A NOTE ON THE USING OF ACCOUNTING DATABASES

Miklos A. Vasarhelyi[*]

and

David C. Yang[**]

December 1994 Revised[***]

[*] KPMG Peat Marwick Professor, Faculty of Management, Rutgers University, Newark, NJ 07102.

[**] Associate Professor, School of Accountancy, College of Business Administration, University of Hawaii at Manoa, Honolulu, HI 96822.

[***] The authors are grateful for the assistance of V. Soybel and E. Faillace. The comments and suggestions of Messrs. Anthony A. Phillips of the FASB, Katherine Brown of Standard & Poor's Compustat Services, Inc. and, William Close of Value Line Data Services and Professors D. H. Bao, W. Baber, C. W. Bastable, W. Beaver, B. Bublitz, T. Harris, J. G. San Miguel, T. J. Mock, J. Ohlson, G. Shillinglaw, M. Stone, and B. N. Srinidhi were greatly appreciated.

A NOTE ON THE USING OF ACCOUNTING DATABASES

ABSTRACT

This paper identifies potential data problems of using accounting databases. To examine data errors, two commonly used accounting databases -- Value Line and Compustat are compared in their qualitative and quantitative features. Data is examined using seven variables over a period of eleven years. Differences found in the data for 1981 are further analyzed by comparing sample data to figures directly drawn from financial statements. Substantial data differences are found. Most of which are attributable to definitional discrepancies and others to direct measurement error. For example, 39.5% of the depreciation figures and 23.2% of the inventory numbers were discrepant by more than 1% of the absolute value of the measure.

The paper also comments on the selection and usage of accounting databases and discusses shortcomings that should be expected in using accounting databases. Finally, recommendations are presented for dealing with these problems to preparers of databases as well as standard setters.

A NOTE ON THE USING OF ACCOUNTING DATABASES

1. INTRODUCTION

Previous research [San Miguel, 1977; Rosenberg and Houglet, 1974; Bennin, 1980; Stone and Bublitz, 1984] has shown that data errors tend to be a problem in the large, machine-readable bases of financial data. San Miguel [1977] examined R & D information in Compustat (CMP) and found 30% of the 256 data points to be discrepant. Rosenberg and Houglet [1974] compared error rates using the Compustat Industrial Tape and the CRSP Monthly Return Tape and found no major errors in CRSP but ten errors out of a possible 6,036 in CMP. Bennin [1980] showed that CMP had improved its data collection over time.

Rosenberg & Houglet [1974] and San Miguel [1977] suggested that when multiple, computerized databases contain similar information, the data should be matched to verify the accuracy of the data. Such a comparison is the most effective and least expensive way of screening for data error. For this study, selected historical cost information in CMP and Value Line (VL) are matched, compared and a subset is verified against the original data in the corporate annual reports. This paper examines the accuracy of selected items in the CMP and VL Databases[1] partially covering the 11 years from 1971 through 1981. These two accounting databases were chosen because they are commonly used by financial professionals [Makin, 1984].[2]

---

[1] The Compustat database is distributed by Standard & Poor's Compustat Services, Inc. An annual industrial tape, dated 8/12/82, is used in this study. The Value Line Database is distributed by Value Line Data Services of Arnold Bernhard & Co., Inc., New York, New York. An annual tape, dated 9/8/82 is used in this study.

[2] Institutional Investor, in conjunction with LINK Resources Corp., a New York market research firm surveyed hundreds of financial professionals, including securities, portfolio managers, pension officers, investment bankers, retail brokers, chief financial officers, treasurers cash managers and risk managers, through both mailed questionnaires and extensive follow-up interviews. Out of more than 100 databases named by respondents to this survey, the following three were most used: 1) Compustat, 2) Value Line, and 3) Dow Jones News/Retrieval.

The purpose of this study is to identify potential data problems of using accounting databases, and discusses shortcomings that should be expected in using accounting databases. Recommendations are also presented for dealing with these problems to preparers of accounting databases as well as standard setters.

