aerobic activity. Aerobic activity should be performed in episodes of at least 10 minutes, and preferably, it should be spread throughout the week. For additional and more extensive health benefits, adults should increase their aerobic physical activity to 300 minutes (5 hours) a week of moderate intensity, or 150 minutes a week of vigorous intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity activity. Additional health benefits are gained by engaging in physical activity beyond this amount. (U.S. Department of Health and Human Services 2008)

By 2015, 30% of adults still engaged in no leisure-time physical activity. About half of the adult population (49.8%) engaged in light or moderate regular physical activity for longer than 150 minutes per week or vigorous physical activity for more than 75 minutes per week, and only about one-third of the adult population (33.6%) engaged in light or moderate regular physical activity for more than 300 minutes per week or vigorous physical activity for more than 150 minutes per week (Office of Disease Prevention and Health Promotion 2017).

Factors positively associated with adult physical activity include high education level, high income, enjoyment of exercise, expectation about the benefits of excercising, belief in the ability to exercise (self-efficacy), history of activity in adulthood, social support from peers,

[1] Please see https://health.gov/paguidelines/guidelines/

family, or spouse, access to and satisfaction with facilities, enjoyable scenery, and safe neighborhoods. The factors negatively associated with adult physical activity include advancing age, low income, lack of time, low motivation, rural residency, the perception of high effort needed for exercise, being overweight or obese, the perception of poor health, and being disabled (Trost et al. 2002).[2]

Researchers have realized that the prevention approaches mainly target individuals with educational and motivational programs have limitations, so they increasingly focus on factors that influence behavior but are outside the individual, such as the built environment (Sallis et al. 2012). Numerous empirical studies revealed that there is a positive association between active living-friendly built environment and physical activity (physical activity mainly refers to walking and bicycling). The factors of built environment that have been found to be positively associated with walking and bicycling include: mixed uses of residential and commercial areas (Cervero and Duncan 2003; Saelens et al. 2003; Frank et al. 2005; Frank et al. 2006; Rundle et al. 2007; Li et al. 2008; Van Dyck et al. 2010; Ewing et al. 2015), residential (household) density (Saelens et al. 2003; Frank et al. 2005; Li et al. 2005; Frank et al. 2006; Van Dyck et al. 2010), regional compactness development (Rundle et al. 2007; Aytur et al. 2008; Ewing et al. 2015), street connectivity (Saelens et al. 2003; Frank et al. 2006; Li et al. 2008; Van Dyck et al. 2010), intersection density (Frank et al. 2005; Li et al. 2005; Ewing et al. 2015), green and open spaces for recreation (Aytur et al. 2008; Li et al. 2005), density of public transit stations (Rundle et al. 2007; Li et al. 2008; Ewing et al. 2015), presence of sidewalks (Davison and Lawson 2006), density of places of employment (Li et al. 2005), block size, gridiron streets (Cervero and Duncan 2003), and enjoyable scenery (Brownson et al. 2001).

[2] From https://www.healthypeople.gov/2020/topics-objectives/topic/physical-activity

Built environment elements, based on different scales or dimensions of interventions, can be classified into three categories: community-scale urban design and land-use policies and practices (e.g., mixed land uses, sidewalk quality, and street connectivity), street-scale urban design and land-use policies and practices (e.g., street lighting, ease and safety of street crossing), and transportation and travel policies and practices (e.g., pedestrian, transit, and light rail access). *The Community Guide* concludes that, among the three interventions, community- and street- scale urban design and land-use policies and practices are effective at promoting physical activity (Heath et al. 2006; The Community Preventive Services Task Force 2016).

To implement community- and street-scale urban design and land-use policies and practices, zoning is the most commonly used and useful tool. "In the United States, the most common means and the best-known form of land-use control is municipal zoning... although a number of other tools (e.g., design review, development agreements, subdivision controls, parking requirements) are also in play... Some 95 percent of American locales, including almost all cities, use zoning, ostensibly to serve the public interest" (Hirt 2014). "The most common form of local land use regulation in this nation is zoning. Simply put, zoning entails separating the land in a particular area into sections, or zones, with different rules governing the activities on that land" (Pendall, Puentes, and Martin 2006). Besides zoning, other legal avenues that affect built environment include environmental regulations to reduce toxic emissions, building and housing codes that set standards for structures, taxation to provide financial incentives for developers to develop pedestrian- and bicyclist-friendly built environment, and spending to provide resources for projects that enhance the built environment (Perdue, Stone, and Gostin 2003).

Zoning at the county and municipal levels is governed by local governments through zoning codes. Zoning codes, which are typically composed of ordinance text and maps, divide local government's jurisdictions into zones and regulate land uses, building structures, development activities, and other aspects of each zone (American Planning Association 2006). Under the 10th Amendment to the Constitution, state legislatures adopted enabling laws to authorize local governments to control land use by adopting comprehensive plans (also called local master plans) and creating zoning districts (Nolon 2006; Schilling and Linton 2005). Zoning codes must conform to local comprehensive plans that are prepared by the local government to describe the current situation and future vision and goals of the community to the local community, including public officials, private citizens, and court judges. The local comprehensive plan advises public officials in making land-use decisions (including zoning) so that decisions can be accepted by private citizens and be respected as reasonable in courts (American Planning Association 2006; Norton 2008). States also authorize local governments to create administrative agencies to review and adjudicate individual proposals for land development (planning boards or commissions) and petitions for relief from zoning regulations (zoning boards of appeal). These agencies are required to hold public hearings on most proposals and petitions, where the public, especially homeowners, can vote for their own interests (Fischel 2004; Nolon 2006).

Zoning ordinances are rooted in public health. Before the birth of modern zoning ordinance, state police regulated land uses based on the common law of public nuisance to protect the public from the harm caused by certain activities. By enabling legislation, states delegate their police power, which includes land-use regulation, to local governments (Schilling and Linton 2005; American Planning Association 2006). The first zoning ordinance took place in New York City in 1916. It was a citywide regulation to stop massive commercial buildings from preventing light and air from reaching the streets below and creating congestion in residential areas. It mainly restricted the height of buildings to a proportion of the lot size (Fischel 2004; Schilling and Linton 2005; Dunlap 2016). In 1922, the Department of Commerce published the *Standard State Zoning Enabling Act* (SZEA), a model law for U.S. states to enable land-use regulations in their jurisdictions. In 1928, the Department of Commerce published the *Standard City Planning Enabling Act* (SCPEA). The two model laws serve as the core foundation for land use controls in the United States (American Planning Association 2006). By 1926, zoning was a new concept that was ruled to be unconstitutional by lower courts. The constitutionality of zoning was upheld by the U.S. Supreme Court in the case of Village of Euclid, Ohio v. Ambler Realty Co. in 1926. In 1922, the Village of Euclid, Ohio, adopted a zoning plan, like many other municipalities at that time, to protect the residential nature of farmland and undeveloped land, which was the majority of Euclid's land, from industrial uses. The zoning plan separated residential and industrial land uses into different districts and put restrictions on the features of buildings (lot area, size, height, etc.) in those districts. The Ambler Realty Company challenged this zoning plan because a tract of undeveloped land the company purchased would have been used for commercial or industrial businesses, which would bring more interest to the company as the company claimed, but the land could only be used for residential and community uses under the zoning plan. The Supreme Court supported the Village of Euclid in 1926. After that, the zoning plan of Euclid became popular and known as "Euclidean zoning" (Schilling and Linton 2005; Nolon 2006; Fischel 2004).

Traditional Euclidean zoning classifies residential, commercial, and industrial land uses into distinct zones without intersection. It is also called single-use zoning because within each zone, the land is used for a single purpose (residential use, commercial use, or industrial use). It aimed to protect public health in the industrial era by separating residents from the pollution and congestion created by rapidly-developing commercial and industrial businesses. This separation, however, created unhealthy communities. The unhealthiness comes from auto reliance and the lack of physical activity. The long distances between places have made it difficult for residents to travel by walking or bicycling. "Poor connectivity, single-use zones, and limited sidewalk infrastructure restrict routine opportunities for physical activity," "highways and other busy roads that bisect neighborhoods are not safe for walking, biking, or skateboarding" (Frank, Engelke, and Schmid 2003; Schilling 2005).

Zoning code reforms that promote active living were triggered by the Smart Growth and New Urbanism movements in the 1990s (Schilling and Linton 2005). Smart Growth is a movement focusing on development strategies. It is a reaction to suburban sprawl. Rather than restricting growth, it calls for compact, sensitive environmental growth (O'Connell 2008; Nolon 2006; Schilling 2005; Schilling and Linton 2005). Specifically, it encourages mixed land uses, compact building design, walkable neighborhoods, a variety of housing and transportation choices, and predictable, fair, and cost-effective developments.[3] Similar ideas can be found in New Urbanism, which is an urban design movement that focused on walkable blocks and streets, nearby housing and shopping, and accessible public spaces.[4] It is evident that traditional Euclidean zoning is a significant barrier to Smart Growth and New Urbanism. Proponents of these movements were interested in reforming zoning, e.g., a few urbanists created their own Smart Code and "other innovators have proposed form-based codes as a complement or perhaps

---

[3] See Smart Growth Online, Smart Growth Principles, at http://smartgrowth.org/smart-growth-principles/
[4] See Congress for the New Urbanism, What is New Urbanism, at https://www.cnu.org/resources/what-new-urbanism

even a substitute for zoning" (Schilling and Linton 2005). Zoning code reform models, different from conventional Euclidean code, include unified development codes, traditional neighborhood development (TND), reverse zoning (also known as TND lite), form-based codes (smart code is a variation of form-based codes) (Schilling 2005). All these code reforms promote highly compact neighborhood, mixed-use and open spaces, and pedestrian-friendly built environment (Schilling 2005; Schilling and Linton 2005; American Planning Association 2006).

Figure 2.1.1 and Figure 2.1.2 show the comparison between Euclidean zoning and transect-based zoning (Tachieva 2012). "The Transect defines a series of zones in transition from sparse rural farmhouses to the dense urban core. Each zone is fractal in that it contains a similar transition from the edge to the center of the neighborhood. The Transect is an important concept in the New Urbanism and smart growth movements" ("Understanding the Basics of Land Use and Planning: Glossary of Land Use and Planning Terms" 2010).

The left panel of Figure 2.1.1 shows a commercial-only area. The left panel of Figure 2.1.2 shows a residential-only area. Both feature low-density buildings, single building types, and single building uses. Transect-based zoning balances the areas with buildings of different kinds and various functions and increases building density. It transforms areas into mixed-use, diverse, and transit-ready communities so that the areas are highly pedestrian- and cyclist-friendly.

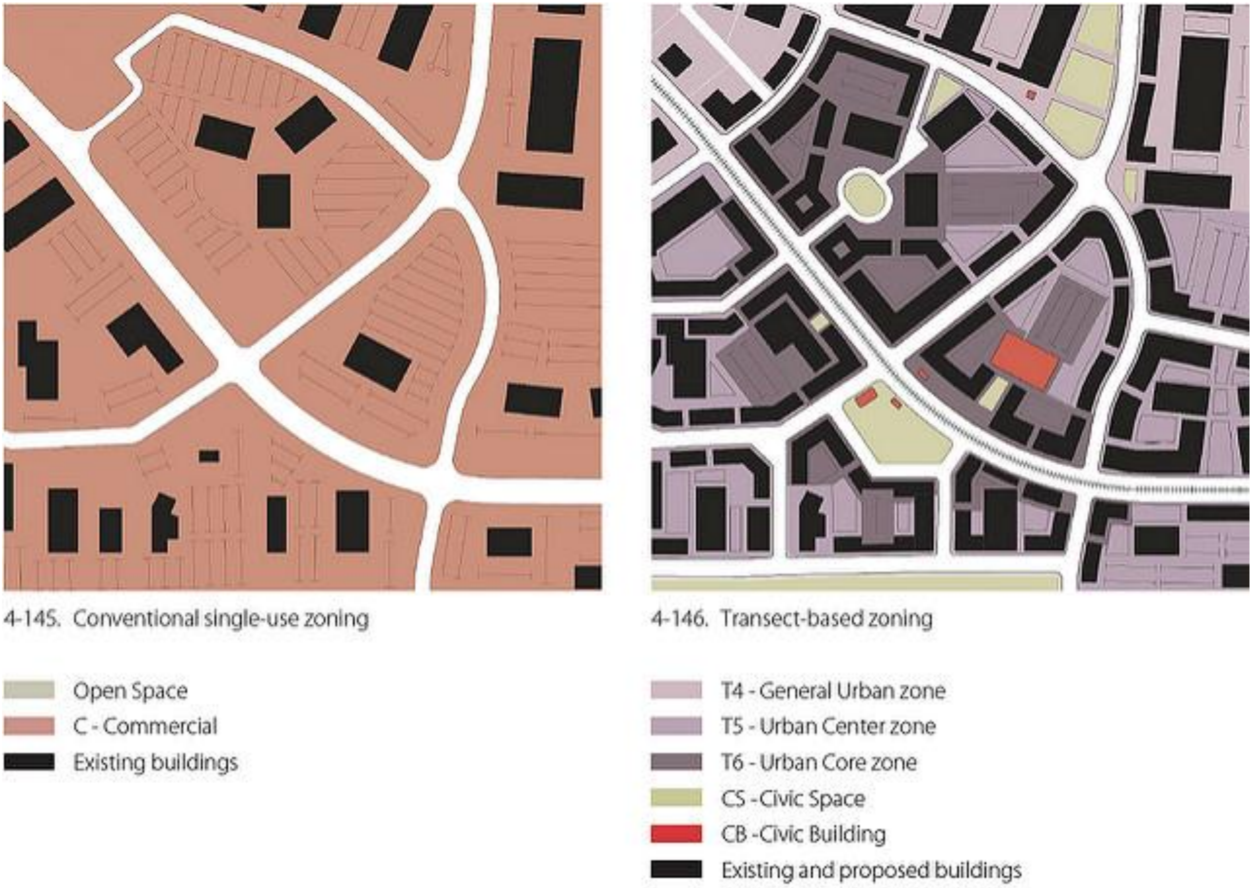
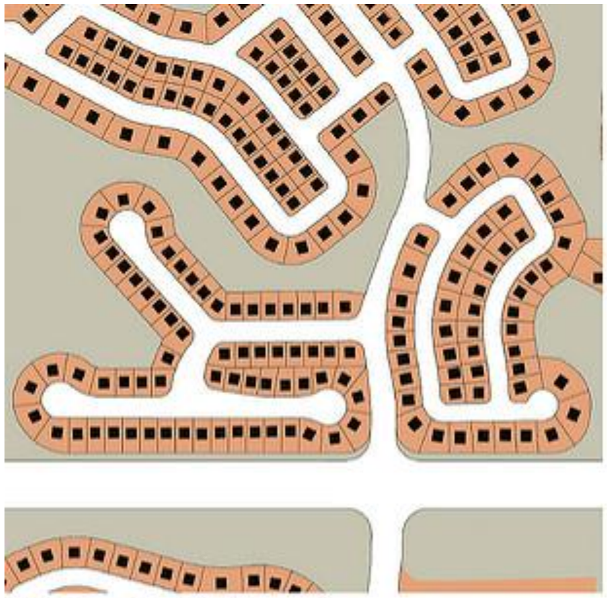


Figure 2.1.1 Euclidean Zoning vs. Transect-Based Zoning 1
From Galina Tachieva, Transect Codes Council Special Edition, August 2012
http://myemail.constantcontact.com/TCC-Special-Edition---Transect-for-Sprawl-Repair.html?soid=1103584053200&aid=NAA8D92NnIY



4-27. Conventional single-use zoning

Open Space
R1 - Single-family Residential
Existing buildings

4-28. Transect-based zoning

T1 - Natural zone
T3 - Sub-Urban zone
T4 - General Urban zone
T5 - Urban Center zone
CS -Civic Space
CB- Civic Building
Existing and proposed buildings

Figure 2.1.2 Euclidean Zoning vs. Transect-Based Zoning 2
From Galina Tachieva, Transect Codes Council Special Edition, August 2012
http://myemail.constantcontact.com/TCC-Special-Edition---Transect-for-Sprawl-Repair.html?soid=1103584053200&aid=NAA8D92NnIY

2 LITERATURE REVIEW

2.1 Association between Zoning and Active Living

Cannon et al. (2013) examined the effects of municipal mixed-use zoning on built walking potential. They evaluated municipal zoning ordinances in 22 California cities to construct the independent variable. They also evaluated potential walking destinations within each zone via Google Earth to build the outcome variable, the daily-use activity measure, which was used to measure walking potential. They found that significant relationships exist between mixed-use zoning ordinances and the percentage of daily-use activities within zones (walking potential). They identified the association between zoning and built environment, rather than the association between zoning and active living outcomes.

A research team led by Chriqui at the University of Illinois at Chicago (UIC) Institute for Health Research and Policy (IHRP) conducted a series of studies on the association between zoning and active living outcomes. They constructed a comprehensive nationwide data set of zoning ordinances. Based on content analysis (Norton 2008), they collected and evaluated zoning codes at the municipal and county levels to measure the elements in zoning codes that positively correlate with active living outcomes. The data cover 4,388 municipalities and unincorporated areas in the 500 most populous U.S. counties (including four consolidated cities) of the 50 U.S. states and represent 72.5% of the U.S. population. The variables they constructed include all the elements of built environment that are suggested in the research literature to have associations with active living outcomes. Variables in the data include code reform, sidewalks, crosswalks, bike-pedestrian connectivity, street connectivity, bike lanes, bike parking, bike-pedestrian trails-paths, mixed use, other walkability, active recreation, passive recreation, pedestrian-oriented zoning, and transit-oriented zoning. They were the first to propose the concept of active living-oriented zoning, which very well describes the zoning data they constructed. The data of active living-oriented zoning summarize the research achievement in the massive amount of literature on the association between built environment and active living outcomes and provide a starting point for future research on the associations among built environment, zoning, and active living outcomes.

Chriqui, Nicholson, et al. (2016) constructed individual physical activity outcome variables (biking, vigorous biking, running/jogging, vigorous running/jogging, and walking) and individual characteristics using the 2011 Behavioral Risk Factor Surveillance System (BRFSS). They merged the individual-level data with population-weighted county-aggregated zoning data. The zoning data they used in this study were constructed for a subset of counties in their data collection framework. They controlled for county-level characteristics (from the American Community Survey (ACS) 2007-2011 5-year estimates) and county-level walkability in their regressions. The variable of walkability partially accounts “for the on-the-ground built environment that individuals within a given county were exposed to” and was constructed using GIS data from NAVTEQ 2011 and ACS 2011 (1-year estimates). They found that the odds of biking and vigorous biking were positively associated with code reform zoning and zoning for bike parking, bike-pedestrian trails/paths, mixed use, active recreation, and passive recreation. The odds of running/jogging were found to be positively associated with zoning for sidewalks, bike lanes, mixed use, and passive recreation. The odds of walking were positively associated with zoning for bike lanes, bike parking, mixed use, and active recreation.

Leider, Chriqui, and Thrun (2016) constructed the variable of individual no leisure-time physical activity and other individual characteristics using 2012 BRFSS. They merged the individual-level data with population-weighted county-aggregated zoning variables for all counties in their data collection framework, which are the same zoning data utilized in this dissertation. They showed that, except zoning for crosswalks, all of the zoning measures were significantly associated with a reduced probability of no leisure-time physical activity.

Chriqui, Leider, et al. (2016a) constructed municipal-level active transportation to work outcomes (transportation to work by walking, biking, or public transportation) and municipal characteristics using ACS 2010-2014 5-year estimates. They merged the data with municipal-level zoning data for all municipalities in their data collection framework, and municipal-level walkability data (NAVTEQ 2013). They found that walking to work was significantly and positively associated with zoning for bike parking, bike-pedestrian trails/paths, other walkability, and mixed use. Biking to work was found to be significantly and positively associated with code reform zoning, zoning for sidewalks, street connectivity, bike lanes, bike parking, bike-pedestrian trails/paths, other walkability, and mixed use.

Thrun, Leider, and Chriqui (2016) used the same data as Chriqui, Leider, et al. (2016a) and found transit-oriented development zoning, a type of code reform, to be significantly and positively associated with taking public transportation to work and active transportation to work (defined as any form of active transportation, including walking, biking, or public transportation use).

2.2 Association between Built Environment and Active Living

In the empirical literature, researchers found positive association between pedestrian- and cyclist-friendly built environment and physical activity (physical activity mainly refers to walking and bicycling). The pedestrian- and cyclist-friendly built environment elements include mixed uses of residential and commercial areas (Cervero and Duncan 2003; Saelens et al. 2003; Frank et al. 2005; Frank et al. 2006; Rundle et al. 2007; Li et al. 2008; Van Dyck et al. 2010; Ewing et al. 2015), residential (household) density (Saelens et al. 2003; Frank et al. 2005; Li et al. 2005; Frank et al. 2006; Van Dyck et al. 2010), regional compactness development (Rundle et al. 2007; Aytur et al. 2008; Ewing et al. 2015), street connectivity (Saelens et al. 2003; Frank et al. 2006; Li et al. 2008; Van Dyck et al. 2010), intersection density (Frank et al. 2005; Li et al. 2005; Ewing et al. 2015), the presence of green and open spaces for recreation (Aytur et al. 2008; Li et al. 2005), the density of public transit stations (Rundle et al. 2007; Li et al. 2008; Ewing et al. 2015), the presence of sidewalks (Davison and Lawson 2006), density of places of employment (Li et al. 2005), block size, the presence of gridiron streets (Cervero and Duncan 2003), and enjoyability of the scenery (Brownson et al. 2001).

Only four papers (Handy, Cao, and Mokhtarian 2005; Handy, Cao, and Mokhtarian 2006; Frank et al. 2007; McCormack and Shiell 2011) address the problem of how unobserved preferences affects the causal effects of built environment on active living outcomes. Below, I introduce these four papers. The authors of the first three studies (Handy, Cao, and Mokhtarian 2005; Handy, Cao, and Mokhtarian 2006; Frank et al. 2007) collected and measured preferences using questionnaires in their surveys and controlled for preferences in their analyses. The last one (McCormack and Shiell 2011) has a difference-in-difference design.

Handy, Cao, and Mokhtarian (2005) attempted to answer the questions of whether neighborhood design influenced people’s travel behavior and travel preferences influenced the choice of what neighborhoods to live in. Their data came from a survey conducted in Northern California and include information about respondents' traveling preferences. The authors found that preferences mostly explained differences in travel behavior between traditional and suburban neighborhoods in the analysis of cross-sectional data. Nevertheless, they found a significant association between changes in travel behavior and changes in the built environment in the quasi-longitudinal analysis, even when controlling for preferences.

Utilizing the same data, Handy, Cao, and Mokhtarian (2006) incorporated travel preferences and neighborhood preferences into the analysis of walking behavior. Both cross-sectional analysis and quasi-longitudinal analysis showed that the built environment had an impact on walking behavior, even after accounting for preferences.

Frank et al. (2007) thought that most studies on the associations between neighborhood design and active and sedentary forms of travel failed "to account for either underlying neighborhood selection factors (reasons for choosing a neighborhood) or preferences (neighborhoods that are preferred) that impact neighborhood choice and behavior." They utilized the information about individual preferences from the travel survey of the project of Strategies for Metropolitan Atlanta's Regional Transportation and Air Quality (SMARTRAQ). They found "factors influencing neighborhood selection and individual preferences, and neighborhood walkability explained vehicle travel distance after controlling for demographic variables." They concluded that "creating walkable environments [might] result in more physical activities and less driving and in lower obesity prevalence for those preferring walkability" (Frank et al. 2007).

MacDonald et al. (2010) assessed the effects of using light rail transit (LRT) systems on BMI, obesity, and weekly recommended physical activity (RPA) levels. To overcome selection bias, they conducted a difference-in-difference research design. They collected data about individuals before (July 2006–February 2007) and after (March 2008–July 2008) the completion of an LRT system in Charlotte, NC. They compared the subjects' BMIs, obesity levels, and physical activity levels pre- and post-LRT construction. They use a propensity score weighting approach to adjust for differences in baseline characteristics among LRT and non-LRT users. They found that the use of LRT to commute to work decreased subjects' BMIs and reduced their odds of becoming obese over time. They also found positive perceptions of one's neighborhood to be associated with low BMIs, reduced odds of obesity, high odds of meeting weekly RPAs through walking, and high odds of meeting RPA levels through vigorous exercise.

Except the above four studies, almost all other studies that focus on the relationship between built environment and physical activity or inactivity, do not take unobserved preferences into account. They are summarized as follows.

Brownson et al. (2001) examined descriptive patterns in perceived environmental and policy determinants of physical activity and the associations between these factors and physical activity. They found neighborhood characteristics, including the presence of sidewalks, enjoyable scenery, heavy traffic, and hills, to be positively associated with physical activity.

Berrigan and Troiano (2002) explored the association between home age and walking behavior among U.S. adults using data from the Third National Health and Nutrition Examination Survey. They found that adults who lived in homes built before 1946 and from 1946 to 1973 were significantly more likely to walk more than one mile and more than 20 times per month than those who lived in homes built after 1973. The association was present among people living in urban and suburban counties, but absent among those living in rural counties.

Cervero and Duncan (2003) used household activity data from the San Francisco region to study the associations between urban environments and nonmotorized travel. They found that, when controlling for the steep terrain that gauge impediments to walking and bicycling, built environment had much weaker influences on walking and bicycling than other control variables.

Ewing et al. (2003) investigated the relationship between urban sprawl and physical activity, obesity, and morbidity. They employed hierarchical models on the data of 1998, 1999, and 2000 BRFSS to establish the association between individual outcomes (physical activity, obesity, BMI, hypertension, diabetes, and coronary heart disease) and the features of individuals and places. They found residents of sprawling counties to be likely to walk less during leisure time, weigh more, and have a greater prevalence of hypertension than residents of compact counties. At the metropolitan level, sprawl was similarly associated with minutes walked, but not with the other variables.

Saelens et al. (2003) utilized survey data to assess the association between neighborhood environment and physical activity. They found "residents of high-walkability neighborhoods reported higher residential density, land use mix, street connectivity, aesthetics, and safety, [and reported] more than 70 more minutes of physical activity and had lower obesity prevalence (adjusted for individual demographics) than did residents of low-walkability neighborhoods" (Saelens et al. 2003).

Frank, Andresen, and Schmid (2004) investigated how obesity relates to community design, physical activity, and time spent in cars. They linked individual data from a travel survey of 10,878 participants in the Atlanta, Georgia region between 2000 and 2002 to some objective measures of land use mix, net residential density, and street connectivity that were developed within a 1-kilometer network distance of each participant's place of residence, and conducted a cross-sectional analysis. They found land-use mix to be negatively correlated with obesity and hours spent in a car per day to be positively correlated with obesity. Moreover, these relationships were stronger among White people than Black people.

Frank et al. (2005) assessed how objectively-measured levels of physical activity were related to objectively-measured aspects of the physical environment around each participant's home while controlling for sociodemographic covariates. The data came from a survey conducted in the 13-county metropolitan Atlanta region between 2001 and 2003 (SMARTRAQ). Frank et al. found that measures of land-use mix, residential density, and intersection density were positively related to the number of minutes of moderate physical activity per day in which one engaged.

Li et al. (2005) linked the data about respondents (average age of 74) from a survey conducted in 56 city-defined neighborhoods in Portland, Oregon to the data from geographical information systems. They employed a cross-sectional hierarchical linear model with neighborhoods as the primary sampling unit and senior residents as the secondary unit. They found walking activity to be positively associated with built environment factors, including the density of places of employment, household density, the presence of green and open spaces for recreation, and the number of street intersections at the neighborhood level, and positively associated with perceptions about the safety of walking and the number of nearby recreational facilities at the residential level.

Gordon-Larsen et al. (2006) investigated the impacts of disparity in access to recreational facilities on physical activities and overweight patterns in adolescents. They linked the data of adolescents in wave I (1994–1995) of the National Longitudinal Study of Adolescent Health to the data of physical activity facilities measured by national databases and satellite data. They found low socioeconomic status and high-minority block groups to have reduced access to facilities, which in turn, was associated with decreased physical activity and increased numbers of overweight. They concluded that inequality in the availability of physical activity facilities might contribute to ethnic and socioeconomic status disparities in physical activities and overweight patterns.

Frank et al. (2006) evaluated the association between a single index of walkability that incorporated land-use mix, street connectivity, net residential density, and retail floor area ratios, with health-related outcomes in King County, Washington. They identified a positive association between walkability and the time spent in physically active travel and a negative association between walkability and BMI.

Rodríguez, Khattak, and Evenson (2006) collected data on the Chapel Hill–Carrboro area of central North Carolina from March through May 2003. They compared various measures of the physical activity of residents of a new urbanist neighborhood to those of a group of conventional suburban neighborhoods in central North Carolina. "The new urbanist neighborhood was a greenfield development built in the late 1990s and early 2000s, and [featured] small lot sizes, office and commercial space within walking distance of most residences, a variety of residential options (single-family homes, townhomes, and condominiums), amenities for pedestrians and bicyclists, reduced building setbacks, and rear alleyways for garages and services like garbage collection and mail delivery" (Rodríguez, Khattak, and Evenson 2006). They found no statistically significant differences.

Berke et al. (2007) examined whether old people that lived in areas that were conducive to walking were more active or less obese than those living in areas where walking was relatively difficult. They linked the data from the Adult Changes in Thought cohort study for a cross-sectional analysis of 936 participants aged 65 to 97 years to the walkability score data (the probability of walking in King County, Washington) from the Walkable and Bikable Communities Project. They found that high walkability scores were associated with significantly more walking for exercise than low walkability scores. However, it was unclear whether the frequency of walking reduced the prevalence of obesity.

Rundle et al. (2007) linked the adult data from a survey conducted within the five boroughs of New York City between January 2000 and December 2002 to the built environment data and employed multilevel modeling for cross-sectional analysis. They found mixed land use, the density of bus stops and subway stops, and population density, but not intersection density, to be significantly negatively associated with BMI after controlling for individual- and neighborhood-level sociodemographic characteristics.

Aytur et al. (2008) combined data from surveys about urban containment policies, BRFSS, the U.S. Census of Population, the National Resources Inventory, and the Texas Transportation Institute Urban Mobility Study to examine whether urban containment policies and the state adoption of growth-management legislation were associated with population levels of leisure-time physical activity (LTPA), and walking or bicycling to work over time. Urban containment policies attempted to manage the location, characteristics, and timing of growth to support a variety of goals, such as compact development, the preservation of green space, and the efficient use of infrastructure. They found strong urban containment policies to be associated with high population levels of LTPA and walking or bicycling to work. Residents of states that adopted growth-management legislation reported significantly more minutes of LTPA per week compared to residents of states without such policies. Weak urban containment policies had inconsistent relationships with physical activity.

Li et al. (2008) linked the survey data from residents aged 50–75 from 120 neighborhoods in Portland, OR in 2006-2007 to the GIS-derived measures of land-use mix, distribution of fast-food outlets, street connectivity, access to public transportation, and green and open spaces. They found high mixed-use land to be positively associated with walking activities and the meeting of physical activity recommendations and negatively associated with overweight/obesity levels. Neighborhoods with high street connectivity, a high density of public transit stations, and many green and open spaces to be related to varying degrees of walking and the meeting of physical activity recommendations.

Van Dyck et al. (2010) used the survey data of 1,200 adults (aged 20–65 years) in Ghent, Belgium between May 2007 and September 2008 to explore whether neighborhood walkability (residential density, land-use mix, and street connectivity) was positively associated with physical activity and whether this association was moderated by neighborhood socioeconomic status. They found that highly-walkable neighborhoods were associated with the long durations of moderate-to-vigorous physical activity, transportation walking and cycling, recreational walking, and the short duration of motorized transport. Neighborhood socioeconomic status was not a significant moderator.

Ewing et al. (2015) argued that much research had data about single regions, which along with the specifications about the different models in each study and the use of different metrics, prevented generalization. To address the problem of the lack of external validity in the literature, they combined several sources of survey data across different survey years to create the largest pooled household travel and built environment dataset[5] and defined a large number of consistent built environmental variables with which to predict five household travel outcomes (car trips,

[5] The data consisted of 664,732 trips by 62,011 households in 15 diverse regions.

walk trips, bike trips, transit trips) and vehicle miles traveled. They employed the hierarchical linear model to account for the dependence of households in the same region on shared regional characteristics and estimated two-part models to account for the excess number of zero values in the distributions of dependent variables. They found walk trip decision and walk trip frequency to be positively associated with activity density within one-quarter mile, the percentage of 4-way intersections within one-quarter mile, land-use entropy within one-half mile of home, transit accessibility to employment within 30 minutes, transit stop density within one-half mile of home, and regional compactness. In addition, walk trip decisions were found to be positively associated with intersection density within one-half mile of home and activity density. Bike trip decisions were positively associated with intersection density, the percentage of 4-way intersections within a mile of one's home, and regional compactness. Bike trip frequency increased with the proportion of 4-way intersections within a mile of one's home and transit stop density within a mile of one's home and declined with regional population.

2.3 Contribution to the Literature

Scholars from different fields (urban planning, public health, transportation, etc.) have used different data and statistical models to reveal the significant positive association between active living-oriented zoning and physical activity and the significant positive association between pedestrian- and cyclist-friendly built environment and physical activity. However, as I read the literature, I found it to lack causal identification. The lack of causal identification means that there are some inherent deficiencies in the research.

As many researchers acknowledge, the literature is short of considering the endogeneity problem. Except for a few studies on built environment and active living outcomes (Handy, Cao, and Mokhtarian 2005; Handy, Cao, and Mokhtarian 2006; Frank et al. 2007; MacDonald et al. 2010), all other research in the literature has the endogeneity problem. The studies on this topic are mainly observational studies in which the researchers use survey data rather than randomized experimental data because they cannot randomly assign people to different areas with different zoning or different built environments to compare results. Observational studies are vulnerable to the endogeneity problem. Specifically, people choose where they live. Individuals with active living preferences may choose to live in places with active living-oriented zoning and built environment. This is called "Tiebout sorting" in economics. Tiebout sorting implies that the observed data is the result of Tiebout equilibrium and market equilibrium. It makes estimates inconsistent and biases the estimated coefficient of active living-oriented zoning or built environment on active living outcomes. It is entirely plausible that some significant effects of zoning and built environment on active living outcomes found in previous research are partly, or even completely, due to people's preferences for active living. Without causal identification, researchers have no idea whether zoning and built environment have impacts on active living outcomes. Much of the literature claims to have policy implications. However, when researchers are unsure of the causal impacts of zoning, how can they make policy suggestions?

The second problem is that the causal relationships among zoning, built environment, and active living outcomes are unclear. Researchers hypothesize that zoning affects active living outcomes by shaping the built environment. However, no researcher has quantitatively verified the mediation of built environment for the effect of zoning on active living outcomes. In addition, it is popular in the related literature to examine the association between zoning/built environment and active transportation to work. However, it may not be the zoning and built environment that causally impact people's transportation methods to work. Instead, people choose their transportation methods to work when making job choices and housing choices. Transportation to work is an integral part of any person's job plan. From the point of view of labor economics, people choose job attributes (compensation, working environment, etc.), living location, and transportation methods to work simultaneously based on their qualifications, budget, and preferences. It is a very complex problem to model people's job choices because numerous unobserved confounders, including people's preferences for different job attributes, housing locations, and transportation methods, are involved in such decision-making. Without a comprehensive structural econometric model of career selection and lifestyle choice, simply putting transportation to work as an outcome variable in a regression, from the perspective of causal inference, is improper.

The third problem is that the regression models in the literature include some covariates that are correlated with the outcome variables but are endogenous (e.g., BMI, income, and employment). This is improper in causal identification. There is a reverse causality problem between BMI and physical activity. Wage and employment are choice results based on Gary Becker's time allocation model. In the time allocation model, being subject to income and time constraints, people choose leisure time and consumption simultaneously to maximize the total utility. Working hours and income are endogenous variables in the time allocation model and will be determined together with leisure time and consumption by first order conditions. Reduced form regressions derived from the time allocation model are supposed to have all endogenous variables as outcomes.

This is the first study that disentangles the causal relationships among zoning, built environment, and active living outcomes. I use the zoning data from the work of Chriqui et al. However, I have a completely different research design from theirs. To identify and estimate the causal effects, I employ instrumental variable identification and mediational analysis. I explore the history and institution of American zoning to search for valid instrumental variables and implement instrumental variable identification and estimation, a signature technique in econometrics, to test endogeneity and estimate the causal effects of active living-oriented zoning on physical activity and sedentary behavior. I employ mediational analysis to quantitatively investigate the mediation of built environment in the causal path from zoning to active living outcomes.

My findings are subject to some limitations. First, because the instrumental variable technique only makes LATE in linear models, the results lack external validity. The treatment effects I estimate is for compliers, a subpopulation whose behaviors change as the values of instrumental variables change. The estimates are not consistent for average treatment effects, so they cannot be generalized to the entire population. Second, I do not aim to examine the causal effects of zoning on traveling behaviors. The instrumental variable identification strategy employed in this study cannot be applied to investigate the causal effects of zoning on traveling behaviors. I only estimate the effects of active living-oriented zoning on physical activity (walking, running, and bicycling) and sedentary behavior (TV watching, music listening, etc.). To investigate the causal effects of zoning on traveling behaviors, researchers need highly complicated structural models that can integrate various location choices into the analysis. Third, I do not examine the effects of code reform or any single zoning provisions; rather, I examine the causal effects of active living-oriented zoning that consists of code reform and other zoning provisions. Different zoning provisions may have different causal effects on active living. The metric of active living-oriented zoning treats all zoning provisions as equal and assumes they are equally likely to affect active living. Future researchers can explore the causal effects of each single zoning provision.

3 DATA AND VARIABLE CONSTRUCTION

The data used in this study come from five sources. Zoning data are used to construct the independent variable—active living-oriented zoning. NAVTEQ 2011 is used to build the independent variable—walkability. The *American Time Use Survey* 2010-2015 is used to create the outcome variables of time usage and individual-level control variables. The *American Community Survey* 2011–2015 5-year estimates are used to construct the county-level control variables. The 1900 census is used to create the instrumental variables.

3.1 Zoning Data

The zoning data are the same as that used in the research of Leider, Chriqui, and Thrun (2016). As discussed in the literature review, this data were contributed to the research field by a research team led by Chriqui at UIC IHRP. This team designed and implemented the whole procedure of collecting and evaluating zoning codes to construct active living-oriented zoning variables. The exposition below mainly draws on the studies of Chriqui, Nicholson, et al. (2016) and Leider, Chriqui, and Thrun (2016).

Zoning codes that were effective as of January 2010 were collected through internet research with telephone and email verification to confirm the adoption of the codes. The sample frame was based on the most populous 496 counties and four consolidated cities in the U.S., which contained 74.3% of the U.S. population based on 2010 census population estimates. Because zoning primarily happens at the municipal level, and county zoning typically covers unincorporated areas, the sample frame included all municipalities and unincorporated areas in these counties/consolidated cities. Due to resource limitations, the sample was restricted to areas that represented more than 0.5% of the population of their given county/consolidated city. The sample was further reduced by 157 municipal jurisdictions for which zoning codes were not electronically available, the community refused to send a copy, or the lack of responses to follow-up calls. Thus, the zoning data cover 4,388 municipalities and unincorporated areas in the 496 counties and four consolidated cities of the 50 U.S. states and represent 72.5% of the U.S. population.

The first step is to evaluate the zoning codes of each jurisdiction. Each jurisdiction's zoning code was evaluated by trained master's level urban planners. The evaluation was regulated and guided by the coding protocol that was developed by the research team. The evaluation results were recorded using a coding tool (forms).[6] First, trained master's level urban planners assessed whether the zoning code was a code reform type of zoning code. Code reform can be classified into three categories: smartcode, full form-based code (non-smartcode), and code reform district(s)/regulations only. Code reform districts/regulations include form-based district/regulation, transect-based district/regulation, new urbanist district/regulation, pedestrian-oriented district/regulation, transit-oriented district/regulation, traditional neighborhood development district/regulation, and other code reform elements. Then the trained master's level urban planners classified each district mentioned in the code into eight mutually-exclusive types of zones/districts (code reform, commercial, mixed use, park/recreation/open space, planned unit development, public/civic/government, residential, and general zoning). For each district, they evaluated the code to determine whether any of the following 11 active living-oriented zoning

[6] For coding forms, please see the online attachment of Chriqui, Nicholson, et al. (2016). For coding protocol, please contact them.

provisions were addressed:[7] sidewalks, crosswalks, bike/pedestrian connectivity, street connectivity, bike lanes, bike parking, trails/paths, mixed use, other general walkability provisions (e.g., traffic calming and pedestrian measures), active recreation, and passive recreation.

The second step is to construct zoning variables for each jurisdiction based on evaluation results. A dichotomous variable, code reform, was created to reflect the evaluation result on code reform. For each provision, a dichotomous variable was created to reflect whether the provision was addressed in any one of the eight zones/districts. Therefore, for each jurisdiction, they derived 12 dichotomous variables.

1. Code reform. It indicates whether the jurisdiction had adopted any type of zoning code reform as of 2010.
2. Sidewalks. It indicates whether the topic of sidewalks was addressed in the zoning code. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of sidewalks.
3. Crosswalks. It indicates whether the topic of marked on-street crossings for pedestrians was mentioned in the zoning code. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of marked on-street crossing for pedestrians.
4. Bike-pedestrian connectivity. It indicates whether there are any sections about connecting or linking sidewalks, paths, trails, bike lanes, or other bicycle- or

[7] As mentioned in the coding protocol, sometimes communities use other phrases when referring to the eleven provisions. For example, communities may use "pedestrian paths (or pathways)" or "walkways" when referring to sidewalks. It is addressed in the protocol and was taken care of by coders. Also, the protocol provides details about how to distinguish semantically similar provisions. For example, when communities mention "bicycle routes," if it is defined by communities as marked lanes for bicycles, it should be coded as the provision of "bike lanes." Otherwise, it should be coded as the provision of "bike-pedestrian trails-paths." The coding protocol instructs coders to use contextual clues and their best judgment.

pedestrian-oriented amenities in the zoning code. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of bike-pedestrian connectivity.

5. Street connectivity. It indicates whether there is any topic about a connective street network with high intersection density and many opportunities to access other uses and developments in the zoning code. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of street connectivity.
6. Bike lanes. It indicates whether there is any topic about designated, marked lanes for bicycles that separate bikes from traffic to allow cyclists safely traveling along the same routes as vehicles. The lanes include cycle tracks (protected bike lanes), sharrows (road lanes with shared lane markings), and bicycle boulevards (designed so that bikes and cars share the road equally and include pavement markings). It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of bike lanes.
7. Bike parking. It indicates whether there is any topic about infrastructure dedicated to providing bike parking opportunities in the zoning code. The infrastructure includes bike racks, bike storage, etc. It is a proxy for street furniture, including benches, streetlights, etc. Zoning codes typically do not have any information about benches and streetlights, but they do have information about bike parking. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of bike parking infrastructure and street furniture.

