Journal of Modern Accounting and Auditing, ISSN 1548-6583
June 2011, Vol. 7, No. 6, 645-651

Annual Earnings Analysis with ARIMA for Future Earnings Prediction

Wuryan Andayani
Brawijaya University, Indonesia

Junaidi
Yogyakarta Technology University, Indonesia

Nurdiono
Lampung State University, Indonesia

This study investigates annual earnings analysis with ARIMA (Autoregressive Integrated Moving Average) for future earnings prediction. Earnings prediction is very important to be used in various aspect of decision making process, such as: investor, creditor, analyst, academicians, practitioners, etc.. Evidence supports the ARIMA model that it is more accurate. It also has a smaller size of error value.

Keywords: annual earnings analysis, future earnings prediction, ARIMA

Introduction

Lorek, Willinger, and Bathke (2008) stated that earnings prediction is very important to be used in various aspects of decision making process, such as: (1) Investor may have good information in helping them to make any investment making decision. Earnings is viewed as an attractive case by investor because it provides information of the future cash flow (FASB, 1994; Elliot, 2006); (2) Earnings can help researchers in comprehending more the time series property of quarterly earnings data for non seasonal companies versus seasonal; (3) Specific knowledge about earnings provides statistic proxy development for earnings persistence (Collins & Khotari, 1989; Lipe & Kormendi, 1994).

Furthermore, according to Foster (1986), the various parties who use financial prediction are the security analysts, loan institutions, and companies' management. Whereas, other parties which conducted earnings prediction are investors, creditors, government, and other parties who may carry out predictions of financial distress condition or a company analysis of a bankrupt condition (Atmini & Andayani, 2005). The research result of McCue (1991) and Atmini and Andayani (2005), revealed that the prediction ability of the earnings model is somewhat bigger than the prediction ability of the cash flow model.

Ball and Brown (1968) conducted a research concerning information content found in the earnings. Their result showed a significant relationship between unexpected earnings and abnormal return. There were several researches about earnings that have been done by Beaver (1970), Lipe (1986), Bernard and Stober (1989). Ball and Watts (1972) examined earnings prediction with run test and serial correlation. Their result indicated that the alteration in the earnings has a random walk model. With the time series mode and a random changing

Wuryan Andayani, lecturer in Accounting Program, Faculty of Economics and Business, Brawijaya University Malang.
Junaidi, lecturer in Accounting Program, Faculty of Economics and Business, Yogyakarta Tecnology University.
Nurdiono, lecturer in Accounting Program, Faculty of Economics and Business, Lampung State University.

pattern, earnings have shown itself as a good factor of predictor.

This study used ARIMA which is an association of AR (Auto Regressive) and MA (Moving Average) which are deferent. This concept is based on assumption that the value of the current data is influenced by the value of the past data. This ARIMA model is known as the Box Jenkins method (Enders, 2004).

This study is related to the AISC (The First Annual Indonesian Scholars Conference in Taiwan 2010) because the analysis tools could improve national competitiveness in the prediction of annual earnings analysis with ARIMA for future earnings prediction. It is very important to be used in various aspect of decision making process, such as investors, creditors, analysts, academicians, practitioners, etc.. This evidence supports the ARIMA model that it is more accurate and has a smaller size of error value.

Literature Review

The Importance of Using ARIMA Model and AR Higher Order

Callen (2009) quoted that several past accounting reviews has describe that the time series of the annual earnings is actually has a random walk character. The character of the time series earnings with a random changing pattern indicates that it has potential to be a predictor. Many companies which are described by using high order process showed that: (1) Time series property for all companies are suitable with the AR(1) process restriction; (2) Accounting literatures illustrated that high order model has a descriptive character than the time series of the annual earnings; (3) Accounting literatures for time series property of interim earnings such as quarter earnings is constructed better as a complex time series process rather than as AR(1); (4) Higher order auto regressive process and ARIMA process can better describe the time series movement as by AR(1).

The ARIMA process and higher order auto regressive (AR) are very important in accounting researches. In accounting, there are permanent accounting shocks such as the changing of standard which is described with a permanent time series process by using AR. There are also shocks which have a transitory characteristic such as special items that are described with MA process. Accounting researches which introduced ARIMA process were of: Beaver, Lambert, and Morse (1980), Ali and Zarowin (1992), and Kumar and Visvanathan (2003).

