Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 71(3) 2017 253–263

A Generalized Measure of Diversity: Application to Longitudinal Data on Crop-groups in North-East India

Utpal Kumar De[1] and Manoranjan Pal[2]
[1]North Eastern Hill University, Shillong
[2]Indian Statistical Institute, Kolkata

Received 04 October 2016; Revised 06 January 2017; Accepted 10 March 2017

SUMMARY

In this paper, we generalize the Herfindahl-Hirschman Index (HHI) using the correlation structure of the individual shares to arrive at a new diversity measure. This is then applied to the cultivation of crops across zones of Assam, India. Before applying these indices we have removed the trend component from the data series. Since the trend effect is eliminated, the indices show the true diversity pattern and reflect the intentions of farmers directly which was not taken into consideration in the earlier measures.

The results show that the diversity indices declined till 1975-76 and then gradually increased up to 1987-88. Thereafter, all the indices became more or less stagnant. It may be because the farmers did not want to take risk or they have less access to the modern farming technology which is also subject to changes over time.

Keywords: Crop diversity, Herfindahl-Hirschman index, Hall and Tideman index, Cropping pattern of Assam.

1. INTRODUCTION

Diversity of cropping pattern is highly associated with the socio-economic, demographic and cultural development of a region. Diversity is thought to have a direct positive relation with the development. However, since development has many dimensions, it may not always be beneficial to advocate for increase of diversity, because it may adversely affect some other issues of development. Some lands may be particularly suitable for special types of crops. So concentration of those crops is more economic for the region. There should be a balance between diversification and concentration which is different for different region. However, when we stick to a given region, the temporal pattern of diversity of cropping pattern will throw some insight if it is analyzed in the perspective of the overall development of the region and not only economic development.

Diversity indicates the degree of spread of activities (like production of commodities, export/import of commodities) to more and more competing items, and relatively in more equitable proportion. On the other hand, the concentration of a particular item indicates the degree of presence of such category in comparison to others in the distribution of population of a place, where diversity is more close to the concept of equality and concentration refers to inequality. Though the concept of inequality and equality was introduced in the area of income distribution, it was later spread also to other variables in economics and other areas of social sciences and even in other natural sciences. Other concepts like diversity, concentration, polarity, segregation etc. came into picture.

The analysis of diversification or concentration finds its importance because of their crucial linkage with the growth and development in the respective fields. In case of economic activity, people become expert in doing some particular activity in which they find more prospects due to growing efficiency in that line. Thus we observe concentration of some

Corresponding author: Utpal Kumar De
E-mail address: utpalkde@gmail.com

activities in a particular zone. Sometimes, human necessity, market conditions, technology, weather etc (as the case may be) help in promoting concentration. Growing concentration of an activity (crop production) may be associated with risk, uncertainty. On the other hand, people (farmers) diversify towards many such activities in order to reduce (distribute) risk and earn sustainable livelihood.

In case of firms in a given industry, concentration measures help to reveal the extent to which a few giant firms have control over the market (monopolistic power or competitive strength). In the context of market or socio-political scenario, too much of concentration may lead to unethical practices. On the other hand, diversity is assumed to be linked with performances. More is the diversity, better is the performance in terms of allocation of resources, rate of return etc. In agriculture, overall returns can be increased either through the expansion of area under cultivation or through rising yields. In any situation however, total agricultural returns can further be increased through diversification of cultivated area from low yielding towards highly remunerative crops (De 2003, De and Chattopadhyay 2010).

In agriculture, crop diversification is an important instrument for its growth, particularly for food and nutrition security, growth of income and employment, poverty alleviation, judicious use of land, water and other resources, sustainable agricultural progress as well as for sustainable environmental management (Singh 2001, De 2003). Therefore, in order to know the benefit accrued to diversity in the agricultural production, one needs to compute a suitable diversity index, which is opposite to that of concentration index.

A variety of measures of diversification have been developed over time and newer measures replaced the older measures in order to avoid the limitation of those measures. Also, measures are developed for their suitability to the particular use. In agriculture, the very crude measure of diversity is the number of crops cultivated and proportion of area under various crops. The simple measure of number of crops however does not speak about the evenness of the distribution of the area under cultivation. Thus the other popular measures were developed and the widely used index was the Hirschman-Herfindahl Index developed independently by both Herfindahl (1950) and Hirschman (1964) (abbreviated as HHI or simply HI as a sort of Herfindahl Index in some studies), which was first used to examine the regional concentration of industries (Theil 1967, Hou and Robinson 2006). It is defined as the sum of squares of proportion of each crop area to gross cultivated area. In symbol HHI = $\sum_{i=1}^{N} P_i^2$; where N is the total numbers of crops and P_i is the proportion of acreage under i^{th} crop to total cropped area. The value of HI ranges from 0 (for perfect diversification) to 1 (for perfect specialization). This concentration index however cannot assume theoretical minimum value for finite number of crops. Also, it gives relatively more weight to larger crop activity. A measure of diversification is obtained by subtracting this value from 1.

Quite a few indices have also been defined in this direction. The Simpson Index is the reciprocal of HI, and is defined as SI = $\frac{1}{\sum_{i=1}^{N} P_i^2}$, which can be traced to a paper by Simpson (1949), titled "Measurement of Diversity" published in Nature. SI represents the number of times one would have to take pairs of individuals at random from a population in order to select a pair of the same species. N is the number of species, or industries, in the population, while P_i is the population weight of each species, firm, industry, or other unit of measurement. The Simpson Index has also been referred to as the Yule Index after the similar measure devised to characterise the vocabulary used by different authors (Yule 1944). Yule (1944) is cited by Simpson (1949) as a key reference for his index, which is a combination of the ideas of Yule (1944) with those of Fisher *et al.* (1943) and Williams (1946).

