

2018

Economic Diversity and Regional Development: Geographical Scale, Structural Measurement, and Modeling Method

Jing Chen

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Chen, Jing, "Economic Diversity and Regional Development: Geographical Scale, Structural Measurement, and Modeling Method" (2018). *Graduate Theses, Dissertations, and Problem Reports*. 7166.
<https://researchrepository.wvu.edu/etd/7166>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Economic Diversity and Regional Development: Geographical Scale, Structural Measurement, and Modeling Method

Jing Chen

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Geography

Randall W. Jackson, Ph.D., Committee Chair

Jamison F. Conley, Ph.D.

Brian J. Cushing, Ph.D.

Trevor M. Harris, Ph.D.

Peter V. Schaeffer, Ph.D.

Department of Geology and Geography

Morgantown, West Virginia

2018

Keywords: economic structure, geographical scale, diversity, economic
clusters, spatial econometrics, regional economic development

Copyright © 2018 Jing Chen

ABSTRACT

Economic Diversity and Regional Development: Geographical Scale, Structural Measurement, and Modeling Method

Jing Chen

The relationship between economic diversity and regional economic performance has been a recurrent topic among regional scientists, economic geographers and development practitioners. However, the empirical understanding of this relationship is often inconsistent with the theoretical assumptions of economic diversity. Thus, to offer suggestions to evaluate the empirical relationship between economic diversity and regional economic performance so that policymakers and economic planners can better understand, formulate and undertake economic development policies, this dissertation examines the roles of (1) geographical scales, (2) structural measurements, and (3) modeling methods in assessing the diversity-performance relationship in the context of U.S. regional economies.

Geographical scales of regional economic systems can alter the empirical understanding of economic diversity on regional economic performance. Although scale-related issues can be study dependent, the diversity-stability relationship is studied among counties, states, Economic Areas, and Metropolitan Statistical Areas in the lower U.S. Based on the result of this multi-level analysis, several general concerns in quantifying regional economic structure are discussed. It is suggested to use functional rather than formal regions as the analytical units to reflect spatial interactions among regions. These units should also be large enough to form meaningful economic systems. In addition, possible temporal variations in the boundaries of functional regions should also be considered.

Structural measurements of economic diversity have long ignored the coexistence of economic specialization and diversity. As such, a novel measure that allows for the interpretation of economic diversity as the presence of multiple specializations is developed. In essence, this measure considers the diversity of clusters for a regional economy. The empirical results indicate that industry and cluster diversity demonstrate different effects on regional economic performance: both industrial and cluster diversity contribute to long-term economic stability, while only cluster diversity stimulates short-term employment growth. Regions thus can pursue high and stable growth by developing diversified specializations.

Modeling methods in existing economic structure research have not considered model uncertainty resulting from the set of control variables and the choice of an appropriate spatial weight matrix when studying the structure-performance relationship. Empirical evidence suggests that the model uncertainty can impact the understanding of economic diversity. A Bayesian Model Average (BMA) method is thus employed to address the model uncertainty. The result of BMA is used to estimate the effects of economic diversity on employment growth and economic stability, and comparisons are made between model estimates with and without model uncertainty.

To my family: past, present and future.

Acknowledgments

I would like to take this opportunity to thank these people who have influenced me greatly throughout my doctoral study.

First of all, I would like to thank my advisor, Dr. Randall Jackson, for his continual encouragement and guidance over the past four years. I am so fortunate that my advisor gives me 100% freedom to discover my research interest and formulate my own dissertation topic. This process is not easy, and I even went astray; but my advisor chose to support me and guide me back to the right road to becoming an independent researcher. Meanwhile, Dr. Jackson is also a geographer as well as a regional scientist, and this dual identity has brought valuable insights to such a multidisciplinary dissertation research.

I would also like to thank my other dissertation committee members, Drs. Trevor Harris (Geography), Jamison Conley (Geography), Peter Schaeffer (Resource Economics), and Brian Cushing (Economics), each of whom has played an important role not only in the writing process of this dissertation but also in shaping my academic career. I truly admire and appreciate their expertise, patience, and particularly, their intellectual inspirations and constructive comments that greatly improved the quality of this dissertation.

I also need to acknowledge my colleagues at the Regional Research Institute and the Department of Geology and Geography. Working with these colleagues is an invaluable and rewarding experience for me. I would like to express a special thank you to Caigan McKenzie for proofreading my dissertation as well as providing other suggestions on my academic writing in the past four years.

Finally, I am indebted to my family members in China. I thank my parents for raising me and teaching me to work hard. Your endless love is essential to my success in completing my Ph.D. degree.

Contents

Abstract	ii
Acknowledgments	iv
List of Figures	viii
List of Tables	ix
1 Introduction	1
1.1 Research Context	1
1.1.1 Growing Interest in Spatial Economics	2
1.1.2 Theoretical Perspectives on Economic Structure	3
1.1.3 Empirical Inconsistencies within Economic Structure Research	4
1.2 Research Questions	5
1.3 Organization	8
2 A Geographical Scale Perspective	9
2.1 Introduction	9
2.2 Background	11
2.2.1 The Definitions of Scale and Scale-Related Problems	11
2.2.2 Analytical Units in Economic Structure Research	13
2.3 Methodology	17
2.3.1 Geographical Units	17
2.3.2 Measuring Diversity	19

2.3.3	Empirical Model	21
2.3.4	Statistical Methods	22
2.4	Empirical Results	25
2.5	Discussion	34
2.6	Chapter Summary	37
3	A Structural Measurement Motivation	38
3.1	Introduction	38
3.2	Background	40
3.2.1	Specialization and Diversity: Two Sides of the Same Coin?	40
3.2.2	Identifying Economic Clusters	41
3.3	Methodology	42
3.3.1	Empirical Frameworks	42
3.3.2	Economic Performance Indicators	43
3.3.3	Measuring Economic Diversity	44
3.3.4	Modeling Methods	46
3.4	Empirical Results	49
3.5	Discussion	53
3.6	Chapter Summary	54
4	A Modeling Method Concern	57
4.1	Introduction	57
4.2	Background	59
4.2.1	Bivariate Statistics	60
4.2.2	The Inclusion of Control Variables	61
4.2.3	The Role of Spatial Spillovers	62
4.3	Methodology	64
4.3.1	Measuring Economic Diversity	65
4.3.2	Potential Control Variables	66
4.3.3	Model Uncertainty	67
4.3.4	Spatial Relationships	71

4.4	Empirical Results	72
4.5	Discussion	76
4.6	Chapter Summary	78
5	Conclusions	80
5.1	Summary and Reflections	80
5.2	Future Research Directions	82

List of Figures

1.1	Independent strategy to solving the research questions	7
1.2	Nested strategy to solving the research questions	7
2.1	Economic Areas in the lower U.S.	18
2.2	Metropolitan Statistical Areas in the lower U.S.	18
2.3	LISA maps of HHI in the contiguous U.S. (six nearest neighbors)	27
2.4	LISA maps of REI in the contiguous U.S. (six nearest neighbors)	28
2.5	Connectivity map of MSAs (queen contiguity)	33
2.6	Connectivity map of MSAs (six nearest neighbors)	33
3.1	Industry diversity in the lower U.S.	47
3.2	Cluster diversity in the lower U.S.	47
3.3	Correlation between industry and cluster diversity.	47
4.1	Three broad groups of modeling methods in economic diversity research	59

List of Tables

1.1	Selective overview of economic diversity research	6
2.1	Different data used in U.S.-based economic diversity research	20
2.2	Description of all the variables	22
2.3	Basic statistics of the HHI and REI variables	26
2.4	Global Moran's I results (six nearest neighbors)	26
2.5	OLS regression results	29
2.6	Lagrange multiplier and likelihood ratio tests (six nearest neighbors)	30
2.7	Estimation of direct, indirect and total effects (six nearest neighbors)	31
2.8	Global Moran's I results across different spatial weight matrices	32
2.9	Estimation of effects across different spatial weight matrices	34
3.1	Description of all the variables	43
3.2	Effect estimates of instability models	50
3.3	Effect estimates of growth models	52
3.4	A typology of specialization and diversification in terms of sectors and clusters	55
4.1	Summary of the economic diversity research that considers spatial effects . .	62
4.2	Posterior inclusion probability of control variables	73
4.3	Posterior inclusion probability of spatial weight matrices	74
4.4	Effect estimates of models that consider industry diversity	75
4.5	Effect estimates of models that consider both industry and cluster diversity	76

Chapter 1

Introduction

“Although basic science is directed at the discovery of general principles, the ultimate value of such knowledge, apart from simple curiosity, lies in our ability to apply it to local conditions and, thus, determine specific outcomes. Although such science may itself be placeless, the application of scientific knowledge in policy inevitably requires explicit attention to spatial variation, particularly when the basis of policy is local.”

Goodchild et al. (2000, p. 142)

1.1 Research Context

“Do not put all your eggs in one basket” is an old proverb that reminds us not to dedicate all of our resources to one project. The origin of this proverb is unclear, but many people attribute it to the novel *Don Quixote*, written by Miguel de Cervantes in the 1600s. Nowadays, such a proverb continues to be useful in our daily lives as well as be beneficial for formulating economic development strategies—for instance, it is risky to make the economy of one region (basket) highly dependent on a limited number of industries (eggs) because it might suffer severely from potential economic downturns.

Compared to the eggs-in-one-basket analogy, this dissertation focuses on the relationship between economic diversity and regional economic performance and evaluates the roles of geographical scales, structural measurements and modeling methods in assessing the diversity-performance relationship. The research context of this dissertation can be viewed in such dimensions as (1) growing interest in spatial economics, (2) theoretical perspectives

of economic structure, and (3) empirical inconsistencies within economic structure research as follows.

1.1.1 Growing Interest in Spatial Economics

In emphasizing the importance of spatial thinking in social science, Goodchild et al. (2000, p. 142) stated that “although basic science is directed at the discovery of general principles, the ultimate value of such knowledge, apart from simple curiosity, lies in our ability to apply it to local conditions and, thus, determine specific outcomes. Although such science may itself be placeless, the application of scientific knowledge in policy inevitably requires explicit attention to *spatial variation*, particularly when the basis of policy is local.”

More specifically, there is growing interest in considering the spatial variation of economic activities both theoretically and empirically. Historically, this interest can find its imprint in the work of location theory (e.g., Hotelling, 1929; Weber, 1929; Christaller, 1933), Isard’s (1956) location and space economy, Alonso’s (1964) bid rent theory, and others. More recently, with the development of the New Economic Geography (e.g., Krugman, 1991b, 1998; Venables, 1996; Fujita et al., 1999), the role of space has been formally introduced in the mainstream economics. Paul Krugman, the 2008 Nobel laureate in economics, further commented that “in recent years there has been a surge of interest in the geographical aspects of development, that is, in the question of where economic activities take place. There is nothing surprising about this interest—or perhaps the surprise is that it took so long for this interest to become a mainstream concern within economics” (Krugman, 1999, p. 142).

In addition to the theoretical development, Goodchild et al. (2000) argued that advances in spatial statistics and spatial econometrics—such as Anselin (1988) and LeSage and Pace (2009)—allow for the interpretation of spatial externalities, spillovers and agglomeration economies resulting from imperfect competition, path dependence and increasing returns. Interestingly, this interpretation also emphasizes two key concepts in spatial data analysis: spatial dependence and spatial heterogeneity (Anselin, 1988). According to Goodchild (2004), the first concept is in line with Tobler’s (1970, p. 236) First Law of Geography: “All things are related, but nearby things are more related than distant things.” By compar-

ison, the second concept stresses variations among the Earth’s surface. For example, the mean or variance of economic diversity is not stationary and varies from space to space. Goodchild (2004, p. 711) further differentiated these two concepts by arguing that “heterogeneity addresses the properties of places taken at a time (a first-order effect in the statistical sense), whereas spatial dependence compares the properties of pairs of places (a second-order effect).”

1.1.2 Theoretical Perspectives on Economic Structure

The theoretical understanding of economic structure and regional economic performance is often viewed from economic diversity and clusters¹. There is a general consensus among regional scientists that a diverse economy is associated with stable economic performance because the economy is not dependent on only a few industries and might suffer less from external economic downturns as suggested by Chinitz (1961) and Conroy (1975), for example. According to Malizia and Ke (1993, p. 222), the concept of diversity refers to “the variety of economic activity which reflects differences in economic structure” at a specific time, while economic stability is defined as “the absence of variation in economic activity over time.” Chinitz (1961, p. 281) further found that “diversified areas exhibit more stability in their growth because their fortunes are not tied to the fortunes of a few industries.” Similarly, Conroy (1975) borrowed the concept of portfolio from the finance literature to explain the diversity-stability relationship. He reasons that, for a given region, every industry can be seen as an independent investment, and a collection of all regional industries can be seen to be an industry portfolio. Accordingly, portfolio risk is greater for a regional economy that has invested in only a few industries.

Moreover, the relationship between regional economic structure and economic growth can be explained by two competing theoretical perspectives². On the one hand, conventional wisdom and much of the previous literature³, such as the Marshall-Arrow-Romer

¹The terms of “clusters” and “specializations” are used interchangeably in this dissertation even though specialized establishments may not be spatially clustered.

²Because agglomeration economies are necessary but not sufficient for clusters (Jackson, 2015), these two perspectives are often discussed with agglomeration economies (e.g., localization and urbanization economies). For a comprehensive review of agglomeration concepts, see Parr (2002).

³For a more comprehensive review of these theories, see Rocha (2004).

(MAR) model and Porter (1990, 1998), hold that intra-industry specialization is preferred for economic growth. As originally proposed by Marshall (1890), Arrow (1962), and Romer (1986), and later formalized by Glaeser et al. (1992), the MAR model hypothesizes that knowledge spillovers tend to be industry specific and that concentrations of similar industries facilitate growth. These knowledge spillovers are known as MAR externalities. As with MAR, Porter (1990, 1998) has suggested that specialization can promote economic growth. On the other hand, Jacobs (1969) suggested that diversity is more conducive to economic growth than specialization. Specifically, knowledge spillovers can arise among diverse firms and economic agents and thus stimulate innovation and growth; these spillovers are referred to as Jacobs' externalities.

1.1.3 Empirical Inconsistencies within Economic Structure Research

Based on the above theoretical assumptions, the role of economic structure in regional economic development has been studied empirically for a long time, with a recurrent focus on its relationship to regional economic stability (See Conroy, 1975; Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006; Deller and Watson, 2016a,b). The empirical relationship between industrial diversity and economic stability at the regional level has been tested, yet the results are often mixed and sometimes conflicting. For instance, a number of authors (e.g., Conroy, 1975; Kort, 1981; Wagner and Deller, 1998; Trendle, 2006) confirmed the negative relationship between diversity and instability, whereas others like Jackson (1984), Attaran (1986), and Deller and Watson (2016b) claimed that this relationship is not always significant.

Similarly, a closely related topic in the literature is the effect of economic diversity on regional unemployment. This topic can be viewed from two seemingly compelling theoretical perspectives, the portfolio theory (Conroy, 1975) and the search theory (Simon, 1988). The portfolio theory suggests that diversity can reduce risk in unemployment and that the positive or negative relationship between economic diversity and stability depends on business cycles. Chiang (2009, p. 952) further explained that “in a prosperous period, a specialized region experiences relatively lower unemployment than a diverse region,” while “a specialized region suffers relatively higher unemployment than a diverse region.” In con-

trast, the search theory claims that economic diversity can reduce unemployment because of greater employment opportunities within diversified economies. Based on these theoretical understandings, Malizia and Ke (1993) and Izraeli and Murphy (2003) verified the negative relationship between diversity and unemployment, while Mizuno et al. (2006) found that the relationship is weak and even non-existent in the case of metropolitan areas in Japan.

1.2 Research Questions

Many authors (e.g., Conroy, 1975; Kort, 1981; Siegel et al., 1995; Wagner and Deller, 1998) consider that the principal factors of the inconsistency between theoretical assumptions and empirical evidence of economic structure include (1) the use of highly-aggregated data sets, (2) inadequate or inappropriate measures of regional economic structure, and (3) overly simplistic statistical modeling methods. Table 1.1 further summarizes previous literature on the economic diversity research in terms of these factors.

This dissertation research aims at answering the following three research questions of geographical scales, structural measurements, and modeling methods in studying the empirical relationship between economic diversity and regional economic performance:

1. What is the most appropriate geographical scale for quantifying regional economic structure? And how do we make that determination for economic development?
2. Has regional economic structure been effectively defined and measured for studying their effects on economic performance? If not, how can these measures be improved?
3. What modeling techniques have been used to study the relationship between economic structure and economic performance? At a time when spatial regression models have been widely accepted in the regional economics literature, have these models been fully used?

These three research questions are situated within a much broader field of literature on economic structure research. Although these questions can be answered independently in Figure 1.1 using the traditional Problem-Plan-Data-Analysis-Conclusion (PPDAC) procedure suggested by MacKay and Ordford (2000), their focuses are not mutually exclusive in

Table 1.1: Selective overview of economic diversity research

Article	Spatial unit	Diversity measure	Modeling technique	Study region	Significant or not
Conroy (1975)	MSA	Portfolio variance	Bivariate correlation	MSAs in the U.S.	Yes
Kort (1981)	MSA	Entropy, national average, ogive, and percentage of durable goods	Bivariate correlation	106 MSAs in the U.S.	Yes
Jackson (1984)	County group	National average, ogive, portfolio variance and percentage of durable goods	Bivariate correlation	Illinois, U.S.	No
Keinath (1985)	EA	Tress Index, Shear Index and coefficient of specialization	Bivariate correlation	U.S.	No
Attaran (1986)	State	Entropy	Bivariate correlation	U.S.	No
Malizia and Ke (1993)	MSA	Entropy	Multivariate regression	All U.S. MSAs	Yes
Wagner and Deller (1998)	State	An input-output based measure	Multivariate regression	U.S.	Yes
Mizuno et al. (2006)	MSA	Herfindahl index	Multivariate regression	118 MSAs in Japan	No
Trendle (2006)	Local Government Area	Entropy	Spatial lag and error models	Queensland, Australia	Yes
Deller and Watson (2016a)	County	Herfindahl, national average, ogive and entropy	Spatial Durbin model	U.S.	Yes
Deller and Watson (2016b)	County	Herfindahl	Geographical weighted regression	U.S.	Not always
Hong and Xiao (2016)	MSA	A multiple specializations indicator and entropy	Spatial lag model	All U.S. MSAs	Yes
Watson and Deller (2017)	County	Herfindahl index	Spatial Durbin model	U.S.	Yes

Notes: MSA = Metropolitan Statistical Area; EA = Economic Area.

essence. For example, the understanding of geographical scales in the first question can help answer the other two questions. In doing so, possible errors can be minimized that result from highly-aggregated data sets. Henceforth, as demonstrated by the nested strategy in Figure 1.2, the conclusions of the first question are used to answer the other two, and the conclusions of the second question are used to answer the third one.

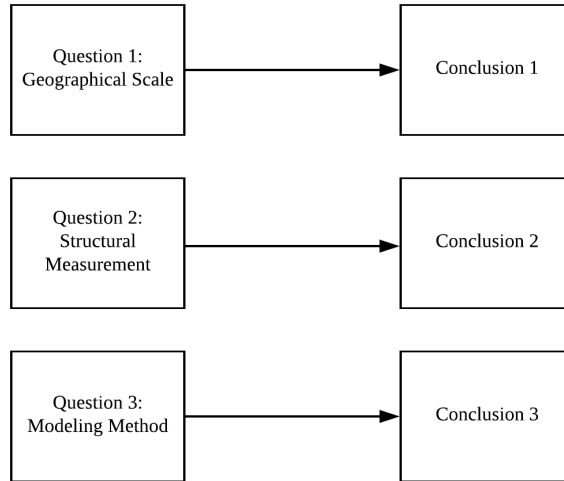


Figure 1.1: Independent strategy to solving the research questions

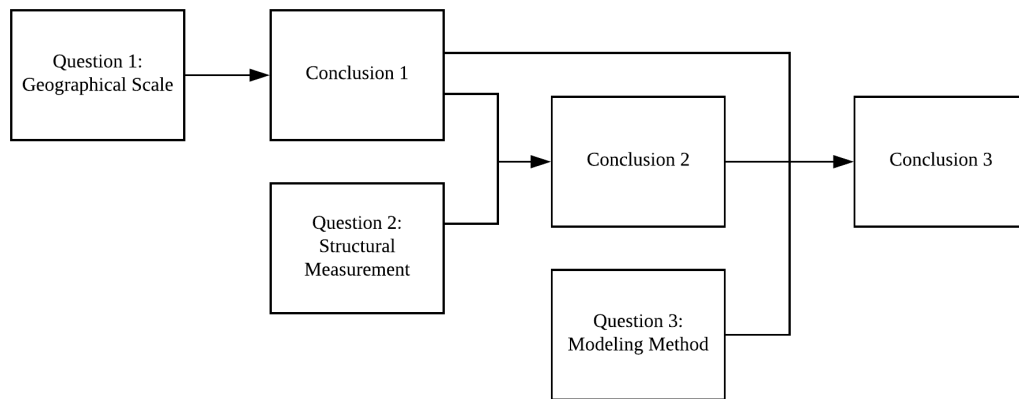


Figure 1.2: Nested strategy to solving the research questions

1.3 Organization

This introductory chapter focuses on the research context and proposes three interconnected research questions on geographical scales, structural measurements and modeling methods in economic structure research. These questions are explored sequentially in Chapters 2-4.

Chapter 2 evaluates the role of geographical scales in assessing the effect of economic diversity on regional economic performance, while keeping the economic structure measure and modeling method unchanged. Specifically, the relationship between economic diversity and regional economic stability is studied empirically using four different geographical units—including counties, states, Metropolitan Statistical Areas (MSAs) and Economic Areas (EAs)—in the context of the lower U.S. Among these units, counties and states are formal or administrative regions, while MSAs and EAs are functional regions. Their differences are further identified and highlighted in such analytical dimensions as full coverage and region size. In addition, the potential spatial spillover effects in the diversity-stability relationship are also considered through spatial econometric models.

Chapter 3 examines existing economic diversity measures in the perspective of the relationship between economic diversity and specialization. As most of these measures do not consider the coexistence of specialization and diversity (Malizia and Ke, 1993; Wagner and Deller, 1998; Jackson, 2015), this chapter develops an alternative measure that allows for interpreting economic diversity as the presence of multiple specializations. Future research thus can adopt this measure and gain additional insights beyond simple levels of employment dispersion between industries.

Chapter 4 overviews previous modeling methods employed in studying the relationship between economic structure and regional economic performance. It is found that these methods have not addressed model uncertainty resulting from the choice of control variables. In that sense, this chapter uses a Bayesian Model Averaging (BMA) method to consider model uncertainty in a spatial context and then assesses the structure-performance relationship.

Chapter 5 summarizes findings from Chapters 2 to 4, discusses their contributions to the literature and suggests future research directions.

Chapter 2

A Geographical Scale Perspective

2.1 Introduction

The role of industrial diversity in regional economic development has been studied for a long time, with a recurrent interest in its relationship to regional economic stability (e.g., Conroy, 1975; Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006; Deller and Watson, 2016a,b; Hong and Xiao, 2016). This interest can be dated back as early as the Great Depression (McLaughlin, 1930; Tress, 1938), and it is hypothesized that industrially diversified regions are associated with stable economic performance. Since then, this hypothesis has been studied both theoretically and empirically (Siegel et al., 1995; Wagner and Deller, 1998; Dissart, 2003; Mack et al., 2007). Likewise, industrial diversity and diversification research also attracts interests from policy makers and governments because of its impacts on regional economic development (Wagner and Deller, 1998; Feser et al., 2014; Jackson, 2015). More recently, scholarly research has focused on the relationship between industrial diversity and regional economic performance during the Great Recession in the late 2000s (e.g., Deller and Watson, 2016a,b; Watson and Deller, 2017).