8. Bike-pedestrian trails-paths. It indicates whether there is any topic about bicycle- or pedestrian-oriented trails or paths, including trails, hiking trails, paths, multi-use paths, pedestrian walking paths, off-road paths, bicycle routes, greenways, walking trails, walking paths, and sidepaths, in the zoning code. It captures whether the jurisdiction had ever put or was putting efforts into promoting the development of bike-pedestrian trails-paths.
9. Other walkability. It indicates whether there are walkability-related topics that were not picked up by any of the other topics in the zoning code. Typically, the topic includes pedestrian safety markers. Specifically, it includes pedestrian safety markers related to vehicular traffic as long as they do not address internal circulation. It also includes pedestrian overpasses or underpasses, boardwalks, riverwalks, parks within walking distance, and pedestrian-oriented streetscape. However, it does not include pedestrian safety markers related to signs, awnings, sidewalk furniture, lighting, landscaping, or skywalks. It is a supplement for other walkability-related zoning measures. It catches the efforts the jurisdiction had put or was putting into promoting the development of walkable communities that other walkability-related zoning measures do not catch.
10. Mixed use. It indicates whether there is any topic about mixed-use buildings within small pockets of the community, including vertical mixed-use buildings and horizontal mixed-use buildings in small areas to encourage walking in the zoning code. It captures whether the jurisdiction had put or was putting efforts into promoting the development of the mixed use of commercial and residential elements.

11. Active recreation. It indicates whether any equipment to support physical activity was ever been mentioned in the zoning code. It includes the markers of active recreation listed in Table 3.1.1. It does not indicate any marker listed in Table 3.1.2. It captures whether the jurisdiction had put or was putting effort into promoting the development of active recreation provision.
12. Passive recreation. It indicates whether there is any topic about any markers of passive recreation that are listed in Table 3.1.3 in the zoning code. Passive recreation markers are different from active recreation markers in that they provide opportunities (e.g., open space) rather than equipment for physical activity. It does not indicate any marker that is listed in Table 3.1.4. It captures whether the jurisdiction had put or was putting effort into promoting the development of passive recreation provisions.

Table 3.1.1 Active Recreation Markers

Playgrounds
Recreation structures
Golf courses or driving ranges, provided they are public (not private country clubs)
Children’s play area
Open space (only if it describes a place where people can be physically active, such as “open space provided for athletic fields”)
Some parks can be counted as active if they list specific provisions, such as “a park with running trails” or “parks with sports amenities.”
Plazas/squares/greens that specify active recreational uses
Exercise or exercise words that use active verbs, such as “jogging” or “running”
Fitness or physical activity
Phrases incorporating the word “recreation” (e.g., recreational facilities, recreational fitness trails, recreational opportunities, outdoor recreation)
Ballfields, playfields, etc.
Sports leagues, sports fields, etc.
Young Men's Christian Association (YMCA)
Athletic facilities, such as swimming pools, basketball courts, tennis courts, and gymnasiums

Note: This table is taken from NCI Code Reform Project Policy Coding Protocol, Board of Trustees, University of Illinois at Chicago, 2014

Table 3.1.2 Active Recreation Markers Not Included

Private accessory uses such as swimming pools, trampolines, basketball goals, etc
Commercial recreation uses
Privately owned, fee-based gyms (e.g., karate studio, private ballet studio)
Carnivals, amusement parks, etc.
Batting cages, go-cart tracks, mini golf, etc.
Temporary (or certain seasonal) recreational uses
Outdoor recreation only devoted to hunting or fishing
Shooting ranges
Recreation related to motorized vehicles (e.g., ATVs, dirt bikes, motor boats, snowmobiles, RVs)

Note: This table is taken from NCI Code Reform Project Policy Coding Protocol, Board of Trustees, University of Illinois at Chicago, 2014

Table 3.1.3 Passive Recreation Markers

Open space (not designed for active recreation, e.g., playgrounds)
Parks (not designed for active recreation, e.g., a park with playgrounds)
Greenways that provides open space or other types of passive recreational uses
Plazas/squares/greens that specify passive recreational uses
Nature preserve (e.g., where visitors can observe wildlife)
Forest preserve, but only if it describes places for visitors to rest, wander, etc.

Note: This table is taken from NCI Code Reform Project Policy Coding Protocol, Board of Trustees, University of Illinois at Chicago, 2014

Table 3.1.4 Passive Recreation Markers Not Included

Anything devoted solely to agriculture uses
Anything devoted solely to silvicultural uses (tree farming)

Note: This table is taken from NCI Code Reform Project Policy Coding Protocol, Board of Trustees, University of Illinois at Chicago, 2014

Among the 12 zoning variables, sidewalks, crosswalks, street connectivity and other walkability are expected to encourage walking, jogging, and running. Bike lanes and bike parking are expected to promote cycling. Bike-pedestrian connectivity, bike-pedestrian trails-paths, code reform, mixed use, active recreation, and passive recreation are supposed to encourage both pedestrian- and cyclist-activities. All of them are expected to discourage sedentary behavior at home, such as TV watching and radio/music listening.

The third step is to construct a population-weighted county-aggregated zoning variable. The construction is the weighted average: for each county, multiply all municipal jurisdictions and unincorporated areas that are identifiable in the sample frame and are within the county by the corresponding population percentage, and then make a summation. Table 3.1.5 shows an example of the weighted average procedure. The county-level variable code reform implies the proportion of the population that was exposed to the code reform in the county or the probability that a randomly-selected individual in the county was exposed to the code reform. The sidewalks variable implies the proportion of the population that was exposed to the promotion of sidewalks in the county, or the probability that a randomly-selected individual in the county was exposed to the promotion of sidewalks. The other 11 county-level zoning variables have similar explanations as that of sidewalks.

Table 3.1.5 Population-Weighted County-Aggregated Zoning Variable (Sidewalks) Construction Instruction

Jurisdiction	The Value of "Sidewalks"	Population Percentage	Population-Weighted Value
Municipality 1	1	33%	0.33
Municipality 2	0	26%	0
Municipality 3	0	15%	0
Municipality 4	1	24%	0.24
Unincorporated Area	1	2%	0.02
County A	0.59	100%	0.59

Note: County A consisted of four municipalities and one unincorporated area identified in the sample frame. Using population percentage as weights, the weighted average result of zoning variable "sidewalks" is 0.59, which is used as the value of "sidewalks" for County A.

In this thesis, I create the zoning variable of active living-oriented zoning to sum up the information of the 12 county-level zoning variables by averaging them. The variable of active living-oriented zoning means the proportion of the population who were exposed to the promotion of active living-oriented zoning, or the probability that a randomly-selected individual in the county was exposed to the promotion of active living-oriented zoning. If this variable equals zero, then no one in the county was exposed to any of the 12 zoning measures because all 12 zoning measures in all jurisdictions within the county equal zero. If this variable equals one, then all the people in the county were exposed to all 12 zoning measures because all 12 of the zoning measures in all jurisdictions within the county equal one.

3.2 NAVTEQ 2011

NAVTEQ 2011 third quarter GIS data were used in combination with other county characteristics that were extracted from ACS 2011 1-year estimates to construct the variable of walkability. The variable of walkability is the same as that used in the studies of Chriqui, Nicholson, et al. (2016) and Leider, Chriqui, and Thrun (2016). The variable of walkability consists of four density metrics: the proportion of four-way intersections to all intersections in the county (NAVTEQ), the total number of intersections in the county (NAVTEQ) divided by the county land area (ACS), housing unit density (ACS), and population density (ACS). To construct the variable of walkability, firstly, summarize the four density metrics. Then standardize the summation by subtracting the mean and then dividing it by the standard deviation. Finally, add a scalar of 1 to the standardized summation to reduce negative values. Walkability is based on the scale created by Slater et al. (2010), which was adapted from the scale created and updated by Ewing and Hamidi (2014) (Chriqui, Nicholson, et al. 2016).

The walkability metric can only be used to partially assess the built environment. The zoning variable, active living-oriented zoning, includes many more elements of built environment than walkability, active recreation (equipment that supports physical activity), passive recreation (opportunities to engage in physical activity, like open space), mixed-use development, bike parking, and bike lanes, to name several. Many of these built environment elements cannot be captured by the walkability metric. To address this point, in the research design, I model the walkability as a partially mediator for the zoning.

3.3 American Time Use Survey (ATUS)

ATUS data for the survey years of 2010-2015 were used to construct dependent variables and individual-level controls. Administered by Census Bureau, ATUS is the first ongoing survey on time use in the U.S. starting from 2003. ATUS participants are randomly selected from households that are completing their participation in the *Current Population Survey* (CPS). They are contacted via phone between two and five months after the last CPS interview for the

ATUS household. Each respondent provides detailed information about his or her activities during a designated 24-hour period that began at 4 a.m. on the designated day and continues through 3:59 a.m. the following day. The designated days are distributed across the days of the week, with 10 percent allocated to each of the weekdays Monday through Friday, 25 percent to Saturdays and 25 percent to Sundays, and they are distributed evenly across the weeks of the year. The number of usable time diaries was about 14,000, or about 1,150 per month, in 2004 and has remained around that level since then (Hofferth, Flood, and Sobek 2015; Hamermesh, Frazis, and Stewart 2005).

Because ATUS provides nationally-representative estimates of Americans' time diaries, covers the full range of nonmarket activities, from childcare to volunteering, and is the only federal survey of time usage, it is widely used by professionals, researchers, and leaders from public and private organizations. The institutions that provide public services and utilize ATUS data include the Bureau of Economic Analysis, the Bureau of Labor Statistics, the Bureau of Transportation Statistics, the Department of Agriculture, United Nations Development Program, etc.[8] An increasing number of economists employ ATUS to conduct studies on time allocation, labor supply, and household production (Hamermesh 2005; Kimmel and Connelly 2006; Meyer and Sullivan 2006; Hamermesh and Donald 2007; Connelly and Kimmel 2007; Kalenkoski, Ribar, and Stratton 2007; Hamermesh and Lee 2007; Zick, Bryant, and Srisukhumbowornchai 2008; Hamermesh 2007; Aguiar and Hurst 2007; Connolly 2008; Hamermesh 2008). In addition, an increasing number of researchers employ ATUS to investigate physical activity, sedentary behavior, and obesity (Zick et al. 2007; Tudor-Locke and Ham 2008; Ham, Kruger, and Tudor-Locke 2009; Dunton et al. 2009; Mullahy and Robert 2010; Sener, Bhat, and

[8] Please see https://www.bls.gov/tus/overview.htm.

Pendyala 2011; Ng and Popkin 2012; Kalenkoski and Hamrick 2013; Colman and Dave 2013; Tudor-Locke et al. 2014; Smith, Ng, and Popkin 2014).

ATUS uses a 3-tier, 6-digit system to code respondents' activities. Table 3.3.1 shows the first-tier activity group of the coding structure. Each first-tier activity group is further disaggregated into second-tier activity groups (more than 100 total) and third-tier activity groups (more than 400 total). For each activity, ATUS records start and end times, the duration of the activity in minutes, a location code, who else was present, and so on. The demographic information about respondents comes from the final CPS interview. ATUS only updates some labor market information, including labor force status, usual hours of work, class of worker, industry, occupation, earnings, and school enrollment (Hofferth, Flood, and Sobek 2015; Hamermesh, Frazis, and Stewart 2005).

Table 3.3.1 ATUS Coding Structure: First-Tier Activity Groups

01 Personal care
02 Household activities
03 Caring for and helping household members
04 Caring for and helping nonhousehold members
05 Working and work-related activities
06 Education
07 Consumer purchases
08 Professional and personal care services
09 Household services
10 Government services and civic obligations
11 Eating and drinking
12 Socializing, relaxing and leisure
13 Sports, exercise, and recreation
14 Religious and spiritual activities
15 Volunteer activities
16 Telephone calls
17 Traveling

Note: From IPUMS American Time Use Survey Data Extract Builder, http://www.atusdata.org (Hofferth, Flood, and Sobek 2015).

I use the duration of activities in minutes (coded based on respondents' answers to the survey) to construct the dependent variables. Table 3.3.2 shows the correspondence between the ATUS 3-tier activity codes and the dependent variables.

The physical activity variable is constructed by the summation of the duration in minutes of "04 biking," "24 running," and "31 walking" within the activity groups "01 participating in sports, exercise, or recreation" (second-tier) and "13 sports, exercise, and recreation" (first-tier). It measures the number of minutes per day the respondent spent on physical activity. Table 3.3.3 shows physical activity examples. The inclusion of biking, running, and walking as physical activities is based on the literature that focuses on the associations between zoning/built environment and the activities of walking and cycling.

The variable of sedentary behavior is constructed by the summation of the duration in minutes of "01 relaxing, thinking," "03 television and movies (not religious)," "05 listening to the radio," and "06 listening to or playing music" that took place at "respondent's home or yard (0101)" or "someone else's home (0103)." These activities are within the activity groups of "03 relaxing and leisure" (second-tier) and "12 socializing, relaxing, and leisure, sedentary behavior" (first-tier). The variable of sedentary behavior measures the number of minutes the respondent spends on sedentary behavior per day.

Table 3.3.4 shows activity examples for sedentary behavior. The inclusion of these categories in sedentary behavior is based on the concept of leisure-time sedentary behavior, which is also known as no leisure-time physical activity and includes light or no physical activity.

The summation helps us avoid unnecessary modeling. Conditional on people choose physical activity (sedentary behavior), there is a choice problem about which type of physical activity (sedentary behavior) to engage in. This choice is noise to my empirical modeling because I do not want to investigate people's preferences for one physical activity (sedentary behavior) over another physical activity (sedentary behavior).

Table 3.3.2 Dependent Variable Construction

ATUS 3-Tier Activity Code	Dependent Variable
13 Sports, Exercise, and Recreation	
01 Participating in Sports, Exercise, or Recreation	
04 Biking	Physical Activity
24 Running	
31 Walking	
12 Socializing, Relaxing, and Leisure	
03 Relaxing and Leisure	
01 Relaxing, thinking	Sedentary Behavior
03 Television and movies (not religious)	
05 Listening to the radio	
06 Listening to or playing music	

Note: The sedentary behavior activities were restricted by the locations of "respondent's home or yard (0101)" or "someone else's home (0103)."

Table 3.3.3 Activity Examples for Physical Activity 2010 - 2015

Activity Code	Activity Examples
130104 Biking	Registering for a bike race
	Jogging
	Running a marathon
130124 Running	Running a race/organized run (2005+)
	Running cross country (2005+)
	Talking to race officials

Note: From IPUMS American Time Use Survey Data Extract Builder, http://www.atusdata.org (Hofferth, Flood, and Sobek 2015).

Table 3.3.4 Activity Examples for Sedentary Behavior 2010 - 2015

Activity Code	**Activity Examples**
120301 Relaxing, Thinking	Breaks at work, unspecified activity Daydreaming Doing nothing/goofing off/wasting time Fantasizing Grieving Hanging around/hanging out (alone) (2004+) Lying around (2007) Reflecting Resting/relaxing/lounging Sitting around Sitting in the hot tub/Jacuzzi/whirlpool (2004+) Sitting in the sauna (2004+) Sunbathing Watching husband assemble lawnmower Watching husband cook dinner Watching wife garden Wondering Worrying/crying
120303 Television and Movies (Not Religious)	Borrowing movies from the library Returning movies to library Setting TiVo/DVR (2011+) Setting the VCR or DVD player Watching a DVD/video/instructional video Watching home movies/home videos Watching TV Watching TV/DVDs on computer (personal interest) (2011+) Watching videos on YouTube (2011+)
120305 Listening to the Radio	Listening to a radio talk show Listening to music on the radio Listening to public radio Listening to the top ten on the radio Listening to podcast (2015+)
120306 Listening to or Playing Music (Not Radio)	Christmas caroling (2004+) Composing music (2005+) Listening to recorded music Listening to records/CDs/DVDs/tapes Listening to someone play the piano Playing musical instrument (leisure) Singing (2004+) Singing karaoke (2004+) Tuning musical instruments

Note: From IPUMS American Time Use Survey Data Extract Builder, http://www.atusdata.org (Hofferth, Flood, and Sobek 2015).

Individual-level controls were constructed using ATUS. They include age, gender, race, marital status, difficulty, and education. Race was classified as non-Hispanic White (reference), non-Hispanic Black, non-Hispanic other, and Hispanic. Marital status was categorized as married (reference), widow/divorced/separated, and single. Any difficulty is a dichotomous variable that indicates whether the respondent had any physical or cognitive difficulty, as measured by an affirmative response to at least one of the CPS's six cognitive difficulties (Hofferth, Flood, and Sobek 2015), which include the following: (1) difficulty in taking care of their own personal needs; (2) blindness or serious difficulty in seeing, even with corrective lenses; (3) deafness or serious difficulty in hearing; (4) difficulty in performing basic activities outside the home alone (excluding temporary health problems, such as broken bones or pregnancies); (5) serious difficulty in walking or climbing stairs; and (6) cognitive difficulties (such as difficulty remembering, concentrating, or making decisions).[9] Education indicates a person's the highest completed level of education, which was classified as less than high school diploma, high school graduate (no college), some college (no degree), college graduate (associate's or bachelor's degree) (reference), and graduate degree (master, professional degrees, doctor). In addition, I restricted the ATUS sample to non-holiday observations. People may have significantly different activities during holidays from usual.

ATUS does not have county identifier so it cannot be directly linked to the county-level zoning data. Because participants in ATUS were selected randomly from households that were

[9] Please see https://www.atusdata.org/atus-action/variables/DIFFANY#description_section.

exiting from participation in CPS, I link ATUS to CPS where county identifiers are available for counties that have a population greater than 100,000[10] to get county identifiers for ATUS[11].

3.4 American Community Survey (ACS)

ACS 2011–2015 5-year estimates were used to construct county-level controls. The county-level controls include the percentage of households in poverty, the percentage of non-Hispanic White people, the percentage of non-Hispanic Black people, the percentage of Hispanic people, median household income, and county land area (square miles). ACS 5-year estimates are more precise than the 1- and 3-year estimates. They were also used to construct the population-weighted county-aggregated zoning metrics because the 5-year estimates capture all jurisdictions nationwide, among which some small jurisdictions are not included in the 1- and 3-year estimates (Chriqui, Nicholson, et al. 2016; U.S. Census Bureau 2008).

3.5 1900 Census: Population, Housing, Agriculture, and Manufacturing Data

I extracted three county-level variables, total number of manufacturing establishments, total number of acres of land in farms, and total land surface area in square miles, from *1900 Census: Population, Housing, Agriculture & Manufacturing Data*, the electronic version of which is available from *National Historical Geographic Information System: Version 11.0*

---

[10] In some cases, CPS also suppresses the identifier for a county with more than 100,000 residents. For instance, if there is 2-county metro area and the adjacent metro county has less than 100,000 residents, the county identifiers of the two counties are suppressed.
[11] U.S. Bureau of Labor Statistics provides detailed instruction on how to link ATUS to CPS: https://www.atusdata.org/atus/resources/linked_docs/Linking_ATUS_and_CPS_files_2-2008.pdf. Also, because ATUS data here were extracted from IPUMS (ATUS-X), we can utilize the variable "CPSIDP," an IPUMS-created identifier for linking across IPUMS CPS and ATUS-X databases

(Minnesota Population Center 2016), to construct two instrumental variables. The instrumental variable manufacturing establishment density was defined as the ratio of the total number of manufacturing establishments to the total number of acres of land on farms. The instrumental variable farmland proportion was defined as the ratio of total number of square miles of land on farms, which was converted from total number of acres of land on farms, to total land surface area in square miles.

3.6 Summary Statistics

Table 3.6.1 shows the characteristics of counties in the zoning data (merged with NAVTEQ 2011, and ACS 2011-2015 5-year estimates to get county characteristics). The data cover 4,388 municipalities and unincorporated areas in 496 counties of 50 states and represents 72.5% of the U.S. population. The average number of jurisdictions within a county was 8.85, with a minimum of 1 and a maximum of 55. On average, a county spanned 1,086 square miles with a population density of 1,209 people per square mile. On average, 13.5 percent of households in the counties were in poverty. About 67.1 percent of the county populations were made up of non-Hispanic White people. Only 11.9 percent were non-Hispanic Black people, and 13.7 percent were Hispanic people. The median county-level household income was $57,600, and the median resident age was 37.8. About 39 percent of the sampled counties were located in the South, 18.3 percent were located in the West, 22.4 percent were located in the Midwest, and 20.2 percent were located in the Northeast. The average county-level walkability scale was 1 (with a maximum of 18). On average, 24.1 percent of the populations of the sampled counties were exposed to code reform zoning. The active living-oriented zoning requirements, from the most prevalent to the least frequent, were for active recreation (81.8%), passive recreation (81.7%), sidewalks (70%), other walkability (e.g., pedestrian orientation and traffic calming) (68.2%), mixed-use development (65.4%), bike-pedestrian trails-paths (57.6%), bike-pedestrian connectivity (42.8%), street connectivity (40.7%), bike parking (40.3%), crosswalks (25.5%), and bike lanes (16.5%).

Table 3.6.2 shows the characteristics of counties in the final analytical sample, where all five data sources were merged. The data cover 2,453 municipalities and unincorporated areas in 251 counties of 37 U.S. states and represents 39.08% of the U.S. population. The average number of jurisdictions within a county was 9.77, with a minimum of 1 and a maximum of 55. On average, a county spanned 1,019 square miles with a population density of 1,543 people per square mile. On average, 13.5 percent of households in the counties were in poverty. About 66.3 percent of the county populations were non-Hispanic White people. Only 11.8 percent were non-Hispanic Black people, and 14.7 percent of the county populations were Hispanic people. The median county-level household income was $58,566 and the median resident age was 37.6. Among the sampled counties, about 38 percent were located in the South, 16.3 percent in the West, 23.9 percent in the Midwest, and 21.5 percent in Northeast. The average county-level walkability scale was 1.08 (with a maximum of 18). On average, 24.3 percent of the populations of the sampled counties were exposed to code reform zoning. The active living-oriented zoning requirements, from the most prevalent to the least prevalent, were active recreation (81.1%), passive recreation (80.8%), sidewalks (69.1%), other walkability (e.g., pedestrian orientation and traffic calming) (66.6%), mixed-use development (62.9%), bike-pedestrian trails-paths (56%), bike parking (39.9%), bike-pedestrian connectivity (39.7%), street connectivity (36.4%), crosswalks (25.2%), and bike lanes (15.2%).

Table 3.6.3 shows the characteristics of counties that are included in the zoning data but excluded from the final analysis. The data cover 1,935 municipalities and unincorporated areas in the 245 counties of 46 U.S. states and represents 33.44% of the U.S. population. The average number of jurisdictions within a county was 7.9, with a minimum of 1 and a maximum of 31. On average, a county spanned 1,156 square miles with a population density of 866 people per square mile. On average, 13.5 percent of households in the counties were in poverty. About 68 percent of the county populations were made up of non-Hispanic White people. Only 12 percent were made up of non-Hispanic Black people, and 12.6 percent were made up of Hispanic people. The median county-level household income was over $56,611, and the median resident age was 37.9 years. Among the sampled counties about 40 percent were located in the South, 20.4 percent were located in the West, 20.8 percent were located in the Midwest, and 18.8 percent were located in the Northeast. The average county-level walkability scale was 0.919 (with a maximum of 8.01). On average, 24 percent of the populations of the sampled counties were exposed to code reform zoning. The active living-oriented zoning requirements, from the most prevalent to the least frequent, were for passive recreation (82.5%), active recreation (82.4%), sidewalks (70.9%), other walkability (e.g., pedestrian orientation and traffic calming) (69.9%), mixed-use development (68%), bike-pedestrian trails-paths (59.3%), bike-pedestrian connectivity (46%), street connectivity (45.1%), bike parking (40.7%), crosswalks (25.8%), and bike lanes (17.7%).

Table 3.6.4 compares the characteristics between counties included in the final analysis and counties excluded from the final analysis. Table 3.6.5 compares the characteristics between the counties included in the final analysis and all counties before being merged with individual-level data and 1900 census data. For continuous variables, I conduct Wilcoxon rank-sum test (also known as the Mann–Whitney U test), which is robust to non-normal distributions, to test the null hypothesis that the sample in the final analysis and the sample excluded from the final analysis have equal medians. For dichotomous variables, I conducted proportional z test to test the null hypothesis that the two samples have identical means (proportions).

Table 3.6.4 shows there are no significant differences between the counties included in the final analysis and those excluded from the final analysis, except the average number of jurisdictions within each county. Counties in the final analysis have significantly more within-county jurisdictions than the counties excluded from the final analysis. The difference in the average number of jurisdictions within each county may be because ATUS were only conducted on metropolitan areas, so the merged data in the final analysis retains larger counties than the data excluded from the final analysis. However, the original sample frame of zoning data was based on the most populous 496 counties and four consolidated cities in the United States, so most of the county characteristics have insignificant differences. Table 3.6.5 shows there are no significant differences between the data about counties included in the final analysis and all counties before being merged with individual-level data and 1900 census data.

The final analytical sample is constructed by merging all five data sources. It is restricted by excluding outlier time usage, which is determined by computing inter-quartile ranges (IQRs) of the time usage of physical activity and sedentary behavior (IQR is the difference between the first and third quartiles of time usage. I calculate two IQRs, one is for physical activity, and the other is for sedentary behavior) and then conservatively using six times the IQRs as the outlier cutoffs. The outlier exclusion can drop those marathon or bike race observations that are not considered to be leisure-time physical activities.

Table 3.6.6 shows the summary statistics for the final analytical sample of physical activity. Only 8.05 percent of the respondents in the sample participated in physical activity. Conditional on participation, people spent 1 hour per day on physical activity on average. The average county-level walkability scale was 1.55 in the full sample and 1.59 in the participation sample. On average, 56.2 percent of the population of the full sample and 59.4 percent of the participation sample was exposed to active living-oriented zoning. Both the full sample and participation sample feature large proportions of non-Hispanic White adults who were married and had a college degrees. There were slightly more females than males in both samples.

Table 3.6.7 presents a summary of the statistics about the final analytical sample of sedentary behavior. About 82 percent of the respondents in the sample engaged in sedentary behavior. Conditional on engagement, people spent nearly 4 hours per day on sedentary behavior on average. The average county-level walkability scale was 1.55 in the full sample and 1.52 in the participation sample. On average, 56.2 percent of the population of the full sample and 55.8 percent of the participation sample was exposed to active living-oriented zoning. Both the full sample and participation sample featured a large proportion of non-Hispanic White adults who were married and had a college degree and included slightly more females than males.

Table 3.6.1 Characteristics of All Counties from the Data of Zoning, NAVTEQ 2011, and ACS 2011-2015 5-Year Estimates (Before Merged with ATUS and 1900 Census)

VARIABLES	Mean	SD	Min	Max
Population Census 2010	467,149	676,856	122,131	9,826,773
% of County Excluded	1.45	3	0	38.8
Number of ALL Municipalities/Unincorporated Areas	13.8	13.2	1	132
Number of Municipalities/Unincorporated Areas Included	8.85	6.64	1	55
Walkability (NAVTEQ/ACS)	1	1	0.716	18
Population Density Per Square Mile 2011	1,209	4,155	7.14	69,572
Percent of Total Intersections That Are 4-Way 2011	0.224	0.0782	0.0812	0.633
Housing Unit Density Per Square Mile 2011	526	1,997	3.38	36,918
Total Intersections Per Square Mile 2011	22.1	28.5	0.291	179
Land Area (Square Miles) Census 2010	1,086	1,794	15	20,057
Active Living-Oriented Zoning	0.512	0.255	0	1
Code Reform Zoning	0.241	0.33	0	1
Sidewalks	0.7	0.324	0	1
Crosswalks	0.255	0.332	0	1
Bike-Pedestrian Connectivity	0.428	0.374	0	1
Street Connectivity	0.407	0.371	0	1
Bike Lanes	0.165	0.29	0	1
Bike Parking	0.403	0.384	0	1
Bike-Pedestrian Trails-Paths	0.576	0.379	0	1
Other Walkability	0.682	0.33	0	1
Mixed Use	0.654	0.338	0	1
Active Recreation	0.818	0.285	0	1
Passive Recreation	0.817	0.288	0	1
% Households in Poverty (ACS)	13.5	4.77	3.72	31.2
% Non-Hispanic White (ACS)	67.1	19.1	3.66	95.3
% Non-Hispanic Black (ACS)	11.9	12.6	0.277	70.1
% Hispanic (ACS)	13.7	14.7	1	95.3
Median Household Income (ACS)	57,600	14,839	30,608	123,453
Median Age (ACS)	37.8	4.34	24.4	57.3
West (ACS)	0.183	0.387	0	1
Midwest (ACS)	0.224	0.417	0	1
Northeast (ACS)	0.202	0.402	0	1
South (ACS)	0.391	0.488	0	1
% ALL Population	74.257			
Number of ALL Municipalities/Unincorporated Areas	6,822			
% Population Included	72.525			
Number of Municipalities/Unincorporated Areas Included	4,388			
Number of Counties	496			
Number of States	50			

Note: The difference between ALL and Included is due to zoning codes were collected only for municipalities or unincorporated areas with populations greater than 0.5% of the county population.

Table 3.6.2 Characteristics of Counties in the Analysis

VARIABLES	Mean	SD	Min	Max
Population Census 2010	497,543	763,734	122,131	9,826,773
% of County Population Excluded	1.53	2.8	0	30.8
Number of ALL Municipalities/Unincorporated Areas	15.2	14.3	1	91
Number of Municipalities/Unincorporated Areas Included	9.77	7.46	1	55
Walkability (NAVTEQ/ACS)	1.08	1.3	0.723	18
Population Density Per Square Mile 2011	1,543	5,365	32	69,572
Percent of Total Intersections That Are 4-Way 2011	0.229	0.0845	0.0852	0.633
Housing Unit Density Per Square Mile 2011	676	2,652	16.4	36,918
Total Intersections Per Square Mile 2011	24.7	33.4	1.08	179
Land Area (Square Miles) Census 2010	1,019	1,640	15	20,057
Active Living-Oriented Zoning	0.498	0.253	0.0169	1
Code Reform Zoning	0.243	0.337	0	1
Sidewalks	0.691	0.317	0	1
Crosswalks	0.252	0.334	0	1
Bike-Pedestrian Connectivity	0.397	0.365	0	1
Street Connectivity	0.364	0.357	0	1
Bike Lanes	0.152	0.283	0	1
Bike Parking	0.399	0.382	0	1
Bike-Pedestrian Trails-Paths	0.56	0.372	0	1
Other Walkability	0.666	0.33	0	1
Mixed Use	0.629	0.344	0	1
Active Recreation	0.811	0.287	0.0549	1
Passive Recreation	0.808	0.291	0.0472	1
Manufacturing Establishment Density Census 1900	5.85	34.2	0	431
Farmland Proportion Census 1900	0.659	0.328	0.00809	1.5
% Households in Poverty (ACS)	13.5	5.13	3.72	31.2
% Non-Hispanic White (ACS)	66.3	20	3.66	95.3
% Non-Hispanic Black (ACS)	11.8	12.1	0.277	65.8
% Hispanic (ACS)	14.7	16.2	1.04	95.3
Median Household Income (ACS)	58,566	15,693	30,608	123,453
Median Age (ACS)	37.6	4.13	24.4	48.7
West (ACS)	0.163	0.37	0	1
Midwest (ACS)	0.239	0.427	0	1
Northeast (ACS)	0.215	0.412	0	1
South (ACS)	0.382	0.487	0	1
% ALL Population	40.031			
Number of ALL Municipalities/Unincorporated Areas	3,813			
% Population Included	39.082			
Number of Municipalities/Unincorporated Areas Included	2,453			
Number of Counties	251			
Number of States	37			

Note: The difference between ALL and Included is due to zoning codes were collected only for municipalities or unincorporated areas with populations greater than 0.5% of the county population.

Table 3.6.3 Characteristics of Counties Excluded from the Analysis

VARIABLES	Mean	SD	Min	Max
Population Census 2010	436,010	574,330	122,166	5,199,971
% of County Population Excluded	1.38	3.2	0	38.8
Number of ALL Municipalities/Unincorporated Areas	12.3	11.9	1	132
Number of Municipalities/Unincorporated Areas Included	7.9	5.52	1	31
Walkability (NAVTEQ/ACS)	0.919	0.524	0.716	8.01
Population Density Per Square Mile 2011	866	2,302	7.14	32,654
Percent of Total Intersections That Are 4-Way 2011	0.219	0.071	0.0812	0.479
Housing Unit Density Per Square Mile 2011	372	915	3.38	12,143
Total Intersections Per Square Mile 2011	19.4	22.2	0.291	163
Land Area (Square Miles) Census 2010	1,156	1,941	26	18,619
Active Living-Oriented Zoning	0.527	0.257	0	0.999
Code Reform Zoning	0.24	0.324	0	1
Sidewalks	0.709	0.331	0	1
Crosswalks	0.258	0.329	0	1
Bike-Pedestrian Connectivity	0.46	0.38	0	1
Street Connectivity	0.451	0.381	0	1
Bike Lanes	0.177	0.296	0	1
Bike Parking	0.407	0.388	0	1
Bike-Pedestrian Trails-Paths	0.593	0.386	0	1
Other Walkability	0.699	0.33	0	1
Mixed Use	0.68	0.331	0	1
Active Recreation	0.824	0.283	0	1
Passive Recreation	0.825	0.285	0	1
% Households in Poverty (ACS)	13.5	4.38	4.31	30.2
% Non-Hispanic White (ACS)	68	18.2	10.3	94.9
% Non-Hispanic Black (ACS)	12	13	0.325	70.1
% Hispanic (ACS)	12.6	13	1	81.8
Median Household Income (ACS)	56,611	13,872	34,299	112,552
Median Age (ACS)	37.9	4.56	27.9	57.3
West (ACS)	0.204	0.404	0	1
Midwest (ACS)	0.208	0.407	0	1
Northeast (ACS)	0.188	0.391	0	1
South (ACS)	0.4	0.491	0	1
% ALL Population	34.226			
Number of ALL Municipalities/Unincorporated Areas	3,009			
% Population Included	33.443			
Number of Municipalities/Unincorporated Areas Included	1,935			
Number of Counties	245			
Number of States	46			

Note: The difference between ALL and Included is due to zoning codes were collected only for municipalities or unincorporated areas with populations greater than 0.5% of the county population.

Table 3.6.4 Characteristic Comparison between Counties Included in the Analysis and Counties Excluded from the Analysis

VARIABLES	Included Median	Excluded Median	z-stat	Prob > z-stat
Population Census 2010	269,314	252,789	1.181	0.238
% of County Population Excluded	0.644	0.632	1.275	0.202
Number of ALL Municipalities/Unincorporated Areas	11	10	2.464	0.014
Number of Municipalities/Unincorporated Areas	8	7	2.859	0.004
Walkability (NAVTEQ/ACS)	0.805	0.812	0.526	0.599
Population Density Per Square Mile 2011	382.136	410.806	0.644	0.519
Percent of Total Intersections That Are 4-Way	0.22	0.213	0.972	0.331
Housing Unit Density Per Square Mile 2011	162.937	179.453	0.271	0.786
Total Intersections Per Square Mile 2011	12.079	12.503	0.465	0.642
Land Area (Square Miles) Census 2010	627.776	652.431	-0.598	0.55
Active Living-Oriented Zoning	0.5	0.562	-1.441	0.15
Code Reform Zoning	0.044	0.046	0.58	0.562
Sidewalks	0.79	0.848	-0.939	0.348
Crosswalks	0.067	0.065	0.365	0.715
Bike-Pedestrian Connectivity	0.308	0.413	-1.543	0.123
Street Connectivity	0.232	0.431	-1.91	0.056
Bike Lanes	0	0	-0.809	0.419
Bike Parking	0.272	0.336	0.275	0.783
Bike-Pedestrian Trails-Paths	0.586	0.734	-1.06	0.289
Other Walkability	0.787	0.852	-1.369	0.171
Mixed Use	0.734	0.815	-1.784	0.074
Active Recreation	1	1	-0.423	0.672
Passive Recreation	1	1	-0.659	0.51
% Households in Poverty (ACS)	13.311	13.709	-0.462	0.644
% Non-Hispanic White (ACS)	70	71.305	-0.694	0.488
% Non-Hispanic Black (ACS)	8.391	7.267	0.539	0.59
% Hispanic (ACS)	8.706	8.146	0.847	0.397
Median Household Income (ACS)	54,989	53,525	1.117	0.264
Median Age (ACS)	37.7	37	0.653	0.514
	Included Mean	Excluded Mean		
West (ACS)	0.163	0.204	-1.172	0.241
Midwest (ACS)	0.239	0.208	0.825	0.409
Northeast (ACS)	0.215	0.188	0.76	0.447
South (ACS)	0.382	0.4	-0.4	0.689
Number of Counties	251	245		
Number of States	37	46		

Note: The difference between ALL and Excluded is due to zoning codes were collected only for municipalities or unincorporated areas with populations greater than 0.5% of the county population.

Table 3.6.5 Characteristic Comparison between Counties Included in the Analysis and All Counties (Counties in the Zoning Data)

VARIABLES	Included Median	All Median	z-stat	Prob > z-stat
Population Census 2010	269,314	263,614.5	0.677	0.499
% of County Population Excluded	0.644	0.632	0.73	0.466
Number of ALL Municipalities/Unincorporated Areas	11	10	1.412	0.158
Number of Municipalities/Unincorporated Areas	8	7	1.638	0.101
Walkability (NAVTEQ/ACS)	0.805	0.811	0.301	0.763
Population Density Per Square Mile 2011	382.136	407.091	0.369	0.712
Percent of Total Intersections That Are 4-Way	0.22	0.216	0.557	0.578
Housing Unit Density Per Square Mile 2011	162.937	168.865	0.155	0.877
Total Intersections Per Square Mile 2011	12.079	12.393	0.266	0.79
Land Area (Square Miles) Census 2010	627.776	646.936	-0.343	0.732
Active Living-Oriented Zoning	0.5	0.526	-0.826	0.409
Code Reform Zoning	0.044	0.045	0.331	0.74
Sidewalks	0.79	0.831	-0.538	0.591
Crosswalks	0.067	0.066	0.209	0.835
Bike-Pedestrian Connectivity	0.308	0.34	-0.884	0.377
Street Connectivity	0.232	0.311	-1.094	0.274
Bike Lanes	0	0	-0.464	0.642
Bike Parking	0.272	0.294	0.157	0.875
Bike-Pedestrian Trails-Paths	0.586	0.674	-0.607	0.544
Other Walkability	0.787	0.82	-0.784	0.433
Mixed Use	0.734	0.779	-1.022	0.307
Active Recreation	1	1	-0.242	0.809
Passive Recreation	1	1	-0.378	0.706
% Households in Poverty (ACS)	13.311	13.559	-0.265	0.791
% Non-Hispanic White (ACS)	70	71.074	-0.398	0.691
% Non-Hispanic Black (ACS)	8.391	7.768	0.309	0.757
% Hispanic (ACS)	8.706	8.258	0.485	0.628
Median Household Income (ACS)	54,989	54,173	0.64	0.522
Median Age (ACS)	37.7	37.5	0.374	0.708
	Included Mean	All Mean		
West (ACS)	0.163	0.183	-0.681	0.496
Midwest (ACS)	0.239	0.224	0.469	0.639
Northeast (ACS)	0.215	0.202	0.432	0.666
South (ACS)	0.382	0.391	-0.229	0.819
Number of Counties	251	496		
Number of States	37	50		

Note: The difference between ALL and Excluded is due to zoning codes were collected only for municipalities or unincorporated areas with populations greater than 0.5% of the county population.