The Industrial diversity was also examined by using Ogive Index (OI), which was computed by taking into account the deviation from the ideal or equal distribution of acreage (Tress 1938). It is defined as OI = $\sum_{i=1}^{N} \left\{ P_i - \left(\frac{1}{N}\right) \right\}^2 \Big/ \left(\frac{1}{N}\right)$; where N is the total number of crops cultivated in the region. Like HHI, it is also subtracted from unity to convert it into an index of diversification and make it comparable with other indices.

Also, Entropy Index (EI) and Modified Entropy Index (MEI) have been widely used in agricultural diversification literature (Shannon 1948, Hackbart and Anderson 1975, Singh *et al.* 1985, Shiyani and

Pandya 1998). The EI is defined as EI $= -\sum_{i=1}^{N} P_i \, Log \, P_i$. Its value varies from 0 (perfect concentration) to Log N (perfect diversity). The upper value of EI can exceed one or be less than one when N is greater or less than the base of logarithm. Thus it does not correspond to any standard scale of measuring degree of diversification. In MEI number of crops, N is considered as the base of logarithm. Symbolically, MEI $= -\sum_{i=1}^{N}(P_i . Log_N \, P_i)$. The lowest and the highest value of MEI is zero (when only one crop is cultivated in the whole area) and one (perfect diversification) respectively. Though it is better than those other measures (due to its uniform scale) it does not consider the change in value due to changes in number of crop activities (Shiyani and Pandya 1998).

Hannah and Kay (1977) stated that most of the indices are special cases of the general index $I_\phi = (\sum_{i=1}^{n} P_i^\phi)^{1/1-\phi}$ where P_i is the share of i^{th} item and ϕ is a parameter, such that $\phi \geq 0$ and $\phi \neq 1$. For $\phi = 2$ the index becomes $1/\sum_{i=1}^{n} P_i^2$, which is the inverse of the Herfindahl Index of measuring disparity. In extreme case of $\phi = 1$ the index becomes the Entropy Index. This general index measures both the number of items and the evenness of item shares, with the parameter ϕ determines the weighting of emphasis on number of items versus evenness. When, ϕ is zero it simply counts the number of items (Patil and Taillie 1982). Using this general index, Tauer and Seleka (1994) did not find significant relationship between the cash receipt variability with the reduction in diversification of agricultural production across 38 US states during 1960 to 1989. Since they used cash receipts to measure diversity across commodities, it does not capture diversity by production practice.

Hoffmann (2007) however tried to provide a distribution based measure of diversity which is more flexible and represents a useful complement to models of generalized feature based diversity, such as Nehring and Puppe's (2002) "Theory of Diversity". There are two fundamentally different generalised models of distribution based diversity indices. The so-called 'Rényi diversity' is an additive and non-concave one parameter generalisation of the Shannon Entropy (Rényi 1961). It is used to measure biodiversity (Ricotta 2003), statistical sampling (Mayoral 1998), economic diversity modelling (Beran 1999) etc. The other generalisation of Shannon entropy, which is concave but not additive was derived by information theorists Havrda and Charvát (1967), statisticians Patil and Taillie (1982), physicist Tsallis (1988) and also used for biodiversity measurement (Keylock 2005).

Also, researchers used a number of inequality indices to examine concentration of the whole in a few items as opposed to diversification (Clarke 1993, Gini 1912, Roll 1992, Tabner 2006). Various concentration and diversification indices are used by Meilak (2008) to examine the export concentration and its link with the size of the economy. Most of these concentration indices exhibit the general form: CI $= \sum P_i W_i$; $i = 1, 2, ..., n$, where W_i = the weight attached to the export share of a particular export category, P_i = the share of export category i and n = the number of export categories.

Most of these indices are based on the shares of each category and used arbitrary weights to the respective items and applicable to cross section data. These measures ignore relative size variations in commodity groups and these can equally describe a country exporting one product and a country exporting x product groups with similar shares.

Though, these measures satisfy the general characteristics of a concentration or diversification measure, it calls for improvement upon such indices in order to incorporate the interdependencies among the variables. Also, while analysing temporal changes, there is general tendency of changes in variables with time. It may be important to adjust with such trend in order to understand the changing diversity due to human effort on consideration of other factors. In this paper, we tried to develop an alternative measure of diversification by taking into account the bi-variate correlation coefficients (of time series of the variables) and excluding the trend effects. Also, we compare the results to examine the improvement in the results obtained by using such measure.

Please note that a*s* in the present case correlations among crops (here crop groups) over time is considered excluding the time trend the emphasis of this measure is on the time series or panel data**.** For cross section comparison of diversity say across zone no question of time effect or correlation at one point would occur and hence simple HHI can be applied.

Review of Some Relevant Measures of Diversification and the Proposed Measure

For a meaningful measure of diversity it is first necessary to convert all the variables into shares. This is to make the measures independent of unit of measurement. Suppose p_i is the share of i^{th} item. Assume without loss of generality that $p_1 > p_2 > \ldots > p_n$, where n is the number of items in the group. Naturally, $\sum p_i = 1$.

If P is a measure of concentration then a measure of diversity may either be $1/P$ (reciprocal of P), or $1–P$ (complementary of P). The value of $1–P$ lies between 0 and 1 as P lies in between 0 and 1. That is why this measure $(1–P)$ is preferred over $1/P$ as a measure of diversification.

Among other indices, Absolute Concentration Ratio is the oldest and a very popular measure which is defined as

$$\text{ACR} = \sum_{i=1}^{K} P_i,$$

where, K is the top K firms in the industry. This is also known as K-firm concentration ratio. Within an industry, K is often taken as a number between 3 and 5, because it is often the case that 3-5 largest firms have about 70%-80% share of the industry. K is thus known as the focal point of concentration. However, K should be fixed if we want to compare among different industries. If firms of all the industries are taken then the value of K may be very large. The corresponding diversity index is Absolute Diversification Index (ADI).

$$\text{ADI} = \mathbf{1} - \sum_{i=1}^{K} \boldsymbol{P_i}.$$

We can define a more general form of most popular Herfindahl-Hirschman Index (HHI) by taking the following weighted sum of P_i:

$$\text{GHI} = \sum_{i=1}^{n} w_i P_i,$$

where w_i is the weight attached to P_i, so that $\sum_{i=1}^{n} w_i$ $\sum_{i=1}^{n} w_i = 1$. HHI is then viewed as the weighted sum of P_i, where the weight of P_i is nothing but P_i itself. Thus smaller firms contribute less to the value of the index. Hall and Tideman (1967) proposed some desirable properties which should be satisfied by a good measure of concentration. HHI satisfies all the axioms proposed by Hall and Tideman. Apart from some of the Hall and Tideman axioms, Hannah and Kay (1977) proposed a set of axioms based on entry of new firms, merger of firms etc.

