Measuring Diversity in Australian Residential Property

by

John Byrne
School of Business IT,
RMIT University, Melbourne, Australia
john.byrne@rmit.edu.au

John Flaherty
Department of Accounting and Finance,
Monash University, Clayton, Victoria
john.flaherty@buseco.monash.edu.au

presented at

Pacific Rim Real Estate Society
Tenth Annual Conference
25-28 January 2004

Thammasat University
Bangkok, Thailand

Abstract

This paper examines the diversity of Australian dwelling type and of Australian property occupancy based on ABS census data. A number of standard diversity measures are employed to calculate an index of diversity. In addition to these measures the standard deviation has been employed to provide an ideal with which to make comparisons. While the results indicate that although Australian dwelling type is becoming more diverse, occupancy type is becoming less diverse.

1. Introduction

This paper reports on a preliminary study into the diversity of Australian residential property in two main areas. First, the diversity of dwelling type is considered, with diversity based on two criteria. These criteria are,

- Dwelling type where the categories are separate houses, attached or town houses, flats or apartments and other type of dwellings
- Type of occupation where the categories are occupied dwellings and unoccupied dwellings [1]

Second, the diversity of home ownership, by ownership type, is calculated. Again two diversity criteria are measured. These criteria are,

- Home ownership type where the categories are owners, purchasers[2], public renters, private renters and other types of tenure
- Type of ownership where the categories are owners and renters

For the purpose of this analysis, perfect diversity may be defined as a uniform distribution among the categories. In other words, if there are 100 dwelling types and four categories; separate homes, attached houses, flats, and other types of dwellings, then there would be 25 dwellings in each category.

In each case diversity is measured over time using Australian Bureau of Statistics (ABS) Census Data. There is a number of diversity measures in the non-housing literature, some of these are discussed in the next section. The population standard deviation is also considered appropriate as it is possible to identify closeness to perfect diversity as defined in the previous paragraph. Closeness to perfect diversity is not immediately obvious with some other measures due to the nature of diversity measures.

In the next section, some of the more common diversity measures mentioned in the various literatures are identified and the reasons for using these measures. In the final section, a diversity index is calculated for each of the categories described above together with a summary of the results.

[1] The term unoccupied is used by the ABS to refer to dwellings that were unoccupied at the time of the census.
[2] Those who are in the process of purchasing their home and still have mortgages.

2. Diversity Measures

There are a number of diversity measures employed in the financial, biological, data mining and environmental literatures. Due to the diversity of applications in these various fields, several measures, many of which overlap, are consistently employed. A survey of the housing literature did not reveal the existence of any suitable diversity measures for the current analysis.

As indicated above, the population standard deviation is considered an appropriate measure to examine diversity. The raw data is converted to proportions in order to calculate the required parameter. This provides a measure of *diversity variability*, which is the term used in this paper.

The biological and environmental literature uses equal weighting of categories under each criterion. Two common measures of diversity are due to Hurlbert (1971) and Simpson (1949). The measure due to Hurlbert is sometimes referred to as an index of diversity.

From Tables 1 and 2 it will be observed that when the Diversity Variability (DV) measure is zero, the distribution of items in each category is equal. As this measure increases in magnitude an upper limit of 0.49 is achieved in the two category case. As the number of categories increase the upper limit is less close to 0.5. The diversity range will always produce a value in the range 0, perfect diversity, to 0.98 imperfect diversity.

Category A	Category B	Proportion Category A	Proportion Category B	Diversity Range	Diversity Variability	Diversity Heterogeneity	Simson's measure	Hurlbert's measure
50	50	0.50	0.50	0.00	0.00	0.50	0.49	0.51
40	60	0.40	0.60	0.20	0.10	0.40	0.52	0.48
30	70	0.30	0.70	0.40	0.20	0.30	0.58	0.42
20	80	0.20	0.80	0.60	0.30	0.20	0.68	0.32
10	90	0.10	0.90	0.80	0.40	0.10	0.82	0.18
1	99	0.01	0.99	0.98	0.49	0.01	0.98	0.02

Table 1: Diversity Measure with Two Categories

A *high* value for Simpson's measure indicates closeness to perfect diversity as defined in this paper. In the two-category case, perfect diversity is represented by a value of 0.49 whereas in the four-category case the value is 0.24. Both Simpson's and Hurlbert's measures produce results that are loosely related. For example Hurlbert's measure is very close to unity minus the Simpson measure. How close it is will depend on the number of categories. When the number of categories is relatively small this relationship holds.