2. COMPUSTAT VS. VALUE LINE DATABASES

Both CMP and VL contain accounting information, Table 1 describes the key contextual differences between the two databases used in this study.

TABLE 1. COMPARING THE DATABASE CONTENTS

	VALUE LINE	COMPUSTAT
# of Companies	1700	6000
Balance Sheet	similar	similar
Income Statement	similar	similar
Statement of Changes	less detail	more detail
Quarterly Information	less detail	more detail
Replacement Cost Data	Yes	No
Ratios	Yes	No
Estimates and Projections	Yes	No
Footnotes	Very limited	A system was developed
Years	1955 on	1964 on

Overall, CMP offers a larger sample of companies at a proportionately larger cost. Specific data content varies between the databases.

2.1 Variables, Sample and Metrics

The *The Accounting Review*, *The Journal of Accounting Research* and *The Journal of Accounting and Economics* were examined for the 1976-1981 period to obtain the frequency[3] of database usage by accounting researchers and choice of variables. Seven variables [total reported assets (ASSET), net sales (SALES), inventory (INVNT), net income before extraordinaries (INBET), current liabilities (LIABL), depreciation, depletion, and amortization (DEPRE), and gross plant (PLANT)] were found to be used most often in empirical studies. Their values were gathered, when available, for the years 1971 to 1981 and provide the sample for this paper.

A relative discrepancy measure D_{it} was used as the scale to compare values for each of the seven variables in the two databases. This measure is defined as:

$$D_{it} = \frac{|V_{itVL} - V_{itCMP}|}{V_{itCMP}}$$

where:

D_{it} = relative discrepancy for variable i in year t, i = 1, 2, ----, 7
V_{it} = value found for variable i in year t using VL or CMP, t = 1971, 1972, ----, 1981

Discrepancies smaller than 1% were ignored [Rosenberg & Houglet, 1974] to avoid confusing discrepancies with rounding.

2.2 Discrepancies

The Database merge led to a common sample of 1479 companies for 1981. Discrepancy categories and occurrences are reported in Table 2.

TABLE 2. COMPARISON OF COMPUSTAT AND VALUE LINE DATABASES - 1981
(1479 COMPANIES MATCHED)

DATA ITEM	ASSET		SALES		INBET		INVNT		DEPRE		LIABL		PLANT	
DISCREPS. NO.	#	%	#	%	#	%	#	%	#	%	#	%	#	%

[3] Ro [1980] and Ro [1981] were the only research studies that used the Value Line Data tape. In both studies Ro developed an initial list of potential control firms from the Value Line Data tape by checking ASR 190's $100 million materiality standard. CMP does not include ASR 190 data.

Equal or <=.01	1,431	96.7	1,331	90.0	1,262	85.3	1,136	76.8	894	60.5	1,305	88.2	1,190	80.5
>.01 but <=.05	14	1.0	47	3.2	39	2.6	9	.6	195	13.2	15	1.0	103	7.0
>.05 but <=.1	3	0.2	18	1.2	29	2.0	4	.3	95	6.4	6	0.4	26	1.7
More than .1	31	2.1	80	5.4	146	9.9	103	7.0	196	13.3	45	3.1	80	5.4
Miss. Values**	0	0.0	3	0.2	3	0.2	227	15.3	99	6.6	108	7.3	80	5.4
Discrep. Rate+	3.3%		10.0%		14.7%		23.2%		39.5%		11.8%		19.5%	
TOTAL	1,479	100	1,479	100	1,479	100	1,479	100	1,479	100	1,479	100	1,479	100

+ Discrepancy rate is defined as the total of discrepancy numbers in category 2 (>.01 but <=.05), 3 (>.05 but <=.1), 4 (>.1) and 5 (missing values) divided by the number of companies matched.