There are many other concentration indices proposed in the literature, for example, Hall and Tideman Index (Hall and Tideman 1967, Hause 1977, Anbarci and Katzman 2005), Entropy measure of concentration (Hart 1971), Comprehensive Concentration Index (CCI), by Horvarth (1970); Hannah and Kay (1977) measure of concentration; Erlat and Akyuz Index of Concentration (2001); U index by Davies (1979) and Linda Index (Linda 1976). One can also define the corresponding diversification index to each of the concentration indices. Since we shall mainly concentrate on Herfindahl Diversity Index (HDI), we refrain ourselves of narrating all these indices.

Following HDI, Douglas Rae (1967, 1968) proposed an index of fractionalization in the party system, which is defined as $D_{DR} = 1 - \sum_{i=1}^{n} P_i^2$, where D^{DR} = the index of fractionalization in the party system and P_i is the proportion of party i of votes given (See also Rae *et al.* 1970, Vayrynen 1970). Greenberg's measure of linguistic diversity (Greenberg 1956) is also same as the above index.

A similar index of fragmentation, defined by Rae and Taylor (1970), is $F_{RT} = 1 - \sum_{i=1}^{n} \left(\frac{n_i}{n}\right)\left(\frac{n_i - 1}{n}\right)$, which is the probability that a randomly selected pair of individuals in a society will belong to different groups. n_i is the number of members of the i^{th} group.

2. THE PROPOSED MEASURE

While measuring the oligopolistic power in the market one should consider the existing cooperation between firms. When there is more cooperation between the firms, there is more power of the firms in the market. Correlations between agricultural variables (here proportion of area under crops) should reduce the diversity index. Almost all the productions of the commodities move in the same direction as time. Time here acts as an intervening variable. So the effect of time from each variable should be removed before finding the correlation structure.

Let us assume that the correlation between area under i^{th} and j^{th} crop over time is ρ_{ij} after eliminating the effect of time. This is in fact the corresponding partial correlation coefficient. The proposed measure is Modified Diversity Index (MDI), defined as $\mathbf{MDI = 1 -} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i \rho_{ij} p_j = \mathbf{1 - P'RP}$, where **R** is the matrix of ρ_{ij} values and **P** is the column vector of shares of area under crops for the current period. Observe that **P'RP** can be taken as a measure of concentration index. There are some salient features of this index.

(a) $0 \leq$ MDI $\leq$ ADI, (assuming that any two variables are non-negatively correlated). It is 0 when $\rho_{ij} = 1$ for all *i* and *j*, i.e., when any two variable move in the same-direction without any error. In this case $s'Rs$ becomes $(\Sigma P_i)^2 = 1$. It is HDI when the variables are completely independent of each other, $\rho_{ij} = 0$ for all *i* and *j*.

(b) $0 \leq$ MDI ≤ 1, assuming that the variables can move freely, i.e., when ρ_{ij} can take any value in the range – 1 to 1. [$\mathbf{P'RP} \leq \mathbf{P'\underline{1}P} = (\sum P_i)^2$ =1, where $\underline{1}$ is the matrix consisting of 1's only. Since **R** is a symmetric non-negative definite matrix, we have $\mathbf{P'RP} \geq 0$.]

3. DESCRIPTION OF DATA AND CALCULATIONS

The present analysis is based on the secondary data on area under various crops since 1951 to 2011 collected from various issues of *Statistical Hand Book* of Assam, *Economic Survey of Assam* and *Reports* from Directorate of Economics and Statistics and Directorate of Agriculture, Government of Assam. Though several studies used earnings from the production of crops in order to compute the diversity index, here we use area under crops for the purpose. This is because the farmers try to maximise their returns from limited land under them through suitable allocation of land among the cultivable crops. Thus the land size allocation to different crops reflects the intention of the farmers which may not be realized through production. Moreover, the area of crop cultivation is more robust than the actual production, which is subject to technology available at the time of production and to the climatic behaviour of nature.

Here instead of considering all individual crops, six major crop groups namely, cereals, pulses, oilseeds, fibres, plantations and fruits and vegetables have been considered. If all the 29 crops grown in Assam are considered, then it will be too laborious. Moreover, despite the fact that some food crops now-a-days became cash crops, with the progress of agriculture (due to irrigation, fertiliser and technological progress) proportion of land allocation towards fruits & vegetables and oilseeds increase over time.

This is done for each year starting from 1951-52 to 2010-11. Thus we have data for 60 years. The following line diagrams give us a clear idea about the trend of area for each of the major groups.

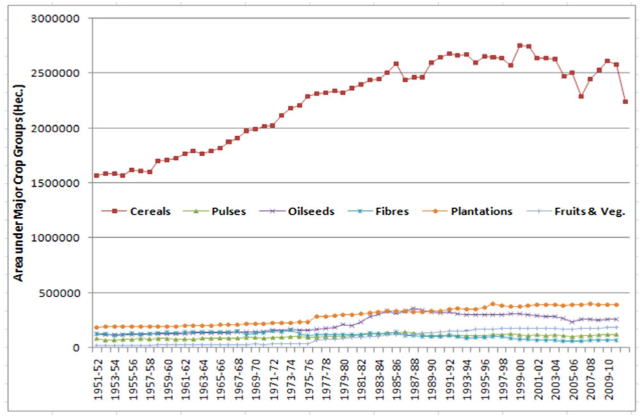

Fig. 1. Trends in the Areas under Major Crop Groups in Assam

It is clear from Fig. 1 that all but one major group have increasing trend. The one which has recorded a decreasing trend is 'Fibres'. Fibres do not come under food crops. The patterns of growth of the groups other than Cereals are not easily discernible because of Cereals itself, which captures over 70 per cent of the total area under cultivation. The trends of other five groups can be clearly visible from Fig. 2 drawn for these five groups only.