Table 3.6.6 Physical Activity Summary Statistics

	Full Sample		Participation Sample		Non-Participation Sample	
VARIABLES	Mean	SD	Mean	SD	Mean	SD
Physical Activity Participation	0.0805	0.272	1	0	0	0
Minutes Spent on Physical Activity	4.83	20.2	60	42.1	0	0
Walkability	1.55	2.39	1.59	2.5	1.54	2.38
Active Living-Oriented Zoning	0.562	0.246	0.594	0.23	0.559	0.247
Manufacturing Establishment Density	12.7	57.2	13.2	60	12.7	56.9
Farmland Proportion 1900	0.586	0.341	0.565	0.346	0.588	0.341
Individual Controls from ATUS						
Age	47.6	17.7	50.3	18.4	47.4	17.6
Male	0.446	0.497	0.482	0.5	0.442	0.497
Female (Reference)	0.554	0.497	0.518	0.5	0.558	0.497
Non-Hispanic White (Reference)	0.586	0.493	0.591	0.492	0.585	0.493
Non-Hispanic Black	0.155	0.362	0.104	0.305	0.159	0.366
Non-Hispanic Other	0.0689	0.253	0.099	0.299	0.0663	0.249
Hispanic	0.19	0.393	0.206	0.404	0.189	0.392
Married (Reference)	0.465	0.499	0.491	0.5	0.463	0.499
Widow, Divorced, Separated	0.264	0.441	0.27	0.444	0.264	0.441
Single	0.271	0.444	0.239	0.427	0.274	0.446
Any Difficulty	0.109	0.311	0.0965	0.295	0.11	0.313
Less than High School	0.145	0.352	0.154	0.361	0.145	0.352
High School Graduate	0.229	0.42	0.182	0.386	0.233	0.423
Some College	0.178	0.383	0.142	0.349	0.182	0.385
College Graduate (Reference)	0.313	0.464	0.333	0.471	0.311	0.463
Graduate	0.135	0.341	0.189	0.392	0.13	0.336
County Controls from ACS						
% Households in Poverty	14	4.87	13.8	4.45	14	4.91
% Non-Hispanic White	56	21.3	53.9	20.7	56.2	21.3
% Non-Hispanic Black	13.1	12.2	12.1	11.5	13.2	12.2
% Hispanic	20.8	19	22.6	18.3	20.6	19
Median Household Income	59,606	14,791	60,973	14,378	59,486	14,821
Median Age	37.1	3.57	37	3.47	37.1	3.58
Land Area (Square Miles) 2010	1,588	2,640	1,741	2,620	1,574	2,642
Census Region						
West (Reference)	0.311	0.463	0.401	0.49	0.303	0.459
Midwest	0.188	0.391	0.152	0.359	0.192	0.394
Northeast	0.223	0.416	0.185	0.389	0.226	0.418
South	0.278	0.448	0.262	0.44	0.28	0.449
Weekday (Reference)	0.5	0.5	0.527	0.499	0.497	0.5
Sunday	0.256	0.436	0.234	0.423	0.258	0.437
Saturday	0.245	0.43	0.239	0.427	0.245	0.43

Month Dummies						
January (Reference)	0.0983	0.298	0.0747	0.263	0.1	0.3
February	0.0817	0.274	0.0665	0.249	0.083	0.276
March	0.0912	0.288	0.0879	0.283	0.0915	0.288
April	0.0823	0.275	0.0965	0.295	0.081	0.273
May	0.0798	0.271	0.0945	0.293	0.0785	0.269
June	0.0848	0.279	0.0889	0.285	0.0845	0.278
July	0.0797	0.271	0.0843	0.278	0.0793	0.27
August	0.0857	0.28	0.109	0.312	0.0836	0.277
September	0.0776	0.268	0.0868	0.282	0.0768	0.266
October	0.0789	0.27	0.0914	0.288	0.0778	0.268
November	0.0816	0.274	0.0691	0.254	0.0827	0.275
December	0.0785	0.269	0.0503	0.219	0.081	0.273
Year Dummies						
Year 2010 (Reference)	0.183	0.386	0.169	0.375	0.184	0.387
Year 2011	0.174	0.379	0.169	0.375	0.175	0.38
Year 2012	0.173	0.378	0.188	0.391	0.171	0.377
Year 2013	0.158	0.365	0.141	0.348	0.159	0.366
Year 2014	0.166	0.372	0.164	0.37	0.166	0.372
Year 2015	0.147	0.354	0.17	0.376	0.145	0.352
Number of States	37		35		37	
Number of Counties	251		213		251	
Number of Observations	24,448		1,969		22,479	

Table 3.6.7 Sedentary Behavior Summary Statistics

	Full Sample		Participation Sample		Non-Participation Sample	
VARIABLES	Mean	SD	Mean	SD	Mean	SD
Sedentary Behavior Participation	0.823	0.381	1	0	0	0
Minutes Spent on Sedentary Behavior	190	187	231	181	0	0
Walkability	1.55	2.39	1.52	2.3	1.68	2.78
Active Living-Oriented Zoning	0.562	0.246	0.558	0.246	0.581	0.244
Manufacturing Establishment Density 1900	12.7	57.2	12.1	54.8	15.8	67.2
Farmland Proportion 1900	0.586	0.341	0.591	0.34	0.567	0.347
Individual Controls from ATUS						
Age	47.6	17.7	48.6	17.9	42.6	15.5
Male	0.446	0.497	0.454	0.498	0.408	0.492
Female (Reference)	0.554	0.497	0.546	0.498	0.592	0.492
Non-Hispanic White (Reference)	0.586	0.493	0.584	0.493	0.596	0.491
Non-Hispanic Black	0.155	0.362	0.16	0.366	0.133	0.339
Non-Hispanic Other	0.0689	0.253	0.0653	0.247	0.0859	0.28
Hispanic	0.19	0.393	0.192	0.394	0.185	0.388
Married (Reference)	0.465	0.499	0.461	0.498	0.484	0.5
Widow, Divorced, Separated	0.264	0.441	0.273	0.445	0.225	0.418
Single	0.271	0.444	0.267	0.442	0.291	0.454
Any Difficulty	0.109	0.311	0.119	0.324	0.0588	0.235
Less than High School	0.146	0.353	0.152	0.359	0.114	0.318
High School Graduate	0.229	0.42	0.242	0.428	0.169	0.375
Some College	0.178	0.383	0.179	0.383	0.176	0.381
College Graduate (Reference)	0.312	0.464	0.304	0.46	0.353	0.478
Graduate	0.135	0.341	0.123	0.329	0.188	0.391
County Controls from ACS						
% Households in Poverty	14	4.87	14.1	4.91	13.8	4.71
% Non-Hispanic White	56	21.3	56.1	21.3	55.9	21
% Non-Hispanic Black	13.1	12.2	13.2	12.2	12.6	11.9
% Hispanic	20.8	19	20.8	19.1	20.8	18.4
Median Household Income	59,606	14,792	59,368	14,779	60,716	14,802
Median Age	37.1	3.57	37.2	3.59	36.9	3.48
Land Area (Square Miles) 2010	1,588	2,640	1,583	2,650	1,607	2,590
Census Region						
West (Reference)	0.311	0.463	0.303	0.46	0.344	0.475
Midwest	0.188	0.391	0.191	0.393	0.179	0.383
Northeast	0.223	0.416	0.226	0.418	0.208	0.406
South	0.278	0.448	0.28	0.449	0.269	0.443
Weekday (Reference)	0.5	0.5	0.493	0.5	0.528	0.499
Sunday	0.256	0.436	0.267	0.442	0.203	0.402
Saturday	0.245	0.43	0.24	0.427	0.268	0.443

Month Dummies						
January (Reference)	0.0983	0.298	0.101	0.301	0.0878	0.283
February	0.0817	0.274	0.084	0.277	0.0713	0.257
March	0.0912	0.288	0.0923	0.29	0.0857	0.28
April	0.0823	0.275	0.0821	0.275	0.0829	0.276
May	0.0798	0.271	0.0765	0.266	0.0954	0.294
June	0.0848	0.279	0.0851	0.279	0.0839	0.277
July	0.0797	0.271	0.0786	0.269	0.0848	0.279
August	0.0857	0.28	0.0841	0.278	0.0929	0.29
September	0.0776	0.268	0.0771	0.267	0.0804	0.272
October	0.0788	0.269	0.0786	0.269	0.0797	0.271
November	0.0816	0.274	0.0835	0.277	0.0725	0.259
December	0.0785	0.269	0.0776	0.267	0.0827	0.275
Year Dummies						
Year 2010 (Reference)	0.183	0.386	0.181	0.385	0.19	0.392
Year 2011	0.174	0.379	0.174	0.379	0.175	0.38
Year 2012	0.173	0.378	0.173	0.379	0.169	0.375
Year 2013	0.158	0.365	0.156	0.363	0.166	0.372
Year 2014	0.165	0.372	0.167	0.373	0.158	0.365
Year 2015	0.147	0.354	0.149	0.356	0.142	0.349
Number of States	37		37		36	
Number of Counties	251		249		228	
Number of Observations	24,458		20,141		4,317	

4 MEDIATIONAL ANALYSIS

4.1 Conceptual Framework

There are generally two approaches that link the built environment to physical activity (Transportation Research Board 2005). One is demand theory, which is used to model individual traveling decisions and is mainly used in economics and transportation research. For example, Cervero and Kockelman (1997) identified three features of the built environment that affect travel demand: density (compact neighborhoods encourage non-motorized travel), diversity (e.g., convenience stores within neighborhoods produce more walk and cycling), and design (e.g., aligning shade trees along sidewalks encourages walking). Demand theory is based on utility maximization. It is powerful at explaining complex behaviors. The other is ecological models, which are mainly used in health behavior research. Ecological models consist of “multiple levels of influence on behavior, from individual and social factors, to institutional, community, built environment, and policy factors” (Sallis et al. 2006; Sallis et al. 2012). They help us understand the interrelations among various factors.

Because I focus on causal inference, I use a directed acyclic graph (DAG) as a conceptual framework to guide the research design. DAG consists of many vertices and edges, with each edge directed from one vertex to another. DAG does not have a loop structure. Different from the other two approaches, DAG shows explicit causal chains among all variables. DAG is widely used in causal inference, for which it is also known as the causal graphic model.

Figure 4.1.1 shows the basic conceptual framework: zoning affects adults’ active living outcomes by shaping the built environment. Path a represents the effect of zoning on built environment. Path b accounts for the effect of built environment on active living outcomes. The variable used to measure the built environment is walkability, which is a standardized summation of four density metrics, including population density, housing unit density, the ratio of four-way intersections to all intersections, and the total number of intersections divided by land area (Chriqui et al. 2016, Slater et al. 2010, Ewing and Hamidi 2013). It can only be used to partially assess the built environment. Other assessments include the condition of the sidewalks, the condition of residential and commercial buildings, and the presence of well-tended yards (Anderson et al. 2013). Due to this limitation, I conduct the research based on Figure 4.1.2.

Figure 4.1.2 shows that the effects of zoning on active living outcomes can be classified into two types. One is the direct effect, which is represented by Path c′, and the other is the indirect effect, which goes along Path a and Path b. The indirect effect is due to the partial mediation of walkability.

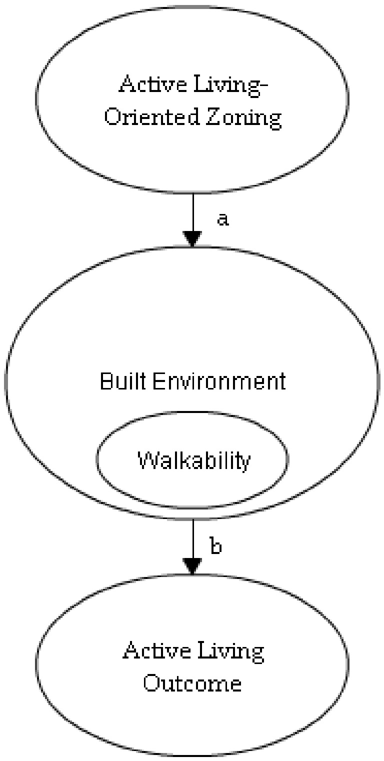


Figure 4.1.1 Conceptual Framework 1

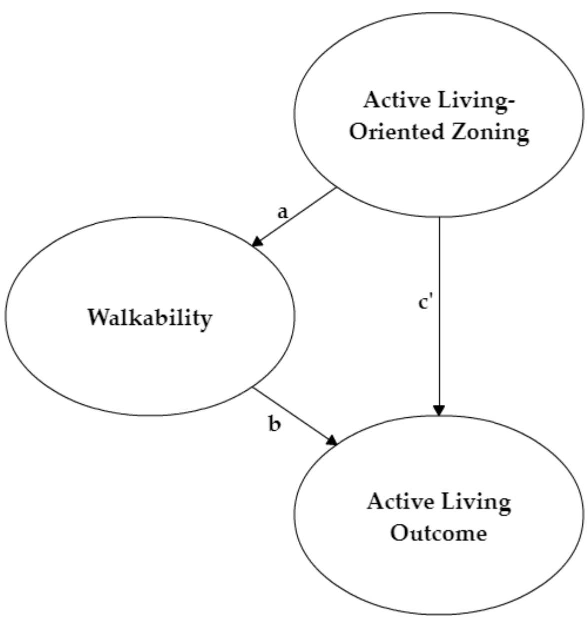


Figure 4.1.2 Conceptual Framework 2

4.2 Basic Model: Two-Part Model

I use a two-part model to characterize adults' active living outcomes

$$E(y|X) = \Pr(y > 0|X) \times E(y|y > 0, X) \tag{1}$$

where y is the time people spent on a certain activity and X is a vector of explanatory variables, including zoning policy and the walkability. $\Pr(y > 0|X)$ is the probability that people will choose to do an activity. $E(y|y > 0, X)$ is the expected time people spend on the given activity. Thus, the expected time people spend on the activity is equal to the product of the two terms. For $\Pr(y > 0|X)$, I use a Probit model and maximum likelihood estimation. For $E(y|y > 0, X)$, I use a linear model $E(y|y > 0, X) = X\beta$ and ordinary least square estimation (OLS). I also do log transformation on y for observations of $y > 0$ and build a linear model, $E(\log(y)|y > 0, X) = X\beta$, because y is highly right-skewed.

Specifically,

$$\text{Prob}(y_{ict} > 0) = \Phi(\beta_0 + \beta_1 z_c + \beta_2 w_c + \beta_3 indi_{itc} + \beta_4 cnty_c + \beta_5 time_t) \tag{2}$$

$$y_{ict} = \delta_0 + \delta_1 z_c + \delta_2 w_c + \delta_3 indi_{itc} + \delta_4 cnty_c + \delta_5 time_t + r_{itc} \text{ for } y_{itc} > 0 \tag{3}$$

$$\log(y_{ict}) = \delta_0 + \delta_1 z_c + \delta_2 w_c + \delta_3 indi_{itc} + \delta_4 cnty_c + \delta_5 time_t + r_{itc} \text{ for } y_{itc} > 0 \tag{4}$$

$\Phi(\cdot)$: standard normal cumulative density function

i: individual index

c: county index

t: time index of three levels consisting of weekday level, month level, and year level (2010, 2011, 2012, 2013, 2014, and 2015)

The time usage y of person *i* in county *c* and year *t* depends on walkability (w_c), active living-oriented zoning (z_c), individual characteristics ($indi_{itc}$), including race, gender, education level, marital status, county characteristics ($cnty_c$), including the percentage of households in poverty, the percentage of non-Hispanic White people, the percentage of non-Hispanic Black people, the percentage of Hispanic people, median household income, county land area of 2010 (square miles), census region, time fixed effect, including weekend fixed effects (Saturday, Sunday), month fixed effects, and year fixed effects.

For the Probit model, I also calculate the average marginal effects based on estimated coefficients using the formula:

$$\text{Average Marginal Effect of } X \text{ on } \text{Prob}(y > 0)$$

$$= \frac{1}{N}\sum_{j=1}^{N}\frac{\partial \widehat{\text{Prob}}(y_j > 0)}{\partial X_j} = \frac{1}{N}\sum_{j=1}^{N}\frac{\partial \Phi(X_j'\hat{\beta})}{\partial X_j} = \frac{1}{N}\sum_{j=1}^{N}\phi(X_j'\hat{\beta})\hat{\beta} \qquad (5)$$

where $\phi(\cdot)$ is standard normal probability density function, the derivative of $\Phi(\cdot)$.

4.3 Three Steps for Mediational Analysis

Mediational analysis can help us understand the mechanism through which the causal variable affects the outcome. I aim to estimate the causal effect of zoning on active living outcomes, and walkability is a mediator.

Following Baron and Kenny (1986), James and Brett (1984), Judd and Kenny (1981), I conduct the mediational analysis in three steps.

Step 1. Exclude the walkability from the basic model and run analysis to estimate the total effect of zoning, which includes both a direct effect and an indirect effect (through walkability). The results are shown in Table 4.5.1 and Table 4.5.2.

Step 2. Estimate the impact of zoning on walkability, which is denoted by “a” in Figure 4.1.2. The results are shown in Table 4.5.3.

Step 3. Include both walkability and zoning in the basic model and run analysis to estimate the effect of walkability and the direct effect of zoning, which are denoted by “b” and “c′,” respectively, in Figure 4.1.2. The results are shown in Table 4.5.4 and Table 4.5.5.

4.4 Sobel Test

Sobel (1982) proposed a test for the significance of the indirect effect of the causal variable. Suppose the marginal effect of zoning on walkability is a, and the marginal effect of walkability on active living outcome is b; the indirect effect of zoning on active living outcome is ab. Based on the Delta method, the standard error of ab is $\sqrt{a^2 s_b^2 + b^2 s_a^2}$, where s_a and s_b are the standard errors of a and b, respectively. The test statistic is formed as $ab/\sqrt{a^2 s_b^2 + b^2 s_a^2}$, which follows standard normal distribution under the null hypothesis that $ab = 0$.

The Sobel test has some limitations here. First, the standard error of ab may not be correct. The derivation of the standard error assumes independence between a and b. When the active living outcome is dichotomous, the calculation of the average marginal effect of walkability b involves the data of zoning, so b may not be independent from the marginal effect of zoning on walkability a. The dependence between a and b will lead to downward bias of the standard error estimate of ab in the Sobel test so that the null hypothesis can be easily rejected. Second, the test is very conservative, so it could have insufficient power to detect the true effect (Mackinnon, Warsi, and Dwyer 1995). Suppose both a and are positive; the sampling distribution of ab will be left-skewed rather than be symmetric at zero. The normal distribution for the test statistic will be improper, so the null hypothesis can hardly be rejected (Kenny 2016).

The limitations of the Sobel test implies researchers cannot purely rely on the Sobel test to draw conclusions about on the mediation. I combine the changes in coefficient magnitudes and the results from the Sobel test to draw the conclusion about the mediation. Hence, the conclusion is robust, even if the assumptions of the Sobel test fail to hold true.

4.5 Discussions and Results

Table 4.5.1 and Table 4.5.2 show the results of Step 1 of the mediational analysis for physical activity and sedentary behavior, respectively. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 2.8 percentage point increase (Probit model) in the chances that an adult engaged in physical activity and a 4.6 percentage point decrease (Probit model) in the chances that an adult chose sedentary behavior, given that other factors remain unchanged. The effects of zoning are statistically significant at the 1% level. Nonetheless, the effect of zoning on the duration people spent on physical activity or sedentary behavior is statistically insignificant.

Table 4.5.3 shows the result of Step 2 of the mediational analysis. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 5.06 unit increase in the walkability metric. The effect of zoning on walkability is statistically significant at the 1% level.

Table 4.5.4 and Table 4.5.5 show the results of Step 3 of the mediational analysis for physical activity and sedentary behavior, respectively. Different from Step 1, I control for both zoning and walkability in the analysis. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 2.1 percentage point increase (Probit model) in the chances that an adult engaged in physical activity and a 2.8 percentage point decrease (Probit model) in the chances that an adult chose sedentary behavior if other factors remain unchanged. A one unit increase in the walkability metric is associated with a 0.1 percentage point increase (Probit model) in the chances that an adult engaged in physical activity and a 0.3 percentage point decrease (Probit model) in the chances that an adult chose sedentary behavior when other factors remain unchanged. Zoning and walkability have no significant effects on the number of minutes people spent on physical activity or sedentary behavior, except that a one-unit increase in the walkability metric is associated with a 0.018 percent increase (OLS) in the number of minutes they spent on physical activity.

Figure 4.5.1 and Figure 4.5.2 summarize the results of the mediational analysis of the probability that people engaged in physical activity and sedentary behavior, respectively. Zoning has significant effects on the decisions of people to participate in physical activity and sedentary behavior. Once I add walkability into the analysis, the effects of zoning decrease in magnitude, but are still significant, and the effects of walkability are significant. Also, zoning has a significant effect on walkability. The Sobel tests show that the indirect effect of zoning is significant on sedentary behavior but insignificant on physical activity. Note that, in physical activity analysis, both a and b are positive. Based on the discussion about the limitations of the Sobel test, the test has insufficient power to correctly reject the null hypothesis that the indirect effect of zoning on physical activity is insignificant. Therefore, I use the changes in magnitudes of the effects to make conclusions in the mediational analysis. Part of the effects of zoning on participation decisions are carried by walkability when controlling for walkability. Walkability is a mediator of the association between zoning and participation decisions. Zoning affects people's decisions to participate in physical activity and sedentary behavior by shaping walkability.

Table 4.5.1 The Mediational Analysis, Step 1: Physical Activity Regression

	Probit		OLS	OLS Log
VARIABLES	COEFF	Marginal Effect	COEFF	COEFF
Active Living-Oriented Zoning	0.190***	0.028***	3.510	0.040
	(0.067)	(0.010)	(4.879)	(0.090)
Number of Observations	24,448	24,448	1,969	1,969

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The full table is in Appendix 5 Column 1, 4, and 6.

Table 4.5.2 The Mediational Analysis, Step 1: Sedentary Behavior Regression

	Probit		OLS	OLS Log
VARIABLES	COEFF	Marginal Effect	COEFF	COEFF
Active Living-Oriented Zoning	-0.187***	-0.046***	-7.719	-0.035
	(0.067)	(0.017)	(6.580)	(0.033)
Number of Observations	24,458	24,458	20,141	20,141

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The full table is in Appendix 7 Column 1, 4, and 6.

Table 4.5.3 The Mediational Analysis, Step 2: Walkability Regression

VARIABLES	Walkability
Active Living-Oriented Zoning	5.06***
	(0.131)
Observations	24,458

Note: In all specifications, I control for county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), census region fixed effects, and intercept. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 2 Column 1.

Table 4.5.4 The Mediational Analysis, Step 3: Physical Activity Regression

	Probit		OLS	OLS Log
VARIABLES	COEFF	Marginal Effect	COEFF	COEFF
Active Living-Oriented Zoning	0.148*	0.021*	0.431	-0.063
	(0.077)	(0.011)	(6.187)	(0.101)
Walkability	0.008*	0.001*	0.538	0.018***
	(0.004)	(0.001)	(0.410)	(0.007)
Number of Observations	24,448	24,448	1,969	1,969

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects. Standard Errors are clustered at county-level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The full table is in Appendix 6 Column 1, 5, and 7.

Table 4.5.5 The Mediational Analysis, Step 3: Sedentary Behavior Regression

	Probit		OLS	OLS Log
VARIABLES	COEFF	Marginal Effect	COEFF	COEFF
Active Living-Oriented Zoning	-0.112*	-0.028*	-5.360	-0.013
	(0.068)	(0.017)	(7.864)	(0.039)
Walkability	-0.014***	-0.003***	-0.495	-0.004
	(0.004)	(0.001)	(0.685)	(0.003)
Number of Observations	24,458	24,458	20,141	20,141

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 8 Column 1, 5, and 7.

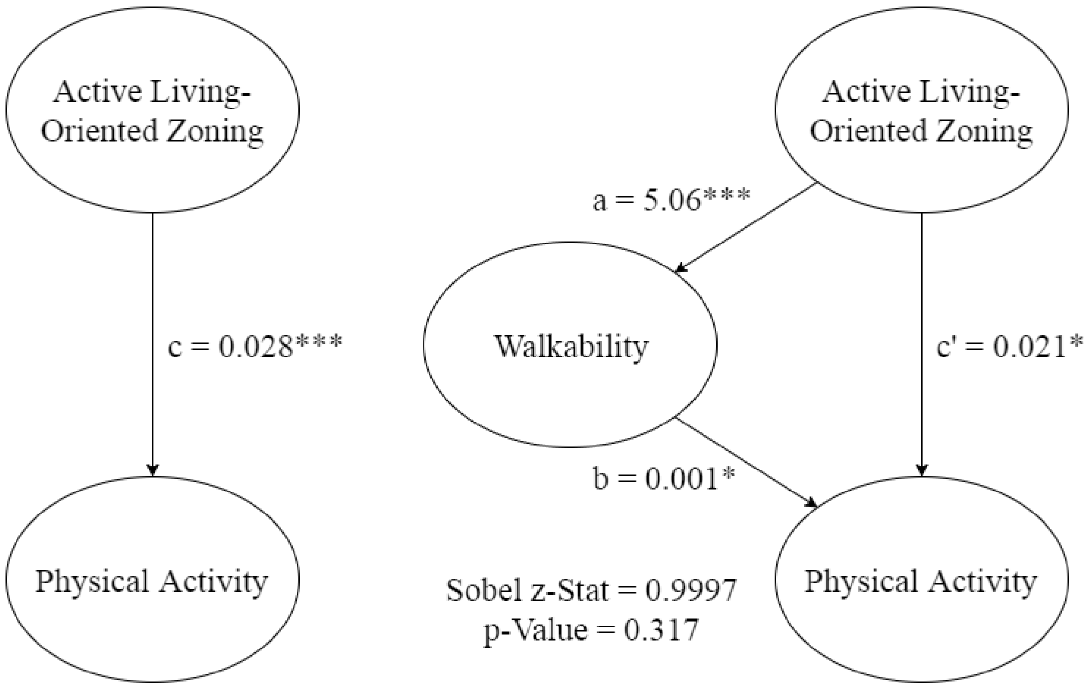


Figure 4.5.1 Physical Activity, Probit Marginal Effects

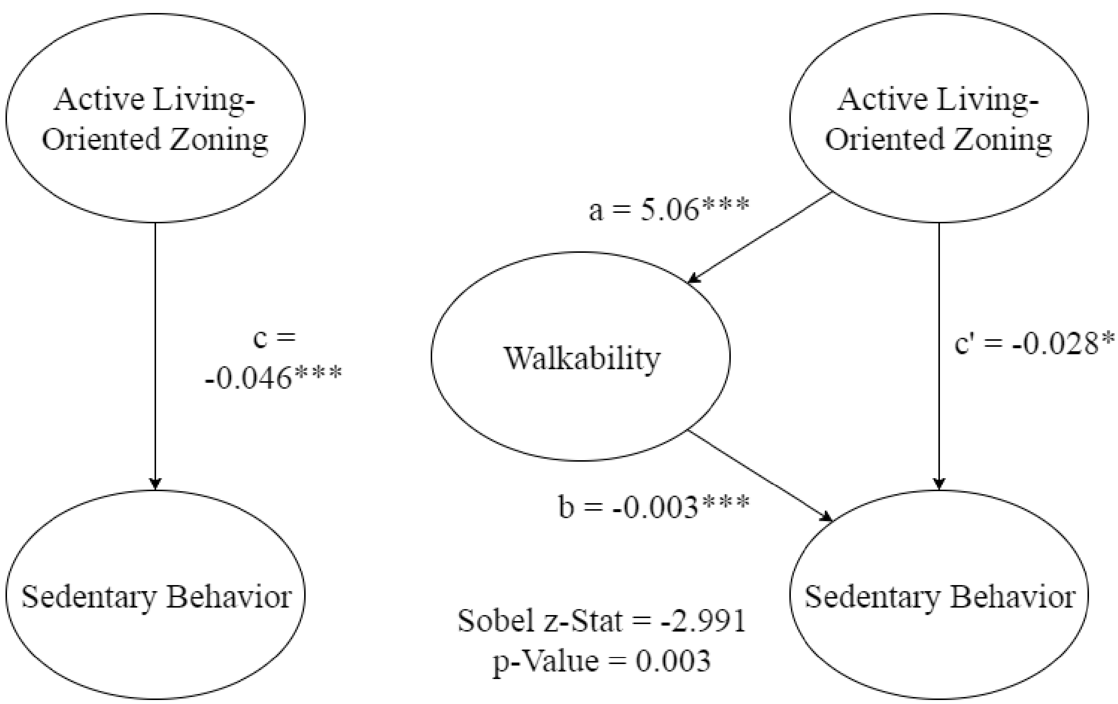


Figure 4.5.2 Sedentary Behavior, Probit Marginal Effects

5 INSTRUMENTAL VARIABLE IDENTIFICATION AND ESTIMATION

5.1 Tiebout Sorting, Tiebout Equilibrium, and Market Equilibrium

The regressions in the mediational analysis are incapable of identifying the causal effects of zoning and built environment on active living outcomes. The reason for this is that they do not take zoning or built environment as a choice variable, which fails to address the sorting process that underlies the demand for zoning and built environment. Built environment and zoning are essentially public goods. Households or individuals can migrate across communities based on their preferences for the public goods provided by communities (Tiebout 1956),[12] which is characterized as "sorting" by economists. Specifically, individuals or households can choose where to live based, in part, on their active living preferences. People who favor active living may choose to live in places with active living-oriented zoning and built environment. The sorting result is the Tiebout equilibrium, in which no one has the incentive to move.

Ideally, researchers should conduct a randomized experiment by randomly assign people to different places with different zoning and built environment conditions and then compare people's active living outcomes. However this kind of research is unrealistic. The data researchers can get are the observed outcomes of the Tiebout equilibrium. The observed built environment and zoning are people's choice results according to people's active living preferences. People's preferences are unobservable to researchers. they hide in the error term of regression. This correlation causes estimated coefficient inconsistencies and biases the estimated

[12] For theoretical rigor, there are three assumptions in Tiebout sorting: every economic agent has the same and full information regarding the whole market; economic agents are free to move; every economic agent faces the same schedule of housing prices.

coefficients of zoning and built environment, which was referred to as the "Tiebout bias" by Goldstein and Pauly (1981). The Tiebout bias is essentially omitted variable bias.

The story does not end here. Tiebout sorting and Tiebout equilibrium are in demand side in the sense that they are resulted from people's preferences on zoning and built environment. The sorting will further affect the supply of zoning. Zoning provision was determined by local officials. However, it can be largely affected by the public through voting at local elections and public hearings held by local administrative agencies that are established under state-enabling laws (Nolon 2006). Fischel (2004) outlined the twentieth-century history of American zoning and found that homeowners dominate zoning content and administration in most jurisdictions through voting. In our model, the migrated people can submit petitions to challenge the local zoning law, attend public zoning hearings held by the planning commission, or vote in local elections. Eventually, they will determine the zoning content based on their active living preferences. Places with relatively more people who favor active living will have more active living-oriented zoning and built environment than those with relatively few people who favor active living. The final provision of zoning and built environment is the market equilibrium determined by the intersection of supply and demand, and the demand is Tiebout equilibrium. Both supply and demand are affected by Tiebout sorting that is driven by unobserved preferences. The data researchers observe is the outcome of market equilibrium, which means the endogeneity problem biases the estimated coefficients of zoning and built environment.

The exposition above is illustrated in Figure 5.1.1.

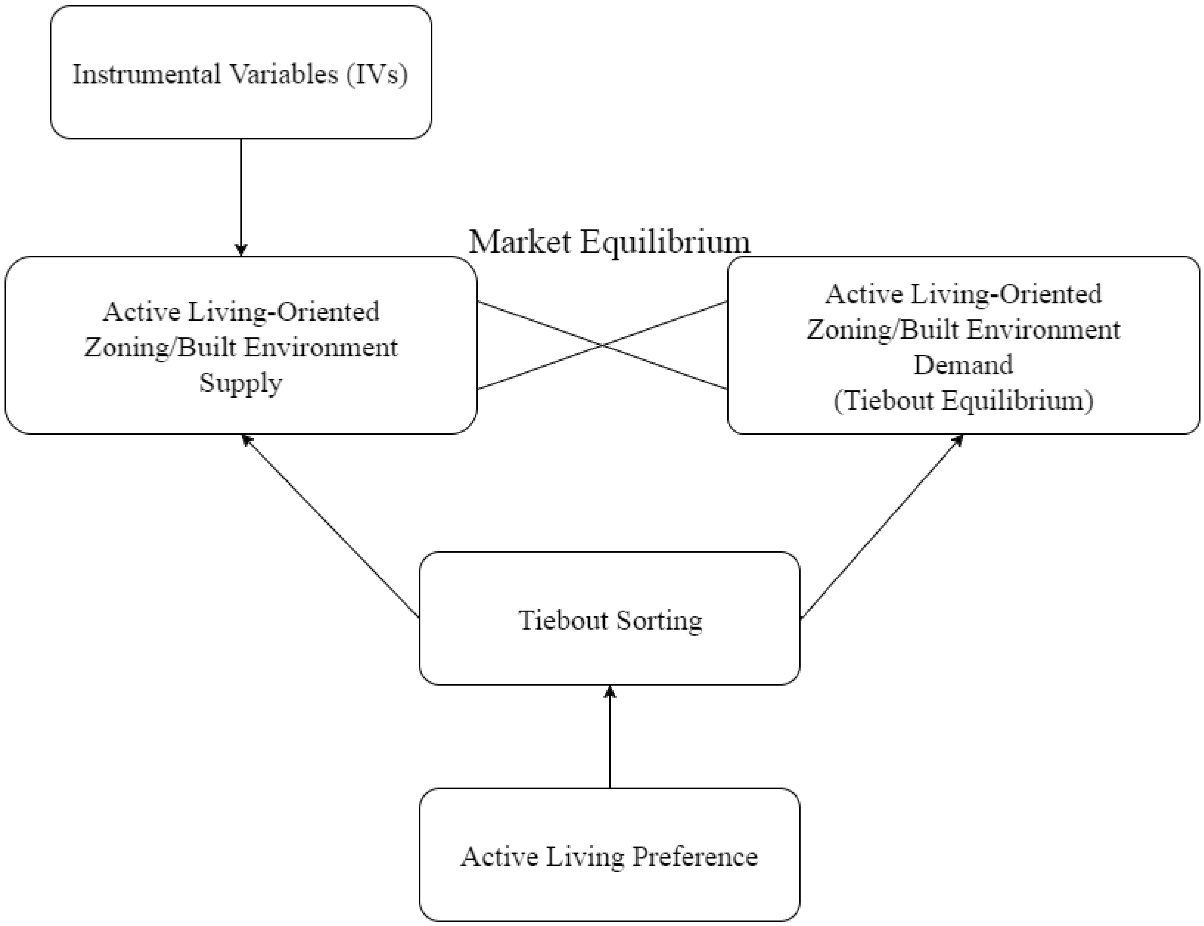


Figure 5.1.1 Tiebout Sorting Illustration

5.2 Traveling Behavior and Sorting

Traveling behavior is different from leisure-time physical activity or sedentary behavior. Leisure-time physical activity and sedentary behavior do not have destinations, and they are purely done for recreation. Goals determine traveling behavior. People typically choose the fastest route or means of transportation that they can afford. People's active living preference play little or no roles in their travel decisions.

The sorting problem still exists for traveling behavior. Rather than based on active living preference, the sorting relies on people's choices of destinations. Consider travel to work. It is an integral part of people's job plans. People choose job attributes (compensation, working environment, etc.), living location, and transportation methods to work simultaneously based on their qualifications, budgets, and preferences. It may not be the zoning or built environment that causally impacts people's work transportation plans. Instead, people choose their traveling methods to work when making job and housing decisions. Consider travel related to education. When adults make housing choice, they take their children's schooling into account. The transportation methods to school are determined together with school selection and housing location based on children's qualifications, budgets, educational preferences toward, and so on.

The destination-reliance of traveling behavior makes it more likely to be the result of self-selection than to be the causal result of zoning and built environment. People decide on which transportation methods to use to get to different destinations when they make housing location choices based on their preferences for various destinations before they move to their houses and travel. Moreover, the sorting problem for traveling behavior is more complicated than leisure-time physical activity and sedentary behavior because various preferences about different destinations are involved. The causal impact of zoning and built environment on traveling behavior requires thorough investigation and could be a research question in the future. In this thesis, I focus on leisure-time physical activity and sedentary behavior.

5.3 Aerobics Regression and the Analysis of Active Living Preference

The active living outcomes include two variables, physical activity, and sedentary behavior. Physical activity includes walking, running, and bicycling. The definition of physical activity is based on the definition of active living-oriented zoning, which mainly consists of pedestrian- and cyclist-friendly zoning provisions. It also follows the literature, which shows a significant association between walking and bicycling, and active living-oriented zoning and built environment. As discussed in previous sections, walking and bicycling can be affected by both policy variables and unobserved active living preference. Nonetheless, some physical activities can be affected by active living preference but are immune to zoning and built environment. Such physical activities may be mainly practiced inside gymnasiums, which do not belong to active living-oriented zoning provisions or community- and street-scale land-use and design. If there exists a significant effect of zoning and built environment on these physical activity outcomes, which should not exist, then this may imply that the observed zoning and built environment data are the choice results of people based partly on their active living preferences (i.e., sorting).

I need to find a comparable physical activity to conduct the analysis. The physical activity should be affected by the same unobserved preferences as walking, running, and bicycling, but immune to zoning and built environment. Human’s preferences are very complicated. Physical activity preferences can be further classified, such as preferences for basketball, football, and muscle training. A basketball fan may not like football at all, and a muscle trainer may not like running or bicycling.

The comparable physical activity here is aerobics. Aerobic refers to the use of oxygen to adequately meet energy demands during exercise via aerobic metabolism. The most typical aerobics include aerobic exercises (e.g., running/jogging, cycling, walking) and aerobic classes (e.g., Zumba). People who like walking, running, and bicycling are likely to be interested in aerobic exercises. Different from the walking, running, and bicycling that is performed outside of rooms, aerobic exercises are usually performed inside gymnasiums with machines (treadmills, elliptical trainers, glider machines, and stationary bicycles). Sometimes they are led by fitness instructors.

Table 5.3.1 shows the regression analysis on aerobics. The models and the covariates are the same as the mediational analysis for physical activity and sedentary behavior. All the coefficients are insignificant. Does this imply there is no sorting?

It is difficult to conclude that there is no sorting. Although exercises are similar, running, walking, bicycling outside rooms and aerobics are fundamentally different from aerobics inside gymnasiums. People who like jogging along streets or walking around communities may have no interest in running on machines inside gymnasiums. Rather than physical activity preferences, the unobserved preference here is active living preference. It is supposed to be that the individual integrates physical activity into the daily life in his/her environment, which is different from intentionally going to gymnasiums to do aerobic exercises, even though the activities are more or less the same (walking, running, and cycling). I proceed to the analysis that takes sorting into account.

Table 5.3.1 The Mediational Analysis, Aerobics Regression

	Step 1				Step 3			
	Probit		OLS	Log OLS	Probit		OLS	Log OLS
VARIABLES	COEFF	Marginal Effect	COEFF	COEFF	COEFF	Marginal Effect	COEFF	COEFF
Active Living-Oriented Zoning	-0.082	-0.001	-9.387	-0.284	-0.213	-0.002	-27.266	-0.591
	(0.252)	(0.002)	(18.621)	(0.330)	(0.277)	(0.002)	(20.349)	(0.359)
Walkability					0.025	0.000	2.167	0.037
					(0.017)	(0.000)	(1.356)	(0.023)
Observations	24,458	24,458	66	66	24,458	24,458	66	66

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 1.

5.4 Instrumental Variable Identification

The solution to the endogeneity problem is to find observed variables as instrumental variables (IVs). Valid IVs extracts the exogenous variations from endogenous explanatory variables that have been confounded by unobserved explanatory variables. In our case, valid IVs can isolate the supply of zoning, at least in part, from the observed market equilibrium outcomes that are partially driven by the unobserved preferences (i.e., sorting). Valid IVs need to satisfy two conditions. The first condition is exclusion restriction; IVs have no association with people's active living preferences conditional on other exogenous explanatory variables. The second condition is that IVs are highly associated with the endogenous variable conditional on all other exogenous explanatory variables. To sum up, valid IVs have no association with individuals' active living outcomes, except through the endogenous variables, conditional on all other covariates.

Because I have two endogenous variables, zoning, and built environment, I need at least two IVs for identification. The IVs are manufacturing establishment density, the ratio of the number of manufacturing establishments to the size of the total land surface, and farmland proportion, the ratio of farmland to the total land surface. Both are county-level variables. They are constructed using *1900 Census: Population, Housing, Agriculture, and Manufacturing Data*.

The finding and construction of IVs are inspired by the early history of American zoning. The first zoning ordinance, which was in New York City in 1916, was aimed to protect residents from rapidly-developing industrial buildings and activities. Traditional Euclidean zoning originated from the zoning plan that was adopted by the Village of Euclid, Ohio in 1922 to prevent industrial development from encroaching on the countryside characteristics of the community. The Village of Euclid, Ohio, at that time, featured residential nature farmland and plenty of undeveloped land. Its zoning plan separated residential and industrial land uses into different districts and put restrictions on the features of buildings (lot area, size, height, etc.) in those districts. It was challenged by the Ambler Realty Company because a tract of undeveloped land the company purchased would have been used for commercial or industry businesses, which would bring more interest to the company, as the company claimed, but could only be used for residential and community uses under the zoning plan. The Supreme Court supported the Village of Euclid in 1926 with the rationale that Euclid's zoning protected the residents of the village from the danger of fire, contagion, and disorder brought by stores, shops, and factories, which led to the constitutionality of zoning (Schilling and Linton 2005; Nolon 2006; Fischel 2004).

The late 19th century and early 20th century was a time when the second wave of the Industrial Revolution, featuring advancements in manufacturing and production technology, took place in the United States. There was a fierce conflict over land use between the rapidly-developing manufacturing industry and the traditional agricultural economy. The newly-emerging industry forces tried to expand their territory to make profit. Farmers tried to protect the nature of countryside from the pollution and congestion the industry buildings would bring. The game between the two forces gave birth to traditional Euclidean zoning. The two forces did not intentionally manipulate pedestrian- and cyclist-friendly built environment and zoning provisions. However, what they did at that time objectively created the first pedestrian- and cyclist-friendly zoning provisions. To expand their territory, industry people asked for densely distributed population (workers), mixed-use spaces, and compactly-designed communities, all of which created ideal conditions for infectious disease crises at that time, when sanitation was inadequate. However, it objectively led to pedestrian- and cyclist-friendly built environment and zoning provisions. Mixed-use spaces directly resulted in the mixed-use zoning provision. Densely distributed populations and compactly-designed communities resulted in pedestrian- and cyclist-oriented zoning provisions, such as sidewalks, crosswalks, street connectivity, bike-pedestrian connectivity, bike-pedestrian trails-paths, bike lanes, bike parking, and other walkability provisions. Conversely, the nature of farmland that farmers wanted to defend featured sparsely distributed buildings, poor connectivity, and long distances between places, all of which led to few pedestrian- and cyclist-friendly zoning provisions. The geographic variations of manufacturing establishment density and farmland proportion represented the original geographic variation of pedestrian- and cyclist-friendly built environment to some extent. In addition, it caused the geographic variation of the strength comparison between farmers and industry people in the early 20th century, which objectively led to the geographic variation of pedestrian- and cyclist-friendly zoning provisions and further affected active living-oriented zoning and built environment afterward.

The most relevant IV identification in the literature is the study of Zhao and Kaestner (2010), which is illustrated in Figure 5.4.1. The authors estimate the causal effects of population density in 1970, 1980, 1990, and 2000 on BMI and obesity from 1976 to 2001 and use the number of planned highway rays emanating from the largest central city of an MSA in *1947 National System of Interstate Highways Plan* as IV for population density. In my study, population density, together with three other density metrics, constitute the variable "walkability,"[13] which is a metric for built environment. I illustrate my identification in Figure

[13] The variable "walkability," which represents built environment, is standardized summation of four density metrics, including population density, housing unit density, the ratio of four-way intersections to all intersections, and the total number of intersections divided by land area.

5.4.2. Although Zhao and Kaestner (2010) and I have different the outcome variables, which implies that the unobserved confounders and the causal mechanism may be different, it is still interesting to compare our IVs. All of us use the geographic variations of historical macro-level physical infrastructure as the source of exogenous variation. The features of time advancement and macro-level metrics make the IVs barely have any association with individual preferences. The differences are the sources of the variation of IVs and the effects of IVs on the current socioeconomic environment. The variation of the number of planned highway rays in the *1947 Highways Plan* is the choice result of humans, and it still has large impacts on today's socioeconomic environment. Rather, the variation of manufacturing establishment density in 1900 is the consequence of a series of historical events. By going through the history of the Industrial Revolution in United States and seeing how the manufacturing industry emerged and developed in the United States during at the latter half of 19th century, I find that the variation can hardly be the choice result of humans. Moreover, farmland proportion in 1900 is likely to be determined by natural resources. The manufacturing industry today is tremendously different from the manufacturing industry in 1900. The structure and the content of today's economy have changed tremendously since 1900. The manufacturing industry and farmland in 1900 have few associations with today's economy. I discuss the IVs in a detailed way in the following sections.