Consider a criterion made up of two categories, perfect diversity would have one-half of the population in each category, whereas imperfect diversity would have all bar one of the population in one category and the remaining one of the population in the other category. As shown in Table 1, Hurlbert's measure gives a value 0.51 for perfect diversity and 0.02 for imperfect diversity in the two category case. The maximum value is different depending on the number of categories. For instance if there are four categories, as in Table 2, then the value for perfect diversity is 0.76 and the value for imperfect diversity is 0.06. So although a higher value indicates greater diversity, the value of the measure for perfect diversity varies depending on the number of categories.

Category A	Category B	Category C	Category D	Diversity Range	Diversity Variability	Simpson	Hurlbert
25	25	25	25	0.000	0.000	0.242	0.758
40	20	20	20	0.200	0.087	0.273	0.727
49	49	1	1	0.480	0.240	0.475	0.525
60	20	20	20	0.333	0.144	0.328	0.672
80	18	1	1	0.790	0.325	0.669	0.331
97	1	1	1	0.960	0.416	0.941	0.059

Table 2: Diversity Measure with Four Categories

When there are two categories, Simpson's measure is 0.98 for imperfect diversity and 0.49 for perfect diversity. In this case the lower the value the greater is the diversity. However, the value of perfect diversity is dependent on the number of categories. When there are four categories then the value for perfect diversity is 0.242 and the value for imperfect diversity is 0.941. With the *diversity variability* measure, the value of perfect diversity is always zero regardless of the number of categories and therefore it is always possible to make a meaningful comparison with perfect diversity as defined in this study. This is not the case with the measures employed by Hurlbert and Simpson.

A basic assumption underlying diversity measures is that, in a two-group population with all units in the first group to begin with, as the proportion in group one (e.g., detached houses) decreases the proportion in group two (e.g., apartments) increases, which results in the population becoming more diverse. An important point concerning diversity is that if the largest group is increasing at the fastest rate over time, relative to other groups, then diversity is decreasing. If, on the other hand, the largest group increases at the slowest rate, then diversity will increase over time.

Measures Discussed in the Financial Literature

There are three financial measures of diversity due to Rajan et al (2000), Scharfstein & Stein (2000) and Burch & Nanda (2003). Rajan et. al, (2000) looked at the returns of companies which were internally diversified, that is, they were involved in a number of different industries, and compared them with a portfolio of firms, each of which was active in a single market segment. They found that the greater the diversity the lower is the diversified firm's value relative to a portfolio of single-segment firms. Diversity is measured by obtaining the resource-weighted opportunities allocated to divisions within the conglomerate firms. A comparison is then made with a portfolio of non-diversified firms using Tobin's Q ratio[3].

Scharfstein and Stein (2000) used a diversity measure to examine the allocation of investment resources within conglomerate firms to determine if resource allocation was

3 Tobin's Q ratio represents the ratio of the market value of the firm to the replacement value of its assets.

more efficient using this type of structure compared with the external capital market. Because resources can flow towards inefficient divisions, within the conglomerate, this can lead to less efficient investment causing the firm's value to be less.

Burch and Nanda (2003) extended this analysis by using spinoff data to reconstruct the diversified firm after the spinoff in order to assess the aggregate improvement in value and attribute improvements in value to changes in diversity. Using direct measures of diversity and industry based proxies this research supports the general findings of Rajan et al and Scharfstein and Stein, that diversity in investment opportunities is a source of value loss for diversified firms

The measure employed by Rajan et al (2000) is a type of coefficient of variation in asset weighted investment opportunities. Scharfstein and Stein (2000) and Burch and Nanda (2003) employ a similar, but simpler measure. All three measures weight the categories under each criterion. In the current study these methods are considered to be inappropriate since any weighting schema would be arbitrary and a direct weighting method is likely to be more consistent with the nature of housing data.

Diversity in Data Mining

Hilderman and Hamilton (1999) consider diversity measures in a data mining context. They describe five properties required of diversity measures. The population standard deviation satisfies all five properties, a minimum value, a maximum value, skewness, invariance and transfer. The following notation DV(x_1, x_2) is a convenient method of representing the calculation of the diversity measure when there are two categories. The number of elements in the first category is represented by x_1 and in the second category by x_2. DV is an abbreviation for diversity variability and represents the population standard deviation of the proportions in each category.

