SHANNON MEASURE OF INDUSTRIAL SPECIALIZATION AT THE STATE LEVEL

Edward Nissan, University of Southern Mississippi
George Carter, University of Southern Mississippi

ABSTRACT

The focus of this research is to use the Shannon (Theil) index and the coefficient of variation in addressing the question of employment diversity at the state level for 2002. The Shannon index was used to place the states into a hierarchy of specialization from the most specialized to the least specialized. The District of Columbia was most specialized while Kentucky was the least specialized state. The coefficient of variation was also used to compare every state with every other state. Again, the District of Columbia was most specialized while Wisconsin was the most diversified state. *JEL classifications: L60, L70, L80, L90*

INTRODUCTION

A good argument can be made for investigating industrial specialization at the state or regional levels because state policies make a difference. According to Blackley (1994), 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. 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 and by implication at the state level.

Wundt (1992) remarked that policymakers benefit from his study for the state of Connecticut 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 by encompassing the 51 states (including DC) to assess the industrial employment diversity between 1972 and 1981. In both studies, the entropy index for measuring diversity (Shannon and Weaver, 1949) was used as the measure of diversification at the 1-digit Standard Industrial Classification (SIC) level. It should be pointed out that diversification is the opposite side of the coin of specialization. Specialization here has the same meaning as business concentration.

Further good arguments for investigating industrial specialization at the state level are supplied by Weingast (1995, 2006). The gist of his idea is that fiscal federalism can explain growth and that healthy competition between state governments is good for growth. In particular, the author's concern is a specific form of federalism, called *market preserving federalism*, with the meaning that the degree of the political system of a country or a state to encroach upon their markets must be limited. Self-enforcing restrictions are required. The economic effects of federalism can be summarized by two public policy choices. The first is political competition whereby a jurisdiction must compete for capital, labor and economic activity through levels of taxation, security of private rights, social amenities, and public goods. The second refers to competition. In a situation that a jurisdiction cartelizes an industry, labor, because of mobility, relocates to other jurisdictions. Similarly in case a jurisdiction attempts to confiscate the wealth of an industry, firms because of their mobility will relocate. In summary, beneficial consequences of federalism results from political decentralization of economic authority unlike when the political authority is at the discretion of the central government with the same consequences as if federalism does not exist. Weingast (1995) provides examples of countries with successful federalism for the last 300 years, "the Netherlands from the late 16^{th} through mid-17^{th} Century, England from the late 17^{th} or early 18^{th} through the mid-19^{th} Century, and the United States from the late 19^{th} Century until the late 20^{th} Century." Similarly, according to Weingast (2009), "China being a *de facto* federal state, has experienced sustained rapid growth. India, having grown slowly for several decades, has experienced high growth in the last decade. In contrast, the large Latin American federal states of Argentina, Brazil and Mexico, have all fared more poorly." How do we account for the differences? The answer is that the institutions that support decentralization give support to the improvement of social welfare.

Ke and Malizia (1993) define diversity as the variety of economic activity of a region. An area with diverse industries should experience stable economic growth and less unemployment than specialized areas. A diverse economy could have many different industries that fluctuate in severity and timing. St. Louis (1980) defines industrial diversity as the presence of a wide variety of industries insulating a region from business cycle swings. In contrast, specialized regions (e.g., Detroit with the automobile industry) that may be subject to boom and bust may experience high unemployment, income instability and migration. Jacobs (1970) believed that firms cluster due to the effects of specialization and diversity to take advantage of agglomeration in geographic proximities. Lim (2004) explains that specialization is locally bounded in metropolitan areas where people are concentrated in small geographic spaces where knowledge can then be transmitted easily among them.

At the single country level, Massell (1964) tackled the issue of fluctuations in exports. Concentration on a narrow range of products for export can explain the fluctuations such as one crop economies. Examples are Ghana with cocoa crops, Mauritius with sugar crops and Sudan with cotton crops. The argument goes that if such economies diversify their exports, earnings would exhibit better stability over time. Feldman and Audretsch (1999) test whether specialization within a narrow set of economic activity is more conducive to spillover of knowledge in comparison with diversity. They support a conclusion that diversity is more conducive to knowledge spillover and innovation than specialization.

This paper, on a similar track as Attaran and Zwick (1987b) at the state level, addresses the question of employment diversity for 2002. While the Attaran and Zwick work on diversification was at the 1-digit Standard Industrial Classification (SIC) system, this paper uses the newer 3-digit North American Industry Classification System (NAICS) for 2002, the latest year for which data was available. The indexes of diversity (to be discussed later) are the Shannon index, known also as Theil's index, and the coefficient of variation. The paper by using some statistical techniques not employed by other researchers will contribute further to the topic of state employment diversity by investigating diversity between major sectors and subsectors by the use of analysis of variance techniques and by comparing each state with other states using F-tests. To the best knowledge of the authors, no such procedures were employed in previous research.

