JRAP 40(1): 71-83.

The Journal of
Regional
Analysis
& Policy

States' Nonagricultural Employment at the 3-Digit North American Industry Classification System (NAICS) Level

Edward Nissan and George Carter
University of Southern Mississippi – USA

Abstract. The focus of this research is to address the question of nonagricultural employment diversity at the state level for 2002, employing the newer 3-digit North American Industry Classification System (3-digit NAICS). The index of diversity used is the Simpson index. A second facet of the paper includes a comparison of states' diversity to that of the United States as the norm. The results indicate that, with a few exceptions, diversity of employment in the majority of states does not differ statistically from employment diversity in the United States as a whole. Further findings indicate that specialization explains employment growth with statistically significant correlation between employment growth and specialization.

1. Introduction

Malizia and Ke (1993) define diversity as the variety of economic activity reflecting the economic structure of a region measured at a specific time and stability as the absence of variation in economic activity. They explain that an area with diverse industries should experience stable economic growth and less unemployment as opposed to specialized areas. Malizia and Ke provide two interpretations. The first is that employment is more severe in some industries than others and that the cyclical timing differs. The second is that a specialized economy could have many different industries that fluctuate in severity and timing. St. Louis' (1980) definition of industrial diversity is the presence of a wide variety of industries. In this way, a region is insulated from business cycle swings in contrast to specialized industries that may be subject to boom and bust. The results are, perhaps, high unemployment, income instability and migration.

Lim (2004) provides a comprehensive summary of the literature concerning agglomeration of firms. Two camps of thought are identified. The first camp claims that knowledge spillovers should not be viewed as the typical reason for the localization of industries. In opposition, the second camp claims that knowledge spillovers are the essential ingredients for localization of industries. Lim expands the review of literature by distinguishing two strands as sources of externalities. They are specialization within specific industries and transmission across industries. The specialization within specific industries was espoused by Marshall (1920). Lim explains that Marshall forwarded the opinion that geographic specialization of industries is a result of their proximity, which enables them to share in transmission of knowledge. The transmission of knowledge across industries, according to Lim, was expounded on by Jacobs (1970). Jacobs believed that local activities in a city are important elements in the process of innovation. Firms have incentive to cluster due to the effects of externalities of specialization and diversity. This is because firms cluster to take advantage of agglomeration in geographic proximities. Paris, London and Berlin are examples of dominant cities because they became large and industrially diversified.

Feldman and Andretsch (1999) researched the question as to whether diversity or specialization of economic activity enhances technological change and its resultant economic growth. There are policy implications for the two alternatives in that if specialization promotes growth, then a narrow set of economic activities in specific geographic regions yields a better innovative outlook. However, better output of innovative activity may be associated with diversity.

Feldman and Andretsch found evidence to reject the specialization thesis and support the diversity thesis.

On a similar theme, Rosenthal and Strange (2003a) studied the extent of agglomerative externalities where they contend that many previous works assumed that agglomeration economies operate at a metropolitan scale. Their results indicate that agglomeration initially weakens very rapidly and tapers off gradually with distance. They say (p. 378) that

> this pattern is consistent with both theoretical models of the internal structure of cities and stylized facts: moving away from a city center, land and house rents, building heights and population density, all decline rapidly at first and slowly thereafter. These findings suggest that agglomeration should ideally be studied at a much more refined geographic level than has been the norm.

In other words, there is a rapid spatial decay of agglomeration economies even within cities. In addition, Rosenthal and Strange (2003b) studied the relationship between agglomeration and hours worked. They showed that the agglomeration and intensity of work are tied together and that the impacts of agglomeration on professional and non-professional workers differ. Professionals work more hours, and they are more drawn to agglomerated areas.

A good argument can be made for investigating employment diversity at the state or regional levels because state policies make a difference. As Blackley (1994) notes, state governments' policies aim to achieve higher economic growth in the long run and, at the same time, reduce volatility of employment in the short run. On a similar theme, Stirboeck (2006), in a study on specialization patterns across European Union regions, declares that specific effects, especially for employment specialization, are evident at the country level. This implies that policies at the country level, and by extension at the state level in the United States, are warranted.

In particular, Wundt's (1992) concern was diversification in the state of Connecticut, with the remark that policymakers benefit from the study by using the results as guidance for policies to identify industries that promote stability. Attaran and Zwick (1987a) did a similar study for the state of Oregon, noting that diversification was viewed after the 1930s depression as an important policy consideration because specialization was a dangerous liability. Attaran and Zwick (1987b) followed their Oregon study to encompass the 51 states (including DC) to assess the industrial employment diversity between 1972 and 1981 at the 1-digit Standard Industrial Classification (SIC) system aggregation.

This paper addresses employment diversity at the state level based on the 3-digit North American Industry Classification System (3-digit NAICS) for 2002, the latest year for which data is available. The index of diversity (to be discussed later) is the Simpson index. A second facet of the paper includes comparisons of states' employment diversity to average state employment diversity as well as diversity as compared to the United States.

2. The Data

The 3-digit NAICS data for 2002 obtained from the Bureau of the Census (2007) was employed for use in calculating the states' employment diversity index. There are 84 sectors to work with. Prominent NAICS 2-digit sectors are Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific and Technical Services; Management of Companies and Enterprises; Administrative and Support and Waste Management and Remediation Services; Education Services; Health Care and Social Assistance; Arts, Entertainment and Recreation; Accommodation and Food Services; and Other Services (except Public Administration).

3. Methodology

3.1. The diversity indexes

Siegel, Johnson and Alwang (1995), henceforth SJA, provide a summary of indexes for measuring diversity (the state of unlikeness). It is a static and positive concept. The basic idea of diversity measures is interregional comparisons. The recommended measures include the Ogive Index, the Entropy Index, Location Quotient, National-Average Index, Portfolio Variance, Input-Output Analysis, which incorporates elements of portfolio theory, and the Herfindahl Index.