Despite a rich body of literature on various aspects of industrial diversity, structural change and economic performance, there is little, if any, empirical analysis to compare the empirical relationship between industrial diversity and regional economic performance across different geographical scales. That said, variations in scale have not been incorporated into the literature, and the majority of current studies have used only a single geographical

scale in quantifying economic structure. For instance, in the U.S. context, states (Attaran, 1986; Wagner and Deller, 1998), counties (Deller and Watson, 2016a,b; Watson and Deller, 2017), Metropolitan Statistical Areas (MSAs; Conroy, 1975; Malizia and Ke, 1993; Hong and Xiao, 2016), and Economic Areas (EAs; Keinath, 1985; Porter, 2003) have been used as the analytical units. Beyond the U.S.-based studies, the corresponding geographical units include metropolitan areas in Japan (Mizuno et al., 2006), cities in China (Fu et al., 2010), and Local Government Areas in Australia (Trendle, 2006). Moreover, numerous scholars (e.g., Harvey, 1968; Watson, 1978; Ruddell and Wentz, 2009; Schaeffer et al., 2011; Kwan, 2012) have stressed the role of varying geographical scales as a generic issue in impacting empirical results. Meanwhile, scale-related issues—such as the modifiable areal unit problem (MAUP; Openshaw and Taylor, 1979; Openshaw, 1984) and the ecological fallacy (Robinson, 1950)—have been examined in specific research questions in regional science and economic geography (e.g. Briant et al., 2010; Resende, 2011; Dapena et al., 2016, 2017, 2018).

Given this, the goal of this chapter is to examine the relationship between industrial diversity and economic stability in the U.S. at multiple geographical scales (including states, counties, MSAs, and EAs) and to discuss several scale-related issues like the Modifiable Areal Unit Problem and functional regions vis-à-vis this relationship. This research is critical because, from a methodological perspective, it offers a multilevel perspective on the diversity-stability relationship. Furthermore, this research is also one of the first studies on geographical scales in economic diversity research. Although the scale-related issues introduced in this chapter have not been fully addressed, explicit attention should be paid to their potential effects when researching economic diversity as well as other regional science topics.

The remainder of this chapter is organized as follows. In the second section, the investigation of the role of geographical scale in assessing the effect of economic diversity draws on two perspectives: (1) the definitions of scale and scale-related problems, and (2) analytical units in economic diversity research, especially in the U.S. context. The third section describes the methodology of this analysis. After that, results are presented and discussed. Finally, the findings from this analysis are summarized.

2.2 Background

2.2.1 The Definitions of Scale and Scale-Related Problems

The concept of scale has various definitions in spatial, temporal, and other dimensions of regional research. Even though only spatial scale is focused on here, Goodchild (2011) suggested that there are three meanings to the concept. First, the term scale may refer to the *cartographic scale* or *representative fraction*, which is calculated as the ratio of a distance on a map to its corresponding distance in the real world—more generally, this ratio is displayed as a scale bar in cartography. By comparison, the second use of the term is the *spatial extent* of a study region. For example, the spatial extent or scale of West Virginia is much larger than that of New York City. Third, scale can also be defined as the *spatial resolution* or *analytical unit* of a study; in other words, scale is the smallest distinguishable part of an object (Tobler, 1988) like pixels in a satellite imagery. For this analysis, the third meaning of scale is used unless explicitly noted otherwise.

In reference to economic diversity research, the spatial scales can be counties (Deller and Watson, 2016a), states (Attaran, 1986; Wagner and Deller, 1998), and Metropolitan Statistical Areas (Kort, 1981; Malizia and Ke, 1993; Hong and Xiao, 2016). Jackson and Sonis (2001) suggested that the choice of spatial resolution or analytical unit is often decided by data availability and political considerations in the study of regional economic systems. Yet, data reporting units and political regions are often inconsistent with the boundaries of these systems. This inconsistency can result in two general scale-related problems, namely (1) the modifiable areal unit problem (MAUP) as proposed by Openshaw and Taylor (1979) and Openshaw (1984) and (2) the uncertain geographical context problem (UGCoP) suggested by Kwan (2012).

The MAUP arises when the analytical units are modifiable in size or spatial arrangement. According to Openshaw (1984), the MAUP includes two interrelated aspects: the scale problem and the zoning problem. First, the scale or aggregation problem occurs when the total number of geographical units changes and causes the statistical properties of aggregated or disaggregated units to vary. Second, the zoning problem refers to the effects resulting from different zoning or combinations of contiguous units. This is because there

are various approaches to aggregating any specified number of zones. The MAUP has been frequently discussed with the ecological fallacy (Robinson, 1950). In Robinson's study, the correlation between the percentage of native born population and the percentage of illiteracy is positive at the individual level but negative at the census division level, demonstrating that statistical relationships cannot be transferred from one scale to another. Peeters and Chasco (2006) further indicated that the concept of ecological fallacy is isomorphic to the MAUP. Methods to reduce impacts of the MAUP involve identifying the best zoning scheme or areal division (e.g., Farmer and Fotheringham, 2011; Kropp and Schwengler, 2016).

In relation to empirical analysis, Anselin (1988, p. 26) further suggested that the MAUP can result in two problems: first, the micro-macro aggregation problem always occurs when the spatial units are not homogeneous or spatial heterogeneity exists; and second, the MAUP can also alter the choice of a spatial weight matrix in spatial regression and the corresponding parameter of the spatial lag variables. The MAUP has also been studied for specific topics in regional science related research (e.g., Resende, 2011; Ouyang et al., 2014; Dapena et al., 2016, 2017, 2018). For instance, Resende (2011) examined the determinants of regional economic growth from 1991 to 2000 in Brazil at four spatial scales (including states, municipalities, micro-regions and spatial clusters) and also indicated that the empirical results vary greatly with the geographical scale. Similarly, when studying labor density and wages in Spain, Dapena et al. (2018) suggested that agglomeration externalities are generated on a local scale, such as local labor markets rather than NUTS-2 or NUTS-3 regions.

In contrast to the well-known MAUP, the uncertain geographical context problem or UGCoP has received much less attention (Kwan, 2012). It is different from the MAUP as it is not about the modifiable units but about the "true spatial configurations." The UGCoP arises when contextual or analytical units that are geographically defined differ from the true geographical context that are usually unknown in most studies to date. For instance, neighborhood regions like census tracts and postal code areas are used as contextual units to detect the geographical distribution of diseases in health research. These regions do not reflect "the actual areas that exert contextual influences on the health outcome under study" (Kwan, 2012, p. 959) because the boundaries of people's activities—such as commuting to

work and social interactions—usually are larger than neighborhoods. Moving beyond health research, Kwan (2012) further argued that any study that examines the effects of area-based attributes on geographical phenomena faces the UGCoP. Of special relevance here is the boundary issue in spatial analysis (e.g., Anselin, 1988; Burt et al., 1996; O’Sullivan and Unwin, 2014). For example, Anselin (1988) discussed the consequence of the boundary value problem within the general frameworks of the spatial autoregressive model (SAR) and the spatial error model (SEM). In addition, the UGCoP also emphasizes the temporal uncertainty in a spatial phenomenon. Because economic activities and interactions vary over space and time, the true spatiotemporal context of these activities becomes uncertain¹.

2.2.2 Analytical Units in Economic Structure Research

The geographical units used in the literature of economic diversity include cities (McLaughlin, 1930), MSAs (Conroy, 1975; Kort, 1981; Hong and Xiao, 2016), county groups (Jackson, 1984), states (Attaran, 1986; Wagner and Deller, 1998), counties (Mack et al., 2007; Deller et al., 2017; Deller and Watson, 2016a,b; Watson and Deller, 2017) and Economic Areas (Keinath, 1985) and can be divided into two types: formal and functional regions². Formal regions like states and counties are defined by government, and the boundaries of these regions are normally stable over time. By comparison, functional regions are defined “based on a greater magnitude of interactions or connections among spatial units within a region than with units outside the region” (Brown and Holmes, 1971, p. 58). One familiar example of functional regions is the U.S. Metropolitan Statistical Areas that have been regularly defined based on population density and commuting ties.

While functional regions have been studied for a long time in geography and regional science (e.g., Philbrick, 1957; Nystuen and Dacey, 1961; Fox and Kumar, 1965; Brown and Holmes, 1971), with the exception of Jackson (1984), much of the economic diversity literature prior to the 1990s has not differentiated functional regions with formal regions.

¹The temporal variations of geographical scales have also been discussed separately as the ecological fallacy in a temporal context by Duque et al. (2006) and the modifiable temporal unit problem (MTUP) by Cheng and Adepeju (2014).

²Although there are other forms of functional economic regions in geographical analysis (e.g., Farmer and Fotheringham, 2011; Kropp and Schwengler, 2016), these regions have not been used to study industrial diversity.

Presumably, the unit of observation in these early studies were determined mainly by availability of data sets at that time. Specifically, when studying the economic diversity of Illinois for the 1959-1976 period, Jackson (1984, p. 106) first articulated the use of functional regions as the analytical units: “individual county data were aggregated into fifteen multicounty regions which exhaust the entire area of the State of Illinois. Each region centres around a major city so as to approximate functional economic areas.” After that, Malizia and Ke (1993) reviewed geographical units used in economic diversity research and further suggested that only function regions should be used.

However, neither Jackson (1984) nor Malizia and Ke (1993) have linked the choice of geographical units with scale-related problems. From a theoretical point of view, formal regions (like counties and states) as the unit of analysis can be seriously affected by the MAUP and the UGCoP. For example, the city of St. Louis in Missouri was built along the western bank of the Mississippi River and shares its boundaries with the state of Illinois. Because state boundaries do not generally hinder economic interactions, the city's businesses might have stronger economic interactions with some businesses in Illinois than with those in the rest of Missouri. These interactions can be local inter-industry flows, face-to-face contact in formal business meetings or frequent informal contact, such as lunch meetings, sports activities, or other social occasions. All of these interactions can bring indirect effects to regional economic performance but cannot be reflected meaningfully in data sets based on formal regions.

In addition to the two general scale-related problems mentioned above, there are another two practical problems regarding county-level data sets. One of these problems is the data issue of difference between place of work and place of residence and many economic diversity studies have not paid attention to it. For instance, the demographic data in the Census are normally measured based on place of residence, whereas the employment data by industry published by the Census County Business Patterns (CBP), the Bureau of Labor Statistics (BLS) Census of Employment and Wages (CEW) and the Bureau of Economic Analysis (BEA) regional data are measured based on place of work³. This inconsistency can greatly

³For more information on these data sets, see U.S. Bureau of Economic Analysis, FAQ: “What is the difference between BEA employment and wages and BLS and Census employment and wages?” (January 12, 2006), https://www.bea.gov/faq/index.cfm?faq_id=104, accessed May 18, 2018.

impact empirical analysis. To mitigate this problem, Dapena et al. (2018, p. 61) indicated that using functional regions that are defined based on population and commuting flows as the analytical units suffer less from this issue because ones place of residence and place of work “are in the same spatial unit.”

The other problem for county-level economic diversity research is that the employment and population of some counties are too small. For example, according to the 2010 Census, the population of Loving County, Texas, is only 115, while that number is 9,818,605 for Los Angeles County, California. This enormous difference makes it almost meaningless to compare their industrial activities and economic performance. More generally, this refers to the small population problem in spatial analysis (Wang and O’Brien, 2005), where small population units (1) are not fair representations of the underlying spatial form or process⁴ and (2) are sensitive to data errors resulting from data collection, data reporting and others. To address this problem, several strategies have been proposed, including removal of small population samples and aggregating small regions into larger ones based on some standards. As for the former, it might exclude valuable samples from the population. By comparison, the latter, in essence, relates to the use of functional regions as the analytical units, which is originally suggested by Jackson (1984) and Malizia and Ke (1993) in the literature of economic diversity.

However, both formal and functional regions have been used in the literature. On the one hand, following the suggestions of Jackson (1984) and Malizia and Ke (1993), some studies (e.g., Trendle, 2006; Mizuno et al., 2006; Hong and Xiao, 2016) have used functional regions—such as MSAs in the U.S. and Local Government Areas in Australia—as the analytical units. For instance, Trendle (2006) used Local Government Areas⁵ to study industrial diversity and also considered the role of spatial spillovers within the diversity-stability relationship by using spatial econometric models.

On the other hand, other studies (e.g., Wagner and Deller, 1998; Mack et al., 2007; Deller and Watson, 2016a,b; Watson and Deller, 2017) have used formal regions like states and counties as the basic units. For example, Wagner and Deller (1998) used states to study

⁴This is similar to the UGCoP.

⁵According to Trendle (2006), Local Government Areas are functional regions and include both urban and rural regions.

diversity and economic performance because of direct policy relevance and data availability. For the same reasons, Deller and Watson (2016a,b) and Watson and Deller (2017) recently used counties as the units of observation. Particularly, Deller and Watson (2016a, p. 1828) admitted that counties are not functional economic areas, and thus “spillover effects across county lines” should be expected. As such, spatial econometric models were employed to account for these spatial spillovers. In addition to the political and data considerations mentioned above, Mack et al. (2007, p. 481) also indicated that such formal regions as counties and states are “more stable units for analysis over time” and the boundaries of these regions are less prone to change than the boundaries of MSAs⁶.

Taken together, even though previous scholars suggested that functional regions—rather than formal regions—should be used to measure regional economic structure, no comparison has been made between these two types of regions in the literature of economic diversity, especially when spatial spillovers are considered. Moreover, since Jackson (1984) and Malizia and Ke (1993) originally suggested the preference for functional regions as the units of observation, this preference, while widely accepted, has not been formally discussed with such scale-related problems as the MAUP yet or interpreted as an artifact of these problems within the literature of economic structure. In that sense, this analysis (1) investigates the relationship between industrial diversity and economic stability in the U.S. context at multiple geographical scales, (2) compares the role of spillover effects in the relationship among functional and formal regions, and (3) discusses the scale-related problems⁷ with the diversity-stability relationship. The next section introduces the methodology to accomplish these tasks.

⁶It should be noted that Mack et al. (2007) first admitted that using counties as analytical units suffers from the MAUP.

⁷The UGCoP is not the main focus of this research because it is difficult to include any true contexts of industrial activities using existing data sets. The UGCoP might be a topic for future research.

2.3 Methodology

2.3.1 Geographical Units

This analysis uses four geographical units, including counties, states, Economic Areas and Metropolitan Statistical Areas. To begin with, county-level data are used to explore the relationship between industrial diversity and economic stability at the most disaggregated level. There are several technical issues that need to be overcome when using county-level data. To ensure consistency, for example, it is necessary to combine 50 independent cities and surrounding counties in Virginia to form 23 county equivalents because of strong economic and social connections. Some newly-established counties like Broomfield, Colorado, were merged with their parent counties. As a result, a total of 3,079 counties or county equivalents in the lower U.S. were included. Next, these counties or county equivalents were spatially aggregated into 48 states and the District of Columbia; state is thus utilized as the second geographical unit.

In addition to counties and states, two functional regions—including EAs and MSAs—are used. On the one hand, EAs are defined by the Bureau of Economic Analysis (BEA) based on county-level population and commuting flows on a regular basis⁸. However, because of government funding reasons, the BEA has stopped updating the boundaries of EAs since 2014⁹. The latest two updates of the Economic Areas were made in 1995 and 2004, and the 1995 version of EAs is used in this analysis. In total, there are 170 EAs covering the lower U.S. (See Figure 2.1).

On the other hand, the relationship between diversity and economic stability is also studied among 359 MSAs. As defined by the U.S. Office of Management and Budget (OMB), the MSAs¹⁰ contain at least one core urban area with a population of 50,000 or more and adjacent counties with at least 25% of the workers living in the county while working in the core urban area. Based on this definition, the delineation of MSA boundaries updates regularly, and the latest three updates were made in 2017, 2015, and 2013 respectively.

⁸For more information about the defining process, see Johnson (1995) and Johnson and Kort (2004).

⁹See “BEA Economic Areas (EAs),” U.S. Bureau of Economic Analysis, <https://www.bea.gov/regional/docs/econlist.cfm>, accessed November 30, 2017.

¹⁰Standard Metropolitan Statistical Area and Standard Metropolitan Area were used historically. For more information about the U.S. metro areas, see Adams et al. (1999).

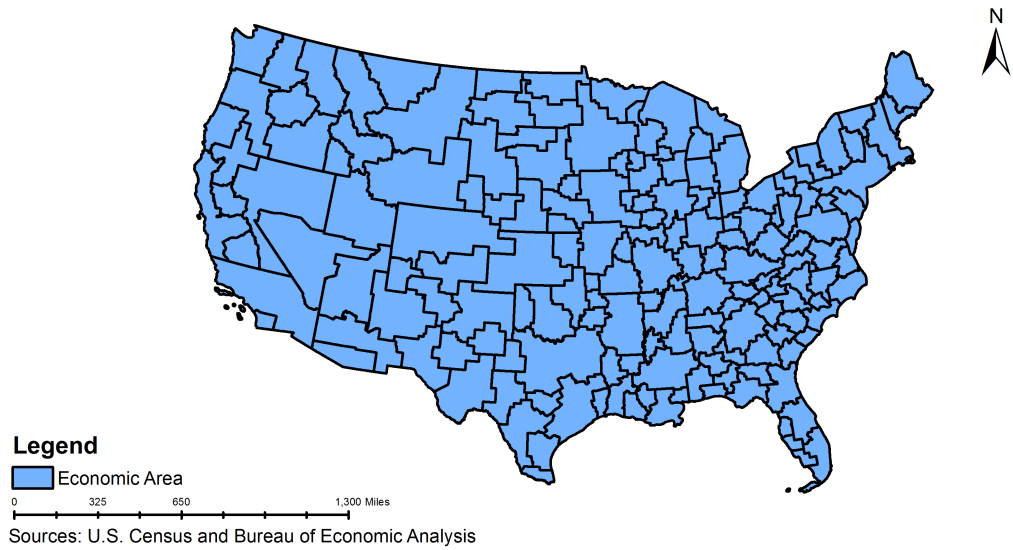


Figure 2.1: Economic Areas in the lower U.S.

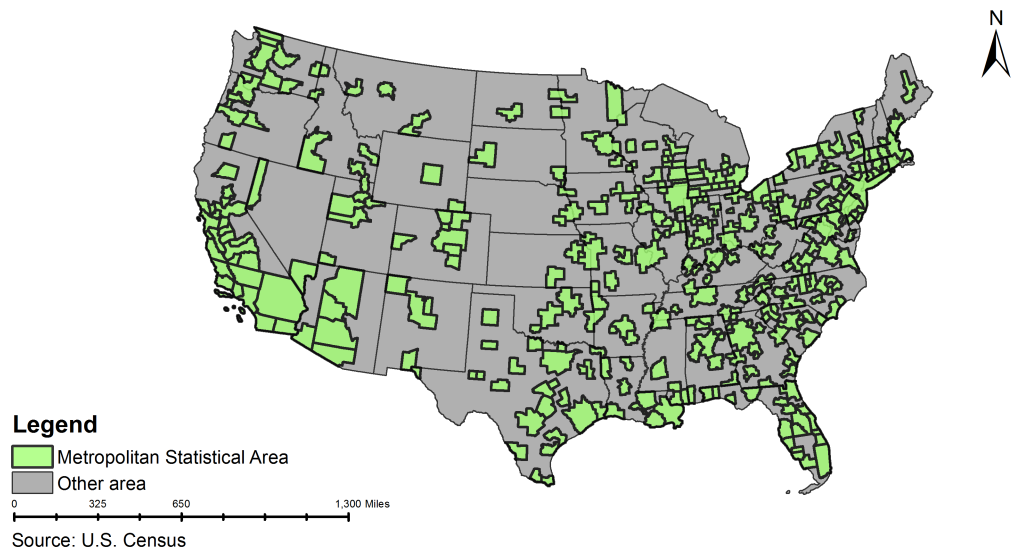


Figure 2.2: Metropolitan Statistical Areas in the lower U.S.

For our purpose, the 1999 version of MSAs was used because it was the basis of data publications in the 2000 Census. As shown in Figure 2.2, MSAs do not cover all the areas in the contiguous U.S and only include urban regions. In that sense, it is not possible to observe the diversity-stability relationship in rural areas using MSA data sets.

2.3.2 Measuring Diversity

As reviewed by previous authors (See Siegel et al., 1995; Dissart, 2003; Mack et al., 2007; Jackson, 2015), there are numerous ways to empirically measure economic diversity like the entropy index, the Herfindahl-Hirschman Index (HHI), the Ogive and the national average. In this analysis, economic diversity is measured as regional employment dispersion using the HHI as follows:

$$HHI_i = \sum_{j=1}^N (e_{ij}/e_i)^2 \quad (2.1)$$

where N is the total number of industries in the i^{th} region; e_{ij} is the number of employment for industry j in the i^{th} region; and e_i is the total number of employment in the i^{th} region.

According to Equation 2.1, a higher HHI indicates the regional economic structure is specialized, while regions with lower HHI values denote that their economies are industrially diversified. HHI ranges from $1/N$ for a perfectly diversified economy to one if all employment is concentrated in one industry. Because the regional science literature assumes that diversified economies are stable in their economic performance, the HHI is expected to be positively associated with the dependent variable or REI. Although HHI has been questioned empirically and theoretically in regional science research (Jackson, 1984, 2015; Wagner and Deller, 1998; Wagner, 2000), this index is still used because it does not require additional data.

Like previous studies (e.g., Jackson, 1984; Malizia and Ke, 1993; Deller and Watson, 2016a,b), economic diversity is calculated based on employment data obtained from the County Business Patterns. Compared with other sources of employment data used in economic diversity research (Table 2.1), the CBP series are preferred for the following three reasons. First, the CBP data are measured by place-of-work and thus are better representatives of the region's industrial structure than employment data by place-of-residence.

Table 2.1: Different data used in U.S.-based economic diversity research

Article	Spatial unit	Period	Industry data source	Level of detail in industries
Conroy (1975)	MSA	1958-1967	BLS	Three-digit SIC manufacturing sectors
Kort (1981)	MSA	1967-1976	BLS	Two-digit SIC nonagricultural sectors
Jackson (1984)	County group	1969-1976	CBP	21 nonagricultural sectors
Keinath (1985)	EA	1971-1978	BEA	14 sectors
Attaran (1986)	State	1972-1981	BLS	Eight nonagricultural sectors
Malizia and Ke (1993)	MSA	1972-1988	CBP	Two-digit SIC sectors
Wagner and Deller (1998)	State	1969-1992	MicroIMPLAN database	SIC sectors in MicroIMPLAN
Deller and Watson (2016a,b)	County	2005-2012	CBP	Three-digit NAICS sectors
Hong and Xiao (2016)	MSA	1998-2010	CBP	Three-digit NAICS sectors
Watson and Deller (2017)	County	2007-2014	CBP	Three-digit NAICS sectors

Notes: MSA = Metropolitan Statistical Area; EA = Economic Area; BLS = Bureau of Labor Statistics; BEA = Bureau of Economic Analysis; CBP = County Business Patterns; SIC = Standard Industrial Classification; NAICS = North American Industry Classification Systems.