** Either database (or both) has missing data.

Out of 10,353 comparisons, 1,284 (12.4%) discrepancies were larger than 1% and 520 (5.02%) had missing fields. These numbers were much larger than expected and required further examination. Particularly striking were the large discrepancies in the depreciation, depletion, amortization, inventory and gross plant figures. A potential explanation to these discrepancies may lie in the fact that definitional differences among the databases are more substantial than the rather similar sample definitions found in the manuals.[4]

In order to examine the source of these data discrepancies a subsample of the 1981 data was drawn and compared to the original financial statements. As coding errors did not explain the full extent of the discrepancies in Table 2, the industry, the foreign currency, and the definitional factors were examined.

2.3 Sources of discrepancies[5]

2.3.1 Explainable Definitional Differences These discrepancies are attributable to three factors:

a. Foreign currency differences - For example, some VL coded data in Canadian dollars while CMP presented them in US dollars;

[4] A sample of the definition comparison from the manuals, a comparison of the different definitions for special industries and the table (analogous to Table 2) for the 1971-1980 period is available from the authors upon request. The results for the longer time period are very similar to the ones displayed in Table 2.

[5] VL and CMP were contacted for their specific coding rules and definitions. VL provided with detailed internal coding rules while CMP declined to provide these for all variables.

b. Industry factors - For the seven variables used in this study, VL and CMP have different definitions for the following industries: bank, savings & loan, insurance, finance, railroad, securities brokerage and utility;

c. Definitional factors - Within the same industry and the same variable, VL and CMP may have different definitions. For example, Income Before Extraordinary Items in VL is after Discontinued Operations while in CMP, it is before Discontinued Operations.

2.3.2. Unexplained Discrepancies There were two sources of unexplained discrepancies:

a. Non-disclosed coding rule differences - The intrinsic heterogeneity of financial statements makes their classification into pre-set categories a difficult task. The best one can hope for is a consistent and fine categorization with some degree of horizontal (across company) and vertical (time-series) comparability. If a discrepancy is found and not explained by definitional differences, it may be the product of a non-detected definitional difference or a data error. Non-detected definitional differences may be the product of poor manual documentation, coding standards that were never documented by the database originators, or the result of detected coding discrepancies that the source is not willing to make public.

Over a period of years, accounting standard changes and detected systematic coding mistakes lead to inevitable time-series heterogeneities. These "systematic inconsistencies" are more serious than pure data errors that result from coding and transcription deficiencies. The systematic inconsistencies would require a major recoding and reorganization of data, a procedure that is cumbersome and dangerous from the standpoint of data integrity. Data errors requires only the correction of specific points in the database.[6]

For example, let us examine the current liability figures. A strong argument can be made both for disclosure as stated as well as for having the current portion of long term debt added back to current liability when this is not done by the company. It is most likely that CMP originally adopted the "as stated" solution and later changed into the "restating" method. This is not stated anywhere in the manual, and if not detected by ourselves would have made this type of discrepancy an "unexplained difference." Currently, both CMP and VL state that the current portion of long term debt

[6] As a general policy neither VL nor CMP will restate past data which have subsequently been changed by the company due to an accounting restatement.

should be treated as a current liability. Four instances of non-compliance to this rule were detected in the data. This created a doubt whether the current portion of long term debt was impounded into current liabilities but not netted out of the total long term debt figure in certain cases.

Another example is that CMP includes 'equipment on rental' in gross plant, while 'dry hole costs' and 'impairment of unproved oil and gas properties' are included in depreciation, depletion and amortization. Neither of these occurrences are specified in CMP's users manual. From the coding instructions sent by Value Line Data Services,[7] we found that `other income' is excluded from net sales while `equity in earnings of unconsolidated subsidiaries' is excluded from SALES, neither of which are specified in VL's users manual. Appendix A lists the items which should be specified in the manuals and the variables to which they pertain.

The problem of an unclear definitions makes it difficult to separate definitional differences from coding errors. Variable definitions are more detailed in CMP than in VL. On the other hand, experienced investment analysts[8] do the VL coding and seem to be allowed more flexibility and judgment than their CMP counterparts.

b. Coding errors - Table 2 indicates the total assets field to be relatively free of error. This indicates a base figure of 3.3% for expected combined undetected discrepancies and coding errors. Table 3 displays the results of the analysis of the subsample where 200 companies were randomly selected from 1479 companies for examination of the 1981 coded data against their annual reports.