For each of the proposed indexes, SJA also points to its strength and deficiency. For instance, the attention of the Entropy and the Ogive indexes is focused on the variety rather than types of sectors. A deficiency of the Portfolio index, according to SJA and Siegel, Alwang and Johnson (1994), is that regions do not have as much control over their portfolios as investors do. Furthermore, the use of variance is not a good measure of diversification, according to St. Louis (1980). St. Louis gives the example that a region could have a rapid growth rate in some industrial sectors to make up for large variance in employment. A region with small variance could have zero or negative growth rate. Malizia and Ke (1993) also explain that portfolio theory refers to trade offs between growth and stability. SJA concludes that definitions of economic diversity and diversification overlap at times and conflict at others where the multiple definitions, meanings and measures contribute to ambiguity and confusion. SJA also conclude that there are two alternative aims to measure diversity. The first is to derive testable hypotheses. The second is to model relationships between changing economic structure and performance.

Wagner (2000) updated the summary of indices done by SJA (1995) by, first, proposing two concepts to define diversity and, second, reviewing some diversity measures. Wagner explains that theory on diversity provides conflicting assessments to policy makers. On one side, the theory suggests that stability can be achieved through more diversity. On the other side, the theory suggests that specialization contributes to growth. The appropriate policy according to Wagner is to have a short-run aim to promote growth and a long-run aim to promote stability with growth.

Wagner (2000) also provides a review of diversity measures by classifying them into four categories. The categories are: (1) equiproportional; (2) type of industries; (3) portfolio; and (4) input-output. Here, the equiproportional indices assume equiproportional levels of economic activity in all industries. Wagner shows that the equiproportional measures are questionable theoretically and empirically. The types of industry measures include percent durable goods, location quotient and shift-share. Wagner shows that the type of industry measures also suffer theoretically and empirically. The third category of measures is the portfolio, where a portfolio is considered efficient if the mean-variance of other portfolios do not give either a higher expected return for the same variance or a lower variance for the same return. Again, Wagner shows some deficiencies as explained earlier in this paper. The final category of suggested measures is the input-output which is the most favored by Wagner because it captures the structure and performance of a region more fully.

Wagner and Dollar (1998) have implemented an input-output approach which can account for inter-industry linkages as well as the structures of the regional economy. This way, policy makers can address a variety of policy issues such as a policy that may affect the structure of the economy of a region. Of importance is whether a given level of diversity at a specific time period leads to growth and stability in a future time period. For this purpose, Wagner and Dollar use five factors that influence economic growth in a region. These factors are markets, labor, taxes, amenities and infrastructure. Accordingly, the input-output scheme used captured the size of the regional economy and the degree of inter-industry linkages. Along with the use of a growth and stability model, the Wagner and Dollar results show statistically that there is an association between higher levels of diversity and higher levels of growth and stability.

In the ecological realm, the question of diversity receives a great deal of attention. For this discipline, according to Pielou (1975), the question revolves around the composition and structure of many-species communities. To this end, after some consideration, Pielou recommends two indexes, one of which is the Simpson (1949) index used in this study.

Let p_i be the proportionate (relative frequency) share received by category i among n well-defined categories, $\sum p_i = 1$. The Simpson (1949) index given by

$$H' = 1 - \sum p_i^2, \; 1 \le i \le n, \tag{1}$$

where H' ranges from 0, when all shares are contained in one sector, to $1-1/n$, where shares are distributed equally among the n sectors. Jacquemin and Berry (1979) explain that H' in equation (1) is a transformation of the Herfindahl business concentration (specialization) index

$$H = \sum p_i^2, \quad 1 \le i \le n, \tag{2}$$

where H ranges between $1/n$, when all shares are distributed equally among the n companies, and 1, when all shares are held by one company (monopoly). There is an apparent relationship between the Herfindahl H of equation (2) and, by implication, the Simpson index of equation (1) with the variance as shown by Grossack (1965). Knowing that $\sum p_i = 1$, the mean $\bar{p}$ is

$$\bar{p} = \frac{\sum p_i}{n} = \frac{1}{n} \tag{3}$$

and the variance, S^2, is

$$S^2 = \frac{\sum (p_i - \bar{p})^2}{n-1} = \frac{\sum (p_i - 1/n)^2}{n-1}. \tag{4}$$

Equation (4) is a handy expression for testing equality of two variances by use of the test statistic

$$F^* = \frac{S_1^2}{S_2^2} \tag{5}$$

and comparing to the tabular F-distribution critical value $F(\alpha, n\text{-}1, n\text{-}1)$ for significance level α. For $\alpha = 0.05$ and n=84, F(.05, 83, 83) $\cong$ 1.46. Equation (5), therefore, can be used, as will be shown later, as a way to test equality of specialization between the states and specialization of the United States.

Another way to portray the results of H of equation (2) is through a normalized "numbers-equivalent," which is the number of sectors m yielding H if the numbers of employees are of equal size. According to Miller (1972), given H and n sectors, m provides the number of equalized sectors to generate a level of H,

$$m = \frac{1}{H}. \tag{6}$$

Routledge (1977) shows that an index of diversity must satisfy some reasonable properties. In particular, it must be one of the indices in the family

$$N_a = \left[\sum_i p_i^a\right]^{\frac{1}{1-a}}, a > 0 \tag{7}$$

where p_i = proportionate share. Accordingly, this family contains the Simpson index

$$N_2 = \frac{1}{\lambda} \tag{8}$$

where $\lambda = \sum p_i^2$ = Simpson index of concentration.

Now in terms of the symbols used in defining equation (1), equation (2) and equation (6),

$$N_2 = \left[\sum_i p_i^2\right]^{\frac{1}{1-2}} = \left[\sum_i p_i^2\right]^{-1} = 1 \Big/ \sum p_i^2 = 1 \big/ \lambda \tag{9}$$

This inverse measure of Simpson's concentration is depicted as $m = 1/H$ (equation 6). Thus, this family of measures satisfies Routledge's (1977) requirement "that diversity ought to be quantified only by an index from the family $\{N_a$ for $a > 0\}$" (p. 507).

