On a Statistical Method to Detect Discontinuity in the Distribution Function of Reported Earnings

Y. Takeuchi

Graduate School of Economics, Osaka University, Toyonaka, Osaka, Japan (takeuchi@econ.osaka-u.ac.jp)

Abstract: In the study of earnings management, researchers attempt to know the reason managers manipulate earnings of a company. Some preceding studies have focused on the distribution of reported earnings, and they treat the evidence of manipulation on earnings as discontinuity (at most cases around zero earnings) in the distribution function of reported earnings. They used statistical tests to detect the discontinuity, though, none of the tests they proposed were shown their derivation and examined their properties. This study tries to find a proper procedure for such kinds of study. In this study, we focus on the test proposed by Burgstahler and Dichev among tests. The test is designed to test the null hypothesis that has continuous distribution function against alternatives that have discontinuity in the distribution function at a certain point by use of empirical distribution (or histogram). The purpose of this study is twofold. First is to derive the test statistics properly. Second is to examine properties of the test using Monte Carlo simulation. The results are as follows. First, we can understand that the test is based on estimators of density (i.e. empirical density) at a certain point for various bin-width and is derived using multinomial distribution. Second, from the results by Monte Carlo simulation, size and power of the test depends on sample size and bin-width. Over moderate sample size, the size of the test is almost correct and the test has good power. In addition, we find that the test is able to detect discontinuity for a small jump in continuous distribution. We also made simulations for various continuous distributions, and we have almost same results. Thus, we conclude that the test is available for various situations.

Keywords: Earnings management; Discontinuity; Distribution function; Bin-width; Discrete approximation

1. INTRODUCTION

In the study of earnings management, researchers attempt to know the reason managers manipulate earnings of a company. Recent survey by McNichols [2000] classified studies on earnings management into three categories; "those based on aggregate accruals, those based on specific accruals and those based on the distribution of earnings after management (p314)." In preceding studies based on the distribution of reported earnings, and they treat the evidence of manipulation (or discretion) on earnings as discontinuity (at most cases, discontinuous around zero earnings) in the distribution function of reported earnings. Although they used statistical tests to detect the discontinuity, none of the tests they proposed were shown their derivation and examined their properties. This study is a trial to find a proper procedure for such kinds of study.

In this study, we focus on the test proposed by Burgstahler and Dichev [1997] among tests because it is only test which is statistically correct.[1] The test is designed to test the null hypothesis that has continuous distribution function against alternatives that have discontinuity in the distribution function at a certain point by use of empirical distribution (or histogram). The purpose of this study is twofold. First is to derive the test statistics properly. Second is to examine properties of the test using Monte Carlo simulation.

The layout of this study is as follows. Section 2 overviews how earning management (discretion) affects reported earnings. Section 3 sketches the test on discontinuity in distribution by Burgstahler and Dichev [1997], and derive its distribution under null hypothesis. Section 4 reports properties of the test by Monte Carlo simulations for several situations. Section 5 gives summary with conclusion.

[1] For detailed discussion, see Kan-no, Takao and Takeuchi [2001].

2. EARNING MANAGEMENT AND ITS EFFECT TO EARNING DISTRIBUTION

Earning decrease is bad information for investors. In fact, news of decrease earnings than expected ones immediately spread markets and make the stock price falling. As maximizing total (market) value of the firm is one of tasks postulated on executives, avoiding decrease earnings or losses is strong incentives for them.

It is supposed that earning management to avoid decrease earning or losses is taken place when earnings become below benchmark value (for example, zero earnings). This situation can be modeled as;

$$r = e \qquad \text{if } e > e^*$$

$$r = \begin{cases} e^* & \text{with } \delta \\ e & \text{with } 1-\delta \end{cases} \qquad \text{if } e \le e^*$$

where e and r are true and reported earnings, respectively. Also e^* represent benchmark value of earnings, δ is probability of managing earnings.

Under this model, even if e follows a continuous distribution, the distribution of reported earnings has a jump (i.e. discontinuous) at the benchmark e^*. Using this result, we can indirectly find a discretionary earnings management by detecting discontinuity in the distribution of reported earnings.