[7] CMP disclosed only internal coding guidelines for the depreciation, depletion and amortization fields.

[8] From Value Line Data Survey.

TABLE 3. COMPARISON OF COMPUSTAT AND VALUE LINE DATABASES AGAINST ANNUAL REPORTS - 1981 DISCREPANCIES SOURCE - BASED ON 200 COMPANIES

DATA ITEM	ASSETS	SALES	INBET	INVNT	DEPRE	LIABL	PLANT
NUMBER OF DISCREPS.	#	#	#	#	#	#	#
DISCREPANCY RATE	%	%	%	%	%	%	%
(1) EXPLAINED DEFINITIONAL DIFFERENCES:							
a. FOREIGN CURRENCY DIFFERENCES							
#	5	5	5	5	5	5	5
%	2.5	2.5	2.5	2.5	2.5	2.5	2.5
b & c. INDUSTRY CLASSIFICATION AND CODING & VARIABLE DEFINITION DIFFERENCES							
#	0	20	7	60	23	14	26
%	0	10	3.5	30	11.5	7	13
(2) UNEXPLAINED DEFINITIONAL DIFFERENCES:							
a & b. NON-DISCLOSED CODING RULE DIFFERENCES & CODING ERRORS							
BY: COMPUSTAT							
#	0	2	6	9	69	1	12
%	0	1	3	4.5	34.5	0.5	6
VALUE LINE							
#	0	1	4	13	9	2	7
%	0	0.5	2	6.5	4.5	1	3.5
VALUE LINE & COMPUSTAT							
#	0	0	0	0	6	0	4
%	0	0	0	0	3	0	2

Table 3 displays large unexplained discrepancies in the inventory and depreciation[9], depletion, amortization fields. These results warn database users about the need for careful examination of archival data coding rules.

2.3.3 Aggregate Error Effects To further evaluate the effects of detected differences, we compared the data by performing a paired-comparison T test. This comparison, after the exclusion of 26 companies with different currencies, can be observed in Table 4. Systematic differences in the net sales, inventories, gross plant, and depreciation, depletion and amortization were noted.

TABLE 4. PAIRED-COMPARISON T TEST OVERALL AND BY INDUSTRIES - 1981

[9] VL tends to collect depreciation figures from the statement of funds while CMP focuses on Tables 3 & 4 of the 10K.

INDUSTRY	VARS	ASSET	SALES	INBET	INVNT	DEPRE	LIABL	PLANT
OVERALL	Sample**	1,453	1,450	1,450	1,226	1,354	1,345	1,373
	_ (in millions)	-6.23	-23.81	-0.86	-6.77	-3.66	-2.38	-5.46
	t Value	-0.78	-2.53*	-1.32	-2.08*	-1.66+	-0.62	-5.71*
AGRICULTURE, FORESTRY & FISHING	Sample	5	4	4	5	4	5	5
	_	0.00	2.10	0.40	0.002	-0.10	0.00	-20.82
	t Value	nv	1.35	0.99	1.00	-1.73	nv	-1.46
MINING	Sample	83	81	81	80	81	83	83
	_	-0.14	10.48	-4.97	2.76	-25.06	0.38	-6.04
	t Value	-0.08	1.56	-1.16	1.51	-2.80*	0.45	-1.13
CONSTRUC.	Sample	22	22	22	22	22	21	22
	_	0.39	-5.02	1.03	9.97	0.42	0.00	-15.02
	t Value	1.00	-0.54	1.66	0.57	1.74+	0.00	-1.07
MANUFACT.	Sample	745	745	745	745	745	745	744
	_	1.32	13.36	-1.09	-0.27	-5.42	-6.70	-2.64
	t Value	1.27	2.34*	-0.99	-0.18	-1.40	-0.99	-4.00*
TRANSPORT., COMMUNIC. & OTHER PUBLIC UTILITIES	Sample	229	299	229	99	229	228	229
	_	6.16	0.23	1.33	-12.3	3.84	4.32	-10.04
	t Value	1.09	0.11	1.20	-2.54*	4.11*	1.22	-3.00*
WHOLESALE & RETAIL TRADE	Sample	133	133	133	131	133	132	133
	_	-0.02	-1.29	-0.48	-6.29	-0.14	0.06	-6.39
	t Value	-0.99	-0.98	-0.84	-0.99	-0.66	0.99	-2.30*
FINANCE, INSURANCE & REAL ESTATE	Sample	140	140	140	51	46	38	63
	_	-81.73	-325.42	-0.96	-120.56	4.68	20.19	-18.44
	t Value	-1.00	-3.66*	-2.32*	-1.69+	1.26	0.98	-1.75+
SERVICES	Sample	62	62	62	62	62	61	62
	_	0.08	7.11	-2.41	-5.66	0.8	0.00	-3.58
	t Value	1.00	1.82+	-1.32	-1.49	1.87+	0.57	-2.32*
OTHERS	Sample	34	34	34	31	32	32	32