Second, the CBP data provide great details on employment for different levels of sectors and geographical areas. Specifically, these data are published annually by the U.S. Census Bureau and contain 2-6 digit North American Industry Classification Systems (NAICS) industrial activities in terms of employment, payroll, and number of establishments at different geographical levels, including the whole country, states, MSAs, counties and zip-code areas. This analysis uses the 3-digit level non-agricultural industries of NAICS (83 sectors) from the CBP. Third, the CBP data suffer less from missing values and have more complete records regarding employment than other sources like the BEA estimates of employment and the BLS data. To protect the confidentiality of its workforce indicators, the U.S. Census Bureau¹¹ uses a series of suppression flags to represent employment ranges: Flag A represents 0-19 employees; Flag B, 20 to 99; and Flag C, 100 to 249. In an earlier study, Deller and Watson (2016a) used data relating to establishments in the CBP where the disclosure rules do not apply. As the sizes of the establishments are unknown, however, an establishment with more than 1,000 employees is treated the same as an establishment with less than ten workers. To circumvent this situation, these employment ranges were

¹¹The BEA and the BLS employment data also have the issue of missing values because of confidential reasons.

estimated with the Upjohn Institute’s “WholeData,” a software that uses Isserman and Westervelt’s (2006) two-stage method to estimate missing values in the CBP. This study thus makes use of the complete CBP data for the year 2000 for the assessment of regional industrial diversity.

2.3.3 Empirical Model

The empirical model can be described as:

$$REI_i = f(HHI_i, CONTROL_i) \quad (2.2)$$

where REI_i stands for regional economic instability (REI) defined by Malizia and Ke (1993); HHI_i is the independent variable of economic diversity; and $CONTROL_i$ is a set of control variables. All the variables on the right hand side of Equation 2.2 reflect the state of the region’s economy for the base year of 2000, while the dependent variable reveals changes from 2000 to 2014.

Regional economic instability is measured using county level employment from 2000 to 2014 from the Bureau of Economic Analysis. According to Malizia and Ke (1993), REI is calculated as the average deviation of total employment from its linear trend (i.e., estimated employment using linear regression) and divided by the trend and can be expressed as:

$$REI_i = \left\{ \sum_{i=1}^N [(E_{it} - E_{it}^{Tr})/E_{it}^{Tr}]^2 / T \right\}^{1/2} \quad (2.3)$$

where i denotes the region index; T is the number of time spans; E_{it} is the actual number of workers for region i at time t ; and E_{it}^{Tr} is the predicted number of workers for region i at time t using a linear trend line. A lower value of REI_i indicate a stable economy over the period of 2000 to 2014.

In addition, based on existing empirical literature (e.g., Malizia and Ke, 1993; Trendle, 2006; Deller and Watson, 2016a; Hong and Xiao, 2016), a set of control variables were retrieved from the 2000 Census¹² and are used to capture the potential demographic, indus-

¹²Note that most of these data were measured by place of residence.

trial and economic differences among regions. Overall, Table 2.2 displays the description of all the variables.

Table 2.2: Description of all the variables

Category	Symbol	Description	Data source
Dependent variable	REI	Average deviation of total employment from its trend divided by trend employment, 2000-2014	BEA
Independent variable	HHI	Indicator of industrial diversity using the Herfindahl Hirschman Index, 2000	CBP
Control variables	Popc	Absolute value of population change, 2000-2014	Census
	Pop	Population size, 2000	Census
	Edu	Percentage of population 25 years or older with at least a bachelors degree, 2000	Census
	Nonwhite	Percentage of non-white population, 2000	Census
	NE	Dummy variable for Northeast states	Census
	South	Dummy variable for Northeast states	Census
	MW	Dummy variable for Midwest states	Census

Notes: BEA = Bureau of Economic Analysis; CBP = County Business Patterns.

2.3.4 Statistical Methods

Exploratory Spatial Data Analysis

Exploratory spatial data analysis (ESDA) offers a set of techniques to describe and visualize spatial association, identify outliers and clusters (or hot spots) and other forms of spatial heterogeneity (Anselin, 1999). For this analysis, the methods of ESDA provide global and local spatial autocorrelation measures that can examine spatial patterns of industrial diversity and economic instability in the contiguous U.S. at multiple scales¹³. The global measure of spatial autocorrelation can be calculated using Morans I index as follows.

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \quad (2.4)$$

where n is the number of observations; z_i and z_j are the deviations from the mean; and w_{ij} is the corresponding element (i, j) in a spatial weight matrix. Moran's I was computed in GeoDa (Anselin et al., 2006) and ranges from -1 to 1. When it is zero, there is no spatial

¹³Mack et al. (2007) also examined the spatial patterns of economic diversity among U.S. counties using global and local spatial autocorrelation measures.

relationship or a random spatial distribution. When I is greater than zero, there is a positive spatial autocorrelation and similar values (high or low) tend to cluster. Conversely, when I is less than zero, there is a negative spatial autocorrelation and dissimilar values have a tendency to cluster.

Although the global Moran's I summarizes the overall spatial autocorrelation of variables or reveals whether values are clustered or not, it fails to tell us where the specific clusters and outliers are. For this purpose, the local Moran's I , a Local Indicator of Spatial Association (LISA; Anselin, 1995), is used to detect localized spatial patterns and can be written as follows:

$$I_i = z_i \sum_j w_{ij} z_j \quad (2.5)$$

where z_i and z_j also denote the deviations from the mean and w_{ij} stands for the i^{th} row and j^{th} column in the spatial weight matrix. The local Moran's I calculated in GeoDa divides regions into five categories of local spatial association between a region and its neighbors:

1. High-High: a region with a high¹⁴ value surrounded by regions with high values;
2. Low-Low: a region with a low value surrounded by regions with low values;
3. Low-High: a region with a low value surrounded by regions with high values;
4. High-Low: a region with a high value surrounded by regions with low values; and
5. Not Significant: a region that is not statistically significant in terms of spatial association.

Spatial Regression Analysis

This analysis uses spatial econometric techniques to study the diversity-stability relationship. As from previous literature, this relationship has been studied using various techniques, such as bivariate techniques (Conroy, 1975; Jackson, 1984; Attaran, 1986), ordinary least squares (OLS) regression (Malizia and Ke, 1993), spatial Durbin model (Deller and

¹⁴High and low values are relative to the mean.

Watson, 2016a), and spatial lag and error models (Trendle, 2006). Because Trendle (2006) and Deller and Watson (2016a) have confirmed the existence of spatial spillovers in the diversity-stability relationship, spatial econometric models seem to be preferred. However, as suggested by previous authors (Anselin, 1988; Duque et al., 2012), the choice of geographical scales can change the estimates of spatial regression models and even the level of spatial autocorrelation. In that sense, the standard OLS and three basic spatial regression models—including the spatial autoregressive model, the spatial error model and the spatial Durbin model—are used. To avoid *ad hoc* model specification, Elhorst’s (2014) approach that includes several Lagrange multiplier (LM) and likelihood ratio (LR) tests is used to select the most appropriate one among these four models (i.e., OLS, SAR, SEM and SDM).

Traditionally, the standard OLS regression estimates the dependent variable as a linear combination of industry diversity and a set of controlled variables. To incorporate spatial dependence effects, the OLS model has been extended into three basic spatial regression models, namely spatial autoregressive model, spatial error model and spatial Durbin model. First, the spatial lag model, or spatial autoregressive model, hypothesizes that the dependent variable at a particular location can be explained not only by the independent variables but also by the dependent variable values of neighbors. Second, the spatial error model assumes that the dependent variable is determined by a set of explanatory variables and that the error terms are spatially correlated. Finally, the spatial Durbin model posits that the dependent variable can be explained by independent variables, spatially lagged independent, and dependent variables. The model extends the SAR model by incorporating spatially lagged independent variables. Overall, Equations 2.6-2.8 present these spatial models and can be estimated through maximum likelihood (ML) procedures.

$$y = \rho W y + X \beta + \varepsilon \tag{2.6}$$

$$y = X \beta + \varepsilon, \varepsilon = \rho W \varepsilon + \mu \tag{2.7}$$

$$y = \rho W y + X \beta + W X \theta + \varepsilon \tag{2.8}$$

Sensitivity Analysis

A spatial weight matrix is required to model the spatial relationship between regions in both ESDA and spatial econometric analysis. According to Anselin (1988) and LeSage and Pace (2009), there are various methods to specify the spatial relationship, including distance-based (k -nearest neighbors or KNN and threshold distance) and contiguity-based (queen and rook contiguity). Among these methods, the most commonly used criterion is queen contiguity, where two regions, i and j , are neighbors if they share at least one point on their boundaries (e.g., the Four Corners Region between Arizona, Colorado, New Mexico and Utah). However, for our purpose, a spatial weight matrix based on six nearest neighbors ($k=6$) is used because it (1) solves the problem of discontinuity of MSAs in the contiguous U.S. and (2) allows each region to have the same number of neighbors (Le Gallo and Ertur, 2003). Although LeSage and Pace (2014) indicated that the choice of spatial weight matrix is of secondary concern compared to model specification, a sensitive analysis of the results is still conducted for spatial weight matrices based on queen contiguity, four, eight and ten nearest neighbors.

2.4 Empirical Results

Basic Statistics

Table 2.3 examines the basic statistics including means, ranges, standard deviations and correlation coefficients across different scales. Particularly, when counties are aggregated into states and EAs, the corresponding range and standard deviation decrease significantly. This can be explained by the scale or aggregation effect in the MAUP. Then, as reflected by the correlation coefficients, diversity and stability are not highly correlated. However, the diversity index is positive no matter what spatial units are used. This suggests that higher levels of diversity indicate greater stability in economic performance and is also consistent with conventional wisdom. Comparatively speaking, the correlation coefficient seems to be higher in the EA- and state-level analyses than the county-level study, meaning that the diversity-stability relationship is stronger among these aggregated regions.

Table 2.3: Basic statistics of the HHI and REI variables

Geographical scale	County	State	Economic Area	Metropolitan Statistical Area
Sample size	3079	49	170	359
HHI mean	0.070	0.035	0.034	0.040
HHI range	0.914	0.057	0.060	0.151
HHI std. dev.	0.061	0.010	0.007	0.016
REI mean	0.035	0.019	0.022	0.024
REI range	0.684	0.053	0.135	0.066
REI std. dev.	0.028	0.009	0.014	0.012
Correlation	0.233	0.357	0.423	0.188

Notes: HHI = Herfindahl Hirschman Index; REI = Regional Economic Instability; Correlation coefficients are significant at 1% level.

ESDA

Table 2.4 shows the global Moran’s I computed based on six nearest neighbors. The geographical scale can greatly change the extent of spatial autocorrelation in both variables. For example, the Moran’s I of the state-level HHI is 0.025, while that number for Economic Areas is almost ten times (0.222) more. Moreover, given that a value of one indicates perfect spatial autocorrelation and a value of zero indicates random distribution, these two variables are not highly clustered. Comparatively, the overall Moran’s I for the REI is more or less greater than that for the HHI at all geographical scales, indicating that economic stability appears to be more clustered than industrial diversity.

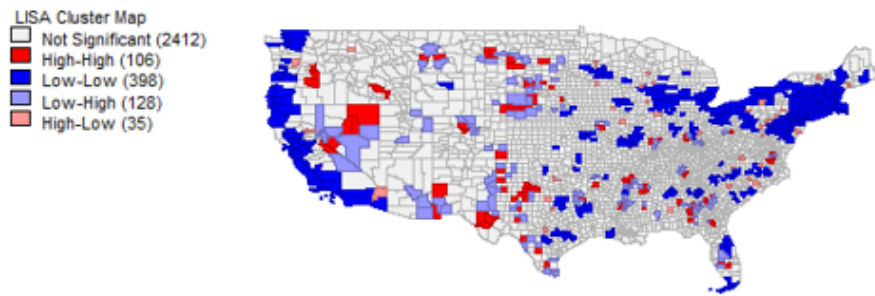
Table 2.4: Global Moran’s I results (six nearest neighbors)

Geographical scale	County	State	Economic Area	Metropolitan Statistical Area
HHI	0.097	0.025	0.222	0.011
REI	0.299	0.444	0.232	0.433

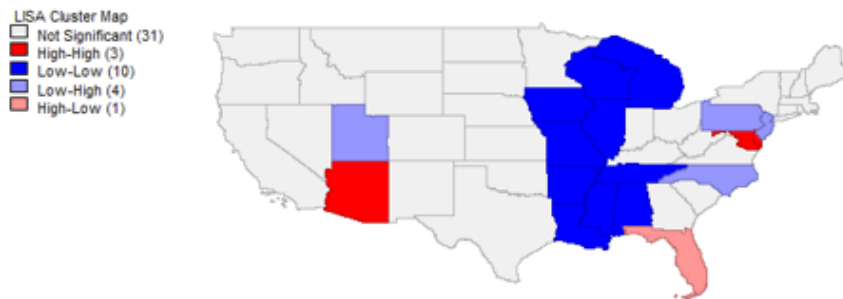
Notes: HHI = Herfindahl Hirschman Index; REI = Regional Economic Instability; All statistics are significant at 1% level.

Figures 2.3 and 2.4 display the LISA cluster maps of the two variables¹⁵. Two findings

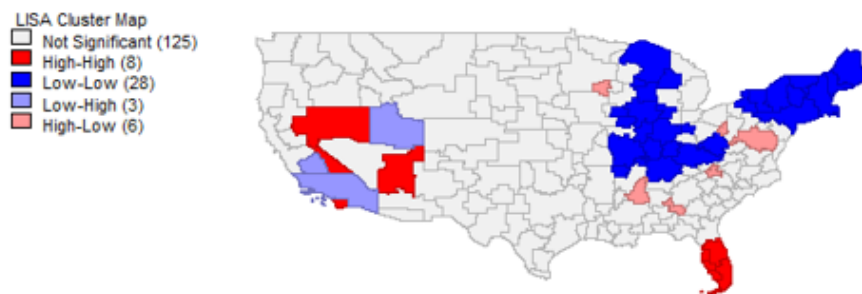
¹⁵The projections in these LISA maps are different from the projections in other maps like Figures 2.1 and 2.2 because GeoDa does not contain any tools to set or change projections. Maps in GeoDa also do not have some cartographic features like direction arrows and scale bars. For more information, see “Answers to Technical GeoDa Questions,” GeoDa Center for Geospatial Analysis, <https://geodacenter.github.io/questions.html>, accessed June 15, 2018



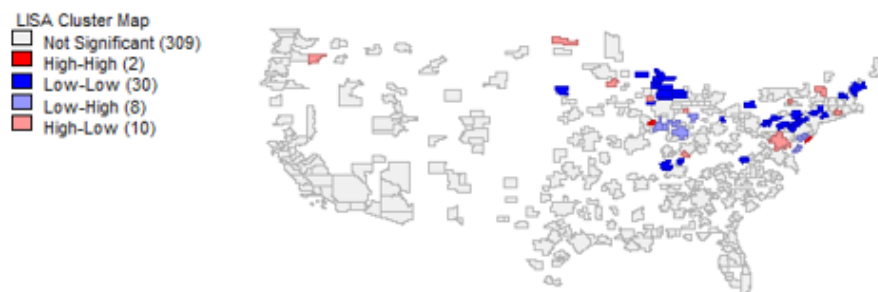
(a) Counties



(b) States



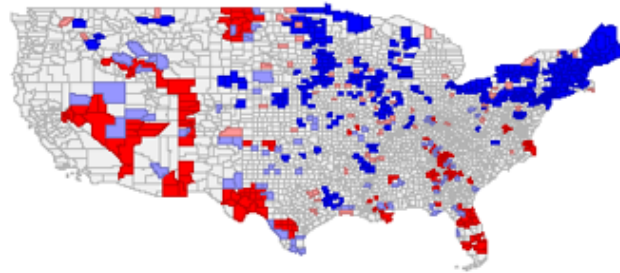
(c) EAs



(d) MSAs

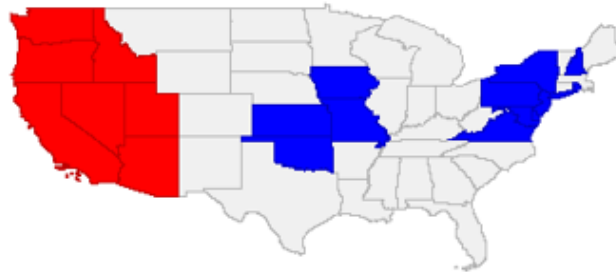
Figure 2.3: LISA maps of HHI in the contiguous U.S. (six nearest neighbors)

LISA Cluster Map
 Not Significant (2398)
 High-High (171)
 Low-Low (368)
 Low-High (82)
 High-Low (60)



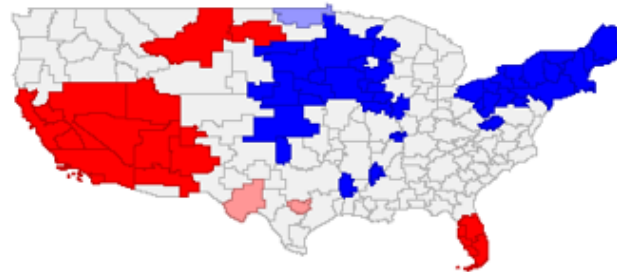
(a) Counties

LISA Cluster Map
 Not Significant (30)
 High-High (7)
 Low-Low (12)
 Low-High (0)
 High-Low (0)



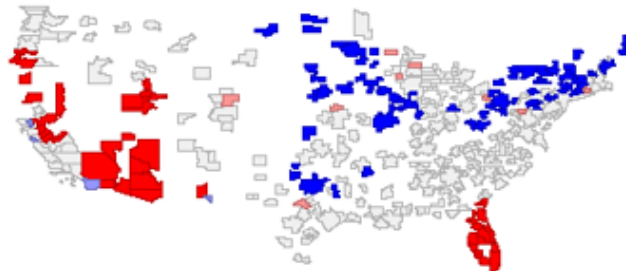
(b) States

LISA Cluster Map
 Not Significant (110)
 High-High (19)
 Low-Low (38)
 Low-High (1)
 High-Low (2)



(c) EAs

LISA Cluster Map
 Not Significant (233)
 High-High (42)
 Low-Low (71)
 Low-High (4)
 High-Low (9)



(d) MSAs

Figure 2.4: LISA maps of REI in the contiguous U.S. (six nearest neighbors)

are most evident from these two maps. First, the spatial distributions of economic stability and industrial diversity vary greatly with the choice of geographical scale. For example, as shown in Figure 2.4, some regions in Florida are significant local clusters of high values (i.e., High-High regions) in terms of the REI variable at the county-, EA- and MSA-level analyses, suggesting that these regions have high levels of economic instability. However, Florida is not a High-High region at the state level. Second, Low-Low regions in terms of both diversity and stability tend to be located in the northeast and central parts of the country at all spatial scales, although the boundaries of these regions differ. The interpretation here is that these industrially diversified regions or clusters experience stable economic performance through the 2000-2014 period.

Spatial Regression Analysis

Table 2.5: OLS regression results

Explanatory variables	County	State	Economic Area	Metropolitan Statistical Area
Diversity	0.194*** (0.000)	0.315 (0.253)	0.673*** (0.000)	0.284*** (0.000)
Popc	0.091*** (0.000)	0.089 (0.189)	0.015*** (0.000)	0.135*** (0.000)
Pop	-0.052*** (0.000)	0.117* (0.063)	0.063 (0.065)	0.024*** (0.279)
Edu	-0.1518*** (0.000)	-0.576*** (0.065)	-0.155 (0.359)	-0.039*** (0.000)
Nonwhite	-0.001 (0.839)	-0.073 (0.510)	-0.072 (0.190)	-0.008 (0.801)
NE	-0.254*** (0.000)	-0.425** (0.046)	-0.513*** (0.000)	-0.387*** (0.000)
South	-0.011 (0.632)	-0.519*** (0.000)	-0.206** (0.004)	-0.214*** (0.000)
MW	-0.190*** (0.000)	-0.545*** (0.003)	-0.370*** (0.000)	-0.314*** (0.000)
Constant	-2.390*** (0.000)	-5.076*** (0.000)	-2.354*** (0.003)	-3.256*** (0.000)
Adjusted R-squared	0.239	0.408	0.412	0.324

Notes: With the exception of the dummy variables, all variables are measured as natural logs; Numbers in the parentheses represent p-values; Significance levels: * for 10%, ** for 5%; *** for 1%.

To provide an initial baseline of the relationship between industrial diversity and economic stability, a normal linear model was estimated using the OLS technique for the four

geographical scales (See Table 2.5). With the exception of state-level analysis, the coefficient of diversity is statistically significant. That is to say, the hypothesized positive relationship between industrial diversity and economic stability cannot be rejected when counties, EAs and MSAs are used as the basic units. Moreover, like existing studies on the diversity-stability relationship, the control variables included in this analysis also play a role in impacting economic stability. Nevertheless, the choice of scale also alters the coefficient and significance of these variables. Finally, as reflected by the adjusted R-squared values, the explanatory power of the model seems to increase when more aggregated units (states and EAs) are used. More specifically, at the county level, the adjusted R-squared term is 0.239, whereas that number is 0.408 at the state level and 0.412 at the EA level.

Table 2.6: Lagrange multiplier and likelihood ratio tests (six nearest neighbors)

Geographical scale	County	State	Economic Area	Metropolitan Statistical Area
LM lag	772.805*** (0.000)	13.666*** (0.000)	11.027*** (0.000)	114.813*** (0.000)
Robust LM lag	74.959*** (0.000)	13.450*** (0.000)	6.985*** (0.008)	48.732*** (0.000)
LM error	799.94*** (0.000)	2.263 (0.132)	4.471** (0.034)	66.455*** (0.000)
Robust LM error	102.102*** (0.000)	2.047 (0.152)	0.429 (0.512)	0.374 (0.540)
LR lag	250.800*** (0.000)	10.798 (0.213)	10.602 (0.225)	35.518*** (0.000)
LR error	254.644*** (0.000)	16.449** (0.036)	16.469** (0.036)	94.756*** (0.000)

Notes: Significance levels: * for 10%, ** for 5%; *** for 1%; Numbers in the parentheses represent p-values.

Because the OLS estimation can ignore the potential spatial dependence and the choice of geographical scale can alter the level of spatial dependence, analyzing the diagnostics for spatial dependence becomes essential. Table 2.6 shows the results of (robust) LM and LR tests. Following Florax et al. (2003), LeSage and Pace (2009) and Elhorst (2014), the best model is chosen among OLS, SAR, SDM and SEM. The LM test results suggest that the non-spatial models are less appropriate than the spatial ones at all geographical scales. Specifically, when counties are used, the LM tests suggest that the SEM is preferred. To further investigate which spatial model offers the best fit or whether the SDM model could

be simplified into the SAR or SEM model, one may perform the likelihood ratio (LR) tests. The LR tests inform us that the SDM best describes the county-level data. Indeed, Deller and Watson (2016a,b) and Watson and Deller (2017) used the SDM in their county-level analysis. By comparison, when states and EAs are used, the LM tests indicate that the SAR better explains the data than does the SEM model, while the LR tests imply that the SDM can be simplified into the SAR. Finally, although the LM tests indicate that the SAR better explains the MSA level data than does the SEM model, the LR tests reject the SAR model in favor of the SDM one.

Table 2.7: Estimation of direct, indirect and total effects (six nearest neighbors)

Geographical scale	County	State	Economic Area	Metropolitan Statistical Area
Model type	SDM	SAR	SAR	SDM
Direct effect	0.182*** (0.000)	0.533* (0.056)	0.730*** (0.000)	0.309*** (0.000)
Indirect effect	0.246* (0.205)	0.824 (0.224)	0.300** (0.012)	0.818*** (0.003)
Total effect	0.428** (0.034)	1.357 (0.121)	1.031*** (0.000)	1.127*** (0.000)

Notes: Numbers in the parentheses represent p-values; Significance levels: * for 10%, ** for 5%; *** for 1%.

