Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, **2015**, *9*, 2559-2564

Open Access

Forecast and Analyze the Telecom Income based on ARIMA Model

Mingzhao Wang*, Yuping Wang, Xiaoli Wang and Zhen Wei

School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, 710071, China

Abstract: With the increasing competition in the telecommunications industry, the operators try their best to increase telecom income *via* various measures, one of which is to set an amount of income as a goal to make the encouragement. Since accurate forecast of income plays an important role in income target setting, this paper builds a time series Autoregressive Integrated Moving Average Model (ARIMA) based on the analysis of income data. Two important issues are involved when setting up the ARIMA model: first, smooth the old data and identify the model, and then estimate the parameters in the model by SPSS software. As a result, we set up an ARIMA (1,2,1) model with order 1 auto-regression, order 2 difference and order 1 lag. Finally, computer simulations are made based on the real data from a telecommunication company and experimental results show that the proposed model fits income data well and performs well in forecasting.

Keywords: ARIMA model, forecast, SPSS, telecom income.

1. INTRODUCTION

With the development of the telecommunications industry, rapid expansions of market demands have gradually become saturated. Operators convert their attention from users to income. They try their best to increase telecom income via various measures, one of which is to set an amount of income as a goal to make the encouragement. Therefore, accurate forecast of income has attracted more and more attention. The income of telecommunications industries can be divided into two parts, business income and subsidy income. Business income can also be further divided into main business income and other business income. Main business income mainly refers to the revenue which includes the sales of finished products and semi-finished products as well as providing industrial services, while other business income refers to part time sales and other business activities beyond the company's basic business. Subsidy income refers to government subsidies which are income-related, and government grants related to assets, etc. The complexity of the income structure determines that it may difficult to forecast the income on details of the revenue since it needs to collect large amounts of data [1-4]. Moreover, if there exists a small deviation in the part of the data, it will be amplified during the accumulation, and the final estimation of income will have a great deviation and thus makes it no sense for the result.

Many factors affect the income, such as the demand of the market, the policy of subsidies, and the invention of new products, etc. There are a lot of issued to be determined if we analyze the income from the relationship between factors and the variation of income. Moreover, most of the factors in this part only have certain relationship other than precise data values. Therefore, most of the previous studies adopt qualitative analysis.

Considering the aforementioned problems, this paper adopts a time series model ARIMA [5, 6] instead of analyzing the revenue breakdown and the factors affected the income. The main advantage of the ARIMA model used for revenue forecasting is that economic rules or social rules of income indicators don't need be considered. Since it is very difficult to get these rules, and some of them can only be quantitatively analyzed, while some can only be qualitatively analyzed, and also some are even unpredictable factors. ARIMA model starts from history data, making use of mathematical methods to mine the regularity and invisible point of data. Only two indicator values are needed for the ARIMA model to predict revenue in future: time and income, which are very easy to get. Meanwhile, ARIMA model has no restriction on the regularity of data. Therefore, seasonal and periodical data can also be applied to the model by certain transformation. Compared to ARIMA model, regression analysis requires that the variance of the sample data are equal and independent, and neural networks is not conducive to be widely used since its operations and the principles are relatively complex although it has high accuracy. In recent years, as a result of its obvious advantages, ARIMA model has applied in many areas [7-9]. It is also used to forecast and analyze telecom income. But the problems are that these available analyses are not exhaustive on the description of the application of the ARIMA model and the explanation of parameter identification is not clear enough, or even skip some of the steps in the model. Moreover, most analyzes of the ARIMA model are made by the Eviews software, and the corresponding explanation and analysis of parameters are also based on the results in connection with Eviews. It is weird that as one of the commonly used analytical software, SPSS software is rarely used for the application of the ARIMA model. Therefore, this paper follows the general process of the ARIMA model, using SPSS to establish the ARIMA time series model, and make a detailed analysis for each parameter, judgment, and step in this model, mining internal rules contained in the data to get more accurate predictive value.

1874-110X/15

2. BACKGROUND KNOWLEDGE

2.1. Introduction of the Time Series

2.1.1. Stationary Time Series

Stationary time series refers to the statistical rules of time series which do not vary over time. Intuitively, a stationary time series can be seen as a fluctuating curve around its average value. Theoretically, there are two kinds of stability, one of which is strictly-sense stationary or completely stationary, while the other is wide-sense stationary or generalized stationary.

The definition of strictly stationary conditions is relatively harsh. The series can be considered stable only when all of the statistical properties of the series do not change over time.

