5-2011

Industrial Diversity and Economic Performance: A Spatial Analysis

Hoa Phu Duy Tran
University of Nebraska-Lincoln

INDUSTRIAL DIVERSITY AND ECONOMIC PERFORMANCE: A SPATIAL ANALYSIS

By

Hoa Phu D. Tran

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Economics

Under the Supervision of Professor James R. Schmidt

Lincoln, Nebraska

May 2011

INDUSTRIAL DIVERSITY AND ECONOMIC PERFORMANCE:
A SPATIAL ANALYSIS

Hoa Phu D. Tran, Ph.D.

University of Nebraska, 2011

Advisor: James R. Schmidt

This study examines the linkage between industrial diversity and economic growth in the 48 contiguous states of the United States. The period of analysis is 1992 through 2009. Five diversity indices are considered and economic growth is measured as the growth rate of nonfarm earnings. Other variables thought to influence economic growth are included in the analysis. They are the growth rate of nonfarm employment, capital, and farm earnings. Tests for the endogeneity of variables are conducted and the need for instrumental variable estimation methods is demonstrated.

First, I consider multivariate model that relates nonfarm earnings growth to the diversity indices and the other variables noted above. The model includes regional fixed effects and time effects but does not allow for spatial dependence among states. The results show that diversity positively influences economic growth. Growth in nonfarm employment and capital are also found to be positively influence economic growth.

Second, I consider two spatial models that allow for a spatial lag and spatial autocorrelation effects among states. The first spatial model assumes a common spatial lag parameter for all states. The second spatial model allows the spatial lag parameter to be unique for each of eight regions within the United States. Two estimation methods are used, the generalized spatial two-state least squares estimator and an instrumental variables estimator along with a spatial heteroskedasticity and

autocorrelation consistent matrix estimator.

The spatial lag parameter is small and statistically insignificant when the parameter is assumed to be the same across regions. However, when the spatial lag parameter is allowed to vary across regions, spatial effects among states are detected and are reasonably strong in some regions. Under both estimation methods for both spatial models, the results provide strong evidence that states with higher levels of diversity experience higher growth rates in nonfarm earnings. Nonfarm employment growth and capital growth are also significant influences upon the growth rate of nonfarm earnings.

Acknowledgements

I owe my deepest gratitude to my advisor, Dr. James R. Schmidt, for his endless encouragement, guidance and support during this research. He has spent numerous times reading my drafts and given me suggestions on how to make it better. It would have been next to impossible to write this dissertation without Dr. James R. Schmidt help and guidance.

My sincere gratitude must also go out to all members of my dissertation committee: Dr. Eric Thompson, Dr. Hendrik Van den Berg, and Dr. John M. Geppert. They had provided tremendous support during this research. I also would like to express my gratitude to Dr. Matthew J. Cushing and Dr. Mary G. McGarvey for their support and encouragement.

I also would like to express my sincere thanks to David Dearmont, Ph.D., an Economist at the Nebraska Department of Revenue for his support and encouragement in this research project.

I also own a debt of gratitude to my wife, Ly Tran, and my parents, Hien Tran and Cham Nguyen, for their unconditional support and encouragement throughout this journey. Without their sacrifices, I would not have able to complete this dissertation.

Table of Contents

List of Tables

Industrial Diversity and Economic Performance: How Strong is the Link? A Spatial Analysis

1. Introduction

The goal of this study is to determine whether the assumption that industrial diversity enhances economic performance holds for regions, and more importantly, to examine how strong is the contribution of high degrees of industrial diversity to regional economic growth.

Industry diversity has long been a research topic in the regional economics literature. The purpose of this study is to investigate what role industry diversity plays in regional growth. Industrial diversity refers to the variety of economic activities that reflect differences in economic structure (Maliza and Ke 1993). If a regional economy is diversified in its economic structure, it may be less affected by an economic downturn. It has been widely assumed that industrial diversity enhances economic performance, the latter being measured by growth rates, per capita income, unemployment rates, or other economic performance indicators.

For more than 200 years traditional economic theory has suggested that specialization permits people to use skills and resources to their best advantage and enables exchange for goods and services which enhances economic growth. The notion of comparative advantage implies that growth requires specialization on industries that the regions have comparative advantage over other regions, which is the opposite of industrial diversification of the economy. Theory also suggests that stability is achieved through diversity. Therefore, theory seems to suggest that there is a tradeoff that regional policy makers are faced with. Regional policy makers can choose to either set policies toward specialization which enhances growth and face the possibility of high volatility in the economy or set policies toward diversifying the economy which gives stability but may lessen the chances of higher economic growth. This tradeoff is one that most, if not all, regional policymakers are unwilling to accept or act upon when designing region economic and fiscal policy. According to the economic theory given above, when policymakers attempt to pursue growth and stability simultaneously, contradictions seem to appear.

Wagner and Deller (1998) suggested that the simultaneous pursuit of growth and stability is not contradictory when viewed in terms of the short-run and long-run. They suggested that short-run policy can be viewed as growth oriented while long-run policy can be viewed as oriented toward regional stability. For short-run goals, policymakers can develop policies that target growth industries. Short-run policies that would promote employment and investment can capitalize on the region's comparative advantage in a few specialized industries that play major roles in leading the region's economic growth. Policies targeting growth in a few specialized industries is only half of the equation, the other half is to promote economic stability in the regions. Policymakers cannot only rely on short-run goals and results which may create a trap where policymakers will not look to consider policies for the long run. It is easy for regional policymakers to fall into the syndrome of "as long as it is not broken under my watch" and just focus on short-run policies that only target growth industries. This can be dangerous because as targeted industries mature and exploit comparative advantage to the highest level, a dampening pressure on growth will develop. Furthermore, if policies for targeted industries fail, the region may be worse off than before the policies were implemented. Thus, long-run policies should be implemented toward diversification in the region to achieve regional economic stability. It is important for policymakers to remember that short-run policies are aimed at promoting growth and long-run policies are aimed at promoting stability. As stability and diversity increase, the potential for economic growth also increases. If regional policymakers can focus on both short-run and long-run policies for the region, then regional growth and stability can be pursued simultaneously.

Specialized economies face high risks when faced with external shocks that affect the specialized industries. For a specialized regional economy, when external shocks affect specialized industries, employees will be laid off, leading to high unemployment in the region since it will be difficult for laid-off employees to find new jobs. Unemployment rates rise, resulting in lower economic performance for the region.

Since traditional economic theory suggests that growth requires specialization, then why do regional scientists widely assume that industrial diversity contributes positively to economic growth? The reasoning is straightforward. As a region becomes more diversified, it becomes less sensitive to fluctuations caused by factors outside the region (Nourse 1968; Richardson 1969). Diversity positions the economic base so that the region can absorb varieties of structural changes in the national economy, for example, changes in national policies concerning international trade. Reliance upon a small group of specialized industries for the majority of regional income is risky due to contractions or reductions in demand for certain goods in an industry (Kort, 1981). Many unpredictable events can cause the demand for goods and services in a specialized industry to shift such as business cycles, policy changes concerning the environment and shocks to trade patterns. Business cycles involve unpredictable shifts over time between periods of rapid economic growth and periods of stagnation or decline which causes shifts in the demand for goods in specialized industries. Environmental policies enacted to reduce emissions of carbon dioxide could tend to put upward pressure on costs of production in a specialized industry, thereby causing the production to fall. Policies concerning international trade such as tariffs and quotas could definitely affect demand conditions. Also, economic conditions at home or abroad that require an expansion or contraction of monetary policy could affect the demand for final goods and could also make imported goods used in the process of production more expensive, thereby leading to a decline in production. The immediate result of a decline in production is the laying off of employees. The laid-off employees may be unable to find alternative jobs in the region if the region is too specialized.

A diverse industrial structure allows the regional economy to respond to more growth opportunities, rather than rely on only a few industries in a specialized structure. A diverse industrial structure provides better employment opportunities and creates more high-paying jobs for the region which attracts and retains highly skilled individuals that will contribute significantly to economic growth. The greater the variety of industries in a region and the more dispersed the regional employment among these industries, the less likely a region is to suffer severe economic decline.

In this study, I focus upon the relationship between regional industrial diversity and economic growth with the 48 contiguous states being used as the regional economies. The time span of analysis is 1992-2009, a period that contained strong economic growth in many states through 2000. From 2001-2003, economic slowdowns occurred in most states with declines in several, then followed by strong growth until the recession that began in late 2007. States are chosen as the regions for the study because many comprehensive policies are enacted at the state level, that is, policies intended to lead a state's economy toward either a diversified industrial structure or a specialized industrial structure.

Chapter 2 offers a survey of theoretical and empirical studies of the relationships between industrial diversity, regional economic growth, and economic instability. The vast majority of the studies support the hypothesis that regional diversity reduces regional economic instability and unemployment. In terms of economic growth, the literature contains mixed results. Until recently, a glaring omission in the literature has been the treatment of spatial correlation. Izraeli and Murphy (2003) and Trendle and Shorney (2004) mention and partially attempt to correct for spatial correlation but not much detail is provided. Recently, Garrett, Wagner, and Wheelock (2007) gave a much more detailed treatment of spatial correlation in a study of state income growth using data from 1977 to 2002. This study will also give careful attention to spatial correlation as well as the endogeneity issue of variables that are used in models of growth. In Chapter 3, I discuss five commonly used industrial diversity indices and describe the data collection process. All five diversity indices are employment-based indices. Chapter 4 is divided into two sections. The first section is devoted to the discussion of multivariate models that relate industrial diversity to economic growth. The second section is devoted to the discussion of spatial models that relate industrial diversity to economic growth. Chapter 5 presents the results for non-spatial and spatial models. Chapter 6 summarizes the findings and offers conclusions.

2. Industrial Diversity and Economic Performance

The linkage between industrial diversity and economic growth has been an important topic in regional economics for many years. The issue of stability of growth also arises when the roles of industrial diversity are examined. The regional literature offers the hypothesis that more diverse areas should experience more stable economic growth and less unemployment compared to less diverse regions. Past research generally shows that industrial diversity promotes growth stability for the region, whereas the link between diversity and levels of growth has been more elusive. Some studies have found significant relationships between diversity and regional economic growth while others have found no relationship. As Wagner and Deller suggested in their 1998 study, the inconsistency of empirical results may be due to small sample sizes or highly aggregated data sets.

2.1. Entropy Index and Per Capita Income

Attaran (1986) explored the issue of industrial diversity and economic performance in U.S. areas and found a negative correlation between diversity and the growth rate of per capita income for the 50 states and District of Columbia during the 10-year period from 1972-1981. This is an unexpected result given the logic outlined above to support our hypothesis.

In the Attaran study, the Entropy function is used as a measure of economic diversity, defined as:

$$D(E_1, E_2, \dots E_n) = -\sum_{i=1}^{n} E_i \, log_2(E_i)$$

where n is the number of economic sectors and E_i is the proportion of total employment of the region that is located in the i^{th} sector. The aim of Attaran's study was to examine the significance of economic diversity, more specifically, to determine whether diversity is correlated with economic performance. Attaran assessed economic performance in terms of two economic variables, unemployment and per capita income. The entropy indices were calculated based on employment data for eight nonagricultural sectors[1].

To test for the existence of a negative relationship between diversification and unemployment, the diversity indices of the 51 study areas for the years 1972 to 1981 were correlated with their corresponding unemployment rates. By doing this, Attaran found that the correlation coefficients were negative as expected, but none of the coefficients were significant. Furthermore, Attaran aggregated the data for all states over the 10-year period and conducted correlation tests. The analysis produced a correlation coefficient of -0.11, indicating an extremely weak but statistically significant negative correlation between diversity and unemployment.

Attaran also tested for the existence of correlation between diversity and per capita income for the same 10-year period. To assess this association statistically, diversity indices of the 51 study areas were correlated with their corresponding per capita incomes. Attaran used the logarithmic form of per capita income in constant dollars with 1967 as the base year. From the correlation tests for each individual year, Attaran concluded that the correlation coefficients for diversity measures and real per capita income were statistically significant for all the years, but the coefficients were negative. Negative correlation implies that lower diversity is associated with higher

[1] The sectors are: durable goods; nondurable goods; construction; transportation; communication and utilities; trade, finance, insurance, and real estate; service and miscellaneous; and government.

levels of per capita income; that is, more specialized regions tend to have higher levels of per capita income than diversified regions.

To further support the result noted above, Attaran found the correlation of the mean diversity indices of 51 study areas and corresponding means of unemployment for the 10-year period to be -0.12, indicating an insignificant relationship. Attaran also found the correlation between mean diversity and mean per capita income for the period of the study to be -0.47, providing evidence of a significant relationship.

The results found by Attaran may have been due to the use of employment-based measures of diversity in tandem with per capita income, a highly aggregated measure of economic activity that includes much more than labor income. Also, the diversity index was based on eight industries which represented a highly aggregated data set. Use of a highly aggregated data set together with significant structural changes in the U.S. economy during the 1972-1981 time span, along with two major inflationary bouts, could have contributed to the detection a of negative relationship between diversity and economic performance. Also, the study by Attaran did not address the econometric problem of omitted variables. Attaran only calculated the correlation between the two variables of interest: diversity indices and unemployment; and diversity indices and the logarithm of real per capita income.

2.2. Input-Output Diversity Index

A study by Wagner and Deller (1998) suggests that higher levels of diversity are statistically associated with higher levels of economic growth as measured by changes in per capita income using averaged data over the long time span of 1969-1991. They also found a negative relationship between economic diversity and a stability measure, that is, higher levels of economic diversification are associated with lower levels of economic instability.

Wagner and Deller considered an alternative approach to conceptualizing diversity, based on a regional input-output model for the 50 states. They implemented the approach by using the regional modeling system MicroIMPLAN (Alward et al. 1989) to construct 51 separate input-output models for each of the 50 states plus the entire U.S. They constructed the diversity measure based on three scalars that describe the regional input coefficients matrix of an input-output model. The first scalar, SI_i, is the measure of the size of the economy, the second scalar, DEN_i, measure the degree of industry imports and the third scalar, C_i, captures the flow of locally produced inputs between endogenous industries.[2] The diversity index is defined as a combination of these three components and they considered both the additive form,

$$ADI_i = (w_1 * SI_i) + (w_2 * DEN_i) + (w_3 * C_i) \sum_{j=1}^{3} w_j,$$

and multiplicative form,

$$MDI_i = SI_i * DEN_i * C_i,$$

of the index. The higher the value of the indices, the higher the degree of diversity for the region.

Wagner and Deller considered two empirical models to test the hypothesis that higher levels of economic diversity result in higher levels of economic stability and growth:

(WD.1) Growth = f(Market, Labor, Taxes, Amenity, Infrastructure, DI)

[2] See Wagner and Deller (1998) for detail discussion of the three scalars: SI, DEN, C.

and

(WD.2) Stability = g(Market, Labor, Taxes, Amenity, Infrastructure, DI).

Instead of assuming linear or logarithmic functional forms for the growth and stability equation, Wagner and Deller employed the Box-Cox estimator.

For the measure of growth and stability, Wagner and Deller chose two characteristics of the regional economy: the unemployment rate and per capita income. The measure of economic performance is defined for the i^{th} region as the average annual growth rate over the period examined:

$$AYGR_i = \frac{\sum_{t=2}^{T}\left[\frac{Y_{it} - Y_{it-1}}{Y_{it-1}}\right] * 100}{T - 1}$$

where T is the number of periods examined and Y is the characteristic of the economy. For their stability measure, they used the variance in the average annual unemployment rate for the same time period.

The results from the Box-Cox estimation of (WD.1) and (WD.2) suggested that the functional form of (WD.1) and (WD.2) is nonlinear since the Box-Cox lambda value equaled 0.63 in the growth equation and 0.41 in the stability equation. The infrastructure variable had a positive and significant coefficient indicating that higher levels of infrastructure stock are associated with higher economic growth. Their main result was the statistical importance of the diversity measures. After accounting for several growth promoting factors, their evidence supports the notion that higher levels of economic diversity lead to higher levels of economic growth as measured by percent changes in per capita income.

Wagner and Deller found a negative and weakly significant coefficient for the diversity measure in the stability equation. The negative coefficient suggested that higher levels of economic diversity are associated with lower levels of economic instability. However, the models explain very small proportions of total variation, about 17 percent in the growth model and about 16 percent in the stability model. The study also measured economic performance as the average growth rate of per capita income which is a measure that includes much more than just labor income, such as capital gains, dividends, and interest which may not have close relationships with employment. Interest, rents, capital gains, and dividends are income components based on place of residence rather than place of work. Also, when the growth rate of per capita income is an average over a 23-year period (1969-1991), it may smooth out the series and perhaps lead to a statistically significant result for the diversity coefficient. Thus, other time spans of data need to be considered.

2.3. Herfindahl Diversity Index, Unemployment, and Per Capita Income

Along the line of Wagner and Deller's study was a study done by Izraeli and Murphy (2003). They also examined the effect of industrial diversity on state unemployment rates and per capita incomes. Izraeli and Murphy hypothesized that well-diversified regional economies should experience lower unemployment rates but well-diversified regions should experience lower per capita incomes compared to regions that have greater industrial concentration (less diversity). By using two panel data sets, one for unemployment and one for per capita income, they found that there exists a strong link between industrial diversity and lower unemployment while the results showed that per capita income is weakly associated with diversity.

Izraeli and Murphy considered two models, one to test for the link between industrial diversity and personal income and one to test for the link between diversity and unemployment. The two models adopted by Izraeli and Murphy have a linear functional form and are defined as:

(IM.1)

$$U_{it} = U(DIV_{it}, USU_t, RPIC_{it}, DEN_{it}, NWT_{it}, TEEN_{it}, OVER65_{it}, POP_{it}, POPCH_{it})$$

(IM.2)

$$RPIC_{it} = RPIC(DIV_{it}, RPICUS_t, DEN_{it}, NWT_{it}, TEEN_{it}, OVER65_{it}, POP_{it}, POPCH_{it})$$

where:

where:	U	- state unemployment rate
	DIV	- measure of the degree of industrial diversity.
	USU	- national unemployment rate
	$RPIC$	- state per capita income (in 1982 dollars)
	DEN	- population density
	NWT	- percent of working-age population that is non-white
	$TEEN$	- percent of working-age population that is 16-19 years of age
	$OVER65$	- the percent of the population 65 years and older
	POP	- the state population
	$POPCH$	- the rate of population growth in a state
	$RPICUS$	- national per capita income (in 1982 dollars); and i and t stand for state i and year t.

Izraeli and Murphy used the Herfindahl index as the measure for industrial diversity. The Herfindahl index for state i at time t is given by

$$DIV_{it} = \sum_{j=1}^{n} (\frac{EMP_{ijt}}{EMP_{it}})^2$$

where EMP_{ijt} is employment in state i in industry j in year t, EMP_{it} is the total state employment in year t, and n is the number of industries in state i in year t. Based on this formulation of the index, the higher the value the less diverse the state's economy would be.

The two dependent variables used by Izraeli and Murphy were the annual unemployment rate and annual per capita personal income (in 1982 dollars). Izraeli and Murphy used two sets of data, the first included 17 states but the time series length for each individual state depended on availability of the data during the span 1960 to 1977.[3] The second series included all annual data during the span 1988-1997 for all 17 states. Izraeli and Murphy considered the level of per capita income instead of the growth rate. By using levels of per capita income over time, Izraeli and Murphy could have encountered the problem of unit roots. The time series may be nonstationary. The authors did not mention the issue of nonstationarity and did not test for unit roots.

Izraeli and Murphy took into account two econometric problems when estimating the (IM.1) and (IM.2) equations. The first problem is omitted variable bias. To address the problem, they considered a more general form for equations (IM.1) and (IM.2):

$$\text{(IM.3)} \qquad y_{it} = X_{it}\beta + \delta_i + \varepsilon_{it}$$

where δ_i is an unobserved fixed effect specific to state i. The fixed effect term, δ_i, is intended to capture idiosyncratic factors specific to a state that are unobservable. The second problem is spatial correlation among the states.

To deal with the problems of omitted variables and spatial correlation, Izraeli and Murphy utilized a two-step procedure. In the first step, they estimated the

[3] See Izraeli and Murphy (2003) for detail on length of spans for individual states.

regression model with fixed effects, saved the residuals and generated a contemporaneous covariance matrix for the model. In the second stage, they re-estimated the fixed effects model using feasible generalized least squares.[4]

Izraeli and Murphy noted that they only partially addressed the spatial correlation issue. They used a restricted version of feasible generalized least squares to partially solve the problem of spatial correlation due to lack of availability of the data. The second sample consisted of 17 states but only 10 time periods, thus it was impossible to generate a nonsingular contemporaneous correlation matrix that would take possible correlation among all of the states into account. In order to get a singular correlation matrix, they grouped the states into four major census regions and came up with a correlation matrix that was invertible. The grouping of 17 states into four regions forced the data set to be highly aggregated which may have produced a biased estimate of the effect of industrial diversity upon economic growth. With respect to the first time span, feasible generalized least squares was not used due to non-uniformity of the sample periods among the states.

After Izraeli and Murphy adjusted for omitted variable bias and spatial correlation, their estimation results for the unemployment rate equation showed that higher degrees of industrial diversity tend to be associated with lower unemployment rates. It is important to note that when Izraeli and Murphy performed the analysis without taking correlation among neighboring states into account, the coefficient on the Herfindahl diversity index was only weakly significant during 1987-1997. When they

---

[4] $\hat{\beta} = (X'(\hat{\Sigma}^{-1} \otimes \mathrm{I})\mathrm{X})^{-1} * (X'(\hat{\Sigma}^{-1} \otimes \mathrm{I})\mathrm{y})$ where $\hat{\Sigma}$ is the contemporaneous cross-correlation matrix estimated with first-stage residuals.

took spatial correlation into account, their analysis revealed the effect of the industrial diversity upon unemployment rate to be highly significant.[5] On the other hand, estimates from the per capita income equation showed mixed results for the 1960-1977 and 1988-1998 spans. For the 1960-1977 spans, the results showed that the Herfindahl diversity index positively correlates with per capita income. That is, the results showed that per capita income is affected negatively by diversity. For the second time span, 1988-1998, the results showed no evidence of a significant relationship between the Herfindahl Index and per capital income.