LeSage and Pace (2009) suggested that the coefficients of spatial models cannot be interpreted directly as marginal effects. Following their suggestion, the direct and indirect effects of industrial diversity on economic instability are estimated (See Table 2.7). The results show that the estimated direct effect in each model is positive, suggesting that industrial diversity contributes to economic stability and is consistent with the theoretical assumption. However, the indirect effect of diversity on instability are different across different geographical scales. To be more specific, for the county-level data, industrial diversity in general has both a positive local effect and a positive total effect. By comparison, when it comes to the state-level analysis, only direct effect is significant. This result is contrary to previous studies on the existence of spatial spillovers in this relationship (Trendle, 2006; Deller and Watson, 2016a,b). Finally, for the functional regions, both direct, indirect and total effects are significant. Particularly, when EAs are used as the analytical units, most of the economic contribution is achieved through local rather than spatial spillovers.

Sensitivity Analysis

Table 2.8: Global Moran’s I results across different spatial weight matrices

Geographical scale	Spatial weight matrix	County	State	Economic Area	Metropolitan Statistical Area
HHI	KNN4	0.103	0.032	0.265	-0.015
	KNN8	0.094	0.015	0.198	0.001
	KNN10	0.087	0.011	0.159	-0.003
	Queen contiguity	0.089	0.080	0.266	0.005
REI	KNN4	0.299	0.496	0.279	0.443
	KNN8	0.278	0.402	0.187	0.416
	KNN10	0.253	0.372	0.177	0.422
	Queen contiguity	0.282	0.483	0.252	0.530

Notes: HHI = Herfindahl Hirschman Index; REI = Regional Economic Instability; All statistics are significant at 1% level; When the queen contiguity spatial weight matrix is used, isolates in weights are removed.

Sensitivity analysis of the results were conducted using queen contiguity and four, eight and ten nearest neighbor ($k = 4, 8, 10$) spatial weight matrices. For brevity, the global Moran’s I and effect estimation results across different spatial weight matrices are shown in Tables 2.8 and 2.9. No matter what method is used to specify the spatial relationship, similar results to those discussed above are arrived at in the analyses of counties, EAs and states—this is in line with LeSage and Pace’s (2014) finding that the choice of a spatial weight matrix is of secondary concern.

Conversely, the MSA-level analysis tells a much different story. Covering only urbanized areas, MSAs in the contiguous U.S. are not always contiguous. As such, some regions have zero neighbors in the queen spatial weight matrix. To illustrate this, Figure 2.5 demonstrates the connectivity map¹⁶ based on the queen contiguity spatial weight matrix for MSAs, where edges between nodes (i.e., MSAs) indicate nodes are neighbors. A total of 51 MSAs have no neighbors and become isolated in this map. Particularly, these MSAs are excluded when calculating the global Moran’s I in GeoDa but are included in spatial regression analysis. By comparison, k-nearest neighbor spatial weight matrices enable each MSA to have the same number of neighbors (See Figure 2.6). To this end, the issue of discontinuity in MSAs results in the difference in model estimation between k nearest neighbor and queen contiguity spatial weight matrices in Tables 2.7 and 2.9. However, other functional regions,

¹⁶Figures 2.5 and 2.6 were made in GeoDa.

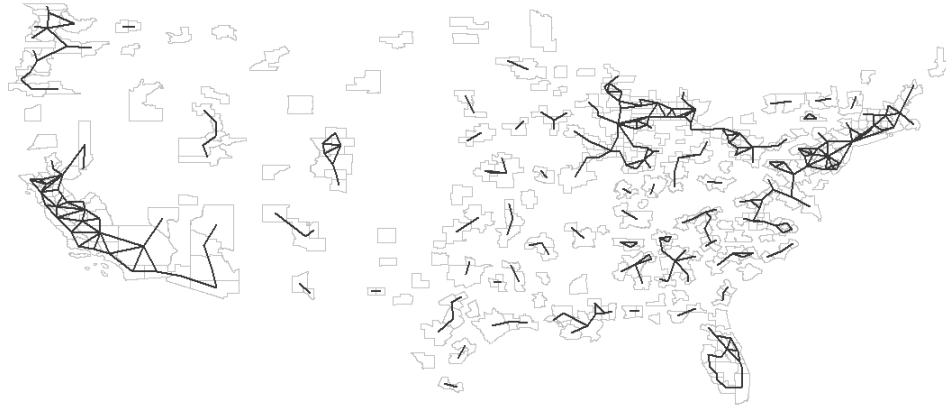


Figure 2.5: Connectivity map of MSAs (queen contiguity)

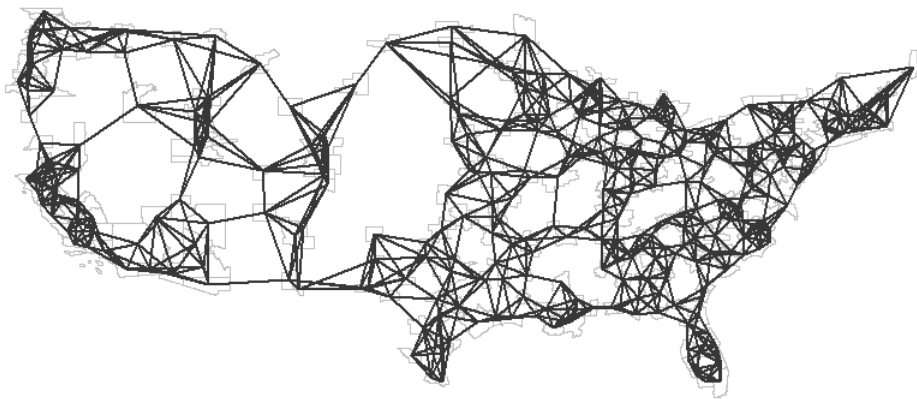


Figure 2.6: Connectivity map of MSAs (six nearest neighbors)

Table 2.9: Estimation of effects across different spatial weight matrices

Spatial Weight Matrix	Estimation	County	State	Economic Area	Metropolitan Statistical Area
KNN4	Model type	SDM	SAR	SAR	SDM
	Direct effect	0.188*** (0.000)	0.584** (0.029)	0.741*** (0.000)	0.317*** (0.000)
	Indirect effect	0.183 (0.217)	0.723 (0.369)	0.265*** (0.008)	0.630*** (0.001)
	Total effect	0.371** (0.017)	1.307 (0.172)	1.007*** (0.000)	0.947*** (0.000)
KNN8	Model type	SDM	SDM	SAR	SDM
	Direct effect	0.182*** (0.000)	0.538** (0.039)	0.715*** (0.000)	0.283*** (0.000)
	Indirect effect	0.340** (0.020)	2.555 (0.172)	0.358** (0.015)	1.228*** (0.000)
	Total effect	0.522*** (0.001)	3.094 (0.118)	1.074*** (0.000)	1.511*** (0.000)
KNN10	Model type	SDM	SDM	SAR	SDM
	Direct effect	0.176*** (0.000)	0.579** (0.042)	0.738*** (0.000)	0.265*** (0.000)
	Indirect effect	0.406*** (0.001)	4.481 (0.544)	0.377** (0.029)	1.128*** (0.000)
	Total effect	0.582*** (0.000)	5.061 (0.502)	1.115*** (0.000)	1.394*** (0.000)
Queen contiguity	Model type	SDM	SAR	SAR	SDM
	Direct effect	0.185*** (0.000)	0.538** (0.049)	0.719*** (0.000)	0.514*** (0.000)
	Indirect effect	0.276* (0.076)	0.660 (0.145)	0.331*** (0.004)	0.354** (0.012)
	Total effect	0.462*** (0.007)	1.198* (0.076)	1.051*** (0.000)	0.869*** (0.000)

Notes: Significance levels: * for 10%, ** for 5%; *** for 1%; Numbers in the parentheses represent p-values.

such as EAs in the U.S., metro areas in Japan (Mizuno et al., 2006) and Local Government Areas in Australia (Trendle, 2006), do cover the whole country, and these regions do not suffer from the discontinuity issue. In that sense, the coverage problem can be different across different studies.

2.5 Discussion

The results of this analysis generally confirm not only that economic diversity is positively associated with employment stability in the U.S. context during the period of 2000-2014 but also that spatial dependence exists in the diversity-stability relationship, no matter what

geographical unit is used. However, the magnitude and significance of the effect of diversity on economic stability vary when different geographical units are used. This variation can possibly impact our understanding of economic structures as well as development policies. In the remainder of this section, some concerns about the choice of spatial units are discussed.

First, to effectively measure regional industry structure, highly disaggregated geographical units are not recommended. Because of the small population problem (Wang and O'Brien, 2005), the data issue of measurement by place-of-residence and place-of-job (Dapena et al., 2018), and the UGCoP (Kwan, 2012), it sometimes can be meaningless to study the economic diversity of a region with a small population and low employment. As illustrated by the ranges and standard deviations in Table 2.3, when county-level data are aggregated, the impact of some extreme values or outliers are excluded. This is known as the scale problem in the MAUP, where aggregated data normally hide much intra-regional heterogeneity. However, in modeling the diversity-stability relationship (See Tables 2.3, 2.5 and 2.7), the county-level model seems not to be better than others. By comparison, aggregated units like EAs and states seem to suffer less from scale-related problems. This result is similar to Resende's (2011) observation that the explanatory power of a municipal model may not be as good as a model that uses states as analytical units when studying economic growth in Brazil. Altogether, data based on highly disaggregated units may not yield intuitive inferences on the effect of economic diversity.

Second, functional regions rather than formal regions are suggested as the analytical units. Although this suggestion has already been advocated by Jackson (1984), Malizia and Ke (1993) and Trendle (2006), several recent studies like Deller and Watson (2016a,b) and Watson and Deller (2017) used formal regions as the analytical units and then spatial econometric techniques to account for spatial spillovers. By comparison, this study uses both formal and functional regions as the basic units and also spatial regression models. As shown in Tables 2.7 and 2.9, the proportion of direct diversity effect is higher for functional regions (EAs) than for formal regions (states), and the diversity-stability relationship in the state-level analysis is not as significant as it is among EAs. In that sense, when formal regions are used as the analytical unit, the diversity-stability relationship might not be fully revealed. From a broader perspective, both EA and state data are spatially aggregated

based on county-level data but with different aggregation schemes. These schemes relate to the zoning problem in the MAUP, and regions aggregated based on functional schemes seems to be preferred to formal schemes.

Third, when functional regions (e.g., MSAs and EAs) are used, potential temporal variations¹⁷ in the boundaries of these regions should also be considered. Due to possible changes in population and commuting flows, the boundaries of MSAs may vary from year to year, and the U.S. Office of Management and Budget regularly updates the boundaries of MSAs. Likewise, the BEA redefined the boundaries of Economic Areas on a regular basis until 2014 when the process was no longer supported because of reductions in government funding. Hence, these outdated EA boundaries might not be used to delimit regional economic systems for recent years. Moreover, Mack et al. (2007) suggested that functional regions are suffered from definitional changes and thus are not suitable for panel and time-series analysis. Attention should be paid to this for future economic diversity research.

Finally, based on all the discussions above, is it possible to choose an appropriate spatial unit for quantifying regional economic structure? It seems that there is not a clear-cut answer to this question, and the spatial scale problem can be study dependent. Particularly, this analysis examines the relationship between industrial diversity and economic stability for the period of 2000-2014 in the contiguous U.S. across different geographical scales. Based on what has been discussed, EAs appear to be good candidates. However, when studying the economic diversity of U.S. urban areas, EAs are no longer the preferred units because they include rural areas. Instead, MSAs are good representatives of the U.S. urban economies. More generally, the choice of spatial scales in delineating regional economic structure varies greatly by study region and period, data availability, and policy relevance among others. In this regard, understanding these scale-related problems and all the potential spatial units of a specific study is an important step in determining the spatial scale(s).

¹⁷These variations also relate to the temporal uncertainty in the UGCoP.

2.6 Chapter Summary

In this chapter, the empirical relationship between industrial diversity and regional economic stability is analyzed at multiple geographical scales in the contiguous U.S. from 2000 to 2014. These scales have been demonstrated here to greatly alter the diversity-stability relationship. Importantly, this analysis confirms the advantages of functional regions over formal regions in modeling economic systems as suggested by Jackson (1984), Malizia and Ke (1993), Trendle (2006) and others. It further stresses the possible changes in the boundaries of functional regions, such as Economic Areas and MSAs (Mack et al., 2007). In addition to these existing understandings, it also suggests not using highly disaggregated geographical units to measure regional economic structure.

Broadly speaking, this analysis introduces several scale-related problems in spatial analysis to the literature of economic diversity, although these problems have not been fully addressed. Other scale-related problems that have not been mentioned here may also impact the empirical understanding of industrial diversity and economic stability. Scale-related problems thus can be study dependent, and the choice of an appropriate geographical unit varies significantly in different research contexts. When quantifying regional economic structures, future researchers and practitioners should have a deep knowledge about all the possible spatial units in the study region and should pay explicit attention to the scale-related problems that are not limited to those mentioned in this analysis.

There are a few potential research directions. While this analysis is based on U.S. regional economies because of data availability, its focus on geographical scale can be applied to other countries as well. For example, extending the study region into developing countries like China might provide additional insights into the role of geographical scale issues both in the analytical and the policy aspects of regional economic development. Moreover, as the disaggregated geographical units or lower level units (counties) are often nested within the upper level units (states and EAs), future research should also consider using spatial hierarchical models, such as Dong et al. (2015) and Lacombe and McIntyre (2017), to study the scale and scope issues in the diversity-stability relationship.

Chapter 3

A Structural Measurement

Motivation

3.1 Introduction

Economic structure is often understood through economic specialization and diversity constructs where economic diversity is defined as, “the variety of economic activity which reflects differences in economic structure at a specific time” Malizia and Ke (1993, p. 222). Most regional economies—such as Detroit and New York—lie on a continuum between pure specialization and complete diversity. In contrast, economic clustering and diversity are assumed to be mutually exclusive (Wagner and Deller, 1998; Wagner, 2000; Deller and Watson, 2016a). This assumption of mutual exclusivity becomes more apparent in the examination of economic diversity measures, such as the Herfindahl Hirschman Index (HHI), where economic diversity is measured as the sum of the squared regional shares of employment for each industry (Wagner, 2000; Trendle, 2006; Chiang, 2009; Hong and Xiao, 2016). In this commonly used metric, higher values for the HHI indicate greater economic specialization or conversely, lower diversity.

However, the dichotomy of economic specialization and diversity has been challenged as regional scientists have reconsidered the definition of economic diversity as the presence of—rather than the absence of—specialization to stress that regional economic systems

can be specialized and diversified simultaneously (See Malizia and Ke, 1993; Wagner and Deller, 1998; Dissart, 2003; Desrochers and Sautet, 2008; Jackson, 2015; Hong and Xiao, 2016). Nevertheless, with the exception of Hong and Xiao (2016), these authors have only established the basic conceptual framework of diversified specializations but have not applied this framework to empirical studies. Specifically, Hong and Xiao (2016) proposed a Multiple Specialization Index (MSI) that allows for the measurement of multiple specializations in economic activities. In Hong and Xiao's study, the MSI is calculated as the ratio of the number of specialized industries to the number of non-zero employment industries in the region. In addition to those, specialized industries have a location quotient (LQ) value greater than a specified cut-off value. However, as suggested by Porter (2003), Spencer et al. (2010) and Delgado et al. (2016), not all industries are specialized. For example, drug stores and elementary schools only serve a local market and should not be considered as candidates for economic specializations.

To address this issue, this chapter develops an alternative measure of economic structure that enables the interpretation of diversified specializations for empirical analysis. As economic specialization can contribute to regional economic growth, and a diverse regional economy is theoretically associated with stable regional economic growth, this measure can leverage the benefits of specialization and diversity concurrently. This chapter also provides an overview of the relationship between economic specialization and diversity within current economic structure measures, including the HHI and Hong and Xiao's (2016) MSI. In addition, policy implications are offered when economic development strategies shift from pure specialization and complete diversification to developing diversified specializations.

Accordingly, Section 3.2 provides a theoretical background on economic specialization and diversity. Methodology is described in Section 3.3, followed by the results and discussion. The final section concludes with the findings of this chapter.

3.2 Background

3.2.1 Specialization and Diversity: Two Sides of the Same Coin?

When it comes to the relationship between specialization and diversity, the traditional view holds that diversity can be interpreted as the absence of specialization; in other words, specialization and diversity are two sides of the same coin. In relation to regional economic performance, regions have to choose between a stable yet slow growth and a high yet risky growth. The traditional view, however, has been challenged by Malizia and Ke (1993), Wagner and Deller (1998), and others. Malizia and Ke (1993) were among the first researchers to consider their coexistence. In their words, “specializations can be the source of competitiveness as well as compensate for one another when business cycles or external shocks occur” (Malizia and Ke, 1993, p. 223). Despite their conceptual advances, Malizia and Ke (1993) still acknowledged that a trade-off exists between growth and instability. In contrast, Wagner and Deller (1998) suggested that short-term economic development strategies can focus on specialization to promote growth, while long-term policy can aim at stability through diversification, indicating that the trade-off no longer exists.

However, the concept of diversified specializations has not been widely applied to measures of economic structure—such as the national average, the ogive, the entropy and the Herfindahl-Hirschman Index (HHI)—for empirical analysis¹. Among these measures, the HHI and the entropy measures have been used more extensively than others in empirical studies. This preference is not only because of computational ease and limited data requirements, as suggested by Wagner (2000), Trendle (2006) and Jackson (2015), but also because the traditional view of the specialization-diversity dichotomy is embedded in these two measures. Specifically, the HHI ranges from $1/N$ for a perfectly diversified economy to 1 if all employment is concentrated in one industry. Similarly, the entropy index reaches its maximum for a one-industry economy, whereas for its minimum, all employment is evenly distributed across sectors. By comparison, other structure measures, such as durable goods percentage and location quotient, decouple economic specialization and diversity; that is to

¹For comparative reviews of these structure measures, see Jackson (1984, 2015), Wagner (2000), and Dissart (2003).

say, unlike the entropy and HHI measures, one cannot simultaneously determine the levels of economic specialization and diversity.

As noted earlier, the only metric that addresses the coexistence of specialization and diversity is Hong and Xiao’s (2016) Multiple Specialization Index, which is calculated as:

$$MSI_i = \frac{\sum_{n=1}^N SP_n}{N} \quad (3.1)$$

where N is the number of sectors in region i ; n is the sector index; and SP_n equals 1 if the location quotient of sector n is greater than the cut-off value; otherwise, it equals 0. The cut-off value was set as the 80th percentile LQ values for each three-digit NAICS (North American Industry Classification Systems) sector, and sectors with large LQs were identified as specializations. The MSI approaches a value of 1 for highly specialized economies and 0 if no sector is specialized. In essence, this index measures the number of specialized industries in a regional economy but is divorced conceptually from any notions of competitiveness, co-location, or interindustry linkages as suggested by Porter (1990, 1998). As such, the remainder of this section visits criteria to identify economic clusters for empirical analysis.

3.2.2 Identifying Economic Clusters

In Hong and Xiao’s (2016) study, the MSI considers industries with non-zero employment as potential economic specializations, whereas, in reality, it is not meaningful to treat sectors that only serve local demand as candidates for economic clusters. As such, the definition of economic clusters is open to discussion. Originally, Porter (1998, p. 226) defined economic clusters as, “a form of network that occurs within a geographical location, in which the proximity of firms and institutions ensures certain forms of commonality and increases the frequency and impact of interactions.” This definition might be useful for case studies of economic clusters like Silicon Valley and Route 126 outside of Boston, but it appears to provide little operational guidance on measuring regional economic clusters for empirical analysis because of its ambiguity (Martin and Sunley, 2003; Yu and Jackson, 2011). As such, a more formal and operational definition of economic clusters is required.

Developing such a definition, of course, is a common step in quantitative analysis on

economic clusters. For example, Porter (2003) classified all industries into three categories:

- *traded industries* that sell goods and services across regions and to other countries;
- *resource-based industries* that are located where the needed resources are found; and
- *local industries* that are industries present in most areas and sell locally.

In this classification, Porter (2003) argued that only traded industries can be economic clusters, although Spencer et al. (2010) suggested that resource-based industries can also be economically specialized. Methodologically, Porter (2003) identified economic clusters based on the geographical correlation of employment between traded industries. In Porter’s (2003, p. 562) words, “if computer hardware employment is nearly always associated geographically with software employment, this provides a strong indication of locational linkages.” The result of Porter’s (2003) method is 29 traded clusters identified in the U.S. More recently, Delgado et al. (2016) extended Porter’s (2003) method by considering co-location patterns of employment and establishments, input-output linkages, and similarity in labor occupation. As a result, they identified 51 traded clusters in the U.S. context, and each cluster is composed of several six-digit NAICS sectors. Additionally, based on Spencer et al. (2010), resource-based industries like coal mining can also be economic clusters in Delgado et al.’s (2016) result.

3.3 Methodology

3.3.1 Empirical Frameworks

To study the effects of economic structure on short-term employment growth and long-term stability in the contiguous U.S., MSAs are used as the basic analytical units because they meaningfully constitute functional economic systems (Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006). The following empirical models are used:

$$REI_i = f(DIV_i, CONTROL_i) \tag{3.2}$$

$$GROWTH_i = g(DIV_i, CONTROL_i) \tag{3.3}$$

where the dependent variables are economic performance indicators, including regional economic instability index (REI) and employment growth rate; and the independent variables are diversity measures and a set of control variables. In both Equations 3.2 and 3.3, the independent variables reveal the state of the regional economy in 2000, while the dependent variables capture short- (2000-2002) and long-term (2000-2014) changes in regional economic performance. All the variables used in this analysis are summarized in Table 3.1.

Table 3.1: Description of all the variables

Category	Symbol	Description	Data source
Dependent variable	REI	Economic instability, 2000-2014	BEA
	GROWTH	Employment growth rate, 2000-2002	BEA
Independent variable	HHIS	HHI of 3-digit NAICS sectors, 2000	CBP
	HHIC	HHI of specialized clusters, 2000	CBP
	MHHI	MHHI=(1+HHIS)×(1+HHIC)	CBP
Control variables	POP	Logged population size, 2000	Census
	EDU	Percentage of the population over 25 with at least a bachelor degree, 2000	Census
	NONWHITE	Percentage of the population non-white, 2000	Census
	INCOME	Per capita income relative to U.S. average, 2000	BEA
	GOODS	Percentage of employment in goods production industries (minus agriculture), 2000	BLS, CEW

Notes: CBP = County Business Patterns; BEA = Bureau of Economic Analysis; BLS = Bureau of Labor Statistics; NAICS = North American Industry Classification Systems; CEW = Census of Employment and Wages; HHI = Herfindahl Hirschman Index

3.3.2 Economic Performance Indicators

Previous research assesses economic performance through many measures like growth rates of per capita income (Wagner and Deller, 1998; Izraeli and Murphy, 2003), unemployment rate (Mizuno et al., 2006; Chiang, 2009), and employment instability index (Conroy, 1975; Kort, 1981; Jackson, 1984). Rather than adopt these measures, however, this analysis uses short-term employment growth and long-term regional economic instability as indicators of economic performance. This is not only because of their popularity in previous literature (Kort, 1981; Malizia and Ke, 1993; Trendle, 2006) and economic development policy discussion (Wagner and Deller, 1998; Deller and Watson, 2016a), but also because they leverage the benefits of economic specialization and diversity at the same time.

Based on Malizia and Ke (1993), regional economic instability (REI) is calculated as:

$$REI_i = \left\{ \sum_{i=1}^N [(E_{it} - E_{it}^{Tr}) \times 100/E_{it}^{Tr}]^2 / T \right\}^{1/2} \quad (3.4)$$

where i denotes the region index; E_{it} is the actual number of workers for region i at time t ; T is the number of observed time spans; and E_{it}^{Tr} is the predicted number of workers for region i at time t using a linear trend line. REI is measured using employment data from the Bureau of Economic Analysis (BEA) during the long-run study period from 2000 to 2014.