Lande (1996) claims that a measure of diversity is ideally nonparametric (distribution free) and statistically accurate, having small bias and sampling variance in samples of moderate size, and be strictly convex. Lande shows that the standard deviation of the Simpson diversity, $1-\lambda = H'$ (equation 1), is not only unbiased, but it also has the smallest standard deviation. The Simpson concentration measure $\lambda = H$ (equation 2) is strictly convex.

Two modes of analysis are employed for comparisons. The first is a comparison of the Herfindahl index, H of equation (2), and, by implication, the Simpson index of equation (1) of each state with the mean of H, $\overline{H}$. This is done by redefining each state's H values in terms of its distance from the mean ($\overline{H}$) in standard deviations (S), denoted as the standardized variable by the transformation

$$Z_i = \frac{H_i - \overline{H}}{S}. \tag{10}$$

For values of Z_i beyond $2S$ from $\overline{H}$, H_i for state i by the empirical rule is considered unusual. For Z_i beyond $3S$, H_i is termed an outlier, which is far enough from the majority of data that it perhaps arose from a specific cause.

The second mode of analysis relies on the concept of what is known as sigma (σ) convergence whereby each state's 84-sectors variance S_i^2 is compared with the sectoral variance of the United States, S^2. Note that the concept of σ-convergence used here is static at a given time period in looking at a cross section of states and not a dynamic process as perceived by Barro and Sala-i-Martin (1992). Barro and Sala-i-Martin find evidence where regions converge toward a national steady state. According to Lall and Yilmaz (2001), the sigma convergence compares, for instance, dispersion of per-capita income to a common rate level. Ray and Dev (2006) focus the idea for use in sigma convergence on the sample variance as defined for this study by equation (4). Goerlich and Mas (2004) contend that equality in a distribution requires the distribution not only to tend toward the first moment (the mean), but that the second moment (the dispersion) must also narrow. Dalgaard and Bastrup (2001) state that the measure for sigma convergence is the coefficient of variation (CV). Friedman (1992) concurs that CV is the only appropriate measure for convergence.

Now, in terms of this paper for a given state i,

$$(CV_i)^2 = \frac{S_i^2}{(\bar{x}_i)^2} = \frac{S_i^2}{\overline{P}_i^2}. \tag{11}$$

Since $\overline{P}_i$ by equation (3) is $1/n_i$ and since the number of sectors n_i = 84 for all the states as well as the number of sectors in the United States, then comparing the convergence of sectoral employment to the norm of the United States, the ratio of state i to the U.S. is

$$\frac{(CV_i)^2}{(CV_{US})^2} = \frac{S_i^2/(1/n)^2}{S_{US}^2/(1/n)^2} = \frac{S_i^2}{S_{US}^2}, \tag{12}$$

tested for significance by the F-test of equation (5).

It should be pointed out that the measure of diversity used in this paper uses the concept of equiproportional level of economic activity as explained by Wagner (2000) and is questioned, as most indexes, both theoretically and empirically. Wagner also shows that the choice of diversity index should be relative to a standard or norm. The Simpson index employed in this paper does not account for a standard or norm. However, the comparisons employed in equation (10), where the norm is the average of the index for all the states, and the comparisons employed in equation (12), where the norm is the U.S., both for the year 2002, provide related standards in the aggregate for comparisons.

4. Results

To provide overall descriptions of the data employed in this research, three appendices are provided. Appendix A displays the 84 nonagricultural employment sectors at the 3-digit NAICS for the United States, giving total employments and their percentages. Among the largest employment sectors are: specialty trade contractors (sector 238) at 4.4 million; professional, scientific and technical services (541) at 7.2 million; administrative and support services (561) at 8.4 million; ambulatory health care services (621) at 4.9 million; hospitals (622) at 5.2 million; and food services and drinking places (722) at 8.3 million.

Appendix B provides descriptive statistics for state employment, giving the total, mean, standard deviation, minimum and maximum. Total employment in the United States in the 84 sectors in 2002 is approximately 109 million. The largest state employments were in California, Florida, New York and Texas, all exceeding 6 million. The largest is in California at 12.6 million. In a similar fashion, Appendix C provides descriptive statistics when the data of Appendix B are transformed to proportions. The proportional mean for all the states is the same 1/84=0.0119. The standard deviations differ, giving a hint for the level of concentration (specialization) among sectors.

Table 1 presents in alphabetical order the state Simpson diversity index, H', with its alternative measure H (the concentration or specialization index) and its numbers-equivalent. The Simpson index ranges from 0.9123 at the lowest, attained by the District of Columbia (DC), to 0.9725 at the highest, for Wisconsin, with the implication that DC is the most sectorally specialized as compared to Wisconsin, the least sectorally specialized. Thus, the larger the H', the more diversified is the state's nonagricultural employment. The complement of the Simpson H' by equation (1) is the Herfindahl H; therefore, DC attains the maximum of H (0.0877), and Wisconsin attains the smallest H (0.0275).

Table 1. Indexes for states at 3-Digit NAICS.