THE DATA

The 3-digit NAICS (North American Industry Classification System) for 2002 obtained from the Bureau of the Census (2007) was employed for use in calculating the states' employment diversity indexes for 84 sectors. The codes of sectors at the 2-digit NAICS for 2002 under consideration are:

However, in order to compare between major sectors and their subsectors in this research, a newer classification is made by aggregating the 84 sectors into major sectors and subsectors in accordance with first-digit, 2-digit and 3-digit codes. Note that the first-digit classification is primarily made for this research because NAICS does not publish classification by 1-digit codes. The first-digit classification is made purely for this research as a convenience to make comparisons between sectors and within subsectors. The classifications accordingly are shown below:

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).

NAICS 2-DIGIT SECTORS

2-Digit Code	Sector
21	Mining
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific and Technical Services
55	Management of Companies and Enterprises
56	Adm & Support & Waste Management &Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)

NAICS Codes	Number of Sectors Three-Digit	Two-Digit	First-Digit
211-238	7	3	1
311-339	21	3	1
423-493	24	5	1
511-562	19	6	1
611-624	5	2	1
711-722	5	2	1
811-814	3	1	1
Total	84	22	7

METHODOLOGY

One of the most common indexes for specialization is the Shannon (entropy) index according to Siegel, Johnson and Alwang (1995) and Pielou (1975). A latest application is by Cunha and Heckman (2007) who used it in their study of the evolution of earnings inequality in the United States. A second measure according to Hannah and Kay (1977) is the coefficient of variation (CV).

Let p_i be the proportionate (relative frequency) share received by industrial sector i among n well-defined sectors, $\sum p_i = 1$. The Shannon-Weaver (1949) index, also labeled as entropy, is given by

$$E = -\sum p_i \log p_i, \qquad 1 \leq i \leq n, \tag{1}$$

where E ranges between "0" when all shares are contained in one sector and $\log n$ when the proportionate shares p_i are equal. Note that the Shannon-Weaver index is sensitive to sectors with smaller employment as compared with the coefficient of variation.

Jacquemin and Berry (1979) as well as Pielou (1975) give preference to Shannon because the index can be decomposed into additive elements depicting "between" and "within" variations among the sectors. Accordingly, the Shannon index of equation (1) can be expressed additively as a weighted average of diversification within sectors as for example diversification at the 3-digit level and the first-digit level.

Patil and Taillie (1982) suggest the use of analysis of variance (ANOVA) to apportion the total diversity "between" and "within" the various sectors rather than the additive scheme available with the Shannon index. For instance, as shown in the Data section and in accordance with ANOVA, there are seven first-digit codes corresponding to the 84 3-digit codes. Thus, for each state there would be seven "between' entities compared to 84 "within" entities. Similarly, when the broad classification is at the 2-digit level, there would be 22 "between" entities compared to 84 "within" entities. Lande (1996) and Keylock (2005) are on board with Patil and Taillie (1982) in the use of ANOVA for partitioning the total diversity into "between" and "within" groups. The ANOVA procedure provides a way to test statistical significance for the "between" and "within" classifications by the use of the F-test

$$F^* = \frac{MSB}{MSW} \tag{2}$$

where the "between" mean square (MSB) and the "within" mean square (MSW) are respectively obtained by dividing the sum of squares for the "between" groups (SSB) and the "within" groups sum of squares (SSW) by the appropriate degrees of freedom.

Another way to portray the results of E in equation (1) is through normalized "numbers-equivalent," which is the number of sectors m yielding E if the number of employees in each sector is of equal size. According to Miller (1972), given an E and n number of sectors, m provides the number of equalized sectors to generate a level of E.

For the E index, when all sectors have equal employment shares,

$$m = \text{antilogarithm}(E). \tag{3}$$

An interesting modification of the Shannon index is made in Campiglio and Caruso (2007). Their concern was the richness and diversity of economic literature viewed through JEL codes for the top-ten generalist journals for the period 2000-2006 to establish a level of diversity among them. There is an apparent analogy between their research and this one. They deal with 10 journals and JEL codes; this research deals with 51 states (with DC) and NAICS codes. One of their indexes for diversity was the Shannon as reported here in equation (1). However, they propose an alternative way to express the index as a standardized measure by dividing by its maximum value, log n. For the Shannon, this alternative for equation (1) is the index

$$E_s = \frac{E}{\log n} = \frac{-\sum p_i \log p_i}{\log n}. \tag{4}$$

Note that as E_s approaches unity, the more is the diversity complete.

The second index to be used is the coefficient of variation (CV) obtained by dividing the standard deviation by the mean. There is an advantage for the coefficient of variation because of its apparent relationship to variances. To show this relationship, knowing that $\sum p_i = 1$, the mean $\overline{p} = 1/n$.