INDUSTRY	VARS	ASSET	SALES	INBET	INVNT	DEPRE	LIABL	PLANT
	_	0.00	0.51	-0.13	0.06	-0.60	0.00	-2.02
	t Value	0.00	1.44	-0.95	0.98	-0.87	-1.00	-1.69

*: significant at the 5% level. _: Mean Difference (in millions)
+: significant at the 10% level. nv: no values
**: companies with missing value in either database was excluded.

A more precise view of the above data was obtained by using industry breakdowns also displayed in Table 4.[10] A substantial portion of the discrepancies are due to **definitional industry differences**, especially in transportation, communication & other utilities, finance, insurance and real estate. They add to the warning made earlier to the user that a careful scrutiny of the raw data to be used in research is essential.

3. METHODOLOGICAL IMPLICATIONS

3.1 The Effects of Different Data Bases on the Cross-Sectional Distributional Properties of Financial Ratios

This section examines the effect of the use of CMP and VL data bases on the cross-sectional distributional properties (mean, variance, skewness, kurtosis[11] and normality) of selected financial ratios. Deakin [1976] investigated the normality of the distributions of eleven commonly used financial ratios over the 1954 to 1972 period for all CMP manufacturing companies and concluded that the normality assumption was generally not tenable except for the debt/asset ratio. Beedles and Simkowitz [1978], in replicating work by McEnally [1974], demonstrated that a seemingly small error rate can have a great effect on findings, especially in studies using higher moments of distribution.

[10] The industry breakdown used in this table follows the one used by Goodman et al. [1982].

[11] The skewness is the third sample moment about the mean, divided by the variance raised to the 3/2 power. Meanwhile the kurtosis is the fourth sample moment about the mean, divided by the variance squared. These four moments are vital to statistical analysis.

Ten[12] of Deakin's 11 financial ratios were computed using CMP and VL. Like Rosenberg and Houglet, we calculated mean, variance, skewness and kurtosis for each ratio. Table 5 shows the four moments of ratios 3, 4, 6, 8 and 10 to be very similar for CMP and VL. However, ratios 1, 2 are significantly different. Consequently, if the same statistical techniques were applied to 1983's ratio 1, these would generate substantially different results contingent on the database used. Another interesting feature of these data is that the normality assumption for all ratios was accepted[13] at the 0.01 significant level which is not consistent with Deakin's results.