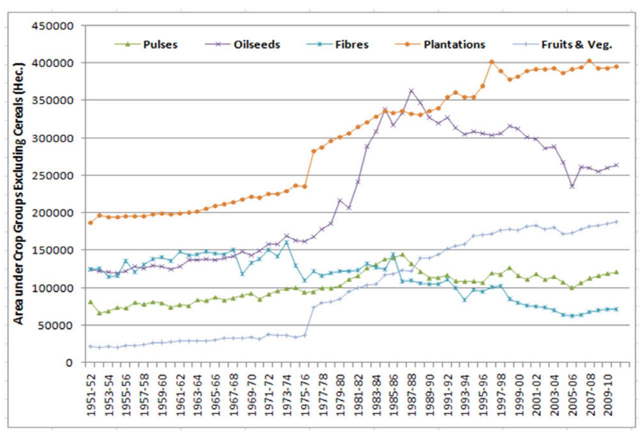

Fig. 2: Trends in the Areas under Major Crops except Cereals

If the patterns of trends are observed, one can notice that there is an increasing movement in all but Fibre up to the year 1987-88. Also the increasing trend is somewhat slow up to 1975-76. After 1987-88, it either decreases or increases at a slower rate than the past. The similarity of the trends would imply high positive correlations among the groups except possibly with the areas under Fibres. Since from the figures it is evident that the trends are similar for each group, the bivariate correlation coefficients among these groups have been computed to check whether it is really so. The correlation matrix is as follows.

Table 1. Bivariate Correlations of the Areas under Major Crop Groups

	Cereals	Pulses	Oil Seeds	Fibres	Plantation Crops	Misc. Crops
Cereals	1	.867**	.905**	-.661**	.929**	.902**
Pulses		1	.914**	-.477**	.819**	.772**
Oil Seeds			1	-.613**	.887**	.880**
Fibres				1	-.834**	-.860**
Plantation Crops					1	.989**
Misc. Crops						1

** Correlation is significant at the 0.01 level (2-tailed).

Area under Fibres is negatively correlated with all other area groups. This was expected from the trend diagram. But, the degrees of correlations, regardless whether negative or positive are too high in absolute value for the areas under crop to be of further effective use without trend corrections. Time seems to be an intervening variable. Thus all the variables for time have been corrected and then diversity index for each year have been computed by using the proposed formula. To compare these indices the diversity indices have been computed using existing method. This leads to the following six types of indices:

1. Herfindahl-Hirschman Diversification Index without correcting the variables for time.

   HDI = $1 - \sum_{i=1}^{n} P_i^2$.

2. Herfindahl-Hirschman Diversification Index after correcting the variables for time (Corrected HDI)

   CHDI = $1 - \sum_{i=1}^{n} P_i^{*2}$, where s_i^* is the share after correcting the variables for time.

3. Entropy Index (EI) = $-\sum_{i=1}^{N} P_i \; Log\, P_i / \log N$, where P_i is the proportion of area under i^{th} crop group. Since the Entropy or corrected Entropy Index takes the maximum value Log N, we divided the values of indices obtained by Log N to make it comparable with HDI or MDI.

4. Entropy Index after correcting the variables for time

   CEI = $-\sum_{i=1}^{N} P_{ij}^{\phi} Log\, P_i^{\phi} / \log N$

   where s_i^* is the share after correcting the variables for time.

5. Modified Diversity Index (MDI)

   MDI = $1 - \sum_{i=1}^{n} \sum_{j=1}^{n} \; p_i \rho_{ij} p_j$ = 1 – **P'RP**,

   where ρ_{ij} is the correlation between i^{th} and j^{th} areas, **P** is the vector of shares and **R** is the correlation matrix of areas.

6. Modified Diversity Index after correcting the variables for time (Corrected MDI)

   CMDI = $1 - \sum_{i=1}^{n} \sum \quad p_i \rho_{ij} p_j$ = 1 – $\mathbf{P}^{\varphi}\mathbf{R}^{\varphi}\mathbf{P}^{\varphi}$

   where P_i^* is the share of area of i^{th} crop group after correcting the variables for time, $\mathbf{P}^*$ is the corresponding vector and $\mathbf{R}^*$ is the correlation matrix of areas after correcting the variables for time.

Here it should be noted that correction for time should be mean preserving. Otherwise the indices are meaningless. More precisely, first we run regression of each variable on time, find the residuals and add each residual with the mean value of the variable.

4. OBSERVATIONS AND DISCUSSION

From the data, first of all the bivariate correlations ρ_{ij}, and the shares P_i have been computed. Thereafter, HDI and MDI are computed. The variables are then

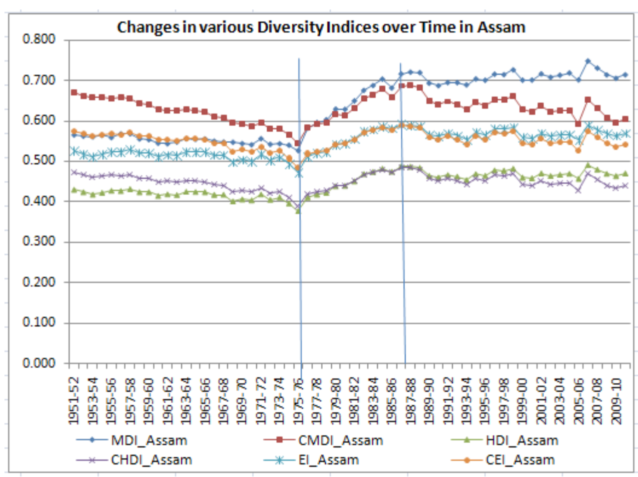

Fig. 3: Trends in the Diversity Indices of the Areas under Major Crop Groups

Table 2. Year wise Diversity Indices of the Areas under Major Crop Groups in Assam