By using a 10-year span for the latter part of the study, Izraeli and Murphy did not have a long enough panel data set to fully account for spatial correlation. As a result, they grouped the 17 states into four regions for their spatial correlation adjustment. The grouping process may present another issue, that is, bias in the selection process. Not only might bias occur in grouping 17 states into four groups, but the process of picking 17 states to begin with also makes the study less than comprehensive in nature. The fact that the state data in the first time span have different lengths also poses a serious problem. In the period from 1960 to 1987, the U.S. as a whole went through several recessions with a pair of severe ones in 1973 and 1981[6]. If the data isn't uniform in length, the characteristics of the data for each state may differ if they include one or more recessions and other states do not.

Izraeli and Murphy considered a general spatial error correlation model but one

[5] Without allowing for spatial correlation, the t-statistic for DIV was 1.63. When allowing for spatial correlation, the t-statistic for DIV was 3.04.
[6] Recessions of 1960, 1969-1970, 1973-1974, and 1981-1982.

might suspect that the per capita income of one state may depend on per capita income of the neighboring states in a direct way, similarly for unemployment rates. Thus, we need to consider other spatial models, such the spatially lagged dependent variable model, for empirical studies that relate industrial diversity to economic performance.

2.4. Industrial Diversity and Employment Stability

Besides the studies discussed above that look at the diversity-economic performance relationship, there is also a literature that looks at the issue of industrial diversity and regional economic instability. Kort (1981) looked at the issue of regional economic instability and diversity using data from 106 Standard Metropolitan Statistical Areas of the U.S. The author used a simple model to test for the linkage:

$$\text{(K.1)} \quad REI_i = a + bDIV_i + \varepsilon_i$$

where REI is an index of regional economic instability and DIV is an index of industrial diversity[7]. REI is a measure that reflects the deviation of non-farm employment from the trend so higher values of REI indicate greater relative economic instability. Kort was concerned with the possibility of heteroskedasticity in the equation explaining instability as a function of diversity, with heteroskedasticity being related to city size. To address this, Kort used weighted least squares by multiplying both sides of the relationship by the square root of the SMSA population. Then (K.1) becomes

$$\text{(K.2)} \quad REI_i * \sqrt{POP_i} = a * \sqrt{POP_i} + bDIV_i * \sqrt{POP_i} + E_i * \sqrt{POP_i} \ .$$

Kort estimated model (K.2) and found that diversity is at least one of the factors that account for instability differences between regions in the U.S. Kort concluded that

[7] See Kort (1981) for formulation of REI and DIV.

the entropy diversity index positively correlates with regional economic instability. That is, regions with lower degrees of diversity tend to have higher levels of economic instability.

The study done by Kort (1981) was criticized by Brewer (1985) for the following reason. After using weighted least squares with transformed data, Kort (1981) computed the R-square measure assuming a single explanatory variable, inflating the explanatory power of the regression to 0.64. As Brewer pointed it out, the explanatory power of Kort's model is only 0.075[8]. Thus, the particular heteroskedasticity adjustment used by Kort did not result in greater explanatory power.

Similar to Kort (1981), Trendle and Shorney (2004) presented a working paper that examined the effect of industrial diversity on economic performance for the Local Government Area (LGA) in Queensland. Trendle and Shorney used a bivariate model to test for a relationship between industrial diversity and regional economic performance. The diversity index used is the entropy index and economic performance is assessed in terms of three variables: employment, unemployment, and per capita income. The data used in the study is from 1996 to 2001. Trendle and Shorney considered five different hypotheses, thus five different regressions were used to examine the relationship between diversity and economic performance. The five hypotheses tested by Trendle and Shorney are:

1. Diversity and employment instability are negatively correlated;
2. Diversity and employment growth are positively correlated;
3. Diversity and the unemployment rate are negatively correlated;
4. Diversity and the instability of the unemployment rate are negatively

[8] See Brewer (1984) for details.

correlated;

5. Diversity and per capita income are positively correlated.

Trendle and Shorney split the data over western and eastern regions in Queensland. The split was due to LGA's in central and western Queensland having small labor forces, narrow industrial bases, and historically low unemployment rates. The analysis of the two data sets for the eastern and western regions supported the hypothesis that industrial diversity has a significant influence on instability and employment growth. However, their analysis did not find that industrial diversity has a significant influence on per capita income. It is important to note that Trendle and Shorney used the level of per capita income rather than the growth rate. The use of levels of per capita income can lead to complicating factors in the analysis due to nonstationary time series that might be present.

Trendle and Shorney (2004) were also concerned with the effects that regional location and economic performance of neighboring regions have upon home regions. That is, they considered the spatial pattern of regional instability. Trendle and Shorney considered the Moran I test for spatial autocorrelation for all the variables used in the analysis including the entropy index of regional diversification. The Moran I-statistic takes the form:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

w_{ij} are the elements in a spatial matrix and $x\ and\ \bar{x}$ are the variable of interest and its mean value, respectively. The spatial weight matrix is a first order contiguity matrix with cells taking a value of 1 if i and j are neighbors and zero otherwise. The results of the Moran I tests suggested that regional economic activities exhibit spatial dependence.

Trendle and Shorney used simple descriptive statistics to explore the relationship between diversity and economic performance and concluded that diversity positively influences economic performance. They also did a preliminary analysis of the spatial pattern of regional economic instability but they did not use estimation procedures that address the spatial correlation problem among regions.

Similarly, Malizia and Ke (1993) explored the relationship between diversity and stability for regions. Malizia and Ke hypothesized that higher diversity leads to less unemployment and greater economic stability. To empirically test the hypothesis, Malizia and Ke used data that included most U.S. metropolitan areas over the time span of 1972 to 1988.

Malizia and Ke used the unemployment rate and a measure of instability as alternative dependent variables. Rather than select one year, they averaged the annual rates of unemployment in 1970, 1980, and 1986. Employment instability is measured as the average deviation from the employment trend and divided by trend employment.[9] As for independent variables, Malizia and Ke used the entropy index as the diversity measure for the region.

Instead of using a bivariate model and the OLS estimation method such as Kort (1981) and Trendle and Shorney (2004), Malizia and Ke included control variables in the cross sectional model. They proposed that the most important factors affecting unemployment and instability are population size, labor force characteristics, and

[9] See Malizia and Ke (1993) for formulation of their employment instability measure.

economic structure. Population size, nonwhite population percentage, female percentage of the labor force, and percentage of adults with college education also appear to have an influence on unemployment and instability.

Malizia and Ke believed that the growth rate of urban employment is also an important factor that determines the stability of the region. The authors' reasoning is that a tradeoff may exist between growth and unemployment or instability. Thus, Malizia and Ke included the growth rate of urban employment in the regression equation to test for the tradeoff. Other control variables that Malizia and Ke thought to have an influence on unemployment rates and instability included social, environmental, and natural geographical factors. To account for regional differences, Malizia and Ke divided the continental U.S. into 11 multistate regions and included the 10 regional dummy variables to account for multistate regions in their analysis[10]. Ordinary least squares was used to estimate the models for the unemployment rate and instability. The results supported the hypothesis that greater diversity leads to lower unemployment rates and less instability for the region.

Malizia and Ke also evaluated the sensitivity of changes in the dependent variables to changes in diversity. They computed the elasticity of the unemployment

---

[10] The 11 multistate regions are :

New England – Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island;
Middle Atlantic – New York, New Jersey, Pennsylvania, West Virginia, Delaware, Maryland;
West Manufacturing Belt – Ohio, Illinois, Michigan, Indiana;
Central Farming – Wisconsin, Minnesota, Iowa, Missouri;
Wheat Belt – the Dakotas, Nebraska, Kansas, Oklahoma;
South Atlantic – Virginia, the Carolinas, Georgia, Florida;
East South Central – Kentucky, Tennessee, Alabama, Mississippi;
West South Central – Arkansas, Louisiana, Texas;
Northern Rockies – Montana, Wyoming, Utah, Colorado;
Pacific Northwest – Idaho, Washington, Oregon;
West Sunbelt – California, Nevada, Arizona, New Mexico.

rate and employment instability with respect to the entropy measure at sample means. The estimates show that both the unemployment rate and employment instability are fairly sensitive to changes in industrial diversity. To put the elasticity into perspective, Malizia and Ke evaluated the elasticity at sample means and reported that a one percent increase in industrial diversity leads to a 1.7 percent reduction in the unemployment rate and a 1.3 percent reduction in instability. Along with Kort (1981) and Malizia and Ke (1993), Conroy (1974) also found that diversity leads to reductions in unemployment rates.

In summary, there is large body of literature that explores the interactions of regional industrial diversity, economic performance, and instability. Most researchers choose to express industrial diversity through measures based on employment shares such as in the entropy and Herfindahl diversity indices. Economic performance measures commonly used are per capita income or the unemployment rate. The vast majority of research comes to the conclusion that regions with high degrees of diversity have lower unemployment rates and lower employment instability. Kort (1981) and Wagner and Deller (1998) conclude that higher levels of industrial diversity are associated with lower levels of economic instability. Izraeli and Murphy (2001), after adjusting for fixed regional effects and partially adjusting for spatial correlation, concluded that higher degrees of diversity tend to be associated with lower unemployment rates. Similar to Izraeli and Murphy, Malizia and Ke (2003) also found that higher diversity leads to lower unemployment rates. In contrast, the relationship between industrial diversity and economic growth has not been consistent in the literature. Attaran (1986) and Izraeli and Murphy (2001) found negative correlation between diversity and the growth rate of personal income while Wagner and Deller (1998) and Trendle and Shorney (2004) found that higher levels of industrial diversity are statistically associated with higher levels of economic growth, as measured by changes in per capita income using averaged data over long time spans. Garrett et al. (2007) also found a positive relationship between industrial diversity and income growth in states. More attention is needed in exploring the relationship between industrial diversity and economic growth.

3. Diversity Indices and Measures of Economic Growth

3.1. Diversity Indices

In regional diversity studies, researchers approach diversity measures in terms of the distribution of employment across industry sectors. I consider five industrial diversity indices that are based on the distribution of employment across industrial sectors. The indices are the most common measures used in empirical studies due to their computational ease and limited demands for data. The first measure is the entropy index, defined for state j as:

$$(1) \qquad Entropy_j = \sum_{i=1}^{m} \frac{E_{ij}}{E_j} * \ln\left(\frac{E_j}{E_{ij}}\right),$$

where E_{ij} is the employment in industry i of state j and E_j is the total employment in state j.

The notion of perfect diversity in a region is defined as equal shares of economic activity across all the industries. If a region experiences perfect diversity then its process of diversification has achieved a maximum, or an equilibrium state. The Entropy index measures diversity against a uniform distribution of employment shares where the norm is equal shares of employment across industry sectors. In the context of the Entropy index, perfect diversity is achieved when the industry shares, E_{ij}/E_j, are the same for all m industries. At the other extreme, perfect specialization exists when all employment is concentrated in just one industry resulting in a zero value for the Entropy index. In the case of m industries, the range for the entropy index is zero to $\ln(m)$. Successively higher values of the index indicate successively higher degrees of diversity. Kort (1981), Attaran (1986), and Malizia and Ke (1993) adopt the Entropy index as their preferred diversity index.

The next two measures of diversity are both referred to as ogive indices and differ by virtue of the penalty function used to weight deviations of industry shares from the norm of a uniform distribution of shares. Jackson (1984) used the ogive index based upon a penalty function of absolute values:

(2) $$OG1_j = \sum_{i=1}^{m} \left| \frac{E_{ij}}{E_j} - \frac{1}{m} \right|$$

while the version based on a quadratic function is

(3) $$OG2_j = \sum_{i=1}^{m} \frac{\left(\frac{E_{ij}}{E_j} - \frac{1}{m} \right)^2}{\frac{1}{m}}.$$

The scale factor *(1/m)* in the $OG2$ index prevents the index from taking on extremely low values inherent in the squaring of small proportions. The lower bound of $OG1$ and $OG2$ is zero and is attained when the employment shares E_{ij}/E_j are the same for all *m* industries which reflect the case of perfect diversity. At the other extreme, the upper bound of $OG1$ is $(m-1)/m$ while the upper bound for $OG2$ is $\frac{(m-1)^2}{m}$, which reflects the case of perfect specialization. Successively higher values of each index indicate successively higher degrees of specialization, with respect to the norm of the uniform distribution of shares. Selection of the uniform distribution as the norm of comparison for economies is somewhat arbitrary and, as noted by Brown and Pheasant (1985), may limit the usefulness and interpretation of indices based on the norm. Nevertheless, these indices have been popular and I use them in the analysis.

The next two diversity indices used in the analysis are both referred to as “national average” indices. They differ from one another according to the penalty function used, as in the above pair of ogive indices. These two indices differ from the Entropy and the ogive indices in that they require a reference economy. The first of the two indices is defined as

$$(4) \qquad NA1_j = \sum_{i=1}^{m} \left| \frac{E_{ij}}{E_J} - \frac{E_{i,US}}{E_{US}} \right|$$

where $E_{i,US}$ is the U.S. employment in industry i and E_{US} is total U.S. employment. Similar to the ogive indices, higher values of the index signal greater levels of specialization of the regional economy with reference to the U.S. industrial structure.

The second of the two national average measures is defined as

$$(5) \qquad NA2_j = \sum_{i=1}^{m} \frac{\left(\frac{E_{ij}}{E_j} - \frac{E_{i,us}}{E_{us}} \right)^2}{\frac{E_{i,us}}{E_{us}}} .$$

In the $NA2$ index, departure of the regional share of employment from the norm established by the U.S. economy is penalized in a quadratic fashion which differs from $NA1$ where the penalty is linear. Similar to $NA1$, higher values of the index signal greater levels of specialization of the regional economy with reference to the U.S. industrial structure.

3.1.1. Data and Industries for the Diversity Indices.

The regions used in this study are the 48 contiguous states of the U.S. The time span of analysis will be 1992 to 2009, a period that witnessed strong economic growth in many states through 2000, followed by slowdowns in most states and declines in several during the early 2000s. Toward the end of the span in late 2007 the U.S. began to experience what is considered to be the worst recession since the Great Depression of the 1930s.

The five industrial diversity indices presented above are calculated for the 48 states using industry-level employment data from the U.S. Bureau of Labor Statistics. I use industries in the NAICS (North American Industrial Classification System) system of industry classifications for the U.S. From the NAICS system, 22 non-farm industries are used. Below is the list of the industries and their NAICS codes.

1. (10000000) Natural Resource and Mining
2. (20000000) Construction
3. (31000000) Durable Goods
4. (32000000) Non-Durable Goods
5. (41000000) Wholesale Trade
6. (42000000) Retail Trade
7. (43220000) Utilities
8. (43400089) Transportation and Warehousing
9. (50000000) Information
10. (55520000) Finance and Insurance
11. (55530000) Real Estate and Rental and Leasing
12. (60540000) Profession, Scientific, and Technical Services
13. (60550000) Management of Companies and Enterprises
14. (60560000) Administrative and Support and Waste Management and Remediation Service
15. (65610000) Educational Services
16. (65620000) HealthCare and Social Assistance
17. (70710000) Arts, Entertainment, and Recreation
18. (70720000) Accommodation and Food Services
19. (80000000) Other Services
20. (90910000) Federal Government
21. (90920000) State Government
22. (90930000) Local Government

3.2. Measures of Economic Growth

I now turn to measures of economic growth. When assessing economic performance in a region, there are many measures that can be employed. Prior research has used levels or growth rates of per capita income (Wagner and Deller, 1998; Izraeli and Murphy, 2001; Garrett et al., 2007). The use of per capita income does present some issues. Per capita income is a broad measure and includes earnings that are not due to employment in the region (via the residence adjustment). Per capita income includes sources such as rent, dividends and interest, the residence adjustment, and transfer payments. None of those sources may have a dependable correspondence with employment levels or with the industry distribution of employment in the region. Rent, dividends and interest, and transfer payments, unlike wages and salaries, are reported by place of residence rather than by place of work.

The economic growth measures should be as conceptually compatible with employment distributions as possible. For that reason, economic growth is measured as the growth rate of real nonfarm earnings. Nominal values are converted to real using the gross state product (GSP) deflator. Since the study is concerned with diversity and growth at the state level, GSP deflators for each state are used rather than the gross domestic product (GDP) deflator. Nonfarm earnings are defined as the sum of wages and salaries, other labor income, and proprietor's income. I do not include farm earnings in the economic growth measure since farm employment is not included in the diversity indices. However, farm earnings may be an important influence upon nonfarm earnings growth in some states and this possibility will be considered below.

All earnings data are from the Bureau of Economic Analysis (BEA). As for the GSP deflators by state, I implicitly derive the deflators by using the ratio of real GSP and nominal GSP by state. The GSP data is from BEA.

4. Model and Methodology

4.1. Multivariate Model of Economic Growth

A prospective model for the growth rate of nonfarm earnings incorporates influences from industrial diversity, employment growth, capital growth, and movements in the farm sector. I also consider fixed effects for each of the multistate regions as classified by the BEA. The BEA classifies the 48 states into eight economic regions based primarily on cross sectional similarities in the states' socioeconomic characteristics. The model is, for state i at time t:

$$GRO_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 NFEM_{it} + \beta_4 K_{it} + \beta_5 FER_{it} + \sum_{j=1}^{7} \delta_j DR_{j,it} + \sum_{j=1}^{\mathrm{T}-1} \phi_j DS_{j,it} + \varepsilon_{it} \quad (6)$$

where GRO is the growth rate in real nonfarm earnings, V is any of the five diversity indices discussed in Chapter 3, $NFEM$ is the growth rate of nonfarm employment, K is the growth rate of capital (see Appendix 2), FER is the growth rate of farm earnings, and the DR_j are dummy variables for the BEA regions[11] (see Appendix 1). The fixed effects are intended to control for region-specific effects that are unobserved and might determine a region's economic growth rate such as climate, geography, traditions, and resource endowments. Thus, the fixed effect terms δ_j capture idiosyncratic factors specific to regions that are unobservable. The ϕ_j capture time effects, which are represented by dummy variables DS_j, when multiple time periods are used in estimating the model (panel data).

Since my hypothesis is that regions with higher degrees of industrial diversity will

---

[11] BEA Regions: New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, and Far West.

experience higher economic growth compared to regions with industrial concentration, I expect the coefficient of the diversity variable to be positive for the Entropy index. Higher values of the Entropy index represent higher degrees of industrial diversity in the region. Conversely, the coefficient of the diversity variable is expected to be negative for the ogive and national average indices since higher values of those indices represent higher degrees of industrial specialization (less diversity).

I expect that nonfarm employment growth will carry a positive coefficient. This is straightforward since the higher the growth rate in nonfarm employment, the higher the growth rate of nonfarm earnings. Similarly, I expect that growth rate of capital will carry a positive coefficient. I also expect real farm earnings growth to carry a positive coefficient. Even though I expect growth in real farm earnings to positively influence the growth of real nonfarm earnings, the significance of the variable is in question, especially when using annual data. Farm earnings are very volatile on a year-to-year basis and can have very irregular patterns due in part to the variability in government farm policies and programs. Other researchers included population as a control variable in their models and found it to be significant contributor to economic growth (Kort (1981); Brewer (1984); Malizia and Ke (1993); Warner and Deller (1998); and Izraeli and Murphy (2001)). All of the studies above chose to measure economic growth by the unemployment rate or growth rate of per capita income. Population may have an important correlation with the unemployment rate and per capita income. However, in this study economic growth is measured by the growth rate of real nonfarm earnings which may not have as strong a correlation with population. Total population contains many persons that are not in the labor force.

There is uncertainty about the temporal dimensions of the prospective relationship between industrial diversity and economic growth. Other studies in this area are inconsistent in their treatments of the timing issue. Some studies used averages of data and some use annual data. Studies that used averages of data over long spans included Attaran (1986), Milizia and Ke (1993), and Wagner and Deller (1998). In this study, I avoid the practice of measuring growth and diversity at different points in time within the same modeling framework. For example, Malizia and Ke (1993) used the average of the unemployment rate for metropolitan areas at three non-contiguous points in time (1970, 1980, and 1986) as their measure of economic activity and related it to the entropy measure computed with industry employment data from 1977. Wagner and Deller (1998) averaged the annual growth rates in per capita income over 1969-1992 for each state and related these 23-year averages to the diversity measure based on data from a single point in time, 1982. Such temporal misalignments and straddles are avoided by making sure that the variables are always measured in a fully contemporaneous fashion with one another.

In this study, I use annual data for the estimation of the model in (6). Using annual data over a long period of time allows us to interpret the marginal effects of diversity, growth of nonfarm employment, growth of capital, and growth of farm earnings, upon state economic growth measured as growth rate of nonfarm earnings. By considering annual data for 48 states over 1992-2009, a panel data set is created in which each state has 18 observations thereby totaling 864 observations across the 48 states. Studies using annual data include Izraeli and Murphy (2001); Garrett et al. (2007). However, Izraeli and Murphy used an unbalanced panel data set. That is, they considered the time span from 1960 to 1986 but not all 17 states in the study had data for the entire time span whereas our panel data set will consist of continuous time spans for all 48 contiguous states from 1992 to 2009.

4.1.1. Estimation Method – Basic Models

When estimating the model in (6) with panel data, cross sectional heteroskedasticity (different error variances for each state) needs to be considered. In the event of heteroskedasticity, t-statistics can be based on "panel-corrected" standard errors that are adjusted for cross-sectional heteroskedasticity and for covariance of the errors between cross sectional units, a method described in Beck and Katz (1995). The panel-corrected covariance matrix given by Beck and Katz is formulated as follows:

(7) $$BK = (X'X)^{-1}[X'(\hat{\Sigma} \otimes \mathrm{I})\mathrm{X}](\mathrm{X'X})^{-1}.$$

$\hat{\Sigma}$ is the $m \times m$ matrix of estimators of the error cross-covariances with i, j element :

(8) $$\hat{\sigma}_{ij} = \frac{\underline{e_i}'\underline{e_j}}{n} = \frac{\sum_{t=1}^{n} e_{it}e_{jt}}{n} \text{ for } i, j = 1, \dots, m \text{ (m cross-sections)}$$

where $\underline{e_i} = y_i - X_i\underline{b}$, the vector of residuals pertaining to cross sectional unit i and $\underline{b}$ is the vector of OLS estimates.

