North American Accounting Studies
Vol. 5, No. 1
2022
PP 1-17

Validating Financial Statement Comparability Assessment in Non-Profit Firms

Andrew Brajcich, Gonzaga University
Daniel Friesner, North Dakota State University

Abstract

In a recent manuscript, Brajcich and Friesner (2022) proposed a new methodology to assess financial comparability in not-for-profit organizations. Their approach utilizes entropy-based information theory, and thus requires few prior assumptions about the formation and implications of financial statement comparability. This manuscript assesses the practical utility of the Brajcich and Friesner (2022) methodology by comparing its results to those generated by nonhierarchical cluster analysis. The analysis was conducted using balance sheets drawn from Washington State critical access hospitals in 2019, using a similar set of hospitals and an identical set of variables outlined in Brajcich and Friesner (2022). We find that our entropy-based results closely mimic those from Brajcich and Friesner (2022), suggesting that the method produces internally consistent results. Additionally, non-hierarchical cluster analysis and the entropy-based methodology produce consistent results, but only when a sufficient number of peer groups is assumed in the non-hierarchical cluster analysis.

Keywords: financial statement comparability, entropy, spreadsheet modelling, not-for-profit firms, cluster analysis

Introduction and Literature Review

A major challenge in the management and regulation of organizations with a not-for-profit tax status is that many of the outcomes typically used to assess the financial success of a for-profit organization (return on equity, net income, debt-equity ratios, profitability ratios, etc.) either do not apply, or do not have the same interpretation, within the context of a not-for-profit firm. More fundamentally, financial accounting metrics are predicated upon a comparison of the organization's performance against a pre-defined standard; that is, a benchmarking exercise. The benchmark is typically based on the performance of a set of peers whose accounting choices are similar to those of the firm being benchmarked. The latter is known as "financial statement comparability". As noted in Chapter 3 of the (Amended) "Statement of Financial Accounting Concepts No. 8", the Financial Accounting Standards Board (FASB) considers comparability as an enhancing qualitative characteristic of all financial statements (FASB 2018).

In organizations with a for-profit tax status, all firms have similar short and long run objectives, which are to maximize profits and shareholder wealth, respectively. But in organizations with not-for-profit tax status, the objectives of the organization, which ultimately lead to higher expenditures and minimal net income after operating (and other) expenses, are not immediately apparent. These objectives may differ substantially from firm to firm, which in turn affects resource allocations, and ultimately determines the information contained on financial statements. The more disparate are not-for-profit firms' objectives, the less comparable are their accounting statements. Thus, it is difficult to identify the subset of not-for-profit firms whose operational and accounting choices are similar, and whose financial accounting statements are truly comparable. These circumstances preclude meaningful benchmarking of those few financial indicators that can be calculated and interpreted consistently across not-for-profit organizations.

In a recent manuscript, Brajcich and Friesner (2022) (hereafter, BF (2022)) proposed a new benchmarking methodology to assess financial comparability in not-for-profit organizations. Their model draws upon information entropy theory, and thus requires few prior assumptions about the formation and implications of financial statement comparability. Unlike the regression-based comparability methods developed by De Franco, Kothari, and Verdi (2011), and subsequently used by numerous other researchers[1], BF (2022) does not rely on the magnitude of firm earnings as the primary indicator of financial statement comparability. This allows their methodology to be applied to firms with not-for-profit tax status. Alternatively, Hoitash, Hoitash, Kurt, and Verdi (2018) and Young and Zeng (2015), defined financial statement comparability using the aggregate perceptions of analysts and corporate board members to define a particular company's peers. Their methodology can be applied to firms regardless of tax status; however, their methods are limited by the use of perceptual data to define peers. The BF (2022) methodology is specifically intended to utilize non-perceptual data culled directly from a firm's accounting statements, and thus is fundamentally different from Hoitash et al. (2018) and Young and Zeng (2015).

A major drawback of the BF (2022) methodology is that few alternative methodologies exist to assess comparability across not-for-profit firms, which use similar financial statement data and require few assumptions about the nature of the firms being assessed. In the absence of an alternative (but equally flexible and parsimonious) methodology, it is difficult to determine whether the BF (2022) methodology i) characterizes comparability in an appropriate manner; ii) whether differences in comparability metrics are due to actual differences in comparability or are simply an artifact of the firm-specific and data-related assumptions underlying the comparability measures; or iii) a combination of i) and ii). As a result, the practical utility of the BF (2022) methodology has not been established.

While no (flexible and parsimonious) alternative methods in the finance and accounting literatures currently exist to appropriate assess the validity and reliability of the BF (2022) methodology, the multivariate analysis literature does offer several viable empirical techniques to assess comparability in a general sense. For example, non-hierarchical cluster analysis is a nonparametric, iterative, flexible methodology that is commonly used to classify decision making units into similar or dissimilar groups based on data collected across these decision making units on several common variables (Hair, Black, Babin, Anderson, and Tatham 2006, pp. 553-627). Non-hierarchical cluster analysis has been used in the literature to identify commonalities across various decision making units (for a set number of variables) in a host of different public and private financial decisions (Bassetto and Kalatzis 2011; Snarr, Friesner and Underwood 2012; Kramaric, Bach, Dumicic, Zmuk, and Zaja 2018; Nakagawa, Kawahara, and Ito 2020). If the BF (2022) methodology produces reasonable results, any comparability assessments generated by their methodology should be generally consistent with those produced by non-hierarchical cluster analysis. As a corollary, if (presumably minor) differences exist in the comparability assessments, those differences should be consistent with corresponding differences in the mathematical underpinnings of each technique.

[1] See Do (2021); Jiu, Hu, and Liu (2021), Choi, Choi, Myers, and Ziebart (2019), Chen and Gong (2018), Chen, Collins, Kravet, and Mergenthaler (2018), Qingyuan and Lumeng (2018), Imhof, Seavey, and Smith (2017), and Kim, Li, Lu, and Yu (2016), among others, for various studies assessing financial statement comparability and linking comparability to other financial phenomena in for-profit firms.