Year	HDI	CHDI	EI	CEI	MDI	CMDI	Year	HDI	CHDI	EI	CEI	MDI	CMDI
1951-52	0.431	0.472	0.527	0.574	0.566	0.672	**1981-82**	0.452	0.451	0.555	0.554	0.650	0.631
1952-53	0.424	0.467	0.517	0.567	0.563	0.663	**1982-83**	0.470	0.468	0.573	0.572	0.676	0.657
1953-54	0.418	0.462	0.512	0.561	0.561	0.659	**1983-84**	0.475	0.473	0.578	0.576	0.688	0.666
1954-55	0.423	0.464	0.517	0.564	0.565	0.660	**1984-85**	0.483	0.481	0.586	0.584	0.704	0.680
1955-56	0.428	0.466	0.523	0.569	0.559	0.656	**1985-86**	0.475	0.472	0.580	0.577	0.682	0.660
1956-57	0.427	0.464	0.523	0.566	0.569	0.658	**1986-87**	0.489	0.485	0.592	0.589	0.717	0.688
1957-58	0.432	0.467	0.528	0.570	0.569	0.657	**1987-88**	0.488	0.484	0.589	0.585	0.722	0.690
1958-59	0.425	0.459	0.521	0.563	0.556	0.644	**1988-89**	0.485	0.480	0.587	0.582	0.720	0.684
1959-60	0.424	0.458	0.521	0.561	0.554	0.640	**1989-90**	0.465	0.457	0.566	0.559	0.694	0.651
1960-61	0.415	0.450	0.511	0.552	0.545	0.629	**1990-91**	0.461	0.452	0.562	0.554	0.688	0.642
1961-62	0.419	0.451	0.515	0.555	0.544	0.627	**1991-92**	0.467	0.458	0.569	0.561	0.696	0.649
1962-63	0.417	0.448	0.514	0.551	0.548	0.624	**1992-93**	0.463	0.453	0.563	0.554	0.696	0.642
1963-64	0.426	0.453	0.523	0.557	0.556	0.630	**1993-94**	0.455	0.442	0.553	0.542	0.690	0.630
1964-65	0.425	0.451	0.522	0.555	0.555	0.625	**1994-95**	0.469	0.457	0.570	0.561	0.705	0.647
1965-66	0.425	0.449	0.522	0.553	0.556	0.622	**1995-96**	0.466	0.452	0.565	0.554	0.702	0.639
1966-67	0.418	0.442	0.515	0.545	0.551	0.612	**1996-97**	0.479	0.467	0.579	0.570	0.716	0.654
1967-68	0.418	0.440	0.515	0.544	0.548	0.607	**1997-98**	0.478	0.465	0.579	0.569	0.715	0.652
1968-69	0.401	0.424	0.498	0.525	0.548	0.595	**1998-99**	0.483	0.470	0.583	0.574	0.727	0.663
1969-70	0.406	0.427	0.503	0.529	0.545	0.592	**1999-00**	0.461	0.443	0.559	0.544	0.702	0.628
1970-71	0.404	0.423	0.500	0.524	0.542	0.585	**2000-01**	0.459	0.440	0.556	0.541	0.702	0.622
1971-72	0.419	0.434	0.517	0.537	0.557	0.597	**2001-02**	0.471	0.453	0.569	0.555	0.717	0.638
1972-73	0.405	0.420	0.503	0.522	0.543	0.579	**2002-03**	0.465	0.443	0.562	0.544	0.709	0.623
1973-74	0.411	0.424	0.509	0.526	0.544	0.579	**2003-04**	0.468	0.446	0.565	0.547	0.714	0.627
1974-75	0.396	0.408	0.492	0.507	0.541	0.565	**2004-05**	0.470	0.446	0.566	0.547	0.719	0.624
1975-76	0.377	0.389	0.472	0.486	0.527	0.544	**2005-06**	0.458	0.428	0.552	0.526	0.702	0.593
1976-77	0.412	0.419	0.512	0.519	0.583	0.583	**2006-07**	0.492	0.471	0.588	0.573	0.749	0.654
1977-78	0.418	0.424	0.519	0.524	0.597	0.591	**2007-08**	0.481	0.456	0.577	0.558	0.731	0.631
1978-79	0.423	0.426	0.523	0.527	0.604	0.594	**2008-09**	0.470	0.441	0.568	0.543	0.715	0.608
1979-80	0.439	0.441	0.540	0.542	0.630	0.617	**2009-10**	0.465	0.433	0.563	0.535	0.706	0.596
1980-81	0.440	0.441	0.543	0.544	0.629	0.614	**2010-11**	0.471	0.440	0.569	0.543	0.715	0.605

corrected for time and translated so that they become mean preserving. The correlations ρ_{ij}^* and the shares P_i^* are computed from corrected mean preserving areas. These may be termed as corrected correlations and corrected shares respectively. In Table 2 we present the six diversity indices.

From Fig. 3 it is clear that all the indices declined till 1975-76 and then gradually increased up to 1987-88. Thereafter, all the indices became more or less stagnant. In case of HDI and Entropy the actual and corrected indices are found to be very close and the corrected values lie below the actual one before 1975-76 and gradually crossed above the actual figures after 1975-76. In case of MDI also CMDI lies below the MDI till 1975-76 and gone above it afterwards and the gap increased continuously. Here, as all the variables in absolute form increased with time, the corrected (against time trend) indices lies above the actual figures in their rising phase (after 1975-76) and lies below the actual figures in their declining phase.