Now,

$$CV = \frac{S}{\overline{p}} \tag{5}$$

$$(CV)^2 = \left(\frac{S}{\overline{p}}\right)^2$$

When comparing two states for their coefficients of variation, say states 1 and 2, and since the number of NAICS codes for each state is the same, $n = 84,$ then $\overline{p}_1 = \overline{p}_2 = \frac{1}{n} = \frac{1}{84}$. Therefore,

$$\frac{(CV_1)^2}{(CV_2)^2} = \left(\frac{S_1}{\bar{p}_1}\right)^2 \bigg/ \left(\frac{S_2}{\bar{p}_2}\right)^2 . \quad (6)$$

Since $\bar{p}_1 = \bar{p}_2 = \frac{1}{n} = \frac{1}{84}$, cancellation yields the result:

$$\frac{(CV_1)^2}{(CV_2)^2} = \frac{S_1^2}{S_2^2} \quad (7)$$

Equation (7) is a handy expression for testing the equality of two coefficient of variation (CV) indexes by use of the test statistic

$$F^* = \frac{S_1^{\ 2}}{S_2^{\ 2}} \quad (8)$$

and comparing it with the tabular F-distribution, $F(\alpha, n-1, n-1)$, for significance level α. For $\alpha = .05$ and n=84, F(.05, 83, 83) = 1.46. Equation (8), therefore, can be used as a way to compare diversity for two states.

RESULTS

Table 1 displays the 84 employment sectors at the 3-digit NAICS for the United States, giving employment percentages. Among the largest employment sectors are 238 (specialty trade contractors) at 4.02 percent; 541 (professional, scientific and technical services) at 6.64 percent; 561 (administrative and support services) at 7.71 percent; 621 (ambulatory health care services) at 4.52 percent, 622 (hospitals) at 4.74 percent; and 722 (food services and drinking places) at 7.62 percent. For convenience, these sectors are highlighted in Table 1, all of which have employment percentages higher than four percent.

TABLE 1
PERCENTAGE OF PAID EMPLOYEES AT 3-DIGIT NAICX FOR THE UNITED STATES

CODE	DESCRIPTION	PERCENT
211	Oil and gas extraction	0.09
212	Mining, except oil and gas	0.18
213	Support activities for mining	0.17
221	Utilities	0.61
236	Construction of buildings	1.53
237	Heavy and civil engineering construction	1.05
238	Specialty trade contractors	4.02
311	Food mfg	1.38
312	Beverage & tobacco product mfg	0.15
313	Textile mills	0.25

314	Textile product mills	0.17
315	Apparel mfg	0.31
316	Leather & allied product mfg	0.04
321	Wood product mfg	0.50
322	Paper mfg	0.45
323	Printing & related support activities	0.66
324	Petroleum & coal products mfg	0.09
325	Chemical mfg	0.78
326	Plastics & rubber products mfg	0.90
327	Nonmetallic mineral product mfg	0.44
331	Primary metal mfg	0.45
332	Fabricated metal product mfg	1.44
333	Machinery mfg	1.08
334	Computer & electronic product mfg	1.16
335	Electrical equipment, appliance, & component mfg	0.45
336	Transportation equipment mfg	1.54
337	Furniture & related product mfg	0.55
339	Miscellaneous mfg	0.69
423	Durable goods merchant wholesalers	3.09
424	Nondurable goods merchant wholesalers	2.11
425	Wholesale electronic markets and agents and brokers	0.23
441	Motor vehicle & parts dealers	1.69
442	Furniture & home furnishings stores	0.49
443	Electronics & appliance stores	0.36
444	Building material & garden equipment & supplies deal.	1.06
445	Food & beverage stores	2.60
446	Health & personal care stores	0.94
447	Gasoline stations	0.85
448	Clothing & clothing accessories stores	1.31
451	Sporting goods, hobby, book, & music stores	0.56
452	General merchandise stores	2.30
453	Miscellaneous store retailers	0.73
454	Nonstore retailers	0.52
481	Air transportation	0.09
483	Water transportation	0.06
484	Truck transportation	1.31
485	Transit & ground passenger transportation	0.37
486	Pipeline transportation	0.04
487	Scenic & sightseeing transportation	0.02
488	Support activities for transportation	0.43
492	Couriers & messengers	0.52
493	Warehousing & storage	0.53
511	Publishing industries (except Internet)	1.00
512	Motion picture & sound recording industries	0.28

515	Broadcasting (except Internet)	0.27
516	Internet publishing & broadcasting	0.04
517	Telecommunications	1.32
518	Internet service providers, web search portals, & data processing	0.47
519	Other information services	0.05
521	Monetary authorities - central bank	0.02
522	Credit intermediation & related activities	3.05
523	Securities intermediation & related activities	0.71
524	Insurance carriers & related activities	2.20
525	Funds, trusts, & other financial vehicles (part)	0.02
531	Real estate	1.21
532	Rental & leasing services	0.57
533	Lessors of nonfinancial intangible assets (exc copyrighted works)	0.02
541	Professional, scientific, & technical services	6.64
551	Management of companies & enterprises	2.39
561	Administrative & support services	7.71
562	Waste management & remediation services	0.30
611	Educational services	0.39
621	Ambulatory health care services	4.52
622	Hospitals	4.74
623	Nursing & residential care facilities	2.60
624	Social assistance	1.95
711	Performing arts, spectator sports, & related industries	0.39
712	Museums, historical sites, & similar institutions	0.11
713	Amusement, gambling, & recreation industries	1.21
721	Accommodation	1.66
722	Food services & drinking places	7.62
811	Repair & maintenance	1.18
812	Personal & laundry services	1.19
813	Religious/grantmaking/civic/professional & similar org	0.82

Source: Bureau of the Census (2007).

