"THE IMPACT OF CEO TURN-OVER ON SECURITY ANALYSTS' FORECAST ACCURACY" - A COMMENT

David T. Doran (E-Mail: dtd1@psu.edu) Penn State University

Abstract

A recent Journal of Applied Business Research article by Sheikholesami, Wilson and Slevin (1998) examined the accuracy of security analysts' earnings forecasts for CEO change firms relative to a control group. The authors applied ANOVA on Value Line percentage forecast error measures and found "marginally significant" results indicating "that precision improved more for CEO change firms than for control firms." Doran (1998) tests for superior methods when scrutinizing forecast error. He finds percentage forecast error data to be severely non-normal, and demonstrates that nonparametric tests based upon ranks are superior to parametric methods. If analysts' earnings forecast precision actually improves more for CEO change firms, test results should be stronger using rank values instead of discrete percentage error measures.

Introduction

A recent article appeared in the Fall 1998 issue of the Journal of Applied Business Research (JABR) by Sheikholesami, Wilson and Slevin (SWS), "The Impact of CEO Turnover on Security Analysts' Forecast Accuracy". The authors used ANOVA to scrutinize Value Line forecast accuracy before vs. after CEO change relative to a control sample. Their measure of earnings divergence (forecast error)was calculated as a percentage error:

APE = $|(F-A)/A|*100$, with A = actual eps, and F = forecast eps.

In testing significance they applied the parametric F test and found "the most important interaction effect" to only be marginally sig-

Readers with comments or questions are encouraged to contact the author via e-mail.

nificant (at 8.4%). The authors do not provide detailed sample descriptive statistics, however the observed mean and standard deviation measures of APE ranged from a high of 101 and 145 (CEO change group, before the change) to a low of 14 and 23 (control group, after the change) across the four sample groups.

In the spirit of Brown and Warner (1980, 1985), Doran (1998) identifies the superior methodological alternatives to be used in studies scrutinizing earnings divergence (e. g., forecast error, forecast bias, earnings performance, etc.). The paper identifies superior (and inferior) methods by introducing positive forecast error at various levels (1% to 10%) and tests the relative power of different methods in identifying divergent earnings.

Consistent with SWS, Doran uses Value Line Data. He documents the severe non-normality of Value Line percentage forecast error data. Large measures of APE can result with seemingly low forecast errors when the denominator is small (at low earnings levels). For example with actual eps=$.01 and forecast eps=$.04; a $.03 forecast error results in APE=300. Such outliers result in a severely leptokurtic distribution. A leptokurtic distribution is one that is peaked with long fat tails. SWS do not control for outliers. This likely is why their observed mean and standard deviation measures are so high for some sample groups and differ drastically across sample blocks.

Doran finds that if parametric methods are applied to percentage forecast error data, some data refinement technique is required in order to control outliers. Otherwise, the severely non-normal sample distribution provides weak parametric test results. However, Doran demonstrates that even with data refinement, using rank values instead of discrete percentage forecast error measures provides more powerful statistical results.

Discussion

Doran's percentage error measure is referred to as "divergent earnings" (DE%) and is calculated (using SWS notation) as: DE %=(A-F)/|F|. DE% differs from APE in that it is signed, deflated by the absolute value of forecasted eps, and not multiplied by 100 (i. e., DE% of 1=APE of 100). However, the distributional implications of the methodology are the same - low deflator levels (whether forecasted or actual eps) cause outliers that result in a severely leptokurtic distribution. For a discussion of the outlier problem, see Fried and Givoly (1984), and Beaver et al. (1979). Doran conducts analysis of DE% where outliers are controlled by applying a simple truncation rule.

Table 1 provides properties of Doran's full DE% sample distribution (2701 cases). The table illustrates how the data can be made less non-normally distributed by controlling outliers. Measures of DE% are truncated to result in a minimum value of -1, and maximum value of +1 (consistent with Philbrick and Ricks (1991)). The full sample distribution of DE% without refinement (truncation) is severely non-normal (Kolmogorov-Smirnov test statistic of 21.9). The kurtosis measure of 1,654 is indicative of an extremely leptokurtic distribution. This is consistent with the outlier problem discussed previously. Outliers are primarily caused by small denominator amounts. The outlier problem is particularly severe when using a percentage error measure because of small observed earnings per share levels. Table 1 also provides the same sample descriptives with the truncation rule applied. Although the sample distribution remains non-normal, it is less severely so. The kurtosis measure declines to 2.22 and the Kolmogorov-Smirnov test statistic drops to 7.7.