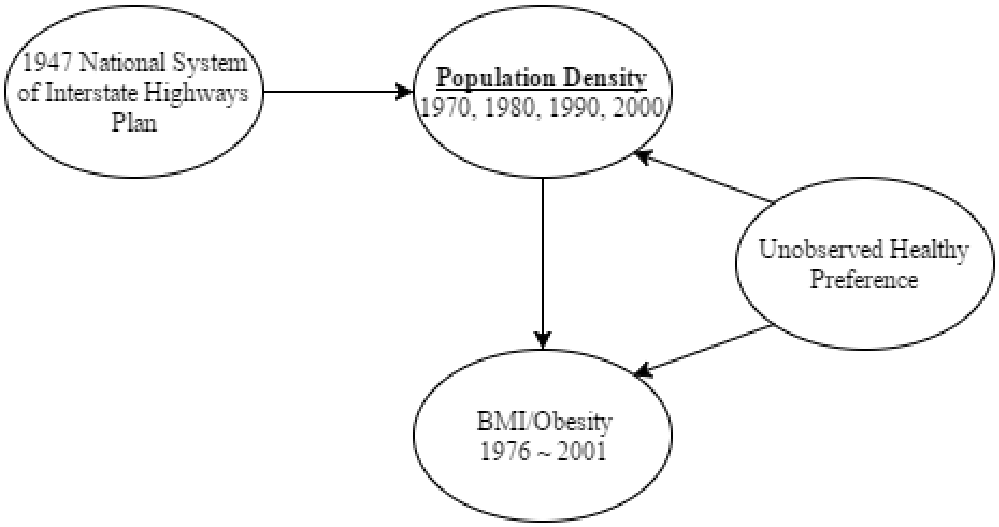


Figure 5.4.1 The Illustration of IV Identification of Zhao and Kaestner (2010)

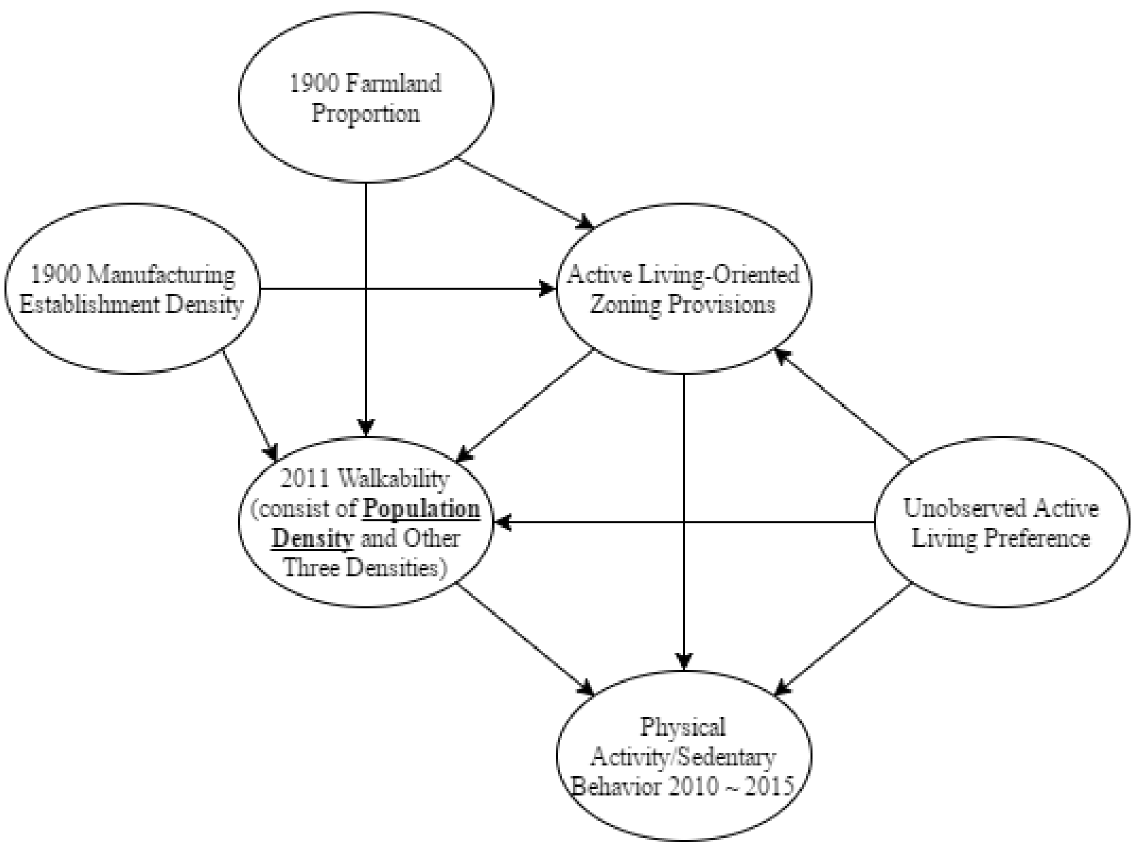


Figure 5.4.2 The Illustration of IV Identification

One concern is that the endogenous variables and IVs are at the county-level but more than a century apart in time and county boundaries change over time. The zoning code and walkability data is from 2010 while the data about manufacturing establishment density and farmland proportion are from 1900. Will the changes in county boundaries over the century have any negative impacts on our analysis?

Counties were usually huge in their early periods. Because the populations in counties increased, the original counties were divided into several smaller counties. New counties were created from one or more small counties. If the counties in 2010 are parts of the original ones from 1900, then the causal chain is still clear. The manufacturing establishment density and farmland proportion in 1900 affected the zoning and built environment in the original county at that time, and the zoning and built environment in the original county affected the zoning and built environment in any part of the original county in 2010. If the county in 2010 is greater in size than the original one in 1900 because incorporated small counties into its domain over time, the causal chain depends on the size comparison between the areas it integrated into its domain and its original area. The areas it integrated into its domain may bring different zoning and built environment that they weaken the correlation between its original zoning and built environment in 1900 and its zoning and built environment in 2010, which implies that the instrumental variables are weak.

By merging all the data, I can successfully identify 251 counties in total. I compare the land surface in square miles based on the data in the 1900 and 2010 censuses for each county. Among the 251 counties, 84 counties increased in terms of land surface, five counties did not change, and 162 counties reduced their sizes. I conduct a paired t-test. The test supports that the land surface for each county is statistically significantly smaller in 2010 than in 1900 at the 0.1% significance level (with a p-value of 0.0028).

The concern that the change of county boundaries causes the instrumental variables to be weak can also be cleared by the IV first-stage regressions that show no weak identification and underidentification.

5.5 The Exclusion Restriction

Valid IVs cannot have any association with unobserved confounder conditional on other covariates in the analysis. This is called exclusion restriction. It is impossible to test the exclusion restriction rigorously, so I have to make it as an assumption. Here, I assume that the 1900 manufacturing establishment density and the 1900 farmland proportion are uncorrelated to individual active living preference from 2010 to 2015 conditional on all controls. This condition implies four assumptions. If any one of the four assumptions fails to hold, then the exclusion restriction will fail to hold and the IVs will be invalid and the IV estimators will be inconsistent.

5.5.1 Assumption 1: IVs Have No Direct Associations with Individual Active Living Preference

The unobserved confounder in our analysis is the active living preference of survey respondents in ATUS 2010-2015. Because the IVs are the county-level manufacturing establishment density in 1900 and the county-level farmland proportion in 1900, it is difficult to imagine that these two factors have any association with individual’s active living preference over 100 years later. The things IVs represent happened more than a century before the respondents took the ATUS survey.

5.5.2 Assumption 2: The Causal Chain Goes From IVs to Endogenous Variables

The invalidation of IVs can also happen if the causal chain goes from endogenous variables to IVs so that unobserved confounders affect IVs through changing endogenous variables. However, this is not true here. The two IVs, manufacturing establishment density and farmland proportion, are variables from 1900. The first zoning ordinance, a citywide regulation in New York City, took place in 1916. In 1922, the Department of Commerce published the model law for U.S. states to enable land-use regulations in their jurisdictions ("Standard State Zoning Enabling Act"). In 1926, the constitutionality of zoning was upheld by the U.S. Supreme Court in the case of Village of Euclid, Ohio v. Ambler Realty Co. In 1928, the Department of Commerce published the *Standard City Planning Enabling Act* (SCPEA) (Fischel 2004; Schilling and Linton 2005; American Planning Association 2006; Nolon 2006). All the land regulations happened later than 1900. Therefore, the causal chain can only go from IVs to endogenous variables.

5.5.3 Assumption 3: The Sources of Variation of IVs Have No Associations with Individual Active Living Preference

The geographic variation of farmland proportion in 1900 came from natural resources. The United States began as a mostly rural nation, with most people living on farms or in small towns and villages. Even in the second half of the 19th century, when the vast expansion of manufacturing industrial plants took place, a majority of Americans relied on agriculture to make their livings in 1900[14] ("Rural Life in the Late 19th Century" 2017). Farmland proportion is a different concept from agricultural productivity. Farmland proportion is primarily determined by the weather patterns (temperature, precipitation, etc.), the quality of soil, the presence of pests, and other natural conditions.[15]

The geographic variation of manufacturing establishments in 1900 came from natural resources (Kim 1995; Kim 1999) and market access (Krugman 1991a; Krugman 1991b; Krugman and Venables 1995; Klein and Crafts 2012). Market access is first addressed in the studies of Krugman (1991b) and Krugman and Venables (1995). They pointed out that the manufacturing industry tends to concentrate on places where it can easily access the intermediates and quickly deliver the final goods. In the second half of the 19th century, market access was determined by the railroad network, which dramatically lowered transportation costs to areas without access to navigable waterways. The railroad construction in the U.S. dramatically increased in the early 1870s and was almost completed by 1900. The locations of railroads were primarily determined by Congress through land grants. From the Congress, four of the five transcontinental railroads were assured the land to lay the tracks and the land to sell to finance the construction. Some small railroads purchased land to lay their tracks from private owners[16] ("Railroads in the Late 19th Century" 2017).

---

[14] Please see *Library of Congress*. http://www.loc.gov/teachers/classroommaterials/presentationsandactivities/presentations/timeline/riseind/rural/ (accessed September 21, 2017).

[15] Farmland proportion is different from agricultural productivity. Agricultural productivity is determined by not only natural resources but also agricultural technology. In the late 19th century, the yield per unit area of farmland increased substantially because of mechanical improvements.

[16] Please see *Library of Congress*. http://www.loc.gov/teachers/classroommaterials/presentationsandactivities/presentations/timeline/riseind/railroad/ (accessed June 6, 2017).

According to Krugman (1991a), when the United States was in an agricultural era, transportation was costly, and the manufacturing industry lacked economies of scale. When the Industry Revolution began, production increased in a few areas where farming populations were concentrated but outside the South. When it came to the 1860s, the coming of railroads made transportation costs fall, technology progress enabled increasing returns for manufacturing, and the nonagricultural population rose. These three factors led to the initial spatial concentration of the manufacturing industry in a small part of the Northeast and the eastern part of the Midwest, which is referred to as the "manufacturing belt" in numerous studies. Geography is path dependent. "In 1900, about $4/5^{th}$ of American manufacturing output was produced in this part of the country which comprised only $1/6^{th}$ of its land area" (Klein and Crafts 2012). Krugman (1991a) comments on the works of other economic historians: "the history of manufacturing location says… Nicholas Kaldor (1972), Paul David (1985), and Brian Arthur (1989) were right -- that increasing returns and cumulative processes are pervasive and give an often decisive role to historical accident." The locations of manufacturing establishments in 1900 were consequences of a series of historical events and can hardly be the results of human choice.

By 1900, five transcontinental railroads connected the Eastern states and the Pacific Coast. As the railroads opened up new areas of the West for settlement, the farmland proliferated throughout the second half of the 19th century. However, the expansion of railroads to the West did not bring the significant concentrations of manufacturing industry in the West. This proves that the locations of manufacturing establishments in 1900 were path dependent and the results of historical contingency and inevitability. This also shows that manufacturing establishment density in 1900 and farmland proportion in 1900 provide independent information.

By the way, the zoning and built environment data in this thesis, as discussed in the introduction, exclusively refer to community-scale and street-scale urban design and land-use policies and practices. They have nothing to do with railroads or highways.

5.5.4 Assumption 4: The Uncontrolled Channels from IVs to Active Living Preference Are Weak Enough to Be Negligible

The two IVs took place more than a hundred years ago. During these one hundred years, numerous things happen. Some things can be the possible channels that connect IVs to active living preference. For example, education could be a possible channel. Places of high historical manufacturing establishment density may be rich or may have plenty of educational institutions that can grant graduate degrees or may require a large number of talented people who have graduate degrees. People with high education levels may gain a lot of health knowledge and be clearly aware of the importance of physical activity on health. Therefore, they may have stronger active living preferences than people with low education levels. However, as long as education is controlled for in the analysis, IVs and active living preference are conditional uncorrelated in the case that education is the only channel that connects them. To make the exclusion restriction suffice, I assume the uncontrolled channels from IVs to active living preference are weak enough to be negligible. The assumption is that either the correlation between the IVs and the uncontrolled channels are very weak, or the correlation between the uncontrolled channels and active living preference are very weak.

Some facts show that the possible channels that go from IVs to active living preference are very weak. First, the manufacturing establishments in 1900 have few associations with today's economy. The manufacturing's share of total U.S. employment has declined steadily over the last 50 years. Many U.S. corporations continue to shift their production facilities overseas (Baily and Bosworth 2014). The once prosperous manufacturing industrial cities in the "manufacturing belt," which were among the largest in the U.S. before World War II, lost their manufacturing industries. For example, Detroit, Michigan filed for bankruptcy in 2013, Chicago, Illinois now is famous for its well-developed financial industry. In addition, they lost their population the most in the country by the end of the 20th century (Hansen, Bryant, and Spencer 2007). For example, based on the census data, from 2000 to 2016, Gary, Indiana lost 25.6% of its population, Flint, Michigan lost 22.1% of its population, Youngstown, Ohio lost 21.6% of its population, and Buffalo, New York lost 12.2% of its population.

Second, the content of the manufacturing industry in 1900 was very much different from that of the current manufacturing industry. Although a system of distributing electrical power was invented in 1880 and electric street railways had been in cities since 1888, because of the low conversion efficiency of fuel to power and the small scale of power plants, electricity was only offered at nighttime and was too expensive to have a significant impact on the economy at that time. Daytime electricity service became common during the early 20th century after the introduction of the AC motor. Steam turbines and internal combustion engines dominated the manufacturing industry. Many aspects of manufacturing that people are now familiar with happened after 1900. The first car was made in the U.S. in 1908. The first moving assembly line was installed in the U.S. in 1913. Nowadays, the manufacturing industry in the U.S. features robots, automation, the IoT (internet of things, direct interconnections over the internet among machines and locations), material science, biotechnology, and artificial intelligence. The textile machines, sewing machines, and steam locomotives of 1900 only exist in museums and antique shops right now. People who work in manufacturing establishments today have completely different working skills, working content, and labor loads than the workers in manufacturing facilities in 1900.

5.6 Covariate Balance Test

In the literature of IV, researchers often conduct covariate balance tests to examines the association between IV and observed characteristics to informally argue for the exogeneity of IVs (Altonji, Elder, and Taber 2002). Suppose the IV is dichotomous. Researchers often split the data of observables into two groups based on the binary values of IVs and conduct mean comparisons between the two groups to show the randomness of IVs (see Angrist and Evans (2017) for an example; the footnotes of Altonji, Elder, and Taber (2005) also provides a list of examples). Researchers believe that if they observe high correlations between IVs and some observables, then they have to be watchful for the correlations between IVs and unobservables.

However, it is unclear why the high correlations between IVs and observables can be a signal for the violation of exclusion restriction. A non-randomly assigned IV that is correlated with observables can still satisfy the exclusion restriction because the exclusion restriction only asks for uncorrelatedness between the IV and unobservables conditional on observables. A randomly assigned IV that is uncorrelated with observables can still violate the exclusion restriction. For example, Evans and Schwab (1995) used Catholic affiliation as an IV to estimate the causal effects of Catholic schooling. Even though being Catholic is randomly assigned, the Catholic faith can change people’s preferences and attitudes toward education, working, and living and regulate people’s behavior in daily life. The randomness of IV only implies the causation of the first-stage regression in two-stage least squares. It has nothing to do with the exclusion restriction (Angrist, Imbens, and Rubin 1996; Imbens 2007).

Here I follow the literature of IV to do correlation analysis on the association between IVs and other covariates used in the mediational analysis (excluding zoning, walkability, and the group of weekday dummies, monthly dummies, and yearly dummies). Table 5.6.1 and Table 5.6.2 show the results. Table 5.6.1 shows that the manufacturing establishment densities in1900 are uncorrelated with most of the individual characteristics except non-Hispanic Black, single, high school graduate, some college, and graduate. Table 5.6.2 shows that the farmland proportions in 1900 are uncorrelated with most of the individual characteristics, except Hispanic, single, less than high school, and high school graduate.

The covariate balance test shows that the two IVs are not randomly distributed. The association between covariates and IVs are plausibly due to the fact that most covariates are intentionally selected based on their high correlations with outcome metrics to reduce bias. As discussed in Section 5.5.4, they could be channels from IVs to active living preference. Nonetheless, by controlling for them, the conditional uncorrelatedness between IVs and active living preference can still be valid in the case that those covariates represent all the strong channels. In fact, the test reminds us to be watchful for the uncontrolled channels. To make the exclusion restriction suffice, it is important to maintain the assumption that the uncontrolled channels from IVs to active living preference are weak enough to be negligible.

To sum up, the covariate balance test has few implication for the exclusion restriction. Passing the test, which means covariates are uncorrelated with IVs, does not imply that the IVs satisfy the exclusion restriction. Failing to pass the test, which means covariates are correlated with IVs, does not imply that the IVs do not satisfy the exclusion restriction. Regardless of whether the IVs pass the test, I have to maintain the assumption from Section 5.5.4—that the uncontrolled channels from IVs to active living preference are weak enough to be negligible.

Table 5.6.1 Correlation Analysis on Manufacturing Establishment Density 1900 and Other Covariates

Manufacturing Establishment Density 1900	Correlation Coefficients	Group 1 Mean of IV	Group 2 Mean of IV	Mean Difference	t-statistic	p-value
Age	-0.001					0.889
Male		13.151	12.209	0.942	1.285	0.199
Non-Hispanic Black		10.682	23.915	-13.233	-11.157	0
Non-Hispanic Other Race		12.593	14.597	-2.004	-1.285	0.199
Hispanic		12.368	14.275	-1.907	-1.88	0.06
Widow, Divorced, Separated		12.808	12.516	0.292	0.361	0.718
Single		9.388	21.729	-12.341	-12.176	0
Any Difficulties		12.744	12.623	0.121	0.106	0.915
Less than High School		12.927	11.583	1.344	1.442	0.149
High School Graduate		13.208	11.127	2.081	2.641	0.008
Some College		13.577	8.834	4.743	5.967	0
Graduate		11.156	22.866	-11.71	-7.8	0
% Households in Poverty	0.173					0
% Non-Hispanic White	-0.132					0
% Non-Hispanic Black	0.139					0
% Hispanic	0.01					0.116
Median Household Income	0.038					0
Median Age	-0.072					0
Land Area (Square Miles) 2010	-0.126					0

Note: For the continuous variable, I calculate the correlation coefficient between Manufacturing Establishment Density 1900 and the variable, and perform t-test directly on the correlation coefficient. For the discrete variable, I do mean comparison of Manufacturing Establishment Density 1900 for Group 1 (the variable equals to 0) and Group 2 (the variable equals to 1), and perform t-test on group means with unequal variance.

Table 5.6.2 Correlation Analysis on Farmland Proportion 1900 and Other Covariates

Farmland Proportion 1900	Correlation Coefficients	Group 1 Mean of IV	Group 2 Mean of IV	Mean Difference	t-statistic	p-value
Age	0.003					0.676
Male		0.587	0.586	0.001	0.337	0.736
Non-Hispanic Black		0.586	0.587	-0.001	-0.019	0.985
Non-Hispanic Other Race		0.586	0.587	-0.001	-0.071	0.944
Hispanic		0.61	0.488	0.122	21.692	0
Widow, Divorced, Separated		0.589	0.579	0.01	2.032	0.042
Single		0.595	0.563	0.032	6.553	0
Any Difficulties		0.586	0.592	-0.006	-0.885	0.376
Less than High School		0.592	0.553	0.039	6.211	0
High School Graduate		0.583	0.596	-0.013	-2.538	0.011
Some College		0.587	0.582	0.005	0.96	0.337
Graduate		0.586	0.591	-0.005	-0.899	0.369
% Households in Poverty	-0.243					0
% Non-Hispanic White	0.29					0
% Non-Hispanic Black	0.031					0
% Hispanic	-0.291					0
Median Household Income	0.146					0
Median Age	0.096					0
Land Area (Square Miles) 2010	-0.397					0

Note: For the continuous variable, I calculate the correlation coefficient between Farmland Proportion 1900 and the variable, and perform t-test directly on the correlation coefficient. For the discrete variable, I do mean comparison of Farmland Proportion 1900 for Group 1 (the variable equals to 0) and Group 2 (the variable equals to 1), and perform t-test on group means with unequal variance.

5.7 IV Asymptotical Bias and the Link between Selection on Observables and Unobservables

Altonji, Elder, and Taber (2005a) proposed a procedure with which to test the exogeneity of the suspect IV. They set up the null hypothesis that the suspect IV satisfies the exclusion restriction. Under the null hypothesis, they calculated the asymptotical bias from the suspect IV. If the asymptotical bias is "substantially different from zero," then "one may be worried that the null hypothesis … is wrong" (Altonji, Elder, and Taber 2005a).

Here, I show how they derived the formula for the asymptotical bias of IV. Suppose $Y = \alpha X_1 + X_2'\gamma + \xi$ where Y is the outcome of interest, X_1 is the potential endogenous variable, X_2 is observables, ξ is unobservables, and γ is defined so that $Cov(X_2, \xi) = 0$, and Z is the suspect IV, they defined β, π, and λ to be the coefficients of the least squares projections

$$proj(Z|X_2) = X_2'\pi, \tag{6}$$

$$proj(X_1|X_2, Z) = X_2'\beta + \lambda Z. \tag{7}$$

Define $\tilde{Z}$ as the residual of the projection of Z on X_2 so that

$$\tilde{Z} \equiv Z - X_2'\pi. \tag{8}$$

They had

$$\hat{\alpha}_{IV} \overset{p}{\rightarrow} \alpha + \frac{Cov(\tilde{Z}, \xi)}{\lambda Var(\tilde{Z})} \tag{9}$$

where $\hat{\alpha}_{IV}$ is the IV estimate of α, and $\frac{Cov(\tilde{Z},\xi)}{\lambda Var(\tilde{Z})}$ is the asymptotical bias.

To estimate $Cov(\tilde{Z}, \xi)$, they utilized a condition,

$$\frac{Cov(Z, X_2'\gamma)}{Var(X_2'\gamma)} = \frac{Cov(Z, \xi)}{Var(\xi)} \tag{10}$$

so that they can derive an estimable formula for the asymptotical bias

$$\frac{Cov(\tilde{Z},\xi)}{\lambda Var(\tilde{Z})}=\frac{Cov(Z-X_2'\pi,\xi)}{\lambda Var(\tilde{Z})}=\frac{Cov(Z,\xi)-\pi Cov(X_2',\xi)}{\lambda Var(\tilde{Z})}$$

$$=\frac{Cov(Z,\xi)}{\lambda Var(\tilde{Z})}=\frac{Var(\xi)}{\lambda Var(\tilde{Z})}\frac{Cov(Z,\xi)}{Var(\xi)}$$

$$=\frac{Var(\xi)}{\lambda Var(\tilde{Z})}\frac{Cov(Z,X_2'\gamma)}{Var(X_2'\gamma)} \tag{11}$$

The arguments $Var(\tilde{Z})$, $Var(X_2'\gamma)$, and λ can be directly estimated. Under the null hypothesis that IV satisfies the exclusion restriction, one can get consistent estimates for γ and $Var(\xi)$ through IV estimation. To get a consistent estimate for $Var(\xi)$, one can estimate $\hat{\xi}=Y-\hat{\alpha}_{IV}X_1+X_2'\hat{\gamma}_{IV}$, and then estimates $Var(\hat{\xi})$.

There are some limitations of this procedure. First, it is an incomplete hypothesis testing procedure. There is neither control for type I error nor derivation of the asymptotical distribution of the estimated asymptotical bias. The estimated asymptotical bias is sample-dependent. A large magnitude of the estimated asymptotical bias can happen by chance. It does not imply that the result is inconsistent with the null hypothesis that IV is exogenous. It is better to derive the asymptotical distribution of the estimated asymptotical bias and follow the standard hypothesis testing procedure to estimate the p-value. Second, to get consistent estimate for $Var(\xi)$, Altonji, Elder, and Taber (2005a) implicitly assumed the homoscedasticity of ξ. The heteroscedasticity of ξ will invalidate the procedure. Third, the correlations between unobservables and the suspect IV are likely to be weaker than the correlations between observables and the suspect IV (Altonji, Elder, and Taber 2005a), so the estimated asymptotical bias is an upper bound for the true asymptotical bias. This may invalidate the procedure because a large estimated upper bound does not mean the true parameter is large or non-zero.

According to Altonji, Elder, and Taber (2005b), there are two reasons for the possibility that the associations between observables and IV are stronger than the associations between unobservables and IV. One is that observables are often selected by intention to reduce bias. They can be highly correlated with outcome metrics and the endogenous variable and, therefore, highly correlated with the suspect IV. However, the condition these authors utilized, $\frac{Cov(Z,X_2'\gamma)}{Var(X_2'\gamma)} = \frac{Cov(Z,\xi)}{Var(\xi)}$, relies on the assumption that the observables are randomly selected. The other is that the error term ξ includes both unobserved confounder and independent white noise. Suppose $\xi = \xi_1 + \xi_2$, where ξ_1 is unobserved confounder and ξ_2 is independent white noise. Altonji, Elder, and Taber (2005a) derived the updated condition

$$\frac{Cov(Z,X_2'\gamma)}{Var(X_2'\gamma)} = \frac{Cov(Z,\xi_1)}{Var(\xi_1)}. \tag{12}$$

Because $Var(\xi) > Var(\xi_1)$ and $Cov(Z,\xi) = Cov(Z,\xi_1)$, they can derive

$$\frac{Cov(\tilde{Z},\xi)}{\lambda Var(\tilde{Z})} = \frac{Var(\xi)}{\lambda Var(\tilde{Z})}\frac{Cov(Z,\xi)}{Var(\xi)} < \frac{Var(\xi)}{\lambda Var(\tilde{Z})}\frac{Cov(Z,\xi_1)}{Var(\xi_1)} = \frac{Var(\xi)}{\lambda Var(\tilde{Z})}\frac{Cov(Z,X_2'\gamma)}{Var(X_2'\gamma)} \tag{13}$$

which implies the estimated asymptotical bias is greater than the true asymptotical bias.

The condition $\frac{Cov(Z,X_2'\gamma)}{Var(X_2'\gamma)} = \frac{Cov(Z,\xi)}{Var(\xi)}$ relies on three strong assumptions. The condition is based on the idea that the associations between observables and the suspect IV can provide some information about the associations between unobservables and the suspect IV. Altonji, Elder, and Taber (2002) provided the assumptions that will lead to the condition. Here, I list the intuition of the assumptions. The intuition comes from the study of Altonji, Elder, and Taber

(2005b) where they utilized a similar condition $\frac{Cov(X_1,X_2'\gamma)}{Var(X_2'\gamma)} = \frac{Cov(X_1,\xi)}{Var(\xi)}$ for the potential endogenous variable to provide a lower bound of the effect of the potential endogenous variable. According to Altonji, Elder, and Taber (2002), the assumptions of the condition for the suspect IV are almost the same as the assumptions about the condition for the potential endogenous variable, except they replaced X_1 with Z.

1. The elements of observables are chosen at random from the full set of factors, including both observables and unobservables, which determine the outcome of interest.
2. The number of observables and factors (including both observables and unobservables) are large, and none of the elements dominate the distribution of IV or the outcome of interest.
3. Suppose one can write $Y = \alpha X_1 + X_2'\gamma + \xi$, where Y is the outcome of interest, X_1 is the endogenous variable, X_2 is observables, and ξ is unobservables. The regression of X_1 on $Y - \alpha X_1$ is equal to the regression of the part of X_1 that is orthogonal to X_2 on the corresponding part of $Y - \alpha X_1$.

Altonji, Elder, and Taber (2005b) used many words to argue that these assumptions "are no more objectionable than" the exogeneity in Gauss-Markov assumptions, but they admitted that these assumptions are strong and unlikely to hold.

Table 5.7.1 shows 2SLS estimates of zoning effects and estimates of potential bias using the methodology of Altonji, Elder, and Taber (2005a). Most of the 2SLS coefficients have huge standard errors, which makes them statistically insignificant suppose they asymptotically follow normal distribution. The huge standard errors may come from either the inefficiency of 2SLS or heteroscedasticity. The asymptotical bias estimates are huge compared to the 2SLS coefficients while they have even huger standard errors than their values. Although Altonji, Elder, and Taber (2005a) did not derive the asymptotical distribution for the bias, I use the normal distribution to derive the significance according to the central limit theorem. Most of the bias estimates are insignificant, except for two of them (Bias Estimate 2 for minutes and Bias Estimate 2 for log(minutes) when IV is farmland proportion) that are weakly significant (significant at the 10% level). Considering the limitations of the methodology, the huge standard errors of bias estimates cast doubt on the method of Altonji, Elder, and Taber (2005a).

Table 5.7.1 2SLS Estimates of Zoning Effects and Estimates of Potential Bias Using Altonji, Elder, and Taber (2005b) Methodology

	Excluded Instruments				
	Manufacturing Establishment Density		Farmland Proportion		
Dependent Variable	2SLS Coefficient	Bias	2SLS Coefficient	Bias	Observations
Physical Activity					
Participation Choice					
Estimate 1	0.046	-1.696	0.033	4.708	24,448
	(0.308)	(7.964)	(0.089)	(11.370)	
Estimate 2	0.044	-1.193	0.037	2.673	24,448
	(0.143)	(4.327)	(0.032)	(4.112)	
Minutes					
Estimate 1	9.676	352.449	10.786	620.737	1,969
	(1,492.294)	(2,034,750.925)	(24.248)	(534.800)	
Estimate 2	9.877	262.358	10.480	417.502*	1,969
	(95.120)	(1,247.424)	(15.142)	(249.020)	
Log(Minutes)					
Estimate 1	0.386	7.022	0.086	14.432	1,969
	(21.697)	(31,974.185)	(0.409)	(10.325)	
Estimate 2	0.332	5.992	0.169	7.834*	1,969
	(1.441)	(17.884)	(0.257)	(4.691)	
Sedentary Behavior					
Participation Choice					
Estimate 1	-0.106	-2.245	-0.220	4.436	24,458
	(0.207)	(7.413)	(0.361)	(8.207)	
Estimate 2	-0.125	-1.372	-0.185***	1.808	24,458
	(0.110)	(5.041)	(0.070)	(2.402)	
Minutes					
Estimate 1	-3.901	273.020	-34.416	496.214	20,141
	(79.862)	(3,574.277)	(70.584)	(2,419.531)	
Estimate 2	-9.046	222.450	-25.287	260.933	20,141
	(29.148)	(433.334)	(21.183)	(312.768)	
Log(Minutes)					
Estimate 1	-0.036	0.885	-0.246	3.664	20,141
	(0.343)	(14.560)	(0.551)	(16.381)	
Estimate 2	-0.071	0.799	-0.183	1.879	20,141
	(0.140)	(2.149)	(0.116)	(1.902)	

Notes:
1. Controls included are described in Table 5.3.1 notes.
2. “Estimate 1” includes the other IV in the set of observables while “Estimate 2” excludes it.
3. Standard Errors are obtained from a 1000-replication bootstrap and are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1.

5.8 IV First Stage

Table 5.8.1 and Table 5.8.2 show the first-stage regression of IV estimation in Step 1 of the mediational analysis, in which there is only one endogenous variable, zoning. Table 5.8.3 and Table 5.8.4 show the first-stage regression of IV estimation in Step 3 of the mediational analysis, in which I regress the two endogenous variables on the two IVs and all other explanatory variables. I use seemingly unrelated regression (SUR) to take the potential correlation between the two error terms of regression equations into account. If there is a significant correlation between the two error terms of the regression equations, then SUR will be more efficient than equation-by-equation ordinary least squares (OLS) because SUR employs feasible generalized least squares with a specific form of the variance-covariance matrix. Otherwise, the estimation results of SUR will be the same as that of equation-by-equation OLS.

The coefficients of the two IVs reflect the case of Village of Euclid, Ohio v. Ambler Realty Co. in 1926. The effect of manufacturing establishment density is positive, which implies the newly developing industry forces at that time challenged traditional Euclidean zoning and sought densely distributed buildings and populations, mixed-use spaces, and compactly-designed communities, all of which objectively promote pedestrian- and cyclist-friendly built environment and zoning. The effect of farmland proportion is negative, which implies the traditional agricultural forces at that time tried to defend the countryside characteristics and Euclidean zoning that feature sparsely distributed buildings and long distances between places. The directions of the coefficients of the two IVs reflect the fierce conflict for land use between the rapidly-developing manufacturing industry and traditional agricultural forces before the zoning ordinance was constitutionalized. The conflict shaped the new zoning and built environment.

The geographic variation of the two forces objectively created the original geographic variation in pedestrian- and cyclist-friendly zoning provisions and further affected active living-oriented zoning and built environment afterward. Note that the zoning metric reflects the policies that were effective as of 2010, including many policies that have not been changed for many years, possibly even decades.

The first-stage analysis is mainly used to investigate the weak identification problem. “The cure can be worse than the disease” (Bound, Jaeger, and Baker 1993; Bound, Jaeger, and Baker 1995). When instrumental variables are only weakly correlated with the endogenous variables, IV estimates are biased and may not be consistent. The problem is also called weak identification. Staiger and Stock (1997) formalized the problem and provided a general rule that the first-stage F-statistic needs to be greater than 10. In my case, I have two endogenous variables and two IVs. The individual first-stage F-statistic is not sufficient for testing the weak identification. Cragg and Donald (1993) created a statistic to assess the whole weakness of instruments for the case of multiple endogenous variables. Stock and Yogo (2005) provided critical values for the Cragg-Donald statistic. Angrist and Pischke (2009) also proposed an F-statistic for the case of multiple endogenous variables.

This case, with two endogenous variables and two IVs, is vulnerable to not only weak identification but also underidentification. Suppose the two IVs are highly correlated with the two endogenous variables, which means there is no weak identification, and the two IVs are linearly correlated; there will be a risk of underidentification. In the extreme case that one IV is a linear function of the other IV, the analysis suffers from underidentification because, essentially, I have only one IV but two endogenous variables. In Section 5.5.3, where I discuss the sources of variation of the two IVs, I show the two IVs provide different information. The farmland proportion in 1900 was mainly determined by natural resources. The manufacturing establishment density in 1900 was determined by natural resources and market access and has the nature of path dependency. To empirically test underidentification, I employ the test derived by Sanderson and Windmeijer (2016).

Different from Stock and Yogo (2005) who formalized the weak identification problem as the coefficients of IVs in the first stage as being local to zero, Sanderson and Windmeijer (2016) formalized the problem as one coefficient being a linear function of the other (i.e., "parameter matrix is local to a rank reduction of one"). Specifically, they considered the form $\pi_1 = \delta\pi_2 + c/\sqrt{n}$, where π_1 and π_2 are the coefficients of the IVs in the first stage, c is a vector of constants, and n is the sample size.

The Sanderson-Windmeijer (SW) first-stage Wald statistic is distributed as Chi2(L1-K1+1), in which L1 is the number of IVs and K1 is the number of endogenous variables, under the null hypothesis that the particular endogenous variable is unidentified. The SW first-stage F-statistic is the F form of the same test statistic, which can be used to test the null hypothesis that the particular endogenous variable is weakly identified (Baum, Schaffer, and Stillman 2016). All the p-values of SW statistics for GMM sample (continuous outcome sample) are very small. I can reject the weak identification and underidentification at the 0.1% significance level.

All the studies regarding weak identification and underidentification are established on the basis of linear regression. For Probit model, I conduct the Wald test for the null hypothesis that the coefficients of IVs are zero. I can reject the null hypothesis at the 0.1% significance level.

Table 5.8.1 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 1: Physical Activity Regression

VARIABLES	IV Probit Active Living-Oriented Zoning	GMM and GMM Log Active Living-Oriented Zoning
Manufacturing Establishment Density	0.001***	0.001***
	(0.000)	(0.000)
Farmland Proportion	-0.164***	-0.175***
	(0.042)	(0.041)
Number of Observations	24,448	1,969
IV Probit First-Stage Weak Identification Test		
F(2, 250)	36.22	
Prob > F(2, 250)	0	
GMM First Stage		
F(2, 212)		52.27
Prob > F(2, 212)		0
Weak Identification Test		
SW F(2, 212)		52.27
Prob > SW F(2, 212)		0
Underidentification Test		
SW Chi2(1)		107.3
Prob > SW Chi2(1)		0

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 3 Column 1 and 2.

Table 5.8.2 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 1: Sedentary Behavior Regression

VARIABLES	IV Probit Active Living-Oriented Zoning	GMM and GMM Log Active Living-Oriented Zoning
Manufacturing Establishment Density	0.001***	0.001***
	(0.000)	(0.000)
Farmland Proportion	-0.164***	-0.163***
	(0.042)	(0.043)
Number of Observations	24,458	20,141
IV Probit First-Stage Weak Identification Test		
F(2, 250)	36.20	
Prob > F(2, 250)	0	
GMM First Stage		
F(2, 248)		33.13
Prob > F(2, 248)		0
Weak Identification Test		
SW F(2, 248)		33.13
Prob > SW F(2, 248)		0
Underidentification Test		
SW Chi2(1)		66.66
Prob > SW Chi2(1)		0

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 4 Column 1 and 2.

Table 5.8.3 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 3: Physical Activity Regression

	IV Probit		GMM and GMM Log	
LABELS	Active Living-Oriented Zoning	Walkability	Active Living-Oriented Zoning	Walkability
Manufacturing Establishment Density	0.001***	0.038***	0.001***	0.037***
	(0.000)	(0.000)	(0.000)	(0.001)
Farmland Proportion	-0.164***	-0.375***	-0.175***	-0.337
	(0.004)	(0.013)	(0.041)	(0.220)
Number of Observations	24,448	24,448	1,969	1,969
IV Probit First-Stage Weak Identification Test				
Chi2(2)	8,594	281,446		
Prob > Chi2(2)	0	0		
GMM First Stage				
F(2, 212)			52.27	1,286
Prob > F(2, 212)			0	0
Weak Identification Test				
SW F(1, 212)			18.14	23.40
Prob > SW F(1, 212)			0.000031	0.000003
Underidentification Test				
SW Chi2(1)			18.62	24.03
Prob > SW Chi2(1)			0.000016	0.000001

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 3 Column 3, 4, 5, and 6.

Table 5.8.4 First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Step 3: Sedentary Behavior Regression

	IV Probit		GMM and GMM Log	
LABELS	Active Living-Oriented Zoning	Walkability	Active Living-Oriented Zoning	Walkability
Manufacturing Establishment Density	0.001***	0.038***	0.001***	0.037***
	(0.000)	(0.000)	(0.000)	(0.001)
Farmland Proportion	-0.164***	-0.375***	-0.163***	-0.374*
	(0.004)	(0.013)	(0.043)	(0.191)
Number of Observations	24,458	24,458	20,141	20,141
IV Probit First-Stage Weak Identification Test				
Chi2(2)	8,599	281,461		
Prob > Chi2(2)	0	0		
GMM First Stage				
F(2, 248)			33.13	570.8
Prob > F(2, 248)			0	0
Weak Identification Test				
SW F(1, 248)			12.48	13.81
Prob > SW F(1, 248)			0.000490	0.000250
Underidentification Test				
SW Chi2(1)			12.56	13.89
Prob > SW Chi2(1)			0.000395	0.000193

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 4 Column 3, 4, 5, and 6.

5.9 IV Probit Model: A Simultaneous Equation System

There are two methods with which to estimate the IV Probit model: control function approach and maximum likelihood estimation. Because the coefficients of the Probit model are difficult to explain, researchers usually transform them into marginal effects on the probability for easy interpretation. Efficiency is generally lost during transformation, so the marginal effects have inflated standard errors, or even lose significance. To preserve the significance of the marginal effects to the greatest extent, I choose maximum likelihood estimation, which is more efficient than the control function approach.

The exposition below mainly draws on the study of Skeels and Taylor (2015). I supplement it with some technical details.

The Probit model of interest is

$$y_j = \begin{cases} 1, \text{if } y_j^* > 0 \\ 0, \text{other wise} \end{cases} \tag{14}$$

$$y_j^* = Y_j'\beta + X_j'\gamma + \epsilon_j \tag{15}$$

where y_j is the outcome variable, y_j^* is a latent variable, Y_j' is a vector of endogenous explanatory regressors, X_j' is a vector of exogenous regressors, j is the index for observation, $j = 1, \dots, N$.

The endogenous regressors Y_j' can be modeled as a linear function of exogenous regressors X_j' and IVs Z_j', which are equivalent to the first-stage regression in two-stage least squares,

$$Y_j' = [X_j', Z_j']\Pi + V_j'. \tag{16}$$

The endogeneity of Y_j' is represented by the vector of the correlation coefficients ρ between ϵ_j and V_j', which have joint normal distribution

$$[\epsilon_j, V_j']' \sim N(0, \Sigma) \tag{17}$$

where $\Sigma = \begin{bmatrix} 1 & \rho' \Omega_{22}^{1/2} \\ \Omega_{22}^{1/2} \rho & \Omega_{22} \end{bmatrix}$ is positive definite.

Equations (5) through (8) constitute the IV Probit model. To estimate the parameters of the system, one needs to find the joint distribution of y_j and Y_j' to form a likelihood function. To do so, I firstly solve for the joint distribution of y_j^* and Y_j'.