3. Burgstahler and Dichev [1997] TEST FOR DETECTING DISCONTINUITY IN THE DISTRIBUTION FUNCTION

Burgstahler and Dichev [1997] proposed a statistical test to find earnings management used by empirical distribution[2] (i.e. histogram) of reported earnings. In this section, I will show a brief sketch of their test and will give another interpretation.

Let X_i $(i = 1,\dots,n)$ are independent random variables with distribution function F. Suppose equal spaced points $-\infty = c_0 < c_1 < \cdots < c_k = \infty$ where $c_j - c_{j-1} = h$ for $j = 2,\dots,k-1$. Then, an empirical frequency in $(c_{j-1}, c_j]$

$$Y_j = \sum_{i=1}^{n} I\{X_i \in (c_{j-1}, c_j]\} \quad (j = 1,\dots,k)$$

follows multinomial distribution with

$$p_j = P(X \in (c_{j-1}, c_j]) = \mathrm{F}(c_j) - \mathrm{F}(c_{j-1}).$$

Hence, $\mathrm{E}(Y_j) = np_j$ and $\mathrm{var}(Y_j) = np_j(1-p_j)$.

Burgstahler and Dichev [1997] introduced the 'smoothness' in the distribution and defined as

$$(p_{j-1} + p_{j+1})/2 = p_j \quad (\text{for } j = 2,\dots,k-1) \qquad (1)$$

and formulate test statistics

$$\tau_{BD} = \frac{(\hat{p}_{j-1} + \hat{p}_{j+1})/2 - \hat{p}_j}{\sqrt{\mathrm{var}((\hat{p}_{j-1} + \hat{p}_{j+1})/2 - \hat{p}_j)}} \qquad (2)$$

where $\hat{p}_j = Y_j / n$ and $\mathrm{var}((\hat{p}_{j-1} + \hat{p}_{j+1})/2 - \hat{p}_j)$ equals

$$\frac{1}{n} p_j(1-p_j) + \frac{1}{4n}(p_{j-1} + p_{j+1})(1 - p_{j-1} - p_{j+1}) + \frac{1}{n} p_j (p_{j-1} + p_{j+1}) \qquad (3)$$

Under null hypothesis of (1), τ_{BD} follows standard normal distribution[3].

Their test uses the notion of 'smoothness' in the probability distribution. However, the relationship in (1) has another interpretation. If F has a density (or is continuously differentiable), limit of density estimates with different bin-width will converge to same finite value.

$$\frac{\hat{p}_j}{h} \to \frac{p_j}{h} \to f\left(\frac{c_{j-1} + c_j}{2}\right) < \infty$$

$$\frac{\hat{p}_{j-1} + \hat{p}_j + \hat{p}_{j+1}}{3h} \to \frac{p_{j-1} + p_j + p_{j+1}}{3h} \to f\left(\frac{c_{j-1} + c_j}{2}\right) < \infty$$

Therefore, (1) is hold if F is continuous and Burgstahler and Dichev [1997] test, hereafter BD test, could be interpreted as a test for detecting discontinuity in a distribution function.

4. MONTE CARLO SIMULATION

4.1 Simulation Setting

To examine properties of the test statistics τ_{BD}, we will make a Monte Carlo experiment. Sample data of X_i $(i = 1,\dots,n)$ is generated based on the distribution function such that,

$$\mathrm{F}(x) = \begin{cases} (1-\delta)\int_{-\infty}^{x} g(x)dx & x < c^* \\ \int_{-\infty}^{c^*} g(x)dx & x = c^* \\ \int_{-\infty}^{x} g(x)dx & x > c^* \end{cases} \qquad (4)$$

---

[2] For the properties of histogram estimate, see Devroye and Györfi [1985].

[3] Bahstheler and Dichev [1997] ignores last term in equation (3).

where $g(x)$, $x \in (-\infty, \infty)$ is a density distribution. Equation (4) shows distribution function $\mathbf{F}$ may have a jump $\delta \int_{-\infty}^{c^*} g(x)dx$ at c^*. In extreme case, $\mathbf{F}$ is continuous (i.e. has a density $g(x)$) when $\delta = 0$, while $\mathbf{F}$ is truncated at c^* when $\delta = 1$. In the experiment, we set seven types of δ; they are 0.0, 0.01, 0.02, 0.05, 0.1, 0.2, and 0.3.