TABLE 1
PROPERTIES OF FULL SAMPLE DISTRIBUTION OF DE%

	Without Truncation	With Truncation
Mean	-.502	-.089
Median	-.029	-.029
Standard Deviation	8.481	.353
Minimum	-389.000	-1.000
Maximum	71.000	1.000
Skewness	-37.200	-.479
Kurtosis	1654.280	2.221
Kolmogorov - Smirnov Z	21.915***	7.740***

Where: DE% = Divergent earnings deflated by the absolute value of forecasted earnings

To accommodate methodological alternatives analysis, Doran randomly draws 200 portfolios of 100 cases each (with replacement) from the full sample. One half of the portfolios are designated as "divergent", while the others are designated "control". Each of the divergent portfolios is matched with a control group portfolio. In order to better isolate the effects of alternative methodologies and minimize differences due to random fluctuation, the 100 matched companion samples are held identical across all analysis. Positive divergent earnings is introduced to all sample cases in each of the 100 "divergent" groups, while each of the 100 matched "control" groups remains consistently unaltered. Statistical tests of observed differences in average DE% between the divergent group and the control group are conducted with 1%, 3%, 5%, 7%, and 10% error introduced. The null hypothesis of no difference in DE% between each of the 100 matched groups is tested at the 1%, 5%, and 10% levels of significance. More powerful methods provide lower incidence of type II error. Type II error occurs with failure to recognize divergent earnings when it in fact exists. The relative power of the methods used is determined by Doran using a chi-square test.

Doran tests for the need when scrutinizing percentage statistical forecast error where parametric to control outliersmethods are used. Table 2, panel A provides evidence that truncation results in more powerful parametric statistical tests when scrutinizing DE%. At forecast error levels of 5% or more the parametric t test is consistently more powerful using the truncated sample data. This indicates that if SWS choose to apply ANOVA to the calculated numerical values of APE, they should use truncation or some other technique in order to make the sample distribution less non-normal. In doing so, parametric methods become more well specified.

As an alternative to using the numerical measures of APE, SWS may have applied nonparametric statistics using rank (ordinal) values instead. Substituting ranks for the calculated percentage forecast error measures mitigates the problems generally associated with outliers. Doran applies the nonparametric Mann Whitney test that uses rank values rather than continuous interval measures of DE%. To determine the need for data normalization if using rank values, Doran conducts Mann Whitney tests of DE% both with and without truncation. Table 2, panel B indicates there is no benefit that results from data truncation under the Mann Whitney test. This result indicates truncation shouldn't be necessary if rank values of APE are substituted for the formula calculated amounts.

Doran tests for relative superiority of statistical method in detecting divergent earnings. He compares the power of nonparametric statistics using rank values to parametric alternatives. Table 3, Panel A provides analysis of statistical method without truncation, while Panel B depicts the results with the -1, +1

TABLE 2
EFFECT OF CONTROLLING FOR OUTLIERS THROUGH TRUNCATION
OBSERVED FREQUENCY (%) - REJECTION OF THE NULL HYPOTHESIS

SIG LEV	0%ERROR			1%ERROR			3%ERROR			5% ERROR			7%ERROR			10%ERROR		
	PANEL A t-TEST OF DE% WITH (W) VS. WITHOUT (WO) TRUNCATION																	
	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2
1%	1	1	1.01	1	3	3.05*	1	4	1.85	1	8	5.70**	2	11	6.66***	5	27	18.00***
5%	3	5	.52	4	5	.12	7	11	.98	9	21	5.65**	11	32	13.06***	14	46	24.38***
10%	13	7	2.00	13	12	.05	17	20	.30	17	34	7.61***	18	45	16.89***	27	65	29.07***
	PANEL B MANN WHITNEY TEST OF DE% WITH (W) VS. WITHOUT (WO)																	
	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2	WO	W	X^2
1%	1	1	.00	2	2	.00	8	8	.00	23	23	.00	42	42	.00	74	74	.00
5%	4	4	.00	8	8	.00	22	21	.03	44	45	.02	70	71	.02	94	94	.00
10%	8	8	.00	13	13	.00	34	34	.00	62	60	.08	86	86	.00	98	98	.00

Where: DE% = Divergent earnings deflated by the absolute value of forecasted earnings
X^2 = Chi-squared distribution value (2-tailed)
*** = Significant at the 1% level
** = Significant at the 5% level
* = Significant at the 10% level

truncation rule applied. Without truncation, the chi-square statistics clearly indicate that the non-parametric Mann Whitney test is superior to the parametric t test. The observed superior performance is consistently significant when divergent earnings of 3% or more is introduced. Panel B of Table 3 again illustrates the superiority of nonparametric statistical analysis even with data refinement. The evidence is not as convincing when 3% forecast error is introduced; however, it is otherwise consistent and pronounced. Doran's statistical method tests indicate that SWS should consider transforming APE measures to rank values before conducting ANOVA. Using rank values should provide more powerful test results.