State	H'	H	m
Alabama	0.9689	0.0311	32.18
Alaska	0.9664	0.0336	29.79
Arizona	0.9632	0.0368	27.14
Arkansas	0.9710	0.0290	34.45
California	0.9652	0.0348	28.71
Colorado	0.9563	0.0437	22.90
Connecticut	0.9680	0.0320	31.21
Delaware	0.9615	0.0385	26.00
D. of Columbia	0.9123	0.0877	11.41
Florida	0.9501	0.0499	20.05
Georgia	0.9672	0.0328	30.45
Hawaii	0.9558	0.0442	22.60
Idaho	0.9641	0.0359	27.89
Illinois	0.9680	0.0320	31.28
Indiana	0.9697	0.0303	33.02
Iowa	0.9713	0.0287	34.82
Kansas	0.9695	0.0305	32.75
Kentucky	0.9705	0.0295	33.93
Louisiana	0.9669	0.0331	30.20
Maine	0.9683	0.0317	31.59
Maryland	0.9611	0.0389	25.74
Massachusetts	0.9656	0.0344	29.10
Michigan	0.9651	0.0349	28.64
Minnesota	0.9705	0.0295	33.86
Mississippi	0.9691	0.0309	32.32
Missouri	0.9694	0.0306	32.72
Montana	0.9644	0.0356	28.09
Nebraska	0.9661	0.0339	29.50
Nevada	0.9268	0.0732	13.66
New Hampshire	0.9694	0.0306	32.64
New Jersey	0.9655	0.0345	28.97
New Mexico	0.9633	0.0367	27.23
New York	0.9654	0.0346	28.89
North Carolina	0.9708	0.0292	34.27
North Dakota	0.9671	0.0329	30.44
Ohio	0.9688	0.0312	32.00
Oklahoma	0.9664	0.0336	29.74
Oregon	0.9688	0.0312	32.08
Pennsylvania	0.9700	0.0300	33.32
Rhode Island	0.9661	0.0339	29.47
South Carolina	0.9669	0.0331	30.23
South Dakota	0.9665	0.0335	29.83
Tennessee	0.9692	0.0308	32.45
Texas	0.9647	0.0353	28.36
Utah	0.9653	0.0347	28.78
Vermont	0.9703	0.0297	33.65
Virginia	0.9620	0.0380	26.33
West Virginia	0.9661	0.0339	29.54
Wisconsin	0.9725	0.0275	36.35
Wyoming	0.9667	0.0333	30.02
United States	0.9681	0.0319	31.37

Note: H' is the Simpson index (equation 1), H is the Herfindahl index (equation 2), and m is its numbers-equivalent (equation 6).
Source: Bureau of Census (2007).

The numbers-equivalent m for the H index by equation (6) corresponds to m = 11.41 for DC and to m = 36.35 for Wisconsin. The meaning of this comparison is that if, for DC, its H of 0.0877 is spread equally among the sectors, the number of sectors dwindles from n = 84 to m = 11.41, a reduction of 86.4 percent. For Wisconsin, the reduction from n = 84 to m = 36.35 is much less at 56.7 percent. A summary of mean and standard deviation for H', H and m for the 51 states is shown in Table 2. Note that for H' and H, the standard deviation is the same, S = 0.0102.

Table 2. Summary of diversification measures.

Index	Mean	Standard Deviation
H'	0.9646	0.0102
H	0.0354	0.0102
m	29.47	4.71

Note: H' is the Simpson index, H is the Herfindahl index, and m is its numbers-equivalent.
Source: Bureau of Census (2007).

Table 3 is a partial rearrangement of Table 1, displaying each state in ascending order of ranking from the least diverse (specialized) to the most diverse as portrayed by H. For comparison purposes, state i specialization as measured by the H index is transformed to the score Z of equation (7)

$$Z_i = \frac{H_i - 0.0354}{0.0102}, \tag{13}$$

where the column identified as Z provides the results. Note, as indicated earlier, the computations by either the Simpson index H' or by the specialization index H would have been the same because the standard deviation for both is S = 0.0102. The only difference is the sign of the Z-score, positive for H' and negative for its complement H.

For DC in Table 3, with Z = 5.14, and for Nevada, with Z = 3.72, the conclusion reached is that their sectoral employment distributions are outliers; each exceeds three standard deviations from the mean. A likely explanation is that DC is highly specialized in government employment, while Nevada is specialized in the gaming and tourism industries. Florida, with Z = 1.42, is somewhat more specialized than the rest of the states.

Note that declaring DC and Nevada as outliers in specialization was based on the empirical rule as provided in the Methodology section, which assumes that the underlying distribution is normal. The distribution

Table 3. States Ranked for Diversity at 3-digit NAICS.

State	H	Z	S	$S^2/S^2(US)$
D. of Columbia	0.0877	5.14	0.0300	3.7918
Nevada	0.0732	3.72	0.0270	3.0681
Florida	0.0499	1.42	0.0213	1.9007
Hawaii	0.0442	0.86	0.0196	1.6189
Colorado	0.0437	0.81	0.0194	1.5898
Maryland	0.0389	0.34	0.0179	1.3490
Delaware	0.0385	0.30	0.0178	1.3292
Virginia	0.0380	0.25	0.0176	1.3056
Arizona	0.0368	0.13	0.0172	1.2483
New Mexico	0.0367	0.13	0.0172	1.2425
Idaho	0.0359	0.05	0.0169	1.1988
Montana	0.0356	0.02	0.0168	1.1865
Texas	0.0353	-0.01	0.0167	1.1690
Michigan	0.0349	-0.05	0.0166	1.1521
California	0.0348	-0.06	0.0165	1.1474
Utah	0.0347	-0.07	0.0165	1.1435
New York	0.0346	-0.08	0.0164	1.1368
New Jersey	0.0345	-0.09	0.0164	1.1321
Massachusetts	0.0344	-0.10	0.0163	1.1240
Rhode Island	0.0339	-0.15	0.0162	1.1028
Nebraska	0.0339	-0.15	0.0162	1.1009
West Virginia	0.0339	-0.15	0.0162	1.0987
Oklahoma	0.0336	-0.18	0.0161	1.0871
Alaska	0.0336	-0.18	0.0161	1.0842
South Dakota	0.0335	-0.19	0.0160	1.0821
Wyoming	0.0333	-0.21	0.0160	1.0716
Louisiana	0.0331	-0.23	0.0159	1.0615
South Carolina	0.0331	-0.23	0.0159	1.0600
North Dakota	0.0329	-0.25	0.0158	1.0488
Georgia	0.0328	-0.26	0.0158	1.0481
Connecticut	0.0320	-0.34	0.0155	1.0081
Illinois	0.0320	-0.34	0.0155	1.0047
Maine	0.0317	-0.37	0.0153	0.9890
Ohio	0.0312	-0.42	0.0152	0.9683
Oregon	0.0312	-0.42	0.0151	0.9645
Washington	0.0311	-0.43	0.0151	0.9598
Alabama	0.0311	-0.43	0.0151	0.9595
Mississippi	0.0309	-0.45	0.0151	0.9528
Tennessee	0.0308	-0.46	0.0150	0.9467
New Hampshire	0.0306	-0.48	0.0149	0.9379
Missouri	0.0306	-0.48	0.0149	0.9340
Kansas	0.0305	-0.49	0.0149	0.9327
Indiana	0.0303	-0.50	0.0148	0.9203
Pennsylvania	0.0300	-0.53	0.0147	0.9063
Vermont	0.0297	-0.56	0.0146	0.8919
Minnesota	0.0295	-0.58	0.0145	0.8824
Kentucky	0.0295	-0.58	0.0145	0.8794
North Carolina	0.0292	-0.61	0.0143	0.8650
Arkansas	0.0290	-0.63	0.0143	0.8573
Iowa	0.0287	-0.66	0.0141	0.8419
Wisconsin	0.0275	-0.78	0.0136	0.7813
United States	0.0319		0.0154	1.0000