When using time series data, the issue of panel unit roots must be considered. I do not expect the panel data to have unit roots since we are dealing with growth rates of variables, not levels. Nevertheless, the panel unit root tests of Im, Pesaran, and Shin (2003) will be applied to the continuous growth variables in (6). The Im, Pesaran, and Shin (IPS) test has a null hypothesis of unit roots in the panel and an alternative of no unit roots. The IPS test assumes individual unit root processes for each of the cross sectional unites in the panel.

For the model in (6), one may suspect the presence of endogeneity in one or more of the independent variables, namely the growth rate of nonfarm employment, the diversity measures, and the growth rate of capital. If endogeneity is present, OLS estimates will be biased and inconsistent. The Hausman test is used to test for endogeneity. The test requires a set of instrumental variables that are correlated with the growth rate of nonfarm employment but not with the error term of the equation in (6), similarly for the growth rate of capital and the diversity measures. The set will consist of the lag the growth rate of of nonfarm employment, the lag of the diversity index, a pair of instrumental variables for the growth rate of capital (discussed below), the growth rate of farm earnings, and all of the dummy variables involved in (6). Farm earnings are considered to be exogenously determined due to the substantial influence of national farm policies and programs.

The lag of the growth rate of capital cannot be used as an instrumental variable because the correlation between the lag and the current value is very low, i.e., the lag of capital is a weak instrument. Thus, I turn to Dagenais and Dagenais (1997) for suggestions of instrumental variables. They proposed a series of higher moment instrument variables. The two higher moment instrument variables that I use in this study are defined by the following vectors:

(9) $$z_1 = \overline{k} * \overline{k}$$

(10) $$z_4 = \overline{k} * \overline{k} * \overline{k} - 3 * \overline{k} * s_k^2,$$

where $\overline{k}$ is the vector of the growth rate of capital observations measured in deviations from their mean, s_k^2 is the maximum likelihood estimator of the variance of the growth rate of capital, and * designates the element-by-element (Hadamard) matrix multiplication operator. Results from Monte Carlo experiments conducted by Dagenais and Dagenais indicated that the combination of z_1 and z_4 performed as well or better than several other groupings of instrumental variables.

Two steps are involved in the Hausman test. In the first, the variables suspected of being endogenous are each regressed upon all exogenous variables and an outside set of instrumental variables. The second regression includes the residuals from each regression in the first step as additional regressors in the original model (6). If the coefficient on any one of the additional regressors are significantly different from zero, then the corresponding original variable is declared to be endogenous. Nonfarm employment growth, capital growth, and the diversity indices will be subjected to the Hausman test.

4.2. Spatial Dependence Models

Studies concerning industrial diversity have used multivariate models to incorporate important control variables when studying the effect of diversity on economic performance, for example in Malizia and Ke (1993) and Izraeli and Murphy (2003). Garrett, Wagner, and Wheelock (2007) included an extensive treatment of spatial correlation in their study of state income growth in the United States using data from 1977-2002. Diversity was included as a control variable in their spatial models but they only considered one index. Besides Garrett et al., earlier studies and analyses of income growth had not given much attention to regional location and the performance of neighboring regions (states). The problem with using panel data that incorporates a location component is that spatial dependence may exist between the observations at each point in time. It is possible that fast and slow growing regions are not distributed randomly across geographic space but that some form of spatial dependence exists. This spatial dependence may be due to some form of spatial process whereby economic growth is transferred from one region to the next, or spillover effects from one state to the next. Regional science theory points out that economic agents may change their decisions depending on market conditions in the region or location as compared to other regions (Elhorst, 2003). That is, market conditions in neighboring states may have positive or negative effects upon the economic conditions in the home states, thereby changing the behavior of economic agents in the home state. These changes may have positive or negative effects on economic growth.

With the 48 states being used as regional economies in the study, there may be spatial dependences in the economic growth patterns of neighboring states. It is likely that economic performance in one state is affected by shocks to economies of neighboring states. For example, to the extent that two states such as New York and Pennsylvania or Michigan and Ohio are significant trading partners, then a demand shock in one state would have repercussions for economic performance in a nearby state (Izraeli and Murphy, 2001). Also, unobserved factors that contribute to economic performance may be spatially correlated across regions at each point in time. Thus, the error term in the model and/or the dependent variable, economic growth rate, of one state may have spatial dependences with neighboring states. In short, by considering the spatial correlation model, we no longer assume that cross sectional units are independent in the spatial sense.

4.2.1. Spatial Models

I will discuss models that allow for spatially lagged dependent variables and spatial autoregressive errors. Even though our application will involve panel data, I will first discuss models for single cross sections and then generalize to panel data. The notation used in the discussion of spatial models will closely follow, and in some instances exactly match, that of Kelejian and Prucha (1998, 2007). For ease of reference, what follows below is a collection of notation declarations and conventions. Notation used by Kelejian and Prucha (1998, 2007) will be denoted by (KP) and it applies to an environment of a single cross section, that is, observations on n cross sectional units in a single time period. The subscript n is used by KP as a reminder that there are n cross sectional units. The KP notation will be generalized to accommodate the extension to panel data. Such changes will be apparent in the list below. Vectors will be underlined.

i - cross sectional unit
n - number of cross sectional units
t - time period
T - number of time periods
$y_{i,n}$ - observation on the dependent variable for cross sectional unit i (KP)
$\underline{y}_n$- $n \times 1$ vector of observations on the dependent variable (KP)
$x_{ij,n}$ - observation on exogenous variable j for cross sectional unit i (KP)
X_n - $n \times k_x$ matrix of observations on k_x exogenous variables (KP)
Y_n- $n \times r_y$ matrix of observations on r_y endogenous variables (KP)

G_n- $n \times s_y$ matrix of observations on s_y instrumental variables
$w_{ij,n}$ - element ij of the $n \times n$ spatial weights matrix W_n with $w_{ii,n}$ = 0 (KP)
$m_{ij,n}$ - element ij of the $n \times n$ spatial weights matrix M_n with $m_{ii,n}$= 0 (KP)
$u_{i,n}$ - spatially correlated error term for cross sectional unit i
$\varepsilon_{i,n}$ - random error term for cross sectional unit i
β_j - coefficient on exogenous variable j
$\underline{\beta}$ - $k_x \times 1$ vector of coefficients on k_x exogenous variables
$\underline{\gamma}$ - $r_y \times 1$ vector of coefficients on r_y endogenous variables
λ - spatial lag coefficient
ρ - spatial autoregressive coefficient

y_{it} - observation on the dependent variable for cross sectional unit i in time t
$\underline{y}_t$ - n x 1 vector of observations on the dependent variable in time t
$\underline{y}$ - $nT \times 1$ vector of observations on the dependent variable, $[\underline{y}_1' \dots \underline{y}_T']'$
x_{itj} - observation on exogenous variable j for cross sectional unit i in time t
$\underline{x}_{tj}$ - $n \times 1$ vector of observations on exogenous variable j in time t
X_t- $n \times k_x$ matrix of observations on k_x exogenous variables at time t, $[\underline{x}_{t1} \dots \underline{x}_{tk_x}]$
Y_{itj}- observation on endogenous variable j for cross sectional unit i in time t
$\underline{Y}_{tj}$- $n \times 1$ vector of observations on endogenous variable j in time t
Y_t- $n \times r_y$ matrix of observations on r_y endogenous variables in time t, $[\underline{Y}_{t1} \dots \underline{Y}_{tr_y}]$
G_t - $n \times s_y$ matrix of observations on s_y instrumental variables in time t, $[\underline{g}_{t1} \dots \underline{g}_{ts_y}]$
X - $nT \times k$ matrix of observations on k_x exogenous variables, $[X_1' \dots X_T']'$
Y- $nT \times r_y$ matrix of observations on r_y endogenous variables, $[Y_1' \dots Y_T']'$
G - $nT \times s_y$ matrix of observations on s_y instrumental variables, $[G_1' \dots G_T']'$
w_{ij} - ij element of the $n \times n$ spatial weights matrix W_n with w_{ii} = 0
m_{ij} - ij element of the $n \times n$ spatial weights matrix m_n with m_{ii} = 0
W - block diagonal $nT \times nT$ matrix $I \otimes W_n$ where I is $T \times T$
M - block diagonal $nT \times nT$ matrix $I \otimes M_n$ where I is $T \times T$
u_{it} - spatially correlated error term for cross sectional unit i in time t
ε_{it} - random error term for cross sectional unit i in time t

In the panel data notation of above, the vectors (or blocks) of observations in $\underline{y}$ (or X) are stacked by time period, not by cross sectional unit. For example, the first block in X

consists of the n observations on the independent variables for the first cross section.

4.2.1.1. Basic Spatial Model

For cross sectional unit i in a single time period, Kelejian and Prucha (1998) write the spatial lag model with exogenous independent variables and first-order spatial autoregressive errors as:

$$(11) \quad y_{i,n} = \sum_{j=1}^{k_x} x_{ij,n}\,\beta_j + \lambda \sum_{j=1}^{n} w_{ij,n}\, y_{j,n} + u_{i,n}$$

$$u_{i,n} = \rho \sum_{j=1}^{n} m_{ij,n}\, u_{j,n} + \varepsilon_{i,n}.$$

The dependent variable's observation in cross sectional unit i is related to the dependent variable's observation in every other cross sectional unit j where the spatial weight, w_{ij}, is nonzero. Similarly, the error term for cross sectional unit i is related to the error terms of every other cross sectional unit j where m_{ij} is non-zero. In this study, it is assumed that $w_{ij,n} = m_{ij,n}$. In matrix format, the spatial model and error process for an entire cross section in a single time period with exogenous independent variables is

$$(12) \quad \underline{y}_n = X_n\underline{\beta} + \lambda W_n\underline{y}_n + \underline{u}_n$$

$$\underline{u}_n = \rho W_n\underline{u}_n + \underline{\varepsilon}_n.$$

The spatial model is used when there is suspicion that economic activity in one state is directly influenced by economic activities in other states. For example, economic growth in a neighboring state stimulates demand in a home state and leads to more exports of goods and services from the home state to other states. A test of $H_0: \lambda = 0$ can be conducted. If H_0 is rejected then there exists spatial dependence among neighboring states.

The spatial model is also used when it is suspected that fast and slow growing states are not distributed randomly across geographic space, but rather have some form of spatial dependence as captured through the spatial autoregressive mechanism for the error terms. The autoregressive error mechanism also accounts for unobserved variables that are related to each state over space. For a single cross section, the Moran test statistic can be used to determine if there is spatial autocorrelation of this type. Special cases of the model are provided where $\lambda = 0$ (spatial autoregressive errors only) or $\rho = 0$ (autoregressive spatial dependence only). In this study, I will estimate the spatial lag model with spatial autoregressive errors.

The spatial model can be extended to allow for a situation where some independent variables are endogenous, that is, they covary with the error term. In this situation, we limit X_n to represent the matrix of observations on k_x exogenous variables in the model and introduce Y_n as the matrix of observations on r_y endogenous variables. The model becomes

$$(13) \quad \underline{y}_n = X_n\underline{\beta} + \lambda W_n\underline{y}_n + Y_n\underline{\gamma} + \underline{u}_n$$

$$\underline{u}_n = \rho W_n\underline{u}_n + \underline{\varepsilon}_n.$$

I will use five different spatial weights matrices in this study. All have dimensions $n \times n$ with the rows and columns represented by states. The first spatial weights matrix is the contiguity matrix in which element w_{ij} is set equal to one if states i and j share a common border and zero if states i and j do not share a common border. In addition, all diagonal elements of W_n, w_{ij} for $i = j$, are equal to zero. The contiguity weights matrix effectively assigns equal shares to spillover effects from all border-touching states while the spillover effects from all other states are ignored.

I also consider a distance-based weights matrix. Distance-based weights matrices have been used in previous studies such as Garrett et al. (2007) and Hernandez (2003). Following Garrett et al. and Hernandez, I use the inverse-distance weights matrix. Under the inverse-distance weighting scheme, the element w_{ij} is set to $1/d_{ij}$, where d_{ij} is the distance between state i and j. Unlike the contiguity weights matrix, the spatial effects upon state i are coming from all other states, rather than just the border-touching states. As the distance between states i and j increases, w_{ij} decreases which effectively gives less weight to states that are farther away. The intuition behind the inverse-distance weighting scheme is straightforward. States that are farther away are thought to have less influence on a particular state economy compared to states that are closer to home. Economic activities sometimes cluster together and one state's activities can have strong linkages to nearby states. For example, economic activity in Virginia may have substantial effects upon economic activity in Maryland but only minor effects, if at all, upon economic activity in Texas. I follow Garrett et al. (2007) and measure distance between state i and j using state population centroids from the 2000 Census of Population. The Bureau of the Census determines the longitudes and latitudes for the population centers of the 50 states. Using this information, one can calculate the distance between state i and state j population centroids.[12]

The third weights matrix is also distance based, where element w_{ij} is the inverse of squared distance between state i and j, that is $w_{ij} = 1/{d_{ij}}^2$. The inverse-distance-squared weights matrix is used by Hernandez (2003). Similar to the inverse-distance weights matrix, as the distance between states i and j increases, w_{ij} decreases so that the influence upon economic growth from states closer to home is greater than states further away from home. The elements in the weights matrix decline at a geometric rate so the influence upon economic activity in state i from activity in state j declines rapidly with distance.

The fourth weights matrix is a combination of the contiguity and inverse-distance weights matrices. Denote the contiguity weights matrix as W_1 and the inverse-distance weights matrix as W_2. The combination weights matrix, W_4, is equal to the element-wise product of W_1 and W_2. Under this weighting scheme, the influence of economic activity in state i upon contiguous state j varies according to the distance between states i and j but there are no effects if states i and j do not have a common border. That is, state economic growth is influenced by surrounding states but stronger influences are provided by states whose population centers are closer. For example, according to this weighting scheme, Nebraska economic growth is more influenced by Kansas and Iowa as compared to Wyoming or South Dakota. The Nebraska population center is closer to Kansas and Iowa population centers as compared to Wyoming and South Dakota population centers. This weights matrix assumes that the six states

[12] Dr. Thomas Garrett of the St. Louis Federal Reserve Bank kindly provided the distance matrix based on population centroids for the census year of 2000.

touching Nebraska have an influence upon Nebraska's economic growth while all other states do not.

Lastly, the fifth weights matrix, W_5, is a combination of the contiguity weights matrix and inverse-distance-squared weights matrix. The combination weights matrix, W_5, is equal to the element-wise product of W_1 and W_3. The structure and behavior of W_5 is similar to that of W_4 but the influence of state i upon neighboring state j declines faster with increasing distance as compared to W_4.

All five weights matrices are row normalized. That is, all elements in each row are restricted to sum to one, $\sum_{j=1}^{n} w_{ij} = 1$, where n is the number of states. This normalization makes the parameter estimates of alternative models comparable. Also, the diagonal elements of all five weights matrices are zero.

4.2.1.2. <u>Spatial Model with Panel Data</u>

For the panel environment of n cross sectional units and T time periods, I follow Hernandez-Murrilo (2003) and assume that the spatial weights matrices W_n and M_n apply in each time period. For panel data with n cross sectional units and T time periods, the spatial model with allowance for endogenous independent variables is written for unit i in time period t as

(14) $$y_{it} = \sum_{j=1}^{k_x} x_{itj}\, \beta_j + \lambda \sum_{j=1}^{n} w_{ij}\, y_{jt} + \sum_{j=1}^{r_y} Y_{itj} \gamma_j + u_{it}$$

$$u_{it} = \rho \sum_{j=1}^{n} m_{ij}\, u_{jt} + \varepsilon_{it}.$$

In matrix format, the model is

(15) $$\underline{y} = X\underline{\beta} + \lambda W \underline{y} + Y\underline{\gamma} + \underline{u}$$

$$\underline{u} = \rho W \underline{u} + \underline{\varepsilon}$$

where the assumption of $w_{ij} = m_{ij}$ has been made. The X and Y data matrices are the stacks (vertical concatenations) of data matrices X_t and Y_t from time periods 1 through T. The spatial weight matrix W is block diagonal with W_n as each of the T blocks.

4.2.2. Estimation of Spatial Models

The spatial models outlined above allow the dependent variable corresponding to each cross sectional unit to depend on a weighted average of that dependent variable in neighboring units. Thus, the spatially lagged dependent variable, $W_n y_n$, is typically correlated with the error term and hence, ordinary least squares is not a consistent estimator for the model.

I will discuss in detail two methods used to estimate the spatial models. The first method is Generalized Spatial Two-Stage Least Squares (GSTSLS) due to Kelejian and Prucha (1998). The second method is the spatial heteroskedasticity and autocorrelation consistent (SHAC) estimator of the covariance matrix in a spatial framework that was developed by Kelejian and Prucha (2007).

4.2.2.1. Estimation of the Basic Spatial Model

For estimating the basic spatial model in (12) by GSTSLS, Kelejian and Prucha suggest that the instrumental variables be composed of a subset of the linearly independent columns of $(X_n, W_n X_n, W_n^2 X_n, \ldots, M_n X_n, M_n W_n X_n, M_n W_n^2 X_n, \ldots)$. If using the assumption of $W_n = M_n$ as I am in this study, the portion of the above list of instrumental variables involving M_n disappears. When W_n is in row-normalized form, then the column of W_nX_n that corresponds to the column of ones in X_n must be dropped since the two columns are linearly dependent.

Let $Z_n = \left(X_n, W_n\underline{y}_n\right)$, $\underline{\delta}_n = (\underline{\beta}', \lambda)'$, and H_n be the data matrix of p instrumental variables. In the first step of GSTSLS, the instrumental variable (or two-stage least squares) estimator of $\underline{\delta}_n$ is computed. It is

(16) $\tilde{\underline{\delta}}_n = (\hat{Z}_n'\hat{Z}_n)^{-1}\hat{Z}_n'\underline{y}_n$

where $\hat{Z}_n = P_nZ_n = \left(P_nX_n, P_nW_n\underline{y}_n\right) = (X_n, \widehat{W_n\underline{y}_n})$, and $P_n = H_n(H_n'H_n)^{-1}H_n'$. $\tilde{\underline{\delta}}_n$ can also be expressed as

$$\tilde{\underline{\delta}}_n = [Z_n'H_n(H_n'H_n)^{-1}H_n'Z_n]^{-1}Z_n'H_n(H_n'H_n)^{-1}H_n'\underline{y}_n.$$

In the second step of GSTSLS, a generalized method of moments estimator (GMM) for the spatial autoregressive coefficient, ρ, is calculated based upon the residuals from the first step. Details are given in Kelejian and Prucha (1998). They note that ρ can be consistently estimated by GMM whether or not the weight matrices for the dependent variable and the error process are equal.

In the third step, an autoregressive transformation is applied to the original variables followed by a final instrumental variables estimator of $\underline{\delta}_n$. With the estimate of ρ from step two, transform Z_n and $\underline{y}_n$ according to

(17) $Z_{n*}(\hat{\rho}_n) = Z_n - \hat{\rho}_nW_nZ_n$

$$\underline{y}_{n*}(\hat{\rho}_n) = \underline{y}_n - \hat{\rho}_nW_n\underline{y}_n.$$

The final instrumental variables estimator of $\underline{\delta}_n$ is

(18) $$\hat{\underline{\delta}}_{F,n} = \left[\hat{Z}_{n*}(\hat{\rho}_n)'\hat{Z}_{n*}(\hat{\rho}_n)\right]'\hat{Z}_{n*}(\hat{\rho}_n)'\underline{y}_{n*}(\hat{\rho}_n)$$

where $\hat{Z}_{n*}(\hat{\rho}_n) = P_n Z_{n*}(\hat{\rho}_n)$. The residuals from estimating the transformed model are

(19) $$\hat{\underline{\varepsilon}}_n = \underline{y}_{n*}(\hat{\rho}_n) - Z_{n*}(\hat{\rho}_n)\hat{\underline{\delta}}_{F,n}.$$

The estimator of the variance of the errors $\varepsilon_{i,n}$ is given by

(20) $$\hat{\sigma}^2_{\varepsilon,n} = \hat{\underline{\varepsilon}}_n'\hat{\underline{\varepsilon}}_n/n$$

and an estimator of the asymptotic covariance matrix of $\hat{\underline{\delta}}_{F,n}$ is provided by

(21) $$\hat{\sigma}^2_{\varepsilon,n}[\hat{Z}_{n*}(\hat{\rho}_n)''\hat{Z}_{n*}(\hat{\rho}_n)']^{-1}.$$

A limitation of the GS2SLS estimation method is that one cannot test the hypothesis of $\rho = 0$.

4.2.2.2. Estimation of the Spatial Model with Endogenous Variables

For the model in (13), let $Z_n = \left(X_n, W_n\underline{y}_n, Y_n\right)$, $\underline{\delta}_n = (\underline{\beta}', \lambda, \underline{\gamma}')'$, and $\overline{X}_n = (X_n, G_n)$ where G_n is the data matrix of a set of s_y instrumental variables that are introduced to assist with treating the endogeneity present in Y_n. Let H_n be the data matrix of p instrumental variables that are composed of a subset of the linearly independent columns of $(\overline{X}_n, W_n\overline{X}_n, W_n^2\overline{X}_n, \dots, M_n\overline{X}_n, M_nW_n\overline{X}_n, M_nW_n^2\overline{X}_n, \dots)$. Under the assumption $W_n = M_n$, any submatrix in H_n that involves M_n is dropped. As before, when W_n is in row-normalized form, then the column of $W_n\overline{X}_n$that corresponds to the column of ones in $\overline{X}_n$ is dropped. Estimation of $\underline{\delta}_n$ proceeds in the same manner as outlined earlier for the basic spatial model.

4.2.2.3. Estimation of the Spatial Model with Panel Data

For the model in (15), the X and Y data matrices are stacks (vertical concatenations) of data matrices X_t and Y_t, respectively, from time periods 1 through T. The spatial weight matrix W is block diagonal with W_n as each block. Let $Z = (X, W\underline{y}, Y)$, $\underline{\delta} = (\underline{\beta}', \lambda, \underline{\gamma}')'$, and $\overline{X} = (X, G)$ where G is the stack of data matrices G_t for the instrumental variables being used to treat endogeneity in the variables of Y. Let H be the data matrix of p instrumental variables that are composed of a subset of the linearly independent columns of $(\overline{X}, W\overline{X}, W^2\overline{X}, \ldots)$. Estimation of $\underline{\delta}$ proceeds in the same manner as outlined earlier for the basic spatial model. If none of the independent variables are endogenous, then G is not required and $\overline{X}$ becomes X.