This manuscript assesses the practical utility of the BF (2022) methodology by applying both the BF (2022) methodology and non-hierarchical cluster analysis to the same data set. If the BF (2022) methodology produces reasonable results, their comparability assessments should be generally consistent with those produced by non-hierarchical cluster analysis. As a corollary, the original BF (2022) methodology used data drawn from critical access hospitals in Washington State during the 2017 fiscal year. This analysis uses balance sheet data drawn from critical access hospitals in Washington State during the 2019 fiscal year, using the identical set of variables outlined in BF (2022). This not only allows for a comparison of comparability estimates between the BF (2022) and cluster analysis methodologies, but also a comparison of results generated by the BF (2022) methodology for similar hospitals over time. In this way, it is possibly to assess both the external and internal consistency of the BF (2022) methodology. Our main findings are twofold. First, the results of the current analysis are very similar to those presented in Brajcich and Friesner (2022). This suggests that their methodology produces internally consistent results. Second, we find that, non-hierarchical cluster analysis and the entropy-based methodology produce consistent results, but only when the researcher assumes a sufficiently large number of peer groups in the non-hierarchical cluster analysis.

Empirical Methodology

Financial Statement Comparability

The BF (2022) methodology assesses financial statement comparability by imputing a common size balance sheet. Using BF (2022)'s terminology, suppose we have a variable Q that embodies these financial characteristics, and which can take one of $j = 1,\ldots,J$ possible values. Each possible realized value for q_j can be expressed theoretically as a probability, and empirically as a relative frequency. That is, q_j, for $j = 1,\ldots,J$, where $0 \leq q_j, \leq 1$, and $\sum_{j=1}^{J} q_j = 1$. Within the context of the common size balance sheet approach, Q represents the distribution of firm resources (or resource flows) across a mutually exclusive and collectively exhaustive set of categories. For example, Q might comprise the distribution of a firm's assets, and the q_js represent the proportion of total firm assets in one of j specific asset categories.

The variable Q is assessed relative to a benchmark (p), whose theoretic and empirical properties are analogous to those of Q. More specifically, we define p_j, $j = 1,\ldots,J$, where $0 \leq p_j, \leq 1$, and $\sum_{j=1}^{J} p_j = 1$. If $q_j = p_j$ for every $j = 1,\ldots,J$, then the firm's realized values for Q are "optimal". Deviations between one or more q_js and the p_js indicate that the firm's distribution for Q is sub-optimal, with greater deviations indicating greater divergence from the benchmark. Since the benchmark p is not directly observed, it must be imputed. BF (2022) draw from information theory (Golan, Judge, and Miller, 1996; pp. 11-12) and impute p using the concept of minimum cross-entropy:

$$min_{p_1 \ldots p_J} CE = \sum_{j=1}^{J} p_j log_2\left(\frac{p_j}{q_j}\right) \tag{1}$$

The premise of the minimum cross-entropy formulation in (1) is to select the benchmark distribution (p) in a manner that assumes as little (a priori) as possible about p (i.e., that the p_js are uniformly distributed), while simultaneously ensuring that the benchmark distribution mimics the observed data (Q) as closely as possible.

Several attributes of the minimum cross-entropy formulation are noteworthy. First, entropy is additive in nature. Hence, it is straightforward to extend the base formulation in (1) to any number of sets of financial characteristics. BF (2022), for example, conduct a benchmarking exercise using two sets of financial characteristics (assets and liabilities), each of which is mutually exclusive and collectively exhaustive:

$$min_{p_1 \ldots p_J, \rho_1 \ldots \rho_K} CE = \sum_{j=1}^{J} p_j log_2\left(\frac{p_j}{q_j}\right) + \sum_{k=1}^{K} \rho_k log_2\left(\frac{\rho_k}{r_k}\right) \qquad (2)$$

where r_k, $k = 1,\ldots,K$ describes the proportion of total liabilities allocated to across one of $k = 1,\ldots,K$ categories, with $0 \le r_k \le 1$, and $\sum_{k=1}^{K} r_k = 1$; ρ_k, $k = 1,\ldots,K$ is the benchmark for each of the k categories, with $0 \le \rho_k \le 1$, and $\sum_{k=1}^{K} \rho_k = 1$; and the remaining variables are as defined previously. Friesner and Brajcich (2022) extend this methodology to incorporate a wider array of financial and non-financial characteristics (4 sets of characteristics instead of 2). They also demonstrate how to test hypotheses about the formation of financial statement comparability. Friesner, Brajcich, Friesner, and McPherson (2022) also extend the formulation in (2) to build a comparability model for critical access hospitals that assesses whether specific department-level productive activities (and expenses they generate from production) impact financial statement comparability across hospitals. Their model includes 4 sets of variables: firm assets, firm liabilities, expenses in the hospital's pharmacy cost center, and expenses in the hospital's medical laboratory cost center.

Second, it is straightforward to extend the cross-entropy formulation to allow for industry wide benchmarking. Given a sample of $i = 1,\ldots,n$ firms in an industry, and assuming two sets of financial characteristics (Q and R, respectively) and benchmarks (p and ρ, respectively), equation (2) can be extended to allow for a single, industry-wide benchmark, to be estimated, against which all firms in the industry can be assessed:

$$min_{p_1 \ldots p_J, \rho_1 \ldots \rho_K} \sum_{i=1}^{n} CE_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{J} p_j log_2\left(\frac{p_j}{q_{ij}}\right) + \sum_{k=1}^{K} \rho_k log_2\left(\frac{\rho_k}{r_{ik}}\right)\right) \qquad (3)$$

The current manuscript utilizes the formulation in (3) to impute the benchmark common size balance sheet and assess comparability across firms. Assessment can occur by i) assessing each individual q_{ij} and r_{ik} against its corresponding benchmark; or ii) assessing overall firm comparability by assessing the relative magnitude of CE_i for a given firm against other firms in the data set. BF (2022) suggest transforming the CE_i variable using the z-score (denoted as $z(CEi)$) and identifying firms with z-score values beyond 2 standard deviations from the mean (i.e.,. z-scores greater than 2 in absolute value) as non-comparable to their peers. Given the current study's objectives, the individual asset and liability benchmarks can collectively be used, along with the overall CE_i scores, to assess the internal consistency of the BF (2022) methodology over time. We focus solely on the overall CE_i scores (rather than the individual asset and liability category benchmarks) to assess external consistency between the BF (2022) and non-hierarchical cluster analysis.