Table 3. Bivariate Correlations of the Diversity Indices of the Areas under Major Crop Groups in Assam

	MDI	CMDI	HDI	CHDI	EI	CEI
MDI	1	0.418**	0.954**	0.408**	0.950**	0.397**
CMDI		1	0.627**	0.983**	0.623**	0.973**
HDI			1	0.642**	0.997**	0.638**
CHDI				1	0.637**	0.998**
EI					1	0.635**
CEI						1

**. Correlation is significant at the 0.01 level (2-tailed).

Another interesting point to be noted is that the MDI is always higher than the HDI and Entropy (EI) in both corrected and uncorrected forms. But all the six indices have similar pattern of movement, first decline to give more emphasis on the food crops for food security concern (up to 1975-76) and then increased though slowly up to 1987-88 and then stabilized with very slow changes in diversity. This is reflected by high correlations among them, as shown in Table 3.

The proposed Modified Diversity Index (MDI) and its corrected version (CMDI) are thus different from and lie above the other four indices. If we compare CMDI and MDI we see prominent contrast between the two indices. The difference between the two values is initially higher than the corresponding other to pairs (EI vs. CEI and HDI vs. CHDI) and this difference gradually decreases at a point in a similar manner. When they increase, the difference between the two is again sharper than the other two pairs. Thus CMDI reflects better diversity pattern.

In conclusion, it can said that after Independence the diversity of crop areas more or less has a decreasing trend up to the period of 1976-77 and then increased up to 1987-88. After 1987-88 it is more or less stagnant. The reason for this is not clear and should be traced in migration patterns, floods or other economic and natural phenomena.

It was already noted that the area under crops for each of the major groups were found by aggregating the areas under individual crops as follows:

Cereals	Rice (Paddy), Maize, Wheat and Other Cereal
Pulses	Gram, Arhar and Other Rabi Pulses
Oil Seeds	Rape and Mustard, Sesamum, Linseed and Castor
Fibres	Jute, Mesta and Cotton
Plantation Crops	Tea, Sugar Cane, Areca Nut and Tobacco
Fruits and Vegetables	Potato, Sweet Potato, Chilli, Turmeric, Tapioca, Banana, Onion and Pine Apple

A closer look in the data reveals that data for some of the individual crops were not available up to 1975-76 and thus the aggregated values were found neglecting the unavailable observations. This further confirms our proposition that the indices should be calculated separately for all the sub-periods. To examine how the correlations changes we may look into the bivariate correlations separately for the three periods as presented in Table 4.

Table 4. Bivariate Correlations of the Areas under Major Crops Separately for the Three Periods

	Cereals	**Pulses**	**Oil Seeds**	**Fibres**	**Plantation Crops**	**Misc Crops**
	Period 1 (1951-52 to 1975-76)					
Cereals	1.000	0.888**	0.966**	0.220	0.980**	0.939**
Pulses		1.000	0.917**	0.257	0.879**	0.863**
Oil Seeds			1.000	0.292	0.957**	0.920**
Fibres				1.000	0.177	0.442*
Plantation Crops					1.000	0.911**
Misc. Crops						1.000
	Period 2 (1976-77 to 1987-88)					
Cereals	1.000	0.891**	0.847**	0.497	0.882**	0.864**
Pulses		1.000	0.952**	0.173	0.982**	0.971**
Oil Seeds			1.000	0.055	0.965**	0.963**
Fibres				1.000	0.165	0.062
Plantation Crops					1.000	0.973**
Misc. Crops						1.000
	Period 3 (1988-89 to 2010-11)					
Cereals	1.000	0.175	0.459*	0.388	-0.142	-0.044
Pulses		1.000	0.359	0.249	0.025	0.085
Oil Seeds			1.000	0.859**	-0.747**	-0.690**
Fibres				1.000	-0.739**	-0.751**
Plantation Crops					1.000	0.920**
Misc. Crops						1.000
	All Periods (1951-52 to 2010-11)					
Cereals	1	.867**	.905**	-.661**	.929**	.902**
Pulses		1	.914**	-.477**	.819**	.772**
Oil Seeds			1	-.613**	.887**	.880**
Fibres				1	-.834**	-.860**
Plantation Crops					1	.989**
Misc. Crops						1

Note: (i) **. Correlation is significant at the 0.01 level (2-tailed), (ii) *. Correlation is significant at the 0.05 level (2-tailed).

The overall correlations indicate inverse relations of fibres with cereals, pulses, oilseeds, plantations and other crops, which is a normal behaviour where cultivation of natural fibre crops have declined over the years while others increased. However, significant positive associations among cereals, pulses, oilseeds and plantation crops are the indications of simultaneous growth without any substitution of land use. Only in the recent past decades minor substations have taken place and thus correlations have been weakened. Though HDI is more or less stagnant over time,

Table 5. Unit Root Test for Stationarity of Various Diversity Indices of Assam

		Coeff	T	Trend	t-Value	AdjR^2	F	Log Likelihood
HDI	Level	-.1832	-2.455	.00026***	2.019	.066	3.047**	184.38
	1st Diff	-1.197***	-9.07	1.22E^{-05}	.137	.585	41.13***	179.13
CHDI	Level	-.203	-2.58	2.27E^{-05}	.254	.08	3.372**	180.45
	1st Diff	-1.227***	-9.33	1.72E^{-05}	.181	.60	43.48***	175.21
EI	Level	-.184	-2.441	.00026*	1.953	.064	2.997*	180.12
	1st Diff	-1.195***	-9.081	7.57E^{-07}	.008	.585	41.24***	175.13
CEI	Level	-.216	-2.67	2.43E^{-05}	.254	.082	3.594**	176.14
	1st Diff	-1.24***	-9.46	1.60E^{-05}	.157	.606	44.75***	171.01
MDI	Level	-.112	-1.904	.00047*	1.794	.03	1.833	165.14
	1st Diff	-1.153***	-8.646	1.68E^{-05}	.138	.561	37.38***	160.75
CMDI	Level	-.148	-2.20	3.14E^{-05}	.251	.05	2.45*	160.6
	1st Diff	-1.212***	-9.184	2.12E^{-05}	.161	.591	42.17***	156.38

Note: Calculated from the estimated Indices on the basis of Area under various crop groups.