In the growth rate model of (6), $V, NFEM$, and K will be tested for endogeneity with the Hausman test. FER is assumed to be exogenous since its levels are strongly influenced by national farm policies. The variables that prove to be endogenous dictate the composition of (6). For example, if V, $NFEM$, and K are endogenous then X is $(\underline{1}\ \ \underline{FER}\ \underline{DR}_1\ \ldots\ \underline{DR}_7\ \ \underline{DS}_1\ \ldots\ \underline{DS}_{T-1})$ and G is $(\underline{V}_{t-1}\ \underline{NFEM}_{t-1}\ \underline{z}_1\ \underline{z}_4)$. The lags of V and $NFEM$ are used as instrumental variables in G and $\underline{z}_1$ and $\underline{z}_4$ are used as the instrumental variables for K in G. Now let $\dot{X} = (\underline{1}\ \underline{FER}\ \underline{DR}_1 \ldots \underline{DR}_7)$ and $\dot{\overline{X}} = \left(\dot{X}, G\right)$. The H matrix to be used is $\left(\dot{\overline{X}}, W\dot{\overline{X}}, W^2\dot{\overline{X}}\right)$. In our estimation work, the dummy variables for the time periods were dropped from X to create $\dot{X}$ due to near linear dependencies that appeared in $W\overline{X}$ and $W^2\overline{X}$ when attempting to estimate the model in (6). If $V, NFEM$, and K are exogenous, the they enter X and $\dot{X}$ while G disappears.

4.2.2.4. SHAC Covariance Matrix Estimator

Kelejian and Prucha (2007) consider the following spatial model for a single cross section

$$(22)\quad \underline{y}_n = X_n\underline{\beta} + \lambda W_n\underline{y}_n + Y_n\underline{\gamma} + \underline{u}_n$$

$$\underline{u}_n = R_n\underline{\varepsilon}_n.$$

where X_n and Y_n contain exogenous and endogenous variables, respectively. The covariance matrix of $\underline{u}_n$ is Σ_n. The error portion of (22) is a much more general specification than in (13) where $R_n = (1 - \rho W_n)^{-1}$. The structural portion of the model is estimated with $\underline{\tilde{\delta}}$, the instrumental variable (or two-stage least squares) estimator of $\delta_n = (\underline{\beta}', \lambda, \gamma')'$ that is also used in the first step of GSTSLS. Recall that $\underline{\tilde{\delta}}$ is

$$(23)\quad \underline{\tilde{\delta}}_n = (\hat{Z}_n'\hat{Z}_n)^{-1}\hat{Z}_n'\underline{y}_n = [Z_n'H_n(H_n'H_n)^{-1}H_n'Z_n]^{-1}Z_n'H_n(H_n'H_n)^{-1}H_n'\underline{y}_n$$

where $Z_n = \left(X_n, W_n\underline{y}_n, Y_n\right)$, $\hat{Z}_n = P_nZ_n$, and $P_n = H_n(H_n'H_n)^{-1}H_n'$. When dealing with covariance matrices of instrumental variable estimators, a common component is the covariance matrix of the cross-products of the instrumental variables and the model errors, here being $n^{-1/2}H_n'\underline{u}_n$. Its covariance matrix is the $p \times p$ matrix $\Psi_n = n^{-1}H_n'\Sigma_nH_n$.

Kelejian and Prucha (2007) develop a spatial heteroskedasticity and autocorrelation consistent (SHAC) estimator of Ψ_n. The r, s element of the $n \times n$ matrix Ψ_n is

$$(24)\quad \psi_{rs,n} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir,n}h_{js,n}\sigma_{ij,n}.$$

Denote the estimation residuals by $\underline{\hat{u}} = \underline{y}_n - Z_n\underline{\tilde{\delta}}_n$. The r,s element of the SHAC estimator $\hat{\psi}_{rs,n}$ is

(25) $\hat{\psi}_{rs,n} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir,n}\, h_{is,n}\hat{u}_{i,n}\hat{u}_{j,n}K(\frac{d_{ij}}{d_{max}})$

where the $K(.)$ denotes a kernel function. The full matrix Ψ_n can also be written as a sum of matrices created by vector products. Let $\underline{h}_g'$ be row g of H_n. $\underline{h}_g$ is the column vector containing the transpose of row g. Then

(26) $\Psi_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\underline{h}_i\underline{h}_j'\sigma_{ij,n}$.

The SHAC estimator $\hat{\Psi}_n$ is given by

(27) $\hat{\Psi}_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\underline{h}_i\underline{h}_j'\hat{u}_{i,n}\hat{u}_{j,n}K(\frac{d_{ij}}{d_{max}})$.

The consistent estimator of the covariance matrix of $\underline{\tilde{\delta}}_n$ is

(28) $\hat{\Phi}_n = n\left(\hat{Z}_n'\hat{Z}_n\right)^{-1}Z_n'H_n(H_n'H_n)^{-1}\hat{\Psi}_n(H_n'H_n)^{-1}H_n'Z_n(\hat{Z}_n'\hat{Z}_n)^{-1}$.

Standard errors and t-statistics for the estimated parameters can be derived from (28).

Kelejian and Prucha (2007) used the Parzen kernel in their simulation work with the SHAC estimator. The Parzen kernel is also used by Lambert and McNamara (2009) in their spatial model of the location determinants of food manufacturers at the county level in the United States. The Parzen kernel will be used in this study and is defined as:

(29) $$K\left(\frac{d_{ij}}{d_{max}}\right) = \begin{cases} 1 - 6(\frac{d_{ij}}{d_{max}})^2 + 6\left|\frac{d_{ij}}{d_{max}}\right|^3 & for \quad 0 \le \left|d_{ij}/d_{max}\right| \le .5 \\ 2\left(1 - \left|\frac{d_{ij}}{d_{max}}\right|^3\right) & for \quad .5 < \left|d_{ij}/d_{max}\right| \le 1 \\ 0 & otherwise \end{cases}$$

where $d_{ij}$is the distance between states i and j. As before, the distance between states i and j is measured as the distance between the two states' population centroids.

Picking d_{max} is an important aspect of the kernel function. Within the spatial framework, d_{max} is essentially the maximum distance from state i to other states that are believed to have an influence upon the economic performance of state i. To come up with d_{max}, Lambert and McNamara (2009) picked $n^{.25}$ as the bandwidth for identifying the group of neighbors (but not necessarily touching) within which to determine d_{max}. The problem with this method of determining d_{max} is that it would produce a non-symmetric SHAC covariance matrix. Notice that d_{max} under Lambert and McNamara's method can be different for each state which will create a non-symmetric matrix in (27) and result in a non-symmetric covariance matrix in (28).

Due to the non-symmetric covariance matrix issue, I have no compelling reason to follow Lambert and McNamara's method to determine d_{max}. Instead of following their suggestions, I use $d_{max} = 500$ (miles). With $d_{max} = 500$, there are, on average, 10 other states involved in the calculation of the covariance contributions from each state in (27). Also, using a constant d_{max} will guarantee a symmetric SHAC covariance matrix.

Kelejian and Prucha (2007) did not develop a panel data version of the SHAC estimator. I will expand the SHAC estimator to panel data by following a strategy similar to that used in expanding the GS2SLS estimator to panel data. In generalizing to panel data, the data matrices are stacks of cross sections and the spatial weight matrix W is block diagonal with W_n as each block. The panel data model is

$$\underline{y} = X\underline{\beta} + \lambda W\underline{y} + Y\underline{\gamma} + \underline{u}$$

$$\underline{u} = R\underline{\varepsilon}$$

where R is the $nT \times nT$ block diagonal matrix

(30) $$R = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R_T \end{bmatrix}$$

and the covariance matrix of $\underline{u}$ is the $nT \times nT$ block diagonal matrix

(31) $$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_T \end{bmatrix}.$$

Consider the vector of cross-products of the instrumental variables and the model errors, $(nT)^{-1/2}H'\underline{u}$. Its covariance matrix is the $p \times p$ matrix $\psi = (nT)^{-1}H' \, \Sigma \, H$, as shown below

(32) $$\Psi = (nT)^{-1}[H_1' \;\; H_2' \;\dots\; H_T'] \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_T \end{bmatrix}$$

$$\Psi = (nT)^{-1}[H_1'\Sigma_1 H_1 + \cdots + H_T'\Sigma_T H_T] = (nT)^{-1}\sum_{t=1}^{T} H_t'\Sigma_t H_t$$

$$\Psi = (nT)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \underline{h}_{i,1}\underline{h}_{j,1}'\sigma_{ij,1} + \cdots + (nT)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \underline{h}_{i,T}\underline{h}_{j,T}'\sigma_{ij,T}.$$

The structural portion of the model is estimated with $\underline{\tilde{\delta}}$, the instrumental variable (or two-stage least squares) estimator of $\underline{\delta} = (\underline{\beta}', \lambda, \gamma')'$ that is also used in the first step of GPTSLS. $\underline{\tilde{\delta}}$ is

(33) $$\underline{\tilde{\delta}} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'\underline{y} = [Z'H(H'H)^{-1}H'Z]^{-1}Z'H(H'H)^{-1}H'\underline{y}_n$$

where $Z = \left(X, W\underline{y}, Y\right)$, $\hat{Z} = PZ$, and $P = H(H'H)^{-1}H'$. Denote the $nT \times 1$ vector of estimation residuals by $\underline{\hat{u}} = \underline{y} - Z\underline{\tilde{\delta}}$. The SHAC estimator is

(34) $\hat{\Psi} = (nT)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\underline{h}_{i,1}\,\underline{h}'_{j,1}\hat{u}_{i,1}\hat{u}_{j,1}K(\frac{d_{ij,1}}{d_1}) + \cdots$

$$+(nT)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\underline{h}_{i,T}\,\underline{h}'_{j,T}\hat{u}_{i,T}\hat{u}_{j,T}K(\frac{d_{ij,T}}{d_T})$$

where $\underline{h}'_{g,t}$ is row g of H_t, $\underline{h}_{g,t}$ is the column vector containing the transpose of row g, and $\hat{u}_{i,t}$ is the residual pertaining to cross sectional unit i in cross section t. Since all distances among population centrods is from one year, the Parzen kernel for each cross section (year) is identical. The consistent estimator of the covariance matrix of $\underline{\tilde{\delta}}$ is

(35) $\hat{\Phi} = (nT)^2\left(\hat{Z}'\hat{Z}\right)^{-1}Z'H(H'H)^{-1}\hat{\Psi}(H'H)^{-1}H'Z(\hat{Z}'\hat{Z})^{-1}$.

Other than the GSTSLS by Kelejian and Prucha (1998) and spatial HAC estimator by Kelejian and Prucha (2007), Conley (1999) also proposed a GMM estimation technique for cross sectional dependence models. Conley considered a spatial model where he relaxed the independence assumptions of the variables between each cross sectional unit. That is, under the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, the data in X for cross sectional unit i is dependent with cross sectional unit j, given that i is not equal to j. Conley noted that the dependency of observations for cross sectional units i and j is unobservable. Thus, the author used the distance between cross sectional units i and j to reflect the unobservable dependency of the dependent variable. The term “distance” can be the physical distance between two cross sectional units or other measures that can be use to account for the unobserved dependency between the two cross sectional units. The spatial weights matrix used by Conley is based on the distances between cross sectional units. The reason Conley uses distance in the spatial weight matrix is because he believes that if cross sectional unit i and j are close to each other, then the observation for the variable in cross sectional unit i may be highly correlated with the observation in cross sectional unit j. As the distance between cross sectional unit i and cross sectional unit j grows, the observations on variables for cross sectional units i and j become closer to being independent.

Conley showed that GMM estimators remain consistent even when the observations on cross sectional unit i are not independent of observations on cross sectional unit j, with i not equal to j. In order to get efficient GMM estimates of $\underline{\beta}$ in the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where the data in X is dependent between cross sectional units, a consistent estimator of the asymptotic covariance matrix is needed. Conley used the weight matrix as a proxy to capture the unobserved dependency effect between cross sectional units to form the consistent estimator of the covariance matrix. The weight matrix is formed with the element in the i, j position being the distance between two cross sectional units i and j.

The GMM estimation method described by Conley does not address estimation methods for models which include a spatial lag. Conley's approach does allow for general forms of error dependence whereas the model in (13) only allows for a first-order autoregressive spatial error structure. Thus, one might argue that Conley's GMM estimation method should be used rather than assume a first-order autoregressive spatial error process. However, the first-order autoregressive spatial error process is a popular assumption and I choose to use the SHAC, rather than Conley's methods, for relaxing the spatial error process assumption.

4.2.3. Region Specific Spatial Model

The spatial models discussed above assume the spatial lag parameter, λ, to be the same for all states. It may be the case that the magnitude of the spatial effect varies across states and regions. For example, states in the East Coast region may have a different level of spatial effect upon one another as compared to states in the Plains region, due to differences in the nature of their regions. Garrett, et al. (2007) extend the traditional spatial correlation model by allowing the spatial lag coefficient to vary across geographic regions of states. In a panel data context, their model is

(36) $$\underline{y} = X\beta + \sum_{k=1}^{Q} \lambda_k W_k \underline{y} + \underline{u}$$

$$\underline{u} = \sum_{k=1}^{Q} \rho_k W_k \underline{u} + \underline{\epsilon}$$

where Q denotes the total number of regions, and $\rho_k \; and \; \lambda_k$ denote the spatial lag and spatial autoregressive coefficients, respectively, for region k. Let $w_k = R_k W_n$ be the spatial weight matrix for region k where R_k is an $n \times n$ matrix whose diagonal element in position i, i is 1 if state i is in the region k and zero otherwise. All off-diagonal elements of R_k are zero. Then $W_k = I_{TxT} \otimes w_k$ is the weights matrix for region k in the panel of T time periods. That is, W_k is a block diagonal matrix with T repetitions of w_k as each block. Other facts of interest are $\sum_{k=1}^{Q} R_k = I_{nxn}$ and $\sum_{k=1}^{Q} w_k = W_n$. With the contiguity, contiguity with inverse-distance, and contiguity with inverse-distance squared weights matrices, the spatial lag model allows the growth in state i in region k to be influenced by the growth in all states that share a common border with state i, regardless if the border-sharing states are in region k or not. For the inverse-distance and inverse-distance-squared weights matrices, the model allows the growth in state i in region k to be influenced by the growth of all other states. If state i is not in region k, then the i^{th} row of w_k is zero.

The model in (36) can be estimated using the SHAC method described in section 4.2.2.4 with the Z, P and H matrices being constructed to reflect the case of regional specific patterns. The Z matrix is defined as $Z = (X, W_1\underline{y}, W_2\underline{y}, \dots, W_Q\underline{y})$, $P = H(H'H)^{-1}H'$, and H is the matrix of p instrumental variables that are composed of a subset of the linearly independent columns of $\left(X, W_1X, \dots, W_qX\right)$. If other endogenous variables are present, then $Y\underline{\gamma}$ is added in (36) and X is replaced by $\overline{X} = (X, G)$ or some similar matrix containing instrumental variables.

The regions used by Garrett et al. are either the four regions or the nine divisions of the U.S. as classified by the Bureau of the Census. The dependent variable is the growth rate of real per capita state personal income and the regressors in X include the labor force participation rate, measures of education attainment and state industrial diversity, state expenditures as percentage of state gross domestic product, and local government revenue as a share of state and local revenue. The data used is from the 48 contiguous states of the United States over the span of 1977 to 2002. Garrett, et al. considered two types of spatial weights matrices, the contiguity matrix and the inverse-distance weights matrix. Distance is measure as the distance between the state population centroids.

Garrett et al. estimated two specifications of spatial models above, one that did not allow for regional differences in the spatial lag coefficient and one that did as in (36). They found that state income growth rates are positively related to the size of a state's labor force, government sectors, and industrial diversity. For the model specification with no regional differences in the spatial lag coefficient, the estimation result revealed strong evidence of spatial correlation. They also found that both spatial weights matrices (contiguity and inverse-distance) yield similar estimates of the spatial lag coefficients and spatial autoregressive coefficients both in terms of magnitude and statistical significance level. Results from their models suggested that a one percentage point increase in the average income growth rate in neighboring states generates a .23 percent increase in state $i's$ income growth rate.

For the second model specification which allows for regional differences in the spatial lag coefficient, Garrett et al. found that state income growth rates exhibited the characteristic of spatial correlation but the spatial lag coefficient varied substantially by region. Dividing the U.S. into four regions (Midwest, West, South, and Northeast), they found that the spatial lag coefficient for the Northeast region is nearly twice as large as the spatial lag coefficients in the other three regions. They believed that the strong spatial correlation in the Northeast region reflected the Northeast region's characteristic of small sized states with large populations, which may lead to higher degrees of spillovers effects in the region.

Separating the U.S. into nine divisions as defined by the Census Bureau, Garrett et al. found evidence of spatial correlation in state income growth rates for some divisions, but not all. They also found that several divisions appear to be affected more strongly by their neighboring states, rather than all states in their region. That is, income growth rates in neighboring states (not necessary in the same region) are more important to the income growth rate of the home state, rather than states in the same region but farther away. The results also showed negative spatial correlation in state income growth rates for the West South Central division when using the inverse-distance weights matrix and showed no spatial correlation when the contiguity weights matrix was used. The reason for these mixed results on the spatial lag coefficient may be due to smaller numbers of states being in some divisions as compared to numbers of states in any one of the four regions.

Garrett et al. considered one industrial diversity measure and two spatial weights matrices in their model specifications. Instead of one diversity measure, I use five industrial diversity measures to explore the effect of diversity upon economic growth. I also consider five spatial weights matrices to see if different weights matrices produce different results. Garrett et al. did not test for endogeneity of the independent variables. It could very well be the case that the labor growth rate is endogenous and if that is the case, then estimation without the use of instrumental variables will be inconsistent. Also, industrial diversity may be another source of endogeneity in their model. Thus, I will perform endogeneity tests for the independent variables in the earnings growth models and compensate for the endogeneity problem, if needed, by using instrumental variable methods. Also, two different methods of spatial estimation are applied in my study to, in some way, assure the robustness of the results.

5. Descriptive Statistics and Estimation Results

5.1. Descriptive Statistics for Diversity Indices and Measures of Economic Growth.

In this chapter, results from estimating the model of nonfarm earnings growth are presented. Prior to discussing estimation results, descriptive statistics for the industrial diversity indices and economic growth are presented. Definitions of the indices appear in equations (1) through (5) in Chapter 3. Table 5.1.1 lists the five most and least diverse state economies (out of 48) as ranked according to the most recent year of annual data, 2009. The high, low, and mean are also calculated. When viewing the index values in Table 5.1.1, recall that higher Entropy values signal greater diversity while lower values of the other four indices signal greater diversity.

In identifying the group of the most diverse states for 2009, there exists minor overlap of states across the two categories of diversity indices (ogive – uniform and national average – national distribution norms) but substantial overlap among the indices within each category. U.S. industry employment shares are not uniform across industries, thus indices based on U.S. shares are expected to differ somewhat from indices based on the norm of uniform shares. Missouri is the only state that is classified as one of the most diverse states according to four of the five indices, the exception being the $OG2$ index. Utah and Georgia are classified as most diverse states according to the Entropy, $OG1$, and $OG2$ indices. For the group of least diverse states, the overlap of states across the five indices is presented. In particular, Wyoming and Nevada are among the least diverse states in all five indices. Vermont is one of the least diverse states according to the Entropy, $OG1$, and $OG2$ indices.

Table 5.1.1

Industrial Diversity Indices - State Extremes, 2009					
Most Diverse					
	Entropy	OG1	OG2	NA1	NA2
	Utah	Missouri	Utah	Missouri	Missouri
	Georgia	Georgia	Georgia	North Carolina	Idaho
	Colorado	Utah	Colorado	Oregon	Illinois
	Missouri	Pennsylvania	Virginia	Texas	California
	Virginia	Nebraska	Illinois	Illinois	Oregon
Least Diverse					
	New Hampshire	Mississippi	New Hampshire	Delaware	New Mexico
	Maine	Vermont	Vermont	New Mexico	Montana
	Wyoming	New Mexico	Maine	Montana	West Virginia
	Vermont	Wyoming	Wyoming	Nevada	Nevada
	Nevada	Nevada	Nevada	Wyoming	Wyoming
High	2.874	0.767	1.305	0.465	1.469
Low	2.634	0.527	0.437	0.058	0.012
Mean	2.793	0.614	0.638	0.181	0.107

Table 5.1.1 also shows the high, low, and the mean of the diversity indices for 2009. It is important to note that the $OG2$ and $NA2$ diversity indices for Wyoming and Nevada are much higher (less diverse) compared to other states. The $OG2$ measure for Nevada is 1.305 while the $OG2$ average value of all states is 0.638. Wyoming is similar to Nevada in the sense that the $NA2$ measure for Wyoming is 1.469 while the average of all states is 0.107. Large values of the $OG2$ and $NA2$ indices indicate that these two states are highly concentrated economies. Wyoming's largest industry, in terms of employment, is local government, which accounted for 16.5 percent of total employment in 2009. In 2009, the five largest industries (out of 22) accounted for 55

percent of employment in Wyoming[13]. For Nevada, accommodation and food services is the largest industry, accounting for 24.5 percent of the state's employment in 2009. The largest four industries (out of 22) in Nevada accounted for roughly 52.4 percent of the state's employment[14].

Table 5.1.2 lists the 48 states in order of growth rates based upon real nonfarm earnings for 2009. The table also shows the high, low, and the mean growth rates for the 48 states. In most cases, states in the most diverse group in Table 5.1.1 generally have economic growth rates in the top 50th percentile. That is, states with higher diversity tend to have relatively higher growth rates. On the other hand, most states in the least diverse group in Table 5.1.1 generally have low economic growth rates, as shown in Table 5.1.2. States such as Mississippi, Vermont, Delaware, and Maine, on average, had slow growth compared to other states and these states appear in the least diverse group shown in Table 5.1.1. Thus, to some degree, the notion that well diversified economies have higher economic growth, the latter being real nonfarm earnings growth, is supported by the patterns in Tables 5.1.1 and 5.1.2.

[13] Local government, accommodation and food services, retail trade, natural resources and mining, and construction.

[14] Accommodation and food services, retail trade, local government, and health care and social assistance.