Wide-sense stationary is defined by using statistical measurements of the characteristics of series. It involves the statistical properties of the series determined by its lower order moments. A time series can be considered wide-sense stationary if its average value and variance functions are constant, and its correlation coefficient does not change over time.

2.1.2. White Noise Series

White noise series is a special stationary series. It is defined as follows: if a random series $\{y_t\}$ is constituted of unrelated random variables, that is, for all $s \neq t$, $Cov(y_s, y_t) = 0$, and then it is called the white noise series. White noise series is a stationary series, whose random variable covariance is 0 at different points. This feature is often called "no memory", which means its changes do not follow any rules and people cannot speculate its direction in future based on its history characteristics. The model can be considered to have achieved good results when its residual series become white noise series and there remains no more information can be identified in the deviance residuals.

2.2. ARIMA (p, d, q) Model

Autoregressive Integrated Moving Average Model (ARIMA model) [10] is commonly used on fitting stationary random series. It is proposed by Box and Jenkins, thus it is also called Box-Jenkins model. Its basic idea can be stated as follows: treat the data series formed by the prediction target with time as a random series, and describe the series with a mathematical model approximately. It can be used to forecast desired values from the past and present values once the model is identified. This model specifies three parameters to analyze the time series: autoregressive order (P), moving average order (q) and the number of differential made to be a stationary series (d), generally called ARIMA (p, d, q). The form can be writen as follows:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (1)$$

where Y_t is the observed value of time series at the stage of t, e_t is the deviation of time series at the stage of t, $\{\varphi_1, \varphi_2, \cdots, \varphi_p\}$ is the autoregressive coefficient, and $\{\theta_1, \theta_2, \cdots, \theta_q\}$ is the moving average coefficient.

When $p = 0$, the model is called the moving average model, denoted as MA(q); when $q = 0$, the model is called autoregressive model, denoted as AR (p).

2.3. Modeling Process

Fig. (**1**) shows the general modeling process of ARIMA. First, smooth the data, and then identify the model, followed by estimating the parameters. After residual test and data fitting, forecast will start.

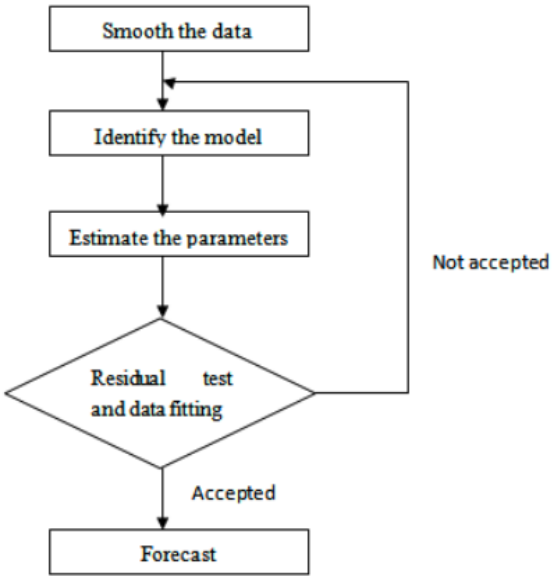

Fig. (1). The process of building ARIMA model.

3. THE APPLICATION OF ARIMA MODEL IN TELECOM INCOME

3.1. Data Collection and Smoothing

The data in this study comes from the income of China Telecom business analysis systems. The data during 2011.1 to 2012.12 of a province will be used as the analysis data, whose income-time chart is shown in Fig. (**2**), while the data from 2013.3 to 2013.5 will be used to validate the fitting of the model to the data, and finally forecast the data in June 2013 by the model.

The time series is required to be a stationary series with zero average value. Here the series only required to satisfy wide-sense stationary. So it is necessary to test the time series on the following three aspects: zero average value, constant variance, and correlation coefficient only related to time interval and independent of specific time. ARIMA model can be used only when time series meets the above three condition. It can be seen from Fig. (**2**) that income series is an annual cycle. It shows an increasing tendency, and the income value does not fluctuating around zero. Therefore, the series does not match ARMAI model since it is not a null means stationary series.