TABLE 5. FOUR MOMENTS OF 10 FINANCIAL RATIOS (1983 Data)

RATIO	DATABASE	SAMPLE	MEAN	VAR.	SKEWNESS	KURTOSIS
1. Curr. Assets / Sales	VL	891	.4695	2.1511	25.9040	720.8865
	CMP	891	.4330	0.3529	17.4787	383.6781
2. Quick Assets / Sales	VL	779	.3212	1.3563	22.6386	569.3801
	CMP	779	.2706	0.8262	7.1516	79.9830
3. Working Cap. / Sales	VL	888	.1657	0.0497	-3.5752	56.6779
	CMP	888	.1659	0.0500	-3.4101	54.7547
4. Curr. Assets / Curr. Liab.	VL	888	1.9949	1.3661	3.7012	28.6292
	CMP	888	1.9988	1.3694	3.6674	28.1866
5. Quick Assets / Curr. Liab.	VL	765	1.2247	0.9724	6.4380	70.2165
	CMP	765	1.2544	0.9745	6.3307	68.4258
6. Curr. Assets / Total Assets	VL	891	.4352	0.0506	0.0194	-1.0504
	CMP	891	.4348	0.0503	0.0171	-1.0434
7. Quick Assets / Total Assets	VL	781	.2635	0.0213	0.9780	1.2323
	CMP	781	.2699	0.0212	0.9674	1.1264
8. Working Cap. / Total Assets	VL	888	.2013	0.0337	0.2203	0.4577
	CMP	888	.2016	0.0338	0.2126	0.4510
9. Net Inc. / Total Assets	VL	994	.0412	0.0036	-1.8769	12.2459
	CMP	994	.0411	0.0041	-2.2962	14.7125

[12] The Cash flow/Total debt ratio was eliminated for the unclear definition of cash flow in VL. This definition is needed for computing the same ratio on the CMP.

[13] A modified version of Kolmogorov-Smirnov D-statistic was used to test the normality assumption.

10. Total Debt / Total Assets	VL	451	.6326	0.0245	0.6683	0.5519
	CMP	451	.6335	0.0246	0.6568	0.5404

3.2 Methodological Implications and the Effect of Research Results

The results of this study indicate that:

1. There is no unanimity among databases on the treatment of specific accounting items, either among items (for example, the treatment of interest income) or among industries.
2. Users must examine carefully the data definitions in the database manuals.
3. There is a certain level of coding error[14] in most databases and fields. It is most likely that these "coding errors" will not substantially affect the results.
4. *Undetected definitional biases* are systematic in nature and bias results. This is particularly dangerous if several different industries are being compared, as the biases may impact the populations being studied differently.
5. Users will find different data fields and fineness of data among databases.
6. When selecting a database the user must consider both its availability and the tradeoffs between content, support [Yang and Vasarhelyi, 1987] and reliability of the data in the database.

4. DISCUSSION AND RECOMMENDATIONS

Substantial differences exist in the interpretation of accounting variables both between databases and across industries. Users must be aware of the definitional discrepancies in the database they will be using. A series of noteworthy definitional differences were found and are summarized in Appendix A. Definitional discrepancies are not mere inconsistencies among accounting databases. In fact, they are a reflection of the latitude available within GAAP classifications as well as a reflection of inter-industry differences and measurement difficulties. Companies which

[14] Out of the sample of "unexplained" discrepancies, we selected and sent to CMP and VL 70 companies whose data had to be reconciled. The outcome was that about 10% of these were found clearly in error and were to be fixed in the database. In addition about 20% of the other discrepancies were deemed as judgmental coding issues that could have been treated one way or another. The major portion of the companies being reconciled had their discrepancies attributed to some internal, and not documented in the manual, coding standard. Finally, about 10-20% of these discrepancies could be attributed to miscellaneous causes.

provide databases should : 1) on a periodic basis provide a full rewrite of their documentation; 2) clearly disclose definitions, definitional changes and time-series inconsistencies in their documentation, and 3) look towards direct data-gathering via electronic media.

Users want clean and consistent data, easy definitions and a reasonable level of detail. Therefore, there is a need for further studies on the comparative characteristics of databases and their error/discrepancy rates. Meanwhile, users will choose databases primarily based on their availability and the existence of the desired data fields. Secondly, the actual treatment of particular industries and detail must be considered at the sample selection stage. Finally, users must not simply rely on the data being provided but must try to understand the implication of the accounting treatments that caused what is identified as **definitional discrepancies** in this study.