The population standard deviation satisfies the minimum value property, since its value will be zero whenever the distribution is the same in each category. For example, DV(5, 5) = DV(10, 10, 10) = DV(50, 50, 50, 50) = 0.

The population standard deviation satisfies the maximum value property, since it achieves its maximum value when all categories, except for one, contain exactly one element. For example, if there are ten elements and two categories the maximum value is DV(9, 1), similarly DV(28, 1, 1) when there are thirty elements and three categories and DV(197, 1, 1, 1) when there are two hundred elements and four categories.

The population standard deviation satisfies the skewness property since a smaller number of categories distributed as unevenly as possible will give a higher value than a larger number of categories distributed as unevenly as possible. For example, DV(199, 1) = 0.4950 > DV(197, 1, 1, 1) = 0.4244.

The population standard deviation satisfies the invariance property since for a given number of categories its value will be the same if the distribution is the same regardless of the way the elements are distributed. For example DV(197, 1, 1, 1) = DV(1, 197, 1, 1) = DV(1, 1, 197, 1) = DV(1, 1, 1, 197) = 0.4244.

The population standard deviation satisfies the transfer property since if elements are moved from a smaller category to a larger category the value of the measure increases. For example, DV(20, 15, 10, 5) < DV(20, 18, 10, 2).

3. Calculation of Diversity Measures

3.1 Calculation of Diversity Measures for Type of Dwelling

Calculation of Diversity Variability (DV) is demonstrated using the following example. Consider the four types of dwellings identified in the 1991 census. The number of each type is shown in column 2 of Table 3.

Type of Dwelling	Number	Proportion	Deviation from mean	Squared Deviation
Separate Dwelling	4,533,946	0.7922	0.54223	0.2940
Town Houses	452,827	0.0791	-0.17088	0.0292
Flats	680,803	0.1190	-0.13104	0.0172
Other Dwellings	55,439	0.0097	-0.24031	0.0578
Sum	**5,723,015**	**1.0000**		**0.3981**
Mean		**0.25**		
Diversity Variability (Standard Deviation of proportions)				**0.3155**

Table 3: ABS Census Data for the Housing Stock in Australia in 1991

Also shown in Table 3 is the proportion of the total in each category. The proportion of Separate Dwellings is 4,533,946/5,723,015 or 0.7922. The mean proportion is 0.25. Deviations from the mean, for each proportion, are squared and summed to obtain a value of 0.3981. The DV measure is then calculated by dividing this value by the number of categories, 4 in this example, and taking the square root of the result. The DV value is 0.3155, suggesting that diversity is imperfect. Perfect diversity, that is items are evenly distributed across categories, in this case would require a DV value of zero. From Table 2 it will be observed that a DV value of 0.433 is obtained when nearly all items are contained in a single category and there are a total of four categories. It will be apparent from Table 3 that almost 80 percent of the data is contained in the category Separate Dwelling, which is reflected in the diversity measure.

The results for the 1996 and 2001 census data are shown in Table 4. These results indicate that the value of the DV measure displays greater diversity in 2001 that in the previous census. The Simpson and Hurlbert measures support this result. In other words the Australian housing stock, based on dwelling type, became more diverse across this period.

Census Date	Separate Houses	Town Houses	Flats	Other Dwellings	Diversity Variation	Simpson	Hurlbert
1996	4,911,909	517,181	816,704	134,024	0.303	0.616	0.384
2001	5,327,309	632,176	923,139	134,274	0.297	0.602	0.398

Table 4: Diversity Measures for 1996 and 2001 Census Data

Clearly this would be a consequence of a faster increase in town house and flat development compared to separate houses. Between the 1991 and 1996 census, the increase in separate dwellings was 8.34%, whereas the increase in town houses and apartments over the same period were 14.21% and 19.96% respectively. Between the 1996 and 2001 census, the increase in separate dwellings was 8.46%, whereas the increase in town houses and apartments over the same period were 22.24% and 13.03% respectively.

Included in the census data are two additional Dwelling Types, dwellings that are occupied but their type is not stated and dwellings that are unoccupied. Variation in the proportion of occupied dwellings compared to not occupied dwellings using the diversity measures illustrated earlier is summarised in Table 5 for each of the last three censuses that were available.

