Regional Research Institute
West Virginia University

Working Paper Series

Economic Diversity and Regional Economic Performance: A Methodological Concern from Model Uncertainty

Jing Chen, Graduate Research Assistant, Regional Research Institute and Department of Geology and Geography, West Virginia University

Working Paper Number 2018-05

Date submitted:March 29, 2018

Key words/Codes: Specialization, Diversity, Economic Structure, Regional Economic Development, Model Uncertainty; R11, R12, C10

Economic Diversity and Regional Economic Performance: A Methodological Concern from Model Uncertainty

Jing Chen*

March 29, 2018

Abstract

Although the role of spatial dependence has been considered in studying the relationship between economic diversity and regional economic performance, the existing literature seldom mentions model uncertainty, which mainly arises from at least two sources. One source of model uncertainty is the choice of an appropriate spatial weight matrix that describes the spatial interactions between two regions, which can be specified in a variety of ways. The second source of model uncertainty is choosing a set of control variables to model the diversity-performance relationship. To overcome these limitations, a Bayesian Model Averaging (BMA) method is used to address model uncertainty when studying the effects of economic diversity on short-term employment growth and long-term economic stability among 359 Metropolitan Statistical Areas (MSA) in the contiguous U.S. The potential spatial spillovers are also considered through spatial regression models. This empirical analysis suggests that ignoring model uncertainty can impact the estimates and our understanding of economic diversity, and it also confirms that economic diversity of neighbors plays an important role in regional economic development.

*Graduate Research Assistant, Regional Research Institute and Department of Geology and Geography, West Virginia University. E–mail: jechen@mail.wvu.edu

1 Introduction

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

Many empirical studies have tested these theoretical assumptions. For example, several findings, including Kort (1981), Malizia and Ke (1993), and Deller and Watson (2016b), support the hypothesis that diversity contributes to economic stability, yet others—such as Jackson (1984), Attaran (1986), and Mizuno et al. (2006)—found that the diversity-stability relationship is insignificant. As such, Malizia and Ke (1993), Wagner and Deller (1998), and others stated that the primary causes of this empirical inconsistency include (1) the inappropriate use of geographical units, (2) poorly defined measures of economic diversity, and (3) overly simplistic modeling methods.

Specifically, numerous geographical units have been used to quantify regional economic structures, such as counties (Deller and Watson, 2016b; Watson and Deller, 2017), states (Attaran, 1986; Wagner and Deller, 1998), Metropolitan Statistical Areas or MSAs in the U.S. (Malizia and Ke, 1993), and Local Government Areas in Queensland, Australia (Trendle, 2006). However, Jackson (1984) and Malizia and Ke (1993) suggested that only functional economic regions (e.g., MSAs) should be used to define these economic structures. Meanwhile, numerous studies have improved existing measures of economic diversity, including input-output based measures in Wagner and Deller (1998) and Siegel et al. (1995) as well as metrics that consider both cluster and industry diversity (e.g. Chen, 2018; Hong and Xiao, 2016).

Another cause of the inconsistency between theoretical assumptions and empirical evidence is the use of simplistic statistical techniques. Although the modeling methods used in the literature of regional economic structure have advanced greatly from bivariate statistics through multiple regression analysis to spatial regression techniques, there is little research on model uncertainty, especially in a spatial context. LeSage and Pace (2009) suggested that model uncertainty can result from at least two sources. Given the variety of methods to specify spatial relationships, one source is the choice of a spatial weight matrix that describes the spatial interactions between two regions. The second source of model uncertainty concerns how to determine the set of control variables to be used to model the diversity-performance relationship. To date, the only study that deals with model uncertainty is Watson and Deller (2017) who used a spatial Bayesian moving average (SBMA) method to determine the set of control variables in studying the effect of industrial diversity on unemployment. However, Watson and Deller (2017) still used counties rather than functional regions as the analytical units, ignored the effect of cluster diversity in studying regional economic structure, and failed to address model uncertainty in the choice of a spatial weight matrix, all of which can impact severely our understanding of the relationship between economic structure and regional economic performance.

