

Methodological Choices In Detecting Divergent Earnings: An Extension

David T. Doran, (Email: dtd1@psu.edu), Penn State University, Erie

Abstract

Divergent earnings are those that differ from expected. Doran (2000) provides evidence that nonparametric tests based upon rank values are superior to parametric alternatives in detecting divergent earnings. He also finds that deflator choice (i.e., forecasted earnings or market price of the stock) is of little importance when superior nonparametric methods are used. This study extends the efforts of Doran (2000) by testing for benefit derived from the common research method of deflating earnings data. The data used here is the same as used by Doran (2000), where Value Line is the source of all earnings data. One hundred independent two sample tests are performed between a positive earnings group and a matched control group. The tests are performed with various levels (1%, 3%, 5%, 7%, and 10 %) of positive actual earnings introduced. Failure to reject the null hypothesis of no positive earnings divergence indicates the existence of Type II error (determined using the nonparametric Mann Whitney test). The Mann Whitney test was performed on the undeflated data, and the same data deflated by: 1) forecasted earnings, and 2) market price of the stock. Difference in frequency of type II error is determined using the Chi-square test. The results generally indicate no significant difference in the ability to identify abnormal divergent earnings when utilizing deflated data. Statistical tests are found here to be at least as powerful when undeflated earnings data are used. There is weak evidence supporting the notion that deflating earnings data inhibits the ability to detect abnormal earnings. These findings indicate that the common practice of deflating earnings data is unnecessary, and may actually weaken the power of statistical tests.

Introduction

Doran (2000) tests for methodological superiority in detecting “divergent earnings” – earnings that differ from expectation. Doran (2000) finds that due to severe nonnormality of divergent earnings data, nonparametric statistical methods that are based upon ranks are superior to parametric alternatives; and, under these nonparametric methods, deflator choice (expected earnings or market price of stock¹) is of little importance.

Christie (1987) argues that when divergent earnings is scrutinized in association with market returns (i.e., earnings response coefficients), stock price is the appropriate deflator choice because it provides dimensional consistency of the independent and dependent variables. Although stock price may be the appropriate deflator in studies of earnings response coefficients, the appropriate deflator in other studies of divergent earnings (“earnings performance”, “forecast error”, etc.)ⁱⁱ remains an open issue.

No study has documented any benefit derived from the common practice of deflating earnings measures. This study contributes to the literature by testing the relative power of statistical tests in detecting abnormal divergent earnings by using undeflated vs. deflated earnings data.

Data and Methodology

The sample data used in this analysis is the same as used in Doran (2000). The source of all expected and actual earnings data is Value Line. Three firms are randomly selected from each week's Value Line Summary and Index

Readers with comments or questions are encouraged to contact the author via email.

publication issued between November 17, 1971 and December 31, 1990. For some weeks, less than three firms are included in the analysis due to lack of requisite information (e.g., forecasted earnings per share (EPS) was "no meaningful figure"). Actual reported earnings were gathered from subsequent issues of Value Line. The full sample includes 2,701 cases. From this full sample, randomly drawn portfolios of 100 cases each (with replacement) are drawn. This process was repeated 200 times. All analysis is conducted using these 200 randomly selected portfolios of 100 cases each.

Divergent earnings (DE) is the undeflated measure, DE\$ is DE deflated by stock price, and DE% is divergent earnings deflated by the absolute value of the EPS forecast. The stock price deflator used is the market price of the particular firm's stock on the date of the Value Line forecast.

Formally stated:

$$DE = (\text{Actual EPS} - \text{Forecasted EPS}) \tag{1}$$

$$DE\$ = (\text{Actual EPS} - \text{Forecasted EPS}) / \text{Stock Price} \tag{2}$$

$$DE\% = (\text{Actual EPS} - \text{Forecasted EPS}) / |\text{Forecasted EPS}| \tag{3}$$

Previous research has documented that analyst forecasts are optimistically biased (see O'Brien (1988), Butler and Lang (1991), Philbrick and Ricks (1991), Abarbanell (1991), and Ali, et al. (1992)). The Value Line forecasts here are consistent with this finding. This results in disproportionately negative observations of DE. For the 2701 full sample cases of DE, 40% are positive, 58% are negative and 2% are zero. Full sample distribution characteristics (e. g., consistently negative full sample means and medians for DE, DE% and DE\$) also document the optimism bias.ⁱⁱⁱ With the Value Line forecasts optimistically biased, the expected value of divergent earnings is negative, and one sample tests of significant difference from zero are inappropriate. To accommodate the optimism bias, two sample tests are conducted.

One half of the portfolios are designated as "abnormal", while the others are designated "control". Positive divergent earnings is introduced to all sample cases in the "abnormal" group, while the "control" group remains consistently unaltered. Positive divergent earnings is introduced at different levels by increasing actual earnings by 1%, 3%, 5%, 7%, and 10%. The 200 randomly drawn portfolios accommodate 100 individual two sample tests. Each of the abnormal 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.