Table 5.1.2

Real Nonfarm Earnings Growth Rate for 48 States in 2009	
Top Half	Bottom Half
Maine	Indiana
Kentucky	Nevada
Rhode Island	Nebraska
Louisiana	New York
Delaware	Ohio
Mississippi	Washington
Connecticut	Kansas
New Jersey	Wyoming
Massachusetts	Iowa
Virginia	Minnesota
South Carolina	Arizona
New Hampshire	Utah
North Carolina	Michigan
Florida	South Dakota
Maryland	Idaho
Oregon	Oklahoma
Alabama	Montana
Georgia	Texas
Illinois	Pennsylvania
Arkansas	New Mexico
California	Vermont
Colorado	Tennessee
North Dakota	West Virginia
Missouri	Wisconsin
High	1.97
Mean	-3.86
Low	-9.15

5.2. Estimation Results for the Nonfarm Earnings Growth Model

Recall the nonfarm earnings growth model in (6) is specified as:

$$(37) \quad GRO_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 NFEM_{it} + \beta_4 K_{it} + \beta_5 FER_{it} + \sum_{j=1}^{7} \delta_j D_{j,it} + \sum_{j=1}^{T-1} \gamma_j DS_{j,it} + \varepsilon_{it}.$$

The panel data set consists of annual data for 48 states over 1992-2009. As noted in Chapter 4, tests of endogeneity for several of the independent variables are needed. If one or more independent variables in a model are endogenous, then two-stage least squares estimation is required to obtain consistent estimators. The Hausman test for endogeneity that was discussed in Section 4.1.1 is used. Table 5.2.1 contains the test results for each variable (t statistics) and the joint test (F statistic) for the group when using each of the diversity indices in turn.

Table 5.2.1: Hausman Test Results

	Hausman Test Statistics				
	Entropy	OG1	OG2	NA1	NA2
Diversity	-1.080	1.843***	0.357	.056	-1.249
Nonfarm Employment	-2.125**	-2.280**	-2.170**	-2.069**	-1.795***
Capital	5.497*	5.533*	5.448*	5.483*	5.570*
Joint Test	10.666*	11.410*	10.209*	10.439*	11.572*

Note: ***/**/* - significant at the .10/.05/.01 levels, respectively.

Based on the test results in Table 5.2.1, instrumental variable methods are used to estimate all of the model specifications. The model variables that require representation by the presence of instrumental variables are the $OG1$ diversity index, the nonfarm employment growth rate ($NFEM$), and capital growth rate (K).

5.2.1. Panel Unit Root Tests

When using panel data for model estimation, checks for nonstationary behavior of the model variables within the panel are needed. Four variables are tested for unit roots: growth rate of nonfarm earnings (GRO), growth rate of nonfarm employment ($NFEM$), growth rate of capital (K), and growth rate of farm earning (FER). I use the Im, Persaran and Sin (2003) test for panel unit roots. The null hypothesis for the Im, Persaran and Sin test is the presence of unit roots. The test assumes individual unit root processes for each cross sectional unit. Table 5.2.2 shows the test results for $GRO, NFEM, K$, and FER. Based on the unit root test results, we reject the null hypothesis of unit roots for all four variables. That is, we estimate (37) using panel data without worrying about the presence of unit roots in the $GRO, NFEM, K$, and FER variables.

Table 5.2.2: Panel Unit Root Tests

	Panel Unit Root Tests			
	GRO	NFEM	K	FER
Im, Pesaran and Shin	-3.86**	-1.67**	-10.61*	-24.89*

Note : Null hypothesis for Im, Pesaran and Sin assumes individual unit root processes.
* - Significant at α=.01
** - Significant at α=.05

5.2.2. Basic Model Results

Table 5.2.3 shows the results from estimating (37) using annual data for the 48 contiguous states over the span of 1992 to 2009. Although available data begins in 1990, one year is lost in creating growth rates and an additional year is lost due to the fact that the lags of nonfarm employment growth and the $OG1$ index are needed as instrumental variables. The t-statistics are based on panel-corrected standard errors that are adjusted for cross sectional heteroskedasticity and for covariance of errors between cross sectional units using the Beck and Katz (1995) method as described in Chapter 4. Besides the issue of cross sectional heteroskedasticity, the issue of endogeneity also requires attention. The Hausman test results in Table 5.2.1 showed that allowance must be made for the endogeneity of nonfarm employment and capital, and $OG1$. Accordingly, all five of the specifications in Table 5.2.3 were estimated with the basic instrumental variables estimator. The set of instrumental variables is the lag of $OG1$, the lag of the nonfarm employment growth rate, and the z_1, and z_4 variables for the capital growth rate. To save space, the coefficients for the region and time dummy variables are not shown in any of the tables containing model estimation results.

The results show that the variables carry expected signs, with a positive coefficient for Entropy and negative coefficients for the pairs of the ogive and national average indices. Recall that an increase in the Entropy index indicates a state moving from a less diverse to more diverse industrial structure, while an increase in the ogive or national average indices indicates a state moving from a more diverse to a less diverse industrial structure. Thus, the estimated coefficients for diversity indices all indicate that higher degrees of diversity promote higher growth rates of nonfarm earnings for the state. The absolute magnitudes of coefficients on diversity indices range from 1.173 for $NA2$ to 3.251 for Entropy. The t-statistics for the diversity indices are high in absolute terms, ranging from 3.27 for Entropy to 5.51 for $NA2$, and all are statistically significant at the 0.01 level. All five models have relatively high R-square measures, indicating that the models fit the data well.

Table 5.2.3

Models of the Growth Rate of Nonfarm Earnings:

Variables	Coefficients				
ENTROPY	3.251 (3.27)*				
OG1		-2.780 (-3.44)*			
OG2			-1.157 (-4.28)*		
NA1				-2.256 (-4.69)*	
NA2					-1.183 (-5.587)*
NFEM	0.840 (11.27)*	0.835 (11.39)*	0.857 (12.40)*	0.849 (12.51)*	0.834 (12.58)*
K	0.250 (4.74)*	0.256 (4.77)*	0.249 (5.34)*	0.247 (5.30)*	0.250 (5.38)*
FER	0.001 (0.68)	0.001 (0.68)	0.001 (0.63)	0.001 (0.51)	0.000 (0.33)
Constant	-9.296 (-3.47)*	1.424 (2.02)**	0.671 (1.36)	0.200 (0.47)	-0.122 (-0.32)
R-Square	0.876	0.877	0.878	0.878	0.879

Note : ***/**/* - significant at the .10/.05/.01 levels, respectively. t-statistics are in parentheses and are based upon standard errors that are adjusted for cross-section heteroskedasticity.

To illustrate the magnitude of the diversity effect, focus on the Entropy coefficient, 3.251, and consider the Entropy index values for 2009 reported in Table 5.1.1. A move from the mean position of the states to the highest degree of diversity would entail an increase of about 0.08 in the Entropy index. Such an increase would add approximately 0.26 percents to the growth rate of nonfarm earnings. If we consider moves from the mean to the highest degrees of diversity across the other four indices, the decreases would be around 0.09, 0.20, 0.12, and 0.10, respectively. The corresponding increases in the growth rate of nonfarm earnings would be approximately 0.25, 0.23, 0.27, and 0.12 percents, respectively. Additional results concerning marginal effects upon growth rates from changes in the diversity indices are shown in Section 5.5.

Not surprisingly, the growth rate of nonfarm employment positively influences the growth rate of nonfarm earnings and is highly significant. The estimated coefficient of $NFEM$ for the five model specifications, utilizing five different diversity indices, were in the range of 0.835 for the model using $OG1$ to 0.857 for the model using $OG2$. The t-statistics are very high, ranging from 11.27 to 12.60. Note that the magnitude of the coefficient for $NFEM$ does not vary much across the five model specifications. This suggests that, on average, an increase of one percent in the growth rate of nonfarm employment contributes to an increase of 0.84 percents in the growth rate of nonfarm earnings. With a high marginal effect and strong statistical significance, nonfarm employment is one of the most important and strongest driving forces for state economies.

Together with the growth rate of nonfarm employment, the growth rate of capital also significantly influences the growth rate of nonfarm earnings. The estimated coefficients for K under the five model specifications range from 0.241 to 0.250. The t-statistics for K are also high, ranging from 4.74 to 5.34, all being significant at the 0.01 level. Similar to $NFEM$, the coefficients for K do not differ much from each other across the five model specifications, with different diversity measures. Generally, an increase of one percent in the growth rate of capital contributes to an increase of 0.25 percents to the growth rate of nonfarm earnings.

Notice that the sum of the coefficients for the nonfarm employment growth rate and capital growth rate are close to one. Furthermore, the ratio of employment to capital is about 3:1, which suggests that if employment and capital growth rates increase by one percent, then employment contributes roughly 0.75 percent to state economic growth and capital contributes roughly 0.25 percent to state economic growth.

Farm earnings growth does not provide a significant influence upon nonfarm earnings growth, both in terms of statistical significance and magnitude. The reason is that farm earnings growth is very volatile from year to year. Thus, on an annual basis, it is difficult to pick up any significant contribution from farm earnings to nonfarm earnings growth.

5.3. Estimation Results for Spatial Models

This section presents the results from estimating spatial dependence models of the growth rate in nonfarm earnings. The basic spatial model is

(38) $$GRO_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 NFEM_{it} + \beta_4 K_{it} + \beta_5 FER_{it} + \lambda \sum_{j=1}^{n} w_{ij} GRO_{jt} + \sum_{j=1}^{7} \delta_j DR_{j,it} + \sum_{j=1}^{T-1} \gamma_j \, DS_{j,it} + u_{it}.$$

When a spatial autoregressive process for the errors is assumed, $\underline{u} = \rho W \underline{u} + \underline{\varepsilon}$, the GSTSLS estimator is used. For the more general assumption concerning the errors, $\underline{u} = R\underline{\varepsilon}$, the SHAC estimator is used.

According to the Hausman test results in Table 5.2.1, the nonfarm employment growth rate $(NFEM)$ and capital growth rate (K) are endogenous, therefore creating the need for instrumental variables. The instrumental variables are $(\underline{NFEM}_{t-1}\ \ \underline{z}_1\ \ \underline{z}_4)$. For the case of $OG1$ index, the set of instrumental variables is $(\underline{NFEM}_{t-1}\ \ \underline{z}_1\ \ \underline{z}_4\ \underline{OG1}_{t-1})$. These vectors comprise the G matrix that is required in the instrumental variable estimator.

For comparison purposes, I first present the results from using the GSTSLS estimator followed by the results from using the SHAC estimator. Prior to discussing the estimation results, it is helpful to review the notation for the spatial weights matrices. W_1 is the contiguity matrix, W_2 is the inverse-distance matrix, W_3 is the inverse-distance-squared matrix, W_4 is the combination of the contiguity and inverse-distance matrices, and W_5 is the combination of the contiguity and inverse-distance-squared matrices.

5.3.1. <u>GSTSLS Estimation Results</u>

This section presents estimation results for the model in (38) obtained by using the GSTSLS estimator. The full matrix of instrumental variables for the GSTSLS estimator is $H = \left(\overline{X}, W\dot{\overline{X}}, W^2\dot{\overline{X}}\right)$ where $\overline{X} = (X, G), X = \left(\underline{1}\ \underline{V}\ \underline{FER}\ \underline{DR}_1 \dots \underline{DR}_7 \underline{DS}_1 \dots \underline{DS}_{17}\right), G =$

$(\underline{NFEM}_{t-1}\ \underline{z}_1\ \underline{z}_4)$, $\dot{\overline{X}} = (\dot{X}, G)$, and $\dot{X} = (\underline{V}\ \underline{FER}\ \underline{DR}_1 \ldots \underline{DR}_7)$. Recall that $OG1$ is endogenous, thus for the case of $V = OG1$ then $G = (\underline{NFEM}_{t-1}\ \underline{z}_1\ \underline{z}_4\ \underline{OG1}_{t-1})$. The fixed effects dummy variables for years were dropped from X to create $\dot{X}$ due to near linear dependencies that occurred when attempting to use $W\overline{X}$ and $W^2\overline{X}$ in H.

Table 5.3.1 contains results for five model specifications using the Entropy measure in conjunction with the five weights matrices.

Table 5.3.1

Spatial Models - Entropy Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-9.027 (-4.17)*	-8.666 (-3.70)*	-8.684 (-3.90)*	-9.051 (-4.22)*	-9.146 (-4.27)*
Entropy	2.914 (3.67)*	2.964 (3.61)*	2.956 (3.64)*	2.938 (3.72)*	3.012 (3.81)*
NFEM	0.746 (11.50)*	0.748 (11.80)*	0.747 (11.84)*	0.756 (11.75)*	0.773 (12.10)*
K	0.344 (8.75)*	0.340 (8.52)*	0.341 (8.66)*	0.34 (8.63)*	0.333 (8.38)*
FER	0.000 (0.53)	0.001 (0.58)	0.001 (0.62)	0.001 (0.62)	0.001 (0.67)
λ	-0.019 (-0.24)	0.104 (0.51)	0.108 (1.04)	-0.025 (-0.35)	-0.021 (-0.32)
ρ	0.219	0.161	0.128	0.220	0.199
R-Square	0.895	0.893	0.894	0.894	0.894

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Similar to the basic (non-spatial) model in Table 5.2.3, the spatial model has a positive coefficient for the Entropy measure. The coefficient ranges from 2.914 for the model using W_1 to 3.012 for the model using W_5. The average of the Entropy coefficients across the five models is 2.96, which is similar to the estimated coefficient in the non-spatial model. The coefficients for entropy are all statistically significant at the .01 level. The spatial model results suggest that an increase of one unit in the Entropy diversity measure contributes to an increase of 2.96 percent in the growth rate of nonfarm earnings. Thus, after allowing for a spatial lag and a spatial autoregressive error in the model, the results continue to show that states with higher levels of diversity experience higher economic growth, as measured by the nonfarm earnings growth rate, holding other factors constant.

Similar to the non-spatial model results presented in Table 5.2.3, nonfarm employment growth continues to play an important role in enhancing state economic growth, as measured by the nonfarm earnings growth rate. The estimated coefficients for $NFEM$ ranges from 0.746 for the model using W_1 to 0.773 for the model using W_5. The average of the estimated coefficients from the five models is 0.75, which is a little smaller when compared to the average of 0.84 in the non-spatial model results. All t-statistics for $NFEM$ are high and significant at the 0.01 level. The estimated marginal effect from $NFEM$ upon the nonfarm earnings growth rate is roughly 0.75 percents.

The growth rate of capital continues to play an important role in influencing the nonfarm earnings growth rate. The results suggest that for an increase of one percent in the growth rate of capital contributes to an increase of 0.34 percents in the nonfarm earnings growth rate. The estimated marginal effects of capital growth upon nonfarm earnings growth are statistically significant at the 0.01 level for all five model specifications.

Notice that the marginal effects of $NFEM$ are lower for the spatial models as compared to the non-spatial model. At the same time, the marginal effects of K are higher in the spatial models as compared to the non-spatial model. Even though the magnitude of the coefficients for $NFEM$ and K change a little between the spatial and non-spatial models, the sum of the coefficients remain close to one. The spatial model continues to suggest that the employment to capital ratio of marginal effects is roughly 3:1.

Farm earnings growth remains insignificant as a contributor to nonfarm earnings growth, both in the statistical sense and in magnitude, as was the case in the non-spatial model. As stated before, the primary reason that farm earnings growth is not significant is because farm earnings are very volatile from year to year.

As for the spatial autoregressive error and spatial lag parameters, they are small in scale. The spatial autoregressive error parameters are positive and range from .128 for the model using W_3 to 0.220 for model using W_4. These parameters can be interpreted in the same way as the autoregressive parameter in a time series model. The estimates of the spatial lag parameters are small, ranging from 0.019 to 0.108 in absolute terms and all are statistically insignificant. Thus, the results for models with the Entropy diversity index suggest that economic growth of neighboring states does not have a significant effect upon a home state's economic growth, as measured by the nonfarm earnings growth rate. Estimation results for the spatial model using the other four indices are similar to those just discussed for the Entropy index. A full discussion of each set of results would become overly redundant. However, it would be helpful to fully discuss results from using one of the other four indices so as to illustrate the similarity of the results that will be apparent when viewing the upcoming tables. The Entropy index is based on the norm of equal shares of economic activity across industries. So are the ogive indices $OG1$ and $OG2$. In contrast, the national average indices $NA1$ and $NA2$ are based on the norm of the industry shares in the U.S. economy. For variety, we select the estimation results for the spatial model using the $NA1$ index for detailed discussion.

Table 5.3.2 presents the estimation results for models containing the $NA1$ diversity index while utilizing the five weights matrices. Under these specifications, the coefficients of $NA1$ carry the expected negative sign and are all statistically significant at the .01 level. The coefficients are consistent across the five model specifications with different weights matrices, and they range from 2.086 for the model using W_1 to 2.192 for the model using W_5, in absolute terms. The average of the estimated coefficients for the $NA1$ index across the five models is 2.13, in absolute terms, which is similar to the average of the estimates from the non-spatial model in Table 5.2.3. Thus, the spatial model continues to suggest that diversity, as measured by the $NA1$ index, plays a positive and significant role in influencing state economic growth.

Similar to the non-spatial model results presented in Table 5.2.3, nonfarm employment growth plays an important role in state economic growth. The estimated coefficients for $NFEM$ range from 0.774 for the model using W_2 to 0.809 for the model using W_5. The average of the estimated coefficients from the five models is 0.78, just a bit smaller than the average of 0.84 from the non-spatial model, but slightly higher than the average of 0.75 from the spatial models that used the Entropy index. All t-statistics for $NFEM$ are high and significant at the .01 level.

Table 5.3.2

Spatial Model - NA1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.441 (-1.04)	0.312 (0.43)	0.081 (0.17)	-0.255 (-0.60)	-0.104 (-0.24)
NA1	-2.086 (-4.89)*	-2.103 (-4.72)*	-2.117 (-4.83)*	-2.133 (-4.97)*	-2.192 (-5.08)*
NFEM	0.779 (11.99)*	0.774 (12.19)*	0.777 (12.39)*	0.790 (12.22)*	0.809 (12.57)*
K	0.323 (8.18)*	0.319 (7.88)*	0.321 (8.03)*	0.310 (7.72)*	0.303 (7.46)*
FER	0.000 (0.41)	0.000 (0.46)	0.000 (0.51)	0.000 (0.51)	0.001 (0.55)
λ	-0.049 (-0.64)	0.165 (0.82)	0.107 (1.02)	-0.019 (-0.26)	-0.006 (-0.09)
ρ	0.266	0.149	0.150	0.230	0.200
R-Square	0.895	0.893	0.894	0.895	0.894

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

The capital growth rate, K, strongly influences the nonfarm earnings growth rate with coefficients that are similar to those from the previous spatial model that used the Entropy index. A one percent increase in the capital growth rate contributes anywhere from a 0.303 to 0.323 percents increase in the nonfarm earnings growth rate. The estimated coefficients for K are statistically significant at the 0.01 level across all five model specifications. Again, the combined contributions from one unit increases in the nonfarm employment and capital growth rates upon the nonfarm earnings growth rate is close to one. Farm earnings growth does not significantly contributes to the nonfarm earnings growth, both in terms of statistical significance and in magnitude. This lack of influence also occurred in the estimates from the Entropy -based model in Table 5.3.1 and the non-spatial model in Table 5.2.3

As for the spatial autoregressive error and spatial lag parameters, the estimates are similar to those in Table 5.3.1. The spatial autoregressive error parameter ranges from 0.149 to 0.266. The spatial lag parameters are small in magnitude and statistically insignificant. The models based upon the NA1 index show no significant spillover effects on an annual basis from growth in neighboring states upon growth in a home state.

Tables 5.3.3, 5.3.4, and 5.3.5 present estimation results for the spatial models using the $NA2$, $OG1$, and $OG2$ indices, respectively. The results show that the estimated coefficients for the $NA2$, $OG1$, and $OG2$ are highly significant, all at the 0.01 level. The results continue to suggest that industrial diversity plays a positive role in enhancing nonfarm earnings growth. The magnitudes of coefficients on the $NA2$ diversity index are similar to each other across the different model specifications that use different weights matrices. The coefficients of the $NA2$ index range from 1.078 to 1.099, in absolute terms, when using the contiguity weights matrix and the combination of contiguity and inverse-distance-squared weights matrix, respectively. Similarly, when using the $OG1$ and $OG2$ indices in turn, the magnitude of coefficients on the diversity indices are similar across the five model specifications with different weights matrices.

The coefficients for $OG1$ range from 2.626 to 2.696, in absolute terms, when using the contiguity weights matrix and the inverse-distance-squared weights matrix, respectively. The coefficients of $OG2$ range from 1.096 to 1.055, in absolute terms, when using the distance-inverse weights matrix and the combination of contiguity and inverse-distance-squared weights matrix, respectively.