This paper hence contributes to the literature on regional economic structure research in several aspects. First, it utilizes a Bayesian Model Average (BMA) method that is different from Watson and Deller's (2017) in order to address model uncertainty in studying the influences of economic diversity on economic stability and employment growth in the context of U.S. regional economies. This method simultaneously addresses model uncertainty (in both sets of control variables and spatial weight matrices) as well as spatial spillovers. Second, based on recent studies (e.g. Chen, 2017, 2018; Hong and Xiao, 2016), this analysis uses MSA as the basic unit to approach regional economic systems and considers both industry and cluster diversity. Finally, this paper also provides a review of previous modeling methods employed in economic structure research, mainly after the 1970s, to compare and contrast their usages and limitations.

The remainder of this paper is organized as follows. Section 2 reviews previous methods used to study the effect of economic diversity on regional economic performance. Following the description of methodology in Section 3, results are presented and discussed. The final section closes with the findings of this paper.

2 Background

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

2.1 Bivariate statistics

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

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

2.2 The inclusion of control variables

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

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

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

2.3 The role of spatial spillovers

Spatial econometric models[1] have been used to assess the potential spatial dependence of regional economic diversity and economic performance by regional scientists, such as Trendle (2006), Hong and Xiao (2016), Deller and Watson (2016b), and Watson and Deller (2017). One of the early works that used spatial regression models to study the relationship between industrial diversity and economic stability is Trendle (2006). He specifically used the spatial autoregressive model (SAR) and the spatial error model (SEM), and the result of Trendle's (2006) analysis confirmed the existence of spatial spillovers in the diversity-stability relationship. Compared to Trendle (2006), Hong and Xiao (2016) used the SAR model to evaluate the performance of industrial diversity on regional economic performance and also proposed a Multiple Specialization Index (MSI) that considers the diversity of economic specializations to emphasize the coexistence of economic diversity and specialization. Furthermore, Deller and Watson (2016b) used the spatial Durbin model (SDM) that captures the spatial dependence in both dependent and independent variables to assess the effect of economic diversity on economic stability among U.S. counties during the Great Recession from 2007 to 2014.

Although the traditional approach is to select a single "best" model for model specification based on various metrics, such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the adjusted R-squared value, and log likelihood value, there is little concern about model uncertainty within economic structure research. For example, the classic trade-off between the inclusion of sufficient independent variables and the inclusion of redundant variables affects the empirical understanding of economic diversity. Fernández et al. (2001a,b) suggested using Bayesian Model Average (BMA) methods to address the model uncertainty issue. Given

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

this, Watson and Deller (2017) applied a Spatial Bayesian Model Averaging (SBMA) method to specify the combination of control variables and used these control variables to model the relationship between industrial diversity and unemployment in the Great Recession with a spatial Durbin model. Watson and Deller (2017) also indicated that the impact of industrial diversity on regional economic performance varied significantly across space and included a heteroscedastic error structure in their estimation.

In a similar vein, this paper uses a BMA method that seeks to control for model uncertainty in both the control variables and the spatial weight matrices. In addition, when assessing the diversity-performance relationship using the BMA results, recent research in the geographical units of regional economic systems and structural measurement of economic diversity is also included. The next section introduces the methodological details.

3 Methodology

The empirical models are expressed as:

$$REI_i = f(DIV_i, CONTROL_i) \tag{1}$$

$$GROWTH_i = g(DIV_i, CONTROL_i) \tag{2}$$

where the dependent variables are the long-term (2000-2014) regional economic instability (REI) index and the short-term (2000-2002) employment growth rate, and the independent variables are economic diversity measures and a set of control variables for the base year 2000. These two empirical models are studied among 359 MSAs in the contiguous U.S. MSA is used as the analytical unit because MSAs form meaningful functional economic systems (Chen, 2017; Jackson, 1984; Malizia and Ke, 1993; Trendle, 2006)).