Table 2 presents the results of the Shannon index E, equation (1), and its alternative measures m, the numbers-equivalent by equation (3) and the normalization procedure E_s outlined in equation (4), shown in columns 1-3. The numbers-equivalent m indicates the number of sectors resulting if all sectors have equal proportion of employment. For Alabama, for instance, the 84 sectors will be reduced to 46.09 sectors if all sectors have equal shares of employment. For Nevada, the 84 sectors are dwindled to m = 28.42. Hence, the more employment is concentrated in fewer sectors, the smaller is the value of m. Note that the smaller the E_s, the more the state is specialized. Also shown in Table 2 is the coefficient of variation (CV) measure represented simply by the standard deviation (S) as a convenient way of portrayal in column 4. As explained in equation (5), CV = $\frac{S}{\bar{p}}$ where

$\bar{p}$ = 1/n = 1/84 for all the states rendering the results to be proportional. For the CV index, the higher the value the more the state is specialized. The ranking of the states by the two measures are shown in the next two columns (columns 5-6). Alabama was ranked 44 by E_s and 38 by CV. Arizona was ranked 11 by E_s and 10 by CV.

The slight difference in the rankings is due, as explained earlier, to the nature of the two indexes. E_s is more sensitive to sectors with low employment as compared to CV. The correlation between the two indexes is very high. The Pearson correlation between E_s and S is -0.974. The negative sign is due to the fact that the two indexes operate in different directions for evaluating employment diversity. On a similar track, the Spearman coefficient of rank correlation is also high at 0.912. Furthermore, to solicit the relationship between the two indexes, a linear regression is performed whereby E_s is a function of S. The result is

$$E_s = 1.017 - 10.379S \quad (9)$$

with a p-value = 0.000 for both coefficients. The residuals are shown in column 7, and the standardized residuals are displayed in column 8. The highest values of the residuals are for DC at -2.6943 and Nevada at 2.6915, somewhat outliers as compared to the rest of the states.

Finally in Table 2 (column 9), there is a display of the percentage of the largest sector in each state. Total employment in the United States in the 84 sectors in 2002 was approximately 109 million. States with large specialized sectors exceeding 10 percent are Colorado (12.72%), District of Columbia (22.24%), Florida (15.86%), Hawaii (12.12%), Montana (10.69%), Nevada (21.54%), New Mexico (10.14%), Utah (10.09%) and Virginia (10.96%). The corresponding standard deviations give a hint for the level of concentration among sectors by noting for instance that DC is 0.0300 and Nevada is 0.0270. Those are the two states with the most specialized sectors.

TABLE 2

SHANNON INDEXES AND COEFFICIENTS OF VARIATION FOR STATES AT 3-DIGIT NAICS

Alphabetical Order	Shannon Index Measures			CV Index	Rank			Standard-ized	Sector
State	E	m	Es	S	E_S	S	Residuals	Residuals	Max%
Alabama	1.6636	46.09	0.8646	0.0151	44	38	0.00431	0.6226	7.61
Alaska	1.6111	40.84	0.8373	0.0161	10	24	-0.01261	-1.8231	8.22
Arizona	1.6161	41.31	0.8398	0.0172	11	10	0.00130	0.1885	9.90
Arkansas	1.6718	46.97	0.8688	0.0143	47	49	0.00020	0.0295	7.26
California	1.6465	44.31	0.8557	0.0165	30	16	0.00994	1.4365	9.24
Colorado	1.5869	38.63	0.8247	0.0194	5	5	0.00904	1.3064	12.72
Connecticut	1.6402	43.68	0.8524	0.0155	24	32	-0.00374	-0.5407	7.11
Delaware	1.5882	38.74	0.8253	0.0178	6	7	-0.00697	-1.0072	9.84
D. of Columbia	1.3220	20.99	0.6870	0.0300	1	1	-0.01864	-2.6943	22.24
Florida	1.5540	35.81	0.8076	0.0213	4	3	0.01166	1.6852	15.86
Georgia	1.6545	45.14	0.8598	0.0158	35	30	0.00677	0.9790	8.81
Hawaii	1.5446	35.04	0.8027	0.0196	3	4	-0.01089	-1.5734	12.12
Idaho	1.5995	39.77	0.8312	0.0169	7	11	-0.01041	-1.5046	8.19