Non-hierarchical Cluster Analysis

This manuscript employs cluster analysis as an alternative means of assessing financial comparability. The advantage of cluster analysis lies in its flexibility and extensive use in the financial literature (Loretz and Moore 2013). For a given set of variables and observations, an iterative clustering algorithm defines cluster membership based on a pre-defined set of criteria, and in doing so defines "similarity" in financial statement information across firms. Consistent with the multivariate analysis literature, this analysis defines similarity using a process known as "Ward's method". Within the context of this analysis, Ward's method creates clusters of not-for-profit firms, or peer groups of related firms, by trying to ensure that very little variation in financial statement information exists across firms in the same cluster (Hair et al. 2006, pp. 588-

589).[2] Should a cluster be comprised of a small group of firms within a population, cluster analysis requires that the sample contains a sufficient number of firms to accurately and precisely represent that small cluster in the sample. Thus, while cluster analysis may not require specific statistical properties of the data, the data should be representative of the population as a whole (Hair et al. 2006, pp. 570-571).

The flexibility inherent in cluster analysis requires an additional assumption. Because cluster analysis merely identifies inter-relationships between variables and/or observations, it does not impose any restrictions on the selection of those variables. Put differently, cluster analysis will attempt to identify similar groups based on any set of variables collected across any set of observations, regardless of whether or not those variables and observations reflect firm performance and/or the industry in which those firms operate. For cluster analysis to provide meaningful results, the researchers must use the existing literature to provide strong theoretical and applied support for the selection of variables and observations used in the analysis (Friesner, McPherson, Schibik, and Brajcich 2018).

Assuming the previous criteria are met, the researcher must also determine which form of cluster analysis to employ (Hair et al. 2006, pp. 553-627). The analysis may be hierarchical or non-hierarchical in nature. Hierarchical cluster analysis attempts to assess the degree of similarity or dissimilarity of a group of variables across a given set of firms (and observations). The goal of hierarchical cluster analysis is to learn more about the inter-relationships that exist between variables, rather than similarities between firms. Non-hierarchical cluster analysis presumes that the researcher has a set of variables that appropriately characterizes firm activity, and seeks to assess the degree of similarity across firms based on these variables. For non-hierarchical analysis, the researcher must also specify the number of clusters one expects in the data, and cluster analysis assigns each firm to one of these clusters. Because the researcher does not know the exact number of clusters, but can reasonably identify a range within which the true number of clusters lies, it is common to repeat the analysis several times, each time specifying a different number of clusters within the expected range. If the cluster analysis produces reasonable results, the results should be relatively consistent across each of these replications. Because we have a well-defined set of financial variables and seek to understand the degree of similarity across firms, this analysis employs non-hierarchical cluster analysis.

Assessing the Internal Consistency of the BF (2022) Methodology

If the BF (2022) methodology produces internally consistent estimates of financial statement comparability, then applying the same methodology to similar groups of firms over multiple time periods should also (in the absence of major economic shocks or major regulatory changes) produce similar benchmarks over time. This provides a simple means to assess internal consistency. We operate under the null hypothesis that the individual asset and liability category benchmarks generated using 2019 data are jointly equal to their corresponding benchmarks generated using 2017 data. To operationalize the test, the 2019 data and the optimal asset and liability category benchmarks from 2019 are used to calculate an overall CE_i score, as outlined in equation (3). We calculate an alternative version of CE_i, denoted as CE_i', by repeating this calculation, but replacing the optimal 2019 benchmark proportions with those reported by BF (2022) using 2017 data. Under the null hypothesis, CE_i should equal CE_i' (and similarly, $z(CE_i)$ should equal $z(CE_i')$). If CE_i, and CE_i' are sufficiently different in magnitude, the null hypothesis

[2] Ward's method specifically attempts to define clusters of peer groups by minimizing the within sum of squared error in each cluster of firms (or within each peer group).

is rejected (and similarly for $z(CE_i)$ and $z(CE_i^{'})$). A matched sample t-test (using a 5 percent significance level) is applied under the null hypothesis that the population means for the two variables are equal.

Comparing the Results across Methodologies and Assessing External Consistency

Each methodology assesses comparability and assigns peers based on a fundamentally different set of assumptions. However, when applied to the same dataset, if both methodologies produce consistent estimates of financial statement comparability, each methodology should consistently assign an individual firm to the same peer group. To identify peer groups across a set of financial statement information, the BF (2022) methodology uses the z-score of the overall CE_i variable ($z(CE_i)$), which is approximately normally distributed with a mean of zero and a standard deviation of 1. As noted above, firms whose CE_i values are further from zero in absolute value are more likely to be considered outliers, and less comparable to other firms in the dataset. The primary measure of the non-hierarchical cluster analysis is the assignment of each firm to one of a fixed number of peer groups. Thus, non-hierarchical cluster analysis produces a discrete (nominal) variable that takes a finite number of values based on the number of peer groups assumed by the researcher. For example, if the researcher assumes the existence of three clusters (or peer groups), the primary outcome is a discrete variable taking values of 1, 2, or 3.