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

$$REI_i = \{\sum_{i=1}^{N}[(E_{it} - E_{it}^{Tr})/E_{it}^{Tr}]^2/T\}^{1/2} \tag{3}$$

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

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

3.1 Measuring economic diversity

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

$$HHIS_i = \sum_{j=1}^{N} (e_{ij}/e_i)^2 \tag{4}$$

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

Moreover, many regional scientists (e.g. Hong and Xiao, 2016; Jackson, 2015; Malizia and Ke, 1993; Wagner and Deller, 1998) reconsidered the relationship between economic diversity and specialization and proposed the concept of diversified specializations. Recently, empirical studies (e.g. Chen, 2018; Hong and Xiao, 2016) have applied this concept to stress the coexistence of specialization and diversity. Compared with Hong and Xiao (2016), Chen (2018) excluded the impact of local industries in identifying economic clusters; for example, utilities and drug stores that only serve local needs should not be regarded as potential economic clusters. Therefore, this analysis uses Chen's (2018) method to consider the diversity of clusters (HHIC) as follows:

$$HHIC_i = \sum_{j=1}^{M} (e_{ij}/e_i)^2 \tag{5}$$

where e_{ij} is the employment for cluster j in the ith region; E_i is the total employment of traded industries[2] in that region, and M is the total number of clusters in region i. Because these clusters are specialized relative to the national average, the location quotients[3] of these clusters should be greater than one. Technically, HHIC ranges from $1/M$ for a perfectly diversified economy to 1 if all employment is concentrated in one cluster. As for the data sources, the industry diversity variable was calculated using the Upjohn Institute's "WholeData" version of County Business Patterns, derived using Isserman and Westervelt's (2006) method. By comparison, based on the same data as well as the cluster identification method of Delgado et al. (2016), the cluster diversity variable is also calculated.

3.2 Potential control variables and spatial weight matrices

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

Similarly, the economic aspect of control variables includes (1) per capita income relative to the U.S. average, (2) percent of households with income below \$20,000, (3) percent of households with income above \$150,000, (4) Gini coefficient of income inequality, (5) per capita income from transfer payments, (6) per capita income from dividends, interest and rent, and (7)

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

[3]LQ is calculated as the ratio of regional employment share to the national employment share of the same sector.

per capita income from proprietorship. Deller et al. (2017) introduced the expected effects of the last three variables for a given region: per capita income from transfer payments introduces stability; per capita income from dividends, interest and rent measures wealth and introduces instability; and finally, per capita income from proprietorship indicates economic dependency on small businesses. These economic variables were collected from the BEA and the Census Bureau for the year 2000.

To capture the industrial differences, the following factors were included: (1) percentage of employment in government sectors, (2) percentage of employment in goods production sectors (minus farming), and (3) percentage of employment in service production sectors. These factors are important because Mizuno et al. (2006) suggested that the economic diversity index does not consider the components of industrial structure. Specifically, Deller et al. (2017) argued that a high dependency on goods-producing sectors contributes to instability[4], whereas the number for employment in service-related and government sectors is positively associated with economic stability. These industry data were obtained from the Bureau of Labor Statistics (BLS), Census of Employment and Wages for the base year of 2000.

Finally, unlike control variables, there are few theoretical foundations concerning the construction of spatial weight matrices. Instead, many empirical analyses use robustness checks to identify the appropriate spatial weight matrix LeSage and Pace (2009, p. 162). According to Anselin (1988), there are various methods (e.g., continuity, distance band, and k nearest neighbors) to specify spatial relationships. For this analysis, seven k nearest neighbor matrices ($k = 3, 4, 5, 6, 7, 8, 9$) are considered. Region i regards its k closest regions in terms of physical distance as neighbors. If region j belongs to these k regions, then the corresponding element in the spatial weight matrix W_{ij} equals one; otherwise, it equals zero.