Illinois	1.6533	45.01	0.8592	0.0155	33	31	0.00306	0.4422	7.70
Indiana	1.6633	46.05	0.8644	0.0148	43	43	0.00099	0.1437	8.10
Iowa	1.6619	45.91	0.8636	0.0141	41	50	-0.00707	-1.0221	7.19
Kansas	1.6533	45.01	0.8592	0.0149	34	42	-0.00317	-0.4579	7.64
Kentucky	1.6818	48.06	0.8740	0.0145	51	47	0.00748	1.0812	8.54
Louisiana	1.6411	43.77	0.8529	0.0159	25	28	0.00091	0.1317	8.66
Maine	1.6420	43.86	0.8533	0.0153	28	33	-0.00492	-0.7106	7.45
Maryland	1.5998	39.79	0.8314	0.0179	8	6	0.00017	0.0245	9.88
Massachusetts	1.6326	42.92	0.8484	0.0163	22	19	0.00056	0.0814	8.29
Michigan	1.6197	41.66	0.8417	0.0166	14	14	-0.00302	-0.4370	7.72
Minnesota	1.6646	46.20	0.8651	0.0145	45	46	-0.00142	-0.2052	7.22
Mississippi	1.6507	44.74	0.8578	0.0151	32	37	-0.00249	-0.3602	7.82
Missouri	1.6619	45.91	0.8636	0.0149	42	41	0.00123	0.1781	8.03
Montana	1.6175	41.45	0.8406	0.0168	13	12	-0.00205	-0.2959	10.69
Nebraska	1.6206	41.75	0.8422	0.0162	15	22	-0.00668	-0.9648	8.10
Nevada	1.4536	28.42	0.7554	0.0270	2	2	0.01862	2.6915	21.54
New Hampshire	1.6571	45.40	0.8611	0.0149	39	40	-0.00127	-0.1833	7.44
New Jersey	1.6313	42.79	0.8478	0.0164	21	18	0.00100	0.1447	8.72
New Mexico	1.6043	40.21	0.8337	0.0172	9	9	-0.00480	-0.6932	10.14
New York	1.6301	42.67	0.8471	0.0164	19	17	0.00030	0.0435	7.95
North Carolina	1.6771	47.55	0.8716	0.0143	49	48	0.00300	0.4342	7.83
North Dakota	1.6293	42.59	0.8467	0.0158	18	29	-0.00633	-0.9145	8.14
Ohio	1.6557	45.26	0.8604	0.0152	37	34	0.00115	0.1656	8.31
Oklahoma	1.6411	43.76	0.8529	0.0161	26	23	0.00299	0.4317	8.85
Oregon	1.6562	45.31	0.8607	0.0151	38	36	0.00041	0.0589	8.46
Pennsylvania	1.6749	47.30	0.8704	0.0147	48	44	0.00596	0.8609	6.97
Rhode Island	1.6288	42.54	0.8464	0.0162	17	21	-0.00248	-0.3577	8.82
South Carolina	1.6477	44.43	0.8563	0.0159	31	27	0.00431	0.6231	9.17
South Dakota	1.6260	42.26	0.8450	0.0160	16	26	-0.00595	-0.8601	8.31
Tennessee	1.6676	46.51	0.8666	0.0150	46	39	0.00527	0.7617	7.98
Texas	1.6426	43.91	0.8536	0.0167	29	13	0.00991	1.4330	9.47
Utah	1.6414	43.79	0.8530	0.0165	27	15	0.00724	1.0463	10.09
Vermont	1.6602	45.73	0.8628	0.0146	40	45	-0.00268	-0.3876	7.45
Virginia	1.6304	42.70	0.8473	0.0176	20	8	0.01296	1.8726	10.96
Washington	1.6549	45.17	0.8600	0.0151	36	35	-0.00029	-0.0422	8.08
West Virginia	1.6331	42.96	0.8487	0.0162	23	20	-0.00018	-0.0253	8.43
Wisconsin	1.6810	47.97	0.8736	0.0136	50	51	-0.00226	-0.3268	7.52
Wyoming	1.6173	41.43	0.8405	0.0160	12	25	-0.01045	-1.5106	9.38
United States	1.6662	46.37	0.8659	0.0154	NA	NA	NA	NA	7.71

Note: Shannon Index E (equation 1), m (equation 3), E_s (equation 4), CV (equation 5) and Residuals (equation 9). Source: Bureau of the Census (2007).

Table 3 displays states in ascending order of ranking from the least diverse (most specialized) to the most diverse (least specialized) as portrayed by E_s where D.C. occupies the first slot as being the most specialized and Kentucky is the last slot as the most diversified. Also, states were displayed in a similar manner by the S index where D.C. again occupies the first slot and Wisconsin the last. As was done by Attaran and Zwick (1987b), this paper classifies the states by their levels of E_s and S from the most specialized to the most diversified. This is done by dividing the states in accordance with their positions in the four quartiles of the E_s and S distributions in Table 3. Accordingly, the boundaries for the E_s and S quartiles are

Highly specialized:	$0.6870 \leq E_s \leq 0.8406$	$0.0167 \leq S \leq 0.0300$
Moderately specialized:	$0.8417 < E_s \leq 0.8530$	$0.0161 \leq S \leq 0.0166$
Moderately diversified:	$0.8533 < E_s \leq 0.8628$	$0.0151 \leq S \leq 0.0160$
Highly diversified:	$0.8636 < E_s \leq 0.8740$	$0.0136 \leq S \leq 0.0150$

The lines in Table 3 identify the states belonging to each of the four quartiles. For the E_s index, there are 13 states starting with DC and ending in Montana considered accordingly to be highly specialized. There are 14 states starting with Michigan and ending with Utah that are moderately specialized. There are 13 states in the moderately diversified standing starting with Maine and ending with Vermont, and finally, there are 11 states starting with Missouri and ending with Kentucky as being in the highly diversified category. By the S index, there were slightly different permutations for the states in the four categories of classifications.

