Error Rates in CRSP and COMPUSTAT: A Second Look
Author(s): Robert Bennin
Source: *The Journal of Finance*, Vol. 35, No. 5 (Dec., 1980), pp. 1267-1271
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2327100

THE JOURNAL OF FINANCE • VOL. XXXV, NO. 5 • DECEMBER 1980

Error Rates in CRSP and COMPUSTAT: A Second Look

ROBERT BENNIN*

I. Introduction

ROSENBERG AND HOUGLET[1] [RH] made an error rate comparison study using the Compustat® Industrial Tape and the CRSP Monthly Return Tape covering the 6.5 years from 1963 to 1968. They found Compustat® to be less accurate, but defended Compustat® on the grounds that Compustat® might consider price information to be the "least important of the data" on the industrial tape. Compustat® has since released its Price Earnings Dividend (PDE) Tape in which the price information is of primary importance. This paper updates the study using this new tape.

II. Advantages of Each Data Base

Both Standard & Poors Compustat® (with its PDE tape) and the Center for Research in Security Prices (with its monthly stock return CRSP tape) sell a data base containing monthly security price information.[2] There are indisputable advantages to each service.

Advantages of the PDE tape

The PDE has a large security data base including OTC, Amex, and NYSE stocks—CRSP monthly only contains NYSE stocks. The PDE lists monthly price high, low, close—CRSP only lists monthly price close. The PDE has balance sheet items available to match its price data from identically ordered companion Compustat® tapes—CRSP is inconveniently ordered to match the balance sheet tapes. The PDE tape is prepared each month—CRSP tapes are only updated twice a year.

Advantages of the CRSP tape

CRSP reports the exact distribution types and dates in its companion Monthly Master tape—The PDE only gives a monthly dividend figure which does not

* Graduate student at the Schools of Business Administration, University of California, Berkeley.

[1] Rosenberg & Houglet, "Error Rates in CRSP and COMPUSTAT Data Bases and Their Implications", Journal of Finance, vol 29 (Sept 1974).

[2] The Compustat PDE tape is maintained by Standard & Poor's Compustat, Inc., Englewood, Colorado.

The CRSP tape is maintained by The Center for Research in Security Prices, University of Chicago.

include rights, warranties, or dividends in the form of another company's stock. CRSP has a larger historical data base, going back to 1926—The PDE data starts in 1962. CRSP retains securities which existed at any time on the NYSE—The PDE drops securities which do not presently meet its requirements.[3]

Although the user might be expected to choose between the services based upon the outlined advantages, another factor persists. That factor: Which source is more accurate?

III. Comparing Data Bases for Accuracy

The monthly return for each NYSE company on the PDE tape was matched against its CRSP monthly return. Unlike the original study, returns including all distributions were used in making the cross tape comparisons.[4] If the dividend or price information needed to calculate a return was missing from the PDE tape, the return was excluded from the study. If a CRSP monthly return was missing, then both that month and the following month were excluded from the study.[5]

The period covered was the 17.5 years from JAN 1962 thru July 1978. Using returns with dividends added the possibility of errors caused by incorrect dividend information. The overall error rate, however, was only ⅓ the rate reported in the earlier study.[6] The results follow in an updated version of [RH]'s Table I.

Table I

Comparison of CRSP & COMPUSTAT® Data Bases

	INDUSTRIALS (1295 Cos)			UTILITIES (99 Cos)		
	#	%	(RH%)	#	%	(RH%)
# of returns matched	170084	100		17376	100	
returns differ by:						
>.05 but, <.2	436	.26	(.45)	4	.02	(.17)
more than .2	31	.02	(.29)	0	0	(.41)
total errors	467	.27	(.74)	4	.02	(.58)

RH% = percentages from previous Rosenberg and Houglet Study.

Notice the large drop in the utility error rate. This was a sensitive area in [RH]'s study, and Compustat® may have devoted more energy to correcting its utility file.

[3] Compustat requirements for Company listing are membership in a major or regional exchange, or an OTC company which commands considerable interest.

[4] The original study used price relatives: $R(t) = [p(t) - p(t-1)]/p(t-1)$

This study uses returns with div.: $R(t) = [p(t) + div(t) - p(t-1)]/p(t-1)$

[5] The next month's return was dropped because CRSP reports a multiple return after a missing return.

[6] Error rate in the Rosenberg study was: $(294/41296) = .0071$

Error rate in this study is: $(471/187460) = .0025$

Four companies showed an unusually high number of dissimilar returns. These companies are:

Company Name	CUSIP	# of diff returns
1 Carter Hawley Hale Store	146227	37
2 Foremost McKesson Inc.	345514	28
3 Lykes Corp	550890	24
4 Sybron Corp	871140	19
		108

A merger had occurred in each case; Compustat® and CRSP elected to report the returns of different partner companies prior to the merge.[7]

There are 33 cases in which trading did not occur on the final day of the month for a particular security. Compustat® and CRSP use different procedures to calculate returns in these situations.[8] After removing the 4 companies and the tradeless dates, table Ib results.