[4] This is consistent with the "durable" goods measure of economic diversity. According to Jackson (1984), durable goods are sensitive to economic fluctuations. It is assumed that during an economic downturn, customers are less likely to purchase such durable goods as automobiles, books and furniture.

3.3 Model uncertainty

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

In regional studies that employ BMA methods, two general approaches have been identified in considering potential spatial dependence and addressing model uncertainty. One approach uses spatial BMA or SBMA via LeSage and Parent's (2007) numerical integration techniques to obtain posterior model probabilities for model specifications (e.g. Crespo Cuaresma et al., 2014; LeSage and Fischer, 2008; Watson and Deller, 2017; Winkler et al., 2015). By comparison, a second approach applies spatial filtering techniques (Getis and Griffith, 2002; Tiefelsdorf and Griffith, 2016) that remove the spatial effects and then consider model uncertainty in the framework of standard (non-spatial) BMA methods. For example, Crespo Cuaresma and Feldkircher (2013) used this approach to study factors that influenced the speed of income convergence in Europe from 1995 to 2005. Crespo Cuaresma and Feldkircher further mentioned that the model uncertainty that results from the appropriate spatial weight matrix can also be solved with the second approach. Furthermore, from a technical point, Crespo Cuaresma and Feldkircher suggested that the use of spatial filtering overcomes the computational difficulties associated with LeSage and Parent's (2007) SBMA. For this reason, this second approach is preferred in this analysis.

Consider such a spatial autoregressive (SAR) model: The SDM can be described as:

$$y = \alpha l_N + \rho W y + X_k \beta_k + \varepsilon \tag{6}$$

where y is the dependent variable for N regions; α is the intercept; l_N is an $N \times 1$ vector of ones; ρ is a scalar that denotes the level of spatial autocorrelation; W is the spatial weight matrix indicating the geographical relationship between any two regions; X_k is an $N \times k$ matrix that includes

k explanatory variables; β_k is the estimated coefficient corresponding to X_k ; and ε is the error term. In Equation 6, the number and identity of the variables in X_k are unknown and come from K potential explanatory variables ($K \geq k$). Any model M_k is contained in a larger set of 2^K possible models. Another model uncertainty arises regarding the appropriate spatial weight matrix to specify the underlying spatial interactions. Suppose Z is the number of potential weighting matrices. The total number of potential models is $Z \times 2^K$. As mentioned before, seven spatial weight matrices and 14 potential control variables are considered. The cardinality of model space is therefore 114,688 (7×2^{14}) in this analysis.

Generally, there are two steps in Crespo Cuaresma and Feldkircher's (2013) approach[5]. First, spatial filtering techniques are used to decompose the data into a purely spatial and a non-spatial component. Specifically, spatial dependence in Equation 6 is removed using an eigenvector decomposition method proposed by Getis and Griffith (2002) and Tiefelsdorf and Griffith (2016). The eigenvectors e_i are included as extra explanatory variables in Equation 6 with the following form:

$$y = \alpha l_N + \sum_{i=1}^{E} \gamma_i e_i + X_k \beta_k + \varepsilon \tag{7}$$

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

Second, the results of spatial filtering are then processed with standard BMA methods. As mentioned earlier, model uncertainty exists in both the spatial weight matrix W and explanatory variables X_k. Following the Bayesian moving average methodology, inference on the parameters can be written as:

$$p(\beta_j|Y) = \sum_{j=1}^{2^k} \sum_{z=1}^{Z} p(\beta_j|Y, M_j^z) p(M_j^z|Y) \tag{8}$$

where Y is the whole data. Note that the focus of this research is not the weighted average but rather the posterior probability of each potential variable. Instead of depending on a single model, BMA calculates the weighted