Census Date	Total Occupied	Not Occupied	Diversity Range	Diversity Variability	Simpson	Hurlbert
1991	5765576	597667	0.81215	0.406075	0.829794	0.170206
1996	6546072	679165	0.81200	0.406001	0.829674	0.170326
2001	7072202	717877	0.81570	0.407847	0.832679	0.167321

Table 5: Diversity Measures for Dwelling Occupation

Table 5 indicates that, based on the criteria of occupancy, the three censuses (1991, 1996 and 2001) diversity has changed very little. For these categories this is not surprising as the proportions of occupied and not occupied remained fairly constant. The 2001 census shows a slight reduction in diversity, which may reflect an anomaly arising from the time the data was collected. It may also reflect the beginning of a trend in the proportion of dwellings occupied relative to not occupied. In fact between the 1991 and 1996 census, the increase in occupied dwellings was 13.54%, whereas the increase in unoccupied dwellings was 13.64%. The corresponding changes from 1996 to 2001 are an increase of 8.04% for occupied and 5.7% for unoccupied. Rates of change are declining.

3.2 Calculation of Diversity Measures for Type of Occupancy

Data type of occupancy rather than housing stock was available from the last six censuses. This meant that trends, if present, could be more easily identified. For the population Type of Occupancy, five categories were reported and the data is given in Table 6A. Again perfect diversity, in other words 20% in each category, would have given a DV Measure of 0.000. The DV Measures for Type of Occupancy over the last six censuses are provided in Table 6B.

The diversity measures for the six censuses are plotted in Figure 1.

Census Year	Owned Outright	Still Purchasing	Public Renter	Private Renter	Other Tenure
1976	1,308,240	1,436,580	202,860	840,420	351,900
1981	1,550,108	1,540,770	228,781	933,800	415,541
1986	1,981,434	1,602,783	280,098	1,058,148	264,537
1991	2,323,295	1,556,550	328,306	1,204,885	351,665
1996	2,689,344	1,656,480	331,296	1,533,056	285,824
2001	2,842,944	1,874,080	318,240	1,633,632	403,104

Table 6A: ABS Census Data by Type of Occupancy

Year	Div. Range	Div. Variability	Simpson	Hurlbert
1976	0.298	0.119	0.271	0.729
1981	0.283	0.118	0.269	0.731
1986	0.331	0.133	0.289	0.711
1991	0.346	0.131	0.286	0.714
1996	0.370	0.139	0.297	0.703
2001	0.357	0.135	0.290	0.710

Table 6B: Diversity Measures by Type of Occupancy for data in Table 6A

Census Year	Owned Outright	Still Purchasing	Public Renter	Private Renter	Other Tenure
1976					
1981	18.49%	7.25%	12.78%	11.11%	18.08%
1986	27.83%	4.02%	22.43%	13.32%	-36.34%
1991	17.25%	-2.88%	17.21%	13.87%	32.94%
1996	15.76%	6.42%	0.91%	27.24%	-18.72%
2001	5.71%	13.14%	-3.94%	6.56%	41.03%