TABLE 3
STATE RANKS BY SHANNON AND CV

State	Es	State	S
District of Columbia	0.6870	District of Columbia	0.0300
Nevada	0.7554	Nevada	0.0270
Hawaii	0.8027	Florida	0.0213
Florida	0.8076	Hawaii	0.0196
Colorado	0.8247	Colorado	0.0194
Delaware	0.8253	Maryland	0.0179
Idaho	0.8312	Delaware	0.0178
Maryland	0.8314	Virginia	0.0176
New Mexico	0.8337	Arizona	0.0172
Alaska	0.8373	New Mexico	0.0172
Arizona	0.8398	Idaho	0.0169
Wyoming	0.8405	Montana	0.0168
Montana	0.8406	Texas	0.0167
Michigan	0.8417	Michigan	0.0166
Nebraska	0.8422	California	0.0165
South Dakota	0.8450	Utah	0.0165
Rhode Island	0.8464	New Jersey	0.0164
North Dakota	0.8467	New York	0.0164
New York	0.8471	Massachusetts	0.0163
Virginia	0.8473	Nebraska	0.0162
New Jersey	0.8478	Rhode Island	0.0162
Massachusetts	0.8484	West Virginia	0.0162

West Virginia	0.8487	Alaska	0.0161
Connecticut	0.8524	Oklahoma	0.0161
Oklahoma	0.8529	South Dakota	0.0160
Louisiana	0.8529	Wyoming	0.0160
Utah	0.8530	Louisiana	0.0159
Maine	0.8533	South Carolina	0.0159
Texas	0.8536	Georgia	0.0158
California	0.8557	North Dakota	0.0158
South Carolina	0.8563	Connecticut	0.0155
Mississippi	0.8578	Illinois	0.0155
Illinois	0.8592	Maine	0.0153
Kansas	0.8592	Ohio	0.0152
Georgia	0.8598	Alabama	0.0151
Washington	0.8600	Mississippi	0.0151
Ohio	0.8604	Oregon	0.0151
Oregon	0.8607	Washington	0.0151
New Hampshire	0.8611	Tennessee	0.0150
Vermont	0.8628	Kansas	0.0149
Missouri	0.8636	Missouri	0.0149
Iowa	0.8636	New Hampshire	0.0149
Indiana	0.8644	Indiana	0.0148
Alabama	0.8646	Pennsylvania	0.0147
Minnesota	0.8651	Vermont	0.0146
Tennessee	0.8666	Kentucky	0.0145
Arkansas	0.8688	Minnesota	0.0145
Pennsylvania	0.8704	Arkansas	0.0143
North Carolina	0.8716	North Carolina	0.0143
Kentucky	0.8740	Wisconsin	0.0136

Note: The lines in the table for E_s and S group states into four quartiles: highly specialized, moderately specialized, moderately diversified and highly diversified. See text for the classification of states. Source: Bureau of the Census (2007)

Table 4 provides the disaggregation of the total sum of squares into “Between” portion and “Within” portion by using the ANOVA scheme outlined in equation (2). The aggregation is done at the first-digit and 2-digit NAICS. Twenty-six of the states had F-values significant (five percent level of significance) at the first-digit level, while all states had significant F-values when the disaggregation is done at the 2-digit level. A plausible explanation is that aggregation at the first-digit level contains a larger number of subsectors than is the case at the 2-digit level. The interest in this exercise is to find out whether diversity between the sectors as aggregates exceeds those within the subsectors that are members in the first-digit or 2-digit classifications. The implication here, especially at the 2-digit level, is that employments categorized by the broader classifications exceed those within the classifications. At the first-digit level, there were seven major sectors encompassing 84 subsectors. At the 2-digit, there were 22 sectors encompassing the 84 sectors.