[5] An R Package that carries these two steps is available at https://modelaveraging.wordpress.com/2010/10/

average of the posterior probability densities, where the weights are the posterior probabilities of each model and can be given by:

$$p(M_j^z|Y) = \frac{p(Y|M_j^z)p(M_j^z)}{\sum_{j=1}^{2^k}\sum_{z=1}^{Z} p(Y|M_j^z)p(M_j^z)} \tag{9}$$

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

$$p(\beta_k|\alpha, \rho, \sigma, M_j) \sim N[\beta_k, \sigma^2(gX_k'X_k)] \tag{10}$$

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

Following Madigan et al. (1995), Raftery et al. (1997), Fernández et al. (2001a,b) and Crespo Cuaresma and Feldkircher (2013), a Markov chain Monte Carlo model composite (MC^3) is employed to obtain the posterior distributions of interest over the model space[6]. A random-walk step is used in every replication of the MC^3 procedure. One can propose an alternative model M' to the current model in each step of the chain by adding (*birth step*) or subtracting (*death step*) a regressor from model M. The chain moves to the proposed model using the following acceptance probability:

$$min\left[1, \frac{p(M'|y)}{p(M|y)}\right] \tag{11}$$

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

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

3.4 Modeling methods

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

$$y = \rho W y + X\beta + W X\gamma + \varepsilon \tag{12}$$

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

The implication of this homoscedasticity suggests that the relationship between economic diversity and regional economic performance is stable across space, whereas Deller and Watson (2016a) found that the effect of economic diversity was more significant in certain parts of the U.S. In this regard, the spatial Durbin model with a heteroscedastic error structure ($\varepsilon \sim N(0, \sigma^2 V)$ where $V \neq I$) in a Bayesian framework is used. Based on LeSage and Pace (2009) and Watson and Deller (2017), the following prior distributions are specified.

$$\pi(\beta) \sim N(C, N) \tag{13}$$

$$\pi(r/v_i) \sim \chi^2 IID(r) \tag{14}$$

$$\pi(1/\sigma^2) \sim \Gamma(d, v) \tag{15}$$

$$\pi(\rho) \sim U[0, 1] \tag{16}$$

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

Finally, LeSage and Fischer (2008) suggested that the coefficients of the variables β in Equation 12 cannot be interpreted as marginal effects directly because of spatial dependence. Instead, following LeSage and Pace (2009), direct, indirect and total effects can be estimated.

4 Empirical Results

For simplicity, three sets of results are presented in this section. The first set is the BMA results (Table 1) that provide insights on model uncertainty in the control variables, and the second set (Table 2) is the posterior probabilities for different spatial weight matrices. The third set (Tables 3 and 4) is the estimated effects of economic diversity on economic instability and employment growth when the control variables are suppressed and the spatial weight matrix is specified.

The results in Table 1 demonstrate significant differences in terms of posterior inclusion probability (PIP). Conceptually, the posterior inclusion probability is calculated as the sum of probabilities of models including variable X_k. A PIP of a variable approaching unity suggests the importance of the variable in explaining the dependent variable. Numerous studies (e.g. Crespo Cuaresma and Feldkircher, 2013; Eicher et al., 2011) have labeled covariates with PIP greater than 0.5 as robust and have suggested including them in the final specification. Hence, robust variables in the instability model are (1) percentage of the population with at least a bachelors degree; (2) per capita income from transfer payments; (3) per capita income from dividends, interest and rent; and (4) percentage of employment in goods production sectors (minus farming). By comparison, in the growth model, the corresponding robust variables (PIP > 0.5) are (1) percentage of non-white population; (2) per capita income relative to the U.S. average; (3) percent of households with income above \$150, 000; (4) per capita income from transfer payments; (5) per capita income from dividends, interest and rent; and (6) percentage of employment in goods production sectors (minus farming). These robust control variables are used to estimate the effects of economic diversity on regional economic instability and employment growth based on Equations 1 and 2. Although Watson and Deller (2017) indicated that the set of control variables is of secondary interest in economic structure research, including these variables is expected to avoid redundant variables that decrease precision of the estimation on the one hand, and overcome

Table 1: Posterior inclusion probability of control variables

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

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

potential bias resulting from omitted variables on the other hand.