Note: H is the Herfindahl index, Z is the Z-score of H, S is state standard deviation and $S^2/S^2(US)$ is the F-ratio for testing equality of variances by equation (5).
Source: Bureau of the Census (2007).

at hand is quite skewed and would likely invalidate the use of empirical rule. In this case, the outliers should be interpreted as values that are clearly separated from other values. The Z measures of the outliers, therefore, indicate extreme deviations. Miller, Jr. (1981) defines an outlier as a single observation that does not conform to the rest of the data.

Perhaps a better approach to stress the variation is to use Chebyshev's Theorem, which does not assume normality. According to Lind, Marchal and Wathen (2010, p. 82), "for a set of observations (sample on population), the proportion of the values that lie within k standard deviations of the mean is at least 1-k^2, where k is any constant greater than one."

By this theorem, at least $1-1/(5.1447)^2 = 1-0.038 = 0.962$ of the values would lie closer to the mean than DC, and at least $1-1/(3.7175)^2 = 1-0.072 = 0.928$ would lie closer than Nevada, so both areas do not conform to the rest of the data. For Florida, with smaller Z score, only $1-1/(1.4242)^2 = 1-0.493 = 0.507$ or more of the values lie closer to the mean.

There are twelve states with positive Z-scores, with the implication that these states are relatively more specialized than the mean. None of the remaining 39 states had a Z-score < -1.00 where the range was from Z=-0.0128 for Texas to Z=-0.7805 for Wisconsin, with the indication that their H index falls within one standard deviation to the left of the mean. An overall conclusion can be reached that, with the exceptions of DC, Nevada and Florida, the remaining states had their H values within $\overline{H} \pm 1S$.

Table 3 also portrays the proportion standard deviation (S) obtained from equation (4) as transposed from Appendix C. Note the complete correspondence in the rankings for both H and S as explained earlier. The column headed by $S^2/S^2(US)$ provides the F-test from equation (5) for testing convergence of states' sectoral employment to the norm of the United States, where $S^2(US) = (0.0154)^2$. Again, the states that show divergence with statistical significance are DC, Nevada and Florida along with Hawaii and Colorado, where the critical value of the test statistic is F* = 1.46. Overall, it seems that these states diverge in their sectoral employment more so than sectoral employment in the United States. No state showed divergence as being less diversified from the U.S. norm in that the critical $F^{**} = 1/F(\alpha, n_1-1, n_2-1)$. For $\alpha = 0.05$, with $F^*(.05, 50, 50) = 1.46$, $F = 1/1.46 = 0.6849$. The lowest F, 0.7813 for Wisconsin, exceeds 0.6849. The overall comparisons, whether the norm is the mean of the states or the norm is the United States, imply that the majority of the states do not display exceptional diversity (specialization) in their sectoral employment.

One interpretation is that, because the states are more similar than different, the diversity index employed may not be a good explanation for differences in relative economic performance. Also, economies could look fairly similar in aggregate, but could look very different at the detailed level. Nevertheless, following the suggestion of Green and Deller (1998) regarding the effects of diversity on economic growth and stability, it is of interest to perform a regression of employment growth on the diversity measure, ceteris paribus.

The results of the regression for the year 2002 using the data for 51 states is

$$\Delta E = -0.0223 + 0.5074H \qquad (14)$$

where ΔE = employment growth for the year 2002, and H = the specialization index H of equation (2).

The regression coefficient is statistically significant at p-value = 0.0045, with the correlation coefficient of r = 0.391. This indicates that the employment specialization index can provide some explanation for employment growth.

When the five states displaying divergence (DC, Nevada, Florida, Hawaii and Colorado) were eliminated for a new regression, the result of regressing employment growth on H is,

$$\Delta E = -0.0419 + 1.0978H. \qquad (15)$$

The coefficient for H is now less significant at p-value = 0.0931 with correlation r = 0.251, indicating significance at $\alpha = 0.10$. Therefore, the regressions of equations (14) and (15) give some support to the proposition that employment specialization could provide some credible explanation for employment growth.

The interaction of specialization H with ΔGSP, the growth of gross state product, is another relationship worth probing. The regression equation for the 51 states is

$$\Delta GSP = 0.0310 + 0.2652H, \; r = 0.145. \qquad (16)$$

Both the regression and correlation coefficients are not significant at p-value = 0.3103. However, the signs are appropriate in that specialization could explain the levels of growth of GSP.

When the regression is done for the 46 states only, the sign of the coefficient for H was reversed, giving,

$$\Delta GSP = 0.0666 - 0.8160H, \; r = -0.117. \qquad (17)$$

The regression and correlation coefficients are not significant at p-value = 0.4403. The explanation for the reversal of sign is that, perhaps, the 46 states that are more diversified rely on the diversification of their employment for their economic stability.