Table 5.3.3

Spatial Model - NA2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.676 (-1.72)***	-0.235 (-0.32)	-0.291 (-0.65)	-0.569 (-1.43)	-0.545 (-1.38)
NA2	-1.078 (-5.83)*	-1.085 (-5.56)*	-1.093 (-5.68)*	-1.096 (-5.91)*	-1.099 (-5.90)*
NFEM	0.772 (12.30)*	0.753 (12.50)*	0.761 (12.70)*	0.785 (12.57)*	0.781 (12.71)*
K	0.313 (7.97)*	0.318 (8.07)*	0.315 (7.97)*	0.303 (7.60)*	0.300 (7.50)*
FER	0.000 (0.27)	0.000 (0.32)	0.000 (0.38)	0.000 (0.37)	0.000 (0.43)
λ	-0.047 (-0.62)	0.100 (0.49)	0.081 (0.78)	-0.041 (-0.57)	-0.032 (-0.48)
ρ	0.264	0.197	0.173	0.252	0.224
R-Square	0.885	0.888	0.889	0.884	0.884

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.3.4

Spatial Models - OG1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	0.761 (1.29)	1.433 (1.75)***	1.224 (1.95)***	0.780 (1.32)	0.883 (1.48)
OG1	-2.626 (-4.05)*	-2.696 (-4.04)*	-2.696 (-4.09)*	-2.634 (-4.08)*	-2.669 (-4.12)*
NFEM	0.730 (11.56)*	0.733 (11.99)*	0.733 (11.98)*	0.739 (11.86)*	0.757 (12.22)*
K	0.350 (8.90)*	0.346 (8.77)*	0.347 (8.87)*	0.348 (8.88)*	0.340 (8.62)*
FER	0.000 (0.51)	0.001 (0.55)	0.001 (0.59)	0.001 (0.59)	0.001 (0.65)
λ	0.028 (0.38)	0.185 (0.98)	0.139 (1.39)	0.014 (0.20)	0.006 (0.10)
ρ	0.172	0.075	0.095	0.181	0.170
R-Square	0.890	0.889	0.890	0.889	0.888

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.3.5

Spatial Models - OG2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.068 (-0.14)	0.386 (0.50)	0.385 (0.75)	-0.004 (-0.01)	0.154 (0.32)
OG2	-1.056 (-4.35)*	-1.062 (-4.25)*	-1.055 (-4.27)*	-1.069 (-4.43)*	-1.096 (-4.54)*
NFEM	0.787 (11.77)*	0.782 (11.98)*	0.783 (12.02)*	0.798 (12.01)*	0.815 (12.33)*
K	0.332 (8.40)*	0.331 (8.29)*	0.333 (8.45)*	0.331 (8.38)*	0.324 (8.13)*
FER	0.000 (0.52)	0.000 (0.58)	0.001 (0.62)	0.001 (0.61)	0.001 (0.67)
λ	-0.055 (-0.71)	0.067 (0.32)	0.078 (0.75)	-0.060 (-0.83)	-0.046 (-0.70)
ρ	0.258	0.208	0.162	0.255	0.225
R-Square	0.895	0.893	0.894	0.894	0.893

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Similar to the models that used the Entropy and $NA1$ indices, nonfarm employment growth and capital growth continue to play a dominant role in promoting nonfarm earnings growth in models that use the $NA2$, $OG1$, and $OG2$ indices. The magnitudes of coefficients are also similar to those reported in Tables 5.3.1 and 5.3.2. Similarly, farm earnings growth does not significantly contribute to nonfarm earnings growth.

As for the spatial autoregressive and spatial lag parameters, the estimates are similar to those in Table 5.3.1 and 5.3.2. The spatial autoregressive error parameters are small for models specified with $NA2$, $OG1$, and $OG2$ indices. The spatial lag parameters are also small and statistically insignificant. Thus, the models specified with $NA1, OG1$, and $OG2$ indices continue to suggest that the spillover effects from the growth of neighboring states upon growth of a home state are insignificant, regardless of which spatial weights matrix is used.

In summary, the spatial model results show that economic diversity is important to nonfarm earnings growth. The influence of economic diversity upon nonfarm earnings growth is positive and statistically significant. The degree of influence of diversity upon nonfarm earnings growth does not seem to differ between models that allow for spatial effects and models that do not allow for spatial effects. This suggests that diversity is an important component in designing state economic policies. Nonfarm employment and capital growth rates are also strong factors that influence nonfarm earnings growth rates and their coefficients sum up close to one.

At first glance, it is surprising that the spatial lag does not significantly influence economic growth. That is, there are no spillover effects from one state to the next on an annual basis. It may be the case that spillover effects do not take place in the short run. That is, if the surrounding states experience high economic growth in a given year, it takes some time for the spillover effects to take place. The state economy needs time to adjust to take advantage of the spillover effects from neighboring states. Thus, spatial correlation may be a longer run phenomenon rather than observable in the short run. For the spatial autoregressive error parameters, the estimated values are small in most of the spatial model specifications. Thus, similar to the spatial lag, the presence of spatial autocorrelation in the spatial model framework is minimal.

5.3.2. SHAC Estimation Results

This section presents estimation results for the model in (38) obtained by using the SHAC estimator with panel data from 1992 to 2009. The H matrix of instrumental variables for the coefficient estimator in (33) is the same as was used in the first step of GPTSLS. The error assumption is $\underline{u} = R\underline{\varepsilon}$ and leads to the use of the SHAC estimator of the covariance matrix given in (35).

Table 5.3.6 presents the estimation results for models using the Entropy diversity index with the five spatial weights matrices. The coefficients on the Entropy index are similar to those in Table 5.3.1 which are from estimating the models with GSTSLS. The t-statistics for the entropy measure are a little bit smaller compared to those estimated with GSTSLS, but remain statistically significant at the 0.01 level. Similarly, the coefficients on the nonfarm employment growth rate $(NFEM)$ and capital growth rate (K) are similar under the SHAC and GSTSLS estimators. The t-statistics for $NFEM$ and K are smaller under the SHAC estimator compared to t-statistics under the GSTSLS estimator, but remain statistically significant at the 0.01 level. Similar to Table 5.3.1, the sum of the estimated coefficients for $NFEM$ and K is close to one regardless of the weight matrix that is used. Farm earnings growth remained insignificant as was the case under the non-spatial model and the spatial model estimated by GSTSLS. Similarly, the spatial lag parameters under the SHAC estimator are similar to those under GSTSLS, that is, the models do not suggest the presence of spillover effects from nearby states upon nonfarm earnings growth in a home state.

Table 5.3.6

SHAC Estimation - Entropy Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-8.888 (-3.69)*	-8.579 (-3.34)*	-8.637 (-3.54)*	-8.987 (-3.74)*	-9.123 (-3.78)*
Entropy	2.892 (3.09)*	2.953 (3.18)*	2.942 (3.16)*	2.950 (3.17)*	3.035 (3.27)*
NFEM	0.733 (6.86)*	0.744 (7.61)*	0.740 (7.52)*	0.749 (7.26)*	0.770 (7.83)*
K	0.346 (3.58)*	0.339 (3.74)*	0.341 (3.77)*	0.337 (3.63)*	0.326 (3.65)*
FER	0.000 (0.60)	0.000 (0.62)	0.000 (0.62)	0.000 (0.61)	0.000 (0.63)
λ	0.004 (0.05)	0.128 (0.57)	0.119 (1.09)	-0.002 (-0.03)	-0.003 (-.04)
R-Square	0.888	0.888	0.890	0.887	0.886
Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.					

Table 5.3.7 presents the estimation results for models containing the $NA1$ diversity index. The coefficients for the $NA1$ index, $NFEM$, and K from SHAC estimation are similar in magnitude to the estimates from the GSTSLS method. The t-statistics for coefficients on $NA1$ are slightly smaller under the SHAC estimator as compared to GSTSLS but the t-statistics for $NFEM$ and K are noticeably lower. Nevertheless, the estimated coefficients for $NFEM$ and K are statistically significant at the 0.01 level. Farm earnings growth continues to be an insignificant influence upon nonfarm earnings growth, as was the case in Table 5.3.2. The spatial lag parameters for the five models are all small and statistically insignificant, as was the case under GSTSLS

estimation. SHAC estimates of the model with the $NA1$ index continue to suggest that the nonfarm earnings growth rate in nearby states does not influence the nonfarm earnings growth rate in home states.

Table 5.3.7

SHAC Estimation - NA1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.410 (-0.68)	0.340 (0.41)	0.067 (0.11)	-0.221 (-0.38)	-0.084 (-0.14)
NA1	-2.023 (-3.60)*	-2.084 (-3.71)*	-2.078 (-3.68)*	-2.076 (-3.73)*	-2.137 (-3.81)*
NFEM	0.765 (7.50)*	0.769 (8.15)*	0.767 (8.15)*	0.780 (8.20)*	0.802 (8.77)*
K	0.327 (3.57)*	0.320 (3.69)*	0.322 (3.70)*	0.308 (3.58)*	0.299 (3.58)*
FER	0.000 (0.48)	0.000 (0.49)	0.000 (0.49)	0.000 (0.51)	0.000 (0.52)
λ	-0.031 (-0.38)	0.184 (0.83)	0.120 (1.09)	-0.000 (-0.00)	0.007 (0.11)
R-Square	0.886	0.888	0.889	0.885	0.885

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = distance^{-1} and border contiguity, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Tables 5.3.8, 5.3.9, and 5.3.10 present the SHAC estimation results for spatial models using the $NA2$, $OG1$, and $OG2$ indices, respectively. The coefficients on the $NA2$, $OG1$, and $OG2$ indices are similar to those in the models estimated by GSTSLS, as shown in Tables 5.3.3, 5.3.4, and 5.3.5. The t-statistics are smaller compared to those estimated by GSTSLS, but all are statistically significant at the 0.01 level. The magnitudes of coefficients for $NA2$, $OG1$, and $OG2$ range from, in absolute terms, 1.082

to 1.116, 2.631 to 2.713, and 1.046 to 1.105, respectively. Notice that the estimated coefficients for each of the diversity indices do not fluctuate much across different models utilizing different weight matrices.

Similar to the estimated models using the Entropy and $NA1$ indices, the nonfarm employment growth rate and capital growth rate positively influence the nonfarm earnings growth rate. Farm earnings growth does not play a insignificant role in determining nonfarm earnings growth for models using $NA2$, $OG1$, and $OG2$ indices. For the spatial lag parameters, the SHAC estimation results show no evidence of spillovers effect from nonfarm earnings growth of nearby states upon nonfarm earnings growth in a home state, a pattern very similar to the results found when using the GSTSLS estimation method.

Table 5.3.8

SHAC Estimation - NA2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.632 (-1.14)	-0.151 (-0.18)	-0.278 (-0.48)	-0.489 (-0.90)	-0.459 (-0.86)
NA2	-1.084 (-3.15)*	-1.082 (-3.18)	-1.088 (-3.18)	-1.111 (-3.19)*	-1.116 (-3.21)*
NFEM	0.758 (7.69)*	0.747 (8.50)*	0.751 (8.47)*	0.780 (8.29)*	0.782 (8.65)*
K	0.317 (3.55)*	0.319 (3.73)*	0.316 (3.69)*	0.300 (3.54)*	0.295 (3.54)*
FER	0.000 (0.30)	0.000 (0.30)	0.000 (0.31)	0.000 (0.33)	0.000 (0.34)
λ	-0.025 (-0.30)	0.135 (0.63)	0.100 (0.91)	-0.018 (-0.23)	-0.012 (-0.18)
R-Square	0.886	0.889	0.889	0.885	0.884

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.3.9

SHAC Estimation - OG1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	0.824 (0.98)	1.450 (1.43)	1.225 (1.45)	0.881 (1.06)	0.990 (1.21)
OG1	-2.631 (-3.64)*	-2.696 (-3.73)*	-2.700 (-3.73)*	-2.669 (-3.69)*	-2.713 (-3.72)*
NFEM	0.719 (6.99)*	0.730 (7.83)*	0.728 (7.69)*	0.733 (7.40)*	0.756 (7.96)*
K	0.350 (3.62)*	0.345 (3.80)*	0.346 (3.80)*	0.343 (3.67)*	0.332 (3.70)*
FER	0.000 (0.58)	0.000 (0.62)	0.000 (0.62)	0.000 (0.60)	0.000 (0.62)
λ	0.046 (0.59)	0.193 (0.92)	0.144 (1.35)	0.032 (0.43)	0.021 (0.32)
R-Square	0.890	0.889	0.890	0.889	0.888

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.3.10

SHAC Estimation – OG2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	0.012 (0.02)	0.457 (0.51)	0.385 (0.54)	0.103 (0.15)	0.246 (0.36)
OG2	-1.046 (-3.54)*	-1.054 (-3.57)*	-1.047 (-3.51)*	-1.073 (-3.63)*	-1.106 (-3.76)*
NFEM	0.774 (7.20)*	0.774 (7.88)*	0.772 (7.69)*	0.791 (7.55)*	0.812 (8.09)*
K	0.334 (3.59)*	0.331 (3.78)*	0.334 (3.80)*	0.327 (3.65)*	0.318 (3.67)*
FER	0.000 (0.59)	0.000 (0.60)	0.000 (0.59)	0.000 (0.59)	0.000 (0.60)
λ	-0.032 (-0.39)	0.101 (0.46)	0.095 (0.89)	-0.034 (-0.44)	-0.026 (-0.37)
R-Square	0.886	0.888	0.889	0.885	0.885

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

In summary, the spatial models utilizing different combinations of diversity indices and weight matrices produce very similar results, in term of coefficient magnitudes, between the SHAC and GSTSLS estimation methods. The SHAC method tends to produce lower t-statistics for diversity, nonfarm employment growth, and capital growth as compared to GSTSLS method, but the estimated coefficients for the three variables are statistically significant at the .01 level.

Under both estimation methods, the results recognize the importance of diversity in state economies. That is, the results provide strong evidence that states with higher levels of economic diversity experience higher growth rates of nonfarm earnings compared to states with lower levels of economic diversity, holding other influences constant. Results from both estimation methods also reveal the great importance of nonfarm employment growth and capital growth. Generally, an increase of one percent in the growth rate of nonfarm employment generates roughly an increase of .75 percent in the growth rate of nonfarm earnings. Also, an increase of one percent in the growth rate of capital generates an increase of roughly .33 percent in the growth rate of nonfarm earnings. The consistency of results across the two estimators suggests that diversity, nonfarm employment growth, and capital growth are positively influencing nonfarm earnings growth, regardless of which diversity measures and spatial weight matrices are being used in the models.

Under both estimation methods, the results provide strong evidence that farm earnings growth does not provide a detectable contribution to nonfarm earnings growth. The magnitudes of the coefficients are small and statistically insignificant under both estimation methods. As for the spatial lag parameters, both methods of estimation provide no evidence of spillover effects. That is, nonfarm earnings growth from nearby states do not seem to effect the growth of nonfarm earnings in a home state.

5.4. Estimation Results For Region Specific Spatial Models

This section presents the results from estimating the region specific spatial model of nonfarm earnings growth using the annual data from 1992 to 2009. The extension of the basic spatial model in (38) to the region specific spatial model with Q regions is

(39) $GRO_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 NFEM_{it} + \beta_4 K_{it} + \beta_5 FER_{it} +$

$$\sum_{k=1}^{Q} \lambda_k \sum_{j=1}^{n} w_{k,ij} GRO_{jt} + \sum_{j=1}^{7} \delta_j DR_{j,it} + \sum_{j=1}^{T-1} \gamma_j \, DS_{j,it} + u_{it}$$

where $w_{k,ij}$ is the ij element of the spatial weight matrix w_k for region k, as defined in Section 4.2.3. In matrix format, the model is

(40) $\underline{y} = X\beta + \sum_{k=1}^{8} \lambda_k W_k \underline{y} + Y\underline{\gamma} + \underline{u}$

where W_k is defined in Section 4.2.3. The error process is assumed to be $\underline{u} = R\underline{\varepsilon}$ and the SHAC estimator is used. The H matrix of instrumental variables is $H = (\overline{X}, W_1\dot{\tilde{X}}, \dots, W_8\dot{\tilde{X}})$ where $\overline{X} = (X, G), X = \left(\underline{1}, \underline{V}, \underline{FER}, \underline{DR}_1 \dots \underline{DR}_7 \underline{DS}_1 \dots \underline{DS}_{17}\right), G = \left(\underline{NFEM}_{t-1}\ \underline{z}_1\ \underline{z}_4\right), \dot{\tilde{X}} = \left(\tilde{X}, G\right)$, and $\tilde{X} = (\underline{V}\ \underline{FER})$. For the case of $V = OG1$, then $\tilde{X} = (\underline{FER})$ and $G = (\underline{NFEM}_{t-1}\ \underline{z}_1\ \underline{z}_4\ \underline{OG1}_{t-1})$ since $OG1$ is endogenous. In essence, the fixed effects and time effects dummy variables are excluded when creating instrumental variables involving the eight weights matrices.

In discussing of the results for the region specific spatial models, we need to keep in mind that for the spatial models discussed in Section 5.3, the spatial lag parameter was assumed to be constant for all states across the United States. Under the fixed spatial lag parameter, the results provided no evidence of spatial effects on nonfarm earnings growth between states.

Table 5.4.1 presents the estimation results for the model in (39) specified with the Entropy index together with the five weights matrices used in our earlier work. In general, the coefficients for the Entropy index, the nonfarm employment growth rate ($NFEM$), and the capital growth rate (K) are statistically significant at 0.01 level. Notice that the magnitude of the coefficients for Entropy, $NFEM$, and K are similar across the five model specifications and are quite similar to the corresponding coefficients in the spatial models discussed in Section 5.3. Specifically, the coefficients for the Entropy index range from 2.71 when using the contiguity weights matrix to 2.98 when using the combination of contiguity and inverse-distance weights matrices. Estimated coefficients for $NFEM$ range from 0.705 when using the contiguity weights matrix to 0.743 for the model that uses the combination of contiguity and inverse-distance weights matrices. Estimated coefficients for K range from 0.312 for the model that uses the contiguity weights matrix to 0.338 when using the inverse-distance-squared weights matrix. The characteristic of the coefficients for $NFEM$ and K summing close to one is once again present in these region specific spatial models. There is no evidence that farm earnings growth plays a significant role in determining nonfarm earnings growth.

Interestingly, when we allow the spatial effects for each of the eight regions to vary, the results provide strong evidence of spatial effects among states. The results vary a little bit depending on the weights matrix being used. For the model specified with the contiguity weights matrix, the results suggest strong spatial effects in all regions except the Southeast. The spatial lag parameters are positive and statistically significant at the 0.01 level for the New England, Great Lakes, Plains, Southwest, Rocky Mountain, and Far West regions and at the 0.10 level for the Mideast region. States in the Rocky Mountain enjoy the highest level of spillover effects from neighboring border-sharing states. At the other extreme, states in the Far West region show the lowest significant level of spillover effects from border-sharing states. The estimated coefficient for the spatial lag parameter in the Far West region is .186.

Table 5.4.1.

Region Specific Spatial Models - Entropy Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-7.768	-7.423	-7.951	-8.283	-8.733
	(-3.17)*	(-3.02)*	(-3.30)*	(-3.40)*	(-3.59)*
Entropy	2.710	2.856	2.759	2.844	2.977
	(2.96)*	(3.13)*	(3.02)*	(3.11)*	(3.22)*
NFEM	0.705	0.721	0.705	0.725	0.743
	(7.89)*	(8.35)*	(8.24)*	(8.23)*	(8.37)*
K	0.312	0.330	0.338	0.314	0.322
	(3.84)*	(3.99)*	(4.06)*	(3.84)*	(3.85)*
FER	0.000	0.001	0.000	0.000	0.000
	(0.52)	(0.71)	(0.52)	(0.46)	(0.41)
λ_1 (New England)	0.217	0.538	0.289	0.155	0.109
	(2.84)*	(3.04)*	(3.17)*	(2.26)**	(1.73)***
λ_2 (Mideast)	0.210	0.468	0.229	0.136	0.082
	(1.74)***	(2.31)**	(1.92)***	(1.27)	(0.81)
λ_3 (Great Lakes)	0.233	0.529	0.305	0.181	0.142
	(2.66)*	(2.77)*	(2.80)*	(2.11)**	(1.68)***
λ_4 (Plains)	0.252	0.511	0.283	0.187	0.137
	(2.76)*	(2.82)*	(2.75)*	(2.09)**	(1.58)
λ_5 (Southeast)	0.118	0.366	0.130	0.052	0.005
	(1.43)	(1.90)***	(1.22)	(0.65)	(0.07)
λ_6 (Southwest)	0.281	0.480	0.333	0.249	0.218
	(2.90)*	(2.42)**	(2.58)*	(2.59)*	(2.34)**
λ_7 (Rocky Mountain)	0.302	0.500	0.294	0.243	0.196
	(3.11)*	(2.81)*	(2.63)*	(2.52)**	(2.05)**
λ_8 (Far West)	0.186	0.552	0.276	0.147	0.108
	(3.00)*	(3.17)*	(3.15)*	(2.40)**	(1.79)***
R-Square	0.889	0.888	0.891	0.889	0.889
Wald Test for $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$	7.907	9.482	9.419	8.259	8.254

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

For the models specified with the inverse-distance weights matrix, the results suggest strong spatial effects in all regions. All spatial lag parameters are statistically significant at the 0.05 level or lower with the exception being the Southeast region where the lag parameter is significant at the 0.10 level. The magnitudes of spatial effects between states are similar across regions. The Wald test statistic cannot reject the null hypothesis that the spatial lag parameter is the same in the eight regions. However, later we pursue pairwise tests of equality between regions and discover some significant differences.

With the exception of states in the Southeast region, the coefficients for the other seven regions are above 0.400, ranging from 0.468 for the Mideast to 0.538 for New England. The coefficient for the Southeast region is 0.366. These results suggests that states located in New England region receive the highest spillover effects on nonfarm earnings growth from the growth of the other 47 states, with more emphasis on states that are closer to home.

For the model using the inverse-distance-squared weights matrix, the results provide strong evidence of spatial correlation in nonfarm earnings growth for states in all regions except for the Southeast. The magnitudes of the spatial lag parameters range from 0.229 in the Mideast to 0.333 in the Southwest. The spatial lag parameters are statistically significant at the 0.01 level for the New England, Great Lakes, Plains, Southwest, Rocky Mountain, and the Far West regions. The spatial lag parameter for the Mideast region is statistically significant at the 0.10 level.

When using the combination of the contiguity and inverse-distance weights matrices, the results provide evidence of spatial correlation in the New England, Great Lakes, Plains, Southwest, Rocky Mountain, and Far West regions. The spatial lag parameters range from 0.146 in the Far West to 0.249 in the Southwest and all are statistically significant at the 0.05 level with exception of the Southwest region at the 0.01 level. By using the contiguity and inverse-distance weights matrices, the growth in state i located in a region is effected by the growth of border-sharing states and weighted by the distances between their population centers.

Lastly, for the model using the combination of contiguity and inverse-distance-squared weights matrices, the results show that only states in the Southwest and Rocky Mountain experience significant spillover effects from border-touching states at the 0.05 level. The spatial lag parameters for the Southwest and the Rocky Mountain are 0.218 and 0.195, respectively. The results also suggest weakly significant spatial effects for states in New England, Great Lakes, and the Far West, all being statistically significant at the 0.10 level.

Overall, the model using the inverse-distance weights matrix provides the strongest evidence of spatial effects in the regions. The results suggest that states' economic growth rates are spatially correlated with the other 47 states, with stronger effects for states that are closer to home, as measured by the distance between the two population centroids. Across the five weights matrices, states in the Southwest region seem to enjoy the highest level of spillover effects as compared to states in other regions. Garrett et al. (2007) used nine Census regions that roughly line up with the eight BEA regions used here. They found that the New England region had the strongest spatial lag parameter when they used the inverse-distance weights matrix.