TABLE 4
F-TESTS FOR DISAGGREGATING (FIRST-DIGIT AND 2-DIGIT) INTO BETWEEN AND WITHIN SUMS OF SQUARES

	First-Digit		2-Digit	
	F	P-Value	F	P-Value
Alabama	1.47	0.198	3.04	0.000
Alaska	4.79	0.000	4.96	0.000
Arizona	1.95	0.083	3.32	0.000
Arkansas	1.66	0.143	1.95	0.022
California	1.66	0.143	5.11	0.000
Colorado	1.88	0.095	4.06	0.000
Connecticut	2.62	0.023	4.76	0.000
Delaware	1.92	0.088	2.50	0.003
District of Colum.	2.97	0.012	12.76	0.000
Florida	1.80	0.111	3.49	0.000
Georgia	1.38	0.233	3.22	0.000
Hawaii	5.36	0.000	10.60	0.000
Idaho	1.63	0.151	3.80	0.000
Illinois	1.73	0.125	4.01	0.000
Indiana	1.56	0.170	2.96	0.000
Iowa	2.29	0.044	2.57	0.002
Kansas	2.20	0.052	2.94	0.001
Kentucky	2.32	0.041	3.05	0.000
Louisiana	3.53	0.004	4.71	0.000
Maine	4.67	0.000	5.63	0.000
Maryland	2.27	0.045	5.55	0.000
Massachusetts	3.02	0.011	5.49	0.000
Michigan	1.29	0.274	3.47	0.000
Minnesota	2.67	0.021	4.95	0.000
Mississippi	2.80	0.016	4.35	0.000
Missouri	2.41	0.034	3.93	0.000
Montana	5.39	0.000	6.47	0.000
Nebraska	1.79	0.111	2.58	0.002
Nevada	4.55	0.001	7.46	0.000
New Hampshire	1.93	0.086	3.76	0.000
New Jersey	1.90	0.092	5.10	0.000
New Mexico	3.75	0.003	4.56	0.000
New York	3.94	0.002	5.16	0.000
North Carolina	1.63	0.151	3.31	0.000
North Dakota	4.77	0.000	4.98	0.000
Ohio	1.92	0.088	3.52	0.000
Oklahoma	2.39	0.036	3.35	0.000
Oregon	1.92	0.089	3.89	0.000
Pennsylvania	2.94	0.012	4.81	0.000
Rhode Island	3.44	0.000	4.25	0.000
South Carolina	1.23	0.299	3.04	0.000
South Dakota	3.91	0.002	4.03	0.000
Tennessee	1.51	0.186	3.01	0.000
Texas	1.80	0.110	3.56	0.000
Utah	1.55	0.174	3.12	0.000
Vermont	3.98	0.002	6.18	0.000
Virginia	1.63	0.149	6.30	0.000
Washington	2.55	0.026	4.40	0.000
West Virginia	5.19	0.000	4.45	0.000

Wisconsin	2.15	0.057	3.26	0.000
Wyoming	8.24	0.000	8.51	0.000
United States	2.62	0.023	4.75	0.000

Note: Disaggregation of Total SS into "Between" and "Within" portions by equation (2). Source: Bureau of the Census (2007).

Table 5 presents the results of F-tests for equality of two variances and, by implication, for equality of their coefficients of variation for specialization as outlined in equations (5) through (8). Here, the states were lined up across the top of the table by the descending magnitudes of their variances obtained as the squares of the standard deviations S (column 4 in Table 2). Upon dividing the variance of each state by the respective variances of the other states, the ratios provide an assessment of that state's specialization as compared to the others in accordance with CV as a measure of specialization. An F-ratio greater than or equal to 1.46 indicates significance at the five percent level.

From Table 5, D.C. is more specialized than every state with the exception of Nevada. Nevada is more specialized than the others. Florida is less specialized than D.C. and Nevada; does not differ in its specialization from Hawaii, Colorado, Maryland and Delaware; and is more specialized than the other states. The list goes on for the other states, culminating in Wisconsin, where it is the least specialized state in accordance with the CV index. The table, therefore, provides hierarchies of employment specialization.

TABLE 5
STATES FOR WHICH THERE IS A SIGNIFICANT DIFFERENCE IN SPECIALIZATION

More Specialized Across Top of Table (Less Specialized in Table Body)																
DC	NV	FL	HI	CO	MD	DE	VA	AZ	NM	ID	MT	TX	MI	CA	UT	NY
FL	FL															
HI	HI															
CO	CO															
MD	MD															
DE	DE															
VA	VA	VA														
AZ	AZ	AZ														
NM	NM	NM														
ID	ID	ID														
MT	MT	MT														
TX	TX	TX														
MI	MI	MI														
CA	CA	CA														
UT	UT	UT														
NY	NY	NY														
NJ	NJ	NJ														
MA	MA	MA														
RI	RI	RI	RI													
NE	NE	NE	NE													
WV	WV	WV	WV													
OK	OK	OK	OK	OK												
AK	AK	AK	AK	AK												
SD	SD	SD	SD	SD												
WY	WY	WY	WY	WY												
LA	LA	LA	LA	LA												
SC	SC	SC	SC	SC												
ND	ND	ND	ND	ND												

GA	GA	GA	GA	GA												
CT	CT	CT	CT	CT												
IL	IL	IL	IL	IL												
ME	ME	ME	ME	ME												
OH	OH	OH	OH	OH												
OR	OR	OR	OR	OR												
WA	WA	WA	WA	WA												
AL	AL	AL	AL	AL												
MS	MS	MS	MS	MS												
TN	TN	TN	TN	TN												
NH	NH	NH	NH	NH												
MO	MO	MO	MO	MO												
KS	KS	KS	KS	KS												
IN	IN	IN	IN	IN	IN											
PA	PA	PA	PA	PA	PA	PA										
VT	VT	VT	VT	VT	VT	VT	VT									
MN	MN	MN	MN	MN	MN	MN	MN									
KY	KY	KY	KY	KY	KY	KY	KY									
NC	NC	NC	NC	NC	NC	NC	NC									
AR	AR	AR	AR	AR	AR	AR	AR	AR								
IA	IA	IA	IA	IA	IA	IA	IA	IA	IA							
WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI	WI

Note: Calculations by equation (8), Significance determined by $F^* \geq 1.46$.
Source: Bureau of the Census (2007).

CONCLUDING REMARKS

This is a paper using a different statistical approach than the one used by Mulligan and Schmidt (2005) which addressed the question of specialization at the nine regional levels. Mulligan and Schmidt incorporated the methodology of the location quotient (LQ). The greater the number of industries the greater the range of the LQs. For their data they used the Annual Surveys of Manufacturers value-added for each of the 20 two-digit standard industry codes (SIC). Their concern similar to this study was to assess the economies of the regions to be more or less balanced in terms of their industry mix.