Note that one would get the negative values of the coefficients in equations 14 through 17 if the preference is for *H´* rather than *H*. As explained earlier, the Simpson diversity index *H´* of equation (1) is the complement of equation (2).

5. Concluding remarks

Assanie and Yücel (2007) enumerate many characteristics that contribute to economic performance, such as amenities, natural resources, labor force characteristics and industrial mix along with industry agglomeration. Groups of companies concentrate geographically because of related technologies and positive spillovers of specialized labor, vital resources and intermediate input suppliers.

This paper provided a literature review of some views which considered the more sectorally diverse employment as the preferred mode and some views which considered the more specialized employment as the preferred mode. The debates among the two camps are ongoing, and each camp provides logical arguments in defense of its views. The paper also provided a detailed description of the industrial mix at the state level using the 3-digit NAICS for employment as a guide. The index (*H*) of specialization chosen is qualified to compare the states' employment diversity with that of the United States as the norm.

The results indicate that, with some exceptions, most states are comparable in their employment diversity (specialization) to the employment diversity (specialization) of the United States. It is recognized that industry clusters, innovation systems and localized economies versus urbanized economies are some of the most important engines to shape states' diversity or specialization. An exhaustive analysis is beyond the purpose of this paper. Instead, the purpose, which could be a contribution to the regional science literature, is the provision of a look at the states' employment diversity (specialization) using the new 3-digit NAICS as the source of analysis.

References

Assanie, L., and M Yücel. 2007. Industry Clusters Shape Texas Economy. Federal Reserve Bank of Dallas *Southwest Economy* (September/October). http://dallasfed.org/research/swe/2007/swe0705b.cfm.

Attaran, M., and M. Zwick. 1987a. Entropy and Other Measures of Industrial Diversification. *Quarterly Journal of Business and Economics* 26: 17-34.

Attaran, M., and M. Zwick. 1987b. An Information Theory Approach to Measuring Industrial Diversification. *Journal of Economic Studies* 16:19-30.

Barro, R.J., and X. Sala-i-Martin. 1990. Convergence. *Journal of Political Economy* 100: 223-251.

Blackley, P.R. 1994. The Impact of Slower Growth and Deindustrialization upon State Output Volatility. *The Review of Regional Studies* 24: 37-53.

Bureau of the Census. 2007. Three-digit North American Industry Classification System www.census.gov/econ/census02/guide/geosumm.htm, accessed on 9/27/2007.

Dalgaard, C-J, and J. Bastrup. 2001. On the Measurement of σ-Convergence. *Economics Letters* 70: 283-287.

Feldman, M.P., and D.B. Andretsch. 1999. Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition. *European Economic Review* 43: 409-429.

Friedman, M. 1992. Do Old Fallacies Ever Die? *Journal of Economic Literature* 30: 2129-32.

Goerlich, F.J., and M.Mas. 2004. Three (Marginal?) Questions Regarding Convergence. *Journal of Economic Studies* 31: 25-38.

Grossack, I.M. 1965. Towards an Integration of Static and Dynamic Measures of Industry Concentration. *Review of Economics and Statistics* 47: 301-308.

Jacquemin, A.P., and C.H. Berry. 1979. Entropy Measure of Diversification and Corporate Growth. *The Journal of Industrial Economics* 27: 359-369.

Jacobs, J. 1970. *The Economy of Cities*. New York: Vintage Books.

Lall, S.V., and S. Yilmaz. 2001. Regional Economic Convergence: Do Policy Instruments Make a Difference? *The Annals of Regional Science* 35: 153-166.

Lande, R. 1996. Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. *Oikos* 76: 5-13.

Lim, U. 2004. Knowledge Spillover, Agglomeration Economics, and the Geography of Innovative Activity: A Spatial Econometric Analysis. *The Review of Regional Studies* 34: 11-36.

Lind, D.A., Marchal, W.G., and Wathen, S.A. 2010. *Statistical Techniques in Business & Economics* (14th ed.). Boston: McGraw-Hill Irwin.

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

Marshall, A. 1920. *Principles of Economics* (8th ed.). London: Macmillan.

Miller, R.A. 1972. Numbers Equivalents, Relative Entropy, and Concentration Ratios: A Comparison using Market Performance. *Southern Economic Journal* 39: 107-112.

Miller, Jr., P.G. *Simultaneous Statistical Inference.* New York: Springer-Verlag.

Pielou, E.C. 1975. *Ecological Diversity*. New York: John Wiley & Sons.

Routledge, R.D. 1979. Diversity Indexes: Which Ones are Admissible? *Journal of Theoretical Biology* 76:503-515.

Ray, S.J., and B. Dev. 2006. σ-Convergence in the Presence of Spatial Effects. *Papers in Regional Science* 85: 217-234.

Rosenthal, S.S., and W.C. Strange. 2003a. Geography, Industrial Organization, and Agglomeration. *The Review of Economics and Statistics* 85: 377-393.

Rosenthal, S.S., and W.C. Strange. 2003b. *Agglomeration, Labor Supply, and Urban Rat Race.* Center for Policy Research Working Paper No. 57: 1-35. www.centerforpolicyresearch.org.

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

Siegel, P.B., J. Alwang, and T.C. Johnson. 1994. Toward an Improved Portfolio Variance Measure of Regional Economic Stability. *The Review of Regional Studies* 24: 71-86.

Simpson, E.H. 1949. Measurement of Diversity. *Nature* 163: 688.

Stirboeck, C. 2006. A Spatial Econometric Analysis of Regional Specialization Patterns across EU Regions. *The Review of Regional Studies* 36: 324-361.

St. Louis, L.V. 1980. A Measure of Regional Diversification and Efficiency. *Annals of Regional Science* 14: 21-31.

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

Wagner, J.E., and S.C. Deller. 1998. Measuring the Effects of Economic Diversity on Growth and Stability. *Land Economics* 74: 541-560.