Conclusion

The results of Doran (1998) suggest that if SWS choose to apply parametric statistical methods when scrutinizing APE, a sophisticated data refinement method that results in a more normal sample distribution is required; other-

TABLE 3
PARAMETRIC t TEST VS. NONPARAMETRIC MANN WHITNEY TEST
OBSERVED FREQUENCY (%) - REJECTION OF THE NULL HYPOTHESIS

SIG LEV	0%ERROR			1%ERROR			3%ERROR			5% ERROR			7%ERROR			10%ERROR		
PANEL A — **t-TEST (t) VS. MANN WHITNEY TEST (MW) OF DE% WITHOUT TRUNCATION**																		
	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2
1%	0	1	1.01	0	2	2.02	1	8	5.70**	1	23	22.92***	2	42	46.62***	5	74	99.61***
5%	3	4	.15	4	8	1.42	7	21	8.14***	9	44	31.45***	11	70	72.23***	14	94	128.82***
10%	13	8	1.33	13	13	.00	17	34	7.61***	17	62	42.37***	18	86	92.63***	27	98	107.54***
PANEL B — **t-TEST (t) vs. WHITNEY TEST (MW) OF DE% WITH TRUNCATION**																		
	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2	t	MW	X^2
1%	1	1	.00	3	2	.21	4	8	1.42	8	23	8.59***	11	42	24.67***	27	74	44.18***
5%	5	4	.12	5	8	.74	11	21	3.72*	21	45	13.03***	32	71	30.45***	46	94	54.86***
10%	7	8	.07	12	13	.05	20	34	4.97**	34	60	13.57***	45	86	37.19***	65	98	36.13***

Where: DE% = Divergent earnings deflated by the absolute value of forecasted earnings

X^2 = Chi-squared distribution value (2-tailed)
*** = Significant at the 1% level
** = Significant at the 5% level
* = Significant at the 10% level

wise, nonparametric rank based methods are more powerful. The suggestion of substituting ranks for discrete data measures and applying ANOVA procedures on the rank value data is consistent with Conover (1980, pp. 337-338):
To analyze an experimental design using the rank transformation, first rank all of the observations together from smallest to largest and then apply the usual analysis of variance to the ranks. The result is a procedure that is only conditionally distribution free. However, it is "robust," which means that the true level of significance is usually fairly close to the approximate level of significance used in the test, no matter what the underlying population distribution might be. The resulting procedure usually has good efficiency (Inman, 1974b, and Conover and Inman, 1976). The recommended procedure in experimental designs for which no parametric test exists is to use the usual analysis of variance on the data and then to use the same procedure on the rank transformed data. If the two procedures give nearly identical results the assumptions underlying the usual analysis of variance are likely to be reasonable and the regular parametric analysis valid. When the two procedures give substantially different results, the analysis on ranks is probably more accurate than the analysis on the data and should be preferred. In such cases the experimenter may want to take a closer look at the data and to look especially for outliers (observations that are unusually large compared with the bulk of the data) or very nonsymmetric distributions. These aberrations in the data affect the analysis of the data to a great extent by changing the level of significance and decreasing the power, but the analysis of the ranks is not affected nearly as much.

The Doran research, coupled with the recommendations of Conover (1980) strongly suggest that substituting rank values for the discrete percentage forecast error measures that SWS use in their analysis is appropriate. If there actually is a relationship between CEO change and analysts' forecast accuracy, conducting ANOVA on the rank values of APE should lead to more powerful (and convincing) results.

Future Research

Studies where scrutiny of earnings divergence (e.g., forecast error, earnings performance, forecast bias, etc.) is of primary importance should closely examine the distributional properties of the sample data. It is likely that the sample distribution is non-normal. Under these circumstances the researcher should perform data refinement techniques to make parametric analysis well specified, or should perform nonparametric analysis using rank values. Otherwise, statistical test results will be relatively weak.

References

1. Beaver, W., R. Clarke, and W. Wright. 1979. "The Association Between Security Returns and the Magnitude of Earnings Forecast Errors." *Journal of Accounting Research* 17, 2: 316-340.
2. Brown, S.J., and J.G. Warner. 1980. "Measuring Security Price Information." *Journal of Financial Economics* 8, 3: pp. 205-258.
3. Brown, S.J., and J.G. Warner. 1985. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics* 14, 1: pp. 3-31.
4. Conover, W.J. 1980. "Practical Nonparametric Statistics, Second Edition." John Wiley and Sons Publishing, Inc.
5. Conover, W.J., and R.L. Inman. 1976. "On Some Alternative Procedures Using Ranks for the Analysis of Experimental Designs." *Communications in Statistics - Theory and Methods*, A5: pp. 1349-1368.
6. Doran, D.T. 1998. "Methodological Choices in Detecting Divergent Earnings." Working Paper No. 98-1, Penn State Behrend School of

Transformation for the Two-Way Classification Model When Interaction May be Present." *The Canadian Journal of Statistics Section C: Applications* 2, pp 227-239.
9. Philbrick, D.R., and W.E. Ricks. 1991. "Using Value Line and IBES Analyst Forecasts in Accounting Research." *Journal of Accounting Research* 29, 2: 397-417.
10. Sheikholeslami, M., M.D. Wilson, and J.R. Selin. 1998. "The Impact of CEO Turnover On Security Analysts' Forecast Accuracy." *Journal of Applied Business Research* 14, 4: 71-75.

Notes

Notes