The characteristics of each primary outcome variable provide a simple and straightforward means to assess whether the two methodologies produce consistent results. If the two methodologies generate consistent results, firms in each discrete peer group produced by non-hierarchical cluster analysis should exhibit a compact, well-defined, and unique range of CE_i scores produced under the BF (2022) methodology. If the two methodologies produce very different characterizations of comparability and peer group formation, there should be no relationship between the magnitude of $z(CE_i)$ (and, by extension, CE_i) and the discrete peer group to which it is assigned by non-hierarchical cluster analysis. Because the $z(CE_i)$ score is an approximately continuous, normally distributed variable, and the peer group designation is a discrete nominal variable, standard one-way analysis of variance (ANOVA) can be used to operationalize this assessment. Within the context of this study, one-way ANOVA operates under the null hypothesis of no mean differences in $z(CE_i)$ scores across the cluster groups generated by non-hierarchical cluster analysis. That is, the null hypothesis assumes that the two methodologies generate fundamentally different characterizations of financial statement comparability. Rejecting the null hypothesis indicates systematic mean differences in $z(CE_i)$ (and CE_i) scores across cluster groupings, which suggests that the two methods generate similar characterizations of financial statement comparability. All hypothesis tests utilize 5 percent significance levels.

Data

The data used in this study are an updated replication of BF (2022). More specifically, BF (2022) use balance sheet data drawn from critical access hospitals operating in Washington State during the year 2017. The current analysis uses analogous data drawn from Washington State critical access hospital financial statements in 2019. This not only allows for a comparison of the BF (2022) methodology to one conducted by non-hierarchical cluster analysis, but also a comparative assessment of the consistency of the methodology over time.[3]

All hospitals operating in the state are required to submit a full set of accounting statements to the Washington State Department of Health on an annual basis, as well as certain financial information on a quarterly basis.[4] For the 2019 reporting cycle, 93 hospitals provided information to the Department of Health, of which 31 held critical access hospital status. The BF (2022) analysis specifically utilizes data drawn from the balance sheets submitted by these critical access hospitals. Following BF (2022), total assets were aggregated into four mutually exclusive and collectively exhaustive categories: cash, non-cash current assets, net property, plant, and equipment, and other non-current assets. Liabilities were also aggregated into four mutually exclusive and collectively exhaustive categories: accounts payable, accrued expenses, other current liabilities, and long-term debts. Five critical access hospitals were eliminated from the analysis for reporting missing, mis-measured, or otherwise unusable data in one of these categories. Thus, we are left with a working sample of 26 observations. We note in passing that the number of critical access hospitals reporting useable data in 2019 is slightly lower than the 36 used in BF (2022). This is likely due to the COVID-19 pandemic, which affected Washington State in early 2020, and impacted the ability of hospitals to report, and the Washington State Department of Health to disclose, hospital financial information.

Results

Table 1 contains the variable names, variable descriptions, and basic descriptive statistics for the data used in the analysis. There is a high degree of consistency between the statistics reported in Table 1 and those reported in BF (2022). For example, at the mean, 2019 critical access hospitals reported approximately 17% of assets in cash, 28% in non-cash current assets, 44% in net property, plant, and equipment, and 12% in other non-current assets. These same percentages in 2017 (as reported by BF (2022)) were 16%, 26%, 46%, and 13%, respectively. Similarly, in 2019, the proportions of liabilities in accounts payable, accrued expenses, other current liabilities, and long-term debt are 8%, 12%, 10%, and 70%. BF (2022) report that in 2017, these same percentages were 11%, 15%, 10%, and 65%, respectively.

Table 1
Variable Names, Definitions, and Descriptive Statistics in 2019

Panel A: Original Data			
Variable	Description	Mean	Standard Deviation
Cash	Assets held in cash	$7,382,392.46	$13,397,045.89
Ncashnc	Assets held in non-cash current assets	$11,167,128.92	$9,393,395.41
Net PPE	Assets held in all net PPE	$20,367,736.04	$20,159,147.53

[3] We note in passing that we also replicated the current analysis using data drawn from 2018 and obtained similar results to those reported in this manuscript. Further information on the 2018 analysis is available from the lead author upon request.

[4] The Department of Health aggregates (and, where necessary, cleans) the data and makes them publicly available on its website: https://doh.wa.gov/data-statistical-reports/healthcare-washington/hospital-and-patient-data/hospital-financial-data.

Other Non-Current Assets	Assets held in all other non-current assets	\$4,967,237.31	\$6,928,405.10
Tassets	Total assets	\$43,884,494.73	\$39,183,699.00
Payable	Liabilities held in accounts payable	\$1,475,610.38	\$1,425,414.40
Accrued	Liabilities held in accrued expenses	\$2,391,233.31	\$2,670,780.98
Othcl	Liabilities held in other current liabilities	\$1,492,376.96	\$1,240,132.39
PLtdebt	Liabilities held in long term debt	\$16,741,483.65	\$16,509,430.55
Tliab	Total Liabilities	\$22,100,704.31	\$19,792,635.46
Panel B: Proportional Data			
Pcash	Proportion of assets held in cash	0.17	
Pncashnc	Proportion of assets held in non-cash current assets	0.28	
Pnppe	Proportion of assets held in net PPE	0.44	
Pothnc	Proportion of assets held in all other non-current assets	0.12	
Ppayable	Proportion of liabilities held in accounts payable	0.08	
Paccrued	Proportion of liabilities held in accrued expenses	0.12	
Pothcl	Proportion of liabilities held in other current liabilities	0.10	
Pltdebt	Proportion of liabilities held in long term debt	0.70	

Table 2 contains the results of the comparability analysis using the BF (2022) methodology. At the optimum, hospitals should attempt to carry 11.7% of assets in cash, 30.1% in non-cash current assets, 49.5% in net property, plant, and equipment, and 8.7% in other non-current assets. The corresponding optimal proportions of liabilities are 6.8% in accounts payable, 12.8% in accrued expenses, 6.9% other current liabilities, and 73.5% in long-term debt. These values are, not surprisingly, quite similar to those of BF (2022), who found the optimal asset percentages in 2017 to be 11.8% of assets in cash, 27.8% in non-cash current assets, 51.7% in net property, plant, and equipment, and 8.7% in other non-current assets. The corresponding optimal proportions of liabilities in 2017 are 8.9% in accounts payable, 14.5% in accrued expenses, 8.0% other current liabilities, and 68.6% in long-term debt.