Generally, ARIMA tends to be more advanced because this method only needs distinctive variable which is also the variable itself in the past, and ARIMA has a smaller defect than the GARCH Model (Nachrowi & Hardius, 2007).

Time Series Analysis

Time series analysis is a set of data which is analyzed within a period of time which has past to predict a condition in the future. Gujarati (1995) confirmed that many human behaviors are influenced by the condition or the time in the past. Enders (2004) stated that econometrics experts have developed a simple model that can do prediction, interpret, and hypotheses testing which related to economics data.

Time series analysis has been developed in various ways such as smoothing method, decomposition method, and ARIMA method by using Box Jenkins approach, and moving average approach. In ARIMA, there are many component p (auto regressive), q (moving average), and d (deferent) which are needed so that the data can be stationer. Therefore, Box Jenkins can give directions that consist of four steps (Povinelli, Bowerman, & O’Connell, 1993):

(1) The identification of p, d, and q value. As a consequence, we use correlogram and partial correlogram;

(2) Auto regressive parameter estimation and moving average component which exist in the model. We can use Least Square method or non linier estimation, but computer program nowadays can simply manage

them easily;

(3) The diagnosis towards the model quality, whether it has suited with the kind of data. It is by testing whether estimation result residual has already got it White Noise character. If the residual has already White Noised, it means that it has got its appropriate model. If it has not, then we may have to search another ARIMA. The appropriate model searching process has to be done persistently;

(4) Forecasting future data by using equations or the chosen model.

Research Design

The data is taken from the Indonesian Capital Market Directory (ICMD) within the period of 1987-2007, with criteria of:

(1) It has got time series data;

(2) Earnings data of companies that are listed in Indonesia Stock Exchange (IDX) within the period of 1987-2007;

(3) The consistent availability of the data needed.

The firms are included in this sample about 15 public firms (Table 1).

Table 1

Names of Companies

No.	Names of companies	Name of Emiten	Period
1	Pt Unggul Indah Corporation Tbk	UNIC	1987-2007
2	Pt Indocement Tunggal Perkasa Tbk	INTP	1987-2007
3	Pt Semen Cibinong Tbk	SMCB	1987-2007
4	Pt Citra Turbindo Tbk	CTBN	1987-2007
5	Pt Jaya Pari Steel Tbk	JPRS	1987-2007
6	Pt Japfa Comfeed Indonesia Tbk	JPFA	1987-2007
7	Pt Berlina Co Ltd Tbk	BRNA	1987-2007
8	Pt Sepatu Bata Tbk	BATA	1987-2007
9	Pt Astra International Tbk	ASII	1987-2007
10	Pt Good Year Indonesia Tbk	GDYR	1987-2007
11	Pt Aqua Golden Mississippi Tbk	AQUA	1987-2007
12	Pt Delta Jakarta Tbk	DLTA	1987-2007
13	Pt Multi Bintang Indonesia Tbk	MLBI	1987-2007
14	Pt Bat Indonesia Tbk	BATI	1987-2007
15	Pt Unilever Indonesia Tbk	UNVR	1987-2007

The method used in this study is the ARIMA method (Box Jenkins). The parameters were calculated by using non seasonal data ARIMA (Gujarati, 1995):

(1) Auto Regressive Model

$$Y_t = Ø_0 + Ø_1 Y_{t-1} + Ø_2 Y_{t-2t} + \ldots Ø_p Y_{t-p} + \epsilon_t \quad (1)$$

where:

Y_t = dependent variable (Net Profit);

Y_{t-1}, Y_{t-2t}, Y_{t-p} = independent variables (variables which have certain lag);

$Ø_0, Ø_1, Ø_2, Ø_p$ = the coefficients of regression;

ϵ_t = error.

The sum of the coefficients of regression is written as “*p*”.

(2) Differencing (degree of differencing)

The condition to conduct ARIMA analysis is the time series data has to be stationer. If the data is not stationer, then they have to be altered by doing a differencing.

Y'_t = one new time series data;

Y_t = one initial time series data;

Y_{t-1} = one initial lag 1 time series data.

If by performing the first differencing, the data has not yet stationer, then we need to perform the second differencing with the formula as below:

$$Y''_t = Y'_t - Y''_{t-1} \tag{2}$$

where:

Y''_t = one time series data with second differencing;

Y''_{t-1} = one new time series data with lag 1.

Generally, the maximum is only until the second order differencing, and differencing is written as “*d*”.