Statistical tests of differences in divergent earnings between the abnormal group and the control group for DE, DE\$, and DE% are conducted. The nonparametric Mann Whitney (rank) test is used. Since only positive abnormal earnings is introduced, the tests of significance are one tailed and the null hypothesis is:

$$H_0: DE, DE\%, DE\$ \text{ of the abnormal group is not greater than } DE, DE\%, DE\$ \text{ of the control group.}$$

The Mann Whitney test is performed 100 times for each divergent earnings measure (DE, DE\$, and DE%), at each level of abnormal divergent earnings (1%, 3%, 5%, 7%, and 10%), and is conducted both at the 1% and 5% levels of significance. This being the case, numeric frequency of acceptance or rejection of the null hypothesis is the same as the percentage acceptance or rejection.

Type I error is observed when the null hypothesis is rejected where zero abnormal divergent earnings has been introduced. Type II error is observed where the null hypothesis is not rejected, but positive abnormal divergent earnings has been introduced. Lower levels of type II error (higher incidence of rejection of the null hypothesis) indicate more powerful methodological choices. That is, lower levels of type II error indicate superiority in detecting abnormal divergent earnings. Differences in frequency of type II error is tested between DE and DE%, and also between DE and DE\$. The Chi-square test is used to identify significant differences in the observed frequency of type II error.

The Chi-square test is performed for each level of introduced abnormal positive earnings, and both Mann Whitney test levels of significance applied. For each Chi-square test there are two groups of 100 cases each. The

number of instances where null is accepted or rejected therefore sums to 100 for each group. The results can be viewed as a 2 x 2 contingency table where the columns are the alternative methodologies scrutinized (e.g., DE vs. DE\$) and the rows are frequency of rejection, vs. frequency of acceptance (100 minus frequency of rejection). This classification scheme accommodates the chi-square test conducted here.^{iv}

Results

Incidence of type I error is scrutinized. With sample size of 100, applying a 95% confidence interval under the binomial distribution results in the following lower and upper limits for observed frequency of type I error: 1% = 0 and 3, and 5% = 1 and 10. In all cases, the observed frequencies of Type I error fall within these 95% confidence limits. With zero abnormal earnings introduced, at the 1% and 5% levels of significance the null hypothesis is rejected with the following frequencies: DE – 1 and 3, DE% - 1 and 4, and DE\$ - 1 and 4.

The frequency of Type II error where positive abnormal earnings has been introduced is shown in Table 1. Panel A documents the incidence of type II error for DE vs. DE%, while Panel B provides a the same information while comparing DE with DE\$.

Panel A indicates that there is only one incidence of significant difference. In the case of 7% introduced abnormal earnings at the 1% Mann Whitney level of significance, statistical tests are more powerful when using undeflated data. Although not statistically significant, with the exception of only two cases (1% abnormal earnings at the 1% and 5% levels of significance), the occurrence of type II error is as infrequent using undeflated data as is observed when deflated by forecasted earnings. There is one instance where the level of type II error is the same - 3% introduced abnormal earnings, at the 1% Mann Whitney level of significance. In the other seven cases, using the undeflated data results in lower incidence of type II error.

Panel B of Table 1 indicates no cases where a significant difference in type II error is observed by using DE rather than DE\$. Again however, although not statistically significant, using undeflated data generally results in lower levels of observed type II error. With the exception of only one case (1% introduced abnormal earnings at the 5% Mann Whitney level of significance), the occurrence of type II error is as infrequent when using the undeflated data as is observed when using data deflated by the stock price. There are two instances where the frequency of type II error is the same (1% and 3% introduced abnormal earnings both at the 1% Mann Whitney level of significance). In the remaining seven cases, using the undeflated data results in lower levels of type II error.

Conclusion

Studies that scrutinize divergent earnings (or forecast error) commonly employ the methodology of deflating divergent earnings measures. This study investigates the impact of this practice on the power of the statistical tests. The results generally indicate there is no significant difference between using undeflated vs. deflated divergent earnings data (deflated by either market price of stock or expected earnings). There is weak evidence indicating that the use of undeflated data may actually provide more powerful statistical tests that are superior in detecting abnormal divergent earnings.

The results indicate that for studies where the key motivation is to correctly identify abnormal divergent earnings (or forecast error), there is no benefit gained from deflating the sample data, and the process of doing so may weaken the ability to detect abnormal divergent earnings when it is actually exists.

Future Research

As indicated in Doran (2000), the results here using Value Line data are likely to be consistent with results of similar analysis performed on alternative earnings data. However, future studies should be conducted using different

Table 1 here (landscape format)

sources of analysts' forecasts (e.g., First Call, Institutional Brokers Estimate System (IBES)), and/or historical based earnings prediction models in order to demonstrate that the results are robust.