Table 3 contains the raw and z-score transformed CE_i values using the 2019 optimal proportions, as well as the raw and CE_i' transformed variables using the optimal 2017 proportions identified by BF (2022). The magnitudes of the estimates across the two sets of calculations vary only slightly. Moreover, the paired t-test fails to reject the null hypothesis of no mean differences between the two variables. This suggests (but does not conclusively prove) that the entropy-based measure of financial statement comparability produces internally consistent results.

Table 2
Analysis of Individual CAHs

CAH	Pcash	Pncashnc	Pothnc	Pnppe	Ppay.	Paccrued	Pothcl	Pltdebt	CE_i	$z(CE_i)$	Variable	Optimal Pj
1	0.125	0.290	0.018	0.567	0.121	0.133	0.234	0.512	0.303	-0.314	Pcash	0.117
2	0.122	0.408	0.214	0.256	0.056	0.132	0.030	0.782	0.251	-0.489	Pncashnc	0.301
3	0.118	0.410	0.000	0.472	0.329	0.194	0.202	0.276	0.598	0.660	Pothnc	0.087
4	0.017	0.550	0.048	0.385	0.069	0.200	0.074	0.657	0.347	-0.170	Pnppe	0.495
5	0.198	0.382	0.056	0.364	0.131	0.125	0.004	0.740	0.294	-0.344		
6	0.457	0.254	0.097	0.193	0.043	0.183	0.104	0.670	0.543	0.479	Ppayable	0.068
7	0.486	0.198	0.011	0.306	0.026	0.216	0.281	0.477	0.864	1.540	Paccrued	0.128
8	0.549	0.180	0.016	0.255	0.213	0.206	0.064	0.516	0.832	1.435	Pothcl	0.069
9	0.084	0.250	0.227	0.439	0.052	0.168	0.016	0.764	0.183	-0.711	Pltdebt	0.735
10	0.075	0.135	0.266	0.524	0.065	0.000	0.066	0.869	0.074	-1.074		
11	0.251	0.175	0.040	0.534	0.062	0.091	0.068	0.780	0.163	-0.779	Min.Entropy	10.353
12	0.055	0.402	0.136	0.407	0.142	0.175	0.128	0.555	0.192	-0.682		
13	0.091	0.178	0.071	0.661	0.032	0.035	0.093	0.840	0.233	-0.545		
14	0.049	0.246	0.261	0.444	0.050	0.130	0.122	0.698	0.199	-0.660		
15	0.105	0.349	0.066	0.480	0.168	0.134	0.170	0.528	0.173	-0.744		
16	0.111	0.204	0.134	0.551	0.038	0.051	0.225	0.687	0.231	-0.553		
17	0.398	0.110	0.051	0.441	0.056	0.164	0.024	0.756	0.429	0.103		
18	0.335	0.196	0.020	0.448	0.074	0.159	0.117	0.651	0.290	-0.358		
19	0.072	0.260	0.005	0.662	0.087	0.067	0.017	0.828	0.399	0.002		
20	0.138	0.196	0.187	0.478	0.023	0.138	0.066	0.773	0.130	-0.887		
21	0.036	0.211	0.513	0.239	0.089	0.047	0.021	0.843	0.779	1.261		
22	0.106	0.243	0.146	0.504	0.157	0.138	0.015	0.690	0.154	-0.809		
23	0.229	0.242	0.066	0.463	0.052	0.130	0.171	0.647	0.133	-0.876		
24	0.298	0.107	0.000	0.595	0.012	0.040	0.050	0.898	0.366	-0.107		
25	0.004	0.486	0.002	0.508	0.022	0.044	0.061	0.873	0.952	1.832		
26	0.001	0.494	0.341	0.164	0.023	0.116	0.049	0.812	1.242	2.791		

Table 3
Assessing the Internal Consistency of the BF (2022) Methodology

	Using 2019 Optimal Benchmarks		Using 2017 Optimal Benchmarks		CE_i-$CE_i^{'}$		$z(CE_i)$ -$z(CE_i^{'})$	
CAH	CE_i	$z(CE_i)$	$CE_i^{'}$	$z(CE_i^{'})$	T-Statistic	Prob.	T-Statistic	Prob.
1	0.303	-0.314	0.247	-0.504	-1.065	0.297	< 0.001	>0.999
2	0.251	-0.489	0.315	-0.293				
3	0.598	0.660	0.478	0.211				
4	0.347	-0.170	0.371	-0.120				
5	0.294	-0.344	0.348	-0.190				
6	0.543	0.479	0.568	0.492				
7	0.864	1.540	0.838	1.329				
8	0.832	1.435	0.775	1.133				
9	0.183	-0.711	0.220	-0.588				
10	0.074	-1.074	0.045	-1.132				
11	0.163	-0.779	0.168	-0.749				
12	0.192	-0.682	0.161	-0.769				
13	0.233	-0.545	0.277	-0.410				
14	0.199	-0.660	0.204	-0.637				
15	0.173	-0.744	0.125	-0.881				
16	0.231	-0.553	0.242	-0.519				
17	0.429	0.103	0.426	0.052				
18	0.290	-0.358	0.264	-0.450				
19	0.399	0.002	0.437	0.087				
20	0.130	-0.887	0.162	-0.766				
21	0.779	1.261	0.847	1.357				
22	0.154	-0.809	0.150	-0.805				
23	0.133	-0.876	0.123	-0.890				
24	0.366	-0.107	0.432	0.071				
25	0.952	1.832	1.061	2.022				
26	1.242	2.791	1.361	2.949				

Table 4 contains the results of the non-hierarchical cluster analysis, as well as the overall entropy-based comparability measure, for each of the critical access hospitals in the data set. This facilitates an initial, relative assessment of consistency across the two approaches to assessing comparability. Examining this table yields several inferences. First, the entropy-based measure indicates that all but 6 of the hospitals in the sample are extremely comparable, with overall comparability measures within one standard deviation of the mean (i.e., $z(CE_i)$ values between -1 and +1). Moreover, of the 6 hospitals whose comparability metrics exceed one in absolute value, only 2 hospitals (hospitals 25 and 26) have overall comparability measures in excess of 1.75 standard deviations from the mean. Only one of these hospitals (hospital 26) is more than two standard deviations from the mean, indicating that the hospital is truly distinct from the others in the data set. These results are consistent with BF (2022), who found that only 2 hospitals out of the 36 they analyzed were non-comparable to their peers. This result is not surprising given the relatively similar geographic and institutional characteristics shared by these firms.