Table 6. Results of Johansen's Cointegration Test on all Six Diversity Indices

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.523319	146.3525	107.3466	0.0000
At most 1 *	0.438304	103.3799	79.34145	0.0003
At most 2 *	0.325583	69.92574	55.24578	0.0015
At most 3 *	0.293481	47.07911	35.01090	0.0017
At most 4 *	0.233602	26.92964	18.39771	0.0025
At most 5 *	0.179836	11.49852	3.841466	0.0007

Trace test indicates 6 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None	0.523319	42.97269	43.41977	0.0559
At most 1	0.438304	33.45413	37.16359	0.1258
At most 2	0.325583	22.84663	30.81507	0.3400
At most 3	0.293481	20.14947	24.25202	0.1592
At most 4	0.233602	15.43112	17.14769	0.0874
At most 5 *	0.179836	11.49852	3.841466	0.0007

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):					
HDI_ASSAM	EI_ASSAM	MDI_ASSAM	CHDI_ASSAM	CEI_ASSAM	CMDI_ASSAM
1097.834	-366.3598	-210.6765	-1179.047	36.57409	473.4758
927.3049	-481.8900	92.87543	-1494.517	1056.822	-120.0704
712.2211	-211.8219	-213.5608	-677.9735	168.0046	248.9695
-187.9881	337.5110	-169.6786	1211.129	-1157.369	45.57378
-986.4401	501.9942	6.681409	-113.8838	402.3373	76.15010
-128.0627	433.5119	-129.5492	526.2351	-782.6095	127.3305

1 Cointegrating Equation					
Normalized cointegrating coefficients					
HDI_ASSAM	EI_ASSAM	MDI_ASSAM	CHDI_ASSAM	CEI_ASSAM	CMDI_ASSAM
1.000000	-0.333711	-0.191902	-1.073975	0.033315	0.431282
	(0.06096)	(0.04410)	(0.20512)	(0.21322)	(0.06321)

Notes: Figures in the parentheses represent standard error, Log likelihood 1722.778

declining MDI indicates lowering diversity pattern over time. However, the CMDI also reflects more or less stagnant diversity of crop in Assam in recent time period.

Table 5 shows that all the indices follow an integrated pattern of order one and virtually no trend. It first declined and then increased after 1875-76 and became stagnant after 1986-87. So a co- integrated pattern is noticed and all six indices have a tendency to converge on a growth rate. Thus, except the value prediction, the trend prospects can be examined by using any one of them. The co-integration test followed by Engle-Granger and Johansen's methods are presented below.

5. CONCLUSION

In this paper we tried to develop a new measure of diversification that reflects the true intension of the farmers and also eliminated the effect of time on the variation in area and proportion of area under crops. Here, we kept in mind that the new measure should follow the general properties of a diversity index and its comparability with other such measures. The results suggest that the new measure provide a better indication of the diversity when comparing the trends. The proposed measure, other than giving a sharper contrast, gives similar trend as the other indices. We in fact generalized the Herfindahl-Hirschman Index (HI) using the correlation structure of the individual shares to arrive at a new diversity measure. This is then applied to the cultivation of crops in Assam, India after removing the trend component.

The results show that the diversity indices declined till 1975-76 and then gradually increased up to 1987-88. Thereafter, all the indices became more or less stagnant. It may be because the farmers do not want to take risk or they have less access to the modern farming technology which is changing over time. North-East India is relatively a backward region in India. The fruits of development, i.e., the technical advancement so far as the innovations of farming technology either do not go to this region or the farmers are not much aware of those. They thus feel reluctant to change the cropping pattern. This may also be due to the lack of variations in the demand pattern of the agricultural goods.

REFERENCES

Adelman, M.A. (1969). Comment on the 'H' concentration measure as a numbers equivalent. *Rev. Eco. Statist.*, **51**, 99-101.

Ahmad, Al-Samarrie and Herman, P. Miller (1967). State differentials in income concentration. *Amer. Eco. Rev.*, **57**, 59-72.

Anbarci, N. and Brett, K. (2005). Socially beneficial mergers: A new class of concentration indices. *GA*, 30144-5591, March 28.

Beran, H. (2005). A generalisation and optimization of a measure of diversity. In: U. Leopold-Wildburger, G. Feichtinger and K.P. Kistner (Eds) *Modelling and Decisions in Economics*, 119-137. *Physica, Heidelberg*.

Clarke, R. (1993). *Industrial Economics*. Cambridge Massachusetts, Blackwell.

Davies, S.W. (1979). Choosing between concentration indices: The iso-concentration curve. *Econometrica,* **46**, 67-75.

De, U.K. (2003). *Economics of Crop Diversification*. Akansha Publishing House, New Delhi.

De, U.K. and Chattopadhyay, M. (2010). Crop diversification by poor peasants and role of infrastructure: evidence from West Bengal. *J. Devt. Agri. Eco.,* **2(9)**, 340-350.

Erlat, G. and Akyuz, O. (2001). Country concentration of Turkish exports and imports over time. *Economics Web Institute*. Downloaded from http://www.luc.edu/orgs/meea/volume3/gerlat/gerlat.html (Accessed on 2-2-2016).

Fisher, R.A., Corbet, S.A. and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Eco.,* **12(1)**, 42-58.

Foldvary, Fred E. (2006). The measurement of inequality, concentration and diversification. *Ind. Econ. J.,* **54(3)**, September-December.

Gini, C. Variabilita' e Mutabilata. (1912). Reprinted in *Memorie di Metodologia Statistica* Rome: Libreria Eredi Virgilio Veschi. 1955. Ed. AV. Pizetti and T. Salvemini; Cited by E. W. Weisstein, Mathworld.com (1999-2004).

Greenberg, Joseph H. (1956). The measurement of linguistic diversity. *Language*, **32(1)**, 109-115.