We formulate τ_{BD} for three types of bin-width h; $h = 0.20\sigma, 0.10\sigma, 0.05\sigma$ where σ is standard error of random variable with pdf $g(x)$. For $g(x)$, we adopt uniform, normal, and Chi-square density.

Each experiment is performed with 1000 replications and for six different sample size n; they are 50, 100, 500, 1000, 5000, and 10000. We employ significance level for the test as 5%. All simulations are made by Gauss for windows (ver. 3.2) with IBM PC compatible PC (Intel Celeron 433 MHz).

4.2 Uniform Case

At the beginning, we examine a primary case in which sample data is drawn from uniform distribution, such that $X \sim U(-\sqrt{3}, \sqrt{3})$. In this case $\sigma = 1$, and we set the benchmark value as $c^* = -1$.

Table 1 shows size and power of BD test at the benchmark value (–1) for different *n*, *h*, and δ. For small sample (*n*=50 and 100), size of the test is not correct (in most case greater than 5%). In addition, the power of the test is poor even in moderate jump of $\delta = 0.30$, i.e. the probability of $c^* = -1$ is $0.30 \times (\sqrt{3} - 1)/\sqrt{12} = 0.063$. On the other hand, for over moderate sample size ($n \geq 500$), the size become almost correct and the power become high enough. For large sample such that ($n \geq 5000$), BD test can almost reject the null hypothesis even for small jump of $\delta = 0.05$, i.e. probability of $c^* = -1$ is 0.011. From Table 1, we can easily find that the power become higher as the bin-width *h* become small.

Results in Table 1 indicate the power only at the benchmark value (at the threshold point). In practice, we seldom know information about the threshold point. Then, we perform BD test around the benchmark value. Table 2 shows rejection rate of the null around the benchmark when *h*=0.10 and $\delta = 0.10$. From this table, except only on two adjacent points, rejection rates are almost same as significance level (5%). However, on two most adjacent points of the benchmark value, even though the null is true, there is high tendency to reject the null. This feature is noteworthy for BD test.

Table 1 Size and power of BD test (uniform case; at *x*=-1).

n	*delta*	*h* 0.20	0.10	0.05
50	0.00	0.075	0.081	0.030
	0.10	0.070	0.061	0.042
	0.20	0.115	0.146	0.149
	0.30	0.204	0.258	0.329
100	0.00	0.055	0.091	0.092
	0.10	0.079	0.096	0.153
	0.20	0.187	0.291	0.409
	0.30	0.396	0.548	0.687
500	0.00	0.053	0.051	0.049
	0.05	0.116	0.152	0.231
	0.10	0.282	0.480	0.645
	0.20	0.816	0.934	0.990
	0.30	0.984	0.999	1.000
1000	0.00	0.056	0.061	0.062
	0.05	0.183	0.293	0.466
	0.10	0.567	0.765	0.931
	0.20	0.984	0.999	0.999
	0.30	1.000	1.000	1.000
5000	0.00	0.052	0.046	0.052
	0.01	0.060	0.106	0.126
	0.02	0.180	0.283	0.444
	0.05	0.687	0.902	0.987
	0.10	0.994	1.000	1.000
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000
10000	0.00	0.041	0.045	0.043
	0.01	0.106	0.164	0.255
	0.02	0.275	0.534	0.761
	0.05	0.938	0.998	1.000
	0.10	1.000	1.000	1.000
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000

(Note) Some of the results are unlisted.

Table 2 Rejection rate around the benchmark (uniform case; *h*=0.10, *delta*=0.10).

midpoint	*n* 500	1000	5000	10000
-1.30	0.048	0.046	0.044	0.049
-1.20	0.042	0.047	0.041	0.055
-1.10	0.278	0.413	0.967	1.000
-1.00	0.480	0.765	1.000	1.000
-0.90	0.186	0.289	0.894	0.995
-0.80	0.044	0.062	0.058	0.069
-0.70	0.054	0.060	0.063	0.058

As we have seen in Table 1, the power is quite poor when sample size is small. One possible reason for this is estimating denominator of equation (2). To check this point, we compare the power using estimated variance with using true variance (Table 3). Even though power of the test with true variance is poor for small sample, small sample bias of the estimator affects the test considerably amount. For example, when n=50, the power of the test reduce to about one quarter if we use the estimated variance.