First, note that

$$Y_j' \sim N([X_j', Z_j']\Pi, \Omega_{22}). \tag{18}$$

Then I calculate the expectation and variance of y_j^*,

$$\begin{aligned} E(y_j^*) &= E(Y_j'\beta + X_j'\gamma + \epsilon_j) \\ &= E(Y_j')\beta + X_j'\gamma \\ &= E([X_j', Z_j']\Pi + V_j')\beta + X_j'\gamma \\ &= [X_j', Z_j']\Pi\beta + X_j'\gamma \end{aligned} \tag{19}$$

$$\begin{aligned} Var(y_j^*) &= Var(Y_j'\beta + X_j'\gamma + \epsilon_j) \\ &= Var(Y_j'\beta + \epsilon_j) \\ &= Var\left(([X_j', Z_j']\Pi + V_j')\beta + \epsilon_j\right) \\ &= Var(V_j'\beta + \epsilon_j) \\ &= \beta' Var(V_j')\beta + Var(\epsilon_j) + 2Cov(V_j', \epsilon_j)\beta \\ &= \beta'\Omega_{22}\beta + 1 + \rho'\Omega_{22}^{1/2}\beta \end{aligned} \tag{20}$$

$$Cov(y_j^*, Y_j) = Cov(Y_j'\beta + X_j'\gamma + \epsilon_j, Y_j)$$

$$= Cov(Y_j'\beta + \epsilon_j, Y_j)$$

$$= \beta' Var(Y_j') + Cov(\epsilon_j, Y_j)$$

$$= \beta' Var(V_j) + Cov(\epsilon_j, V_j)$$

$$= \beta'\Omega_{22} + \Omega_{22}^{1/2}\rho. \quad (21)$$

So

$$[y_j^*, Y_j'] \sim N\left(\left[[X_j', Z_j']\Pi\beta + X_j'\gamma, [X_j', Z_j']\Pi\right], \begin{bmatrix} \beta'\Omega_{22}\beta + 1 + \rho'\Omega_{22}^{1/2}\beta & \beta'\Omega_{22} + \Omega_{22}^{1/2}\rho \\ \beta'\Omega_{22} + \Omega_{22}^{1/2}\rho & \Omega_{22} \end{bmatrix}\right)$$

$$\sim N\left(\left[[X_j', Z_j']\Pi\beta + X_j'\gamma, [X_j', Z_j']\Pi\right], \begin{bmatrix} 1 & \beta' \\ 0 & I_{22} \end{bmatrix} \Sigma \begin{bmatrix} 1 & 0 \\ \beta & I_{22} \end{bmatrix}\right). \quad (22)$$

For bivariate normal distribution, I have the following conclusion. Suppose $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and $\rho(X, Y) = \rho$,

$$Y|X \sim N\left(\mu_Y + \rho\sigma_Y\left(\frac{X - \mu_X}{\sigma_X}\right), \sigma_Y^2(1 - \rho^2)\right). \quad (23)$$

Therefore, I can get

$$y_j^* | Y_j \sim N\left(Y_j'\beta + X_j'\gamma + (Y_j' - [X_j', Z_j']\Pi)\Omega_{22}^{-1/2}\rho, 1 - \rho'\rho\right). \quad (24)$$

I then derive the joint distribution of $[y_j, Y_j]$

$$f(y_j, Y_j) = f(y_j|Y_j) f(Y_j)$$

$$= \{\text{Prob}(y_j = 0|Y_j)\}^{1-y_j} [\text{Prob}(y_j = 1|Y_j)]^{y_j} f(Y_j)$$

$$= \{\text{Prob}(y_j^* \leq 0|Y_j)\}^{1-y_j} \{\text{Prob}(y_j^* > 0|Y_j)\}^{y_j} f(Y_j) \quad (25)$$

where

$$\text{Prob}(y_j^* \leq 0|Y_j)$$

$$= \text{Prob}\left(\frac{y_j^* - \left(Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho\right)}{\sqrt{1-\rho'\rho}}\right.$$

$$\left.\le \frac{-\left(Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho\right)}{\sqrt{1-\rho'\rho}}\middle| Y_j\right)$$

$$= \Phi\left(\frac{-\left(Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho\right)}{\sqrt{1-\rho'\rho}}\right) \tag{26}$$

$$\text{Prob}\left(y_j^* > 0 \middle| Y_j\right)$$

$$= \text{Prob}\left(\frac{y_j^* - \left(Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho\right)}{\sqrt{1-\rho'\rho}}\right.$$

$$\left.> \frac{-\left(Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho\right)}{\sqrt{1-\rho'\rho}}\middle| Y_j\right)$$

$$= \Phi\left(\frac{Y_j'\beta + X_j'\gamma + \left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1/2}\rho}{\sqrt{1-\rho'\rho}}\right) \tag{27}$$

and

$$f\left(Y_j\right) = \frac{1}{2\pi}(\det\Omega_{22})^{-1/2}\exp\left[-\frac{1}{2}\left(Y_j' - [X_j', Z_j']\Pi\right)\Omega_{22}^{-1}\left(Y_j' - [X_j', Z_j']\Pi\right)'\right]. \tag{28}$$

Finally, I get the log-likelihood function

$$L(\beta, \gamma, \Pi, \rho, \Omega_{22}; y, Y, X, Z)$$

$$= \ln\left(\prod_j f(y_j, Y_j)\right)$$

$$= \ln\left(\prod_j \{\text{Prob}(y_j^* \leq 0|Y_j)\}^{1-y_j}\{\text{Prob}(y_j^* > 0|Y_j)\}^{y_j} f(Y_j)\right)$$

$$= \sum_j \left((1 - y_j)\ln\{\text{Prob}(y_j^* \leq 0|Y_j)\} + y_j \ln\{\text{Prob}(y_j^* > 0|Y_j)\} + \ln f(Y_j)\right). \qquad (29)$$

By maximize the log likelihood function, I can get estimators for $\hat{\beta}$, $\hat{\gamma}$, $\widehat{\Pi}$, $\hat{\rho}$, and $\widehat{\Omega}_{22}$. The policy effects are $\hat{\beta}$, which are the coefficients of Y_j. The hypothesis test on $H_0: \rho = 0$ provides an endogeneity test. If $\rho = 0$, then there is no endogeneity.

The predicted probability for observation *j* is

$$\widehat{\text{Prob}}(y_j = 1|Y_j) = \widehat{\text{Prob}}(y_j^* > 0|Y_j)$$

$$= \Phi\left(\frac{Y_j'\hat{\beta} + X_j'\hat{\gamma} + \left(Y_j' - [X_j', Z_j']\widehat{\Pi}\right)\widehat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1 - \hat{\rho}'\hat{\rho}}}\right). \qquad (30)$$

Therefore, the average predicted probability is

$$\frac{1}{N}\sum_{j=1}^{N} \widehat{\text{Prob}}(y_j = 1|Y_j) = \frac{1}{N}\sum_{j=1}^{N} \Phi\left(\frac{Y_j'\hat{\beta} + X_j'\hat{\gamma} + \left(Y_j' - [X_j', Z_j']\widehat{\Pi}\right)\widehat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1 - \hat{\rho}'\hat{\rho}}}\right). \qquad (31)$$

The marginal effects of endogenous variables are

$$\frac{1}{N}\sum_{j=1}^{N} \frac{\partial \widehat{\text{Prob}}(y_j = 1|Y_j)}{\partial Y_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N} \frac{\partial \Phi\left(\frac{Y_j'\hat{\beta} + X_j'\hat{\gamma} + \left(Y_j' - [X_j', Z_j']\widehat{\Pi}\right)\widehat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1 - \hat{\rho}'\hat{\rho}}}\right)}{\partial Y_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\beta} + \widehat{\Omega}_{22}^{-1/2}\hat{\rho}\right)\phi\left(\frac{Y_j'\hat{\beta} + X_j'\hat{\gamma} + \left(Y_j' - \left[X_j', Z_j'\right]\widehat{\Pi}\right)\widehat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1-\hat{\rho}'\hat{\rho}}}\right). \tag{32}$$

5.10 General Method of Moments

For linear regression, there are four techniques with which to implement IV estimation: generalized instrumental variables estimation, two-stage least squares (2SLS), control function approach, and general method of moments (GMM). GMM outperforms the other methods in efficiency and dealing with heteroskedasticity, so I use GMM.

Suppose the equation of interest is

$$y_j = X_j'\beta + u_j. \tag{33}$$

Researchers can write it in matrix form

$$y = X\beta + u \tag{34}$$

where $X = [Y\ Z_2]$, X is the matrix of K regressors, Y is the matrix of K_1 endogenous regressors, and Z_2 is the matrix of $(K - K_1)$ exogenous regressors.

There are L_1 IVs, represented by the matrix Z_1. Because both Z_1 and Z_2 are exogenous, I denote them as $Z = [Z_1\ Z_2]$. In literature, IVs Z_1 are called excluded instruments, and exogenous regressors Z_2 are called included instruments. Suppose Z has L columns, $L - L_1 = K - K_1$.

If $L_1 = K_1$, then there are exactly the same number of IVs and endogenous regressors, and the system is "exactly identified"; if $L_1 > K_1$, then it is "overidentified"; if $L_1 < K_1$, then it is "underidentified."

The generalized instrumental variable estimation is used to calculate coefficients in one step. $\hat{\beta}_{IV} = (X'P_Z X)^{-1}X'P_Z y$ where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix of Z.

The 2SLS is used to calculate coefficients in two steps. Note that $\hat{\beta}_{IV} = (X'P_Z X)^{-1}X'P_Z y = (X'P_Z P_Z X)^{-1}X'P_Z y = ((P_Z X)'P_Z X)^{-1}(P_Z X)'y$ because $P_Z = Z(Z'Z)^{-1}Z'$ is a symmetric idempotent matrix. Because P_Z is also a projection matrix, $P_Z X$ is the projection of X on the column space of Z, which implies that researchers can write $\hat{\beta}_{IV} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'y$, where $\hat{X}$ is the fitted value when researchers regress X on Z. Therefore, the two stages are, first, regress the endogenous variables on the IVs and all other exogenous explanatory variables to get fitted values, and second, replace endogenous variables in the original regression with fitted values. The control function approach involves another two stages. The first stage is the same as that in 2SLS. In the second stage, the residual from the first stage is added to the original regression as a regressor.

For binary choice models, exponential models, and the case that endogenous variables nonlinearly enter into original regression, 2SLS requires more assumptions than the linear-in-parameter model while the control function does not. However, for linear-in-parameter regression and the case that endogenous variable is discretely distributed, the control function approach imposes additional assumptions about the relationship between the endogenous variable and the IVs compared to 2SLS. The control function approach offers us some conveniences that 2SLS does not. In the second stage of the control function approach, researchers can directly test the endogeneity of the endogenous variable without implementing the Hausman specification test. Researchers can also implement robust inference to address the heteroskedasticity of the error term. Note that neither the control function approach nor 2SLS can make consistent estimators for the standard errors of the coefficients (Wooldridge 2007; Wooldridge 2015).

The exposition below mainly draws on the study of Baum, Schaffer, and Stillman (2003).

The exogeneity of Z means $E\left(Z_j'u_j\right) = 0$. Define $g_j(\beta) = Z_j'u_j = Z_j'\left(y_j - X_j\beta\right)$ where $g_j(\cdot)$ is $L \times 1$. Hence, the exogeneity of Z implies $E\left(g_j(\beta)\right) = 0$. There are L moment conditions (orthogonality conditions) satisfied at the true value β.

For each population moment condition, there is a sample moment condition. One can write L sample moments as

$$\bar{g}(\beta) = \frac{1}{N}\sum_{j=1}^{N} g_j(\beta) = \frac{1}{N}\sum_{j=1}^{N} Z_j'u_j = \frac{1}{N}\sum_{j=1}^{N} Z_j'\left(y_j - X_j\beta\right) = \frac{1}{N}Z'u. \tag{35}$$

Choosing an estimator for β that solves $\bar{g}\left(\hat{\beta}\right) = 0$ is called method of moments (MM).

Note that I have L equations and K parameters. If the equation system is underidentified ($L < K$, or equivalently $L_1 < K_1$ because $L - L_1 = K - K_1$), then it is unsolvable. If the equation system is overidentified ($L > K$, or equivalently $L_1 > K_1$), then generally, it is impossible to find a $\hat{\beta}$ that sets all L sample moment conditions to be exactly zero. The GMM solution is to take an $L \times L$ weighting matrix W and use it to construct a quadratic form

$$J(\beta) = N\bar{g}(\beta)'W\bar{g}(\beta) \tag{36}$$

and find an estimator that minimizes $J(\beta)$. $J(\beta)$ is GMM objective function.

By solving the K first order conditions

$$\frac{\partial J\left(\hat{\beta}\right)}{\partial \hat{\beta}} = 0 \tag{37}$$

one can get the GMM estimator

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'y. \tag{38}$$

There are as many GMM estimators as there are different weighting matrices. However, if the weighting matrices only differ by a constant multiplier, then the GMM estimators will be the same because of the minimization programming.

The GMM estimators are all consistent, but not all efficient. The optimal choice of weighting matrix should be the one that minimizes the asymptotic variance of the estimator. To figure out the optimal weighting matrix, I first find the asymptotic variance of the estimator.

$$\begin{aligned}\hat{\beta}_{GMM} &= (X'ZWZ'X)^{-1}X'ZWZ'y \\ &= (X'ZWZ'X)^{-1}X'ZWZ'(X\beta + u) \\ &= \beta + (X'ZWZ'X)^{-1}X'ZWZ'u. \end{aligned} \tag{39}$$

So

$$\begin{aligned}Var(\hat{\beta}_{GMM}) &= Var(\beta + (X'ZWZ'X)^{-1}X'ZWZ'u) \\ &= Var((X'ZWZ'X)^{-1}X'ZWZ'u) \\ &= (X'ZWZ'X)^{-1}X'ZWZ'Var(u)ZWZ'X(X'ZWZ'X)^{-1}. \end{aligned} \tag{40}$$

Let

$$Q_{XZ} = E\left(X_j'Z_j\right) \tag{41}$$

$$S = \frac{1}{N}E(Z'uu'Z) = \frac{1}{N}E(Z'\Omega Z) \tag{42}$$

where S is the covariance matrix of the moment conditions g.

Hence,

$$Asy.Var\left(\hat{\beta}_{GMM}\right) = \frac{1}{N}(Q_{XZ}'WQ_{XZ})^{-1}(Q_{XZ}'WSWQ_{XZ})(Q_{XZ}'WQ_{XZ})^{-1}. \tag{43}$$

By setting $W = S^{-1}$, I get the asymptotic variance minimized

$$Asy.Var\left(\hat{\beta}_{EGMM}\right) = \frac{1}{N}(Q_{XZ}'S^{-1}Q_{XZ})^{-1}. \tag{44}$$

The corresponding estimator is called efficient GMM estimator

$$\hat{\beta}_{EGMM} = (X'ZS^{-1}Z'X)^{-1}X'ZS^{-1}Z'y. \tag{45}$$

However, no one knows S. To estimate S, researchers need to determine the value of Ω in advance.

The technique here is very similar to the feasible generalized least squares. I use a consistent estimator of residual u_j to form a consistent estimator of Ω,

$$\hat{\Omega} = \begin{bmatrix} \hat{u}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{u}_N^2 \end{bmatrix} \tag{46}$$

where $\hat{u}_j$ is a consistent estimator of u_j. Any consistent estimator of β can make a consistent estimator of the residual.

Therefore, the efficient GMM estimator can be obtained through two steps:

1. Estimate the equation using IV to get consistent estimator $\hat{u}_j$ of the residual.
2. Plug $\hat{u}_j$ into Equation (38) to form the consistent estimator of Ω.

$$\hat{\beta}_{FE2SGMM} = \left(X'Z\left(Z'\hat{\Omega}Z\right)^{-1}Z'X\right)^{-1} X'Z\left(Z'\hat{\Omega}Z\right)^{-1}Z'y \tag{47}$$

with $\hat{W} = \hat{S}^{-1} = \left(\frac{1}{N}Z'\hat{\Omega}Z\right)^{-1}$.

$\hat{\beta}_{FE2SGMM}$ is the feasible efficient two-step GMM estimator with asymptotic variance

$$Asy.Var\left(\hat{\beta}_{FE2SGMM}\right) = \left(X'Z\left(Z'\hat{\Omega}Z\right)^{-1}Z'X\right)^{-1}. \tag{48}$$

The implementation procedure is robust to heteroskedasticity. In the presence of heteroskedasticity, $\hat{\beta}_{FE2SGMM}$ is efficient while the IV estimator (2SLS) is inefficient and the standard IV estimates of the standard errors are inconsistent, preventing valid inference.

5.11 Testing for Endogeneity: Hausman Specification Test and C Test

In *Introductory Econometrics*, Wooldridge noted (Wooldridge 2009, p. 511) that "…an important cost of performing IV estimation when x and u are uncorrelated: the asymptotic variance of the IV estimator is always larger, and sometimes much larger than the asymptotic variance of the OLS estimator." This statement reminds researchers of the price they have to pay when using IV estimation to solve the endogeneity problem. Because the endogeneity in regressions makes estimators biased and inconsistent and researchers usually set unbiasedness and consistency as priority, researchers choose to sacrifice efficiency to get a consistent estimator through IV estimation. IV estimation can always give researchers a consistent estimator, given that the IVs are valid, regardless of whether there is an endogeneity problem in the regressions of the mediational analysis. However, if the regression in the mediational analysis can give researchers an unbiased and consistent estimator, then they do not need to pay the price of efficiency to do IV estimation. It is possible that the active living preference has little effect on driving people to seek active living-oriented zoning and built environment, or the effect is too small to consider. Nowadays, researchers rely on statistical significance to judge whether effects exist (even though this type of judgment is debatable). It is highly plausible that the estimated coefficients from IV estimation are insignificant due to the low efficiency of IV estimation, which leads researchers to conclude that causal effects do not exist when they truly exist. Therefore, it is important to know whether the association between active living preference and zoning/built environment is prominent enough to be worth to pay attention to.

Researchers cannot directly test the association between the unobserved confounder and the suspect endogenous variables. However, researchers can compare the OLS estimator and IV estimator. The null hypothesis is that the suspect endogenous variables are uncorrelated with the error term. Suppose valid IVs exist. IV estimation can make a consistent estimator under both the null hypothesis and the alternative hypothesis. OLS can make a consistent estimator under the null hypothesis but an inconsistent estimator under the alternative hypothesis. If the null hypothesis is true, then the OLS estimate and IV estimate are expected to be very similar. If a big difference between the OLS estimate and IV estimate is observed, then the truth of the null hypothesis is questionable. Because estimates are sample-dependent and estimators have sampling distribution, instead of measuring the difference between the magnitudes of the estimates, one should measure the probability that the difference between the estimators exceeds the difference between the estimates (i.e., p-value). If the p-value is greater than type I error rate, it implies that, given that the null hypothesis is true, researchers have great chances to observe the current difference between the estimators or the even bigger difference than the current difference. The current observation is consistent with the null hypothesis. Therefore, researchers cannot reject the null hypothesis.

The test is called Hausman specification test. The test is based on the assumption that IVs are valid. The test statistic is a precision matrix-scaled quadratic form in the differences between the OLS estimator and the IV estimator. The exposition below mainly draws on the study of Baum, Schaffer, and Stillman (2003). Specifically,

$$H = N\left(\hat{\beta}^c - \hat{\beta}^e\right)' \left(V\left(\hat{\beta}^c\right) - V\left(\hat{\beta}^e\right)\right)^{-} \left(\hat{\beta}^c - \hat{\beta}^e\right) \tag{49}$$

where $\hat{\beta}^c$ is the IV estimator that is consistent under both the null and the alternative hypotheses, $\hat{\beta}^e$ is the OLS estimator that is consistent under the null hypothesis but inconsistent under the alternative hypothesis, $V\left(\hat{\beta}^c\right)$ and $V\left(\hat{\beta}^e\right)$ are consistent estimates of the asymptotic variance of $\hat{\beta}^c$ and $\hat{\beta}^e$, respectively, and operator – denotes a generalized inverse. $\hat{\beta}^e$ is more efficient than $\hat{\beta}^c$. H follows Chi-square distribution with the degree of freedom being the number of regressors being tested for endogeneity.

In the GMM framework, researcher can form the heteroskedasticity-robust test statistic. The GMM objective function $J(\beta)$ evaluated at the efficient GMM estimator $\hat{\beta}_{EGMM}$ follows $\chi^2(L-K)$ if L orthogonality conditions are valid. $J(\hat{\beta}_{EGMM})$ is called the Hansen J statistic. To form the test statistic of the endogeneity test, researcher can perform GMM twice to construct two Hansen J statistics. One treats the suspect endogenous variable as exogenous, which is similar to OLS, the more efficient estimation, to form Hansen J statistic that follows $\chi^2(L+1-K)$ under the null hypothesis that $(L+1)$ orthogonality conditions are valid. The other one treats the suspect endogenous variable as endogenous, which is similar to IV estimation, to form a Hansen J statistic that follows $\chi^2(L-K)$ under both the null and the alternative hypotheses. Under the null hypothesis, both the two Hansen J statistics will make consistent estimators so that their difference is not supposed to be significantly large. Therefore, researchers form the test statistic as the difference between the two Hansen J statistics, which is called the GMM distance test or C test. Because the test statistic is the difference between the two Chi-square statistics ($\chi^2(L+1-K)$ and $\chi^2(L-K)$), it will follow $\chi^2(1)$.

Note that the C test is based on the assumption that IVs are valid as well. If the IVs are invalid, then researchers cannot answer the question of whether the potential endogenous variables are truly endogenous.

5.12 Overidentification Test

In Step 1 of the mediational analysis with IV estimation, there is only one endogenous variable, zoning, but two IVs. The model is overidentified. The overidentification allows researchers to test the exogeneity of each IV as long as the other IV satisfies the exclusion restriction.

Researchers can utilize the C test again. They can perform GMM twice to construct two Hansen J statistics; one uses the whole set of IVs, and the other excludes the suspect IV. The null hypothesis is that the suspect IV satisfies the exclusion restriction as long as the remaining IV is exogenous. Under the null hypothesis, both GMM estimations have valid orthogonality conditions and can produce consistent estimators; the two Hansen J statistics, which follow $\chi^2(L-K)$ and $\chi^2(L-1-K)$, respectively, are supposed to be very similar. The C test statistic is the difference between the two Hansen J statistics. The C test statistic follows $\chi^2(1)$. Researchers calculate the probability that the difference equals to or exceeds the observed difference between the two Hansen J statistics (i.e., p-value). If the probability is greater than the type I error rate, then this implies that researchers have great chances to observe the current difference between the two Hansen J statistics or the even greater difference than the current difference, given that the null hypothesis is true. The current observation is consistent with the null hypothesis. Therefore, researchers cannot reject the null hypothesis.

Note that the overidentification test for the suspect IV is based on the assumption that the other IV is valid. If the other IV is invalid, then researchers cannot answer the question of whether the suspect IV satisfies the exclusion restriction.

5.13 Discussions and Results

Before I investigate the results, it is important to make two concepts clear: average treatment effect (ATE) and local average treatment effect (LATE). To make discussions clear, I focus on linear-in-parameter models. If the error term has zero expectation conditional on covariates (no endogeneity and zero unconditional expectation of the error term) in the regression, then the estimated coefficients are unbiased and consistent estimators for the ATE of the population. In IV estimation, the estimated coefficients are no longer consistent estimators for ATEs. Rather, they are consistent estimators for LATEs, which are ATEs on "compliers," a subpopulation whose active living outcomes change as the values of the IVs change.

The concept "complier" comes from experimental design. In experimental design, although researchers can never identify whom the compliers are, they can identify the proportion of never-takers, always-takers, and compliers in the sample under the assumptions that there are no defiers and random assignment. In addition, researchers can identify the average outcomes for never-takers and always-takers, and the ATE for compliers (LATE) under the assumption of exclusion restriction. When experimental subjects are randomly assigned to a treatment group and control group, under the assumption that there are no defiers (also known as the monotonicity assumption), the subjects who do not take treatment in the treatment group can be identified as never-takers, and the subjects who take treatment in the control group can be identified as always-takers. Because of random assignment, which implies that the treatment group and control group are probabilistic equivalents, the proportion of never-takers in the treatment group and the proportion of always-takers in the control group can be generalized to the whole sample. Therefore, the proportion of compliers can be calculated. Under the exclusion restriction, which implies that the treatment assignment has no effect on outcomes except through treatment status, the average results for never-takers in the treatment group and always-takers in control group are unbiased and consistent estimators for never-takers and always-takers in the whole sample, respectively. Hence, the ATE for compliers can be calculated.

In the econometrics of instrumental variable identification, IVs and endogenous variables can be viewed as treatment assignment and treatment status, respectively. The assumption of exclusion restriction is the same. The first-stage requirement is stronger than the assumption that there are no defiers (monotonicity) (Imbens 2007). However, there is a subtle difference—that the random assignment seems to not be strongly required for IVs. The counterpart of the random assignment of treatment in econometrics is zero correlation between IVs and error terms conditional on other covariates in the first-stage regressions, which implies the first-stage regressions are causal. The random assignment assumption requires investigation of the IV assignment mechanism. Although some texts and research papers include random assignment as a requirement for valid IVs (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996; Imbens 2007), the causal interpretation of the IV estimators does not require this assumption (Angrist, Imbens, and Rubin 1996). Moreover, in most applications of IV identification in the empirical literature, it is difficult to argue the causality of the first-stage regressions. For example, to estimate the causal effect of schooling on earnings, Angrist and Krueger (1991) used quarter of birth as an IV. Card (2001) used the distance to college as an IV. Parents can use contraceptive devices to manipulate the quarter of birth according to their preferences for children's schooling. Households sort across housing locations based on their preferences, so the distance to college is a choice variable. Both quarter of birth and distance to college are hardly uncorrelated with error terms in first-stage regressions, even conditional on observables. In this thesis, the farmland proportions of counties in 1900 depend on natural resources, and the manufacturing establishment densities of counties in 1900 are the consequences of a series of historical events.

Though I cannot rigorously prove that there are zero correlations between the IVs and the error terms conditional on other covariates in the first-stage regressions, I assume that the first stages are causal.

Suppose the endogenous variable x and the IV z are dichotomous variables; the IV estimator is

$$\hat{\beta} = \frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{E[x_i|z_i = 1] - E[x_i|z_i = 0]} = E\left[\left(y_i(1) - y_i(0)\right)\middle|x_i(1) - x_i(0) = 1\right] \qquad (50)$$

where $E[x_i|z_i = 1] - E[x_i|z_i = 0] \neq 0$ characterizes compliers (Angrist, Imbens, and Rubin 1996; Imbens 2007). The estimator is ATE on compliers (i.e., LATE). With other explanatory variables, the estimator is LATEs averaged across covariate cells (Angrist and Pischke 2009, p. 198). If the endogenous variable is continuous, then it can be characterized as a multi-valued variable that takes on values of 0, 1, …, M. The IV estimator is

$$\hat{\beta} = \sum_{m=1}^{M} \lambda_m \, E\left[\left(y_i(m) - y_i(m-1)\right)\middle|x_i(1) \geq m > x_i(0)\right] \qquad (51)$$

where $\lambda_m = \frac{\Pr\left(x_i(1) \geq m > x_i(0)\right)}{\sum_{m=1}^{M} \Pr\left(x_i(1) \geq m > x_i(0)\right)}$ (Imbens 2007). It is a weighted average of LATE over each value of the endogenous variable. If the IV is continuous, then the IV estimator can be characterized using the integration of marginal treatment effect (Heckman and Vytlacil 2005). Here, to make the discussions simple, I characterize the continuous IV as multi-valued variable that takes on the values z_0, z_1, …, z_Q. The IV estimator is

$$\hat{\beta} = \sum_{q=1}^{Q} \lambda_q \, E\left[\left(y_i(1) - y_i(0)\right)\middle|x_i(z_q) - x_i(z_{q-1}) = 1\right] \qquad (52)$$

where $\lambda_q = \frac{\left(p(z_q)-p(z_{q-1})\right)\sum_{l=q}^{Q}\pi_l(z_l-E(z_i))}{\sum_{q=1}^{Q}\left\{\left(p(z_q)-p(z_{q-1})\right)\sum_{l=q}^{Q}\pi_l(z_l-E(z_i))\right\}}$, $p(z) = \Pr(x_i = 1|z_i = z)$ with $p\left(z_{q-1}\right) \leq p\left(z_q\right)$, and $\pi_q = \Pr\left(z_i = z_q\right)$ (Imbens 2007). It is a weighted average of LATE over each value of the IV. Hence, the estimator from GMM can be interpreted as an average effect of zoning on the duration of physical activity (sedentary behavior) for compliers, consisting of two types of weighted averaging. The first is a weighted average of LATE of zoning over different levels of zoning. The second is, given any specific level of zoning, say 0.5, a weighted average of LATE, where the averag is over those of people who would have been exposed to the zoning of at least 0.5 if higher 1900 manufacturing establishment density or lower 1900 farmland proportion had been in effect for them, and who would have been exposed to the zoning of less than 0.5 had they been subject to the lower 1900 manufacturing establishment density or the higher 1900 farmland proportion. The zoning level 0.5 means 50 percent of the county population was exposed to the development of active living-oriented zoning.

For the IV Probit model, things are more complicated than for linear-in-parameter models. It is impossible to compare the IV Probit model to the experiment design. If the endogenous variable and the IV are dichotomous in the IV Probit model, then researchers can still define metrics for never-takers, always-takers, and compliers using the parameters of the simultaneous equations (Angrist and Pischke 2009, Section 4.6.3). Nonetheless, the estimators in the IV Probit model are for population parameters. The marginal effects calculated from the estimators in the IV Probit model are consistent estimators for ATEs.

Table 5.13.1 and Table 5.13.2 show the results of Step 1 of the mediational analysis with IV estimation for physical activity and sedentary behavior, respectively. All the overidentification tests show insignificant results, which implies that given that one of the IVs satisfies the exclusion restriction, the other one also satisfies the exclusion restriction. Almost all the endogeneity tests show insignificant differences in the orthogonality conditions between with and without IVs except the Probit model of sedentary behavior, which implies the correlation between unobserved active living preference and zoning is too weak to need to take care of, except when people decide whether to engage in sedentary behavior.

Firstly, let us focus on the binary choice model. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 2.1 percentage point increase (IV Probit model) in the chances that an adult engaged in physical activity and a 1.4 percentage point decrease (IV Probit model) in the likelihood that an adult chose sedentary behavior, given that other factors remain unchanged. The effects include both the direct effect and the possible indirect effect that goes through walkability.

The marginal effect of zoning on physical activity from the IV Probit model (0.021) is similar to its counterpart from the Probit model (0.028) in magnitude, which is consistent with the fact that the endogeneity test shows insignificant endogeneity. The marginal effect of zoning on sedentary behavior in the IV Probit model (-0.014) is smaller than its counterpart from the Probit model (-0.046), which is consistent with the fact that the endogeneity causes upward bias, and the endogeneity test shows significant endogeneity.

The marginal effects in the IV Probit model are much less significant than the coefficients. This is due to the inefficiency of the marginal effect of calculation. To calculate the marginal effects, I estimate the IV Probit model firstly. Then I plug the estimates of the IV Probit model into the formula of the marginal effects

$$\frac{1}{N}\sum_{j=1}^{N}\left(\hat{\beta}+\hat{\Omega}_{22}^{-1/2}\hat{\rho}\right)\phi\left(\frac{Y_j'\hat{\beta}+X_j'\hat{\gamma}+\left(Y_j'-\left[X_j',Z_j'\right]\hat{\Pi}\right)\hat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1-\hat{\rho}'\hat{\rho}}}\right) \tag{32}$$

Note that, in Equation (24), as long as one of the parameters ($\hat{\beta}$, $\hat{\gamma}$, $\hat{\Pi}$, $\hat{\rho}$, $\hat{\Omega}_{22}$) is insignificant, the calculated marginal effects can hardly be significant. In other words, the marginal effect calculation incorporates the sampling errors of the estimators of all parameters in the simultaneous equation system. IV estimation is less efficient than estimation without IVs. The formula-based marginal effect calculation worsens the situation so that the calculated marginal effects are weakly significant (for physical activity), or even insignificant (for sedentary behavior). The marginal effects allow researchers to interpret the coefficient easily. However, researchers should still turn to the significance of the coefficients in the IV Probit model whenever they need to make judgments based on statistical significance.Now, let us turn to the linear models. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 9.957 unit increase (GMM) or a 29.7 percent increase (GMM Log) in the minutes people spent on physical activity and a 8.9 percent decrease (GMM) in the minutes people spent on sedentary behavior.

The estimated effects from GMM and GMM Log have greater magnitudes than those from OLS and OLS Log, respectively. Even though IV estimation is less efficient than non-IV estimation, the GMM estimates have greater standard errors than the OLS estimates. The increases in magnitudes are so large so that some GMM estimates are statistically significant.

The changes in magnitudes of the coefficients are the combined results of endogeneity, measurement error (Harmon and Oosterbeek 2000), and the heterogeneous treatment effects for different subpopulations. The possible endogeneity in our study causes upward bias, so the IV estimation should have pulled the magnitude down to its true level. The measurement error of an explanatory variable, in a linear model, tends to bias the coefficient toward zero, so IV estimation should pull the magnitude up to its true level (Angrist and Krueger 2001). Here, the measurement error comes from the impossibility of linking ATUS respondents to the zoning metrics of their specific jurisdictions or areas (Leider, Chriqui, and Thrun 2016). The heterogeneous treatment effects imply that the ATEs for compliers (LATE) can be larger or smaller than the ATEs for the whole population. Because the endogeneity tests show insignificant endogeneity, the increase in magnitude is likely caused by the measurement error of zoning and the LATE.

Table 5.13.3 shows the result of Step 2 of the mediational analysis. After controlling for IVs, I find that zoning still has a significant effect on walkability. A one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 0.379 unit increase in the walkability metric.

Table 5.13.4 and Table 5.13.5 show the results of Step 3 of the mediational analysis with IV estimation for physical activity and sedentary behavior, respectively. I control for both zoning and walkability in the analysis. Step 3 of the mediational analysis is the key to establishing the mediation of walkability. Because most of the effects of zoning and walkability on the minutes (continuous outcome) are not significant in analysis either without or with IVs, it is impossible to establish the mediation of walkability for the minutes (continuous outcome).[17] I focus the discussion on binary choice for participation.

The standard errors of the estimates from the IV Probit model are much greater than those from the Probit model. Most of the estimates from the IV Probit model are insignificant. The huge standard errors are due to two facts: the low efficiency of IV estimation compared to

[17] The insiginificance can be caused by various factors, including relatively small sample size, the high correlation between predicted walkability and predicted zoning (since we use the same set of IVs). Given the significant effect of zoning on the minutes in Step 1 with IV estimation, it is mostly plausible to be caused by the high correlation between predicted endogenous variables.

non-IV estimation, and the high collinearity between predicted zoning and predicted walkability, which comes from the fact that I use the same set of covariates for zoning and walkability in IV first stage.

For physical activity analysis, the insignificant coefficients from the IV Probit model in Step 3 (Table 5.13.4) make it difficult to draw any conclusions about the mediation of walkability. Fortunately, the endogeneity test shows insignificant endogeneity, which means the estimates from the Probit model are consistent. Because the endogeneity test in Step 1 of physical activity analysis (Table 5.13.1) also suggests that the Probit model estimates are consistent, I can still rely on the results of the Probit model to establish the mediation of walkability. Also, the marginal effects from the IV Probit model in Step 3 are very similar in magnitude to those from the Probit model in Step 3. I illustrate the mediational analysis with IV Probit for physical activity in Figure 5.13.1. Once walkability is added to the analysis, the marginal effect of zoning on physical activity participation decreases from 0.021 to 0.020, which suggests that walkability is a mediator for zoning.

For sedentary behavior analysis, again, there are the huge standard errors of the estimates from the IV Probit model in Step 3 (Table 5.13.5). The explanation is the same as that of the physical activity analysis. The endogeneity test shows significant endogeneity at the 10% significant level, which means estimates from the Probit model are inconsistent. As the endogeneity test in Step 1 of sedentary behavior analysis (Table 5.13.2) also suggests the Probit model estimates are inconsistent, I can no longer rely on the results of the Probit model to establish the mediation of walkability. The coefficients of zoning from the IV Probit model in both Step 1 (Table 5.13.2) and Step 3 are significant while the coefficient of walkability from the IV Probit model in Step 3 is insignificant, which implies that zoning has a significant effect on sedentary behavior participation but walkability does not. Also, although the marginal effects of zoning calculated from the coefficients from the IV Probit model are insignificant because of the inefficiency of the calculation procedure, they are the same in Step 1 and Step 3 (-0.014), which implies that controlling for walkability in the analysis does not change the effects of zoning. Hence, the mediation of walkability does not hold in sedentary behavior. I illustrate the mediational analysis with the IV Probit model for sedentary behavior in Figure 5.13.1.

The mediational analysis for sedentary behavior establishes the mediation of walkability with the Probit model but does not do so with the IV Probit model. Because the endogeneity tests in the IV Probit model for sedentary behavior show significant endogeneity in both Step 1 and Step 3, the mediational effect of walkability in the causal path from zoning to sedentary behavior is actually due to the sorting, which is based on people's preferences for active living. When the IV Probit model estimates consistent causal parameters for the population, the mediational effect of walkability disappears.

The failure of mediation of walkability in the causal path from zoning to sedentary behavior implies that built environment elements that cannot be captured by the walkability metric play important roles in people's sedentary behavior decisions. The zoning variable, active living-oriented zoning, includes many more elements of built environment than walkability, including active recreation (equipment to support physical activity), passive recreation (opportunities for physical activity like open space), mixed-use development, bike parking, and bike lanes. Many of these zoning provisions cannot be captured by the walkability metric, which is constituted by four density metrics: population density, housing unit density, the ratio of four-way intersections to all intersections, and the total number of intersections divided by land area.

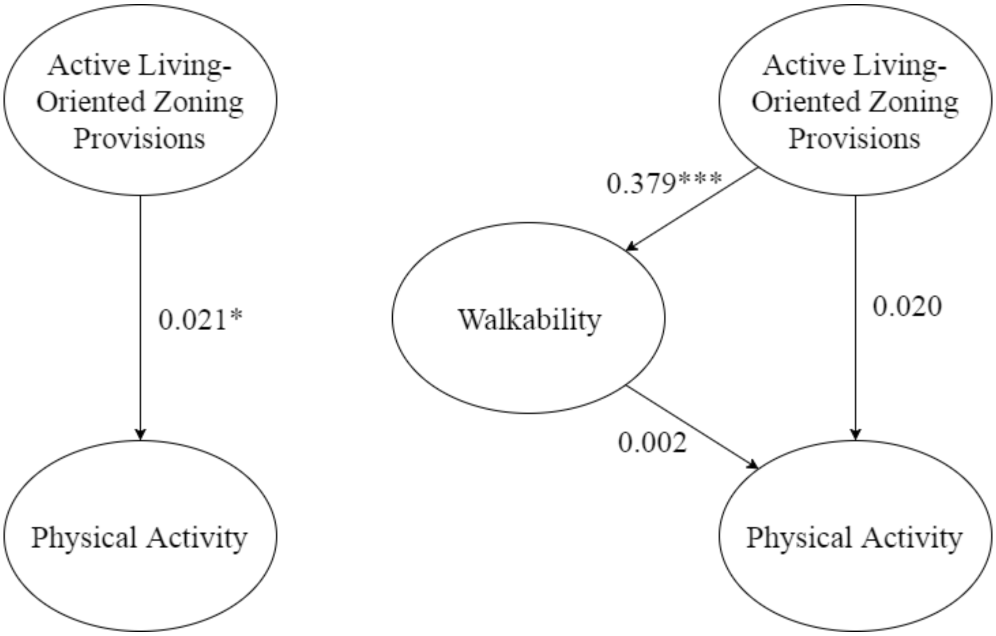


Figure 5.13.1 Physical Activity, IV Probit Marginal Effects

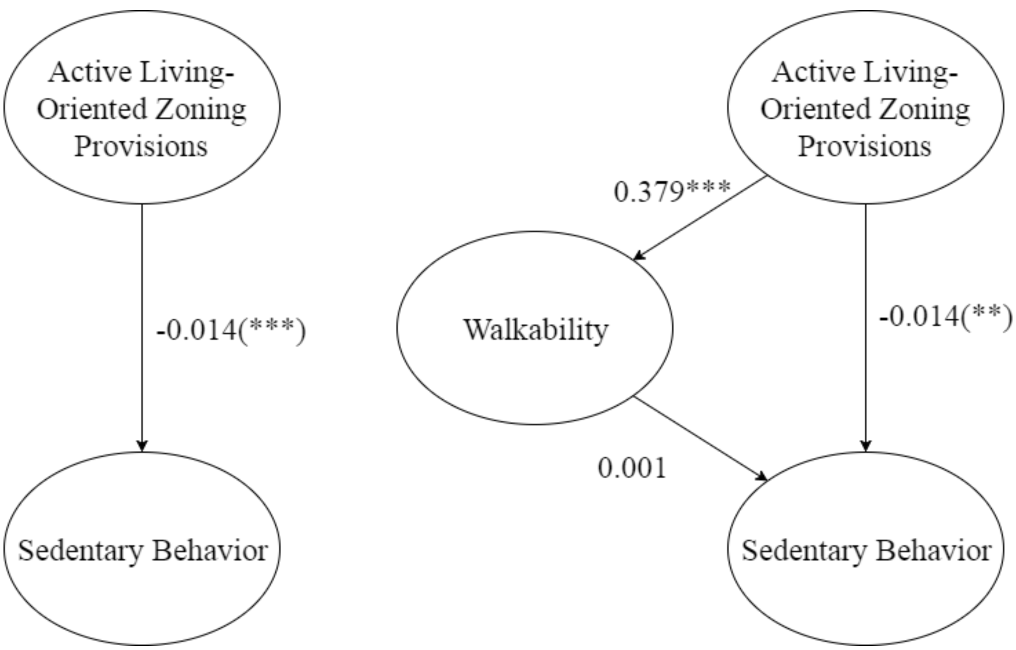


Figure 5.13.2 Sedentary Behavior, IV Probit Marginal Effects
Note: The asterisks in parentheses indicate the significance of corresponding estimated coefficients from IV Probit because the marginal effects calculation procedure is inefficient.

I focus the discussion on the mediational analysis with the IV Probit model because most of the effects of zoning and walkability on the minutes of physical activity and sedentary behavior (continuous outcome) are not significant in analysis either without or with IVs.[18] However, even the effects of zoning and walkability on the minutes are significant, the mediational analysis with IV in linear models may not be valid. IV estimation in linear models makes consistent estimators for average treatment effects for compliers (LATE). In linear models with IV estimation, the compliers of zoning and walkability in Step 3 can be different. The estimated causal effects can be the average effects on different subpopulations, so the mediational analysis is invalid.