Table 6C: Percentage change between Successive Censuses

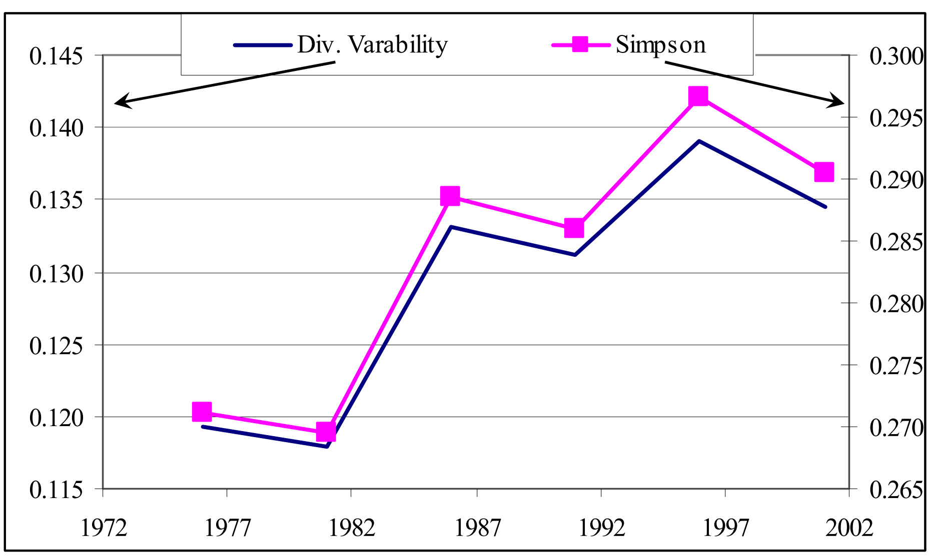

Figure 1: Plot of DV and Simpson Measures for Type of Occupancy

Diversity Variability (DV) is represented by the axis on the left of Figure1 and Simpson's measure by the axis on the right. Clearly there is an increase in DV Measure for Type of Occupancy indicating that there is less diversity in this population across time. Simpson's measure produces an almost identical result, allowing for the difference in scale.

The change in Outright Ownership from the 1976 to 1981 census resulted in an increase by this category of 18.49%. Between the 1981 and 1986 census this category increased by 27.83% and thereafter, while continuing to increase between successive censuses, did so at a slower rate in each successive census. The categories of Public Renter and Other Tenure are small relative to the other three categories and do not greatly influence the diversity measure. Growth rates for the category Private Renter has increased steadily (refer Table 6C). The categories, Outright Ownership and Private Renting, increased at a faster rate than the category Still Purchasing. This explains the decrease in diversity over time.

Census data was also available for occupancy based on the categories of Owners and Renters. These data, together with the diversity measures are provided in Table 7. Owners include those that own outright and those that are still purchasing. Renters include public renters and private renters. The results for the two groups over all the censuses are shown in Table 7. Some idea of the trend in the DV measure may be observed from Figure 2.

Census Date	Total Owners	Total Renters	Total	Diversity Range	Diversity Variability	Simpson	Hurlbert
1976	2,744,820	1,043,280	3,788,100	0.449	0.225	0.601	0.399
1981	3,090,878	1,162,581	4,253,489	0.453	0.227	0.603	0.397
1986	3,584,217	1,338,246	4,922,463	0.456	0.228	0.604	0.396
1991	3,879,845	1,533,490	5,413,335	0.433	0.217	0.594	0.406
1996	4,345,824	1,864,352	6,210,176	0.400	0.200	0.580	0.420
2001	4,717,024	1,951,872	6,668,896	0.415	0.207	0.586	0.414