Raftery et al. (1997), Eicher et al. (2011), and Crespo Cuaresma et al. (2014) further classified robust variables, according to their PIP values, into four categories: weak (50-75%), substantial (75-95%), strong (95-99%) and decisive (above 99%) variables. In that sense, Table 1 also denotes the robustness category of these control variables. It seems that the overall PIP values of robust variables are higher in the growth model than those values in the instability model. Interestingly, the PIP values of employment in goods producing industries in both models are greater than 0.50, indicating that this variable is significantly associated with the dependent variables. By comparison, there are several control variables with low PIP values, such as Gini coefficient of income inequality. However, this does not undermine the validity of the theoretical assumptions but implies that this variable does not help us explain the variations in the dependent variables here.

Table 2: Posterior inclusion probability of control variables

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

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

Table 2 reports the posterior probabilities associated with k nearest neighbor weight matrices with $k = 3, 4, 5\ 8, 9$. Comparing the results of these two models, the instability model seems to lend strong support to the four nearest neighbor spatial weight matrix (KNN4), while the growth model appears to favor the six nearest neighbor spatial weight matrix (KNN6). As noted earlier, few theoretical perspectives have provided guidance on the specification of spatial weight matrix, and the matrix with highest posterior probability is used. That said, KNN4 is used for the instability model and KNN6 for the growth model.

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

Table 3: Effect estimates of growth models

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

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

To supplement Table 3, Models 5-8 in Table 4 consider the impact of both industry and cluster diversity on economic stability and employment growth. For the instability models, it seems that only cluster diversity always contributes to economic stability in Models 5 and 6. Focusing on the t-statistics of its estimated direct, indirect and total effects, cluster diversity in the suppressed model appears to be more significantly associated with economic instability than that in the saturated model. By comparison, the signs of industry diversity in Models 5 and 6 are inconsistent with the theoretical assumption that industrial diversity enhances regional economic stability. This result seems to suggest the use of clusters rather than industries in order to assess economic diversity. Similarly, in the growth models, the diversity of clusters seems to be more associated with employment growth than the industrial diversity variable.

5 Discussion

The empirical results provide several interesting points worthy of note. First, although closely related to the work of Watson and Deller (2017), who studied the relationship between industrial diversity and unemployment using a spatial Bayesian Moving Average method developed by LeSage and Parent (2007), this paper differs from Watson and Deller's (2017) study in the

Table 4: Effect estimates of growth models

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

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

following aspects: (1) a spatial filtering-based BMA method that considers model uncertainty in the choice of control variables and spatial weight matrices is employed; (2) rather than focusing on the relationship between unemployment and economic diversity, this paper concentrates on long-term economic stability and short-term employment growth to leverage the benefits of both economic diversity and specialization; and (3) this paper also uses functional regions as the analytical units and considers both industrial and cluster diversity on regional economic performance to minimize the impacts of factors that lead to the empirical inconsistency with theoretical assumptions of economic structure, such as highly geographical datasets and inappropriate measure of economic diversity.

Second, building on the recent work of economic structure research (e.g. Chen, 2017; Hong and Xiao, 2016; Jackson, 2015), this study includes both industrial and cluster diversity as indicators of economic diversity to study its effect on regional economic performance. Historically, numerous studies

(Malizia and Ke, 1993; Wagner and Deller, 1998; Watson and Deller, 2017) had viewed economic diversity as the diversity of economic activities across industries, therefore treating them the same. By comparison, recent studies (Delgado et al., 2016; Porter, 2003; Spencer et al., 2010) have begun to take a cluster perspective of regional economic structure. Particularly, Porter (2003, p. 562) suggested using clusters rather than industries as the basic units to assess economic diversity because of "the externalities across related industries within clusters." In this study, when comparing with the effects of industry and cluster diversity in Tables 3 and 4, cluster diversity seems to be more supportive to the theoretical foundations of economic specialization and diversity (Conroy, 1975; Glaeser et al., 1992; Porter, 1990, 1998) than industrial diversity. However, cluster and industrial diversity may overlap but are not identical essentially because industries may or may not form economic clusters. To this end, both are important elements of economic diversity and should be considered in studying the relationship between economic diversity and regional economic performance.