Employment growth for region i at time t is calculated as the percent rate of increase in total employment from $t - 1$ to t .

$$GROWTH_i = \frac{(E_{it} - E_{it-1}) \times 100}{E_{it-1}} \quad (3.5)$$

where E_{it} is the actual number of workers for region i at time t and E_{it-1} is the number of workers at time $t - 1$. Like REI, the growth variable is also calculated based on the BEA data sets but for the short term studying period from 2000 to 2002.

3.3.3 Measuring Economic Diversity

Similar to Jackson (1984), Malizia and Ke (1993) and Hong and Xiao (2016), the diversity variable is measured based on data from County Business Patterns (CBP). CBP is published annually by the U.S. Census Bureau and contains employment by the two to six-digit NAICS sectors for different levels of geographical regions like states, counties and zip-code areas. For confidentiality reasons, the U.S. Census Bureau uses data ranges for the number of employment for some sectors. Values to replace these ranges, however, were estimated in the Upjohn Institute’s “WholeData” version of CBP, derived using Isserman and Westervelt’s (2006) method. The complete data of 2000 were accessed and used to assess regional economic diversity.

In this analysis, three HHI-based diversity measures of economic activities are used. The first measure quantifies the level of employment dispersion between 87 three-digit NAICS

sectors, and this is the traditional measure of economic diversity used in Chiang (2009), Mizuno et al. (2006) and others. The HHI of sectors (HHIS) are defined as:

$$HHIS_i = \sum_{j=1}^N (e_{ij}/E_i)^2 \quad (3.6)$$

where e_{ij} is the employment of industry j in region i ; E_i is the total number of people employed in the i^{th} region; and N stands for the number of industries. In Equation 3.6, the value of HHIS is between $1/N$ and 1. Smaller values of the index suggest greater dispersion or diversity in economic activities (i.e., employment).

To interpret economic diversity as the presence of multiple specializations, the second measure of economic diversity, based on Delgado et al.'s (2016) cluster template², revises the commonly used HHI in Equation 3.6 as follows³:

$$HHIC_i = \sum_{j=1}^M (e_{ij}/E_i)^2 \quad (3.7)$$

where e_{ij} denotes the employment of cluster j in MSA i ; E_i is the total employment of traded industries; and M is the number of clusters in the same region. Similarly, the HHI of clusters (HHIC) ranges from $1/M$ to 1, and smaller values of HHIC suggest a greater level of diversity within clusters. Unlike the first diversity measure (i.e., HHIS), and Hong and Xiao's (2016) MSI, this measure of economic diversity is calculated using the cluster definition of Delgado et al. (2016) and also excludes the impact of local industries. In addition, these clusters also should be concentrated relative to the nation; namely, the LQ of these clusters are greater than 1⁴. The LQ for cluster j in region i is defined as:

$$LQ_{ij} = \frac{e_{ij}/E_i}{e_{Nj}/E_N} \quad (3.8)$$

where e_{ij} stands for employment for cluster j in region i ; E_i is the total employment of

²Purdue Center for Regional Development (2007) also developed a cluster template for regional analysis, but the emphasis of this template is mainly rural regions; for example, compared to Porter (2003) and Delgado et al. (2016), jewelry clusters were excluded.

³Slaper et al. (2018) also developed an empirical measure of cluster diversity but did not emphasize the coexistence of specialization and diversity.

⁴The cut-off value of 1 has been criticized by numerous scholars (e.g., Ellison and Glaeser, 1997; Carroll et al., 2008; Tian, 2013). Future research might consider other cut-off values or the revised versions of LQ.

traded industries in region i ; e_{Nj} represents the national total employment of cluster j ; and E_N is the total employment of traded industries for the whole study region. For a given cluster, an LQ greater than 1 indicates that the region has a higher concentration than the nation.

The mean of HHIS for U.S. MSAs is 0.049 (range, 0.030 to 0.179), while the mean of HHIC is 0.140 (range, 0.043 to 0.680). Figures 3.1 and 3.2 present the geographical distributions of industry and cluster diversity, where MSAs with greater than the average diversity values are differentiated from those with less than average diversity values. Comparing these two figures, 182 of the 359 MSAs have high levels of both industry and cluster diversity, while 92 MSAs have low levels of industry and cluster diversity. In contrast, 49 MSAs are industrially diversified with limited cluster diversity, whereas 36 MSAs have strong cluster diversity but with low levels of industrial diversity. To further assess the relationship between industry and cluster diversity, Figure 3.3 displays the scatterplot of the HHIC and the HHIS with the fitted linear trend line. As reflected by the R^2 value, the industry diversity measure only explains 36.7% of variation in the diversity index of clusters. This weak correlation is understandable, as a higher degree of industry diversity might not necessarily display greater diversity within clusters.

Third, to consider the joint effect of both HHIC and HHIS, a multiplicative HHI (MHHI) is defined as:

$$MHHI_i = (1 + HHIC_i) \times (1 + HHIS_i) \quad (3.9)$$

where HHIS measures the economic diversity of sectors; and HHIC quantifies the diversity of economic clusters for a given region. Ideally, if the base economy is most diversified at both sector and cluster levels, then MHHI reaches its minimum; conversely, if the economy has a single industry that forms only one cluster, the value of MHHI is 4.

3.3.4 Modeling Methods

As suggested and confirmed by Trendle (2006), Deller and Watson (2016a), and others, spatial dependence does exist in the way in which economic structure itself impacts regional economic performance, whereas the ordinary least squares (OLS) model would ignore this

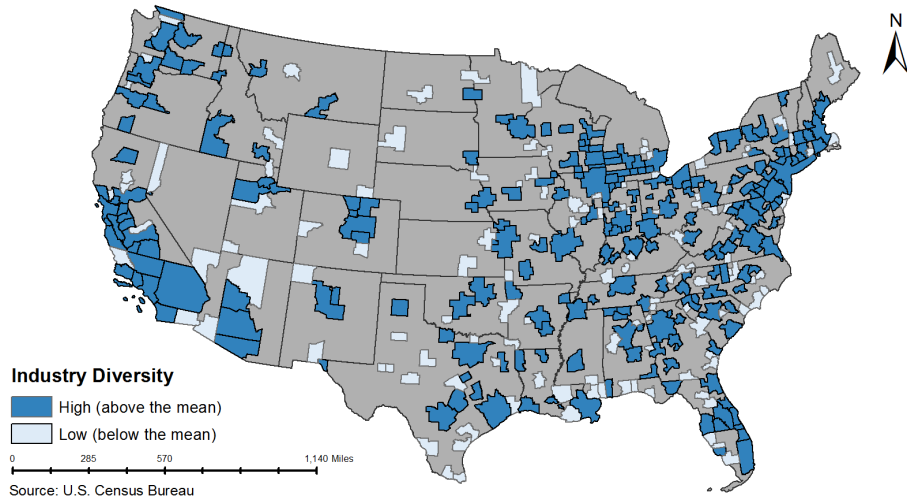


Figure 3.1: Industry diversity in the lower U.S.

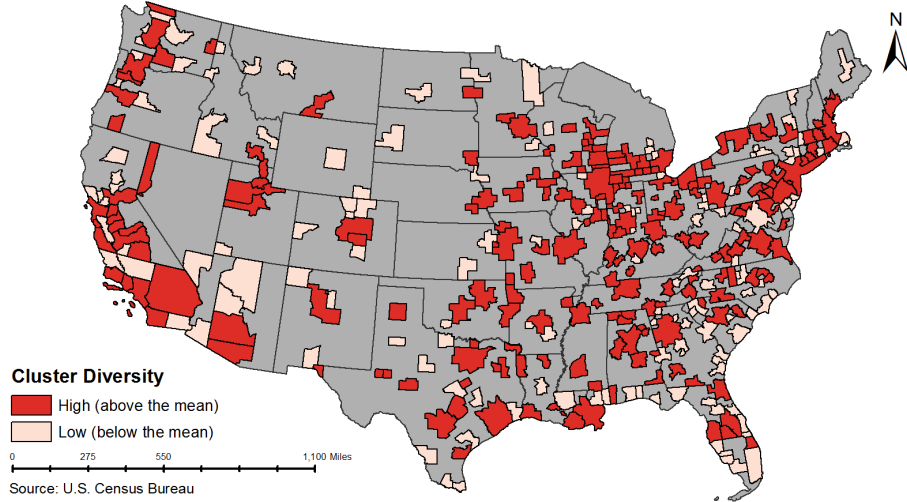


Figure 3.2: Cluster diversity in the lower U.S.

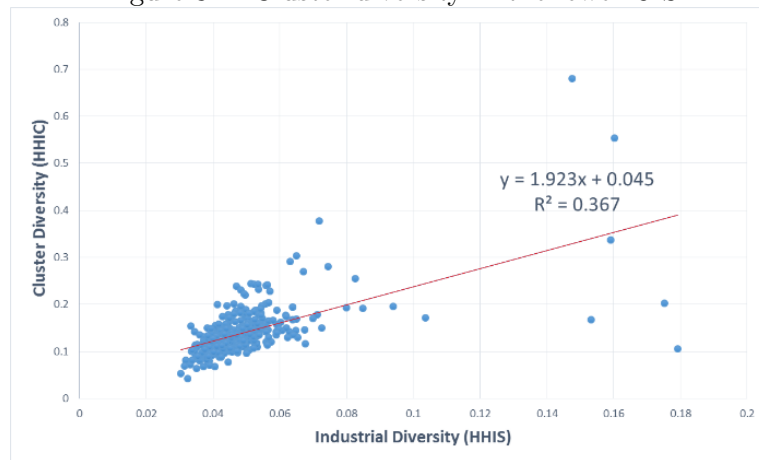


Figure 3.3: Correlation between industry and cluster diversity.

dependence and thus result in inaccurate estimates. The spatial Durbin model (SDM) is such an approach that incorporates spatial dependence in both dependent and independent variables when compared to the spatial autoregressive model and the spatial error model. Moreover, LeSage and Pace (2009) suggested that the SDM should be used when one believes that there might exist omitted variables that demonstrate spatial autocorrelation.

The SDM can be described as:

$$y = \rho W y + X\beta + WX\gamma + \varepsilon \quad (3.10)$$

where y is the dependent variable for region i ($i=1, \dots, N$); X is a matrix of independent variables; β is a vector of estimated coefficients of the independent variable; ρ is a coefficient that describes the strength of the spatial autocorrelation in the dependent variable; γ^5 is a vector of estimated coefficients of the spatially lagged, independent variables WX ; and ε is the error term. The term W denotes the spatial weight matrix and reflects the geographic relationship that can be specified using various methods like distance- and contiguity-based ones. In this analysis, queen contiguity is used, where two regions i and j are neighbors if they share at least one point or side on their boundaries. The corresponding element in the spatial weight matrix W_{ij} is 1, and W_{ij} equals 0 otherwise.

Bayesian spatial econometric techniques rather than maximum likelihood methods are used. LeSage and Pace (2009, p. 150) indicated that Bayesian spatial econometric techniques outperform maximum likelihood methods as “in small samples parameters may exhibit asymmetry or heavy tailed distributions that deviate from normality.” For this reason, Equation 3.10 is estimated using the Bayesian Markov Chain Monte Carlo (MCMC) method. Generally, let y denote the whole data and θ represent a vector of parameters of interest. The posterior distribution of the parameters, $\pi(\theta | y)$, is expressed as:

$$\pi(\theta | y) = \frac{\pi(y | \theta)\pi(\theta)}{\pi(y)} \quad (3.11)$$

where $\pi(\theta)$ is the prior probability density function for θ ; $\pi(y | \theta)$ is the density function for

⁵To differentiate in Equation 3.11, γ is used here.

y when the parameter value is θ ; and $\pi(y)$ is a constant term and normalizes the posterior distribution. Because $\pi(y)$ is free from the parameter vector θ , Equation 3.11 is summarized as the Bayesian phrase, “the posterior is proportional to the likelihood times the prior” and can also be rewritten as:

$$\pi(\theta | y) \propto \pi(y | \theta)\pi(\theta) \quad (3.12)$$

Markov chain Monte Carlo (MCMC) methods are used to draw inferences regarding the parameters. Specifically, the Gibbs sampling procedure is used to generate samples of θ^6 and ρ , whereas the Metropolis-Hasting algorithm is used to generate ρ^7 . The prior distribution of the β parameter is a multivariate normal distribution with a mean of zero and a covariance of $10,000I_k$: The prior values for σ come from an inverse gamma distribution with both shape and scale parameter as 0. The prior values for the ρ parameter come from a univariate normal distribution with a mean of 0 and a standard deviation of 10,000. Each model was run for 56,000 iterations with the initial 6,000 discarded as burn-in iterations. The removal of these iterations is useful because the initialized values of the parameters might be unstable.

3.4 Empirical Results

Instability Models

Table 3.2 presents the estimation results of three instability models that use three different diversity measures (HHIS, HHIC and MHHI). As the coefficient of the spatially lagged dependent variable, ρ , is statistically significant in each model, the economic stability of an MSA can be impacted by the instability of neighboring MSAs. Unlike the OLS regressions, the estimated β s in Equation 3.10 cannot be directly interpreted as marginal effects because of potential spatial dependence in the variables (LeSage and Pace, 2009). Instead, following LeSage and Pace (2009), the direct, indirect, and total effects were estimated. Note that the direct and indirect effects here can possibly move in opposite directions.

⁶Strictly speaking, the β term here stands for the coefficients of the independent variables as well as the spatially lagged independent variables.

⁷For more information about these algorithms, see Lacombe (2008) and LeSage and Pace (2009).

Table 3.2: Effect estimates of instability models

(a) Model 1	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
HHIS	9.605***	3.088	5.191	0.962	14.796**	1.987
POP	0.101***	2.765	0.097	1.649	0.198***	2.629
EDU	-0.021**	-2.071	-0.017	-1.009	-0.038*	-1.696
NONWHITE	-0.006	-1.098	0.012	1.521	0.005	0.615
INCOME	0.009**	2.063	-0.006	-0.850	0.003	0.372
GOODS	0.014*	1.878	-0.037***	-3.076	-0.023	-1.570
ρ					0.402***	9.643
(b) Model 2	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
HHIC	5.197***	5.749	4.500***	2.995	9.697***	4.949
POP	0.085**	2.424	0.070	1.322	0.155**	2.352
EDU	-0.021**	-2.187	-0.014	-0.966	-0.036*	-1.783
NONWHITE	-0.006	-1.001	0.010	1.297	0.004	0.475
INCOME	0.009**	2.031	-0.006	-0.935	0.003	0.383
GOODS	0.012	1.615	-0.037***	-3.365	-0.025*	-1.889
ρ					0.351***	8.085
(c) Model 3	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
MHHI	2.128***	5.175	1.438**	2.183	3.566***	3.926
POP	-0.063	-1.224	-0.018	-0.236	-0.081	-0.759
EDU	-0.024**	-2.442	-0.021	-1.317	-0.045**	-2.130
NONWHITE	-0.003	-0.604	0.010	1.334	0.007	0.794
INCOME	0.011**	2.459	-0.004	-0.572	0.007	0.850
GOODS	0.006	0.832	-0.039***	-3.271	-0.033**	-2.283
ρ					0.373***	8.920

Notes: HHIS=Herfindahl Hirschman Index of Sectors; HHIC=Herfindahl Hirschman Index of Clusters; MHHI=Multiplicative Herfindahl Hirschman Index; Significance levels: * for 10%, ** for 5%; *** for 1%.

Focusing on the control variables, the population size in Models 1 and 2 is found to contribute to regional economic instability as reflected in the estimated direct and total effects. An MSA with a larger population appears to increase the economic instability of the MSA and its neighbors. By comparison, the population size is statistically insignificant in Model 3. Meanwhile, the estimated direct and total effects of the education attainment variable are negative and significant in Models 1-3, indicating that a better educated population in an MSA seems to reduce its economic instability. Conversely, the indirect effect estimate is insignificant. Furthermore, the nonwhite variable seems to be insignificant and the direct of the income variable is significant in Models 1-3. Finally, the effect of employment in goods-producing sectors on economic stability is mixed when different diversity measures are used.

Turning to the diversity variable, the estimated direct and total effects of economic diversity are positive and statistically significant in Model 1. This result is in line with the portfolio theory that a region's economic diversity contributes to its stability in economic activities. Conversely, the positive indirect effect is not significant as reflected in the *t* statistics. This, however, is in conflict with previous studies on the spatial spillover effects of economic diversity as suggested by Trendle (2006) and Deller and Watson (2016a). In Model 2, the estimated direct, indirect and total effects of economic diversity are significant. Similarly, the positive and significant direct, indirect and total effects in Model 3 suggest that the economic instability of an MSA can be influenced by its own industry mix as well as its neighbors' economic structures. Overall, after controlling for industrial, demographic and economic variables, diversity is still positively associated with economic stability, although different measures of diversity can alter the impact and statistical significance of diversity.

Growth Models

Similar to the case of instability models, the direct, indirect, and total effects of economic diversity on employment growth were estimated in Models 4-6 in Table 3.3. The spatially lagged dependent variable is statistically significant in each model, indicating that MSAs with high employment growth can encourage the employment growth of their neighbors.

In Models 4-6, the estimated direct effects of all the control variables (population size, education attainment, nonwhite, income and goods) are significant, although there are variations in the statistical significance of the indirect and total effects. Population size is positively and significantly associated with employment growth in terms of direct and total effects, indicating that population size contributes to employment growth in MSAs during the study period from 2000 to 2002. As the population variable was logged, these estimated effects can be directly interpreted as elasticity. For example, in Model 4, an increase of 10 percent in population size would have a direct effect of a six percent employment increase and the total effect of a five percent increase in employment growth, *ceteris paribus*. Similarly, the MSAs with a higher education attainment usually have greater employment growth. Moreover, counteracting effects occur in the nonwhite variable and the income variable. The direct effect of these two variables seems to reduce job creation, while their

Table 3.3: Effect estimates of growth models

(a) Model 4	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
HHIS	7.117	0.811	-21.411	-1.504	-14.294	-0.742
POP	0.616***	5.646	-0.099	-0.623	0.516***	2.596
EDU	0.106***	3.527	-0.051	-1.144	0.055	0.932
NONWHITE	-0.072***	-4.039	0.077***	3.470	0.005	0.210
INCOME	-0.069***	-5.202	0.039**	2.068	-0.030	-1.324
GOODS	-0.151***	-6.645	-0.062**	-1.914	-0.213***	-5.625
ρ					0.301***	15.407
(b) Model 5	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
HHIC	6.361**	2.309	-3.512	-0.807	2.849	0.503
POP	0.581***	5.481	-0.153	-0.980	0.428**	2.263
EDU	0.105***	3.567	-0.053	-1.218	0.052	0.892
NONWHITE	-0.070***	-4.007	0.074***	3.407	0.004	0.180
INCOME	-0.069***	-5.311	0.042**	2.284	-0.027	-1.193
GOODS	-0.157***	-6.954	-0.066**	-2.096	-0.223***	-5.891
ρ					0.306***	14.503
(c) Model 6	Direct effect	t-statistic	Indirect effect	t-statistic	Total effect	t-statistic
MHHI	3.296***	2.689	-2.344	-1.271	0.952	0.376
POP	0.343**	2.240	-0.003	-0.013	0.340	1.133
EDU	0.106***	3.564	-0.051	-1.137	0.055	0.910
NONWHITE	-0.065***	-3.674	0.074***	3.336	0.008	0.345
INCOME	-0.068***	-5.144	0.042**	2.189	-0.026	-1.113
GOODS	-0.169***	-7.397	-0.048	-1.428	-0.218***	-5.593
ρ					0.319***	13.154

Notes: HHIS=Herfindahl Hirschman Index of Sectors; HHIC=Herfindahl Hirschman Index of Clusters; MHHI=Multiplicative Herfindahl Hirschman Index; Significance levels: * for 10%, ** for 5%; *** for 1%.

indirect or spatial spillover effects appear to stimulate employment growth. Finally, employment in goods-producing sectors is likely to reduce employment growth as demonstrated by the negative and significant direct and total effects in all models.

Focusing on the diversity variable, in Models 5 and 6, the direct effect seems to contribute to short-term employment growth and is in line with conventional wisdom, the MAR externalities, as well as the theories of Porter (1990, 1998) that specialization can promote economic growth. In comparison, the indirect or spatial spillover effect in Models 4-6 is negative but insignificant. Finally, although the total effect is insignificant in Models 4-6, there appears to be a trade-off between direct and indirect effects as these two effects move in the opposite direction.

3.5 Discussion

In this analysis, the effects of economic structure on economic stability and growth are studied using three HHI-based economic structure measures. The first two measures quantify the level of employment dispersion among (1) three-digit NAICS sectors (HHIS) and (2) specialized clusters (HHIC). To consider the interplay of these two measures, the MHHI is also used. Based on the empirical results, there are several interesting points for discussion.

First, both industry and cluster diversity seem to contribute to economic stability. In Models 1-2, the estimated direct and total effects of the diversity variable (HHIS or HHIC) is positive and significant. When the diversity of both sectors and clusters are considered in Model 3, the estimated direct, indirect and total effects of economic diversity on economic stability are positive and significant. As such, this analysis confirms that not only industrial diversity but also the diversity of specialized clusters are positively associated with economic stability. In short, for a given region, both industry and cluster portfolios can contribute to economic stability.

Second, unlike the instability models, the effect of industrial diversity on economic growth differs from that effect of cluster diversity. In Model 4, the effect of industrial diversity is statistically insignificant and in line with Hong and Xiao (2016), suggesting that the overall industry diversity hardly impacts employment growth. By comparison, as demonstrated in the estimated direct effect in Model 5, the diversity of clusters is positively and significantly associated with employment growth. Perhaps, this result can be explained by the MAR externalities and theories of Porter (1990, 1998) rather than Jacobs' (1969) externalities. Similarly, the direct effect of the diversity variable in Model 6 is positive and significant.

Third, as reflected in both instability and growth models, measuring the economic diversity of specialized clusters and industrial diversity simultaneously contributes to revealing more effects more revealed effects of diversity on economic stability and growth than measuring industrial diversity alone. Many authors (Conroy, 1975; Kort, 1981; Siegel et al., 1995; Wagner and Deller, 1998) have considered inappropriate measures of economic diversity as one of the factors that contribute to the inconsistency between theoretical assumption and

empirical evidence of the relationship between economic structure and regional economic performance. Interpreting economic diversity as both industry and cluster diversity could improve existing economic diversity measures. Although only HHI is used here, these improvements can also be applied to other diversity measures, such as the national average and the entropy index.

Finally, of special relevance here is research on related and unrelated variety⁸ in economic geography, such as Frenken et al. (2007) and Boschma et al. (2011), where the entropy index is used to measure different types of variety. Specifically, based on the Standard Industrial Classification (SIC) scheme⁹, Frenken et al. (2007) measured unrelated variety as the average employment entropy across two-digit industries and total variety as the average employment entropy between five-digit industries. The corresponding related variety equals the difference between the total entropy and the unrelated entropy. Similarly, Boschma et al. (2011) indicated that related and unrelated variety can also be defined based on Porter's (2003) definition of clusters; namely, unrelated variety can be measured among clusters and total variety can be calculated among sectors. As such, the unrelated variety measure of Boschma et al. (2011) displays technical similarities with the cluster diversity measure (HHIC). However, these studies neither explicitly emphasize that economic clusters should be specialized relative to the nation ($LQ \geq 1$), nor aimed at interpreting economic diversity as the presence of multiple specializations, which is the course pursued in this chapter.