Wundt, B.D. 1992. Reevaluating Alternative Measures of Industrial Diversity as Indicators of Regional Cyclical Variations. *The Review of Regional Studies* 22: 59-73.

Appendix A. Number and Percentage of Nonagricultural Sector Paid Employees at 3-digit NAICS for the U.S.

Code	Description	Paid Employees	%
211	Oil and gas extraction	98,491	0.09
212	Mining, except oil and gas	197,499	0.18
213	Support activities for mining	186,731	0.17
221	Utilities	663,013	0.61
236	Construction of buildings	1,669,376	1.53
237	Heavy and civil engineering construction.	1,143,194	1.05
238	Specialty trade contractors	4,380,035	4.02
311	Food mfg	1,507,436	1.38
312	Beverage & tobacco product mfg	159,995	0.15
313	Textile mills	268,610	0.25
314	Textile product mills	182,597	0.17
315	Apparel mfg	342,689	0.31
316	Leather & allied product mfg	45,404	0.04
321	Wood product mfg	540,561	0.50
322	Paper mfg	491,548	0.45
323	Printing & related support activities	718,206	0.66
324	Petroleum & coal products mfg	102,490	0.09
325	Chemical mfg	853,520	0.78
326	Plastics & rubber products mfg	981,070	0.90
327	Nonmetallic mineral product mfg	483,288	0.44
331	Primary metal mfg	490,526	0.45
332	Fabricated metal product mfg	1,572,909	1.44
333	Machinery mfg	1,173,647	1.08
334	Computer & electronic product mfg	1,261,065	1.16
335	Electrical equipment, appliance, & component mfg	494,340	0.45
336	Transportation equipment mfg	1,680,127	1.54
337	Furniture & related product mfg	597,094	0.55
339	Miscellaneous mfg	757,495	0.69
423	Durable goods merchant wholesalers	3,374,171	3.09
424	Nondurable goods merchant wholesalers	2,298,584	2.11
425	Wholesale electronic markets and agents and brokers	246,685	0.23
441	Motor vehicle & parts dealers	1,845,496	1.69
442	Furniture & home furnishings stores	535,029	0.49
443	Electronics & appliance stores	391,000	0.36
444	Building material & garden equipment & supplies dealers	1,160,016	1.06
445	Food & beverage stores	2,838,653	2.60
446	Health & personal care stores	1,024,429	0.94
447	Gasoline stations	926,792	0.85
448	Clothing & clothing accessories stores	1,426,573	1.31
451	Sporting goods, hobby, book, & music stores	611,144	0.56
452	General merchandise stores	2,504,364	2.30
453	Miscellaneous store retailers	800,722	0.73
454	Nonstore retailers	571,438	0.52

Appendix A (continued).

Code	Description	Paid Employees	%
481	Air transportation	102,708	0.09
483	Water transportation	66,407	0.06
484	Truck transportation	1,427,475	1.31
485	Transit & ground passenger transportation	399,514	0.37
486	Pipeline transportation	38,385	0.04
487	Scenic & sightseeing transportation	24,372	0.02
488	Support activities for transportation	471,958	0.43
492	Couriers & messengers	569,554	0.52
493	Warehousing & storage	572,485	0.53
511	Publishing industries (except Internet)	1,089,585	1.00
512	Motion picture & sound recording industries	303,134	0.28
515	Broadcasting (except Internet)	291,361	0.27
516	Internet publishing & broadcasting	40,049	0.04
517	Telecommunications	1,440,141	1.32
518	Internet svc providers, web search portals, & data processing	514,046	0.47
519	Other information services	57,818	0.05
521	Monetary authorities - central bank	22,367	0.02
522	Credit intermediation & related activities	3,321,461	3.05
523	Securities intermediation & related activities	770,128	0.71
524	Insurance carriers & related activities	2,394,464	2.20
525	Funds, trusts, & other financial vehicles (part)	20,651	0.02
531	Real estate	1,314,813	1.21
532	Rental & leasing services	619,615	0.57
533	Lessors of nonfinan. intang. assets (exc copyrighted works)	27,016	0.02
541	Professional, scientific, & technical services	7,243,495	6.64
551	Management of companies & enterprises	2,605,301	2.39
561	Administrative & support services	8,410,979	7.71
562	Waste management & remediation services	332,019	0.30
611	Educational services	430,436	0.39
621	Ambulatory health care services	4,924,908	4.52
622	Hospitals	5,173,329	4.74
623	Nursing & residential care facilities	2,830,908	2.60
624	Social assistance	2,122,379	1.95
711	Performing arts, spectator sports, & related industries	423,732	0.39
712	Museums, historical sites, & similar institutions	123,107	0.11
713	Amusement, gambling, & recreation industries	1,315,379	1.21
721	Accommodation	1,813,326	1.66
722	Food services & drinking places	8,307,625	7.62
811	Repair & maintenance	1,285,405	1.18
812	Personal & laundry services	1,296,525	1.19
813	Religious/grantmaking/civic/professional & similar org	893,519	0.82

Source: Bureau of the Census (2007).

Appendix B. Descriptive statistics of paid employees by state at 3-digic NAICS.