There are also implications for previous studies that have incorporated deflated earnings data and had statistically weak results. Repeating such efforts with undeflated earnings data may strengthen the outcome. 📖

References

1. Abarbanell, J., "Do Analysts' Earnings Forecasts Incorporate Information in Prior Stock Price Changes", *Journal of Accounting and Economics*, Vol. 14 (June 1991): 147-165.
2. Ali, A., A. Klien, and J. Rosenfield. 1992. "Do Analysts Properly Use Information about Permanent and Transitory Earnings Components in Setting Their Forecasts of Annual EPS?" *The Accounting Review* 67, 1: 183-198.
3. Asquith, P., P. Healy and K. Palepu. 1989. "Earnings and Stock Splits." *The Accounting Review* 64, 3: 387-403.
4. 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.
5. Bowen, R.M., M. Johnson, T. Shelvin, and D. Shores. 1992. "Determinants of Quarterly Earnings Announcements." *Journal of Accounting Auditing and Finance* 7, 4 (Fall): 395-422.
6. Brown, L.D., P.A. Griffin, R.L. Hagerman, and M.E. Zmijewski. 1987. "Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings." *Journal of Accounting and Economics* 9 (April): 61-87.
7. Brown, L.D., P.A. Griffin, R.L. Hagerman, & M.E. Zmijewski. 1987a. "An Evaluation of Alternative Proxies for the Market's Assessment of Unexpected Earnings." *Journal of Accounting & Economics* 9 (July): 159-193.
8. Brown, L.D., and K.J. Kim. 1991. "Timely Aggregate Analyst Forecasts as Better Proxies for Market Earnings Expectations." *Journal of Accounting Research* 29, 2 (Autumn): 382-385.
9. Brown, L.D., and M.S. Rozeff. 1978. "The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings." *Journal of Finance* 33 (March): 1-16.
10. Butler, K.C., and L.H. Lang. 1991. "The Forecast Accuracy Of Individual Analysts: Evidence of Systematic Optimism and Pessimism." *Journal of Accounting Research* 29 (Spring): 150-156.
11. Christie, A.A. 1987. "On Cross-Sectional Analysis in Accounting Research." *Journal of Accounting and Economics*
12. Conover, W.J. 1980. "Practical Nonparametric Statistics, Second Edition." John Wiley and Sons Publishing, Inc.
13. Doran, David T., and Robert Nachtmann. 1988. "The Association of Stock Distribution Announcements and Earnings Performance." *Journal of Accounting, Auditing and Finance* 3, 2 (Spring): 113-132.
14. Doran, David T. 1994. "Stock Splits: Tests of Earnings Signaling and Attention Directing Hypotheses Using Analyst Forecasts and Revisions." *Journal of Accounting, Auditing and Finance* 9, 3 (Summer): 411-422.
15. Doran, David T. 1995. "Earning Performance and Interim Reporting." *Journal of Applied Business Research* 11, 1: 67-72.
16. Doran, David T. 2000. "Methodological Choices In Detecting Divergent Earnings." *Journal of Applied Business Research* 16, 3: 113-123.
17. Fried, D., and D. Givoly. 1982. "Financial Analysts' Forecasts: A Better Surrogate for Market Expectations." *Journal of Accounting and Economics* 4: 85-107.
18. O'Brien, P.C. 1988. "Analysts' Forecasts as Earnings Expectations." *Journal of Accounting and Economics* 10, 1: 52-83.
19. 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.
20. Williams, P.A. 1996. "The Relation Between a Prior Earnings Forecast by Management and Analyst Response to a Current Management Forecast", *The Accounting Review* 71, 1: 607-621.

Endnotes

- ¹ These are commonly used deflators. See for example Williams (1996), Doran (1994), Philbrick and Ricks (1991), and Doran (1995).
- ² For example studies of relative accuracy of forecast agents (see Brown and Rozeff (1978), Fried and Givoly (1982), Brown et. al. (1987), and Philbrick and Ricks (1991)). There are also studies of earnings signaling associated with particular events (see Doran and Nachtmann (1988), Asquith, et. al., (1989), and Doran (1994).
- ³ For a detailed analysis of the full sample data, see Doran (2000) Table 1.
- ⁴ For a detailed discussion of the chi-square test applied here, see Conover (1980), p.144.

Notes

ⁱ These are commonly used deflators. See for example Williams (1996), Doran (1994), Philbrick and Ricks (1991), and Doran (1995).

ⁱⁱ For example studies of relative accuracy of forecast agents (see Brown and Rozeff (1978), Fried and Givoly (1982), Brown et. al. (1987), and Philbrick and Ricks (1991)). There are also studies of earnings signaling associated with particular events (see Doran and Nachtmann (1988), Asquith, et. al., (1989), and Doran (1994).

ⁱⁱⁱ For a detailed analysis of the full sample data, see Doran (2000) Table 1.

^{iv} For a detailed discussion of the chi-square test applied here, see Conover (1980), p.144.

Notes