The exposition below, which is mainly based on the lecture notes of Imbens (2007), shows that the compliers of two endogenous variables in the same IV estimation can be different. To make discussions clear, suppose the two endogenous variables, Y_1 and Y_2, are dichotomous. I have

$$y_j = Y_{1j}\beta_1 + Y_{2j}\beta_2 + X_j'\gamma + \epsilon_j$$

$$Y_{1j} = 1\{X_j'\Pi_{X1} + Z_j'\Pi_{Z1} + v_{1j} > 0\}$$

$$Y_{2j} = 1\{X_j'\Pi_{X2} + Z_j'\Pi_{Z2} + v_{2j} > 0\} \tag{53}$$

where y is the outcome, X is the vector of included instruments (exogenous covariates), and Z is the vector of excluded instruments (IVs). If $\Pi_{Z1} > 0$, then the compliance types for Y_1 and Y_2 depend on v_1 and $v_2$respectively, and can be classified as

[18] The mediational analysis first step with IV estimation shows significant effects of zoning on the minutes. But the mediational analysis third step with IV estimation shows insignificant effects of zoning on the minutes, which is due to the highly collinearity between predicted zoning and predicted walkability since we use the same set of covariates in IV first stage.

$$\text{unit } j \text{ is a } = \begin{cases} \text{never} - \text{taker} & & \text{if } \ v_{1j} < -X_j'\Pi_{X1} - Z_j'\Pi_{Z1} \\ \text{complier} & \text{for } Y_1 & \text{if } \ -X_j'\Pi_{X1} - Z_j'\Pi_{Z1} \leq v_{1j} < -X_j'\Pi_{X1} \\ \text{always} - \text{taker} & & \text{if } \ -X_j'\Pi_{X1} \leq v_{1j} \end{cases}$$

$$\text{unit } j \text{ is a } = \begin{cases} \text{never} - \text{taker} & & \text{if } \ v_{2j} < -X_j'\Pi_{X2} - Z_j'\Pi_{Z2} \\ \text{complier} & \text{for } Y_2 & \text{if } \ -X_j'\Pi_{X2} - Z_j'\Pi_{Z2} \leq v_{2j} < -X_j'\Pi_{X2} \, . \\ \text{always} - \text{taker} & & \text{if } \ -X_j'\Pi_{X2} \leq v_{1j} \end{cases} \tag{54}$$

For the same individual j, $(\Pi_{X1}\ \Pi_{Z1}\ v_{1j})$ and $(\Pi_{X2}\ \Pi_{Z2}\ v_{2j})$ can be different, which implies that if j is a complier for Y_1, then j could be a never-taker or always-taker for Y_2.

Table 5.13.1 The Mediational Analysis with Instrumental Variables, Step 1: Physical Activity Regression

	Probit		IV Probit		OLS	GMM	OLS Log	GMM Log
VARIABLES	COEFF	Marginal Effect	COEFF	Marginal Effect	COEFF	COEFF	COEFF	COEFF
Active Living-Oriented Zoning	0.190**	0.028**	0.301**	0.021*	3.510	9.957*	0.040	0.297**
	(0.067)	(0.010)	(0.083)	(0.012)	(4.879)	(5.641)	(0.090)	(0.115)
Number of Observations	24,448	24,448	24,448	24,448	1,969	1,969	1,969	1,969
IV Probit								
Endogeneity Test								
Chi2(1)			1.854					
Prob > Chi2(1)			0.173					
Overidentification Test								
Chi2(1)			0.0341					
Prob > Chi2(1)			0.854					
GMM								
Endogeneity Test								
Chi2(1)						1.019		1.942
Prob > Chi2(1)						0.313		0.163
Overidentification Test								
Manufacturing Establishment Density								
Chi2(1)						0.00224		0.518
Prob > Chi2(1)						0.962		0.472
Farmland Proportion								
Chi2(1)						0.00224		0.518
Prob > Chi2(1)						0.962		0.472

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 5.

Table 5.13.2 The Mediational Analysis with Instrumental Variables, Step 1: Sedentary Behavior Regression

	Probit		IV Probit		OLS	GMM	OLS Log	GMM Log
VARIABLES	COEFF	Marginal Effect	COEFF	Marginal Effect	COEFF	COEFF	COEFF	COEFF
Active Living-Oriented Zoning	-0.187***	-	-	-0.014	-7.719	-12.537	-0.035	-0.089**
	(0.067)	(0.017)	(0.089)	(0.022)	(6.580)	(9.843)	(0.033)	(0.041)
Number of Observations	24,458	24,458	24,458	24,458	20,141	20,141	20,141	20,141
IV Probit								
Endogeneity Test								
Chi2(1)			17.20					
Prob > Chi2(1)			0.000					
Overidentification Test								
Chi2(1)			2.458					
Prob > Chi2(1)			0.117					
GMM								
Endogeneity Test								
Chi2(1)						0.254		2.305
Prob > Chi2(1)						0.614		0.129
Overidentification Test								
Manufacturing Establishment Density								
Chi2(1)						1.062		1.826
Prob > Chi2(1)						0.303		0.177
Farmland Proportion								
Chi2(1)						1.062		1.826
Prob > Chi2(1)						0.303		0.177

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 7.

Table 5.13.3 The Mediational Analysis with Instrumental Variables, Step 2: Walkability Regression

VARIABLES	(1)	(2)
Active Living-Oriented Zoning	5.06***	0.379***
	(0.131)	(0.0359)
Manufacturing Establishment Density		0.037***
		(0.0000904)
Farmland Proportion		-0.314***
		(0.0119)
Observations	24,458	24,458
Adjusted R-squared	0.485	0.948

Note: In all specifications, I control for county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), census region fixed effects, and intercept. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 2.

Table 5.13.4 The Mediational Analysis with Instrumental Variables, Step 3: Physical Activity Regression

	Probit		IV Probit		OLS	GMM	OLS Log	GMM Log
VARIABLES	COEFF	Marginal Effect	COEFF	Marginal Effect	COEFF	COEFF	COEFF	COEFF
Active Living-Oriented Zoning	0.148*	0.021*	0.255	0.020	0.431	10.866	-0.063	0.064
	(0.077)	(0.011)	(0.278)	(0.040)	(6.187)	(20.012)	(0.101)	(0.335)
Walkability	0.008*	0.001*	0.003	0.002	0.538	-0.041	0.018***	0.011
	(0.004)	(0.001)	(0.012)	(0.002)	(0.410)	(0.871)	(0.007)	(0.015)
Number of Observations	24,448	24,448	24,448	24,448	1,969	1,969	1,969	1,969
IV Probit Endogeneity Test								
Chi2(1)			0.243					
Prob > Chi2(1)			0.886					
GMM Endogeneity Test								
Active Living-Oriented Zoning								
Chi2(1)						0.248		0.137
Prob > Chi2(1)						0.619		0.711
Walkability								
Chi2(1)						1.105		0.414
Prob > Chi2(1)						0.293		0.520

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 6.

Table 5.13.5 The Mediational Analysis with Instrumental Variables, Step 3: Sedentary Behavior Regression

	Probit		IV Probit		OLS	GMM	OLS Log	GMM Log
VARIABLES	COEFF	Marginal Effect	COEFF	Marginal Effect	COEFF	COEFF	COEFF	COEFF
Active Living-Oriented Zoning	-0.112*	-0.028*	-0.868**	-0.014	-5.360	-37.305	-0.013	-0.266*
	(0.068)	(0.017)	(0.367)	(0.092)	(7.864)	(27.008)	(0.039)	(0.152)
Walkability	-0.014***	-0.003***	0.018	0.001	-0.495	1.257	-0.004	0.009
	(0.004)	(0.001)	(0.017)	(0.004)	(0.685)	(1.207)	(0.003)	(0.007)
Number of Observations	24,458	24,458	24,458	24,458	20,141	20,141	20,141	20,141
IV Probit Endogeneity Test								
Chi2(1)			4.691					
Prob > Chi2(1)			0.0958					
GMM Endogeneity Test								
Active Living-Oriented Zoning								
Chi2(1)						1.421		3.421
Prob > Chi2(1)						0.233		0.0644
Walkability								
Chi2(1)						1.279		1.221
Prob > Chi2(1)						0.258		0.269

Note: In all specifications, I control for individual's demographic information: age, a group of dummies for race (Black, Hispanic, and others, White is the reference group), a dummy for gender, a group of dummies for education level (less than high school, high school graduate, some college, and graduate school, college graduate is the reference group), a group of dummies for marital status (widow, divorced, and separated, single, married is the reference group), county characteristics (% of households in poverty, % non-Hispanic White, % non-Hispanic Black, % Hispanic, median household income, county land area 2010 (square miles)), two dummies indicating "Sunday" and "Saturday," census region fixed effects, month fixed effects, year fixed effects, and intercept. Standard Errors are clustered at county-level. *** p<0.01, ** p<0.05, * p<0.1. The full table is in Appendix 8.

6 POLICY SIMULATION

6.1 Methods

To gain thorough understanding of how zoning policy will affect our leisure-time physical activity and sedentary behavior, I conduct prediction based on estimated models that have consistent and statistically significant estimates of the coefficients of zoning. I only focus on models from Step 1 of the mediational analysis, where walkability is excluded from the analysis, because the coefficients of zoning in Step 1 include both the direct and indirect effects of zoning. The indirect effects come from the mediation of walkability.

Table 6.1.1 summarizes the coefficients of active living-oriented zoning on active living outcomes in Step 1 of the mediational analysis.

Table 6.1.1 The Coefficients of Active Living-Oriented Zoning from Step 1 of the Mediational Analysis

Models	Consistent Estimator	Consistent Estimator from IV Estimation	Note
Probit Model of Physical Activity Participation	0.190***	0.301***	Insignificant endogeneity
Linear Model of the Minutes of Physical Activity	Insignificant	9.957*/29.7%***	Insignificant endogeneity
Probit Model of Sedentary Behavior Participation	N/A	-0.529***	Significant endogeneity
Linear Model of the Log of the Minutes of Sedentary Behavior	Insignificant	-8.9%**	Insignificant endogeneity

Note: The significance of endogeneity comes from endogeneity test, where the test statistic is constructed by IV estimation. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The average predicted probability from the Probit model is

$$\frac{1}{N}\sum_{j=1}^{N}\widehat{\text{Prob}}\left(y_j = 1\middle|Y_j\right) = \frac{1}{N}\sum_{j=1}^{N}\Phi\left(Y_j'\hat{\beta} + X_j'\hat{\gamma}\right). \tag{55}$$

The average predicted probability from the IV Probit model is

$$\frac{1}{N}\sum_{j=1}^{N}\widehat{\text{Prob}}\left(y_j = 1\middle|Y_j\right) = \frac{1}{N}\sum_{j=1}^{N}\Phi\left(\frac{Y_j'\hat{\beta} + X_j'\hat{\gamma} + \left(Y_j' - \left[X_j', Z_j'\right]\widehat{\Pi}\right)\widehat{\Omega}_{22}^{-1/2}\hat{\rho}}{\sqrt{1 - \hat{\rho}'\hat{\rho}}}\right). \tag{31}$$

The average predicted minutes from OLS is

$$\frac{1}{N}\sum_{j=1}^{N}\hat{y}_j = \frac{1}{N}\sum_{j=1}^{N}X_j'\hat{\beta}_{OLS}\,. \tag{56}$$

The average predicted minutes from GMM is

$$\frac{1}{N}\sum_{j=1}^{N}\hat{y}_j = \frac{1}{N}\sum_{j=1}^{N}(P_Z X)_j'\hat{\beta}_{FE2SGMM} \tag{57}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ and $P_Z X$ is the IV first-stage prediction for all regressors including both exogenous and endogenous explanatory variables.

The average predicted minutes from GMM Log is

$$\frac{1}{N}\sum_{j=1}^{N}\hat{y}_j = \frac{1}{N}\sum_{j=1}^{N}\exp\left\{(P_Z X)_j'\hat{\beta}_{FE2SGMM}\right\}\exp\left\{y_j - (P_Z X)_j'\hat{\beta}_{FE2SGMM}\right\} \tag{58}$$

where $y_j - (P_Z X)_j'\hat{\beta}_{FE2SGMM}$ is the predicted residual of the outcome equation.

All predictions are implemented within the analytical sample: the full sample for the binary choice model, the participation sample for the continuous outcome model.

6.2 Results

Table 6.2.1 and Table 6.2.2 show the prediction results. I visualize the prediction results in Figure 6.2.1, Figure 6.2.2, Figure 6.2.3, and Figure 6.2.4. All predictions fit our intuition. As the percentage of the population that was exposed to the promotion of active living-oriented zoning increases, the proportion of the adult population who took part in physical activity increases and the time adults spent on physical activity increases, the proportion of the adult population that choose sedentary behavior decreases, and the time adults spent on sedentary behavior decreases. Compared to the real situation, if all places promoted active living-oriented zoning, then they could increase the proportion of the adult population that engages in physical activity by 1.25 percentage points (Probit model) or 0.95 percentage point (IV Probit model) and the time that adults spend engagning in physical activity by 4 minutes per day (GMM) or 8 minutes per day (GMM Log). It could also decrease the proportion of the adult population that engages in sedentary behavior by 0.6 percentage points (IV Probit) and the time that adults spend engagin in sedentary behavior by 9 minutes per day (GMM Log).

Table 6.2.1 The Simulation Results on Physical Activity

	Physical Activity Prediction			
	Probability		Minutes	
Active Living-Oriented Zoning	Probit	IV Probit	GMM	GMM Log
0	0.066***	0.069***	54.075***	50.378***
	(0.005)	(0.006)	(3.655)	(3.581)
0.1	0.068***	0.071***	55.071***	51.898***
	(0.004)	(0.006)	(3.106)	(3.109)
0.2	0.071***	0.073***	56.066***	53.465***
	(0.004)	(0.005)	(2.563)	(2.613)
0.3	0.073***	0.075***	57.062***	55.079***
	(0.003)	(0.004)	(2.031)	(2.099)
0.4	0.076***	0.077***	58.058***	56.742***
	(0.002)	(0.003)	(1.524)	(1.583)
0.498	0.078***	0.079***	59.033***	58.420***
	(0.002)	(0.002)	(1.083)	(1.127)
0.5	0.078***	0.079***	59.053***	58.455***
	(0.002)	(0.002)	(1.075)	(1.119)
0.6	0.081***	0.081***	60.049***	60.219***
	(0.002)	(0.002)	(0.790)	(0.888)
0.7	0.084***	0.083***	61.045***	62.037***
	(0.002)	(0.002)	(0.855)	(1.129)
0.8	0.087***	0.085***	62.040***	63.910***
	(0.003)	(0.003)	(1.213)	(1.700)
0.9	0.090***	0.088***	63.036***	65.839***
	(0.004)	(0.004)	(1.689)	(2.416)
1	0.093***	0.090***	64.032***	67.827***
	(0.005)	(0.006)	(2.206)	(3.216)
Sample Average	0.0805		60	
Observations	24,448	24,448	1,969	1,969

Note: All predictions are implemented within the analytical sample: the full sample for the binary choice model, the participation sample for continuous outcome model. Standard errors are calculated using the delta method, *** p<0.01, ** p<0.05, * p<0.1

Table 6.2.2 The Simulation Results on Sedentary Behavior

	Sedentary Behavior Prediction	
	Probability	Minutes
Active Living-Oriented Zoning	IV Probit	GMM Log
0	0.831***	242.263***
	(0.013)	(5.783)
0.1	0.830***	240.119***
	(0.011)	(4.783)
0.2	0.829***	237.993***
	(0.009)	(3.821)
0.3	0.827***	235.887***
	(0.007)	(2.918)
0.4	0.826***	233.799***
	(0.005)	(2.124)
0.498	0.825***	231.771***
	(0.003)	(1.583)
0.5	0.825***	231.729***
	(0.003)	(1.576)
0.6	0.823***	229.678***
	(0.003)	(1.526)
0.7	0.822***	227.645***
	(0.004)	(1.987)
0.8	0.820***	225.630***
	(0.005)	(2.693)
0.9	0.819***	223.633***
	(0.007)	(3.486)
1	0.817***	221.654***
	(0.010)	(4.308)
Sample Average	0.823	231
Observations	24,458	20,141

Note: All predictions are implemented within the analytical sample: the full sample for the binary choice model, the participation sample for continuous outcome model. Standard errors are calculated using the delta method, *** p<0.01, ** p<0.05, * p<0.1

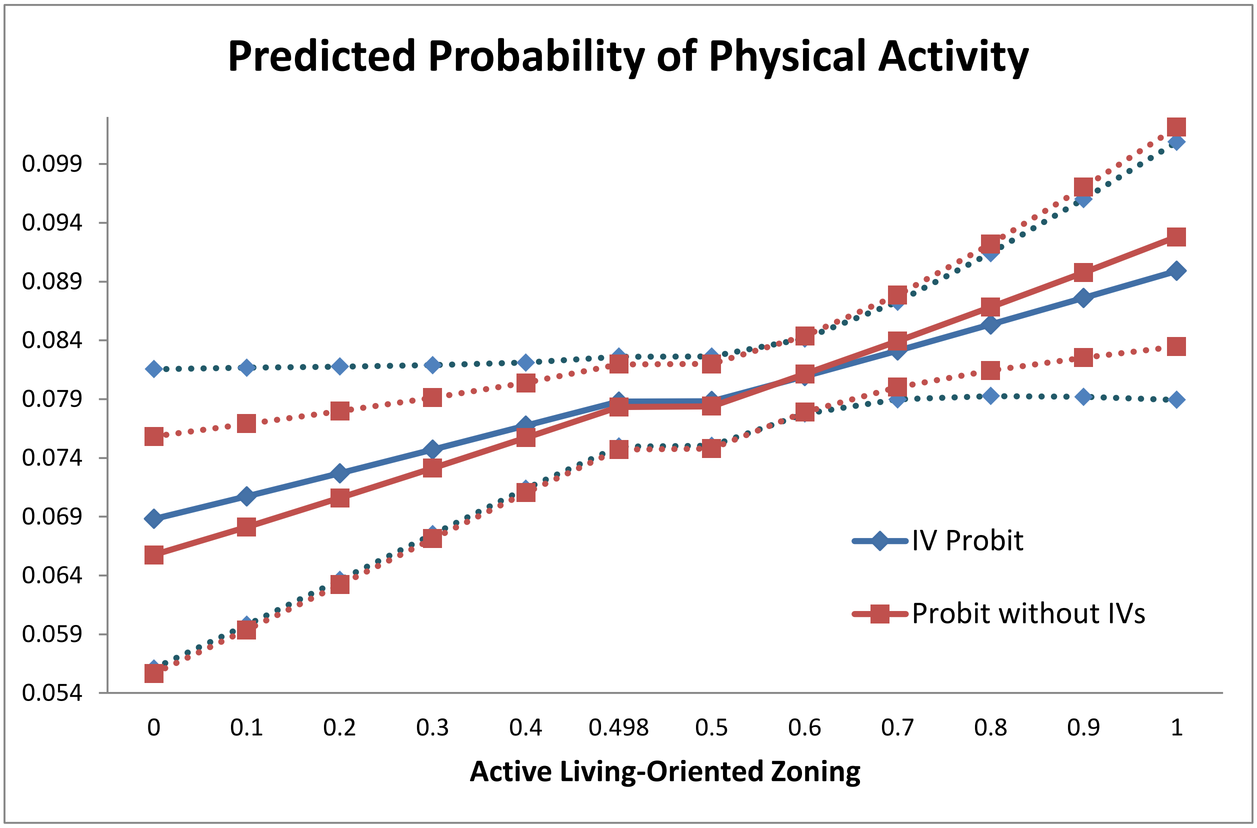


Figure 6.2.1 The Predicted Probability of Physical Activity in the Full Analytical Sample
Note: The sample mean of physical activity participation in the full analytical sample is 0.0805.
Dotted Lines indicate confidence intervals.

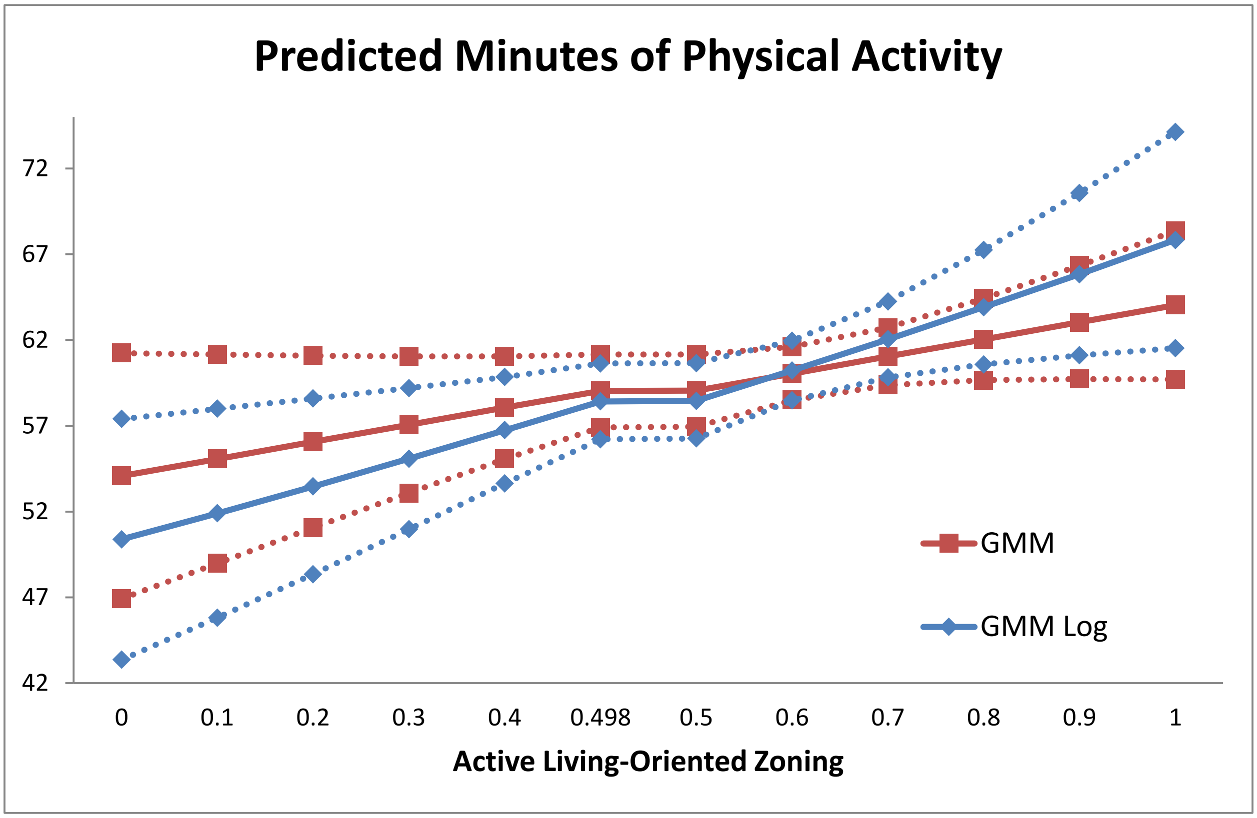


Figure 6.2.2 The Predicted Minutes of Physical Activity within the Analytical Sample of Physical Activity Participation

Note: The sample mean of the minutes of physical activity in the analytical sample of physical activity participation is 60. Dotted Lines indicate confidence intervals.

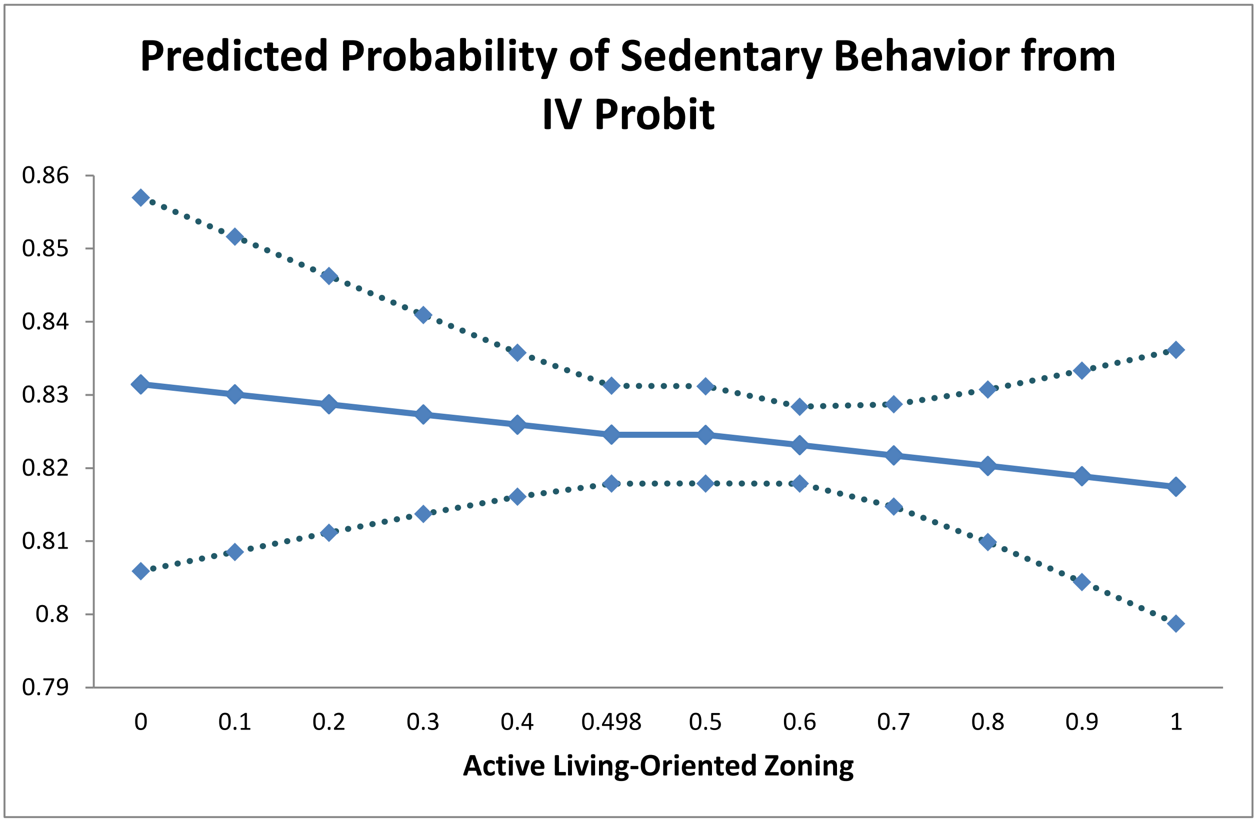


Figure 6.2.3 The Predicted Probability of Sedentary Behavior in the Full Analytical Sample
Note: The sample mean of sedentary behavior participation in the full analytical sample is 0.823. Dotted Lines indicate confidence interval.

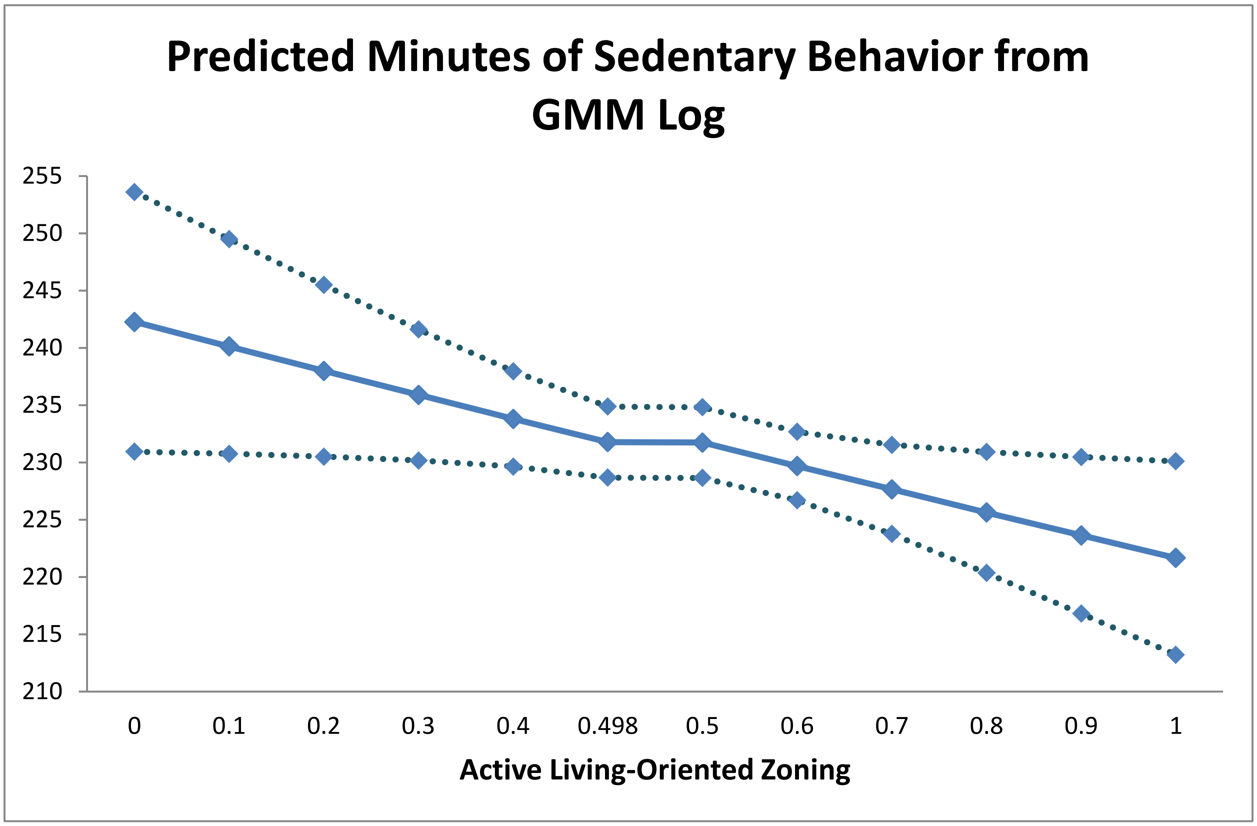


Figure 6.2.4 The Predicted Minutes of Sedentary Behavior within the Analytical Sample of Sedentary Behavior Participation

Note: The sample mean of the minutes of sedentary behavior in the analytical sample of sedentary behavior participation is 231. Dotted Lines indicate confidence interval.

7 CONCLUSION

This is the first study to investigate the causal effects of zoning on active living outcomes and quantitatively verify the underlying mechanisms (i.e., the mediation of built environment in the causal path from zoning to active living outcomes). I introduce the causal inference to the literature of the associations among zoning, built environment, and active living outcomes. Previous literature shows that there is a significant positive association between active living-oriented zoning and active living outcomes. Because people choose where they live, the observed zoning and built environment can be the choice results of people. Whether the association is causal, however, has not been established in the literature. While previous studies hypothesize that zoning affects active living outcomes by shaping the built environment, no study has quantitatively verified the mediation of built environment.

I exploit the plausibly exogenous variation in active living-oriented zoning and built environment caused by manufacturing establishment density and farmland proportion in 1900 that represents the conflict between the rapidly-developing manufacturing industry and the traditional agricultural economy in the late 19th century. I explore the history and the institution of American zoning and conduct statistical analysis to argue for the validity of the two IVs. By utilizing the technique of GMM, an IV estimation technique, I test the endogeneity of active living-oriented zoning and estimate its causal effects on active living outcomes.

Based on IV estimation, a one percent increase in the population that was exposed to the promotion of active living-oriented zoning is associated with a 2.8 percentage point increase (IV Probit) in the likelihood that an adult will engage in physical activity, a 9.957 unit increase (GMM) or 29.7 percent increase (GMM Log) in the minutes an adult spent on physical activity, a

1.4 percentage point decrease (IV Probit) in the chances that an adult chose sedentary behavior, and a 8.9 percent decrease (GMM Log) in the minutes an adult spent on sedentary behavior. Although the marginal effects of zoning on participation probabilities are nearly insignificant due to the inefficiency of IV estimation and the marginal effect calculation procedure, the corresponding estimated coefficients in IV Probit are significant, which means the effects of zoning are significant. I conduct endogeneity tests, which are also based on IV estimation. For sedentary behavior participation, the endogeneity is significant. For other models, the endogeneity is insignificant. Nonetheless, IV estimation for linear models corrects the measurement errors caused by the impossibility of linking survey respondents to the zoning metrics of the specific areas of their county. Note that the marginal effects of zoning on the participation probabilities are causal effects on the whole population while the marginal effects of zoning on the minutes are the weighted average of LATE in the sense of the causal effects on compliers whose behaviors change as the values of IVs change.

The significant effects of active living-oriented zoning on physical activity and sedentary behavior align with our expectations very well. The metric of active living-oriented zoning consists of 12 zoning provisions that are supposed to encourage physical activity and discourage sedentary behavior. Specifically, sidewalks, crosswalks, street connectivity, and other walkability provisions are expected to encourage walking, jogging, and running. Bike lanes and bike parking are expected to promote cycling. Bike-pedestrian connectivity, bike-pedestrian trails-paths, code reform, mixed-use spaces, active recreation, and passive recreation are expected to promote both pedestrian- and cyclist-activities. Meanwhile, all of these zoning provisions are expected to discourage sedentary behavior at home, such as TV watching and radio/music listening.

In physical activity participation analysis, walkability is a mediator for zoning. This is concluded from the Probit model because the endogeneity tests show insignificant endogeneity in the physical activity participation analysis. In the sedentary behavior participation analysis, walkability is not a mediator for zoning. This is concluded from IV Probit because the endogeneity tests show significant endogeneity in the sedentary behavior participation analysis. For the duration of physical activity and sedentary behavior, the mediation of walkability for zoning cannot be established because most of the coefficients are insignificant in both OLS and GMM. In addition, I show that the mediational analysis in linear models with IV estimation can be invalid because the compilers of zoning and walkability in Step 3 can be different.

The failure of mediation of walkability in the causal path from zoning to sedentary behavior participation implies that some built environment elements that cannot be captured by the walkability metric play important roles in people's sedentary behavior participation decision. The zoning variable, active living-oriented zoning, includes many more elements of built environment than walkability, active recreation (equipment to support physical activity), passive recreation (opportunities for physical activity, like open space), mixed-use development, bike parking, and bike lanes, to name a few. Many of these zoning provisions cannot be captured by the walkability metric, which is constituted by four density metrics, including population density, housing unit density, the ratio of four-way intersections to all intersections, and the total number of intersections divided by land area.

Based on the estimated models that have consistent and significant coefficients of zoning, I conduct policy simulation to make within-analytical sample predictions. Compared to the real situation, if all places promoted active living-oriented zoning, then this would increase the proportion of the adult population that engages in physical activity by 1.25 percentage points

(Probit) or 0.95 percentage point (IV Probit) and the time that adults spend doing physical activity by 4 minutes per day (GMM) or 8 minutes per day (GMM Log). It could decrease the proportion of the adult population that engages in sedentary behavior by 0.6 percentage point (IV Probit) and the time that adults spend doing sedentary behavior by 9 minutes per day (GMM Log).

Overall, the results of this dissertation suggest that active living-oriented zoning, indeed, causally promotes physical activities, such as walking, running, and bicycling, and causally discourages sedentary behaviors at home, such as TV watching and music listening, partly by shaping the built environment. These findings enhance our understanding of the correlations between active living-oriented zoning and leisure-time physical activity/no leisure-time physical activity found by previous research (Chriqui, Nicholson, et al. 2016; Leider, Chriqui, and Thrun 2016).

This dissertation serves as a sound foundation for policy and strategic plan makers to address pedestrian- and cyclist-friendly zoning provisions to promote active living. Although the movement of zoning code reform started in the 1990s, most places have not changed their zoning regulations for decades. Today, traditional Euclidean zoning still dominates most communities across the country. An increasing number of professionals, researchers, and leaders of public and private organizations are realizing that traditional Euclidean zoning has resulted in the urban sprawl and a lack of active living. They advocate for local governments to upgrade built environment to promote physical activity. The Committee on Childhood Obesity Prevention Actions for Local Governments recommends that improvements to the built environment (e.g., a network of sidewalks and street crossing, a well-connected network of off-street trails and paths for pedestrians and bicyclists, etc.) can encourage walking and bicycling for transportation and recreation (Transportation Research Board, Institute of Medicine, and National Research Council 2009). The Committee on Accelerating Progress in Obesity Prevention suggests that "communities, organizations, community planners, and public health professionals should encourage physical activity by enhancing the physical and built environment, rethinking community design, and ensuring access to places for such activity" (Institute of Medicine 2012). The U.S. National Physical Activity Plan insists that "we must be intentional in the design and development of our communities to make it easier for people to be active on a daily basis" (National Physical Activity Plan Alliance 2016). They provide detailed strategies for the development of active living-friendly built environment. This study supports that implementing and updating zoning laws could be an effective way to upgrade the built environment to promote physical activity. Future zoning policy should address pedestrian- and cyclist-friendly zoning provisions.

REFERENCES

Aguiar, Mark, and Erik Hurst. 2007. "Life-Cycle Prices and Production." *American Economic Review* 97 (5): 1533–59. doi:10.1257/aer.97.5.1533.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2002. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." https://pdfs.semanticscholar.org/efee/be6138776d1f0405185a501409cf15515ea0.pdf.

———. 2005a. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* XL (4): 791–821. doi:10.3368/jhr.XL.4.791.

———. 2005b. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1). The University of Chicago Press: 151–84. doi:10.1086/426036.

American Planning Association. 2006. *Planning and Urban Design Standards*. John Wiley & Sons.

Anderson, James M., John M. MacDonald, Ricky Bluthenthal, and J. Scott Ashwood. 2013. "Reducing Crime by Shaping the Built Environment with Zoning: An Empirical Study of Los Angeles." *University of Pennsylvania Law Review*. http://www.scopus.com/inward/record.url?eid=2-s2.0-84874044836&partnerID=tZOtx3y1.

Angrist, Joshua D., and William N. Evans. 2017. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review*. American Economic Association. Accessed August 6. doi:10.2307/116844.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal

Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434). Taylor & Francis, Ltd.American Statistical Association: 444. doi:10.2307/2291629.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4). Oxford University Press: 979–1014. doi:10.2307/2937954.

———. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4): 69–85. doi:10.1257/jep.15.4.69.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. http://press.princeton.edu/titles/8769.html.

Aytur, Semra A., Daniel A. Rodriguez, Kelly R. Evenson, and Diane J. Catellier. 2008. "Urban Containment Policies and Physical Activity." *American Journal of Preventive Medicine* 34 (4): 320–32. doi:10.1016/j.amepre.2008.01.018.

Baily, Martin Neil, and Barry P. Bosworth. 2014. "US Manufacturing: Understanding Its Past and Its Potential Future." *Journal of Economic Perspectives* 28 (1): 3–26. doi:10.1257/jep.28.1.3.

Baron, Reuben M., and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173–82. doi:10.1037/0022-3514.51.6.1173.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. 2003. "Instrumental Variables and GMM: Estimation and Testing." *Stata Journal* 3 (1): 1–31.

———. 2016. *IVREG2: Stata Module for Extended Instrumental variables/2SLS and GMM Estimation*. *Statistical Software Components*.

Berke, Ethan M., Thomas D. Koepsell, Anne Vernez Moudon, Richard E. Hoskins, and Eric B. Larson. 2007. “Association of the Built Environment With Physical Activity and Obesity in Older Persons.” *American Journal of Public Health* 97 (3): 486–92. doi:10.2105/AJPH.2006.085837.

Berrigan, David, and Richard P Troiano. 2002. “The Association between Urban Form and Physical Activity in U.S. Adults.” *American Journal of Preventive Medicine* 23 (2 Suppl): 74–79. http://www.ncbi.nlm.nih.gov/pubmed/12133740.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. “Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak.” *Journal of the American Statistical Association*. doi:10.1080/01621459.1995.10476536.

Bound, John, David a Jaeger, and Regina Baker. 1993. “The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables.” *NBER Technical Paper Series*.

Brownson, Ross, Elizabeth Baker, Robyn Housemann, Laura Brennan, and Stephen Bacak. 2001. “Environmental and Policy Determinants of Physical Activity in the United States.” *American Journal of Public Health* 91 (12). American Public Health Association: 1995–2003. doi:10.2105/AJPH.91.12.1995.

Cannon, C. L., S. Thomas, R. D. Treffers, M. J. Paschall, L. Heumann, G. W. Mann, D. O. Dunkell, and S. Nauenberg. 2013. “Testing the Results of Municipal Mixed-Use Zoning Ordinances: A Novel Methodological Approach.” *Journal of Health Politics, Policy and*

Law 38 (4): 815–39. doi:10.1215/03616878-2208612.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127–60. http://links.jstor.org/sici?sici=0012-9682%28200109%2969%3A5%3C1127%3AETRTSP%3E2.0.CO%3B2-Q.

Cervero, Robert, and Michael Duncan. 2003. "Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area." *American Journal of Public Health* 93 (9). American Public Health Association: 1478–83. doi:10.2105/AJPH.93.9.1478.

Cervero, Robert, and Kara Kockelman. 1997. "Travel Demand and the 3Ds: Density, Diversity, and Design." *Transportation Research Part D: Transport and Environment* 2 (3): 199–219. doi:10.1016/S1361-9209(97)00009-6.

Chriqui, Jamie F., Julien Leider, Emily Thrun, Lisa M. Nicholson, and Sandy Slater. 2016. "Communities on the Move: Pedestrian-Oriented Zoning as a Facilitator of Adult Active Travel to Work in the United States." *Frontiers in Public Health* 4 (April). Frontiers: 71. doi:10.3389/fpubh.2016.00071.

Chriqui, Jamie F., Lisa M. Nicholson, Emily Thrun, Julien Leider, and Sandy J. Slater. 2016. "More Active Living-Oriented County and Municipal Zoning Is Associated With Increased Adult Leisure Time Physical Activity--United States, 2011." *Environment and Behavior* 48 (1). SAGE Publications: 111–30. doi:10.1177/0013916515611175.

Colman, Gregory, and Dhaval Dave. 2013. "Exercise, Physical Activity, and Exertion over the Business Cycle." *Social Science & Medicine* 93 (September): 11–20. doi:10.1016/j.socscimed.2013.05.032.

Connelly, Rachel, and Jean Kimmel. 2007. *Spousal Influences on Parents' Non-Market Time*

Choices. http://ftp.iza.org/dp2894.pdf.

Connolly, Marie. 2008. “Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure.” *Journal of Labor Economics* 26 (1): 73–100. doi:10.1086/522067.

Cragg, John G., and Stephen G. Donald. 1993. “Testing Identifiability and Specification in Instrumental Variable Models.” *Econometric Theory*. doi:10.1017/S0266466600007519.

Davison, Kirsten Krahnstoever, and Catherine T Lawson. 2006. “Do Attributes in the Physical Environment Influence Children’s Physical Activity? A Review of the Literature.” *The International Journal of Behavioral Nutrition and Physical Activity* 3 (July). BioMed Central: 19. doi:10.1186/1479-5868-3-19.

Dunlap, David. 2016. “Zoning Arrived 100 Years Ago. It Changed New York City Forever.” *The New York Times*, July 26. http://www.nytimes.com/2016/07/26/nyregion/new-yorks-first-zoning-resolution-which-brought-order-to-a-chaotic-building-boom-turns-100.html?_r=1.