The LQ is a measure of the share of occupations in a regional economy compared to their share in the national economy. When LQ = 1.00, the implication is that a region is self-sufficient in the particular occupation as compared to the nation. When LQ < 1, the implication is that the region is less than self-sufficient, and LQ > 1 implies that the region is more than self-sufficient. This type of approach differs from the one pursued in this paper in that it compares the specialization of employment in each sector to that of the nation. When LQ > 1 for a given sector in a region, it implies that the region is more specialized in that given sector than the nation. If the majority of the sectors in a region have LQ < 1, the implication is a diversified employment contrary to when few sectors record LQ > 1 with the implication for specialization. The approach used in this paper, by calculating two indexes of diversity, is to provide an overall picture of employment diversity for each state and provides a means of comparisons between states.

This paper examined the industrial mix at the state level using the 3-digit NAICS for employment as a guide. The two indexes of diversity that are employed, the Shannon and the coefficient of variation, are the best indexes for this purpose. The Shannon index, E_s, and the coefficient of variation, CV, were used to group states in accordance with their levels of E_s and S into four quartiles from the most specialized to the most diversified. Furthermore, the coefficient of variation, because of its statistical comparability with the variance as shown by equations (5) through (7), was in turn used to test equality of variances by comparing each state with each other state. The results provide a detailed survey in comparing states for employment diversity.

Conroy (1975) and Brewer and Moomaw (1985) are in agreement that specialization can create economic well-being and economic growth. They also find advantages in diversification as opposed to concentration. Diversity is preferred for its stabilization of employment and incomes and reducing reliance on exportation of concentrated products. Two states compared for specialization that stand at the two extremes in accordance to the E_s index are Nevada and Kentucky. Nevada depends a great deal on tourism and recreation while Kentucky's employment is highly diversified. In a down-turn in the business cycle, tourism and recreation industries experience a down-turn also. This point was clearly demonstrated by Berzon (2009). Berzon explains that strategies of casino companies are shifting toward slower growth. New building projects are halted for at least ten years. This is happening because of the collapsing economy in the last part of this decade. Nevada can suffer a great deal in terms of employment and incomes while the impact on Kentucky might be less severe. On the other hand, Nevada's economy because of specialization can contribute to its economic growth much faster than Kentucky. This happened in the last decade in Nevada where some $30 billion was invested in casinos and hotels.

REFERENCES

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.

Berzon, A., 2009. "Vegas Casinos Fold on Expansion Plans." *The Wall Street Journal* (October 5): A1, A4.

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

Brewer, H.L. and R.L. Moomaw. 1985. "A Note on Population Size, Industrial Diversification, and Regional Economic Instability." *Urban Studies* 22: 349-354.

Bureau of the Census. 2007. "Three-digit North American Industry Classification System." http://www.census.gov/econ/census02/guide/geosumm.htm.

Campiglio, L., and R. Caruso, 2007. "Where Economics Has Been Headed? Multiple Identities and Diversity in Economic Literature—Evidence from Top Journals over the Period 2000-2006, A First Note."
SSRN: http://ssrn.com/abstract=1008133.

Conroy, M.E., 1975. "Alternative Strategies for Regional Industrial Diversification." *Journal of Regional Science* 14: 31-46.

Cunha, F. and J.J. Heckman. 2007. "The Evolution of Inequality, Heterogeneity and Uncertainty in Labor Earnings in the U.S. Economy." NBER Working Paper W13526. http://ssrn.com/abstract=1024130.

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

Hannah, L. and J.A. Kay. 1977. *Concentration in Modern Industry: Theory, Measurement and the U.K. Experience* The Macmillan Press Ltd: London.

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*. Vintage Books: New York.

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

Keylock, C.J. 2005. "Simpson Diversity and the Shannon-Weaver Index as Special Cases of Generalized Entropy." *Oikos* 109: 203-205.

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

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

Massell, B.F. 1964. "Export Concentration and Fluctuations in Export Earnings: A Cross-Section Analysis." *The American Economic Review* 54: 47-63.

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

Mulligan, G.F. and C. Schmidt. 2005. "A Note on Localization and Specialization." *Growth and Change* 36: 565-576.

Patil, G.P., and C. Taillie. 1982. "Diversity as a Concept and its Measurement." *Journal of the American Statistical Association* 77: 548-561.

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

Shannon, C.E. and W. Weaver. 1949. *The Mathematical Theory of Communication* University of Illinois Press: Urbana.

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

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.

Weingast, B.R. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development." *The Journal of Law, Economics & Organization* 11: 1-31.

Weingast, B.R. 2006. "Second Generation Fiscal Federalism: Implications for Decentralized Democratic Governance and Economic Development." SSRN: http://ssrn.com/abstract=1153440 60 pages.

Weingast, B.R. 2009. "Second Generation Fiscal Federalism: The Implications for Fiscal Incentives." *Journal of Urban Economics* 65: 279-293.

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