Table 3 Power comparison with variances (uniform case; h=0.05, *delta*=0.10, at x=-1).

n	BD(est.) A	BD(true) B	A/B
50	0.042	0.187	0.225
100	0.153	0.387	0.395
500	0.645	0.823	0.784
1000	0.932	0.972	0.959

4.3 Normal Case

It seems that the uniform case we have seen in 4.2 is easy to detect discontinuity because its distribution function has linear function. Next, we examine a more complicated case such that sample data is drawn from standard normal distribution. We also set the benchmark value as $c^* = -1$, because the second derivative of the distribution function $F = \Phi$ is most steep there.

Table 4 shows size and power of BD test. It is easily find that the pattern of the simulation results in Table 4 is quite similar to that in Table1.

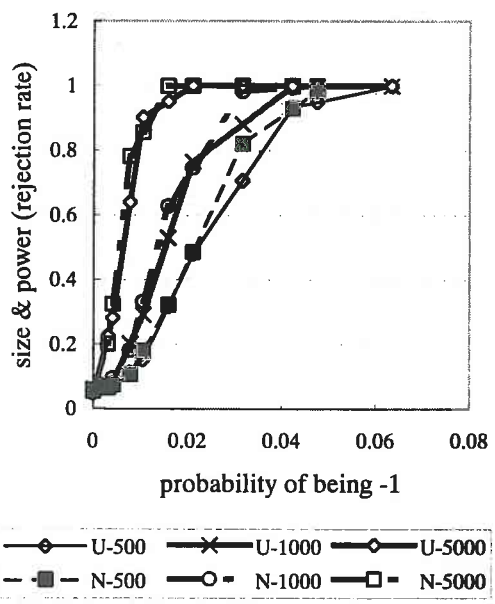

Figure 1 Power comparison on uniform (U-n) and Normal (N-n) case; controlled probability jump.

To make fair comparison with uniform case, we control the probability at benchmark value. For uniform case, $\delta = 0.20$ means that probability of being (-1) is about 0.042. On the other hand, for normal case $\delta = 0.30$ corresponds that probability of being (-1) is $0.30 \times \Phi(-1) = 0.048$. Figure 1 represents power functions of uniform case and normal case against amount of probability jump at (-1).

Table 4 Size and power of BD test (normal case; at x=-1).

n	*delta*	h = 0.20	0.10	0.05
50	0.00	0.086	0.082	0.015
	0.10	0.045	0.052	0.025
	0.20	0.083	0.101	0.089
	0.30	0.141	0.179	0.201
100	0.00	0.055	0.098	0.062
	0.10	0.059	0.091	0.099
	0.20	0.135	0.219	0.292
	0.30	0.261	0.383	0.521
500	0.00	0.049	0.052	0.071
	0.05	0.080	0.105	0.164
	0.10	0.203	0.322	0.477
	0.20	0.580	0.820	0.934
	0.30	0.905	0.983	0.999
1000	0.00	0.055	0.047	0.057
	0.05	0.136	0.181	0.327
	0.10	0.377	0.629	0.812
	0.20	0.909	0.983	1.000
	0.30	0.999	1.000	1.000
5000	0.00	0.040	0.057	0.067
	0.01	0.065	0.104	0.108
	0.02	0.116	0.202	0.336
	0.05	0.471	0.782	0.945
	0.10	0.967	0.999	1.000
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000
10000	0.00	0.044	0.052	0.042
	0.01	0.081	0.111	0.204
	0.02	0.215	0.310	0.567
	0.05	0.818	0.968	1.000
	0.10	1.000	1.000	1.000
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000

(Note) Some of the results are unlisted.