The moving average model is formulated as:

$$Y_t = W_0 + \epsilon_{t-1} - W_1\epsilon_{t-1} - W_2\epsilon_{t-1} + Wq\epsilon_{t-q} \tag{3}$$

where:

Y_t = Dependent variable;

W_1, W_2, Wq = coefficients;

ϵ_t = Error;

$\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q}$ = the error value of lag 1, 2, and so on.

The coefficient (parameter) of the model is written as “*q*”.

The whole three components above are called ARIMA (*p*, *d*, *q*) and the formulas can be combined to be ARIMA (1, 1, 1).

$$Y_t = \varnothing_1 Y_{t-1} + \epsilon_t + W_1\epsilon_{t-1} \text{ or } ARIMA\ (1, 1, 1) \tag{4}$$

$$Y_t = (1 + \varnothing_1)\ Y_{t-1} - \varnothing Y_{t-2} + \epsilon_t + W_1\epsilon_{t-1} \tag{5}$$

The Steps of Box Jenkins:

(1) Identifying the model, is a step on determining whether the time series data is already stationer or not. If not, then it may have to be altered so that it can be stationer by doing through differencing, which is the difference between period *t* and period *t-1*. The correlogram of the time series data that has been stationer generally will extremely decrease if lag (*k*) is getting bigger.

(2) If the time series data is already stationer, then we may formulate the model as well as testing the goodness of fit, in order to search the best model.

(3) Predicting the model. To measure the prediction error, we have to focus on the dispersion. According to Foster (1986), there are two kinds of measuring of dispersion that can be done, they are the mean absolute error (MABE) and the mean square error (*MSE*).

$$MABE_i = \frac{1}{N}\sum_{t=1}^{N}|\ Y_{i,t} - E(Y_{i,t})\ | \tag{6}$$

$$MSE_i = \frac{1}{N}\sum_{t=1}^{N}|\ Y_{i,t} - E(Y_{i,t})^2\ | \tag{7}$$

$Y_{i,t}$ = the realization of the forecast variable in period *t* for companies;

$E\ (Y_{i,t})$ = the forecast variable in period *t* for company *i;*

N = the sum of forecasting that being tested.

Analysis

Stationery Testing

The stationery testing shows that UNIC Company has not yet stationer, which is shown by the value of ADF (Augmented Dickey Fuller), that is -2.693603 > critical value -3.8304. The ADF value is above the critical value on any degree. After conducting the first differencing, the ADF value is -4.073854 < -3.8572; and on another level of degree. Therefore, the data has already stationer. The stationer testing procedure is done for the whole testing by using the Box Jenkins model.

Identifying Tentative Model Used

The pattern of AR (1), AR (2), and MA (2) are not significant. After estimating the tentative model, then we may search for the best model. Table 2 shows that the fit model, where AR (1) and MA (2) are significant, accompanied by F statistic of 12.27093 (with a significance of 0.000696), and *R* squared of 0.6206555, as well as Adjusted *R* squared of 0.570076. Beside, the model also has a minimum *RMSE*, which is one of the criteria to evaluate the prediction error. The formula to calculate the *RMSE* is:

$$RMSE = \sqrt{\sum_{t=1}^{N}(Yt - Y')^2 \;/\; N} \quad (8)$$

Table 2

ARIMA Estimation Model

No.	Companies	Model	oefficient	AR (1)	AR (2)	MA (1)	MA (2)	Adj. *R* Sq.	*F*
1	UNIC	ARIMA	-1,005.068	-0.714631			-0.979557	0.570076	12.27093
		(1, 1, 2)	(-0.907593)	(-3.869688)			(-1935.368)		
2	INTP	ARIMA	71,167.02	-1.615983			-2.610314	0.773302	29.99475
		(1, 1, 2)	(0.865177)	(-4.313163)			(-1.841821)		
3	SMCB	ARIMA	-3,334.540	-0.660105			-0.871984	0.366067	6.197085
		(1, 0, 2)	(-0.369613)	(-3.125874)			(-9.9748)		
4	CTBN	ARIMA	33,529.56	-0.660650		-1.971309		0.911187	83.07659
		(1, 1, 1)	(1.023284)	(-3.216880)		(-3.326182)			
5	JPRS	ARIMA	-2,149.919	-0.762274			-0.979982	0.621807	15.79740
		(1, 0, 2)	(-1.283248)	(-10.09650)			(-3828.939)		
6	JPFA	ARIMA	5,568.032	-0.782141			-0.843145	0.416946	7.435975
		(1, 0, 2)	(1.152740)	(-4.166633)			(-5039132)		
7	BRNA	ARIMA	-3,619.830	-0.636085			-1.342109	0.691518	21.17516
		(1, 0, 2)	(-1.532058)	(-3.137440)			(-19.67419)		
8	BATA	ARIMA	-58,638.07	-0.830444			-0.979988	0.583903	13.62960
		(1, 0, 2)	(-1.316581)	(-8.942278)			(-2622.926)		
9	ASII	ARIMA	25,728.67		-1.155409		0.836450	0.000216	1.001727
		(2, 1, 2)	(0.066510)		(-4.244129)		(8.788780)		
10	GDYR	ARIMA	728.6732	-0.601252		-0.989840		0.838147	42.42745
		(1, 2, 1)	(0.9472201)	(-3.279951)		(-3518.943)			
11	AQUA	ARIMA	-324.3570	-0.689476			-0.979998	0.668650	18.15260
		(2, 1, 2)	(-0.687570)	(-4.699448)			(-2428.629)		
12	DLTA	ARIMA	-458.2072	-0.410785			-0.975132	0.797612	34.49855
		(1, 1, 2)	(-0.375970)	(-3.108364)			(-37.92573)		
13	MLBI	ARIMA	277.4483	-0.940434			-1.201814	0.663722	17.77672
		(1, 1, 2)	(0.689226)	(-6.625267)			(-12.63079)		
14	BATI	ARIMA	1,430.759	-0.802475		-0.989565		0.825467	38.83651
		(1, 2, 1)	(0.911687)	(-4.050066)		(-3716.248)			
15	UNVR	ARIMA	-749.7570	-0.475920			-0.914534	0.772749	28.20331
		(1, 2, 2)	(-0.069637)	(-2.907742)			(-28.85513)		

We obtain the ARIMA model (1, 1, 2) and so the formula is set as follows:

$$D\ (Earnings,\ 2) = -1005.068 + (-0.714631)\ AR\ (1) + (-0.979957)\ MA\ (2) \quad (9)$$

Further, the ARIMA estimation model for *earnings* of all samples of companies can be seen in Table 2.

The best model estimation result for earnings of all samples of companies can be seen on the tables. As being stated before, ARIMA is a combination of three components; they are the auto regressive (AR) or "*p*", differencing or "*d*", and the moving average (MA) or "*q*". So the ARIMA model will be:

$$D\ (Earnings,\ 2) = -1005.068 - 0.714631\ AR\ (1) - 0.979957\ MA\ (2) \quad (10)$$

From Table 2, we can see that all samples of companies have AR and MA pattern. AR is a current period data that is influenced by the residual value of the precedent period. The empirical result shown on Table 2 indicated that the ARIMA model can give a better and more accurate result. The result of this study is supported by Nachrowi and Hardius (2007) who stated that the ARIMA model is more superior to the GARCH. Beside, it is also supported by Callen (2009) who stated that ARIMA can better illustrate the time series movement.

Forecasting With the Model

After obtaining the best model, then we can view the standard error by seeing the size of RMSE, MABE, and MAPE as shown in Table 3.

Table 3
RMSE, MABE, and MAPE

No.	Names of companies	RMSE	MABE	MAPE
1	UNIC	59,857.47	41,606.06	101.6361
2	INTP	1,409.308	1,082,438	146.5877
3	SMCB	209,765.7	95,939.80	117.4914
4	CTBN	6,393,451	2,979,181	134.8139
5	JPRS	60,691.90	39,971.04	117.5957
6	JPFA	104,644.9	56,653.92	288.7707
7	BRNA	619,760.5	367,196.3	257.0443
8	BATA	1,840,540	1,221,354	544.2106
9	ASII	1,407,616	1,200,298	5,092.156
10	GDYR	55,575.28	39,227.43	98.00591
11	AQUA	29,500.80	22,791.83	100.080
12	DLTA	55,553.19	39,386.09	103.8596
13	MLBI	58,559.72	37,094.69	99.98709
14	BATI	80,338.2	113,394.2	802.2741
15	UNVR	306,959.4	262,380.9	113.4359

Because the earnings data is set in millions, so we found that the error is small. It indicates that ARIMA has a small size of error, and it is also supported by the residual plot. The distance between the actual value and the fitted value is small, where the actual point is stated near the fitted point. Only on certain years that some of the companies gain a far distance between its actual value and fitted value, and it is presume that the company maybe carrying out an earnings management. The result of the residual plot, RMSE, MABE, and MAPE can be seen on the tables.