State	Employees (000)	Mean (000)	St Dev. (000)	Min	Max
Alabama	1,512.58	18.01	22.85	17	115,141
Alaska	217.83	2.59	3.50	0	17,906
Arizona	1,950.12	23.22	33.60	34	193,110
Arkansas	928.11	11.05	13.25	72	67,406
California	12,598.72	149.98	208.12	912	1,164,306
Colorado	1,966.56	23.41	38.24	0	250,188
Connecticut	1,506.79	17.94	23.33	122	107,079
Delaware	362.02	4.31	6.44	0	35,608
District of Columbia	375.59	4.47	11.28	0	83,525
Florida	6,155.49	73.28	130.87	42	976,071
Georgia	3,225.89	38.40	50.93	374	284,221
Hawaii	418.82	4.99	8.22	0	50,781
Idaho	446.51	5.32	7.54	0	36,570
Illinois	5,075.57	60.42	78.45	734	390,759
Indiana	2,446.28	29.12	36.19	58	198,244
Iowa	1,203.25	14.32	17.03	0	86,498
Kansas	1,061.73	12.64	15.81	0	81,113
Kentucky	1,420.27	16.91	20.54	75	121,346
Louisiana	1,522.39	18.12	24.19	15	131,764
Maine	474.03	5.64	7.27	0	35,295
Maryland	2,007.25	23.90	35.95	343	198,351
Massachusetts	2,891.48	34.42	47.27	51	239,737
Michigan	3,820.24	45.48	63.23	350	294,889
Minnesota	2,294.50	27.32	33.24	91	165,575
Mississippi	866.93	10.32	13.05	0	67,797
Missouri	2,263.69	26.95	33.74	70	181,732
Montana	296.04	3.52	4.97	0	31,635
Nebraska	731.67	8.71	11.84	0	59,242
Nevada	931.89	11.09	25.17	0	200,684
New Hampshire	528.98	6.30	7.90	0	39,344
New Jersey	3,507.39	41.75	57.55	224	305,845
New Mexico	548.45	6.53	9.43	0	55,616
New York	7,082.65	84.32	116.45	306	563,081
North Carolina	3,146.25	37.46	45.12	127	246,304
North Dakota	250.16	2.98	3.95	0	20,372
Ohio	4,630.50	55.12	70.27	422	384,996
Oklahoma	1,135.22	13.51	18.25	149	100,507
Oregon	1,270.43	15.12	19.24	38	107,517
Pennsylvania	4,776.50	56.86	70.12	612	333,054
Rhode Island	396.77	4.72	6.43	0	34,993
South Carolina	1,483.12	17.66	23.55	0	136,031
South Dakota	290.34	3.46	4.66	0	24,138
Tennessee	2,204.15	26.24	33.07	89	175,963
Texas	7,717.65	91.88	128.68	983	731,139
Utah	872.56	10.39	14.39	0	88,010
Vermont	245.58	2.92	3.58	0	18,289
Virginia	2,827.41	33.66	49.82	224	309,824
Washington	2,123.86	25.28	32.09	139	171,589
West Virginia	540.64	6.44	8.74	0	45,570
Wisconsin	2,300.63	27.39	31.36	71	172,924
Wyoming	178.47	2.12	2.85	0	16,745
United States	109,029.93	1,297.98	1,681.37	20,651	8,410,979

Source: Bureau of the Census (2007).

Appendix C. Descriptive statistics of proportion of paid employees by state at 3-digit NAICS.

State	Mean	St Dev.	Minimum	Maximum
Alabama	0.0119	0.0151	0.0000	0.0761
Alaska	0.0119	0.0161	0.0000	0.0822
Arizona	0.0119	0.0172	0.0000	0.0990
Arkansas	0.0119	0.0143	0.0000	0.0726
California	0.0119	0.0165	0.0001	0.0924
Colorado	0.0119	0.0194	0.0000	0.1272
Connecticut	0.0119	0.0155	0.0000	0.0711
Delaware	0.0119	0.0178	0.0000	0.0984
D.C.	0.0119	0.0300	0.0000	0.2224
Florida	0.0119	0.0213	0.0000	0.1586
Georgia	0.0119	0.0158	0.0000	0.0881
Hawaii	0.0119	0.0196	0.0000	0.1212
Idaho	0.0119	0.0169	0.0000	0.0819
Illinois	0.0119	0.0155	0.0001	0.0770
Indiana	0.0119	0.0148	0.0000	0.0810
Iowa	0.0119	0.0141	0.0000	0.0719
Kansas	0.0119	0.0149	0.0000	0.0764
Kentucky	0.0119	0.0145	0.0001	0.0854
Louisiana	0.0119	0.0159	0.0000	0.0866
Maine	0.0119	0.0153	0.0000	0.0745
Maryland	0.0119	0.0179	0.0000	0.0988
Massachusetts	0.0119	0.0163	0.0000	0.0829
Michigan	0.0119	0.0166	0.0000	0.0772
Minnesota	0.0119	0.0145	0.0000	0.0722
Mississippi	0.0119	0.0151	0.0000	0.0782
Missouri	0.0119	0.0149	0.0000	0.0803
Montana	0.0119	0.0168	0.0000	0.1069
Nebraska	0.0119	0.0162	0.0000	0.0810
Nevada	0.0119	0.0270	0.0000	0.2154
New Hampshire	0.0119	0.0149	0.0000	0.0744
New Jersey	0.0119	0.0164	0.0000	0.0872
New Mexico	0.0119	0.0172	0.0000	0.1014
New York	0.0119	0.0164	0.0000	0.0795
North Carolina	0.0119	0.0143	0.0000	0.0783
North Dakota	0.0119	0.0158	0.0000	0.0814
Ohio	0.0119	0.0152	0.0000	0.0831
Oklahoma	0.0119	0.0161	0.0000	0.0885
Oregon	0.0119	0.0151	0.0000	0.0846
Pennsylvania	0.0119	0.0147	0.0001	0.0697
Rhode Island	0.0119	0.0162	0.0000	0.0882
South Carolina	0.0119	0.0159	0.0000	0.0917
South Dakota	0.0119	0.0160	0.0000	0.0831
Tennessee	0.0119	0.0150	0.0000	0.0798
Texas	0.0119	0.0167	0.0001	0.0947
Utah	0.0119	0.0165	0.0000	0.1009
Vermont	0.0119	0.0146	0.0000	0.0745
Virginia	0.0119	0.0176	0.0001	0.1096
Washington	0.0119	0.0151	0.0001	0.0808
West Virginia	0.0119	0.0162	0.0000	0.0843
Wisconsin	0.0119	0.0136	0.0000	0.0752
Wyoming	0.0119	0.0160	0.0000	0.0938
United States	0.0119	0.0154	0.0002	0.0771

Source: Bureau of the Census (2007).