Dunton, G F, D Berrigan, R Ballard-Barbash, B Graubard, and A A Atienza. 2009. “Joint Associations of Physical Activity and Sedentary Behaviors with Body Mass Index: Results from a Time Use Survey of US Adults.” *International Journal of Obesity* 33 (12): 1427–36. doi:10.1038/ijo.2009.174.

Evans, William N., and Robert M. Schwab. 1995. “Finishing High School and Starting College: Do Catholic Schools Make a Difference?” *The Quarterly Journal of Economics* 110 (4). Oxford University Press: 941–74. doi:10.2307/2946645.

Ewing, Reid, and Shima Hamidi. 2014. “Measuring Urban Sprawl and Validating Sprawl Measures.” Washington, DC. https://gis.cancer.gov/tools/urban-sprawl/sprawl-report-short.pdf.

Ewing, Reid, Tom Schmid, Richard Killingsworth, Amy Zlot, and Stephen Raudenbush. 2003. “Relationship between Urban Sprawl and Physical Activity, Obesity, and Morbidity.” *American Journal of Health Promotion : AJHP* 18 (1). SAGE Publications: 47–57. doi:10.4278/0890-1171-18.1.47.

Ewing, Reid, Guang Tian, JP Goates, Ming Zhang, Michael Greenwald, Alex Joyce, John Kircher, and Wlliam Greene. 2015. “Varying Influences of the Built Environment on Household Travel in 15 Diverse Regions of the United States.” *Urban Studies* 52 (13). SAGE Publications: 2330–48. doi:10.1177/0042098014560991.

Fischel, William. 2004. “An Economic History of Zoning and a Cure for Its Exclusionary Effects.” *Urban Studies* 41 (2). SAGE Publications: 317–40. doi:10.1080/0042098032000165271.

Frank, Lawrence D., Martin A. Andresen, and Thomas L. Schmid. 2004. “Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars.” *American Journal of Preventive Medicine* 27 (2): 87–96. doi:10.1016/j.amepre.2004.04.011.

Frank, Lawrence D., Peter O. Engelke, and Thomas L. Schmid. 2003. *Health and Community Design : The Impact of the Built Environment on Physical Activity*. 2nded. Island Press.

Frank, Lawrence D., James F. Sallis, Terry L. Conway, James E. Chapman, Brian E. Saelens, and William Bachman. 2006. “Many Pathways from Land Use to Health: Associations between Neighborhood Walkability and Active Transportation, Body Mass Index, and Air Quality.” *Journal of the American Planning Association* 72 (1). Taylor & Francis Group : 75–87. doi:10.1080/01944360608976725.

Frank, Lawrence D., Thomas L. Schmid, James F. Sallis, James Chapman, and Brian E. Saelens. 2005. “Linking Objectively Measured Physical Activity with Objectively Measured Urban

Form: Findings from SMARTRAQ." *American Journal of Preventive Medicine* 28 (2): 117–25. doi:10.1016/j.amepre.2004.11.001.

Frank, Lawrence D, Brian E. Saelens, Ken E. Powell, and James E. Chapman. 2007. "Stepping towards Causation: Do Built Environments or Neighborhood and Travel Preferences Explain Physical Activity, Driving, and Obesity?" *Social Science & Medicine* 65 (9): 1898–1914. doi:10.1016/j.socscimed.2007.05.053.

Goldstein, G.S., and M.V. Pauly. 1981. "Tiebout Bias on the Demand for Local Public Goods." *Journal of Public Economics* 16 (2): 131–43. doi:10.1016/0047-2727(81)90021-9.

Gordon-Larsen, Penny, Melissa C Nelson, Phil Page, and Barry M Popkin. 2006. "Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity." *Pediatrics* 117 (2): 417–24. doi:10.1542/peds.2005-0058.

Ham, Sandra A, Judy Kruger, and Catrine Tudor-Locke. 2009. "Participation by US Adults in Sports, Exercise, and Recreational Physical Activities." *Journal of Physical Activity & Health* 6 (1): 6–14. http://www.ncbi.nlm.nih.gov/pubmed/19211953.

Hamermesh, Daniel S. 2005. "Why Not Retire? The Time and Timing Costs of Market Work." *Working Papers*. University of Michigan, Michigan Retirement Research Center. https://ideas.repec.org/p/mrr/papers/wp104.html.

———. 2007. "Time to Eat: Household Production under Increasing Income Inequality." *American Journal of Agricultural Economics* 89 (4): 852–63. doi:10.1111/j.1467-8276.2007.01012.x.

———. 2008. "Direct Estimates of Household Production." *Economics Letters* 98 (1). Elsevier: 31–34. https://ideas.repec.org/a/eee/ecolet/v98y2008i1p31-34.html.

Hamermesh, Daniel S., and Stephen Donald. 2007. "The Time and Timing Costs of Market

Work." *NBER Working Papers*. National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/13127.html.

Hamermesh, Daniel S., Harley Frazis, and Jay Stewart. 2005. "Data Watch: The American Time Use Survey." *Journal of Economic Perspectives*. doi:10.1257/0895330053148029.

Hamermesh, Daniel S., and Jungmin Lee. 2007. "Stressed Out on Four Continents: Time Crunch or Yuppie Kvetch?" *Review of Economics and Statistics* 89 (2): 374–83. doi:10.1162/rest.89.2.374.

Handy, Susan, Xinyu Cao, and Patricia Mokhtarian. 2005. "Correlation or Causality between the Built Environment and Travel Behavior? Evidence from Northern California." *Transportation Research Part D: Transport and Environment* 10 (6): 427–44. doi:10.1016/j.trd.2005.05.002.

Handy, Susan, Xinyu Cao, and Patricia L. Mokhtarian. 2006. "Self-Selection in the Relationship between the Built Environment and Walking: Empirical Evidence from Northern California." *Journal of the American Planning Association* 72 (1). Taylor & Francis Group : 55–74. doi:10.1080/01944360608976724.

Hansen, Jeff, Joseph D. Bryant, and Thomas Spencer. 2007. "WHICH WAY FORWARD?" *AL.com and The Birmingham News* . http://blog.al.com/bn/2007/03/which_way_forward.html.

Harmon, Colm, and Hessel Oosterbeek. 2000. "The Returns to Education: A Review of Evidence, Issues and Deficiencies in the Literature." *CEE Discussion Papers*. Centre for the Economics of Education, LSE. https://ideas.repec.org/p/cep/ceedps/0005.html.

Heath, Gregory W., Ross Brownson, Judy Kruger, Rebecca Miles, Kenneth E. Powell, and Leigh T. Ramsey. 2006. "The Effectiveness of Urban Design and Land Use and Transport Policies and Practices to Increase Physical Activity: A Systematic Review." *Journal of Physical Activity and Health* 3 (s1): S55–76. doi:10.1123/jpah.3.s1.s55.

Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3). The Econometric Society: 669–738. doi:10.2307/3598865.

Hirt, Sonia A. 2014. *Zoned in the USA: The Origins and Implications of American Land-Use Regulation*. *Cornell University Press*. Ithaca ; London: Cornell University Press,. doi:10.1080/02665433.2015.1063322.

Hofferth, Sandra L., Sarah M. Flood, and Matthew Sobek. 2015. "American Time Use Survey Data Extract System: Version 2.5 [Machine-Readable Database]." Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota. https://www.atusdata.org/atus/index.shtml.

Imbens, Guido W. 2007. "Lecture 05: Instrumental Variables with Treatment Effect Heterogeneity: Local Average Treatment Effects." In *NBER Lecture What's New in Econometrics*. https://www.mendeley.com/library/.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467. doi:10.2307/2951620.

Institute of Medicine. 2012. *Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation*. Washington, DC, DC: The National Academies Press. doi:10.17226/13275.

James, Lawrence R., and Jeanne M. Brett. 1984. "Mediators, Moderators, and Tests for Mediation." *Journal of Applied Psychology*. doi:10.1037/0021-9010.69.2.307.

Judd, C. M., and D. A. Kenny. 1981. "Process Analysis: Estimating Mediation in Treatment

Evaluations." *Evaluation Review*. doi:10.1177/0193841X8100500502.

Kalenkoski, Charlene M., and Karen S. Hamrick. 2013. "How Does Time Poverty Affect Behavior? A Look at Eating and Physical Activity." *Applied Economic Perspectives and Policy*. doi:10.1093/aepp/pps034.

Kalenkoski, Charlene M., David C. Ribar, and Leslie S. Stratton. 2007. "The Effect of Family Structure on Parents' Child Care Time in the United States and the United Kingdom." *Review of Economics of the Household* 5 (4): 353–84. doi:10.1007/s11150-007-9017-y.

Kenny, David A. 2016. "Mediation." http://davidakenny.net/cm/mediate.htm.

Kim, Sukkoo. 1995. "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S." *Quarterly Journal of Economics*. doi:10.2307/2946643.

———. 1999. "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880–1987." *Regional Science and Urban Economics*. doi:10.1016/S0166-0462(98)00010-6.

Kimmel, Jean, and Rachel Connelly. 2006. "Is Mothers' Time With Their Children Home Production or Leisure?" *IZA Discussion Papers*. Institute for the Study of Labor (IZA). https://ideas.repec.org/p/iza/izadps/dp2058.html.

Klein, Alexander, and Nicholas Crafts. 2012. "Making Sense of the Manufacturing Belt: Determinants of U.S. Industrial Location, 1880-1920." *Journal of Economic Geography*. doi:10.1093/jeg/lbr023.

Krugman, Paul. 1991a. "History and Industry Location: The Case of the Manufacturing Belt." *The American Economic Review* 81 (2). American Economic Association: 80–83. http://www.jstor.org/stable/2006830.

———. 1991b. "Increasing Returns and Economic Geography." *Journal of Political Economy*.

doi:10.1086/261763.

Krugman, Paul, and Anthony J. Venables. 1995. "Globalization and the Inequality of Nations." *The Quarterly Journal of Economics*. doi:10.2307/2946642.

Leider, Julien, Jamie F. Chriqui, and Emily Thrun. 2016. "Associations between Active Living-Oriented Zoning and No Adult Leisure-Time Physical Activity in the U.S." *Preventive Medicine*, June. doi:10.1016/j.ypmed.2016.06.029.

Li, Fuzhong, John Fisher, Ross Brownson, and Mark Bosworth. 2005. "Multilevel Modelling of Built Environment Characteristics Related to Neighbourhood Walking Activity in Older Adults." *Journal of Epidemiology and Community Health* 59 (7). BMJ Group: 558–64. doi:10.1136/jech.2004.028399.

Li, Fuzhong, Peter A. Harmer, Bradley J. Cardinal, Mark Bosworth, Alan Acock, Deborah Johnson-Shelton, and Jane M. Moore. 2008. "Built Environment, Adiposity, and Physical Activity in Adults Aged 50–75." *American Journal of Preventive Medicine* 35 (1): 38–46. doi:10.1016/j.amepre.2008.03.021.

MacDonald, John M., Robert J. Stokes, Deborah A. Cohen, Aaron Kofner, and Greg K. Ridgeway. 2010. "The Effect of Light Rail Transit on Body Mass Index and Physical Activity." *American Journal of Preventive Medicine* 39 (2): 105–12. doi:10.1016/j.amepre.2010.03.016.

Mackinnon, David P., Ghulam Warsi, and James H. Dwyer. 1995. "A Simulation Study of Mediated Effect Measures." *Multivariate Behavioral Research* 30 (1). NIH Public Access: 41. doi:10.1207/s15327906mbr3001_3.

McCormack, Gavin R., and Alan Shiell. 2011. "In Search of Causality: A Systematic Review of the Relationship between the Built Environment and Physical Activity among Adults."

International Journal of Behavioral Nutrition and Physical Activity 8 (1). BioMed Central: 125. doi:10.1186/1479-5868-8-125.

Meyer, Bruce D., and James X. Sullivan. 2006. “Consumption, Income, and Material Well-Being After Welfare Reform.” *NBER Working Papers*. National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/11976.html.

Minnesota Population Center. 2016. *National Historical Geographic Information System: Version 11.0 [Database]*. Minneapolis: University of Minnesota. doi:10.18128/D050.V11.0.

Mullahy, John, and Stephanie A. Robert. 2010. “No Time to Lose: Time Constraints and Physical Activity in the Production of Health.” *Review of Economics of the Household* 8 (4). Springer US: 409–32. doi:10.1007/s11150-010-9091-4.

National Physical Activity Plan Alliance. 2016. “U.S. National Physical Activity Plan.” SC. http://physicalactivityplan.org/docs/2016NPAP_Finalforwebsite.pdf.

Ng, S. W., and B. M. Popkin. 2012. “Time Use and Physical Activity: A Shift Away from Movement across the Globe.” *Obesity Reviews* 13 (8): 659–80. doi:10.1111/j.1467-789X.2011.00982.x.

Nolon, John R. 2006. “Historical Overview of the American Land Use System: A Diagnostic Approach to Evaluating Governmental Land Use Control.”

Norton, Richard K. 2008. “Using Content Analysis to Evaluate Local Master Plans and Zoning Codes.” *Land Use Policy* 25 (3): 432–54. doi:10.1016/j.landusepol.2007.10.006.

O’Connell, Lenahan. 2008. “Exploring the Social Roots of Smart Growth Policy Adoption by Cities.” *Social Science Quarterly* 89 (5): 1356–72. doi:10.1111/j.1540-6237.2008.00581.x.

Office of Disease Prevention and Health Promotion. 2017. “Healthy People 2020 Data for

Physical Activity." https://www.healthypeople.gov/2020/data-search/Search-the-Data#topic-area=3504;

Pendall, Rolf, Robert Puentes, and Jonathon Martin. 2006. "From Traditional to Reformed: A Review of the Land Use Regulations in the Nation's 50 Largest Metropolitan Areas." *Brookings Institution Research Brief*, 1–39. http://www.brookings.edu/metro/pubs/20060802_Pendall.pdf.

Perdue, Wendy Collins, Lesley A Stone, and Lawrence O Gostin. 2003. "The Built Environment and Its Relationship to the Public's Health: The Legal Framework." *American Journal of Public Health* 93 (9). American Public Health Association: 1390–94. doi:10.2105/AJPH.93.9.1390.

"Railroads in the Late 19th Century." 2017. *Library of Congress*. Accessed June 6. http://www.loc.gov/teachers/classroommaterials/presentationsandactivities/presentations/timeline/riseind/railroad/.

Rodríguez, Daniel A., Asad J. Khattak, and Kelly R. Evenson. 2006. "Can New Urbanism Encourage Physical Activity?: Comparing a New Urbanist Neighborhood with Conventional Suburbs." *Journal of the American Planning Association* 72 (1). Taylor & Francis Group : 43–54. doi:10.1080/01944360608976723.

Rundle, Andrew, Ana V Diez Roux, Lance M Free, Douglas Miller, Kathryn M Neckerman, and Christopher C Weiss. 2007. "The Urban Built Environment and Obesity in New York City: A Multilevel Analysis." *American Journal of Health Promotion : AJHP* 21 (4 Suppl). SAGE Publications: 326–34. doi:10.4278/0890-1171-21.4S.326.

"Rural Life in the Late 19th Century." 2017. *Library of Congress*. Accessed September 21. http://www.loc.gov/teachers/classroommaterials/presentationsandactivities/presentations/ti meline/riseind/rural/.

Saelens, Brian E., James F. Sallis, Jennifer B. Black, and Diana Chen. 2003. "Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation." *American Journal of Public Health* 93 (9). American Public Health Association: 1552–58. http://www.ncbi.nlm.nih.gov/pubmed/12948979.

Sallis, James F., Robert B. Cervero, William Ascher, Karla A. Henderson, M. Katherine Kraft, and Jacqueline Kerr. 2006. "AN ECOLOGICAL APPROACH TO CREATING ACTIVE LIVING COMMUNITIES." *Annual Review of Public Health* 27 (1). Annual Reviews: 297–322. doi:10.1146/annurev.publhealth.27.021405.102100.

Sallis, James F., Myron F. Floyd, Daniel A. Rodríguez, and Brian E. Saelens. 2012. "Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease." *Circulation* 125 (5).

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190 (2): 212–21. doi:10.1016/j.jeconom.2015.06.004.

Schilling, Joseph. 2005. "Creating a Regulatory Blueprint for Healthy Community Design | Icma.org." Washington, DC . http://icma.org/en/icma/knowledge_network/documents/kn/Document/2810/Creating_a_Regulatory_Blueprint_for_Healthy_Community_Design.

Schilling, Joseph, and Leslie S. Linton. 2005. "The Public Health Roots of Zoning: In Search of Active Living's Legal Genealogy." *American Journal of Preventive Medicine* 28 (2 SUPPL. 2): 96–104. doi:10.1016/j.amepre.2004.10.028.

Sener, Ipek, Chandra Bhat, and Ram Pendyala. 2011. "When, Where, How Long, and with

Whom Are Individuals Participating in Physically Active Recreational Episodes?” *Transportation Letters* 3 (3). Maney Publishing: 201–17. doi:10.3328/TL.2011.03.03.201-217.

Skeels, Christopher L., and Larry W. Taylor. 2015. “Prediction in Linear Index Models with Endogenous Regressors.” *Stata Journal* 15 (3). College Station, TX, TX: Stata Press: 627–44. http://www.stata-journal.com/article.html?article=st0397.

Slater, Sandy, Reid Ewing, Lisa Powell, Frank Chaloupka, Lloyd Johnston, and Patrick O’Malley. 2010. “The Association between Community Physical Activity Settings and Youth Physical Activity, Obesity, and Body Mass Index.” *The Journal of Adolescent Health* 47 (5): 496–503. doi:10.1016/j.jadohealth.2010.03.017.

Smith, Lindsey P., Shu Wen Ng, and Barry M. Popkin. 2014. “No Time for the Gym? Housework and Other Non-Labor Market Time Use Patterns Are Associated with Meeting Physical Activity Recommendations among Adults in Full-Time, Sedentary Jobs.” *Social Science & Medicine* 120 (November): 126–34. doi:10.1016/j.socscimed.2014.09.010.

Sobel, Michael E. 1982. “Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models.” *Sociological Methodology* 13. Wiley: 290. doi:10.2307/270723.

Staiger, Douglas, and James Stock. 1997. “Instrumental Variables Regression with Weak Instruments.” *Econometrica*. doi:http://www.jstor.org/stable/2171753.

Stock, James H, and Motohiro Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *The National Bureau of Economic Research*. doi:citeulike-article-id:1459542.

Tachieva, Galina. 2012. “TCC Special Edition | Transect for Sprawl Repair.” *Transect Codes Council*. http://myemail.constantcontact.com/TCC-Special-Edition---Transect-for-Sprawl-

Repair.html?soid=1103584053200&aid=NAA8D92NnIY.

The Community Preventive Services Task Force. 2016. “Physical Activity: Built Environment Approaches Combining Transportation System Interventions with Land Use and Environmental Design.” *The Community Guide*. https://www.thecommunityguide.org/findings/physical-activity-built-environment-approaches.

Thrun, Emily, Julien Leider, and Jamie F. Chriqui. 2016. “Exploring the Cross-Sectional Association between Transit-Oriented Development Zoning and Active Travel and Transit Usage in the United States, 2010–2014.” *Frontiers in Public Health* 4 (June). doi:10.3389/fpubh.2016.00113.

Tiebout, Charles M. 1956. “A Pure Theory of Local Expenditures.” *Journal of Political Economy* 64 (5). The University Press of Chicago : 416–24. doi:10.1086/257839.

Transportation Research Board. 2005. “Does the Built Environment Influence Physical Activity? Examining the Evidence - Special Report 282.” *Built Environment*, January, 1–270. http://www.nap.edu/catalog/11203/does-the-built-environment-influence-physical-activity-examining-the-evidence.

Transportation Research Board, Institute of Medicine, and National Research Council. 2009. *Local Government Actions to Prevent Childhood Obesity*. Edited by Lynn Parker, Annina Catherine Burns, and Eduardo Sanchez. Washington, DC, DC: The National Academies Press. doi:10.17226/12674.

Trost, Stewart G., Neville Owen, Adrian E. Bauman, James F. Sallis, and Wendy Brown. 2002. “Correlates of Adults’ Participation in Physical Activity: Review and Update.” *Medicine and Science in Sports and Exercise* 34 (12): 1996–2001.

doi:10.1249/01.MSS.0000038974.76900.92.

Tudor-Locke, Catrine, and Sandra A Ham. 2008. “Walking Behaviors Reported in the American Time Use Survey 2003-2005.” *Journal of Physical Activity & Health* 5 (5): 633–47. http://www.ncbi.nlm.nih.gov/pubmed/18820341.

Tudor-Locke, Catrine, John M. Schuna, Peter T. Katzmarzyk, Wei Liu, Karen S. Hamrick, and William D. Johnson. 2014. “Body Mass Index: Accounting for Full Time Sedentary Occupation and 24-Hr Self-Reported Time Use.” *PLoS ONE*. doi:10.1371/journal.pone.0109051.

U.S. Census Bureau. 2008. “A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know.” Washington, DC. https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf.

U.S. Department of Health and Human Services. 2008. “2008 Physical Activity Guidelines for Americans.” https://health.gov/paguidelines/guidelines/.

“Understanding the Basics of Land Use and Planning: Glossary of Land Use and Planning Terms.” 2010. Sacramento, CA. http://elkgrovecity.org/UserFiles/Servers/Server_109585/File/Departments/Planning/Projects/General Plan/2010_-_landuseglossary.pdf.

Van Dyck, Delfien, Greet Cardon, Benedicte Deforche, James F. Sallis, Neville Owen, and Ilse De Bourdeaudhuij. 2010. “Neighborhood SES and Walkability Are Related to Physical Activity Behavior in Belgian Adults.” *Preventive Medicine* 50: S74–79. doi:10.1016/j.ypmed.2009.07.027.

Wooldridge, Jeffrey M. 2007. “Lecture 06: Control Function and Related Methods.” In *NBER*

Lecture What's New in Econometrics.

———. 2009. *Introductory Econometrics: A Modern Approach*. 4thed. South Western, Cengage Learning.

———. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50 (2). University of Wisconsin Press: 420–45. doi:10.3368/jhr.50.2.420.

Zhao, Zhenxiang, and Robert Kaestner. 2010. "Effects of Urban Sprawl on Obesity." *Journal of Health Economics* 29 (6): 779–87. doi:10.1016/j.jhealeco.2010.07.006.

Zick, Cathleen D., W. Keith Bryant, and Sivithee Srisukhumbowornchai. 2008. "Does Housework Matter Anymore? The Shifting Impact of Housework on Economic Inequality." *Review of Economics of the Household* 6 (1): 1–28. doi:10.1007/s11150-007-9020-3.

Zick, Cathleen D., Ken R. Smith, Barbara B. Brown, Jessie X. Fan, and Lori Kowaleski-Jones. 2007. "Physical Activity during the Transition from Adolescence to Adulthood." *Journal of Physical Activity & Health* 4 (2): 125–37. http://www.ncbi.nlm.nih.gov/pubmed/17570883.

APPENDICES

Appendix 1

The Mediational Analysis, Aerobics Regression, Full Table

	Step 1			Step 3		
VARIABLES	Probit	OLS	Log OLS	Probit	OLS	Log OLS
Active Living-Oriented Zoning	-0.082	-9.387	-0.284	-0.213	-27.266	-0.591
	(0.252)	(18.621)	(0.330)	(0.277)	(20.349)	(0.359)
Walkability				0.025	2.167	0.037
				(0.017)	(1.356)	(0.023)
Age	0.009***	-0.142	0.001	0.009***	-0.158	0.001
	(0.003)	(0.380)	(0.007)	(0.003)	(0.380)	(0.006)
Male	-0.418***	-13.239	-0.353	-0.419***	-15.401	-0.390
	(0.101)	(13.071)	(0.269)	(0.101)	(12.696)	(0.265)
Non-Hispanic Black	-0.334**	30.465	0.714**	-0.330**	33.591	0.768**
	(0.159)	(19.380)	(0.354)	(0.158)	(20.491)	(0.371)
Non-Hispanic Other	0.017	-1.232	0.077	0.015	-1.149	0.078
	(0.166)	(15.211)	(0.273)	(0.166)	(14.445)	(0.260)
Hispanic	-0.292**	1.713	-0.034	-0.287**	3.639	-0.001
	(0.140)	(24.546)	(0.432)	(0.140)	(24.239)	(0.423)
Widow, Divorced, Separated	-0.091	0.907	0.032	-0.094	-1.587	-0.011
	(0.106)	(7.469)	(0.140)	(0.105)	(7.738)	(0.142)
Single	-0.138	0.773	0.099	-0.149	-8.912	-0.067
	(0.123)	(20.153)	(0.365)	(0.120)	(22.910)	(0.401)
Any Difficulty	-0.353**	12.543	-0.011	-0.354**	13.207	0.000
	(0.168)	(20.870)	(0.371)	(0.168)	(21.243)	(0.383)
Less than High School	0.079	12.185	0.288	0.084	19.212	0.408
	(0.146)	(19.010)	(0.307)	(0.146)	(21.049)	(0.342)
High School Graduate	-0.321**	-46.695**	-0.802**	-0.317**	-46.109**	-0.792**
	(0.159)	(21.027)	(0.329)	(0.158)	(21.369)	(0.337)
Some College	0.158	-17.595	-0.365	0.159	-17.108	-0.356
	(0.126)	(13.300)	(0.222)	(0.126)	(12.910)	(0.215)
Graduate	0.309***	-10.320	-0.279	0.308***	-7.342	-0.228
	(0.106)	(11.980)	(0.221)	(0.107)	(11.821)	(0.215)
% Households in Poverty	-0.019	-0.517	-0.027	-0.027	-0.776	-0.031
	(0.024)	(3.176)	(0.057)	(0.026)	(3.320)	(0.059)
% Non-Hispanic White	-0.008	2.147*	0.028	-0.008	2.113*	0.028
	(0.007)	(1.143)	(0.019)	(0.007)	(1.148)	(0.019)
% Non-Hispanic Black	-0.019*	3.009*	0.049*	-0.019*	2.672*	0.044
	(0.010)	(1.600)	(0.028)	(0.010)	(1.567)	(0.027)
% Hispanic	-0.010	2.840**	0.044*	-0.011	2.651**	0.041*
	(0.007)	(1.322)	(0.024)	(0.007)	(1.307)	(0.024)
Median Household Income	-0.000	0.001	0.000	-0.000	0.001	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Median Age	-0.005	0.894	0.028	-0.005	0.865	0.027
	(0.013)	(1.886)	(0.032)	(0.013)	(1.901)	(0.033)
Midwest	-0.044	-9.632	-0.271	-0.093	-16.738	-0.393*
	(0.177)	(13.437)	(0.236)	(0.179)	(12.021)	(0.213)
Northeast	-0.026	-26.299*	-0.539**	-0.134	-40.936***	-0.790***

	(0.178)	(14.668)	(0.257)	(0.195)	(14.881)	(0.244)
South	0.083	-37.654**	-0.732**	0.071	-36.159**	-0.706**
	(0.155)	(16.293)	(0.288)	(0.160)	(15.680)	(0.272)
Land Area (Square Miles) 2010	0.000	-0.002	-0.000	0.000	-0.003	-0.000
	(0.000)	(0.002)	(0.000)	(0.000)	(0.003)	(0.000)
Sunday	-0.395***	1.332	0.064	-0.393***	4.387	0.117
	(0.131)	(11.955)	(0.214)	(0.131)	(12.116)	(0.214)
Saturday	-0.551***	10.508	0.339	-0.551***	9.085	0.315
	(0.138)	(16.186)	(0.320)	(0.138)	(16.188)	(0.319)
February	0.045	-15.538	-0.278	0.042	-18.007	-0.320
	(0.161)	(16.193)	(0.293)	(0.161)	(16.808)	(0.306)
March	0.081	-16.782	-0.276	0.078	-22.322	-0.371
	(0.186)	(20.809)	(0.416)	(0.187)	(21.281)	(0.419)
April	0.059	-0.953	-0.084	0.060	6.765	0.048
	(0.168)	(18.095)	(0.308)	(0.168)	(18.780)	(0.326)
May	-0.257	-28.089	-0.604	-0.253	-27.514	-0.594
	(0.206)	(24.479)	(0.501)	(0.206)	(25.376)	(0.521)
June	-0.371	-43.497	-0.765	-0.372	-41.243	-0.726
	(0.249)	(29.332)	(0.562)	(0.250)	(28.611)	(0.552)
July	-0.183	-40.402*	-0.637	-0.186	-45.082*	-0.718*
	(0.203)	(22.252)	(0.396)	(0.203)	(22.975)	(0.410)
August	-0.030	-32.854*	-0.696*	-0.033	-37.530**	-0.776**
	(0.172)	(18.662)	(0.377)	(0.173)	(17.261)	(0.348)
September	-0.218	22.037	0.288	-0.220	18.763	0.231
	(0.196)	(27.612)	(0.456)	(0.197)	(27.207)	(0.457)
October	-0.099	4.494	0.237	-0.102	2.457	0.202
	(0.196)	(17.054)	(0.302)	(0.196)	(16.701)	(0.298)
November	0.044	-16.756	-0.154	0.045	-16.948	-0.157
	(0.180)	(16.020)	(0.293)	(0.180)	(15.723)	(0.293)
December	-0.093	-10.957	-0.089	-0.090	-12.890	-0.122
	(0.193)	(24.595)	(0.429)	(0.193)	(23.935)	(0.417)
Year 2011	-0.168	-8.706	-0.223	-0.167	-3.022	-0.126
	(0.149)	(21.950)	(0.400)	(0.149)	(22.665)	(0.414)
Year 2012	0.042	15.798	0.174	0.045	21.273	0.268
	(0.139)	(16.656)	(0.270)	(0.139)	(17.156)	(0.288)
Year 2013	0.184	-9.965	-0.249	0.183	-3.804	-0.143
	(0.137)	(15.853)	(0.292)	(0.137)	(16.201)	(0.299)
Year 2014	-0.091	-14.104	-0.386	-0.090	-10.724	-0.328
	(0.155)	(18.934)	(0.334)	(0.155)	(18.331)	(0.326)
Year 2015	0.040	-13.564	-0.289	0.040	-7.991	-0.194
	(0.143)	(14.281)	(0.233)	(0.143)	(14.191)	(0.237)
Constant	-0.984	-183.916	0.457	-0.676	-151.461	1.013
	(1.424)	(174.294)	(2.865)	(1.454)	(178.318)	(2.959)
Observations	24,458	66	66	24,458	66	66
R-squared		0.705	0.755		0.722	0.769

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix 2

Walkability Regression, Full Table

VARIABLES	Walkability	Walkability
Active Living-Oriented Zoning	5.06***	.379***
	-0.131	(.0359)
1900 Manufacturing Establishment Density		.037***
		(.0000904)
1900 Farmland Proportion		-.314***
		(.0119)
% Households in Poverty	.275***	-.0417***
	(.00889)	(.00216)
% Non-Hispanic White	-.0247***	-.0625***
	(.00194)	(.00152)
% Non-Hispanic Black	-.0426***	-.0588***
	(.0016)	(.0017)
% Hispanic	-.0255***	-.0494***
	(.00193)	(.00148)
Median Household Income	.0000714***	-.0000127***
	(3.01e-06)	(6.10e-07)
Median Age	.00623**	.0031***
	(.00287)	(.000954)
Midwest	1.8***	.922***
	(.0426)	(.0204)
Northeast	4.26***	1.11***
	(.0804)	(.0311)
South	.541***	.562***
	(.0318)	(.0202)
Land Area (Square Miles) 2010	-.0000292***	-.0000169***
	(1.88e-06)	(9.93e-07)
Constant	-8.55***	7.02***
	(.554)	(.179)
Observations	24,458	24,458
Adjusted R-squared	.485	.948

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix 3

First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Physical Activity Regression, Full Table

	Step 1		Step 3		Step 3	
	IV Probit	GMM and GMM Log	IV Probit	IV Probit	GMM and GMM Log	GMM and GMM Log
VARIABLES	Zoning	Zoning	Zoning	Walkability	Zoning	Walkability
1900 Manufacturing Establishment Density	0.001***	0.001***	0.001***	0.038***	0.001***	0.037***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
1900 Farmland Proportion	-0.164***	-0.175***	-0.164***	-0.375***	-0.175***	-0.337
	(0.042)	(0.041)	(0.004)	(0.013)	(0.041)	(0.220)
Age	0.000	-0.000	0.000	-0.001***	-0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Male	0.001	-0.002	0.001	-0.014**	-0.002	-0.018
	(0.002)	(0.007)	(0.002)	(0.007)	(0.007)	(0.026)
Non-Hispanic Black	0.013**	-0.014	0.013***	0.031***	-0.014	0.024
	(0.005)	(0.014)	(0.003)	(0.011)	(0.014)	(0.053)
Non-Hispanic Other	-0.001	-0.011	-0.001	-0.030**	-0.011	-0.032
	(0.004)	(0.010)	(0.004)	(0.015)	(0.010)	(0.032)
Hispanic	0.009**	-0.003	0.009***	0.015	-0.003	0.030
	(0.004)	(0.009)	(0.003)	(0.011)	(0.009)	(0.025)
Widow, Divorced, Separated	0.006*	-0.009	0.006**	-0.003	-0.009	-0.020
	(0.004)	(0.010)	(0.003)	(0.009)	(0.010)	(0.021)
Single	0.004	-0.010	0.004	-0.047***	-0.010	-0.063*
	(0.004)	(0.010)	(0.003)	(0.010)	(0.010)	(0.035)
Any Difficulty	-0.009**	-0.020	-0.009**	-0.008	-0.020	-0.025
	(0.004)	(0.013)	(0.003)	(0.012)	(0.013)	(0.033)
Less than High School	-0.009**	-0.012	-0.009***	0.019	-0.012	-0.020
	(0.004)	(0.012)	(0.003)	(0.012)	(0.012)	(0.028)
High School Graduate	-0.007	0.006	-0.007**	-0.012	0.006	-0.011
	(0.004)	(0.011)	(0.003)	(0.010)	(0.011)	(0.037)
Some College	-0.009**	-0.009	-0.009***	0.015	-0.009	0.046
	(0.004)	(0.011)	(0.003)	(0.011)	(0.011)	(0.031)
Graduate	-0.005	-0.020**	-0.005	0.002	-0.020**	0.027
	(0.004)	(0.009)	(0.003)	(0.012)	(0.009)	(0.036)
% Households in Poverty	-0.017**	-0.015*	-0.017***	-0.048***	-0.015*	-0.043
	(0.007)	(0.008)	(0.001)	(0.002)	(0.008)	(0.038)
% Non-Hispanic White	-0.011***	-0.010***	-0.011***	-0.067***	-0.010***	-0.062**
	(0.004)	(0.003)	(0.000)	(0.001)	(0.003)	(0.025)
% Non-Hispanic Black	-0.007*	-0.006*	-0.007***	-0.062***	-0.006*	-0.055**
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.025)
% Hispanic	-0.011***	-0.010***	-0.011***	-0.054***	-0.010***	-0.049**
	(0.004)	(0.003)	(0.000)	(0.001)	(0.003)	(0.022)
Median Household Income	-0.000***	-0.000**	-0.000***	-0.000***	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	-0.008**	-0.007	-0.008***	0.000	-0.007	-0.003
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.014)
Midwest	-0.080	-0.068	-0.080***	0.891***	-0.068	0.825**
	(0.056)	(0.055)	(0.004)	(0.016)	(0.055)	(0.330)
Northeast	-0.336***	-0.339***	-0.336***	0.985***	-0.339***	1.094**
	(0.057)	(0.061)	(0.004)	(0.014)	(0.061)	(0.471)

South	0.067	0.086	0.067***	0.587***	0.086	0.533**
	(0.059)	(0.056)	(0.004)	(0.014)	(0.056)	(0.270)
Land Area (Square Miles) 2010	-0.000	-0.000	-0.000***	-0.000***	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sunday	-0.004	0.002	-0.004	-0.010	0.002	0.053
	(0.003)	(0.008)	(0.002)	(0.009)	(0.008)	(0.039)
Saturday	-0.001	0.000	-0.001	0.034***	0.000	0.048**
	(0.003)	(0.008)	(0.002)	(0.009)	(0.008)	(0.024)
February	0.004	-0.006	0.004	0.040**	-0.006	0.071*
	(0.005)	(0.018)	(0.005)	(0.017)	(0.018)	(0.041)
March	0.010**	0.030*	0.010**	0.036**	0.030*	0.054
	(0.004)	(0.016)	(0.005)	(0.016)	(0.016)	(0.034)
April	0.005	0.017	0.005	0.051***	0.017	0.089*
	(0.005)	(0.017)	(0.005)	(0.017)	(0.017)	(0.047)
May	0.006	0.004	0.006	0.014	0.004	0.014
	(0.004)	(0.016)	(0.005)	(0.017)	(0.016)	(0.039)
June	0.006	0.023	0.006	0.032*	0.023	0.093*
	(0.005)	(0.017)	(0.005)	(0.016)	(0.017)	(0.048)
July	0.008*	0.021	0.008*	-0.000	0.021	0.045
	(0.005)	(0.015)	(0.005)	(0.017)	(0.015)	(0.046)
August	0.005	0.015	0.005	0.020	0.015	0.012
	(0.004)	(0.017)	(0.005)	(0.016)	(0.017)	(0.037)
September	-0.001	-0.014	-0.001	0.023	-0.014	0.046
	(0.004)	(0.014)	(0.005)	(0.017)	(0.014)	(0.032)
October	0.005	0.005	0.005	0.036**	0.005	0.021
	(0.005)	(0.016)	(0.005)	(0.017)	(0.016)	(0.043)
November	0.004	0.004	0.004	0.025	0.004	-0.000
	(0.005)	(0.016)	(0.005)	(0.017)	(0.016)	(0.057)
December	0.003	0.006	0.003	-0.000	0.006	0.012
	(0.006)	(0.018)	(0.005)	(0.017)	(0.018)	(0.055)
Year 2011	0.000	-0.014	0.000	0.005	-0.014	-0.002
	(0.003)	(0.010)	(0.003)	(0.012)	(0.010)	(0.026)
Year 2012	-0.000	-0.029**	-0.000	0.001	-0.029**	-0.078*
	(0.003)	(0.013)	(0.003)	(0.012)	(0.013)	(0.040)
Year 2013	0.002	-0.000	0.002	0.000	-0.000	-0.006
	(0.004)	(0.012)	(0.003)	(0.012)	(0.012)	(0.023)
Year 2014	-0.000	-0.024**	-0.000	0.004	-0.024**	-0.019
	(0.003)	(0.011)	(0.003)	(0.012)	(0.011)	(0.034)
Year 2015	-0.001	-0.019*	-0.001	-0.011	-0.019*	-0.063*
	(0.004)	(0.011)	(0.003)	(0.012)	(0.011)	(0.036)
Constant	2.540***	2.338***	2.540***	8.010***	2.338***	7.626**
	(0.468)	(0.508)	(0.034)	(0.118)	(0.508)	(3.434)
Observations	24,448	1,969	24,448	24,448	1,969	1,969

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix 4

First-Stage Analysis for the Mediational Analysis with Instrumental Variables, Sedentary Behavior Regression, Full Table

	Step 1		Step 3		Step 3	
	IV Probit	GMM and GMM Log	IV Probit	IV Probit	GMM and GMM Log	GMM and GMM Log
VARIABLES	Zoning	Zoning	Zoning	Walkability	Zoning	Walkability
1900 Manufacturing Establishment Density	0.001***	0.001***	0.001***	0.038***	0.001***	0.037***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
1900 Farmland Proportion	-0.164***	-0.163***	-0.164***	-0.375***	-0.163***	-0.374*
	(0.042)	(0.043)	(0.004)	(0.013)	(0.043)	(0.191)
Age	0.000	0.000	0.000	-0.001***	0.000	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.001	0.002	0.001	-0.014*	0.002	-0.015
	(0.002)	(0.002)	(0.002)	(0.007)	(0.002)	(0.014)
Non-Hispanic Black	0.013**	0.010*	0.013***	0.031***	0.010*	0.033
	(0.005)	(0.005)	(0.003)	(0.011)	(0.005)	(0.026)
Non-Hispanic Other	-0.001	-0.001	-0.001	-0.029**	-0.001	-0.029*
	(0.004)	(0.005)	(0.004)	(0.015)	(0.005)	(0.016)
Hispanic	0.009**	0.009**	0.009***	0.014	0.009**	0.012
	(0.004)	(0.005)	(0.003)	(0.011)	(0.005)	(0.014)
Widow, Divorced, Separated	0.006*	0.007*	0.006**	-0.003	0.007*	-0.002
	(0.004)	(0.004)	(0.003)	(0.009)	(0.004)	(0.010)
Single	0.004	0.003	0.004	-0.047***	0.003	-0.039**
	(0.004)	(0.004)	(0.003)	(0.010)	(0.004)	(0.018)
Any Difficulty	-0.009**	-0.010**	-0.009**	-0.008	-0.010**	-0.013
	(0.004)	(0.005)	(0.003)	(0.012)	(0.005)	(0.009)
Less than High School	-0.009**	-0.011**	-0.009***	0.019	-0.011**	0.017
	(0.004)	(0.005)	(0.003)	(0.012)	(0.005)	(0.013)
High School Graduate	-0.007	-0.009**	-0.007**	-0.012	-0.009**	-0.010
	(0.004)	(0.004)	(0.003)	(0.010)	(0.004)	(0.012)
Some College	-0.009**	-0.009**	-0.009***	0.015	-0.009**	0.008
	(0.004)	(0.004)	(0.003)	(0.011)	(0.004)	(0.011)
Graduate	-0.005	-0.006	-0.005	0.002	-0.006	0.005
	(0.004)	(0.004)	(0.003)	(0.012)	(0.004)	(0.018)
% Households in Poverty	-0.017**	-0.018**	-0.017***	-0.048***	-0.018**	-0.048
	(0.007)	(0.008)	(0.001)	(0.002)	(0.008)	(0.036)
% Non-Hispanic White	-0.011***	-0.011***	-0.011***	-0.067***	-0.011***	-0.068**
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.028)
% Non-Hispanic Black	-0.007*	-0.007*	-0.007***	-0.062***	-0.007*	-0.064**
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.029)
% Hispanic	-0.011***	-0.011***	-0.011***	-0.054***	-0.011***	-0.055**
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.025)
Median Household Income	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	-0.008**	-0.008**	-0.008***	0.000	-0.008**	0.002
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.016)
Midwest	-0.080	-0.077	-0.080***	0.890***	-0.077	0.914**
	(0.056)	(0.057)	(0.004)	(0.016)	(0.057)	(0.364)
Northeast	-0.336***	-0.334***	-0.336***	0.985***	-0.334***	0.989**