Table 7: SD Measures for Owners and Renters only

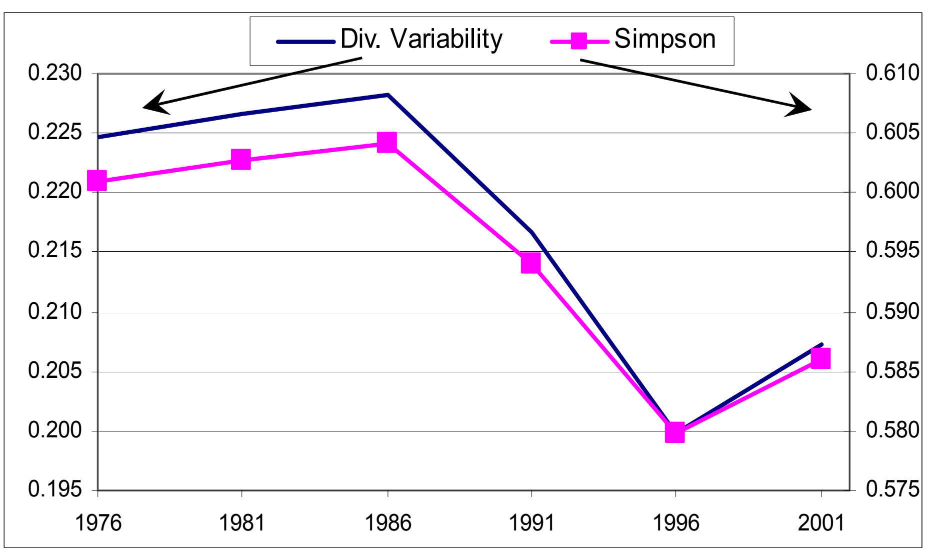

Figure 2: Plot of DV and Simpson Measures for Owners and Renters only

It is clear from Figure 2 that the general trend in both measures, Diversity Variability (DV) and Simpson, is downwards, in other words there is a general increase in diversity over time. Note that in this case there are two categories and if each category contained 50%, then the DV value would be zero and the Simpson measure would equal 0.49 (refer Table 1). For the most recent census the trend has been reversed slightly. This latter movement is a consequence of a greater percentage increase in the larger category (Owners) relative to the smaller category (Renters). Table 8 contains the percentage changes. Previously the smaller category tended to increase at a faster rate than the larger category, which explains the increase in diversity throughout the earlier period.

Census	Owners	Renters
1976 - 1981	12.61%	11.44%
1981 - 1986	15.96%	15.11%
1986 - 1991	8.25%	14.59%
1991 - 1996	12.01%	21.58%
1996 - 2001	8.54%	4.69%

Table 8: Percentage Changes Over Time for Owners and Renters

The census data indicates a strong upward trend in renters, relative to owners. However, this trend was not dominant in the census data contained in Table 6A, despite using the same data. The data in Table 7 aggregates the two ownership types, outright owners and still purchasing, and the two rental categories, public and private. Rates of increase within the categories for which a diversity measure is calculated has implications for the magnitude of the diversity index. Applications of diversity measures should take account of such outcomes.

Conclusion

Measures of diversity may be used in a range of applications. They are particularly useful in identifying trends in data containing many sub groups or categories. The intention of this paper is to draw attention to these measures and examine their suitability for application in the housing market. A significant amount of data now resides in both government and private databases that could be analysed using these measures.

It is the view of the authors that diversity measures may be used to examine trends in the housing market over time. The measures may also be employed to make comparisons across capital cities and between neighborhoods. They may also be employed to make meaningful comparisons between countries where it is appropriate to do so.

Appendix – Diversity Measures

Diversity Range = Proportion in largest group – proportion in smallest group

Diversity Variability (Standard Deviation) $= \sqrt{\dfrac{\sum_{i=1}^{k}\left(Group_i - mean\right)^2}{k}}$ k = number of categories

Diversity heterogeneity = 1- proportion attributed to largest group

Simpson $= \dfrac{\sum_{i=1}^{k} n_i(n_i - 1)}{n(n-1)}$

n_i = the number in each category
n = total number in all categories
k = number of categories

Hurlbert $= \dfrac{n}{n-1} * \left(1 - \sum_{i=1}^{k} p_i^2\right)$

p_i = the proportion in each category
n = total number in all categories
k = number of categories

Dissimilarity Index $DI = \dfrac{1}{2}\sum_{i=1}^{n}\left|\dfrac{R_{1i}}{R_1} - \dfrac{R_{2i}}{R_2}\right|$

where R_j ~ population of dwelling type j in regional area
Rji ~ population of dwelling type j in neighborhood i
n ~ number of neighborhoods in regional area.

The dissimilarity index is useful when comparing two groups (e.g. dwelling types) across neighborhoods within a particular geographic region. The index indicates the percentage of one grouping that should be added to the second grouping to produce an even distribution across neighborhoods. DI ranges from 0, evenly distributed, to 100, representing maximum differences (total differentiation).

References

Australian Bureau of Statistics, various publications.

Burch, T. and Nanda, V., (2003), Divisional diversity and the conglomerate discount. Journal of Financial Economics (70) 69-98.

Chambers, J., Cleveland, W., Kleiner, B. and Tukey, P., (1983), Graphical methods for data analysis. Wadsworth.

Hilderman, R. and Hamilton, H., (1999), Heuristic measures of interestingness. Proceedings of the Third European Conference on the Principles of Data Mining and Knowledge Discovery, 233-241.

Hurlbert, S., (1971), The non-concept of species diversity: a critique and alternative parameters. Ecology (52) 577-586.

Montgomery, D., (1985), Introduction to Statistical Quality Control. John Wiley and Sons, New York.

Rajan, R., Servaes, H. and Zingales, L., (2000), The cost of diversity: the diversification discount and inefficient investment. Journal of Finance (55) 35-80.

Scharfstein, D. and Stein, J., (2000), The dark side of internal capital markets: divisional rent-seeking and inefficient investment. Journal of Finance (55) 2537-2564.

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