Table 5.4.2 presents the estimation results for the model (39) specified with the $NA1$ index together with the five weights matrices. The results for these models generally agree with the results for the spatial models discussed in Section 5.3. That is, the estimated coefficients for the $NA1$ index, nonfarm employment growth rate, and capital growth rate are statistically significant at the 0.01 level. Their magnitudes are similar across the five model specifications and are similar to those discussed in Section 5.3. The sum of the coefficients for the nonfarm employment growth rate and capital growth rate are close to one and there is no evidence that farm earnings growth plays a significant role in influencing nonfarm earnings growth.

For the model specified with the $NA1$ index and contiguity weights matrix, the results provide evidence of spatial correlation for states in seven of the regions with the exception being the Southeast. This pattern is similar to the results in Table 5.4.1 when using the Entropy index. The spatial lag parameter is strongest for states in the Southwest and weakest for states in the Far West with magnitudes of 0.294 and 0.181, respectively.

Table 5.4.2

Region Specific Spatial Models - NA1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	0.076	1.329	0.417	0.064	0.061
	(0.12)	(1.44)	(0.62)	(0.11)	(0.10)
NA1	-2.063	-2.144	-2.113	-2.169	-2.255
	(-3.81)*	(-3.98)*	(-3.88)*	(-3.95)*	(-4.02)*
NFEM	0.718	0.766	0.747	0.744	0.766
	(8.16)*	(9.42)*	(9.19)*	(8.70)*	(9.02)*
K	0.316	0.315	0.317	0.311	0.311
	(3.98)*	(4.10)*	(4.15)*	(4.00)*	(4.00)*
FER	0.000	0.000	0.000	0.000	0.000
	(0.23)	(0.35)	(0.27)	(0.18)	(0.14)
λ_1 (New England)	0.183	0.540	0.276	0.129	0.089
	(2.42)**	(3.04)*	(3.05)*	(1.89)***	(1.41)
λ_2 (Mideast)	0.275	0.519	0.293	0.205	0.154
	(2.33)**	(2.69)*	(2.50)**	(1.93)***	(1.53)
λ_3 (Great Lakes)	0.191	0.487	0.264	0.150	0.117
	(2.06)**	(2.64)*	(2.30)**	(1.65)***	(1.31)
λ_4 (Plains)	0.211	0.488	0.282	0.166	0.127
	(2.31)**	(2.78)*	(2.70)*	(1.86)***	(1.46)
λ_5 (Southeast)	0.083	0.410	0.149	0.023	-0.025
	(1.06)	(2.26)**	(1.40)	(0.29)	(-0.33)
λ_6 (Southwest)	0.294	0.619	0.399	0.269	0.245
	(3.02)*	(3.18)*	(2.99)*	(2.76)*	(2.57)**
λ_7 (Rocky Mountain)	0.275	0.554	0.326	0.228	0.200
	(2.76)*	(3.09)*	(2.86)*	(2.32)**	(2.08)**
λ_8 (Far West)	0.181	0.534	0.271	0.137	0.101
	(2.85)*	(3.22)*	(3.02)*	(2.17)**	(1.58)
R-Square	0.891	0.888	0.890	0.890	0.889
Wald Test for $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$	11.916	10.526	9.785	10.982	12.531***

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

For the model specified with the $NA1$ index and inverse-distance weights matrix, the results provide strong evidence of spatial correlation for states in all eight regions. All spatial lag parameters for the eight regions are statistically significant at the 0.01 level with exception of the Southwest which is statistically significant at the 0.05 level. The magnitudes of the parameters range from 0.410 for the Southeast to 0.619 for the Southwest.

For the model using the $NA1$ index with inverse-distance-squared weights matrix, the results provide evidence of spatial correlation for states in all regions except for the Southeast. The magnitudes of the spatial lag parameter range from 0.264 the Great Lakes to .399 for the Southwest.

For the model using the $NA1$ index with the combination of the contiguity and inverse-distance weights matrices, the results show evidence of spatial correlation for states in the Southwest, Rocky Mountain, and Far West regions. The estimated coefficients range from 0.137 for the Far West to 0.269 for the Southwest and all are statistically significant at the 0.05 level or lower. The results suggest weak spatial correlation for states in the New England, Mideast, Great Lakes, and Plains regions with spatial lag parameters that are statistically significant at the 0.10 level. Recall that by using the combination of the contiguity and inverse-distance weights matrices, the growth in state i located in region j is affected by the growth of border-sharing states in region j and weighted by the distances between their population centroids.

For the model using the $NA1$ index with the combination of the contiguity and inverse-distance-square weights matrices, the results show that only states in the Southwest and Rocky Mountain regions experience spillover effects from border-touching states. The spatial lag parameters for the Southwest and Rocky Mountain regions are 0.245 and 0.200, respectively, and are statistically significant at the 0.05 level.

Overall, the magnitudes of spatial correlation within each of the eight regions are similar for each of the models in Table 5.4.2. As was the case in Table 5.4.1, the model using the inverse-distance weights matrix provided the strongest evidence of spatial effects among states. Again, the results seem to suggest that states' economic growth rates are spatially correlated with the other 47 states, and the degree of spatial effects is a function of distance between the two states population centroids. Similar to the results in Table 5.4.1, based on the Entropy index, the results based on the $NA1$ index seem to suggest that states in the Southwest region experience the highest level of spillover effects compared to states in other regions.

Tables 5.4.3, 5.4.4, and 5.4.5 present the estimation results for the region specific spatial models using the $NA2$, $OG1$, and $OG2$ indices, respectively.

Table 5.4.3

Region Specific Spatial Models - NA2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	-0.382	0.797	0.122	-0.407	-0.343
	(-0.63)	(0.86)	(0.18)	(-0.69)	(-0.59)
NA2	-1.042	-1.070	-1.076	-1.068	-1.102
	(-3.13)*	(-3.24)*	(-3.23)*	(-3.18)*	(-3.25)*
NFEM	0.702	0.724	0.731	0.726	0.755
	(8.00)*	(8.63)*	(8.87)*	(8.41)*	(8.79)*
K	0.327	0.328	0.316	0.323	0.314
	(3.95)*	(4.06)*	(4.02)*	(3.96)*	(3.90)*
FER	0.000	0.000	0.000	0.000	0.000
	(0.02)	(0.22)	(0.12)	(0.01)	(0.02)
λ_1 (New England)	0.170	0.551	0.289	0.114	0.079
	(2.25)**	(3.16)*	(3.21)*	(1.69)***	(1.26)
λ_2 (Mideast)	0.229	0.567	0.313	0.167	0.131
	(1.99)**	(2.86)*	(2.68)*	(1.60)	(1.31)
λ_3 (Great Lakes)	0.159	0.519	0.291	0.116	0.092
	(1.71)***	(2.67)*	(2.53)**	(1.26)	(1.01)
λ_4 (Plains)	0.140	0.500	0.266	0.089	0.063
	(1.63)	(2.72)*	(2.50)**	(1.01)	(0.71)
λ_5 (Southeast)	0.069	0.416	0.197	0.016	-0.013
	(0.867)	(2.21)**	(1.83)***	(0.20)	(-0.18)
λ_6 (Southwest)	0.253	0.562	0.370	0.225	0.210
	(2.65)*	(2.81)*	(2.83)*	(2.38)**	(2.26)**
λ_7 (Rocky Mountain)	0.251	0.537	0.318	0.220	0.184
	(2.75)*	(3.00)*	(2.90)*	(2.38)**	(2.00)**
λ_8 (Far West)	0.165	0.586	0.294	0.123	0.093
	(2.66)*	(3.38)*	(3.28)*	(1.96)***	(1.46)
R-Square	0.893	0.890	0.891	0.892	0.890
Wald Test for $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$	8.254	7.667	4.930	8.712	8.402

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = distance^{-1} and border contiguity, and W_5 = distance^{-2} and border contiguity. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.4

Region Specific Spatial Models - OG1 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	1.351	2.099	1.101	1.178	1.139
	(1.67)***	(1.88)***	(1.21)	(1.45)	(1.37)
OG1	-2.501	-2.543	-2.472	-2.569	-2.642
	(-3.39)*	(-3.43)*	(-3.35)*	(-3.47)*	(-3.54)*
NFEM	0.691	0.706	0.691	0.712	0.732
	(8.02)*	(8.57)*	(8.27)*	(8.31)*	(8.43)*
K	0.316	0.334	0.344	0.320	0.325
	(3.90)*	(4.09)*	(4.09)*	(3.87)*	(3.84)*
FER	0.000	0.000	0.000	0.000	0.000
	(0.52)	(0.64)	(0.50)	(0.48)	(0.43)
λ_1 (New England)	0.220	0.542	0.267	0.150	0.103
	(2.95)*	(3.07)*	(2.96)*	(2.26)**	(1.68)***
λ_2 (Mideast)	0.287	0.525	0.244	0.179	0.107
	(2.47)**	(2.60)*	(2.05)**	(1.72)***	(1.69)***
λ_3 (Great Lakes)	0.252	0.556	0.310	0.196	0.107
	(3.08)**	(2.90)*	(2.97)*	(2.46)**	(1.07)
λ_4 (Plains)	0.283	0.494	0.264	0.214	0.158
	(3.04)*	(2.74)*	(2.63)*	(2.40)**	(2.02)**
λ_5 (Southeast)	0.125	0.363	0.094	0.045	0.168
	(1.53)	(1.91)***	(0.88)	(0.57)	(1.97)**
λ_6 (Southwest)	0.285	0.510	0.307	0.233	0.000
	(3.03)*	(2.52)**	(2.39)**	(2.51)**	(0.010)
λ_7 (Rocky Mountain)	0.301	0.483	0.274	0.241	0.196
	(3.19)*	(2.69)*	(2.47)**	(2.55)**	(2.10)**
λ_8 (Far West)	0.184	0.536	0.265	0.147	0.113
	(3.00)*	(3.08)*	(3.02)*	(2.46)**	(1.91)***
R-Square	0.890	0.889	0.891	0.890	0.890
Wald Test for $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$	10.466	11.033	12.471***	10.767	10.615

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.5

Region Specific Spatial Models - OG2 Index					
Variables	W_1	W_2	W_3	W_4	W_5
Constant	0.447	1.231	0.451	0.356	0.318
	(0.66)	(1.24)	(0.59)	(0.53)	(0.46)
OG2	-0.959	-1.003	-0.958	-1.000	-1.038
	(-3.42)*	(-3.56)*	(-3.38)*	(-3.54)*	(-3.63)*
NFEM	0.718	0.744	0.724	0.739	0.757
	(7.89)*	(8.52)*	(8.35)*	(8.21)*	(8.35)*
K	0.316	0.327	0.338	0.322	0.330
	(3.90)*	(4.04)*	(4.12)*	(3.93)*	(3.95)*
FER	0.000	0.000	0.000	0.000	0.000
	(0.48)	(0.63)	(0.48)	(0.42)	(0.37)
λ_1 (New England)	0.211	0.516	0.281	0.150	0.103
	(2.73)*	(2.92)*	(3.04)*	(2.16)**	(1.62)
λ_2 (Mideast)	0.164	0.445	0.215	0.106	0.062
	(1.30)	(2.21)**	(1.81)***	(0.96)	(0.61)
λ_3 (Great Lakes)	0.217	0.484	0.283	0.167	0.128
	(2.52)**	(2.53)**	(2.60)*	(1.98)**	(1.54)
λ_4 (Plains)	0.233	0.474	0.263	0.173	0.124
	(2.55)**	(2.62)*	(2.55)**	(1.93)***	(1.43)
λ_5 (Southeast)	0.102	0.316	0.107	0.038	-0.009
	(1.23)	(1.65)***	(1.01)	(0.48)	(-0.12)
λ_6 (Southwest)	0.270	0.488	0.329	0.235	0.204
	(2.78)*	(2.43)**	(2.53)**	(2.46)**	(2.21)**
λ_7 (Rocky Mountain)	0.293	0.482	0.283	0.235	0.190
	(2.97)*	(2.70)*	(2.52)**	(2.41)**	(1.97)**
λ_8 (Far West)	0.179	0.529	0.261	0.136	0.096
	(2.88)*	(3.04)*	(2.95)*	(2.22)**	(1.57)
R-Square	0.890	0.889	0.891	0.890	0.890
Wald Test for $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$	7.890	12.154***	10.337	8.334	8.470

Note: Weights matrices: W_1 = border contiguity, W_2 = distance^{-1}, W_3 = distance^{-2}, W_4 = border contiguity and distance^{-1}, and W_5 = border contiguity and distance^{-2}. t-statistics are in parentheses. ***/**/* indicate significance at 10/5/1 percent levels, respectively.

The magnitude and t-statistics of the coefficients for diversity indices, $NFEM$, K, and FER presented in Tables 5.4.3, 5.4.4, and 5.4.5 are similar to the spatial model estimates presented in Tables 5.3.3, 5.3.4, and 5.3.5. The estimated coefficients for the diversity indices, $NFEM$, and K are consistent across different model specifications with different weight matrices and all are statistically significant at the 0.01 level. Farm earnings growth continues to play an insignificant role in determining the growth rate of nonfarm earnings.

For the spatial lag parameters, the estimates and patterns of significance are different across each of the five models that use different weights matrices. Estimates of the models containing $NA2$, $OG1$, and $OG2$ diversity indices provide strong evidence of spatial correlation among the states for all eight regions when using the inverse-distance and inverse-distance-squared weights matrices. In terms of magnitudes of coefficients, models that use the inverse-distance matrix provide the highest estimates of the spatial lag, as seen in Tables 5.4.3 through 5.4.5. Thus, estimates of the spatial lag parameters are highest when allowing states to be influenced by 47 other states, with more weight being given to state that are closer to the home state. Of the eight regions, states in the Southeast region seen to experience the least amount of spillover effects from neighboring states.

In summary, after allowing spatial effects to vary by region, the model estimates continue to provide strong evidence that industrial diversity positively influences the growth rate of nonfarm earnings. Furthermore, the significance of nonfarm employment growth rates and capital growth rates are consistent throughout the study.

The Wald tests in Tables 5.4.1 through 5.4.5 show that the spatial lag parameters do not significantly vary across regions. The null hypothesis that all spatial lag parameters are equal for the eight regions cannot be rejected at the 0.05 level. However, the Wald test is very broad and differences between pairs of regions might be obscured. Equality tests for each pair of regions were performed and the results are shown in Tables 5.4.6 through 5.4.10. The tests show that spatial effects are not much different across regions, with the exception of the Southeast. The results suggest that spatial effects for states in the Southeast region differ from those in several other regions. Recall from Table 5.4.1 through 5.4.5 that the spatial lag parameter in the Southeast region was either statistically insignificant or small in magnitude. Thus, the pair-wise test results involving the Southeast may just be an artifact of the lack of spatial effects present in the Southeast.

Table 5.4.6

Tests for Equality of Spatial Lag Parameters - Entropy Index	New England	Mideast	Great Lakes	Plains	Southeast	Southwest	Rocky Mountain
(A) Contiguity weights							
New England							
Mideast	0.00						
Great Lakes	0.03	0.05					
Plains	0.17	0.15	0.05				
Southeast	1.66	0.81	2.51	3.58***			
Southwest	0.41	0.32	0.20	0.08	2.84***		
Rocky Mountain	0.80	0.61	0.50	0.29	4.70**	0.04	
Far West	0.20	0.05	0.42	0.82	1.30	1.15	1.98
(B) Inverse-distance weights							
New England							
Mideast	0.64						
Great Lakes	0.01	0.33					
Plains	0.14	0.21	0.04				
Southeast	6.16**	1.29	3.29***	4.23**			
Southwest	0.36	0.01	0.18	0.10	1.52		
Rocky Mountain	0.19	0.09	0.08	0.02	2.43	0.03	
Far West	0.03	0.64	0.05	0.22	5.19**	0.47	0.28
(C) Inverse-distance-squared weights							
New England							
Mideast	0.44						
Great Lakes	0.03	0.55					
Plains	0.01	0.32	0.07				
Southeast	4.29**	1.13	4.29**	5.00**			
Southwest	0.17	0.73	0.05	0.23	3.88**		
Rocky Mountain	0.00	0.36	0.01	0.02	3.64***	0.11	
Far West	0.03	0.24	0.12	0.01	4.45**	0.30	0.04
(D) Inverse-distance and contiguity weights							
New England							
Mideast	0.04						
Great Lakes	0.09	0.18					
Plains	0.15	0.24	0.01				
Southeast	1.92	0.77	3.01***	3.51***			
Southwest	0.91	0.85	0.39	0.35	3.81***		
Rocky Mountain	0.88	0.88	0.38	0.34	4.71**	0.00	
Far West	0.01	0.01	0.21	0.30	2.38	1.24	1.27
(F) Inverse-distance-squared and contiguity weights							
New England							
Mideast	0.09						
Great Lakes	0.15	0.33					
Plains	0.12	0.28	0.00				
Southeast	2.06	0.64	3.21***	3.10***			
Southwest	1.26	1.25	0.47	0.58	4.33**		
Rocky Mountain	0.85	0.98	0.27	0.35	4.33**	0.04	
Far West	0.00	0.08	0.19	0.15	2.61	1.38	0.96

Note: ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.7

Tests for Equality of Spatial Lag Parameters - NA1 Index							
	New England	Mideast	Great Lakes	Plains	Southeast	Southwest	Rocky Mountain
(A) Contiguity weights							
New England							
Mideast	0.78						
Great Lakes	0.01	0.61					
Plains	0.11	0.40	0.05				
Southeast	1.94	4.07**	2.07	3.63***			
Southwest	1.20	0.03	0.84	0.65	5.03**		
Rocky Mountain	0.91	0.00	0.68	0.48	5.75**	0.03	
Far West	0.00	0.86	0.01	0.15	2.94***	1.49	1.21
(B) Inverse-distance weights							
New England							
Mideast	0.11						
Great Lakes	0.42	0.10					
Plains	0.69	0.14	0.00				
Southeast	5.81**	1.96	0.93	1.67			
Southwest	0.62	0.77	1.27	1.56	4.60**		
Rocky Mountain	0.03	0.13	0.44	0.64	3.51***	0.33	
Far West	0.01	0.02	0.24	0.30	2.80***	0.56	0.04
(C) Inverse-distance-squared weights							
New England							
Mideast	0.05						
Great Lakes	0.02	0.08					
Plains	0.01	0.02	0.04				
Southeast	3.21***	2.62	1.78	3.86**			
Southwest	1.22	0.73	1.12	1.10	5.44**		
Rocky Mountain	0.30	0.09	0.31	0.23	4.34**	0.37	
Far West	0.00	0.06	0.01	0.02	3.15***	1.33	0.35
(D) Inverse-distance and contiguity weights							
New England							
Mideast	0.66						
Great Lakes	0.05	0.28					
Plains	0.20	0.16	0.03				
Southeast	2.18	3.85**	2.65	4.13**			
Southwest	1.90	0.28	1.04	0.92	6.06**		
Rocky Mountain	1.08	0.05	0.55	0.43	5.81**	0.13	
Far West	0.01	0.50	0.02	0.14	3.49***	1.86	1.05
(F) Inverse-distance-squared and contiguity weights							
New England							
Mideast	0.53						
Great Lakes	0.10	0.13					
Plains	0.21	0.07	0.01				
Southeast	2.52	3.54***	3.10***	4.27**			
Southwest	2.40	0.56	1.18	1.19	6.96*		
Rocky Mountain	1.37	0.17	0.58	0.56	6.25**	0.15	
Far West	0.03	0.31	0.04	0.11	3.74***	2.16	1.16

Note: ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.8

Tests for Equality of Spatial Lag Parameters - NA2 index

	New England	Mideast	Great Lakes	Plains	Southeast	Southwest	Rocky Mountain
(A) Contiguity weights							
New England							
Mideast	0.36						
Great Lakes	0.01	0.42					
Plains	0.15	0.83	0.05				
Southeast	2.05	2.82***	1.30	1.20			
Southwest	0.72	0.04	0.70	1.29	3.81***		
Rocky Mountain	0.86	0.04	0.84	1.63	5.29**	0.00	
Far West	0.00	0.41	0.00	0.12	2.48	0.93	1.44
(B) Inverse-distance weights							
New England							
Mideast	0.04						
Great Lakes	0.11	0.20					
Plains	0.52	0.61	0.04				
Southeast	3.76***	3.23***	1.14	1.56			
Southwest	0.01	0.00	0.12	0.37	2.27		
Rocky Mountain	0.03	0.09	0.03	0.19	2.31	0.05	
Far West	0.16	0.04	0.38	0.92	4.55**	0.05	0.26
(C) Inverse-distance-squared weights							
New England							
Mideast	0.09						
Great Lakes	0.00	0.05					
Plains	0.10	0.29	0.08				
Southeast	1.68	1.67	1.14	1.07			
Southwest	0.58	0.23	0.42	0.98	2.89***		
Rocky Mountain	0.12	0.00	0.07	0.38	2.30	0.21	
Far West	0.01	0.04	0.00	0.14	1.95	0.51	0.08
(D) Inverse-distance and contiguity weights							
New England							
Mideast	0.34						
Great Lakes	0.00	0.23					
Plains	0.10	0.64	0.09				
Southeast	1.99	2.67	1.53	1.11			
Southwest	1.30	0.24	0.90	1.71	4.50**		
Rocky Mountain	1.43	0.24	0.97	1.97	5.73**	0.00	
Far West	0.02	0.22	0.01	0.19	2.77***	1.20	1.30
(F) Inverse-distance-squared and contiguity weights							
New England							
Mideast	0.34						
Great Lakes	0.02	0.14					
Plains	0.04	0.46	0.09				
Southeast	1.78	2.35	1.62	1.07			
Southwest	1.80	0.43	1.03	1.87	4.92**		
Rocky Mountain	1.36	0.23	0.75	1.56	5.08**	0.05	
Far West	0.04	0.15	0.00	0.14	2.56	1.46	1.06

Note: ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.9

Tests for Equality of Spatial Lag Parameters - OG1 index							
	New England	Mideast	Great Lakes	Plains	Southeast	Southwest	Rocky Mountain
(A) Contiguity weights							
New England	0.42						
Mideast	0.17	0.12					
Great Lakes	0.53	0.00	0.15				
Plains	1.52	2.84***	3.74***	5.04**			
Southeast	0.46	0.00	0.11	0.00	3.17***		
Southwest	0.78	0.02	0.30	0.04	4.90**	0.03	
Rocky Mountain	0.27	1.15	1.11	1.87	1.11	1.49	2.27
Far West							
(B) Inverse-distance weights							
New England							
Mideast	0.05						
Great Lakes	0.03	0.10					
Plains	0.49	0.14	0.52				
Southeast	6.91*	3.88**	4.96**	4.36**			
Southwest	0.09	0.02	0.17	0.03	2.30		
Rocky Mountain	0.49	0.18	0.59	0.02	2.24	0.06	
Far West	0.01	0.01	0.05	0.27	5.47**	0.05	0.32
(C) Inverse-distance-squared weights							
New England							
Mideast	0.08						
Great Lakes	0.29	0.48					
Plains	0.00	0.05	0.36				
Southeast	5.84**	2.77***	7.60*	7.19*			
Southwest	0.15	0.29	0.00	0.19	4.48**		
Rocky Mountain	0.01	0.09	0.16	0.02	4.70**	0.09	
Far West	0.00	0.05	0.34	0.00	6.50**	0.18	0.01
(D) Inverse-distance and contiguity weights							
New England							
Mideast	0.10						
Great Lakes	0.35	0.03					
Plains	0.62	0.12	0.05				
Southeast	2.07	2.12	4.82**	5.59**			
Southwest	0.76	0.22	0.14	0.03	3.97**		
Rocky Mountain	1.00	0.33	0.24	0.08	5.40**	0.01	
Far West	0.00	0.13	0.53	0.87	3.06***	1.00	1.34
(F) Inverse-distance-squared and contiguity weights							
New England							
Mideast	0.00						
Great Lakes	0.51	0.27					
Plains	0.67	0.36	0.01				
Southeast	2.11	1.29	4.90**	5.31**			
Southwest	1.16	0.69	0.19	0.13	4.41**		
Rocky Mountain	1.04	0.64	0.15	0.09	4.96**	0.01	
Far West	0.02	0.00	0.42	0.57	3.46***	1.06	0.97

Note: ***/**/* indicate significance at 10/5/1 percent levels, respectively.