Another set of important inferences can be drawn from the final columns in the table, which use non-hierarchical cluster analysis to classify hospitals into peer groups. Given the small number of hospitals included in this study, cluster analysis examined peer group membership assuming a small range of clusters. More specifically, the results in table 4 assume an analysis using 2 clusters or groups of peers, a 3 cluster model (3 peer groups), a 4 cluster model (4 peer groups), and a 5 cluster model (5 peer groups). All four models give similar results, with the 3, 4, and 5 peer group cluster solutions effectively disaggregating the 2 cluster solution into smaller subsets.[5] This trend also implies that the cluster analysis is analyzing the data in an internally consistent manner.

A final inference that can be drawn from Table 4 pertains to the consistency in assessing comparability across the two methods. The hospitals with entropy-based comparability scores outside of the [-1,+1] interval do not appear to be consistently assigned to the same peer groups as other hospitals who also have scores outside of the [-1,+1] range. It is worth noting in passing, however, that hospitals 21 (which has z-score value of 1.261) and 26 (which has the largest z-score value in the sample, 2.791) are consistently assigned to the same non-hierarchical cluster analysis peer groups. Concomitantly, hospital 25, whose $z(CE_i)$ measure of 1.832 is the second largest in the sample, is always assigned to the same peer groups as hospitals 22 ($z(CE_i)$: -0.809) and 5 ($z(CE_i)$: -0.344).

The third inference drawn from Table 4 is imprecise in that it does not involve hypothesis testing and the conclusions drawn do not attach probability values to those specific conclusions. To address this concern, Table 5 applies one-way ANOVA to assess whether specific cluster groups have systematic (i.e., statistically significant) differences in entropy-based comparability scores. Table 5 Panel A assesses differences between the two methods using the cluster analysis that assumes the existence of only 2 peer groups. The F-test shows no significant difference in the z-score between the two cluster groups. Thus, if there are only two peer groups, cluster analysis and the entropy-based information theory measure do not provide consistent

---

[5] Note that the cluster number is irrelevant; only firms belonging to the same cluster is relevant when interpreting the results. So, for example, hospitals 1 and 2 are assigned to cluster peer group 2 in the 2 cluster model. In the 3 cluster model, they are assigned to peer group 1, in the 4 cluster model they are assigned to peer group 4, and in the 5 cluster model they are assigned to peer group 2. When interpreting the results, the primary inference is that these two hospitals fall into the same peer group, regardless of how many clusters are assumed to exist.

Table 4
Comparison of the BF (2022) and Cluster Analysis Results

CAH	pcash	pncashnc	pothnc	pnppe	ppayable	paccrued	pothcl	pltdebt	$z(CE_i)$	Cluster2	Cluster3	Cluster4	Cluster5
1	0.125	0.290	0.018	0.567	0.121	0.133	0.234	0.512	-0.314	1	3	3	3
2	0.122	0.408	0.214	0.256	0.056	0.132	0.030	0.782	-0.489	2	2	1	1
3	0.118	0.410	0.000	0.472	0.329	0.194	0.202	0.276	0.660	1	3	3	3
4	0.017	0.550	0.048	0.385	0.069	0.200	0.074	0.657	-0.170	1	2	3	4
5	0.198	0.382	0.056	0.364	0.131	0.125	0.004	0.740	-0.344	2	1	4	4
6	0.457	0.254	0.097	0.193	0.043	0.183	0.104	0.670	0.479	1	3	2	5
7	0.486	0.198	0.011	0.306	0.026	0.216	0.281	0.477	1.540	1	3	2	5
8	0.549	0.180	0.016	0.255	0.213	0.206	0.064	0.516	1.435	1	3	2	5
9	0.084	0.250	0.227	0.439	0.052	0.168	0.016	0.764	-0.711	2	2	1	4
10	0.075	0.135	0.266	0.524	0.065	0.000	0.066	0.869	-1.074	2	1	4	2
11	0.251	0.175	0.040	0.534	0.062	0.091	0.068	0.780	-0.779	2	1	4	2
12	0.055	0.402	0.136	0.407	0.142	0.175	0.128	0.555	-0.682	1	3	3	4
13	0.091	0.178	0.071	0.661	0.032	0.035	0.093	0.840	-0.545	2	1	4	2
14	0.049	0.246	0.261	0.444	0.050	0.130	0.122	0.698	-0.660	2	2	1	4
15	0.105	0.349	0.066	0.480	0.168	0.134	0.170	0.528	-0.744	1	3	3	3
16	0.111	0.204	0.134	0.551	0.038	0.051	0.225	0.687	-0.553	2	1	4	2
17	0.398	0.110	0.051	0.441	0.056	0.164	0.024	0.756	0.103	2	1	2	2
18	0.335	0.196	0.020	0.448	0.074	0.159	0.117	0.651	-0.358	1	3	2	2
19	0.072	0.260	0.005	0.662	0.087	0.067	0.017	0.828	0.002	2	1	4	2
20	0.138	0.196	0.187	0.478	0.023	0.138	0.066	0.773	-0.887	2	1	4	2
21	0.036	0.211	0.513	0.239	0.089	0.047	0.021	0.843	1.261	2	2	1	1
22	0.106	0.243	0.146	0.504	0.157	0.138	0.015	0.690	-0.809	2	1	4	4
23	0.229	0.242	0.066	0.463	0.052	0.130	0.171	0.647	-0.876	1	3	4	2
24	0.298	0.107	0.000	0.595	0.012	0.040	0.050	0.898	-0.107	2	1	4	2
25	0.004	0.486	0.002	0.508	0.022	0.044	0.061	0.873	1.832	2	1	4	4
26	0.001	0.494	0.341	0.164	0.023	0.116	0.049	0.812	2.791	2	2	1	1

results. Table 5, Panels B and C assess differences between the two methods using the cluster analysis with a 3 peer group solution, and a 4 peer group solution, respectively. Again, the F-test shows no significant difference between cluster group membership and the z-score of the entropy based comparability measure.