3.6 Chapter Summary

This chapter develops a measure to interpret the diversity of economic specializations and emphasizes the coexistence of economic diversity and specialization. It also studies the effects of industry and cluster diversity on regional economic performance. The empirical evidence confirms that both industry and cluster diversity can contribute to economic stability, yet also suggests that only cluster diversity promotes employment growth. Together with Hong and Xiao (2016), this analysis empirically confirms that one region can simulta-

⁸ For a comprehensive review of recent research on related variety, see Content and Frenken (2016).

⁹ In the U.S., the SIC has been replaced by the North American Industry Classification Systems (NAICS) since 1998.

neously pursue economic growth and stability by promoting diversified specializations.

Table 3.4: A typology of specialization and diversification in terms of sectors and clusters

	Specialize sector(s)	Diversify sector(s)
Specialize cluster(s)	Type 1: Reinforcement	Type 2: Supplement
Diversify cluster(s)	Type 3: Replication	Type 4: Transplantation

The findings of this analysis also remind economic development researchers and practitioners to consider the underlying relationship between targeted sectors and their linked clusters in industrial recruitment activities. Based on Martin and Sunley (2006) and Boschma et al. (2017), Table 3.4 enumerates four possible types of economic structure-based development policies that consider this relationship. Reinforcement (Type 1) represents pure specialization at both sector and cluster levels. After adopting economic development policies of this type, regions would experience faster economic growth in the short term. Yet, it is not recommended to specialize only in these clusters in the long run, because a limited number of specializations might be severely impacted by external economic downturns, or “do not place a region’s employment eggs in one industry or cluster basket.” By comparison, Supplement (Type 2) illustrates a case that the targeted sectors improve the diversity of sectors while reducing the diversity of clusters. Development policies of this type are common nowadays. For example, Jackson’s (2015) clusters and diversification strategy (CADS) can be used to identify sectors that fail to support existing economic clusters in terms of supply deficits. Economic development policies focusing on these sectors can promote sector-level economic diversity while decreasing cluster-level diversity; in other words, existing economic clusters would be supplemented in terms of supply. Replication (Type 3) demonstrates a situation where the targeted sectors enhance the economic diversity of clusters while reducing the diversity of sectors. Frenken et al. (2007) indicated to import sectors that are closely related to existing economic structure to be potential clusters. Transplantation (Type 4) indicates complete diversification in the lens of both sectors and clusters. This diversification process can be referred to as importing popular or advanced sectors without fully considering existing regional economic structures when comparing it to Type

3 development policies. Although theoretically feasible in industry targeting and recruiting procedures, it is not suggested as “one should take existing regional competences as building blocks to broaden the economic base of the region” (Frenken et al., 2007, p. 696).

There are several potential directions for future research. The relationship between cluster and industry diversity should be studied. For example, which one has the priority in regional economic development? Although Desrochers and Sautet (2008) and Hong and Xiao (2016) suggested that overly specialized economies should enhance industry diversity first and then promote specializations as a diversified economy is the prerequisite for the emergence of diverse specializations, this preference has not been examined empirically. Apart from the HHIC—a cross-fertilization of the cluster template of Delgado et al. (2016), location quotient, and the Herfindahl Hirschman Index—developed in this chapter, it is also interesting to develop other measures to meaningfully quantify cluster diversity (e.g., Slaper et al., 2018). In addition, based on the identified typology of structure-based development strategies in Table 3.4, future research could further develop this typology in such analytical dimensions as key actors, industry targeting methods and risks. With a deeper understanding of economic structures, both specialization and diversity can better benefit regional economic development.

Chapter 4

A Modeling Method Concern

4.1 Introduction

Regional scientists, economic geographers and planners have studied the relationship between economic diversity and regional economic performance for many decades, both theoretically and empirically (e.g., Conroy, 1975; Kort, 1981; Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006; Hong and Xiao, 2016). Within the literature of regional economic diversity, the portfolio theory hypothesizes that diversified economies usually display greater stability in their economic performance and less volatility from external downturns (Conroy, 1975). On the other hand, conventional wisdom and previous theories—such as the MAR externalities (Glaeser et al., 1992) and Porter’s (1990; 1998) economic clusters—hold that economic specialization can promote economic growth. In contrast, Jacobs (1969) has argued that it is diversity that contributes to growth.

Many empirical studies have tested these theoretical assumptions, while the results are often mixed (Jackson, 1984; Attaran, 1986; Mizuno et al., 2006). Malizia and Ke (1993), Wagner and Deller (1998) and others stated that the primary causes of this empirical inconsistency include (1) the use of inappropriate use of geographical units, (2) poorly defined measures of economic diversity, and (3) overly simplistic modeling methods. Specifically, numerous geographical units have been used to quantify regional economic structures, such as counties (Deller and Watson, 2016a; Watson and Deller, 2017), states (Attaran, 1986; Wagner and Deller, 1998), Metropolitan Statistical Areas or MSAs in the U.S. (Malizia and

Ke, 1993; Hong and Xiao, 2016), and Local Government Areas in Queensland, Australia (Trendle, 2006). However, Jackson (1984) and Malizia and Ke (1993) suggested that only functional economic regions (e.g., MSAs) should be used to define these economic structures because of such issues as the Modifiable Areal Unit Problem (Openshaw and Taylor, 1979; Openshaw, 1984) and the uncertain geographical context problem (Kwan, 2012). Meanwhile, numerous studies have improved existing measures of economic diversity, including input-output based measures in Wagner and Deller (1998) and Siegel et al. (1995) and metrics that consider the diversity of economic specializations and industries (e.g., Hong and Xiao, 2016).

Another cause of the inconsistency between theoretical assumptions and empirical evidence is the use of simplistic statistical techniques. Although the modeling methods in the literature of regional economic structure have advanced greatly from bivariate statistics through multivariate regression to spatial regression techniques, there is little research on model uncertainty, especially in a spatial context. LeSage and Pace (2009) suggested that model uncertainty can result from at least two sources. Given the variety of methods to specify spatial relationships, one source is the choice of an appropriate spatial weight matrix that describes the spatial interactions between two regions. However, LeSage and Pace (2014) recently suggested that the model estimates are insensitive to the spatial weight matrix as long as the model is specified correctly. By comparison, the second aspect of model uncertainty still bothers economic structure researchers and concerns how to determine the set of control variables to be used to model the diversity-performance relationship. To date, the only study that deals with model uncertainty is Watson and Deller (2017) who used a Spatial Bayesian Model Averaging (SBMA) method to determine the set of control variables in studying the effect of industrial diversity on unemployment. Nevertheless, Watson and Deller (2017) still used counties (i.e., administrative regions) rather than functional regions as the analytical units and ignored the effect of cluster diversity in studying regional economic structure, both of which can impact severely our understanding of the relationship between economic structure and economic performance.

This chapter then contributes to the literature on regional economic structure research in several aspects. First, it utilizes a Bayesian Model Averaging (BMA) method that is

different from Watson and Deller’s (2017) in order to address model uncertainty in studying the influences of economic diversity on economic stability and employment growth in the context of U.S. This method simultaneously addresses model uncertainty and spatial spillovers. To further evaluate the role of model uncertainty in studying the relationship between economic diversity and regional economic performance, comparisons are made between model estimates with and without addressed model uncertainty. Second, based on recent studies like Hong and Xiao (2016) and findings of Chapters 2 and 3 in the dissertation, this analysis uses MSAs as the basic units to approach regional economic systems and considers both industry and cluster diversity. Finally, this chapter also provides a review of previous modeling methods employed in economic structure research, mainly after the 1970s, to compare and contrast their usages and limitations.

The remainder of this chapter is organized as follows. Section 4.2 reviews previous methods used to study the effect of economic diversity on regional economic performance. Following the description of methodology in Section 4.3, results are presented and discussed. The final section closes with the findings of this chapter.

4.2 Background

Based on previous studies of regional economic structure in the last five decades, three broad groups of modeling methods can be identified, including (1) bivariate statistics, (2) multivariate regression, and (3) spatial econometric models. These methods, along with several examples of each, are displayed in Figure 4.1. The rest of this section reviews these methods as follows.

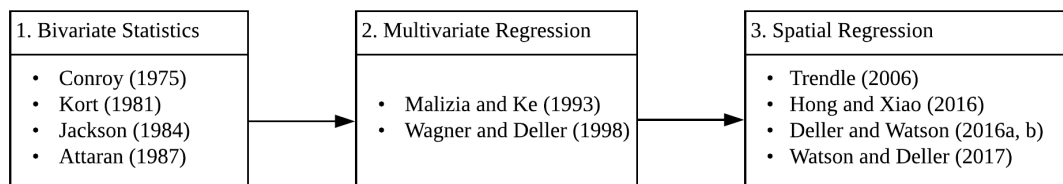


Figure 4.1: Three broad groups of modeling methods in economic diversity research

4.2.1 Bivariate Statistics

The first group of empirical studies (e.g., Conroy, 1975; Kort, 1981; Jackson, 1984; Attaran, 1986) has employed bivariate statistics to study the impact of economic diversity on regional economic performance. For example, Conroy (1975) used several bivariate techniques—such as the Person correlation coefficients and bivariate regression—to study the relationship between economic diversity and stability. Using data from 52 MSAs during the period of 1958-1967, Conroy found that economic diversity contributes to stability. Kort (1981) also examined the extent to which industrial diversity affects economic stability among 106 MSAs in the U.S. using bivariate regression; he further considered the possibility of heteroscedasticity as a matter of city size and used a weighted linear regression to study diversity and stability. Kort concluded that economic diversity explains the differences in regional economic instability. Conversely, Jackson (1984) studied the relationship between economic diversity and stability in the case of Illinois counties using simple correlations and found that this relationship is insignificant. Similarly, Attaran (1986) also assessed the impacts of economic diversity on unemployment, economic instability, and economic growth among the 50 states plus Washington, D.C. Using correlation indices, Attaran found that these impacts are insignificant or even non-existent.

The dependent variables in these four studies are economic performance indicators, such as employment growth and unemployment rate, while the only explanatory variable is industrial diversity. As such, other factors that might influence regional economic performance have not been controlled for; region size, for instance, may affect economic stability and large regions tend to demonstrate greater stability in their economic performance than do small ones. In an econometrics textbook, Stock and Watson (2007, p. 478) used the term control variable to “describe a variable that is included in a regression model to control for a factor that, if omitted from the regression, would lead to omitted variable bias for the coefficient of interest.” In the literature on regional economic structure, Malizia and Ke (1993, p. 226) argued that “control variables are needed to reduce estimation bias” resulting from the use of bivariate techniques. To this end, including control variables becomes necessary to understand the effect of economic diversity on regional economic performance.

4.2.2 The Inclusion of Control Variables

The second group of studies has used multivariate statistics to consider the impact of control variables; examples are Malizia and Ke (1993) and Wagner and Deller (1998). Particularly, Malizia and Ke (1993) used multiple linear regression to study the influence of economic diversity on unemployment and economic stability among 282 MSAs in the U.S. In addition to the diversity variable, Malizia and Ke included several control variables—such as population size, labor force characteristics, and industry employment percentages—in their cross-sectional models. The control variables to include in the final estimation were determined through partial correlation both individually and in combination. The problem of heteroscedasticity was also considered, but Malizia and Ke found that this was not a problem in their analysis. Additionally, the employment growth rate was included in Malizia and Ke's work to test the hypothesized negative relationship between economic growth and stability. As a result, their analysis confirmed that industrially diversified regions experience low unemployment rates and high economic stability.

Wagner and Deller (1998) studied state-level diversity in the U.S. and its impact on economic stability and growth. Their control variables were selected based on Duffy's (1994) five broad factors that influence regional economic performance and a series of principal component analyses. Compared to Malizia and Ke (1993), Wagner and Deller proposed that long-term development goals should focus on diversification while short-term goals should focus on growth; in other words, the trade-off between stability and growth no longer exists. Thus, growth rate was excluded from the control variables in Wagner and Deller's analysis.

Taken together, the dependent variables in these two studies are indicators of regional economic performance, whereas the independent variables include a diversity measure and a set of control variables that capture the demographic, economic and industrial differences among regions. Although studies in this group include the impact of control variables, they have not considered the effect of spatial dependence. Ignoring this dependence might result in the misspecification of economic diversity and thus result in inappropriate economic development policies. As such, spatial econometric techniques might be more appropriate.

4.2.3 The Role of Spatial Spillovers

Table 4.1: Summary of the economic diversity research that considers spatial effects

Article	Spatial regression models	Spatial dependence	Spatial heterogeneity	Model uncertainty
Trendle (2006)	SAR and SEM	✓		
Hong and Xiao (2016)	SAR	✓		
Deller and Watson (2016a)	SDM	✓		
Deller and Watson (2016b)	GWR	✓	✓	
Watson and Deller (2017)	SDM with a heteroscedastic error	✓	✓	✓

Notes: ✓=Yes; SAR = Spatial autoregressive model; SEM = Spatial error model; SDM = Spatial Durbin model; GWR = geographical weighted regression.

Spatial econometric models¹ have been used to assess the spatial dependence of regional economic diversity and economic performance by regional scientists, such as Trendle (2006), Hong and Xiao (2016), Deller and Watson (2016a,b), and Watson and Deller (2017) in Table 4.1. According to Anselin (1988), geographical data sets display two general properties that should be addressed in spatial data analysis: spatial dependence and spatial heterogeneity². One of the early works that used spatial regression models to study the relationship between industrial diversity and economic stability is Trendle (2006). He specifically used the spatial autoregressive model (SAR) and the spatial error model (SEM). The result of Trendle's (2006) analysis confirmed the existence of spatial dependence within the diversity-stability relationship. Compared to Trendle (2006), Hong and Xiao (2016) used the SAR model to evaluate the performance of industrial diversity on regional economic performance and also proposed a Multiple Specialization Index (MSI) that considers the diversity of economic specializations to emphasize the coexistence of economic diversity and specialization. Furthermore, Deller and Watson (2016a) used the spatial Durbin model (SDM) that captures the spatial dependence in both dependent and independent variables to assess the effect of economic diversity on economic stability among U.S. counties during the Great Recession from 2007 to 2014. Similarly, Deller and Watson (2016b) used the method of Geographi-

¹For an overview of spatial regression models, see Anselin (1988) and LeSage and Pace (2009).

²Chapter 1 provides a formal introduction of these two concepts.

cal Weighted Regression, or GWR³, to study the relationship between economic diversity and stability and found that the relationship is not stable across space. In that sense, apart from spatial dependence, spatial heterogeneity with the underlying diversity-stability relationship is also considered.

Although the traditional approach is to select a single “best” model for model specification based on various metrics, such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the adjusted R-squared value, and log likelihood, there is little concern about model uncertainty within economic structure research. For example, the classic trade-off between the inclusion of sufficient independent variables and the inclusion of redundant variables affects the empirical understanding of economic diversity. Fernández et al. (2001a,b) suggested using Bayesian Model Averaging (BMA) methods to address the model uncertainty issue. Given this, Watson and Deller (2017) applied a Spatial Bayesian Model Averaging method to specify the combination of control variables and used specified control variables to model the diversity-unemployment relationship with a spatial Durbin model. In addition, because Deller and Watson (2016b) indicated that the impact of industrial diversity on regional economic performance varied significantly across space, Watson and Deller (2017) also included a heteroscedastic error structure in their estimation to account for the spatial heterogeneity within the diversity-performance relationship.

In a similar vein, this chapter uses a BMA method that seeks to consider model uncertainty. Comparisons between model estimates with and without addressed model uncertainty are also made. In addition, when assessing the diversity-performance relationship using the BMA results, recent research into the geographical units of regional economic systems and structural measures of economic diversity is also included. The next section introduces the methodological details.

³For more information about the technical details of the GWR method, see Fotheringham et al. (2002).

4.3 Methodology

The empirical models are expressed as:

$$REI_i = f(DIV_i, CONTROL_i) \quad (4.1)$$

$$GROWTH_i = g(DIV_i, CONTROL_i) \quad (4.2)$$

where the dependent variables are the long-term (2000-2014) regional economic instability (REI) index and the short-term (2000-2002) employment growth rate, and the independent variables are economic diversity measures and a set of control variables for the base year 2000. These two empirical models are studied among 359 MSAs in the contiguous U.S. As suggested by previous authors (e.g., Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006) and Chapter 2 in this dissertation, MSA is used as the analytical unit because MSAs can form meaningful functional economic systems, and the focus of this chapter is not to investigate the diversity-performance for the whole U.S. regions.

According to Kort (1981), Jackson (1984), Malizia and Ke (1993) and others, regional economic instability is measured as:

$$REI_i = \left\{ \sum_{i=1}^N [(E_{it} - E_{it}^{Tr}) / E_{it}^{Tr}]^2 / T \right\}^{1/2} \quad (4.3)$$

where i is the region index; E_{it} is the observed number of employment for region i at time t ; T is the number of observed time spans; and E_{it}^{Tr} is the predicted number of employment for region i at time t using a linear trend line. By comparison, employment growth at time t is measured as the growth rate of total employment from $t - 1$ to t . Both the instability and growth variables were calculated using data from the U.S. Bureau of Economic Analysis.

Although the empirical models and dependent variables are specified, this analysis still encounters four issues: (1) economic diversity measurement, (2) model uncertainty, (3) potential control variables, and (4) spatial relationships. The rest of this section discusses and addresses these issues.

4.3.1 Measuring Economic Diversity

As reviewed by Wagner (2000), Dissart (2003) and Jackson (2015), economic structure literature has defined economic diversity through various metrics like the national average, the ogive, the Herfindahl Hirschman Index (HHI) and the entropy index. Among these metrics, the entropy index and the HHI have been used more widely than others. As the focus of this analysis is not to compare different measures of industrial diversity, the HHI of sectors (HHIS) is used and can be calculated as:

$$HHIS_i = \sum_{j=1}^N (e_{ij}/e_i)^2 \quad (4.4)$$

where e_{ij} is the employment for industry j in the i^{th} region; E_i is the total employment in the i^{th} region; and N is the total number of industries in the i^{th} region. This index reaches its maximum of one for a one-sector economy and approaches to its minimum of $1/N$ if all sectors are evenly distributed in terms of employment.

Moreover, many regional scientists (e.g., Malizia and Ke, 1993; Wagner and Deller, 1998; Jackson, 2015; Hong and Xiao, 2016) reconsidered the relationship between economic diversity and specialization and proposed the concept of diversified specializations. Recently, empirical studies—such as Hong and Xiao (2016) and Chapter 3 in this dissertation—have applied this concept to stress the coexistence of specialization and diversity. Specifically, compared with Hong and Xiao (2016), the analysis in Chapter 3 excluded the impact of local industries in identifying economic clusters; for example, utilities and drug stores that only serve local needs should not be regarded as potential economic clusters. Therefore, this analysis uses the method used in Chapter 3 to consider the diversity of clusters (HHIC) as follows:

$$HHIC_i = \sum_{j=1}^M (e_{ij}/e_i)^2 \quad (4.5)$$

where e_{ij} is the employment for cluster j in the i^{th} region; E_i is the total employment of traded industries⁴ in that region; and M is the total number of clusters in region i . Be-

⁴For more information about the definition of traded and local industries, see Porter (2003) and Delgado et al. (2016).

cause these clusters are specialized relative to the national average, the location quotients (LQs)⁵ of these clusters should be greater than one. HHIC ranges from $1/M$ for a perfectly diversified economy to one if all employment is concentrated in one cluster. As for the data sources, the industry diversity variable is calculated using the Upjohn Institute’s “Whole-Data” version of County Business Patterns, derived using Isserman and Westervelt’s (2006) method. Based on the same data as well as Delgado et al.’s (2016) cluster classification method, the cluster diversity variable is also calculated.

4.3.2 Potential Control Variables

Building on Trendle (2006), Deller and Watson (2016a), Watson and Deller (2017), and Deller et al. (2017), the demographic, economic, and industrial differences between MSAs are considered as potential control variables in this study. The demographic factors include (1) population, (2) percentage of the population greater than 25 years of age with at least a bachelors degree, (3) percentage of the population over 65, and (4) percentage of the nonwhite population. These data describe the general demographic characteristics of regions, and no specific hypotheses are offered relative to regional economic performance. In addition, the demographic data are from the 2000 Census.

Similarly, the economic aspect of control variables includes (1) per capita income relative to the U.S. average, (2) percent of households with income below \$20,000, (3) percent of households with income above \$150,000, (4) Gini coefficient of income inequality, (5) per capita income from transfer payments, (6) per capita income from dividends, interest and rent, and (7) per capita income from proprietorship. Deller et al. (2017) introduced the expected effects of the last three variables for a given region: per capita income from transfer payments introduces stability; per capita income from dividends, interest and rent measures wealth and introduces instability; and finally, per capita income from proprietorship indicates economic dependency on small businesses. These economic variables were collected from the BEA and the Census Bureau for the year 2000.

To capture the industrial differences, the following factors were included: (1) percentage

⁵As mentioned in Chapter 3, LQ is calculated as the ratio of regional employment share to the national employment share of the same sector.

of employment in government sectors, (2) percentage of employment in goods production sectors (minus farming), and (3) percentage of employment in service production sectors. These factors are important because Mizuno et al. (2006) suggested that the economic diversity index does not consider the components of industrial structure. Specifically, Deller et al. (2017) argued that a high dependency on goods-producing sectors contributes to instability⁶, whereas the number for employment in service-related and government sectors is positively associated with economic stability. These industry data were obtained from the Bureau of Labor Statistics (BLS), Census of Employment and Wages for the base year of 2000.

4.3.3 Model Uncertainty

Since the 1990s, Bayesian Model Averaging methods have been introduced in economic growth literature (e.g., Sala-i-Martin, 1997; Fernández et al., 2001a,b; Sala-i-Martin et al., 2004) to address the model uncertainty issue regarding the choice of explanatory variables and model specification. More recently, scholars have considered the role of spatial spillovers and explored various aspects of model uncertainty at regional levels (See LeSage and Parent, 2007; LeSage and Fischer, 2008; Crespo Cuaresma and Feldkircher, 2013; Crespo Cuaresma et al., 2014). A growing body of literature in regional science has gone beyond economic growth research and has used BMA methods to determine the set of control variables to address model uncertainty (e.g., Parent and LeSage, 2012; Winkler et al., 2015; Watson and Deller, 2017).

In regional studies that employ BMA methods, two general approaches have been identified in considering potential spatial dependence and addressing model uncertainty. One approach uses spatial BMA or SBMA via LeSage and Parents (2007) numerical integration techniques to obtain posterior model probabilities for model specifications. Examples of researchers who have employed this approach include LeSage and Fischer (2008), Crespo Cuaresma et al. (2014), Winkler et al. (2015), and Watson and Deller (2017). By comparison, a second approach applies spatial filtering techniques (Getis and Griffith, 2002;

⁶This is consistent with the “durable” goods measure of economic diversity. According to Jackson (1984), durable goods are sensitive to economic fluctuations. During an economic downturn, customers are assumed to be less likely to purchase such durable goods as automobiles, books and furniture.

Tiefelsdorf and Griffith, 2016) that remove the spatial effects and then consider model uncertainty in the framework of standard (non-spatial) BMA methods. For example, Crespo Cuaresma and Feldkircher (2013) used this approach to study factors that influenced the speed of income convergence in Europe from 1995 to 2005. Crespo Cuaresma and Feldkircher further mentioned that the model uncertainty that results from the appropriate spatial weight matrix can also be solved with the second approach, although LeSage and Pace (2014) indicated that the model estimates are insensitive to the choice of spatial weights⁷. Furthermore, from a technical point, Crespo Cuaresma and Feldkircher suggested that the use of spatial filtering overcomes the computational difficulties associated with LeSage and Parent’s (2007) SBMA. For this reason, this second approach is preferred in this analysis.