Table 5.4.10

Tests for Equality of Spatial Lag Parameters - OG2 index							
	New England	Mideast	Great Lakes	Plains	Southeast	Southwest	Rocky Mountain
(A) Contiguity weights							
New England							
Mideast	0.19						
Great Lakes	0.00	0.21					
Plains	0.07	0.35	0.04				
Southeast	2.02	0.33	2.55	3.47***			
Southwest	0.35	0.63	0.25	0.13	2.95***		
Rocky Mountain	0.71	1.04	0.59	0.39	4.79**	0.04	
Far West	0.20	0.02	0.28	0.54	1.57	1.03	1.78
(B) Inverse-distance weights							
New England							
Mideast	0.69						
Great Lakes	0.11	0.13					
Plains	0.35	0.10	0.01				
Southeast	8.65*	2.16	3.31***	5.01**			
Southwest	0.08	0.13	0.00	0.02	2.94***		
Rocky Mountain	0.15	0.12	0.00	0.01	3.57***	0.00	
Far West	0.02	0.65	0.19	0.38	6.48**	0.14	0.22
(C) Inverse-distance-squared weights							
New England							
Mideast	0.55						
Great Lakes	0.00	0.45					
Plains	0.05	0.27	0.05				
Southeast	5.24**	1.39	4.29**	5.16**			
Southwest	0.19	0.87	0.15	0.38	4.48**		
Rocky Mountain	0.00	0.41	0.00	0.05	4.06**	0.15	
Far West	0.07	0.24	0.06	0.00	4.81**	0.41	0.06
(D) Inverse-distance and contiguity weights							
New England							
Mideast	0.21						
Great Lakes	0.04	0.33					
Plains	0.08	0.38	0.01				
Southeast	2.28	0.47	3.05***	3.48***			
Southwest	0.75	1.05	0.39	0.36	3.81***		
Rocky Mountain	0.81	1.19	0.45	0.42	4.78**	0.00	
Far West	0.04	0.09	0.17	0.24	2.43	1.15	1.28
(F) Inverse-distance-squared and contiguity weights							
New England							
Mideast	0.21						
Great Lakes	0.09	0.42					
Plains	0.06	0.35	0.00				
Southeast	2.47	0.58	3.31***	3.15***			
Southwest	1.09	1.36	0.48	0.58	4.41**		
Rocky Mountain	0.82	1.21	0.34	0.43	4.53**	0.02	
Far West	0.01	0.13	0.17	0.13	2.63	1.33	1.06

Note: ***/**/* indicate significance at 10/5/1 percent levels, respectively.

5.5. How Strong is the Link between Diversity and Economic Growth?

From Sections 5.1 through 5.4, estimates of the nonfarm earnings growth models, whether non-spatial or spatial, show that states with well diversified economies are better performers economically as compared with states with less diverse economies, holding other factors constant. From a statistical point of view, the estimated coefficients for the diversity indices are generally significant and carry the expected sign, signaling that industrial diversity positively influences nonfarm earnings growth rates. The models confirm the positive influence from diversity, but how strong is the influence upon nonfarm earnings growth rates?

I first consider movements of diversity index values from their means to the most diverse state in 2009. The movements will be joined with coefficients from the basic spatial models and region specific models to examine the degree of importance of diversity's effect upon the growth rate of nonfarm earnings. I also consider the movements of diversity index values from the 25th percentile to the 75th percentile (from least to most diverse).

Table 5.5.1 shows the increases in the growth rate of nonfarm earnings due to movements in diversity indices from the average to most diverse state, using the coefficients from the basic spatial models that were estimated with the SHAC method. From Table 5.5.1, the model estimates suggests that an increase from the mean to the highest value (an increase of 0.081) for the Entropy index contributes anywhere from 0.234 to 0.246 percents to the growth rate of nonfarm earnings, depending on which of the five weights matrix is used. Similarly, the increase in the growth rate of nonfarm earnings due to movements from the mean to most diverse states ranges from 0.249 to 0.263 for the $NA1$ index, 0.100 to 0.103 for the $NA2$ index, 0.219 to 0.228 for the $OG1$ index, and 0.210 to 0.222 for the $OG2$ index.

Table 5.5.1

Increases in the Growth Rate of Nonfarm Earnings Due to Movements from Mean to Most Diverse - SHAC Estimator

	Spatial Weights Matrix				
	W_1	W_2	W_3	W_4	W_5
Entropy	0.234	0.239	0.238	0.239	0.246
NA1	0.249	0.256	0.256	0.255	0.263
NA2	0.103	0.103	0.103	0.106	0.106
OG1	0.229	0.235	0.235	0.232	0.236
OG2	0.210	0.212	0.210	0.216	0.222

Table 5.5.2 shows the increases in the growth rate of nonfarm earnings due to movements in diversity indices from the mean to most diverse state, using estimates from the region specific spatial models. With respect to the Entropy index, a movement from the mean to most diverse state contributes anywhere from a 0.220 to 0.241 percents increase in the growth rate of nonfarm earnings, depending on the spatial weight matrix used. Similarly, the effects from the $NA1$, $NA2$, $OG1$, and $OG2$ indices upon the growth rate of nonfarm earnings do not differ much from their counterparts based upon the basic spatial models.

Table 5.5.2

Increases in the Growth Rate of Nonfarm Earnings Due to Movements from Mean to Most Diverse - Region Specific Models

	Spatial Weights Matrix				
	W_1	W_2	W_3	W_4	W_5
Entropy	0.220	0.231	0.223	0.230	0.241
NA1	0.254	0.264	0.260	0.267	0.277
NA2	0.099	0.102	0.102	0.101	0.105
OG1	0.218	0.221	0.215	0.224	0.230
OG2	0.193	0.202	0.192	0.201	0.208

Table 5.5.3 shows the increases in the growth rate of nonfarm earnings due to movements in diversity indices from the first quartile to third quartile (rank from least diverse to most diverse state) in 2009 based upon the estimates from the basic spatial models. The movement in diversity index values from the first quartile to third quartile for the Entropy, $NA1$, $NA2$, $OG1$, and $OG2$ indices are 0.038, 0.070, 0.059, 0.052, and 0.099, respectively. The movement from first quartile to third quartile of the Entropy index contributes anywhere from 0.109 to 0.115 percents to the growth rate of nonfarm earnings. When using the, $NA1$, $NA2$, $OG1$, and $OG2$ indices, the increase in the growth rate of nonfarm earnings ranges from 0.142 to 0.150, 0.062 to 0.064, 0.131 to 0.136, and 0.103 to 0.109, respectively, across the five weights matrices.

Table 5.5.3

Increases in the Growth Rate of Nonfarm Earnings Due to Movements from First Quartiles to Third Quartile - SHAC Estimator					
	Spatial Weights Matrix				
	W_1	W_2	W_3	W_4	W_5
Entropy	0.109	0.111	0.111	0.111	0.115
NA1	0.142	0.146	0.146	0.146	0.150
NA2	0.062	0.062	0.062	0.064	0.064
OG1	0.131	0.134	0.135	0.133	0.136
OG2	0.103	0.104	0.103	0.106	0.109

Lastly, Table 5.5.4 shows the increases in the growth rate of nonfarm earnings due to movements in diversity indices from the first quartile to third quartile (from least diverse to most diverse state) based upon the estimates from the region specific spatial models. The movements in the Entropy index from the first quartile to the third quartile contribute anywhere from 0.102 to 0.112 percent increases in the growth rate of nonfarm earnings. When using the $NA1, NA2, OG1$, and $OG2$ indices, the increase in the growth rate of nonfarm earnings ranges from 0.145 to 0.158, 0.060 to 0.063, 0.127 to 0.136, and 0.093 to 0.103, respectively, across the five weights matrices. Overall, the effects based on estimates from the region specific spatial models do not differ much from those based upon the basic spatial models.

Table 5.5.4

Increases in the Growth Rate of Nonfarm Earnings Due to Movements from First Quartiles to Third Quartile - Region Specific Models

	Spatial Weights Matrix				
	W_1	W_2	W_3	W_4	W_5
Entropy	0.102	0.108	0.104	0.107	0.112
NA1	0.145	0.150	0.148	0.152	0.158
NA2	0.060	0.061	0.061	0.061	0.063
OG1	0.128	0.129	0.126	0.132	0.136
OG2	0.095	0.099	0.095	0.099	0.103

To put the effects from diversity into perspective, the average growth rate in nonfarm earnings from 1991 to 2009 was 2.62 percent. Thus, the partial effects from the diversity indices are relatively small compared to the annual average growth rate. Even though the estimation results show that diversity is statistically significant in influencing nonfarm earnings growth, the effect is small. It is much more important for states to focus on policies that would increase the growth rates of nonfarm employment and capital, since the results suggest that an increase of one percent in the growth rates of nonfarm employment (capital) will contribute increases of roughly 0.70 (0.39) percents to the growth rate of nonfarm earnings. In the long run, it is ideal if states could set policies toward increasing growth of employment and capital and, at the same time, diversify the state economy since they all positively influence overall economic growth.

6. Conclusion

The main goal of this study was to determine the link between industrial diversity and economic growth within the states of the U.S. The measure of economic growth used in the analysis was confined to the labor portions of the income stream, in accordance with using industry employments as the ingredients in indices of industrial diversity. The relationship was analyzed by considering both spatial and non-spatial model specifications. While the models suggest that the link between diversity and economic growth is significant in the statistical sense, it is minor in terms of scale. The estimation results show that the effect from diversity upon economic growth in a state is complementary, in a positive sense, to the dominant influence provided by overall growth in nonfarm employment and capital. These findings suggest that efforts to diversify state economies will generate long-term benefits but maintenance of steady overall growth in employment and capital should be focused on, at least in the short run. In virtually all of the spatial and non-spatial models, the estimated coefficients for nonfarm employment growth and capital growth show sums that are close to one. These findings go in line with Wagner and Deller (1998), which stated that local economies should focus on policies that focus on growth of employment in the short-run while long-run economic policy should be focused on diversifying the local economy.

Farm earnings play an insignificant role in influencing nonfarm earnings growth. This is because the volatility of the farm sector from year to year makes the linkages to the nonfarm portion of economies very difficult to detect in short time frames.

Another Interesting finding was that the spatial lag parameters were statistically insignificant when the same parameter was assumed to apply for all states in the U.S. The finding suggests that economic growth in neighboring states are not dependent on their neighbors. It may be that the linkage between neighboring states do not take place on an annual basis. That is, states may need time to adjust to take advantage of the spillovers from neighboring states. Local businesses may need extended periods of time to adjust to take advantage of growth occurring in neighboring states. Or, use of the same parameter for all states may not be appropriate.

I then extended the basic spatial model and allowed for the spatial effects to vary across regions in the U.S. In doing so, I found that states do experience spillover effects from neighboring states. There is strong evidence that states in the Southeast region experience little or no spatial effects from common-bordered states. In contrast, the Southwest, Great Lakes, Rocky Mountain, Plains, New England, Mideast and Far West regions experience significant spillovers effects from both the common-bordered states and all states in the U.S. as a whole.

Appendix 1. BEA Regions

New England
Connecticut
Maine
Massachusetts
New Hampshire
Rhode Island
Vermont

Mideast
Delaware
Maryland
New Jersey
New York
Pennsylvania

Great Lakes
Illinois
Indiana
Michigan
Ohio
Wisconsin

Plains
Iowa
Kansas
Minnesota
Missouri
Nebraska
North Dakota
South Dakota

Southeast
Alabama
Arkansas
Florida
Georgia
Kentucky
Louisiana
Mississippi
North Carolina
South Carolina
Tennessee
Virginia
West Virginia

Southwest
Arizona
New Mexico
Oklahoma
Texas

Rocky Mountain
Colorado
Idaho
Montana
Utah
Wyoming

Far West
California
Nevada
Oregon
Washington

Appendix 2. Capital Stocks of States

The method proposed by Garofalo and Yamarik (2002) is followed to build a capital stock series for each of the 48 states. The Bureau of Economic Analysis provides capital stock estimates at year end for the U.S. as a whole. To build a capital stock series for each state, Garofalo and Yamarik use annual earnings data at the industry level for each state to estimate the capital stock for the industry in a given state. Fortunately, we are able to match the industry line codes in the BEA earnings figures to the line codes of the Fixed Asset Table 3.1ES. from BEA. Table A1 below shows the bridge for the line codes between the earnings table and the fixed asset accounts table. Since the capital stock data are year end values and annual earnings data are midyear values, a conversion of the capital stock data is needed in order to time align with earnings data. I average every two years of capital stock data to create the midyear values. That is, the capital stock values at midyear for the U.S. in 2009 is the average of year end data from 2008 and 2009. I converted capital stock data to midyear values before applying the apportionment procedure proposed by Garofalo and Yamarik. The procedure is represented by the following equations:

(1) $$k_{i,j}(t) = \frac{y_{i,j}(t)}{Y_i(t)} K_i(t)$$

$$k_j(t) = \sum_{i=1}^{20} k_{i,j}(t)$$

where i is the industry $(i = 1, \dots, 20)$, t is the year, and j is the state $(j = 1, \dots, 48)$. I use 20 industries in this study as listed in Chapter 3 with the exception that government is not broken down into Federal, State, and Local categories. The components in (1) are:

$k_{i,j}(t)$ - capital stock K for industry i in state j in year t
$y_{i,j}(t)$ - earnings in industry i for state j in year t
$Y_i(t)$ - U.S. earnings in industry i in year t
$K_i(t)$ - U.S. capital stock for industry i in year t
$k_j(t)$ - total capital stock for state j in year t.

For the real estate industry, the BEA data for fixed assets includes owner-occupied property which does not generate measured income. Thus, I subtract owner-occupied property from the BEA estimate in the real estate industry so that the capital series for the real estate industry is in line with the earnings estimates. The figure for owner-occupied property is available in the BEA's Fixed Asset Table 5.1.

Table A1: Bridge Between Earnings and BEA's Fixed Asset Table 3.1ES

Table 3.1ES Line Code	Earnings Line Code	Industry
4	100	Forestry, fishing, and related activities
5	200	Mining
9	300	Utilities
10	400	Construction
12	510	Durable goods
24	530	Nondurable goods
33	600	Wholesale trade
34	700	Retail trade
35	800	Transportation and warehousing
44	900	Information
49	1000	Finance and insurance
55*	1100	Real estate and rental and leasing
58	1200	Professional, scientific, and technical services
62	1300	Management of companies and enterprises
63	1400	Administrative and waste management services
66	1500	Educational services
67	1600	Health care and social assistance
72	1700	Arts, entertainment, and recreation
75	1800	Accommodation and food services
78	1900	Other services, except government
**	2000	Government

* - Line 55 from Fixed Asset Table 3.1ES includes owner-occupied property and it is to be deducted from the real estate industry's reported capital stock. Data for owner-occupied property comes from line 11 of Fixed Asset Table 5.1.

** Line 19 of BEA Fixed Asset Table 1.1

REFERENCES

Attaran, M. 1986. Industrial Diversity and Economic Performance in U.S. Areas. *The Annals of Regional Science* 20: 44-54.

Beck, N. and J. Katz. 1995. What to Do (and Not to Do) with Time-Series-Cross-Section Data in Comparative Politics*.* *American Political Science Review* 89: 634-647.

Breusch, T.S., and A. R. Pagan. 1979. A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica* 48*:* 1287-1294.

Brown, D.J. and J. Pheasant. 1985. A Sharpe Portfolio Approach to Regional Economic Analysis. *Journal of Regional Science* 25: 51-63.

Conley, T.G. 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics* 92: 1-45.

Conroy, M.E. 1974. Alternative Strategies for Regional Industrial Diversification. *Journal of Regional Science* 14: 31-46.

Conroy, M.E. 1975. The Concept and Measurement of Regional Industrial Diversification. *Southern Economic Journal* 41: 492-505.

Crone, T.M. and A. Clayton-Matthews. 2005. Consistent Economic Indexes for the 50 States. *The Review of Economics and Statistics* 87: 593-603.

Dagenais, M.G. and D.L. Dagenais. 1997. Higher Moment Estimators for Linear Regression Models With Errors in the Variables. *Journal of Econometrics* 76: 193-221.

Das, D., H.H. Kelejian, and I.R. Prucha. 2003. Finite Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances. *Regional Science* 82: 1-26.

Dissart, J.C. 2003. Regional Economic Diversity and Regional Economic Stability: Research Results and Agenda. *International Regional Science Review* 26*:* 423-446.

Elhorst, J.P. 2003. Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review* 26: 244-268.

Garcia-Mila, T., and T.J. McGuire. 1993. Industrial Mix as a Factor in the Growth and Variability of States' Economies. *Regional Science and Urban Economics* 23*:* 731-48.

Gasper, A.G. and S. Yamarik. 2002. Regional Convergence: Evidence From a New State-by-State Capital Stock Series. *The Review of Economics and Statistics* 84: 316-323.

Garrett, T.A., G.A. Wagner, and D. C. Wheelock 2007. Regional Disparities in the Spatial Correlation of State Income Growth, 1977-2002. *Annals of Regional Science* 41: 601-618.

Godfrey, L. G. 1978. Testing for Multiplicative Heteroscedasticity. *Journal of Econometrics* 8: 227-236.

Hernandez-Murillo, R. 2003. Strategic Interaction in Tax Policies Among States. *The Federal Reserve Bank of St. Louis Review* 85: 47-56.

Hess, G.D. and I. Shigeru. 1997. Measuring and Comparing Business-Cycle Features. *Journal of Business & Economic Statistics* 15: 432-444.

Im, K.S., M. H. Pesaran, and Y. Shin. 2003. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115: 53-74.

Jackson, R.W. 1984. An Evaluation of Alternative Measures of Regional Industrial Diversification. *Regional Studies* 18: 103-112.

Kelejian, H.H. and I.R. Prucha. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Real Estate Finance and Economics* 17: 99-121.

Kelejian, H.H. and I.R. Prucha. 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40: 509-533.

Kelejian, H.H. and I.R. Prucha. 2007. HAC Estimation in a Spatial Framework. *Journal of Econometrics* 140: 131-154.

Kort, J.R. 1981. Regional Economic Instability and Industrial Diversification in the U.S. *Land Economics* 57: 596-608.

Lambert, D. M. and K. T. McNamara. 2009. Location Determinants of Food Manufacturers in the United States, 2000-2004: Are Nonmetropolitan Counties Competitive?. *Agricultural Economics* 40: 617-630.

Malizia, E.E. and S. Ke. 1993. The Influence of Economic Diversity on Unemployment and Stability. *Journal of Regional Science* 33: 221-235.

Nissan, E. and G. Carter. 2006. The Measurement of Employment Diversity For States and Regions. *Journal of Economics and Finance* 30: 186-197.

Owang, M.T., D.E. Rapach, and H.J. Wall. 2009. States and the Business Cycle. *Journal of Urban Economics* 65: 181-194.

Owyang, M.T., J. Piger, and H.J. Wall. 2005. Business Cycle Phases in U.S. States. *The Review of Economics and Statistics* 87: 604-616.

Siegel, P.B., T.G. Johnson, and J. Alwang. 1995. Regional Economic Diversity and Diversification. *Growth and Change* 26: 261-284.

Stock, J. H., and M. W. Watson. 1989. New Indexes of Coincident and Leading Economic Indicators. *NBER Macroeconomics Annual* 4: 351-394.

Trendle, B. and G. Shorney. 2004. The Effect of Industrial Diversification on Regional Economic Performance. *Working Paper No. 20.* Queensland Government.

Thompson, E. and R. Shaffer. 1996. Targeting Occupations for Displaced Workers: A Test in Wisconsin Regions. *The Review of Regional Studies* 26: 147-163.

Wagner, J.E. and S.C. Deller. 1998. Measuring the Effects of Economic Diversity on Growth and Stability. *Land Economics* 74: 541-546.

Watson, M. W. 1994. Business-Cycle Durations and Postwar Stabilization of the U.S. Economy. *The American Economic Review* 84: 24-46.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48: 817-838.

U.S. Bureau of Economic Analysis. 2009. <http://www.bea.gov/regional/spi/>.

U.S. Bureau of Labor Statistics. 2009. <http://stats.bls.gov/sae/data.htm>.