Third, Crespo Cuaresma and Feldkircher's (2013) BMA approach has been employed successfully to address simultaneously model uncertainty from the choice of control variables and spatial weight matrix. The comparison between the saturated and suppressed models in Tables 3 and 4 suggests that ignoring this model uncertainty can impact our understanding of the relationship between economic diversity and regional economic performance. As such, together with Watson and Deller (2017), this study suggests that future economic structure research can consider such model uncertainty to better understand economic diversity. In terms of the modeling method, both Crespo Cuaresma and Feldkircher's (2013) and LeSage and Parent's (2007) approaches provide solid technical foundations to address model uncertainty.

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

6 Conclusions

In this paper, a Bayesian Model Average method is employed in the economic structure research to address model uncertainty when studying the effects of economic diversity on long-term economic stability and short-term employment growth among 359 MSAs in the contiguous U.S. Compared to previous studies back to the 1970s, this method considers the impacts of control variables, spatial dependence of the dependent and independent variables, and, more importantly, addresses model uncertainty resulting from the set of control variables and spatial weight matrix. A spatial Durbin model with a heteroscedastic error structure is also employed to estimate the effect of economic diversity. It is expected that the methodology used in this paper can benefit future empirical research on economic diversity. Moreover, building on the work of recent economic structure research, this paper also uses functional economic regions to approach regional economic systems and considers the diversity of economic clusters.

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

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

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Attaran, M. (1986). Industrial Diversity and Economic Performance in U.S. Areas. *The Annals of Regional Science*, 20(2):44–54.

Chen, J. (2017). Geographical Scale, Industrial Diversity and Regional Economic Stability. RRI Working Paper Series 2017-03. Regional Research Institute, West Virginia University.

Chen, J. (2018). Interpreting Economic Diversity as the Presence of Multiple Specializations. RRI Working Paper Series 2018-02. Regional Research Institute, West Virginia University.

Conroy, M. E. (1975). The Concept and Measurement of Regional Industrial Diversification. *Southern Economic Journal*, 41(3):492–505.

Crespo Cuaresma, J., Doppelhofer, G., and Feldkircher, M. (2014). The Determinants of Economic Growth in European Regions. *Regional Studies*, 48(1):44–67.

Crespo Cuaresma, J. and Feldkircher, M. (2013). Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe. *Journal of Applied Econometrics*, 28(4):720–741.

Delgado, M., Porter, M. E., and Stern, S. (2016). Defining Clusters of Related Industries. *Journal of Economic Geography*, 16(1):1–38.

Deller, S. and Watson, P. (2016a). Spatial Variations in the Relationship between Economic Diversity and Stability. *Applied Economics Letters*, 23(7):520–525.

Deller, S. C., Conroy, T., and Watson, P. (2017). Women Business Owners: a Source of Stability during the Great Recession? *Applied Economics*, 49(56):5686–5697.

Deller, S. C. and Watson, P. (2016b). Did Regional Economic Diversity Influence the Effects of the Great Recession? *Economic Inquiry*, 54(4):1824–1838.

Dissart, J. C. (2003). Regional Economic Diversity and Regional Economic Stability: Research Results and Agenda. *International Regional Science Review*, 26(4):423–446.

Duffy, N. E. (1994). the Determinants of State Manufacturing Growth Rates: a TwoDigitLevel Analysis. *Journal of Regional Science*, 34(2):137–162.

Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2011). Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants. *Journal of Applied Econometrics*, 26(1):30–55.

Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100(2):381–427.

Fernández, C., Ley, E., and Steel, M. F. J. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16(5):563–576.