Lastly, Table 5, Panel D conducts the same analysis under the assumption of a cluster analysis model with 5 peer groups. The peer groups identified in Table 5, Panel D are analogous to those identified in Panels A through C of Table 5. However, there are now statistically significant differences across the cluster peer groups and the entropy-based comparability scores. Given this finding of statistical significance, it is interesting to refer back to Table 4 to examine the specific hospitals (and their associated financial statement characteristics) that are assigned to each of these five peer groups. Peer groups 1 and 5 contain small groups of hospitals with much larger entropy based comparability scores relative to groups 2 through 4. In particular, peer group 1 includes three hospitals, one of which is hospital 26, which has the largest entropy based comparability score. The second is hospital 21, which also has one of the larger, positive entropy based comparability score. The third hospital in this peer group (hospital 2) has a liability structure that matches the other two hospitals. Peer group 5 contains three hospitals, including hospitals 7 and 8, which carry entropy-based comparability scores in the range of 1.540 and 1.435. The third hospital in the peer group (hospital 6) has an asset structure that is similar to the other two hospitals. Peer group 2 consists of hospitals with either extremely small positive or moderate negative entropy based comparability values (i.e., $z(CE_i)$ values ranging from 0.103 to -1.074). Most hospitals in this peer group have $z(CE_i)$ values between -0.107 and -0.887. These hospitals tend to have slightly higher proportions of their assets allocated to net property, plant and equipment. Most hospitals in peer group 2 also have liabilities that are disproportionately allocated to long term debt. Peer group 3 consists of 3 hospitals (hospitals 1, 3, and 15) which exhibit different entropy based comparability measures, but which have very comparable values across all of the asset categories and several (accrued expenses and other current liabilities), not all liability categories. More specifically, the hospitals in peer group 3 exhibit very different proportions of liabilities in accounts payable and long term debt. Peer group 4 consists of all remaining firms, and appears to be an "other" category of firms. Like peer group 2, hospitals in peer group 4 tend to have most of their assets allocated to physical infrastructure and liabilities allocated to long term debts. However, they tend to have a fundamentally different distribution of other assets and liabilities compared to those in peer group 2.

Table 5
One-Way Analysis of Variance

Variable	Cluster	Obvn.	Mean	Std. Dev.	F-Statistic	Prob.
Panel A: Two Cluster/Peer Group Solution						
$z(CE_i)$	1	10	0.097	0.885	0.147	0.704
	2	16	-0.061	1.089		
Panel B: Three Cluster/Peer Group Solution						
$z(CE_i)$	1	11	-0.28744	0.799534	0.858	0.437
	2	6	0.337053	1.407544		
	3	9	0.126619	0.933191		
Panel C: Four Cluster/Peer Group Solution						

$z(CE_i)$	1	5	0.439	1.549	1.779	0.181
	2	5	0.640	0.829		
	3	5	-0.250	0.563		
	4	11	-0.376	0.806		
Panel D: Five Cluster/Peer Group Solution						
$z(CE_i)$	1	3	1.188	1.641	4.246	0.011
	2	10	-0.507	0.407		
	3	3	-0.133	0.720		
	4	7	-0.221	0.933		
	5	3	1.151	0.584		

Conclusions

The primary objective of this manuscript is to present a case study that compares the results generated through BF (2022)'s entropy-based information theory to those generated by non-hierarchical cluster analysis. As a corollary, the case study used the same variables and an extremely similar data set to that employed in the original BF (2022) study, which also allows for an initial assessment of the internal consistency of their methodology. The primary findings of the study are twofold. First, the results of the current analysis (which utilize data drawn from Washington State critical access hospitals in 2019) closely match those generated in the original BF (2022) study, which use an analogous sample of hospitals in 2017. This implies that their method produces internally consistent assessments of financial statement comparability among similar samples of firms over time.

Second, the BF (2022) method produces results that are similar to those generated by cluster analysis. However, an important caveat to this conclusion is that the cluster analysis must assume a sufficient number of peer groups to allow those with extreme entropy-based comparability methods (which are identified as non-comparable) to end up in their own peer group. This conclusion is highly consistent with the assumptions and mathematical principles underlying both techniques. The entropy measure assumes as little a priori knowledge as possible about which firms are, or are not, comparable. Put slightly differently, firms are assumed to be comparable unless sufficient empirical information exists to over-rule this assumption. This effectively means that the BF (2022) methodology identifies firms who are outliers. Concomitantly, non-hierarchical cluster analysis requires the research to assume a set number of clusters or peer groups at the start of the analysis, and the analysis populates those pre-defined clusters/peer groups given the financial data selected by the researcher. Thus, non-hierarchical cluster analysis only distinguishes peer groups to the extent that the research *expects* there to be fundamentally distinct peer groups. For the two methods to provide consistent results, the research must specify a sufficient number of peer groups to allow the cluster solution to place the non-comparable outliers identified by the BF (2022) methodology in the same or similar groups.

The primary implications of this study are also threefold. First, since the BF (2022) methodology produces comparability metrics that, under certain conditions, are consistent with non-hierarchical cluster analysis, as well as previous applications of the technique, there is evidence to suggest that the BF (2022) methodology has practical utility. Second, while the BF (2022) methodology has been described as a method to assess financial statement comparability, it is more appropriately characterized as a technique to eliminate firms that are non-comparable to other firms in a data set. Thus, it is most appropriately applied to situations where some prescreening of firms in a dataset has occurred to ensure that these firms have at least rough commonalities. Otherwise, the methodology may produce "false negative" results by failing to identify firms that may be non-comparable to their peers. Third, the conditions under which the two methods produce consistent results are critically contingent on the ability of the researcher to select an appropriate number of clusters/peer groups for the non-hierarchical cluster analysis. In practice, this may or may not be likely, or even feasible. Since the BF (2022) methodology does not require this assumption, it provides unique information in the assessment of financial statement comparability. Researchers are advised to use both techniques in tandem to fully assess comparability in their data.