Conclusion

From the empirical result indicates that the ARIMA model is more accurate and has a smaller size of error value. This is supported by the study conducted by Nachrowi and Hardius (2007) where they stated that the

ARIMA model is more superior to the GARCH. It is also supported by Callen (2009) which also confirmed that the ARIMA model can better illustrate the time series movement. Therefore, it can be said that ARIMA is very important to be used in accounting research (Callen, 2009; Beaver et al., 1980; Ali & Zarowin, 1992; Kumar & Visvanathan, 2003; Lorek et al., 2008).

This research is hopefully useful for accounting literatures illustrated by Callen (2009) that high order model has a descriptive character than the time series of the annual earnings. This are also useful for time series property of interim earnings such as quarter earnings is constructed better as a complex time series process rather than as AR (1) and higher order auto regressive process and ARIMA process can better describe the time series movement as by AR (1).

There are some limitations in this study: First, the period observation is from 1987 until 2007; Second, the samples are about 15 public firms. The two limitations may be effect on generalization of the results.

References

Ali, A., & Zarowin, P. (1992). The role of earnings level in annual earnings-return studies. *Journal of Accounting Research, 30*, 286-296.

Atmini, S., & Andayani, W. (2005, September 15-16). The benefit of earnings and cash flow to predict financial distress condition in textile mill products and apparel and other textile product. Paper presented at *SNA VIII Solo*, 460-474.

Ball, R., & Brown, P. (1968, Autumn). An empirical evaluation of acconting income numbers. *Journal of Accounting Research,* 159-178.

Ball, R., & Watt, R. (1972, June). Some time properties of accounting income. *Journal of Finance,* 663-682.

Beaver, W., Lambert, R., & Morse, D. (1980). The information content of security prices. *Journal of Accounting and Economics, 2*(1), 3-28.

Beaver, W. H. (1970). The time series behavior of earnings supplement. *Journal of Accounting Research,* 663-682.

Bernard, V., & Stober, L. (1989, October). The nature and amount of information in cash flows and accrual. *The Accounting Review,* 624-652.

Callen, J. L. (2009). Shocks to shocks: A theoretical foundation for the information content of earnings. *Contemporary Accounting Research, 26*(1), 135-166.

Collins, D., & Kothari, S. (1989). An analysis of intertemporal and cross sectioal determinants of earnings response coefficient. *Journal of Accounting Economics, 11*, 143-182.

Elliot, W. B. (2006). Are investors influenced by pro forma emphasis and reconciliations in earnings announcements? *The Accounting Review, 81*(1), 113-133.

Enders, W. (2004). *Applied econometric time series* (2nd ed.). J. Wiley.

Financial Accounting Standards Board (FASB). (1994). *Statements of financial accounting concepts: Accounting standards.* Illinois: Burr Ridge, Irwin.

Foster, G. (1986). *Financial statement analysis* (2nd ed.). Prentice Hall International.

Gujarati, D. N. (1995). *Basics econometrics* (3rd ed.). McGraw Hill International Editions.

Kumar, K. R., & Visvanathan, G. (2003). Information content of defferred tax valuation allowance. *The Accounting Review, 78*(2), 471-490.

Lipe, R., & Kormendi, R. (1994). Mean reversion in annual earnings and its implications for security valuation. *Rev Quant Finance, 4*, 27-46.

Lipe, R. C. (1986). The information contained in the components of earnings. *Journal of Accounting Research, 24*, 37-64.

Lorek, K., Willinger, G., & Bathke, A. (2008). Statistically based quarterly earnings expectation models for non seasonal firms. *Rev Quant Finance Accounting, 31*, 105-119.

McCue, M. J. (1991). The use of cash flow to analyze financial distress in california hospitals. *Hospitals and Health Service Administration, 36*, 223-241.

Nachrowi, N. B., & Hardius, U. (2007). Prediksi IHSG dengan model GARCH dan model ARIMA. *Jurnal Ekonomi dan Pembangunan Indonesia, 7*(2).

Povinelli, R. J., Bowerman, B. L., & O'Connell, R. T. (1993). *Forecasting and time series an applied approach* (3rd ed.). Belmot, California: Duxbury Press.