Getis, A. and Griffith, D. A. (2002). Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis*, 34(2):130–140.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126–1152.

Hong, S. and Xiao, Y. (2016). The Influence of Multiple Specializations on Economic Performance in U.S. Metropolitan Areas. *Sustainability*, 8(9):963–17.

Isserman, A. M. and Westervelt, J. (2006). 1.5 Million Missing Numbers: Overcoming Employment Suppression in County Business Patterns Data. *International Regional Science Review*, 29(3):311–335.

Jackson, R. W. (1984). An Evaluation of Alternative Measures of Regional Industrial Diversification. *Regional Studies*, 18(2):103–112.

Jackson, R. W. (2015). Are Industry Clusters and Diversity Strange Bedfellows? *Review of Regional Studies*, pages 1–17.

Jacobs, J. (1969). *The Economy of Cities*. Vintage Books.

Kort, J. R. (1981). Regional Economic Instability and Industrial Diversification in the US. *Land Economics*, 57(4):596–608.

LeSage, J. P. and Fischer, M. M. (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis*, 3(3):275–304.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.

LeSage, J. P. and Parent, O. (2007). Bayesian Model Averaging for Spatial Econometric Models. *Geographical Analysis*, 39(3):241–267.

Ley, E. and Steel, M. F. J. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression. *Journal of Applied Econometrics*, 24(4):651–674.

Madigan, D., York, J., and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.

Malizia, E. E. and Ke, S. (1993). The Influence of Economic Diversity on Unemployment and Stability. *Journal of Regional Science*, 33(2):221–235.

Mizuno, K., Mizutani, F., and Nakayama, N. (2006). Industrial Diversity and Metropolitan Unemployment Rate. *The Annals of Regional Science*, 40(1):157–172.

Parent, O. and LeSage, J. P. (2012). Determinants of Knowledge Production and Their Effects on Regional Economic Growth. *Journal of Regional Science*, 52(2):256–284.

Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, 68(2):73–93.

Porter, M. E. (1998). *On Competition*. Harvard Business Press.

Porter, M. E. (2003). The Economic Performance of Regions. *Regional Studies*, 37(6-7):549–578.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, 92(437):179–191.

Sala-i-Martin, X. (1997). I Just Ran Four Million Regressions. *American Economic Review*, 87(2):178–183.

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94(4):813–835.

Siegel, P. B., Johnson, T. G., and Alwang, J. (1995). Regional Economic Diversity and Diversification. *Growth and Change*, 26(2):261–284.

Spencer, G. M., Vinodrai, T., and Gertler, M. S. (2010). Do Clusters Make a Difference? Defining and Assessing Their Economic Performance. *Regional Studies*, 44(6):697–715.

Stock, J. and Watson, M. W. (2007). *Introduction to Econometrics, 2nd ed.* Pearson Education, Inc., Boston, MA.

Tiefelsdorf, M. and Griffith, D. A. (2016). Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A*, 39(5):1193–1221.

Trendle, B. (2006). Regional Economic Instability: the Role of Industrial Diversification and Spatial Spillovers. *The Annals of Regional Science*, 40(4):767–778.

Wagner, J. E. (2000). Regional Economic Diversity: Action, Concept, or State of Confusion. *Journal of Regional Analysis and Policy*, 30(2):1–22.

Wagner, J. E. and Deller, S. C. (1998). Measuring the Effects of Economic Diversity on Growth and Stability. *Land Economics*, 74(4):541–17.

Watson, P. and Deller, S. (2017). Economic Diversity, Unemployment and the Great Recession. *The Quarterly Review of Economics and Finance*, 64:1–11.

Winkler, R., Deller, S. C., and Marcouiller, D. (2015). Recreational Housing and Community Development: A Triple Bottom Line Approach. *Growth and Change*, 46(3):481–500.

Zellner, A. (1986). On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. In Goel, P. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier Science Publishers, Inc.