While interesting and impactful conclusions were generated through the completion of this study, several study limitations exist which should be addressed in future research. First, both the current study and the BF (2022) study examine financial statement comparability in critical access hospitals. These hospitals are highly regulated, and as such may exhibit a higher degree of homogeneity across firms than not-for-profit firms operating in other industries. As a result, the entropy-based information methods may have greater difficulty differentiating between comparable and non-comparable firms. In a highly homogenous sample of hospitals, there may be few non-comparable firms to begin with, which is why i) only 1-2 firms were found to be non-comparable to their peers; and 2) there was little correlation between the entropy-based on non-hierarchical cluster analysis methods when a small number of peer groups is assumed. Applications of one or both methods to other types of inherently less homogenous firms with not-for-profit tax status may yield disparate results. A second limitation is that the study was conducted using a small data set. As noted previously, small datasets limit the number of possible clusters that can be assumed. Replication of this work with much larger data sets, which allows the researcher to assume a larger number of clusters in the non-hierarchical cluster analysis, may yield very different inferences from what was reported in this manuscript. Third, the study authors chose to use financial variables (drawn exclusively from firm balance sheets) defined in a previous study to assess the internal consistency of the BF (2022) methodology. However, the balance sheet is only one of many different financial statements upon which comparability may be characterized. Replications of this analysis that use a broader array of financial variables and/or non-financial variables may generate results that differ from, and improve upon, those reported in this manuscript.

References

Bassetto, C., and A. Kalatzis. 2011. Financial distress, financial constraint and investment decision: Evidence from Brazil. *Economic Modelling* 28 (1-2): 264-271.

Brajcich, A., and D. Friesner. 2022. Assessing financial statement comparability using spreadsheet modelling. *Journal of Theoretical Accounting Research* 18 (1): 87-113.

Chen, A., and J.J. Gong. 2018. Accounting comparability, financial reporting quality, and the pricing of accruals. *Advances in Accounting* 45 (1). http://dx.doi.org/10.2139/ssrn.3455830

Chen, C., D. Collins, T. Kravet, and R. Mergenthaler. 2018. Financial statement comparability and the efficiency of acquisition decisions. *Contemporary Accounting Research* 35 (1): 164-202.

Choi, J., S. Choi, L. Myers, and D. Ziebart. 2019. Financial statement comparability and the informativeness of stock prices about future earnings. *Contemporary Accounting Research* 36 (1): 389-417.

De Franco, G., S. Kothari, and R. Verdi. 2011. The benefits of financial statement comparability. *Journal of Accounting Research* 49 (4): 895-931.

Do, T. 2021. Financial statement comparability and corporate debt maturity. *Finance Research Letters* 40: Article 101693. https://doi.org/10.1016/j.frl.2020.101693

Financial Accounting Standards Board. 2018. Statement of accounting concepts number 8. Conceptual framework for financial reporting chapter 1, the objective of general purpose financial reporting, and chapter 3, qualitative characteristics of useful financial information. https://fasb.org/page/PageContent?pageId=/standards/concepts-statements.html

Friesner, D., and A. Brajcich. 2022. Do non-financial characteristics impact financial statement comparability? *Journal of Theoretical Accounting Research* forthcoming.

Friesner, D., A. Brajcich, K. Friesner, and M. McPherson. 2022. *Assessing Similarity in Production from Financial Accounting Statements using Comparability Analysis.* Working paper, North Dakota State University.

Friesner, D., M. McPherson, T. Schibik, and A. Brajcich. 2018. Identifying peers in international tax competition. *Applied Economics Letters* 25 (9): 584-587.

Golan, A., G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* New York, NY: John Wiley and Sons.

Hair, J., W. Black, B. Babin, R. Anderson, and M. Tatham. 2006. *Multivariate Data Analysis, 6th Edition*. Upper Saddle River, NJ: Prentice Hall Publishers.

Hoitash, R., U. Hoitash, A. Kurt, and R. Verdi. 2018. *An Input-based Measure of Financial Statement Comparability*. Working paper, SSRN. http://dx.doi.org/10.2139/ssrn.3208928

Imhoff, M., S. Seavey, and D. Smith. 2017. Comparability and cost of equity capital. *Accounting Horizons* 31 (2): 125-138.

Jiu, L., S. Hu, and Y. Liu. 2021. Does financial statement comparability mitigate corporate frauds in an emerging market? Evidence from China. *Asia-Pacific Journal of Accounting and Finance* forthcoming. https://doi.org/10.1080/16081625.2021.1976229

Kim, J., L. Li, L. Lu, and Y. Yu. 2016. Financial statement comparability and expected cash risk. *Journal of Accounting and Economics* 61 (2-3): 294-312.

Kramarić, T., M. Bach, K. Dumičić, B. Žmuk, and M. Žaja. 2018. Exploratory study of insurance companies in selected post-transition countries: Non-hierarchical cluster analysis. *Central European Journal of Operations Research* 26: 783–807.

Loretz, S., and P. Moore. 2013. Corporate tax competition between firms. *International Tax and Public Finance* 20 (5): 725-752.

Nakagawa, K., T. Kawahara, and A. Ito. 2020. Asset allocation strategy with non-hierarchical clustering risk parity portfolio. *Journal of Mathematical Finance* 10: 513-524. https://www.scirp.org/journal/paperinformation.aspx?paperid=103344

Qingyuan, L., and W. Lumeng. 2018. Financial statement comparability and corporate tax avoidance. *China Journal of Accounting Studies* 6 (4): 448-473.

Snarr, H., D. Friesner, and D. Underwood. 2012. Evaluating evolutionary changes in state TANF policies. *Applied Economics Letters* 19 (17): 1753-1758.

Young, S., and Y. Zeng. 2015. Accounting Comparability and the Accuracy of Peer-Based Valuation Models. *The Accounting Review* 90 (6): 2571–2601.