Table Ib

Comparison of CRSP & COMPUSTAT® Data
(4 Cos, tradeless dates removed)

	INDUSTRIALS (1291 Cos)			UTILITIES (99 Cos)		
	#	%	(RH%)	#	%	(RH%)
# of returns matched	169316	100		17376	100	
returns differ by:						
>.05 but, <.2	301	.18	(.45)	4	.02	(.17)
more than .2	25	.01	(.29)	0	0	(.41)
total errors	326	.19	(.74)	4	.02	(.58)

Because of pressure to produce a timely tape, it was thought that errors would occur most often in the most recent data. The more recent data also was at a disadvantage since users had less time to uncover errors. Table 2 shows that the breakdown of yearly differences does not confirm this suspicion.

Table II

# of differences by year

yr	1962	1963	1964	1965	1966	1967	1968	1969	1970
#diff	23	17	17	19	46	56	32	30	37
yr	1971	1972	1973	1974	1975	1976	1977	1978 (thru Aug)	
# diff	14	3	9	10	5	5	3	4	

[7] Foremost McKesson — CRSP uses Foremost Dairies
Compustat uses McKesson & Robbins
Sybron — CRSP uses Ritter
Compustat uses M F Patterson
Carter Hawley Hale — CRSP uses Emporium Capwell
Compustat uses Broadway-Hales
Lykes — CRSP uses Lykes Bros.
Compustat uses other companies

[8] PDE uses the last traded price of the month. CRSP uses the BID-ASK if they report a return.

Table III shows the breakdown of differences between reported returns.

Table III

# of returns differing by x%

diff	.05	.06	.07	.08	.09	.10	.11	.12	.13	.14
# pts	92	58	27	27	21	12	17	6	9	11
diff	.15	.16	.17	.18	.19	.20	.21	.22	.23	.24>
# pts	5	6	4	6	4	2	1	2	2	18

For each return where a difference occurred, the two sources were then compared with prices reported by ISL in the Standard & Poor's Daily Stock Record.[9] Using ISL as the criterion for "correct" entries, Tables IV and V list the breakdown of errors.

Table IV

Cause of Error

Cause:	CRSP prices	PDE prices	Rts, War, Div	No Match	Not Found
	24	264	17	11	14

Table V

Error Source by Year

yr	62	63	64	65	66	67	68	69	70	71
error by:										
CRSP	0	5	2	0	0	1	4	2	1	3
PDE	19	11	11	17	41	47	27	27	35	11
yr	72	73	74	75	76	77	78 (thru July)			
error by:										
CRSP	0	0	2	1	2	1	0			
PDE	0	8	5	1	2	0	2			

The Bank & Quotation Record and Moody's Dividend Record were also used to check for the cause of error. Using these sources no appreciable variation from Table IV occurred and the majority of errors still arose from PDE prices.[10]

IV. Conclusion

Since the Rosenberg and Houglet study of error rates in CRSP and Compustat,® Compustat® has issued a Price-Dividend-Earnings tape. Now that Compustat®

[9] Standard & Poor's Corporation, "ISL Daily Stock Price Index", 1962–1978.

[10] When the Bank & Quotation Record and Moody's Dividend Record are used the following table results:

Table IVb

Cause of Error

Cause:	CRSP prices	PDE prices	Rts, War, Div	No Match	Not Found
	35	254	20	7	14

has a tape primarily dealing with security prices, it seems timely to re-examine the comparative accuracy of the two services. In cross checking the tapes, 471 rates of return were found to differ by more than .05. This amounts to a 2/3 reduction in the percentage of differing rates of return since the original study. The number of discrepancies does, however, reffirm the usefulness of [RH]'s proposed data comparison by a user group.

Compustat® errors drop markedly after 1970, indicating that it has improved its data collection. But CRSP still stands out as the more reliable service by a factor of 13 prior to 1975. Since 1975, the two services are equally reliable. The cross checking technique reveals a Compustat® error rate of 1/1000, and a CRSP error rate of 1/10000 over the years 1962–1978.[11]

The fact that Compustat® does so well in the later years suggests that Compustat® will become as accurate as CRSP. One problem persists with the PDE tape: it does not include rights offerings, warrants, options, or dividends in the form of another company's stock in its dividend figure.

[11]The errors found in this study will be forwarded to both services and presumably will be corrected in their newer tapes.