Consider such a spatial autoregressive (SAR) model as:

$$y = \alpha l_N + \rho W y + X_k \beta_k + \varepsilon \quad (4.6)$$

where y is the dependent variable for N regions; α is the intercept; l_N is an $N \times 1$ vector of ones; ρ is a scalar that denotes the level of spatial autocorrelation; W is the spatial weight matrix indicating the geographical relationship between any two regions; X_k is an $N \times k$ matrix that includes k explanatory variables; β_k is the estimated coefficient corresponding to X_k ; and ε is the error term. In Equation 4.6, the number and identity of the variables in X_k are unknown and come from K potential explanatory variables ($K \geq k$). Any model M_k is contained in a larger set of 2^K possible models.

According to LeSage and Pace (2009), another model uncertainty arises regarding the appropriate spatial weight matrix to specify the underlying spatial interactions. Unlike control variables, there are few theoretical foundations on the construction of spatial weight matrices, and Anselin (1988) suggested that there are various methods (e.g., continuity, distance band and k nearest neighbors) to specify spatial relationships. As such, many

⁷Although the choice of spatial weights (e.g., k nearest neighbors, contiguity and distance-band based methods) does not matter (LeSage and Pace, 2014), there are differences between spatial weight matrices and other cross-sectional dependence matrices based on non-geographical factors, such as socio-cultural indicators. Crespo Cuaresma and Feldkircher’s (2013) approach can also be used among these cross-sectional dependence matrices.

empirical analyses use robustness checks based on “the log-likelihood function values” to identify the appropriate spatial weight matrix (LeSage and Pace, 2009, p. 162). However, more recently, LeSage and Pace (2014) indicated that estimates and inferences of spatial econometric models are insensitive to the spatial weight matrix used in the model. Although this type of model uncertainty does not matter that much, a spatial weight matrix has to be specified for a particular analysis. Specifically, similar to the control variables, the “best” spatial weight matrix is chosen based on the probabilities of seven k nearest neighbor matrices ($k = 3, 4, \dots, 8, 9$). In these matrices, region i regards its k closest regions in terms of physical distance as neighbors. Region i regards its k closest regions in terms of physical distance as neighbors. If region j belongs to these k regions, then the corresponding element in the spatial weight matrix W_{ij} equals one; otherwise, it equals zero. Overall, there are seven spatial weight matrices and 14 potential control variables considered. The cardinality of model space is therefore 114,688 (7×2^{14}) in this analysis.

Generally, there are two steps in Crespo Cuaresma and Feldkircher’s (2013) approach⁸. First, spatial filtering techniques are used to decompose the data into a purely spatial and a non-spatial component. Specifically, spatial dependence in Equation 4.6 is removed using an eigenvector decomposition method proposed by Getis and Griffith (2002) and Tiefelsdorf and Griffith (2016). The eigenvectors e_i are included as extra explanatory variables in Equation 4.6 with the following form:

$$y = \alpha l_N + \sum_{i=1}^E \gamma_i e_i + X_k \beta_k + \varepsilon \quad (4.7)$$

where each eigenvectors e_i spans one of the spatial dimensions. Moreover, this step can also reduce the degree of multicollinearity and further “separate spatial effects from the ‘intrinsic’ impact the employed regressors exert on the dependent variable” (Crespo Cuaresma and Feldkircher, 2013, p. 723).

Second, the results of spatial filtering are then processed with standard BMA methods. As mentioned earlier, model uncertainty exists in both the spatial weight matrix W and explanatory variables X_k . Following the Bayesian Model Averaging methodology, inference

⁸An R Package that carries these two steps is available at <https://modelaveraging.wordpress.com/2010/10/>, accessed March 15, 2018.

on the parameters can be written as:

$$p(\beta_j|Y) = \sum_{k=1}^{2^k} \sum_{z=1}^Z p(\beta_j|Y, M_j^z)p(M_j^z|Y) \quad (4.8)$$

where Y is the whole data. Note that the focus of this research is not the weighted average but rather the posterior probability of each potential variable. Instead of depending on a single model, BMA calculates the weighted average of the posterior probability densities, where the weights are the posterior probabilities of each model and can be given by:

$$p(M_j^z|Y) = \frac{p(Y|M_j^z)p(M_j^z)}{\sum_{k=1}^{2^k} \sum_{z=1}^Z p(Y|M_j^z)p(M_j^z)} \quad (4.9)$$

where $p(M_j^z)$ denotes the prior distribution of M_j^z . For a given model, a non-informative prior on α and σ , and a g-prior on β are used as follows:

$$p(\beta_k|\alpha, \rho, \sigma, M_j) \sim N[\beta_k, \sigma^2(gX_k'X_k)] \quad (4.10)$$

with $g = 1/\max\{N, K^2\}$. Fernández et al. (2001a,b) indicated that using Zellner's (1986) g-prior simplifies the computational process. Finally, based on Ley and Steel (2009), a binominal-beta prior distribution is used for the prior distribution of M_j^z .

Following Madigan et al. (1995), Raftery et al. (1997), Fernández et al. (2001a,b), and Crespo Cuaresma and Feldkircher (2013), a Markov chain Monte Carlo model composite (MC^3) is employed to obtain the posterior distributions of interest over the model space⁹. A random-walk step is used in every replication of the MC^3 procedure. One can propose an alternative model M' to the current model in each step of the chain by adding (*birth step*) or subtracting (*death step*) a regressor from model M . The chain moves to the proposed model using the following acceptance probability:

$$\min \left[1, \frac{p(M'|y)}{p(M|y)} \right] \quad (4.11)$$

Otherwise, the chain stays in the current model. In that sense, the posterior probabilities of

⁹See Crespo Cuaresma and Feldkircher (2013) for a more detailed description of this method.

explanatory variables and spatial weight matrices are calculated based on the models visited by the MC^3 rather than the whole model space. As detailed later, the explanatory variables and spatial weight matrix are specified based on these posterior probabilities; in other words, model uncertainty can be addressed. In addition, the results of this specification are used to model the relationship between economic diversity and regional economic performance based on the method introduced in the next subsection.

4.3.4 Spatial Relationships

As suggested by Trendle (2006), Deller and Watson (2016a), Watson and Deller (2017) and others, the spatial spillover effects cannot be ignored when studying the effect of economic diversity on regional economic performance. In this analysis, the spatial Durbin model is used because it considers the spatial effects in both dependent and independent variables. More formally, the SDM posits that the variations of the dependent variable can be explained by the spatially lagged dependent and independent variables and a set of independent variables with the following form:

$$y = \rho W y + X\beta + WX\gamma + \varepsilon \quad (4.12)$$

where y is the dependent variable; X is a matrix of independent variables; β is a vector of estimated coefficients of the independent variable; ρ is a coefficient that describes the strength of the spatial autocorrelation in the dependent variable; γ is a vector of estimated coefficients of the spatially lagged independent variables WX ; and ε is the error term that follows a homoscedastic pattern ($\varepsilon \sim N(0, \sigma^2 \times I)$).

The implication of this homoscedasticity suggests that the relationship between economic diversity and regional economic performance is stable across space. Yet, Deller and Watson (2016b) found that the effect of economic diversity was more significant in certain parts of the U.S.(i.e., spatial heterogeneity) using the GWR method—a local regression technique that returns a parameter estimate for each observation. As the purpose of this analysis is not to derive the local estimate but to compare model estimates between with and without addressed model uncertainty, the GWR model is not preferred here. Rather,

the SDM with a heteroscedastic error structure ($\varepsilon \sim N(0, \sigma^2 V)$ where $V \neq I$) in a Bayesian framework provides the magnitude and significance of an average effect over space (i.e., global regression) and thus should be used. Based on LeSage and Pace (2009) and Watson and Deller (2017), the following prior distributions are specified.

$$\pi(\beta) \sim N(C, N) \tag{4.13}$$

$$\pi(r/v_i) \sim \chi^2 IID(r) \tag{4.14}$$

$$\pi(1/\sigma^2) \sim \Gamma(d, v) \tag{4.15}$$

$$\pi(\rho) \sim U[0, 1] \tag{4.16}$$

The parameters β , ρ and γ can be drawn sequentially in a Bayesian framework. This analysis used 56,000 draws with the first 6,000 as the burn-ins. The removal of these burn-ins is useful because the initial values of the parameters might be unstable.

Finally, LeSage and Fischer (2008) and LeSage and Pace (2009) suggested that the coefficients of the variables β in Equation 4.12 cannot be interpreted as marginal effects directly because of spatial dependence. Instead, following LeSage and Pace (2009), direct, indirect and total effects can be estimated.

4.4 Empirical Results

For simplicity, three sets of results are presented. The first set is the BMA results (Table 4.2) that provide insights into model uncertainty in the control variables, and the second set (Table 4.3) is the posterior probabilities for different spatial weight matrices. The third set (Tables 4.4 and 4.5) is the estimated effects of economic diversity on economic growth and instability when the control variables are suppressed and the spatial weight matrix is specified.

The results in Table 4.2 demonstrate significant differences in terms of posterior inclusion probability (PIP). Conceptually, the posterior inclusion probability is calculated as the sum of probabilities of models including variable X_k . A PIP of a variable approaching

Table 4.2: Posterior inclusion probability of control variables

Variable	Instability model	Robustness category	Growth model	Robustness category
Logged population	0.076		0.562	Weak
Percentage of the population greater than 25 years old with at least a bachelors degree	0.757	Substantial	0.077	
Percentage of the population over 65	0.392		0.103	
Percentage of nonwhite population	0.035		0.936	Substantial
Per capita income relative to the U.S. average	0.121		0.999	Decisive
Percent of households with income below \$20, 000	0.104		0.241	
Percent of households with income above \$150, 000	0.309		0.987	Strong
Gini coefficient of income inequality	0.044		0.083	
Per capita income from transfer payments	0.585	Weak	0.998	Decisive
Per capita income from dividends, interest and rent	0.771	Substantial	0.995	Decisive
Per capita income from proprietorship	0.048		0.198	
Percentage of employment in government sectors	0.044		0.084	
Percentage of employment in goods production sectors (minus farming)	0.711	Weak	0.989	Strong
Percentage of employment in service production sectors	0.040		0.084	

Notes: Calculations are based on standard Markov chain Monte Carlo model composition (MC^3) sampling with 100 thousand burn-ins and 1 million draws; PIP values greater than 0.5 are in bold.

unity suggests the importance of the variable in explaining the dependent variable. Numerous studies (Eicher et al., 2011; Crespo Cuaresma and Feldkircher, 2013) have labeled covariates with PIP greater than 0.5 as robust and have suggested including them in the final specification. Hence, robust variables in the instability model are (1) percentage of the population with at least a bachelors degree; (2) per capita income from transfer payments; (3) per capita income from dividends, interest and rent; and (4) percentage of employment in goods production sectors (minus farming). By comparison, in the growth model, the corresponding robust variables ($PIP \geq 0.5$) are (1) percentage of nonwhite population; (2) per capita income relative to the U.S. average; (3) percent of households with income above \$150,000; (4) per capita income from transfer payments; (5) per capita income from dividends, interest and rent; and (6) percentage of employment in goods production sectors (minus farming). These robust control variables are used to estimate the effects of economic diversity on regional economic instability and employment growth based on Equations 4.1 and 4.2. Although Watson and Deller (2017) indicated that the set of control variables is of secondary interest in economic structure research, including these variables is expected

to avoid redundant variables that decrease precision of the estimation on the one hand, and overcome potential bias resulting from omitted variables on the other hand.

Raftery et al. (1997), Eicher et al. (2011), and Crespo Cuaresma et al. (2014) further classified robust variables, according to their PIP values, into four categories: weak (50-75%), substantial (75-95%), strong (95-99%) and decisive (above 99%) variables. In that sense, Table 4.2 also denotes the robustness category of these control variables. It seems that the overall PIP values of robust variables are higher in the growth model than those values in the instability model. The PIP values of employment in goods producing industries in both models are greater than 0.50, indicating that this variable is significantly associated with the dependent variables. By comparison, there are several control variables with low PIP values, such as Gini coefficient of income inequality. However, this does not undermine the validity of the theoretical assumptions but implies that this variable does not help us explain the variations in the dependent variables here.

Table 4.3: Posterior inclusion probability of spatial weight matrices

Variable	Instability model	Growth model
KNN3	0.000	0.001
KNN4	0.824	0.001
KNN5	0.163	0.425
KNN6	0.000	0.570
KNN7	0.016	0.000
KNN8	0.01	0.000
KNN9	0.000	0.001

Note: For each spatial weight matrix, posterior probability is calculated as the sum of posterior probabilities of models containing the eigenvectors of that matrix.

Table 4.3 reports the posterior probabilities associated with k nearest neighbor weight matrices with $k = 3, 4, \dots, 8, 9$. Comparing the results of these two models, the instability model seems to lend strong support to the four nearest neighbor spatial weight matrix (KNN4), while the growth model appears to favor the six nearest neighbor spatial weight matrix (KNN6). As noted earlier, few theoretical perspectives have provided guidance on the specification of spatial weight matrix, and the matrix with highest posterior probability is used for each model. That said, KNN4 is used for the instability model and KNN6 for the growth model.

Table 4.4: Effect estimates of models that consider industry diversity

	Direct	Indirect	Total
(1) Saturated instability model			
Industry diversity (HHIS)	0.051 (1.635)	0.010 (0.158)	0.061 (0.842)
(2) Suppressed instability model			
Industry diversity (HHIS)	0.024** (1.934)	0.067*** (2.765)	0.092*** (3.513)
(3) Saturated growth model			
Industry diversity (HHIS)	-0.048 (-0.549)	-0.370 (-1.596)	-0.418 (-1.642)
(4) Suppressed growth model			
Industry diversity (HHIS)	-0.023 (-0.256)	-0.200 (-0.903)	-0.224 (-0.952)

Note: Significance levels: * for 10%, ** for 5%; *** for 1%; t statistics in parentheses.

Table 4.4 presents the estimated effects of industrial diversity on economic instability and employment growth. Besides the suppressed models (Models 2 and 4) that include the appropriate set of control variables from Table 4.2, a saturated version of both the instability and growth models is provided in Models 1 and 3. According to LeSage and Pace (2009, p. 184), saturated models include all variables during estimation. For the instability models, the results of Models 1 and 2 seem to support the portfolio theory that industrial diversity is positively associated with economic stability. Specifically, the estimated direct, indirect and total effects are positive and significant in Model 2, while only the direct effect in Model 1 is significant. By comparison, except for the total effect in Model 3, the effect of diversity on employment growth is not significant. In other words, the results of the growth models appear to indicate that industrial diversity only barely stimulates employment growth.

To supplement Table 4.4, Models 5-8 in Table 4.5 consider the impact of both industry and cluster diversity on economic stability and employment growth. For the instability models, it seems that only cluster diversity always contributes to economic stability in Models 5 and 6. Focusing on the t-statistics of its estimated direct, indirect and total effects, cluster diversity in the suppressed model appears to be more significantly associated with economic instability than that in the saturated model. By comparison, the signs of industry diversity in Models 5 and 6 are inconsistent with the theoretical assumption that industrial diversity enhances regional economic stability. This result seems to suggest the

Table 4.5: Effect estimates of models that consider both industry and cluster diversity

	Direct	Indirect	Total
(5) Saturated instability model			
Industry diversity (HHIS)	-0.007 (-0.202)	-0.179** (-2.197)	-0.187** (-2.026)
Cluster diversity (HHIC)	0.026** (2.069)	0.078 (3.083)	0.104*** (3.955)
(6) Suppressed instability model			
Industry diversity (HHIS)	0.022 (0.553)	-0.027 (-0.347)	-0.005 (-0.061)
Cluster diversity (HHIC)	0.036*** (2.758)	0.096*** (3.687)	0.133*** (4.672)
(7) Saturated growth model			
Industry diversity (HHIS)	-0.101 (-0.920)	-0.456 (-1.557)	-0.557* (-1.738)
Cluster diversity (HHIC)	0.026 (0.723)	0.031 (0.058)	0.058 (0.677)
(8) Suppressed growth model			
Industry diversity (HHIS)	0.013 (0.365)	-0.096 (-1.266)	-0.082 (-0.952)
Cluster diversity (HHIC)	0.024** (1.934)	0.067*** (2.765)	0.092*** (3.513)

Note: Significance levels: * for 10%, ** for 5%; *** for 1%; t statistics in parentheses.

use of clusters rather than industries in assessing economic diversity. Similarly, in the growth models, the diversity of clusters seems to be more associated with employment growth than the industrial diversity variable.

4.5 Discussion

The empirical results provide several interesting points worthy of note. First, although closely related to the work of Watson and Deller (2017), who studied the relationship between industrial diversity and unemployment using a Spatial Bayesian Model Averaging method developed by LeSage and Parent (2007), this analysis differs from Watson and Deller's (2017) study in the following aspects: (1) a spatial filtering-based BMA method that considers model uncertainty in the choice of control variables and spatial weight matrices is employed; (2) rather than focusing on the relationship between unemployment and economic diversity, this analysis concentrates on long-term economic stability and short-term employment growth to leverage the benefits of both economic diversity and specialization; and (3) it also uses functional regions as the analytical units and considers both industrial

and cluster diversity on regional economic performance to minimize the impacts of factors that lead to the empirical inconsistency with theoretical assumptions of economic structure, such as highly geographical datasets and inappropriate measure of economic diversity.

Second, building on recent work in economic structure research, such as Jackson (2015), Hong and Xiao (2016), and findings of Chapter 3, this study includes both industrial and cluster diversity as indicators of economic diversity to study its effect on regional economic performance. Historically, numerous studies (Malizia and Ke, 1993; Wagner and Deller, 1998; Watson and Deller, 2017) had viewed economic diversity as the diversity of economic activities across industries, therefore treating them the same. By comparison, recent studies (Porter, 2003; Spencer et al., 2010; Delgado et al., 2016) have begun to take a cluster perspective of regional economic structure. Particularly, Porter (2003, p. 562) suggested using clusters rather than industries as the basic units to assess economic diversity because of “the externalities across related industries within clusters.” In this study, when comparing the effects of industry and cluster diversity in Tables 4.4 and 4.5, cluster diversity seems to be more supportive to the theoretical foundations of economic specialization and diversity (See Conroy, 1975; Glaeser et al., 1992; Porter, 1990, 1998) than industrial diversity. However, cluster and industrial diversity may overlap but are not identical essentially because industries may or may not form economic clusters. To this end, both are important elements of economic diversity and should be considered in studying the relationship between economic diversity and regional economic performance.

Third, Crespo Cuaresma and Feldkircher’s (2013) BMA approach has been employed successfully to address simultaneously model uncertainty from choice of control variables and spatial spillovers. The comparison between saturated and suppressed models in Tables 4.4 and 4.5 suggests that ignoring the model uncertainty can impact our understanding of the relationship between economic diversity and regional economic performance. Together with Watson and Deller (2017), this study suggests that future economic structure research can consider such model uncertainty to better understand economic diversity. In terms of the modeling method, both Crespo Cuaresma and Feldkircher’s (2013) and LeSage and Parent’s (2007) approaches provide solid technical foundations to address model uncertainty.

Finally, after the model uncertainty is considered, spatial spillovers still exist within the

diversity-performance relationship. That is to say, regional economic development should consider this spatial effect and encourage collaboration between regions. Based on the empirical results, promoting one regions diversity of economic clusters can encourage long-term economic stability of the region as well as its neighbors. On the other hand, specializing economic clusters can also bring spatial spillovers to neighbors and thus promote their employment growth. As such, neighbor regions might be regarded as a source of economic development and collaboration policies.

4.6 Chapter Summary

In this chapter, a Bayesian Model Averaging method is employed to address model uncertainty when studying the effects of economic diversity on long-term economic stability and short-term employment growth among 359 MSAs in the contiguous U.S. Compared to previous studies back to the 1970s, this method considers the impacts of control variables, spatial dependence of the dependent and independent variables, and, more importantly, addresses model uncertainty resulting from the set of control variables and spatial weight matrix. Meanwhile, a spatial Durbin model with a heteroscedastic error structure is also employed to estimate the effect of economic diversity. The methodology used in this chapter is expected to reduce the extent of misspecification in the modeling process and thus benefit future empirical research on economic diversity. Building on the work of recent economic structure research, this analysis also uses functional regions to assess regional economic systems and considers the diversity of economic clusters.

Significantly, the empirical results of this chapter suggest that ignoring model uncertainty can alter our understanding of economic diversity on regional economic performance and that after controlling for regional attributes and considering model uncertainty, the spatial spillovers of both industrial and cluster diversity still exist. Future economic development should consider the impact of spatial spillovers. In addition, this analysis confirms that industrial and cluster diversity exert two different mechanisms on regional economic performance.

Future research can consider the following three avenues. First, it is interesting to ex-

exploit other approaches to defining the cross-sectional dependence beyond k nearest neighbor spatial weight matrices. For example, inter-regional trade flows can be used to construct the spatial weight matrix, and this should help in understand the nature of spatial interactions within the diversity-performance relationship. Second, future research can also compare Crespo Cuaresma and Feldkircher's (2013) and LeSage and Parent's (2007) approaches that deal with model uncertainty within a spatial context in such dimensions as computational costs, usages and limitations. Finally, with more available data sets that might impact regional economic performance, it is also meaningful to include them as potential explanatory variables, such as women business ownership (Deller et al., 2017), and expand the economic diversity-performance literature beyond traditional thinking of economic diversity for economic development.

Chapter 5

Conclusions

5.1 Summary and Reflections

This dissertation research has explored geographical scales, structural measurements and modeling methods in economic structure research in the context of the U.S. regional economies. Yet, it is not the main purpose of the dissertation to conduct an *ad hoc* empirical analysis on the relationship between economic structure and regional economic performance and provide specific economic development suggestions for the country. Instead, the empirical inconsistency with the theoretical assumptions of economic structure is focused in the perspectives of geographical scales, structural measurements, and modeling methods in Chapter 2-4. The suggestions and improvements regarding these perspectives are expected to assist our understanding in economic diversity and hence benefit regional development.

More specifically, Chapter 2 evaluates the relationship between economic diversity and economic instability across four geographical scales—including counties, states, MSAs and EAs. Based on the empirical results in this chapter, geographical scales are found to greatly impact our understanding of the empirical relationship between economic diversity and economic stability. Although scale-related problems can be study dependent, functional regions rather than formal regions are recommended to approach regional economic systems. Equally important is region size because regions should be large enough in terms of population and employment to quantify meaningful economic structures. In addition, possible temporal variations in the boundaries of functional regions should also be considered.

Chapter 3 develops an alternative measure of economic structure that can interpret economic diversity as the presence of multiple specializations. When comparing with other measures, this measure considers the impact of industries that only serve local demands and also excludes these industries as the candidates for economic clusters. Moreover, the empirical results in this chapter suggest that both industry and cluster diversity contribute to long-term economic stability, while only cluster diversity stimulates short-term employment growth. In that sense, this chapter confirms that one region can simultaneously pursue a high and stable growth by developing diversified specializations.

Chapter 4 employs a BMA method to consider model certainty resulting from the set of control variables as well as the choice of an appropriate spatial weight matrix. The results of the BMA procedure are used to assess the relationship between economic diversity (including both industrial and cluster diversity introduced in Chapter 3) and economic performance. The empirical evidence indicates that ignoring model uncertainty can influence the effect of diversity on regional economic performance. For this reason, considering model uncertainty can improve our understanding of economic diversity and thus benefit development policies.

Methodologically, although Chapters 2-4 use spatial econometric techniques for model specification and comparison¹, there are nuances among the spatial regression models employed in these chapters. More specifically, Chapter 2 relies on maximum likelihood methods, whereas Chapter 3 uses Bayesian spatial economic models. Both of these two chapters consider the spatial dependence in the diversity-performance relationship. In addition to this spatial dependence, Chapter 4 further introduces a heteroscedastic error term in its modeling method; that means, both spatial dependence and spatial heterogeneity (Anselin, 1988) are considered within the diversity-performance relationship.