Hackbart, M.M. and Anderson, D.A. (1975). On measuring economic diversification. *Land Economics,* **51(4)**, 374-378.

Hall, M. and Tideman, L. (1967). Measures of concentration. *J. Amer. Statist. Assoc.* **62**, 162-68.

Hannah, L. and Kay, J.A. (1977). *Concentration in Modern Industry: Theory, Measurement and the U.K. Experience*. MacMillan Press, London.

Hart, P.E. (1971). Entropy and other measures of concentration. *J. Roy. Statist. Soc., Series A* (General) **133**, 73-85.

Hart, P.E. (1975). Moment distribution in economics: an exposition, entropy and other measures of concentration. *J. Roy. Statist. Soc., Series A* (General), **138**, 423-434.

Hause, J.C. (1977). The measurement of concentrated industrial structure and the size distribution of firms. *Ann. Eco. Social Measure.*, **6**, 73-107.

Havrda, J. and Charvát, F. (1967). Quantification method of classification processes: The concept of structural α-entropy. *Kybernetika*, **3**, 30-35.

Herfindahl, O.C. (1950). *Concentration in the US steel industry*, Doctoral dissertation Unpublished. Columbia University.

Hirschman, A.O. (1964). The paternity of an index. *Amer. Econ. Rev.*, **54**, 761-762.

Hoffmann, S. (2006). Concavity and additivity in diversity measurement: re-discovery of an unknown concept. Univ., FEMM.

Hou, K. and Robinson, D.T. (2006). Industry concentration and average stock returns. *J. Finance,* **61(4)**, 1927-1956.

Hovarth, J. (1970). Suggestion for a comprehensive measure of concentration. *South. Eco. J.*, **36**, 446-452.

Hart, P.E. and Clarke, R. (1980). *Concentration in British Industry 1935-75*. Cambridge University Press, Cambridge.

Keylock, C. (2005). Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. *Oikos*, **109**, 203-206.

Kwoka, J. (1985). The Herfindahl index in theory and practice. *Antitrust Bull.*, **30**, 915-947.

Linda, R. (1972). Oligopolistic competition and international competitive planning. *Economie Appliquee*, **2-3**, 325-472.

Linda, R. (1976). M´ethodologie de l'analyse de la concentration appliqu´ee `al'´etude Des Secteurs et des march´es, *European Commission*.

Mansfield, Edward D. (1992). The concentration of capabilities and international trade. *International Organization,* **46(3)**, 731-764.

Mayoral, M. (1998). Rényi's entropy as an index of diversity in simple-stage cluster sampling. *J. Inform. Sci.*, **105**, 101-114.

Miller, R.A. (1967). Marginal concentration ratios and industrial profit ratios: some empirical results of oligopoly behavior. *South. Eco. J.*, **34**, 259-267.

Meilak, C. (2008). Measuring export concentration: The implications for small states. *Bank Valletta Rev.*, **37**, 35-48.

Nehring, K. and Puppe's, C. (2002). A theory of diversity. *Econometrica* **70**, 1155-1198.

Patil, G.P. and C. Taillie (1982). Diversity as a concept and its measurement. *J. Amer. Statist. Assoc.,* **77(379)**, 548-561.

Douglas, Rae. (1967). *Political Consequences of Electoral Laws*. Yale University Press, New Haven, 53-58.

Douglas, Rae (1968). A note of the fractionalization of some European party systems. *Comparative Pol. Studies,* **1(3)**, 414-416.

Douglas, R. and Taylor M. (1970). *The Analysis of Political Cleavages*. Yale University Press, New Haven, 24-33.

Ray, J.L. and Singer, J.D. (1973). Measuring the concentration of power in the international system. *Sociol. Methods Res.,* **1(4)**, 403-437.

Rényi, A. (1961). On measures of entropy and information. In: *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley*, 547-561.

Ricotta, C. (2003). On parametric evenness measures. *J. Theo. Biol.,* **222**, 189-197.

Roll, R. (1992). Industrial structure and the comparative behaviour of international stock market indices. *J. Finance,* **47(1)**, 3-41.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell. Sys. Tech. J.,* **27(3)**, 379-423.

Shiyani, R.L. and Pandya, H.R. (1998). Diversification of agriculture in Gujarat: A spatio-temporal analysis. *Ind. J. Agric. Eco.,* **53(4)**, 627-639.

Simpson, E.H. (1949). Measurement of diversity. *Nature*, **163**, 688.

Jain, A.J., Singh, KK. and Sain, I. (1985). Diversification of Punjab agriculture: An econometric analysis. *Ind. J. Agric. Eco.*, **50(3)**, 298-303.

Singh, R.B. (2001). Welcome address, presented at the expert consultation on crop diversification in the Asia pacific region held at FAO Regional Office for Asia and the Pacific in Bangkok, Thailand, 4-6 July, 2000.

Tabner, I.T. (2006). Benchmark concentration: Value weights versus equal weights in the FTSE 100 Index, SSRN Working Paper.

Tauer, L.W. and Seleka, T.B. (1994). *Agricultural Diversity and Cash Receipt Variability for Individual States*. Cornell Agricultural Economics Staff Paper No 94-1, January, Cornell University: New York.

Theil, H. (1967). *Economics and Information Theory*. North-Holland Publishing Co., Amsterdam.

Tress, R.C. (1938). Unemployment and diversification of industry. *The Manchester School,* **9**, 140-152.

Tsallis, C. (1988). Possible generalization of boltzmann-gibbs statistics. *J. Statist. Phys.,* **52(1/2)**, 479-487.

Vayrynen, Raimo (1970). Analysis of party systems by concentration, fractionalization, and entropy measures. *Scandinavian Pol. Studies*. Bind 7 (1972).

Williams, C.B. (1946). Yule's characteristic and the index of diversity. *Nature,* **157(13)**, 482.

Yule, G.U. (1944). *Statistical Study of Literary Vocabulary*. Cambridge University Press, Cambridge.