	(0.057)	(0.057)	(0.004)	(0.014)	(0.057)	(0.457)
South	0.067	0.067	0.067***	0.587***	0.067	0.607*
	(0.059)	(0.060)	(0.004)	(0.014)	(0.060)	(0.317)
Land Area (Square Miles) 2010	-0.000	-0.000	-0.000***	-0.000***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sunday	-0.004	-0.004	-0.004	-0.010	-0.004	-0.006
	(0.003)	(0.003)	(0.002)	(0.009)	(0.003)	(0.009)
Saturday	-0.001	-0.003	-0.001	0.034***	-0.003	0.038*
	(0.003)	(0.003)	(0.002)	(0.009)	(0.003)	(0.020)
February	0.004	0.002	0.004	0.040**	0.002	0.031**
	(0.005)	(0.006)	(0.005)	(0.017)	(0.006)	(0.014)
March	0.010**	0.011**	0.010**	0.035**	0.011**	0.030*
	(0.004)	(0.005)	(0.005)	(0.016)	(0.005)	(0.017)
April	0.005	0.007	0.005	0.052***	0.007	0.033*
	(0.005)	(0.006)	(0.005)	(0.017)	(0.006)	(0.018)
May	0.006	0.004	0.006	0.014	0.004	-0.003
	(0.004)	(0.004)	(0.005)	(0.017)	(0.004)	(0.013)
June	0.006	0.007	0.006	0.032*	0.007	0.023
	(0.005)	(0.005)	(0.005)	(0.016)	(0.005)	(0.015)
July	0.008*	0.009*	0.008*	-0.000	0.009*	-0.006
	(0.005)	(0.005)	(0.005)	(0.017)	(0.005)	(0.015)
August	0.005	0.004	0.005	0.019	0.004	0.016
	(0.004)	(0.004)	(0.005)	(0.016)	(0.004)	(0.015)
September	-0.001	0.001	-0.001	0.022	0.001	0.014
	(0.004)	(0.005)	(0.005)	(0.017)	(0.005)	(0.010)
October	0.005	0.005	0.005	0.036**	0.005	0.035**
	(0.005)	(0.006)	(0.005)	(0.017)	(0.006)	(0.014)
November	0.004	0.000	0.004	0.024	0.000	0.020
	(0.005)	(0.006)	(0.005)	(0.017)	(0.006)	(0.021)
December	0.003	0.004	0.003	-0.001	0.004	-0.021
	(0.006)	(0.006)	(0.005)	(0.017)	(0.006)	(0.019)
Year 2011	0.000	0.001	0.000	0.005	0.001	0.002
	(0.003)	(0.003)	(0.003)	(0.012)	(0.003)	(0.010)
Year 2012	-0.000	0.003	-0.000	-0.000	0.003	0.001
	(0.003)	(0.003)	(0.003)	(0.012)	(0.003)	(0.007)
Year 2013	0.002	0.003	0.002	0.000	0.003	0.001
	(0.004)	(0.004)	(0.003)	(0.012)	(0.004)	(0.009)
Year 2014	-0.000	0.001	-0.000	0.004	0.001	-0.000
	(0.003)	(0.004)	(0.003)	(0.012)	(0.004)	(0.010)
Year 2015	-0.001	-0.000	-0.001	-0.011	-0.000	-0.014
	(0.004)	(0.005)	(0.003)	(0.012)	(0.005)	(0.015)
Constant	2.541***	2.549***	2.541***	8.007***	2.549***	8.058**
	(0.468)	(0.467)	(0.034)	(0.118)	(0.467)	(3.486)
Observations	24,458	20,141	24,458	24,458	20,141	20,141

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix 5

The Mediational Analysis with Instrumental Variables, Step 1: Physical Activity Regression, Full Table

	Probit	IV Probit		OLS	GMM	Log OLS	Log GMM
VARIABLES	Physical Activity	Physical Activity	Zoning	Physical Activity	Physical Activity	Physical Activity	Physical Activity
Active Living-Oriented Zoning	0.190***	0.301***		3.510	9.957*	0.040	0.297***
	(0.067)	(0.083)		(4.879)	(5.641)	(0.090)	(0.115)
Age	0.007***	0.007***	0.000	-0.027	-0.027	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.058)	(0.057)	(0.001)	(0.001)
Male	0.076***	0.076***	0.001	9.094***	9.117***	0.121***	0.124***
	(0.025)	(0.025)	(0.002)	(1.997)	(1.967)	(0.036)	(0.035)
Non-Hispanic Black	-0.185***	-0.187***	0.013**	1.298	1.362	0.022	0.019
	(0.042)	(0.042)	(0.005)	(3.046)	(2.990)	(0.054)	(0.053)
Non-Hispanic Other	0.139***	0.140***	-0.001	0.942	1.053	0.014	0.014
	(0.041)	(0.041)	(0.004)	(3.357)	(3.310)	(0.053)	(0.052)
Hispanic	0.042	0.040	0.009**	0.063	0.074	0.017	0.010
	(0.030)	(0.030)	(0.004)	(3.109)	(3.041)	(0.049)	(0.047)
Widow, Divorced, Separated	-0.020	-0.021	0.006*	5.108**	5.157**	0.066	0.067*
	(0.030)	(0.030)	(0.004)	(2.245)	(2.213)	(0.041)	(0.041)
Single	0.027	0.024	0.004	8.586***	8.500***	0.119***	0.117***
	(0.032)	(0.033)	(0.004)	(2.433)	(2.403)	(0.044)	(0.044)
Any Difficulty	-0.114**	-0.113**	-0.009**	-6.617*	-6.476*	-0.180***	-0.169***
	(0.044)	(0.044)	(0.004)	(3.821)	(3.757)	(0.062)	(0.062)
Less than High School	0.012	0.015	-0.009**	0.192	0.369	-0.016	-0.015
	(0.037)	(0.037)	(0.004)	(2.838)	(2.827)	(0.058)	(0.057)
High School Graduate	-0.146***	-0.144***	-0.007	0.882	0.932	-0.020	-0.024
	(0.031)	(0.031)	(0.004)	(3.160)	(3.108)	(0.055)	(0.054)
Some College	-0.146***	-0.144***	-0.009**	-2.904	-2.801	-0.047	-0.045
	(0.033)	(0.033)	(0.004)	(3.053)	(3.025)	(0.051)	(0.050)
Graduate	0.123***	0.123***	-0.005	-3.401	-3.362	-0.016	-0.018
	(0.030)	(0.030)	(0.004)	(2.145)	(2.109)	(0.038)	(0.037)
% Households in Poverty	-0.010	-0.010	-0.017**	-0.345	-0.368	-0.004	-0.004
	(0.006)	(0.006)	(0.007)	(0.452)	(0.436)	(0.009)	(0.008)
% Non-Hispanic White	-0.004	-0.002	-0.011***	-0.244	-0.175	-0.003	0.000
	(0.002)	(0.002)	(0.004)	(0.157)	(0.146)	(0.003)	(0.003)
% Non-Hispanic Black	-0.002	-0.001	-0.007*	-0.194	-0.141	-0.000	0.002
	(0.003)	(0.003)	(0.004)	(0.209)	(0.210)	(0.004)	(0.004)
% Hispanic	-0.002	-0.001	-0.011***	-0.239	-0.165	-0.002	0.001
	(0.003)	(0.003)	(0.004)	(0.189)	(0.177)	(0.003)	(0.003)
Median Household Income	-0.000	-0.000	-0.000***	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	-0.000	0.001	-0.008**	-0.218	-0.196	-0.003	-0.002
	(0.004)	(0.004)	(0.004)	(0.280)	(0.273)	(0.005)	(0.005)
Midwest	-0.152***	-0.140**	-0.080	1.948	2.514	-0.006	0.031
	(0.053)	(0.055)	(0.056)	(3.687)	(3.435)	(0.068)	(0.067)
Northeast	-0.170***	-0.142***	-0.336***	6.851**	8.396***	0.072	0.149**
	(0.042)	(0.044)	(0.057)	(3.416)	(2.723)	(0.065)	(0.061)
South	-0.132***	-0.141***	0.067	-2.837	-3.590	-0.057	-0.080
	(0.046)	(0.045)	(0.059)	(3.303)	(3.352)	(0.062)	(0.064)
Land Area (Square Miles) 2010	-0.000	-0.000	-0.000	0.001***	0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Sunday	-0.074***	-0.074***	-0.004	12.660***	12.647***	0.187***	0.181***
	(0.026)	(0.026)	(0.003)	(2.472)	(2.424)	(0.039)	(0.038)
Saturday	-0.052*	-0.053*	-0.001	13.026***	12.972***	0.208***	0.204***
	(0.030)	(0.030)	(0.003)	(2.083)	(2.069)	(0.038)	(0.037)
February	0.048	0.047	0.004	-6.604	-6.505	-0.079	-0.069
	(0.062)	(0.062)	(0.005)	(4.821)	(4.707)	(0.074)	(0.073)
March	0.126**	0.125**	0.010**	-1.337	-1.555	0.003	0.001
	(0.060)	(0.060)	(0.004)	(4.339)	(4.310)	(0.064)	(0.063)
April	0.243***	0.243***	0.005	1.133	1.044	-0.013	-0.012
	(0.057)	(0.057)	(0.005)	(4.808)	(4.680)	(0.075)	(0.074)
May	0.245***	0.245***	0.006	1.109	1.146	0.049	0.057
	(0.063)	(0.063)	(0.004)	(5.243)	(5.181)	(0.074)	(0.072)
June	0.181***	0.180***	0.006	-7.516	-7.663*	-0.101	-0.100
	(0.060)	(0.060)	(0.005)	(4.695)	(4.638)	(0.077)	(0.076)
July	0.185***	0.184***	0.008*	-4.741	-4.756	-0.044	-0.038
	(0.060)	(0.060)	(0.005)	(4.652)	(4.595)	(0.072)	(0.070)
August	0.288***	0.288***	0.005	2.124	2.121	0.058	0.056
	(0.060)	(0.060)	(0.004)	(5.164)	(5.082)	(0.075)	(0.074)
September	0.215***	0.215***	-0.001	-3.481	-3.424	-0.053	-0.047
	(0.053)	(0.053)	(0.004)	(5.279)	(5.192)	(0.083)	(0.081)
October	0.220***	0.220***	0.005	0.312	0.292	0.025	0.026
	(0.054)	(0.054)	(0.005)	(5.995)	(5.913)	(0.085)	(0.084)
November	0.064	0.064	0.004	-3.845	-3.837	-0.084	-0.086
	(0.064)	(0.064)	(0.005)	(4.726)	(4.640)	(0.081)	(0.079)
December	-0.086	-0.086	0.003	-3.004	-3.017	-0.050	-0.047
	(0.076)	(0.076)	(0.006)	(5.764)	(5.688)	(0.093)	(0.092)
Year 2011	0.024	0.024	0.000	2.121	2.290	0.009	0.012
	(0.040)	(0.040)	(0.003)	(3.016)	(2.990)	(0.050)	(0.050)
Year 2012	0.088*	0.088*	-0.000	2.413	2.618	0.018	0.023
	(0.046)	(0.046)	(0.003)	(2.944)	(2.899)	(0.054)	(0.054)
Year 2013	-0.024	-0.024	0.002	0.575	0.595	0.010	0.013
	(0.039)	(0.039)	(0.004)	(3.006)	(2.968)	(0.047)	(0.046)
Year 2014	0.020	0.020	-0.000	-1.458	-1.221	-0.024	-0.012
	(0.043)	(0.043)	(0.003)	(3.497)	(3.472)	(0.056)	(0.056)
Year 2015	0.106***	0.107***	-0.001	-0.584	-0.427	-0.007	-0.002
	(0.041)	(0.041)	(0.004)	(3.111)	(3.048)	(0.057)	(0.056)
1900 Manufacturing Establishment Density			0.001***				
			(0.000)				
1900 Farmland Proportion			-0.164***				
			(0.042)				
Constant	-1.405***	-1.645***	2.540***	86.402***	75.195***	4.044***	3.532***
	(0.434)	(0.429)	(0.467)	(25.381)	(22.238)	(0.493)	(0.467)
Observations	24,448	24,448	24,448	1,969	1,969	1,969	1,969

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix 6

The Mediational Analysis with Instrumental Variables, Step 3: Physical Activity Regression, Full Table

VARIABLES	Probit	IV Probit			OLS	GMM	Log OLS	Log GMM
	Sedentary Behavior	Sedentary Behavior	Zoning	Walkability	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior
Active Living-Oriented Zoning	0.148*	0.255			0.431	10.866	-0.063	0.064
	(0.077)	(0.278)			(6.187)	(20.012)	(0.101)	(0.335)
Walkability	0.008*	0.003			0.538	-0.041	0.018***	0.011
	(0.004)	(0.012)			(0.410)	(0.871)	(0.007)	(0.015)
Age	0.007***	0.007***	0.000	-0.001**	-0.027	-0.027	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.058)	(0.057)	(0.001)	(0.001)
Male	0.077***	0.076***	0.001	-0.014	9.111***	9.124***	0.122***	0.122***
	(0.025)	(0.025)	(0.002)	(0.014)	(1.991)	(1.973)	(0.036)	(0.035)
Non-Hispanic Black	-0.186***	-0.187***	0.013**	0.031	1.219	1.359	0.020	0.022
	(0.042)	(0.042)	(0.005)	(0.028)	(3.051)	(2.991)	(0.054)	(0.053)
Non-Hispanic Other	0.140***	0.140***	-0.001	-0.030*	0.953	1.052	0.015	0.016
	(0.041)	(0.041)	(0.004)	(0.016)	(3.354)	(3.310)	(0.053)	(0.052)
Hispanic	0.041	0.040	0.009**	0.015	0.003	0.056	0.015	0.015
	(0.030)	(0.030)	(0.004)	(0.013)	(3.106)	(3.064)	(0.049)	(0.048)
Widow, Divorced, Separated	-0.020	-0.021	0.006*	-0.003	5.025**	5.166**	0.063	0.065
	(0.030)	(0.030)	(0.004)	(0.009)	(2.253)	(2.223)	(0.041)	(0.040)
Single	0.025	0.024	0.004	-0.047**	8.395***	8.508***	0.113**	0.114***
	(0.033)	(0.033)	(0.004)	(0.020)	(2.425)	(2.409)	(0.044)	(0.044)
Any Difficulty	-0.114**	-0.113**	-0.009**	-0.008	-6.625*	-6.438*	-0.180***	-0.178***
	(0.044)	(0.044)	(0.004)	(0.010)	(3.825)	(3.840)	(0.062)	(0.063)
Less than High School	0.013	0.014	-0.009**	0.019	0.273	0.378	-0.013	-0.012
	(0.037)	(0.037)	(0.004)	(0.015)	(2.833)	(2.833)	(0.058)	(0.057)
High School Graduate	-0.145***	-0.144***	-0.007	-0.012	0.970	0.911	-0.017	-0.018
	(0.031)	(0.031)	(0.004)	(0.013)	(3.174)	(3.140)	(0.056)	(0.055)
Some College	-0.146***	-0.145***	-0.009**	0.015	-2.956	-2.785	-0.049	-0.047
	(0.033)	(0.033)	(0.004)	(0.012)	(3.051)	(3.046)	(0.051)	(0.050)
Graduate	0.122***	0.122***	-0.005	0.002	-3.615*	-3.356	-0.023	-0.020
	(0.030)	(0.031)	(0.004)	(0.020)	(2.134)	(2.115)	(0.037)	(0.037)
% Households in Poverty	-0.012*	-0.010	-0.017**	-0.048	-0.523	-0.352	-0.010	-0.008

	(0.006)	(0.007)	(0.007)	(0.036)	(0.502)	(0.546)	(0.010)	(0.010)
% Non-Hispanic White	-0.004	-0.003	-0.011***	-0.067**	-0.239	-0.168	-0.002	-0.002
	(0.002)	(0.003)	(0.004)	(0.027)	(0.157)	(0.205)	(0.003)	(0.004)
% Non-Hispanic Black	-0.002	-0.002	-0.007*	-0.062**	-0.175	-0.136	0.000	0.001
	(0.003)	(0.003)	(0.004)	(0.029)	(0.210)	(0.230)	(0.004)	(0.004)
% Hispanic	-0.002	-0.001	-0.011***	-0.054**	-0.233	-0.157	-0.002	-0.001
	(0.002)	(0.004)	(0.004)	(0.025)	(0.189)	(0.244)	(0.003)	(0.004)
Median Household Income	-0.000	-0.000	-0.000***	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	-0.000	0.001	-0.008**	0.000	-0.223	-0.191	-0.003	-0.003
	(0.004)	(0.005)	(0.004)	(0.016)	(0.282)	(0.291)	(0.005)	(0.005)
Midwest	-0.166***	-0.148**	-0.080	0.891**	0.969	2.703	-0.039	-0.018
	(0.053)	(0.072)	(0.056)	(0.360)	(3.922)	(5.265)	(0.071)	(0.093)
Northeast	-0.205***	-0.162	-0.336***	0.985**	4.263	8.809	-0.014	0.041
	(0.050)	(0.115)	(0.057)	(0.459)	(4.760)	(9.140)	(0.079)	(0.157)
South	-0.136***	-0.139***	0.067	0.587*	-3.143	-3.629	-0.067	-0.073
	(0.045)	(0.045)	(0.059)	(0.313)	(3.310)	(3.456)	(0.061)	(0.062)
Land Area (Square Miles) 2010	-0.000	-0.000	-0.000	-0.000	0.001***	0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sunday	-0.074***	-0.074***	-0.004	-0.010	12.680***	12.630***	0.188***	0.187***
	(0.026)	(0.026)	(0.003)	(0.009)	(2.475)	(2.449)	(0.039)	(0.039)
Saturday	-0.053*	-0.053*	-0.001	0.034*	12.958***	12.973***	0.206***	0.206***
	(0.030)	(0.030)	(0.003)	(0.018)	(2.087)	(2.070)	(0.037)	(0.037)
February	0.047	0.047	0.004	0.040**	-6.683	-6.454	-0.082	-0.079
	(0.062)	(0.062)	(0.005)	(0.017)	(4.826)	(4.831)	(0.074)	(0.074)
March	0.126**	0.125**	0.010**	0.036**	-1.329	-1.557	0.004	0.001
	(0.060)	(0.060)	(0.004)	(0.017)	(4.335)	(4.311)	(0.064)	(0.063)
April	0.243***	0.243***	0.005	0.051**	1.182	1.070	-0.011	-0.012
	(0.057)	(0.057)	(0.005)	(0.023)	(4.804)	(4.712)	(0.075)	(0.074)
May	0.246***	0.245***	0.006	0.014	1.191	1.150	0.051	0.051
	(0.063)	(0.063)	(0.004)	(0.013)	(5.234)	(5.183)	(0.074)	(0.073)
June	0.180***	0.180***	0.006	0.032**	-7.454	-7.660*	-0.099	-0.102
	(0.060)	(0.060)	(0.005)	(0.014)	(4.699)	(4.639)	(0.077)	(0.076)
July	0.185***	0.184***	0.008*	-0.000	-4.565	-4.769	-0.038	-0.041
	(0.060)	(0.060)	(0.005)	(0.012)	(4.660)	(4.602)	(0.073)	(0.071)
August	0.288***	0.288***	0.005	0.020	2.178	2.100	0.059	0.058

	(0.060)	(0.060)	(0.004)	(0.016)	(5.171)	(5.100)	(0.075)	(0.074)
September	0.215***	0.215***	-0.001	0.023*	-3.598	-3.389	-0.057	-0.055
	(0.053)	(0.053)	(0.004)	(0.012)	(5.289)	(5.246)	(0.083)	(0.082)
October	0.220***	0.220***	0.005	0.036***	0.293	0.306	0.025	0.025
	(0.054)	(0.054)	(0.005)	(0.014)	(5.997)	(5.920)	(0.085)	(0.084)
November	0.064	0.064	0.004	0.025	-3.806	-3.816	-0.083	-0.083
	(0.064)	(0.064)	(0.005)	(0.020)	(4.721)	(4.661)	(0.081)	(0.080)
December	-0.086	-0.086	0.003	-0.000	-2.957	-3.023	-0.049	-0.050
	(0.076)	(0.076)	(0.006)	(0.022)	(5.765)	(5.689)	(0.093)	(0.092)
Year 2011	0.024	0.024	0.000	0.005	2.193	2.283	0.011	0.012
	(0.040)	(0.040)	(0.003)	(0.009)	(3.023)	(2.995)	(0.050)	(0.050)
Year 2012	0.088*	0.088*	-0.000	0.001	2.441	2.620	0.019	0.021
	(0.046)	(0.046)	(0.003)	(0.006)	(2.944)	(2.900)	(0.054)	(0.054)
Year 2013	-0.023	-0.024	0.002	0.000	0.633	0.592	0.012	0.012
	(0.039)	(0.039)	(0.004)	(0.008)	(3.017)	(2.969)	(0.047)	(0.046)
Year 2014	0.020	0.020	-0.000	0.004	-1.425	-1.214	-0.023	-0.020
	(0.043)	(0.043)	(0.003)	(0.009)	(3.503)	(3.476)	(0.056)	(0.057)
Year 2015	0.107***	0.107***	-0.001	-0.011	-0.590	-0.428	-0.007	-0.005
	(0.041)	(0.041)	(0.004)	(0.012)	(3.116)	(3.049)	(0.057)	(0.056)
1900 Manufacturing Establishment Density			0.001***	0.038***				
			(0.000)	(0.001)				
1900 Farmland Proportion			-0.164***	-0.375*				
			(0.042)	(0.196)				
Constant	-1.325***	-1.551**	2.540***	8.010**	92.817***	73.214	4.259***	4.019***
	(0.444)	(0.724)	(0.467)	(3.480)	(27.476)	(47.401)	(0.508)	(0.800)
Observations	24,448	24,448	24,448	24,448	1,969	1,969	1,969	1,969

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1

Appendix 7

The Mediational Analysis with Instrumental Variables, Step 1: Sedentary Behavior Regression, Full Table

	Probit	IV Probit		OLS	GMM	Log OLS	Log GMM
VARIABLES	Sedentary Behavior	Sedentary Behavior	Zoning	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior
Active Living-Oriented Zoning	-0.187***	-0.529***		-7.719	-0.035	-12.537	-0.089**
	(0.067)	(0.089)		(6.580)	(0.033)	(9.843)	(0.041)
Age	0.013***	0.013***	0.000	2.433***	0.011***	2.420***	0.011***
	(0.001)	(0.001)	(0.000)	(0.084)	(0.000)	(0.083)	(0.000)
Male	0.122***	0.121***	0.001	43.830***	0.192***	43.728***	0.191***
	(0.017)	(0.017)	(0.002)	(2.145)	(0.011)	(2.137)	(0.011)
Non-Hispanic Black	0.044	0.049	0.013**	50.312***	0.155***	50.447***	0.158***
	(0.032)	(0.033)	(0.005)	(4.226)	(0.018)	(4.200)	(0.018)
Non-Hispanic Other	-0.008	-0.012	-0.001	-0.766	-0.068***	-1.081	-0.071***
	(0.044)	(0.044)	(0.004)	(4.046)	(0.022)	(4.025)	(0.022)
Hispanic	0.008	0.013	0.009**	-5.215	-0.022	-5.811	-0.024
	(0.031)	(0.031)	(0.004)	(3.777)	(0.023)	(3.680)	(0.022)
Widow, Divorced, Separated	-0.067***	-0.062***	0.006*	21.119***	0.096***	21.300***	0.097***
	(0.023)	(0.024)	(0.004)	(3.189)	(0.016)	(3.167)	(0.016)
Single	0.087***	0.095***	0.004	38.374***	0.194***	38.383***	0.195***
	(0.024)	(0.024)	(0.004)	(3.093)	(0.016)	(3.106)	(0.016)
Any Difficulty	0.185***	0.181***	-0.009**	77.415***	0.276***	77.402***	0.279***
	(0.040)	(0.040)	(0.004)	(4.666)	(0.018)	(4.635)	(0.018)
Less than High School	0.293***	0.284***	-0.009**	59.751***	0.254***	59.505***	0.250***
	(0.039)	(0.039)	(0.004)	(4.172)	(0.019)	(4.153)	(0.018)
High School Graduate	0.212***	0.205***	-0.007	48.939***	0.230***	48.871***	0.227***
	(0.027)	(0.027)	(0.004)	(3.092)	(0.015)	(3.095)	(0.015)
Some College	0.084***	0.078**	-0.009**	21.138***	0.108***	21.515***	0.110***
	(0.031)	(0.031)	(0.004)	(3.350)	(0.017)	(3.321)	(0.017)
Graduate	-0.179***	-0.176***	-0.005	-29.764***	-0.152***	-29.535***	-0.152***
	(0.035)	(0.034)	(0.004)	(3.278)	(0.020)	(3.260)	(0.020)
% Households in Poverty	0.001	-0.000	-0.017**	1.406**	0.006*	1.454**	0.006**
	(0.006)	(0.006)	(0.007)	(0.601)	(0.003)	(0.597)	(0.003)
% Non-Hispanic White	0.001	-0.004	-0.011***	0.407**	0.002*	0.340*	0.001
	(0.002)	(0.003)	(0.004)	(0.191)	(0.001)	(0.196)	(0.001)
% Non-Hispanic Black	0.003	-0.000	-0.007*	0.498*	0.003**	0.432	0.002*
	(0.002)	(0.003)	(0.004)	(0.283)	(0.002)	(0.273)	(0.001)
% Hispanic	0.003	-0.002	-0.011***	0.444**	0.002*	0.375*	0.002
	(0.002)	(0.003)	(0.004)	(0.225)	(0.001)	(0.221)	(0.001)
Median Household Income	-0.000	-0.000	-0.000***	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	0.012***	0.009**	-0.008**	0.588	0.005*	0.528	0.004
	(0.003)	(0.004)	(0.004)	(0.476)	(0.003)	(0.473)	(0.003)
Midwest	0.082*	0.046	-0.080	7.356	0.023	6.789	0.014
	(0.049)	(0.053)	(0.056)	(4.812)	(0.028)	(4.922)	(0.028)
Northeast	0.022	-0.062	-0.336***	6.788	0.019	5.420	0.002
	(0.050)	(0.048)	(0.057)	(4.321)	(0.023)	(4.750)	(0.023)
South	0.057	0.084*	0.067	7.474*	0.038*	8.235**	0.042*
	(0.040)	(0.049)	(0.059)	(3.987)	(0.023)	(3.967)	(0.023)

Land Area (Square Miles) 2010	0.000	0.000	-0.000	0.001**	0.000***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sunday	0.191***	0.190***	-0.004	50.041***	0.253***	49.735***	0.253***
	(0.023)	(0.023)	(0.003)	(2.916)	(0.015)	(2.895)	(0.015)
Saturday	-0.025	-0.025	-0.001	43.632***	0.227***	43.541***	0.226***
	(0.021)	(0.021)	(0.003)	(2.783)	(0.014)	(2.776)	(0.014)
February	0.008	0.011	0.004	1.947	-0.001	1.810	0.001
	(0.044)	(0.044)	(0.005)	(4.984)	(0.025)	(4.963)	(0.025)
March	-0.027	-0.023	0.010**	-11.252**	-0.064***	-11.061**	-0.061***
	(0.046)	(0.046)	(0.004)	(5.128)	(0.022)	(5.101)	(0.022)
April	-0.070	-0.068	0.005	-13.890***	-0.061**	-14.203***	-0.062***
	(0.044)	(0.044)	(0.005)	(4.961)	(0.024)	(4.937)	(0.024)
May	-0.197***	-0.195***	0.006	-19.092***	-0.078***	-18.366***	-0.071***
	(0.045)	(0.044)	(0.004)	(5.113)	(0.026)	(5.048)	(0.025)
June	-0.081*	-0.077*	0.006	-16.127***	-0.080***	-15.567***	-0.075***
	(0.047)	(0.047)	(0.005)	(5.275)	(0.024)	(5.232)	(0.024)
July	-0.117***	-0.113***	0.008*	-15.930***	-0.075**	-15.695***	-0.074**
	(0.041)	(0.041)	(0.005)	(5.716)	(0.029)	(5.690)	(0.029)
August	-0.138***	-0.137***	0.005	-17.858***	-0.084***	-17.467***	-0.082***
	(0.042)	(0.042)	(0.004)	(5.667)	(0.028)	(5.631)	(0.028)
September	-0.100**	-0.099**	-0.001	2.520	-0.010	3.499	-0.008
	(0.047)	(0.047)	(0.004)	(5.712)	(0.028)	(5.620)	(0.027)
October	-0.088*	-0.084*	0.005	-7.609	-0.051*	-8.216	-0.053*
	(0.047)	(0.047)	(0.005)	(5.639)	(0.027)	(5.595)	(0.027)
November	0.011	0.012	0.004	-5.883	-0.048*	-5.422	-0.047*
	(0.043)	(0.042)	(0.005)	(5.415)	(0.026)	(5.379)	(0.026)
December	-0.122**	-0.120**	0.003	-0.795	-0.013	-0.394	-0.010
	(0.048)	(0.048)	(0.006)	(5.065)	(0.023)	(5.036)	(0.023)
Year 2011	0.024	0.022	0.000	3.687	0.011	3.352	0.008
	(0.032)	(0.032)	(0.003)	(3.791)	(0.019)	(3.770)	(0.018)
Year 2012	0.028	0.026	-0.000	3.378	0.030	3.206	0.026
	(0.034)	(0.034)	(0.003)	(4.405)	(0.021)	(4.387)	(0.021)
Year 2013	-0.018	-0.018	0.002	0.757	0.015	0.762	0.011
	(0.029)	(0.029)	(0.004)	(3.862)	(0.018)	(3.849)	(0.018)
Year 2014	0.050*	0.049*	-0.000	0.092	0.005	0.219	0.005
	(0.029)	(0.029)	(0.003)	(3.986)	(0.020)	(3.969)	(0.020)
Year 2015	0.038	0.036	-0.001	-1.325	0.002	-1.477	-0.002
	(0.032)	(0.031)	(0.004)	(4.571)	(0.023)	(4.565)	(0.023)
1900 Manufacturing Establishment Density			0.001***				
			(0.000)				
1900 Farmland Proportion			-0.165***				
			(0.042)				
Constant	-0.263	0.490	2.541***	-66.508*	3.636***	-55.478	3.786***
	(0.379)	(0.461)	(0.467)	(33.909)	(0.191)	(36.293)	(0.182)
Observations	24,458	24,458	24,458	20,141	20,141	20,141	20,141

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1

Appendix 8

The Mediational Analysis with Instrumental Variables, Step 3: Sedentary Behavior Regression, Full Table

	Probit	IV Probit	IV Probit	IV Probit	OLS	GMM	Log OLS	Log GMM
VARIABLES	Sedentary Behavior	Sedentary Behavior	Zoning	Walkability	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior	Sedentary Behavior
Active Living-Oriented Zoning	-0.112*	-0.868**			-5.360	-0.013	-37.305	-0.266*
	(0.068)	(0.367)			(7.864)	(0.039)	(27.008)	(0.152)
Walkability	-0.014***	0.018			-0.495	-0.004	1.257	0.009
	(0.004)	(0.017)			(0.685)	(0.003)	(1.207)	(0.007)
Age	0.013***	0.013***	0.000	-0.001*	2.434***	0.011***	2.435***	0.011***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.084)	(0.000)	(0.084)	(0.000)
Male	0.121***	0.122***	0.001	-0.014	43.814***	0.192***	43.892***	0.193***
	(0.017)	(0.017)	(0.002)	(0.014)	(2.149)	(0.011)	(2.147)	(0.011)
Non-Hispanic Black	0.044	0.052	0.013**	0.031	50.352***	0.156***	50.585***	0.158***
	(0.032)	(0.033)	(0.005)	(0.029)	(4.227)	(0.018)	(4.227)	(0.018)
Non-Hispanic Other	-0.010	-0.012	-0.001	-0.029*	-0.818	-0.069***	-0.854	-0.069***
	(0.044)	(0.044)	(0.004)	(0.016)	(4.047)	(0.022)	(4.026)	(0.022)
Hispanic	0.008	0.016	0.009**	0.014	-5.181	-0.022	-4.895	-0.019
	(0.031)	(0.032)	(0.004)	(0.013)	(3.769)	(0.023)	(3.812)	(0.023)
Widow, Divorced, Separated	-0.066***	-0.060**	0.006*	-0.003	21.149***	0.097***	21.335***	0.098***
	(0.023)	(0.024)	(0.004)	(0.009)	(3.184)	(0.016)	(3.154)	(0.016)
Single	0.091***	0.096***	0.004	-0.047**	38.494***	0.195***	38.536***	0.196***
	(0.024)	(0.024)	(0.004)	(0.020)	(3.111)	(0.016)	(3.113)	(0.016)
Any Difficulty	0.185***	0.177***	-0.009**	-0.008	77.418***	0.276***	77.085***	0.274***
	(0.040)	(0.041)	(0.004)	(0.010)	(4.667)	(0.018)	(4.640)	(0.018)
Less than High School	0.291***	0.279***	-0.009**	0.019	59.727***	0.254***	59.357***	0.251***
	(0.039)	(0.039)	(0.004)	(0.015)	(4.170)	(0.019)	(4.153)	(0.018)
High School Graduate	0.211***	0.202***	-0.007	-0.012	48.920***	0.229***	48.619***	0.227***
	(0.027)	(0.027)	(0.004)	(0.013)	(3.090)	(0.015)	(3.128)	(0.015)
Some College	0.083***	0.074**	-0.009**	0.015	21.136***	0.108***	20.826***	0.106***
	(0.031)	(0.031)	(0.004)	(0.012)	(3.349)	(0.017)	(3.403)	(0.017)
Graduate	-0.176***	-0.178***	-0.005	0.002	-29.685***	-0.151***	-29.951***	-0.153***
	(0.034)	(0.034)	(0.004)	(0.020)	(3.295)	(0.020)	(3.299)	(0.020)

% Households in Poverty	0.005	-0.007	-0.017**	-0.048	1.532**	0.007**	0.958	0.003
	(0.006)	(0.010)	(0.007)	(0.036)	(0.615)	(0.003)	(0.789)	(0.004)
% Non-Hispanic White	0.000	-0.006	-0.011***	-0.067**	0.393**	0.002*	0.148	-0.000
	(0.002)	(0.004)	(0.004)	(0.027)	(0.188)	(0.001)	(0.289)	(0.002)
% Non-Hispanic Black	0.002	-0.001	-0.007*	-0.062**	0.475*	0.003*	0.351	0.002
	(0.002)	(0.004)	(0.004)	(0.029)	(0.280)	(0.002)	(0.299)	(0.002)
% Hispanic	0.003	-0.005	-0.011***	-0.054**	0.429*	0.002*	0.172	0.000
	(0.002)	(0.005)	(0.004)	(0.025)	(0.221)	(0.001)	(0.318)	(0.002)
Median Household Income	-0.000	-0.000	-0.000***	-0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median Age	0.012***	0.007	-0.008**	0.000	0.590	0.005*	0.416	0.004
	(0.003)	(0.006)	(0.004)	(0.016)	(0.474)	(0.003)	(0.505)	(0.003)
Midwest	0.108**	-0.016	-0.080	0.890**	8.209*	0.031	2.724	-0.012
	(0.048)	(0.094)	(0.056)	(0.360)	(4.928)	(0.028)	(6.605)	(0.036)
Northeast	0.085	-0.205	-0.336***	0.985**	8.801*	0.037	-4.379	-0.065
	(0.053)	(0.159)	(0.057)	(0.458)	(5.240)	(0.028)	(10.916)	(0.059)
South	0.064	0.093	0.067	0.587*	7.753*	0.041*	8.626*	0.048*
	(0.040)	(0.062)	(0.059)	(0.313)	(3.944)	(0.022)	(4.465)	(0.027)
Land Area (Square Miles) 2010	0.000	0.000	-0.000	-0.000	0.001**	0.000***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sunday	0.191***	0.189***	-0.004	-0.010	50.046***	0.253***	49.967***	0.252***
	(0.023)	(0.023)	(0.003)	(0.009)	(2.914)	(0.015)	(2.907)	(0.015)
Saturday	-0.024	-0.025	-0.001	0.034*	43.659***	0.227***	43.533***	0.226***
	(0.021)	(0.021)	(0.003)	(0.018)	(2.776)	(0.014)	(2.784)	(0.014)
February	0.010	0.011	0.004	0.040**	1.996	-0.001	2.006	-0.001
	(0.044)	(0.044)	(0.005)	(0.017)	(4.987)	(0.025)	(4.959)	(0.025)
March	-0.026	-0.020	0.010**	0.035**	-11.236**	-0.064***	-10.980**	-0.062***
	(0.046)	(0.046)	(0.004)	(0.017)	(5.127)	(0.022)	(5.106)	(0.022)
April	-0.069	-0.067	0.005	0.052**	-13.879***	-0.061**	-13.738***	-0.060**
	(0.044)	(0.044)	(0.005)	(0.024)	(4.963)	(0.024)	(4.960)	(0.024)
May	-0.198***	-0.192***	0.006	0.014	-19.109***	-0.078***	-18.983***	-0.077***
	(0.045)	(0.044)	(0.004)	(0.013)	(5.114)	(0.026)	(5.074)	(0.026)
June	-0.080*	-0.076	0.006	0.032**	-16.069***	-0.080***	-15.976***	-0.079***
	(0.047)	(0.047)	(0.005)	(0.014)	(5.278)	(0.024)	(5.240)	(0.024)
July	-0.117***	-0.110***	0.008*	-0.000	-15.919***	-0.075**	-15.659***	-0.072**
	(0.041)	(0.041)	(0.005)	(0.012)	(5.715)	(0.029)	(5.682)	(0.029)
August	-0.138***	-0.135***	0.005	0.019	-17.853***	-0.084***	-17.806***	-0.084***

	(0.042)	(0.042)	(0.004)	(0.016)	(5.668)	(0.028)	(5.626)	(0.028)
September	-0.100**	-0.099**	-0.001	0.022*	2.537	-0.010	2.536	-0.010
	(0.047)	(0.047)	(0.004)	(0.012)	(5.713)	(0.028)	(5.699)	(0.028)
October	-0.087*	-0.083*	0.005	0.036***	-7.585	-0.051*	-7.470	-0.050*
	(0.047)	(0.046)	(0.005)	(0.014)	(5.645)	(0.027)	(5.649)	(0.028)
November	0.012	0.012	0.004	0.024	-5.843	-0.048*	-5.948	-0.049*
	(0.043)	(0.042)	(0.005)	(0.019)	(5.418)	(0.026)	(5.399)	(0.026)
December	-0.121**	-0.118**	0.003	-0.001	-0.799	-0.013	-0.629	-0.012
	(0.048)	(0.048)	(0.006)	(0.022)	(5.067)	(0.023)	(5.033)	(0.023)
Year 2011	0.023	0.022	0.000	0.005	3.648	0.011	3.685	0.011
	(0.032)	(0.032)	(0.003)	(0.009)	(3.797)	(0.019)	(3.790)	(0.019)
Year 2012	0.027	0.026	-0.000	-0.000	3.345	0.029	3.423	0.030
	(0.034)	(0.034)	(0.003)	(0.006)	(4.411)	(0.021)	(4.385)	(0.021)
Year 2013	-0.019	-0.017	0.002	0.000	0.730	0.014	0.832	0.015
	(0.029)	(0.029)	(0.004)	(0.008)	(3.862)	(0.018)	(3.850)	(0.018)
Year 2014	0.050*	0.049*	-0.000	0.004	0.092	0.005	0.117	0.006
	(0.029)	(0.029)	(0.003)	(0.009)	(3.986)	(0.020)	(3.965)	(0.020)
Year 2015	0.037	0.036	-0.001	-0.011	-1.365	0.001	-1.376	0.001
	(0.032)	(0.031)	(0.004)	(0.012)	(4.583)	(0.023)	(4.560)	(0.023)
1900 Manufacturing Establishment Density			0.001***	0.038***				
			(0.000)	(0.001)				
1900 Farmland Proportion			-0.164***	-0.375*				
			(0.042)	(0.196)				
Constant	-0.404	1.209	2.541***	8.007**	-70.201**	3.602***	-4.111	4.126***
	(0.350)	(0.951)	(0.467)	(3.480)	(35.010)	(0.196)	(64.758)	(0.349)
Observations	24,458	24,458	24,458	24,458	20,141	20,141	20,141	20,141

Note: Standard Errors are clustered at county-level, *** p<0.01, ** p<0.05, * p<0.1

VITA

Xuan Yin

https://www.linkedin.com/in/xuanyin/

EDUCATION

Ph.D. in Economics, University of Illinois at Chicago, Chicago, IL, 2017

M.A. in Economics, University of Illinois at Chicago, Chicago, IL, 2014

M.S. in Mathematics, with Probability and Statistics concentration, University of Illinois at Chicago, Chicago, IL, 2013

Master in Economics, Shanghai University of Finance and Economics, Shanghai, China, 2010

Bachelor in Economics, Shanghai University of Finance and Economics, Shanghai, China, 2006

EXPERIENCE

Data Mining Research Intern: Econometric Forecasting, Data Mining Service Center, Bosch	Palo Alto, CA May 2016 – Jul 2016
Graduate Research Assistant, Institute for Health Research and Policy (IHRP), University of Illinois at Chicago	Chicago, IL May 2014 – Apr 2016
Graduate Research Assistant, Urban Transportation Center (UTC), CUPPA, University of Illinois at Chicago	Chicago, IL Sep 2013 – April 2014
Visiting Graduate Student, Department of Economics, University of Chicago	Chicago, IL 2012 - 2013
Graduate Teaching Assistant, Department of Economics, University of Illinois at Chicago	Chicago, IL Aug 2011 – May 2013
Graduate Research Assistant, Institute for Health Research and Policy (IHRP), University of Illinois at Chicago	Chicago, IL Jan 2012 – May 2012
Graduate Research Assistant, Institute of Government and Public Affairs (IGPA), University of Illinois	Chicago, IL Aug 2010 – May 2011

Graduate Teaching Assistant, Department of Economics, University of Illinois at Chicago — Chicago, IL, Aug 2009 – May 2010

PUBLICATION

Liu, Feng & Yin, Xuan. 2012. "Does Weather Affect Health Behaviors? Evidence from Temperature and Cigarette Smoking." *Smoking: Health Effects, Psychological Aspects and Cessation* (pp.75-82). New York: Nova Biomedical Books.

Merriman, David & Yin, Xuan. 2011. "Illinois' Economy: Past Trends, Current Challenges, Future Prospects." *The Illinois Report 2011* (pp.9-17). Urbana, Illinois: Institute of Government and Public Affairs