Taken together, as illustrated in Chapter 4, the improvements of geographical scales, structural measurements, and modeling methods are not mutually exclusive. Henceforth, to better understand the effect of economic diversity on regional economic performance, it is suggested to consider scale-related issues, diversity of clusters, and model uncertainty simultaneously.

¹According to LeSage and Pace (2009, p. 126), spatial econometric modeling can also be used for “estimation and inference about parameters” and “prediction of an out-of-sample set of observations.”

5.2 Future Research Directions

Moving beyond the empirical inconsistency addressed in this dissertation, future economic structure research can consider these three directions.

First, given that a large volume of empirical research across multiple branches of social sciences (including economic structure research) considers spatial dependence and spatial heterogeneity of economic activities, future economic structure research should consider how this spatial form takes place from an analytical standpoint and, what we should do with the spatial spillovers in the perspective of public policy. In this dissertation, it is suggested to pay explicit attention to the geographical scale issue of economic activities when applying the cross-fertilization of spatial statistics and spatial economics. Analytically, this issue is inherent in the *area analysis* (or *regional*) tradition in geography (Pattison, 1964) that deals with such questions as determining the spatial extent and boundaries of regions. The issue of geographical scales raised in Jackson (1984) and Malizia and Ke (1993) and emphasized in this dissertation is a step in which to delineate regional economic systems and shed light on the underlying spatial process of economic activities. This emphasis on process is expected to continue and will benefit our understanding in economic structure and regional development policies as well.

Second, future research can explore different aspects of regional economic structure and thus strengthen our understanding in an eclectic manner². Historically, economic structure research has focused on the Growth Pole (Perroux, 1950), economic base theory (Tiebout, 1956), industrial districts (Isard, 1956), key sectors (Schultz, 1977; Hewings, 1982), industrial complexes (ÓhUallacháin, 1984), industry clusters (Porter, 2003; Delgado et al., 2016), economic diversity (Conroy, 1975; Jackson, 1984; Malizia and Ke, 1993; Deller and Watson, 2016a,b), and related and unrelated variety (Frenken et al., 2007; Content and Frenken, 2016), all of which can diversify our understanding of economic structure and benefit regional development. Hence, future regional economic development should benefit from diverse aspects of economic structure with more granularity and nuance being explored and exploited by both researchers and practitioners.

²Interestingly, analyzing one region's economic structure can reflect the *locational* tradition of geography which relates to in-depth analyses of a place for certain phenomena or activities (Pattison, 1964).

Third, a major limitation of this dissertation research is that it has not addressed the role of time and has only focused on cross-sectional data, given a growing interest in the inclusion of temporal variations of economic structure (e.g. Izraeli and Murphy, 2003; Chiang, 2009; Neffke et al., 2011; Watson and Deller, 2017). For example, Watson and Deller (2017) found that the spatial spillover effect of diversity on unemployment is only significant at the height of the Great Recession using multiple cross-sectional analyses; that is to say, the spatial effect varies temporally. Likewise, Neffke et al. (2011) suggested that young and mature industries have different diversity needs (i.e., MAR or Jacob's externalities).

Broadly speaking, to stress the importance of time and space in social science research, Goodchild et al. (2000, p. 142) indicated that "space provides the framework for the integration of different social processes and, hence, different domains of social science. Reductionist traditions in science have led to our current arrangements in which different classes of processes are studied largely in isolation, in distinct disciplines, and often without specific attention to space and time. In the real world, these separate processes interact in a *spatiotemporal* context. New tools and data sources now allow treatment of the more general case." In the field of economic structure research, Krugman (1991a) also emphasized the role of path dependence in deciding the process of clustering; that is to say, history matters in shaping regional economic structure. Empirically, the temporal variations of economic structures are expected to provide additional insights besides simple cross-sectional analysis. Goodchild (2004, p. 172) further suggested that the spatial "process is much easier to infer from longitudinal data, with its representation of the sequence of events, than from cross-sectional data." Therefore, it is suggested to put economic structure research into a spatiotemporal context. Nevertheless, from a practical perspective, with the development of spatial statistics and spatial econometrics, collecting, representing, and analyzing such a large volume of spatiotemporal data sets at a fine industry level will be a longstanding research agenda for governments, institutions, and industries.

Bibliography

- Adams, J. S., VanDrasek, B. J., and Phillips, E. G. (1999). Metropolitan Area Definition in the United States. *Urban Geography*, 20(8):695–726.
- Alonso, W. (1964). *Location and Land Use*. Harvard University Press.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27(2):93–115.
- Anselin, L. (1999). Interactive Techniques and Exploratory Spatial Data Analysis. In Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., editors, *Geographical Information Systems: Principles, Techniques, Management and Applications*, pages 251–264. New York: Wiley.
- Anselin, L., Syabri, I., and Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38(1):5–22.
- Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies*, 29(3):155–173.
- Attaran, M. (1986). Industrial Diversity and Economic Performance in U.S. Areas. *The Annals of Regional Science*, 20(2):44–54.
- Boschma, R., Coenen, L., Frenken, K., and Truffer, B. (2017). Towards a Theory of Regional Diversification: Combining Insights from Evolutionary Economic Geography and Transition Studies. *Regional Studies*, 51(1):31–45.

- Boschma, R. A., Minondo, A., and Navarro, M. (2011). Related Variety and Regional Growth in Spain. *Papers in Regional Science*, 38:241–257.
- Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations? *Journal of Urban Economics*, 67(3):287–302.
- Brown, L. A. and Holmes, J. (1971). The Delimitation of Functional Regions, Nodal Regions, and Hierarchies by Functional Distance Approaches. *Journal of Regional Science*, 11(1):57–72.
- Burt, J., Barber, G., and Rigby, D. (1996). *Elementary Statistics for Geographers*. The Guilford Press.
- Carroll, M. C., Reid, N., and Smith, B. W. (2008). Location Quotients versus Spatial Autocorrelation in Identifying Potential Cluster Regions. *The Annals of Regional Science*, 42(2):449–463.
- Cheng, T. and Adepeju, M. (2014). Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PLOS One*, 9(6):e100465.
- Chiang, S.-h. (2009). The Effects of Industrial Diversification on Regional Unemployment in Taiwan: Is the Portfolio Theory Applicable? *The Annals of Regional Science*, 43(4):947–962.
- Chinitz, B. (1961). Contrasts in Agglomeration: New York and Pittsburgh. *The American Economic Review*, 51(2):279–289.
- Christaller, W. (1933). *Central Places in Southern Germany*. [Translation into English by Carlisle W. Baskin in 1966]. Prentice-Hall.
- Conroy, M. E. (1975). The Concept and Measurement of Regional Industrial Diversification. *Southern Economic Journal*, 41(3):492–505.
- Content, J. and Frenken, K. (2016). Related Variety and Economic Development: A Literature Review. *European Planning Studies*, 24(12):2097–2112.

- Crespo Cuaresma, J., Doppelhofer, G., and Feldkircher, M. (2014). The Determinants of Economic Growth in European Regions. *Regional Studies*, 48(1):44–67.
- Crespo Cuaresma, J. and Feldkircher, M. (2013). Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe. *Journal of Applied Econometrics*, 28(4):720–741.
- Dapena, A. D., Morollón, F. R., Pires, M. d. M., and Gomes, A. d. S. (2017). Convergence in Brazil: New Evidence Using a Multilevel Approach. *Applied Economics*, 49(50):5050–5062.
- Dapena, A. D., Vázquez, E. F., and Morollón, F. R. (2016). The Role of Spatial Scale in Regional Convergence: The Effect of MAUP in the Estimation of β -Convergence Equations. *The Annals of Regional Science*, 56(2):473–489.
- Dapena, A. D., Vázquez, E. F., and Morollón, F. R. (2018). Labor Density and Wages in Spain: Evidence from Geographically Disaggregated Data. *Growth and Change*, 49(1):55–70.
- Delgado, M., Porter, M. E., and Stern, S. (2016). Defining Clusters of Related Industries. *Journal of Economic Geography*, 16(1):1–38.
- Deller, S. C., Conroy, T., and Watson, P. (2017). Women Business Owners: a Source of Stability during the Great Recession? *Applied Economics*, 49(56):5686–5697.
- Deller, S. C. and Watson, P. (2016a). Did Regional Economic Diversity Influence the Effects of the Great Recession? *Economic Inquiry*, 54(4):1824–1838.
- Deller, S. C. and Watson, P. (2016b). Spatial Variations in the Relationship between Economic Diversity and Stability. *Applied Economics Letters*, 23(7):520–525.
- Desrochers, P. and Sautet, F. (2008). Entrepreneurial Policy: The Case of Regional Specialization vs. Spontaneous Industrial Diversity. *Entrepreneurship Theory and Practice*, 32(5):813–832.

- Dissart, J.-C. (2003). Regional Economic Diversity and Regional Economic Stability: Research Results and Agenda. *International Regional Science Review*, 26(4):423–446.
- Dong, G., Harris, R., Jones, K., and Yu, J. (2015). Multilevel Modelling with Spatial Interaction Effects with Application to an Emerging Land Market in Beijing, China. *PLoS ONE*, 10(6):e0130761–19.
- Duffy, N. E. (1994). The Determinants of State Manufacturing Growth Rates: A Two-Digit-Level Analysis. *Journal of Regional Science*, 34(2):137–162.
- Duque, J. C., Anselin, L., and Rey, S. J. (2012). The Max-P-Regions Problem. *Journal of Regional Science*, 52(3):397–419.
- Duque, J. C., Artís, M., and Ramos, R. (2006). The Ecological Fallacy in a Time Series Context: Evidence from Spanish Regional Unemployment Rates. *Journal of Geographical Systems*, 8(4):391–410.
- Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2011). Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants. *Journal of Applied Econometrics*, 26(1):30–55.
- Elhorst, J. P. (2014). Matlab Software for Spatial Panels. *International Regional Science Review*, 37(3):389–405.
- Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in US Manufacturing Industries: a Dartboard Approach. *Journal of Political Economy*, 105(5):889–927.
- Farmer, C. J. Q. and Fotheringham, A. S. (2011). Network-Based Functional Regions. *Environment and Planning A*, 43(11):2723–2741.
- Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100(2):381–427.
- Fernández, C., Ley, E., and Steel, M. F. J. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16(5):563–576.

- Feser, E., Mix, T., White, M., Poole, K., Markley, D., and Pages, E. (2014). Economic Diversity in Appalachia: Statistics, Strategies, and Guides for Action. *Washington DC: Appalachian Regional Commission.*
- Florax, R. J., Folmer, H., and Rey, S. J. (2003). Specification Searches in Spatial Econometrics: the Relevance of Hendry's Methodology. *Regional Science and Urban Economics*, 33(5):557–579.
- Fotheringham, A. S., Brunson, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- Fox, K. A. and Kumar, T. K. (1965). The Functional Economic Area: Delineation and Implications for Economic Analysis and Policy. *Papers of the Regional Science Association*, 15(1):57–85.
- Frenken, K., Van Oort, F., and Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, 41(5):685–697.
- Fu, S., Dong, X., and Chai, G. (2010). Industry Specialization, Diversification, Churning, and Unemployment in Chinese Cities. *China Economic Review*, 21(4):508–520.
- Fujita, M., Krugman, P. R., Venables, A. J., and Fujita, M. (1999). *The Spatial Economy: Cities, Regions and International Trade*, volume 213. Wiley Online Library.
- Getis, A. and Griffith, D. A. (2002). Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis*, 34(2):130–140.
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126–1152.
- Goodchild, M. F. (2004). GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers*, 94(4):709–714.
- Goodchild, M. F. (2011). Scale in GIS: An Overview. *Geomorphology*, 130(1):5–9.

- Goodchild, M. F., Anselin, L., Appelbaum, R. P., and Harthorn, B. H. (2000). Toward Spatially Integrated Social Science. *International Regional Science Review*, 23(2):139–159.
- Harvey, D. W. (1968). Pattern, Process, and the Scale Problem in Geographical Research. *Transactions of the Institute of British Geographers*, 45:71–78.
- Hewings, G. J. D. (1982). The Empirical Identification of Key Sectors in an Economy: A Regional Perspective. *The Developing Economies*, 20(2):173–195.
- Hong, S. and Xiao, Y. (2016). The Influence of Multiple Specializations on Economic Performance in U.S. Metropolitan Areas. *Sustainability*, 8(9):963.
- Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39(153):41–57.
- Isard, W. (1956). *Location and Space-Economy*. MIT Press.
- Isserman, A. M. and Westervelt, J. (2006). 1.5 Million Missing Numbers: Overcoming Employment Suppression in County Business Patterns Data. *International Regional Science Review*, 29(3):311–335.
- Izraeli, O. and Murphy, K. J. (2003). The Effect of Industrial Diversity on State Unemployment Rate and Per Capita Income. *The Annals of Regional Science*, 37(1):1–14.
- Jackson, R. W. (1984). An Evaluation of Alternative Measures of Regional Industrial Diversification. *Regional Studies*, 18(2):103–112.
- Jackson, R. W. (2015). Are Industry Clusters and Diversity Strange Bedfellows? *Review of Regional Studies*, 45:113–129.
- Jackson, R. W. and Sonis, M. (2001). On the Spatial Decomposition of Forecasts. *Geographical Analysis*, 33(1):58–75.
- Jacobs, J. (1969). *The Economy of Cities*. Vintage Books.
- Johnson, K. P. (1995). Redefinition of the BEA Economic Areas. *Survey of Current Business*, 75:75–81.

- Johnson, K. P. and Kort, J. R. (2004). 2004 Redefinition of the BEA Economic Areas. *Survey of Current Business*, 84(11):68–75.
- Keinath, W. F. (1985). The Spatial Component of the Post-Industrial Society. *Economic Geography*, 61(3):223–240.
- Kort, J. R. (1981). Regional Economic Instability and Industrial Diversification in the US. *Land Economics*, 57(4):596–608.
- Kropp, P. and Schwengler, B. (2016). Three-Step Method for Delineating Functional Labour Market Regions. *Regional Studies*, 50(3):429–445.
- Krugman, P. R. (1991a). History and Industry Location: The Case of the Manufacturing Belt. *The American Economic Review*, 81(2):80–83.
- Krugman, P. R. (1991b). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499.
- Krugman, P. R. (1998). What’s New about the New Economic Geography? *Oxford Review of Economic Policy*, 14(2):7–17.
- Krugman, P. R. (1999). The Role of Geography in Development. *International Regional Science Review*, 22(2):142–161.
- Kwan, M.-P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*, 102(5):958–968.
- Lacombe, D. J. (2008). An Introduction to Bayesian Inference in Spatial Econometrics. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1244261. SSRN Working Paper. Retrieved November 15th, 2017.
- Lacombe, D. J. and McIntyre, S. G. (2017). Hierarchical Spatial Econometric Models in Regional Science. In Jackson, R. W. and Schaeffer, P. V., editors, *Regional Research Frontiers - Vol. 2*, pages 151–167. Springer, Cham, Cham, Switzerland.

- Le Gallo, J. and Ertur, C. (2003). Exploratory Spatial Data Analysis of the Distribution of Regional per Capita GDP in Europe, 1980-1995. *Papers in Regional Science*, 82(2):175–201.
- LeSage, J. P. and Fischer, M. M. (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis*, 3(3):275–304.
- LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.
- LeSage, J. P. and Pace, R. K. (2014). The Biggest Myth in Spatial Econometrics. *Econometrics*, 2(4):217–249.
- LeSage, J. P. and Parent, O. (2007). Bayesian Model Averaging for Spatial Econometric Models. *Geographical Analysis*, 39(3):241–267.
- Ley, E. and Steel, M. F. J. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression. *Journal of Applied Econometrics*, 24(4):651–674.
- Mack, E. A., Grubestic, T. H., and Kessler, E. (2007). Indices of Industrial Diversity and Regional Economic Composition. *Growth and Change*, 38(3):474–509.
- MacKay, R. J. and Ordford, R. W. (2000). Scientific Method, Statistical Method and the Speed of Light. *Statistical Science*, 15(3):254–278.
- Madigan, D., York, J., and Allard, D. (1995). Bayesian Graphical Models for Discrete Data. *International Statistical Review*, 63(2):215–232.
- Malizia, E. E. and Ke, S. (1993). The Influence of Economic Diversity on Unemployment and Stability. *Journal of Regional Science*, 33(2):221–235.
- Marshall, A. (1890). *Principles of Economics*. Macmillan, New York, 1st edition.
- Martin, R. and Sunley, P. (2003). Deconstructing Clusters: Chaotic Concept or Policy Panacea? *Journal of Economic Geography*, 3(1):5–35.

- Martin, R. and Sunley, P. (2006). Path Dependence and Regional Economic Evolution. *Journal of Economic Geography*, 6(4):395–437.
- McLaughlin, G. E. (1930). Industrial Diversification in American Cities. *The Quarterly Journal of Economics*, 45(1):131–149.
- Mizuno, K., Mizutani, F., and Nakayama, N. (2006). Industrial Diversity and Metropolitan Unemployment Rate. *The Annals of Regional Science*, 40(1):157–172.
- Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J., and Olander, L.-O. (2011). The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. *Regional studies*, 45(1):49–65.
- Nystuen, J. D. and Dacey, M. F. (1961). A Graph Theory Interpretation of Nodal Regions. *Papers of the Regional Science Association*, 7(1):29–42.
- ÓhUallacháin, B. (1984). The Identification of Industrial Complexes. *Annals of the Association of American Geographers*, 74(3):420–436.
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. GeoBooks, Norwich, UK.
- Openshaw, S. and Taylor, P. J. (1979). A Million or So Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In Wrigley, N., editor, *Statistical Applications in the Spatial Sciences*, pages 127–144. Pion, London, UK.
- O’Sullivan, D. and Unwin, D. (2014). *Geographic Information Analysis*. John Wiley & Sons.
- Ouyang, Y., Wentz, E. A., Ruddell, B. L., and Harlan, S. L. (2014). A MultiScale Analysis of SingleFamily Residential Water Use in the Phoenix Metropolitan Area. *Journal of the American Water Resources Association*, 50(2):448–467.
- Parent, O. and LeSage, J. P. (2012). Determinants of Knowledge Production and Their Effects on Regional Economic Growth. *Journal of Regional Science*, 52(2):256–284.
- Parr, J. B. (2002). Missing Elements in the Analysis of Agglomeration Economies. *International Regional Science Review*, 25(2):151–168.

- Pattison, W. D. (1964). The Four Traditions of Geography. *Journal of Geography*, 63(5):211–216.
- Peeters, L. and Chasco, C. (2006). Ecological Inference and Spatial Heterogeneity: An EntropyBased Distributionally Weighted Regression Approach. *Papers in Regional Science*, 85(2):257–276.
- Perroux, F. (1950). Economic Space: Theory and Applications. *The Quarterly Journal of Economics*, 64(1):89–104.
- Philbrick, A. K. (1957). Principles of Areal Functional Organization in Regional Human Geography. *Economic Geography*, 33(4):299–336.
- Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, 68(2):73–93.
- Porter, M. E. (1998). *On Competition*. Harvard Business Press.
- Porter, M. E. (2003). The Economic Performance of Regions. *Regional Studies*, 37(6-7):545–546.
- Purdue Center for Regional Development (2007). Unlocking Rural Competitiveness: The Role of Regional Clusters. Available at: <https://pcrd.purdue.edu/index.php>. Retrieved November 15th, 2017.
- Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, 92(437):179–191.
- Resende, G. M. (2011). Multiple Dimensions of Regional Economic Growth: The Brazilian Case, 1991- 2000. *Papers in Regional Science*, 90(3):629–662.
- Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15(3):351–357.
- Rocha, H. O. (2004). Entrepreneurship and Development: The Role of Clusters. *Small Business Economics*, 23(5):363–400.

- Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5):1002–1037.
- Ruddell, D. and Wentz, E. A. (2009). Multi-Tasking: Scale in Geography. *Geography Compass*, 3(2):681–697.
- Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94(4):813–835.
- Sala-i-Martin, X. X. (1997). I Just Ran Four Million Regressions. *American Economic Review*, 87(2):178–183.
- Schaeffer, P. V., Jackson, R. W., and Bukenya, J. O. (2011). Regional Science Reconsidered. *The Review of Regional Studies*, 41:161–177.
- Schultz, S. (1977). Approaches to Identifying Key Sectors Empirically by Means of Input-Output Analysis. *The Journal of Development Studies*, 14(1):77–96.
- Siegel, P. B., Johnson, T. G., and Alwang, J. (1995). Regional Economic Diversity and Diversification. *Growth and Change*, 26(2):261–284.
- Simon, C. J. (1988). Frictional Unemployment and the Role of Industrial Diversity. *The Quarterly Journal of Economics*, 103(4):715–728.
- Slaper, T., Harmon, K., and Rubin, B. (2018). Industry Clusters and Regional Economic Performance: A Study Across US Metropolitan Statistical Areas. *Economic Development Quarterly*, 32(1):44–59.
- Spencer, G. M., Vinodrai, T., and Gertler, M. S. (2010). Do Clusters Make a Difference? Defining and Assessing Their Economic Performance. *Regional Studies*, 44(6):697–715.
- Stock, J. and Watson, M. W. (2007). *Introduction to Econometrics, 2nd ed.* Pearson Education, Inc., Boston, MA.
- Tian, Z. (2013). Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method. *Journal of Regional Analysis and Policy*, 43(2):186.

- Tiebout, C. M. (1956). The Urban Economic Base Reconsidered. *Land Economics*, 32(1):95–99.
- Tiefelsdorf, M. and Griffith, D. A. (2016). Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A*, 39(5):1193–1221.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1):234–240.
- Tobler, W. R. (1988). Resolution, Resampling, and All That. *Building Databases for Global Science*, 12:9–137.
- Trendle, B. (2006). Regional Economic Instability: The Role of Industrial Diversification and Spatial Spillovers. *The Annals of Regional Science*, 40(4):767–778.
- Tress, R. C. (1938). Unemployment and the Diversification of Industry. *The Manchester School*, 9(2):140–152.
- Venables, A. J. (1996). Equilibrium Locations of Vertically Linked Industries. *International Economic Review*, 37(2):341–359.
- Wagner, J. E. (2000). Regional Economic Diversity: Action, Concept, or State of Confusion. *Journal of Regional Analysis & Policy*, 30(2):1–22.
- Wagner, J. E. and Deller, S. C. (1998). Measuring the Effects of Economic Diversity on Growth and Stability. *Land Economics*, 74(4):541–556.
- Wang, F. and O’Brien, V. (2005). Constructing Geographic Areas for Analysis of Homicide in Small Populations: Testing Herding-Culture-of-Honor Proposition. In Wang, F., editor, *Geographic Information Systems and Crime Analysis*, pages 84–101. IGI Global.
- Watson, M. K. (1978). The Scale Problem in Human Geography. *Geografiska Annaler Series B Human Geography*, 60(1):36–47.
- Watson, P. and Deller, S. (2017). Economic Diversity, Unemployment and the Great Recession. *The Quarterly Review of Economics and Finance*, 64:1–11.

- Weber, A. (1929). *Theory of the Location of Industries [translated by CJ Friedrich from Weber's 1909 book]*. The University of Chicago Press.
- Winkler, R., Deller, S. C., and Marcouiller, D. (2015). Recreational Housing and Community Development: A Triple Bottom Line Approach. *Growth and Change*, 46(3):481–500.
- Yu, J. and Jackson, R. W. (2011). Regional Innovation Clusters: A Critical Review. *Growth and Change*, 42(2):111–124.
- Zellner, A. (1986). On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. In Goel, P. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier Science Publishers, Inc.