Depreciation and inventory numbers are substantially different. Their analysis indicates definitional difference effects to be a more important factor than the "error" effect previously emphasized in the literature. Their comprehension may allow adjustments for detected discrepancies (e.g. film rights in inventories), which may serve to decrease the variance of findings.

New technologies now allow the collection and online maintenance of much finer information at lower cost and their subsequent interface with other data sources. The problems evidenced in this paper are the result of this changing technology. Standard setting bodies must serve as agents for disclosure of standard homogeneity requirements across variables and industries. This would place the FASB not only in the role of regulator of measurement methods, but also as the setter of industry-by-industry disclosure standards and the issuer of guidelines for financial statements to be supplied in electronic medium.

APPENDIX A

Non-disclosed Coding Rule Differences

(A) INCOME BEFORE EXTRAORDINARY ITEMS

1. CMP includes "subsidiary's preferred stock dividend requirements," VL excludes.
2. CMP excludes "non-recurring credit related to accounting change, net," VL includes.

(B) SALES

1. VL includes "royalty income" and "excise taxes," CMP excludes.
2. VL excludes "equity in earnings of unconsolidated subsidiaries," CMP excludes.

(C) INVENTORIES

1. VL excludes "deposits and/or advances on material purchases," CMP excludes.
2. VL excludes "short-time timber leases", CMP includes.
3. CMP includes "securities purchases under agreement to resell", VL excludes.
4. CMP does not provide inventories for utility companies, but VL does.

(D) DEPRECIATION AND AMORTIZATION

1. CMP Includes "dry hole costs", VL excludes.
2. VL includes "amortization of deferred cost" and "depreciation on discontinued operations," CMP excludes.
3. CMP includes "leasehold impairment provision against income" and "impairment of unproved oil and gas properties", VL excludes.

(E) GROSS PLANT

1. VL excludes "rental equipments and parts", CMP includes.
2. VL excludes "airline companies' deposits and advances on flight equipment", "construction in progress and funds for construction, equipment leased to others", and "real estate companies' and land developers' land held for development and sale," CMP excludes.

REFERENCES

Beedles, W. L. and Simkowitz, M. A., "A Note on Skewness and Data Errors," **Journal of Finance**, (March 1978), pp. 288-292.

Bennin, R., "Error Rates in CRSP and Compustat: A Second Look," **The Journal of Finance**, December 1980, pp. 1267-1271.

Deakin, E. B., "Distributions of Financial Accounting Ratios: Some Empirical Evidence," **The Accounting Review**, (January 1976), pp. 90-96.

Goodman, H., Phillips, A., Burton, J. C. & Vasarhelyi, M. A., **Illustrations and Analysis of Disclosures of Inflation Accounting Information**, (AICPA, 1982).

Makin, C., "The World of Databases," **Institutional Investor**, (October 1984), pp. 343-352.

McEnally, R. W., "A note on the Return Behavior of High Risk Common Stocks," **Journal of Finance**, (March 1974), pp. 199-202.

Ro, B. T., "The Disclosure of Replacement Cost Accounting Data and Its Effect on Transaction Volumes," **The Accounting Review**, (January 1981), pp. 70-84.

Ro, B. T., "The Adjustment of Security Returns to the Disclosure of Replacement Cost Accounting Information," **Journal of Accounting and Economics**, (August 1980), pp. 159-189.

Rosenberg, B. & Houglet, M., "Error Rates in CRSP and Compustat DataBases and Their Implications," **The Journal of Finance**, September, 1974, pp. 1303-1310.

San Miguel, J. G., "The Reliability of R & D Data in Compustat and 10-K Reports," **The Accounting Review**, July, 1977, pp. 638-641.

Stone, M. and Bublitz, B., "An Analysis of the Reliability of the FASB Data Bank of Changing Pricing and Pension Information," **The Accounting Review** (July 1984), pp. 469-473.

Yang, D. C. and M. A. Vasarhelyi, "Selecting Financial Accounting Databases for Management Information Systems," **Proceedings of the Twentieth Hawaii International Conference on System Sciences** (January 1987), pp. 406-412.