For both cases, means and variances of distributions from which sample data has drawn are same. From Figure 1, we can find that two power functions have almost same form for each sample size. Thus, we may conclude that BD test seems to be robust for the distribution by which sample data is generated. It might be noted that we can perfectly detect the discontinuity if probability jump is 0.06 for n=500, 0.04 for n=1000, and 0.02 for n=5000.

4.4 Chi-square Case

In this subsection, we examine a case that is more difficult, i.e. tougher case, than normal case. Suppose sample data is drawn from Chi-square distribution. We assume $g(x) \sim \chi^2(8)$. In this case, distribution mean is eight and distribution variance is 16. Then, we set h as 0.8, 0.4, and 0.2. As in other two simulations, we set the benchmark value as mean minus one standard deviation, that is to say $c^* = 4$.

Table 5 shows size and power of BD test. The pattern of the simulation results in Table 5 is also quite similar to that in Table1 and Table 4.

Table 5 Size and power of BD test (Chi-square case; at x=4).

		h		
n	*delta*	0.8	0.4	0.2
50	0.00	0.063	0.083	0.052
	0.10	0.059	0.060	0.050
	0.20	0.071	0.084	0.089
	0.30	0.102	0.151	0.139
100	0.00	0.057	0.080	0.104
	0.10	0.059	0.078	0.098
	0.20	0.107	0.144	0.206
	0.30	0.184	0.257	0.420
500	0.00	0.052	0.044	0.067
	0.05	0.093	0.090	0.107
	0.10	0.170	0.229	0.337
	0.20	0.474	0.648	0.853
	0.30	0.793	0.921	0.988
1000	0.00	0.053	0.048	0.044
	0.05	0.125	0.133	0.209
	0.10	0.319	0.393	0.626
	0.20	0.769	0.913	0.985
	0.30	0.977	0.998	1.000
5000	0.00	0.070	0.047	0.054
	0.01	0.142	0.083	0.084
	0.02	0.217	0.180	0.206
	0.05	0.536	0.570	0.814
	0.10	0.934	0.987	0.998
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000
10000	0.00	0.127	0.058	0.047
	0.01	0.237	0.084	0.120
	0.02	0.359	0.256	0.421
	0.05	0.833	0.864	0.981
	0.10	0.999	1.000	1.000
	0.20	1.000	1.000	1.000
	0.30	1.000	1.000	1.000

(Note) Some of the results are unlisted.

However, there are two exceptions. First, for n=50 and $\delta = 0.30$, the power of h=0.2, 0.139, is less than that of h=0.4, 0.151. A possible reason for this is to adopt narrow bin-width. By using narrow bin-width relatively to small sample size, $\hat{p}_j$ might become poor estimation. As a result, estimation on denominator in (2) becomes poor, then, it could not rejected so often. The fact that the power of h=0.4 is 0.299 and that of h=0.2 is 0.401 when we use true variances may confirm this point.

Second, for n=10000, size is far from true value, 5%, regardless large sample size. In this case, what bin-width is too large to distinguish continuous increment and jump is one plausible explanation.

5. SUMMARY AND CONCLUSION

We can summarize the results as follows. First, we can understand that the test is based on estimators of density (i.e. empirical density) at a certain point for various bin-width and is derived using multinomial distribution. Second, from the results by Monte Carlo simulation, size and power of the test depends on sample size and bin-width. Over moderate sample size, the size of the test is almost correct and the test has good power. In addition, we find that the test is able to detect discontinuity for a small jump in continuous distribution. We also made simulations for various continuous distributions, and we have almost same results. Thus, we conclude that the test is available for various situations

6. REFERENCES

Burgstahler, D., and I. Dichev, Earnings management to avoid earnings decreases and losses, *Journal of Accounting and Economics*, 24, 99–126, 1997.

Devroye, L. and L. Györfi, *Nonparametric Density Estimation: The L_1 View*, John Wiley, 356 pp., New York, 1985.

Kan-no, H., H. Takao and Y. Takeuchi, Statistical models for earnings management, unpublished manuscript, Osaka University, 2001.

McNichols, M.F., Research design issues in earnings management studies, *Journal of Accounting and Public Policy*, 19, 313–345, 2000.