The method of generating non-stationary time series stationary involves logarithm and differential. The series Z_t can be obtained by logarithmic the original time series Y_t once and then differential it twice. Z_t will be divided into two sub-series on an annual basis [11]. Then, verify the

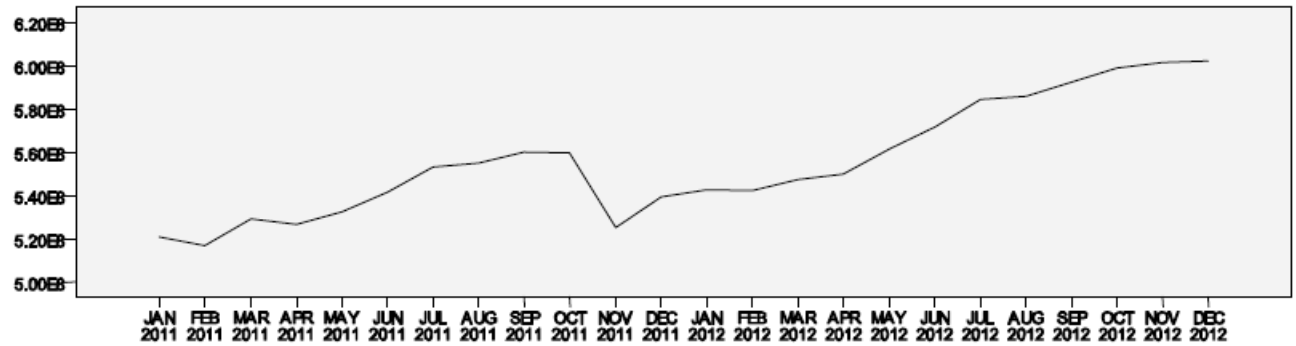

Fig. (2). Income-time chart of a province of China Telecom from 2011.1 to 2012.12.

smooth of the new series on three aspects: average value, variance, and correlation coefficient as follow.

Tables (**1-3**) show the average values of sub-series, the variance of sub-series, and the correlation coefficient of sub-series, respectively. In Table **1**, column 2 shows the number of data in the sub-series, columns 3, 4, and 5 lists the average value, the standard deviation, and the variance of the two sub-series. As shown in Table **2**, the data in row 1 involves group variances while row 2 involves variances inside groups. Columns 2 to 5 lists the square of deviance, the degree of freedom, and the F equals to the ratio of group variances and variances inside groups. The significance coefficient Sig. always equals a significance level of 0.05 or 0.01. Negate the hypothesis when the significant coefficient below the critical value, which means that there are no significant differences in the average value in the different groups. Tables **1** and **2** show that the average value of two sub-series in 2011 and 2012 is close to 0, that the variability of the average value and variance is small, and that the significant coefficient 0.665 > 0.05. As a result, the related data shows that the hypothesis is true. This means that time has no significant influence on income. The data listed in row 1 in Table **3** shows that the correlation coefficient of sub-series in 2011 and 2012 is 0.244

Table 1. Test of average values of sub series.

Year	N	Average Value	Standard Deviation	Variance
2011	10	0.0034	0.04220	0.01335
2012	12	0.0021	0.01108	0.00320
Total	23	0.0004	0.02891	0.00616

Table 2. Test of variance of sub series.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.000	1	0.000	0.193	0.665
Within Groups	0.017	20	0.001	----	----
Total	0.018	21	----	----	----

Table 3. Test of correlation coefficient of sub series.

		2011	2012
2011	Correlation Coefficient	1.000	0.244
	Sig. (2-tailed)	.----	0.325
	N	10	10
2012	Correlation Coefficient	0.244	1.000
	Sig. (2-tailed)	0.325	-----.
	N	10	12

(below 0.3), while the data in row 2 involves the probability of the assumption (the sub-series of 2011 and 2012 are uncorrelated) that the correlation coefficient equals 0 is true. The two data shows that the two sub-series have weak correlation and almost have no dependency. Through the above analysis, we know that after logarithmic the original time series Y_t once and differential it twice, Z_t has become a stationary series and the ARIMA model can be used.

3.2. Model Identification

Identification of the model involves two steps: first, determine which model should be adopted, AR(P) model , MA(q) model or ARIMA(p, d, q) model, and then determine the value of p, d, q in Eq.(1). The two identifications depend on the property of ACF and PACF, which is shown in Table **4**. Figs. (**3** and **4**) involve the analysis of autocorrelation and partial correlation about series Z_t. It indicates that the ARIMA (p, d, q) model can be adopted as the two figures with tailing. As a result, the above analysis indicates that differential the original time series Y_t twice can obtain stationary series Z_t, thus parameter d equals to 2. Note that the autocorrelation and partial correlation begin to decay from k = 2, thus p and q equal to 1 or 2. Therefore, we get four possible models: ARIMA(1,2,1) model, ARIMA(1,2,2) model, ARIMA (2,2,1) model , and ARIMA (2,2,2).

Table 4. Criteria of order determination of ARIMA model.

ACF	PACF	Order Determination of Model
Tailing	Order p tailing	AR(p)model
Order q tailing	Tailing	MA(q)model
Tailing	Tailing	ARIMA(p,q)model

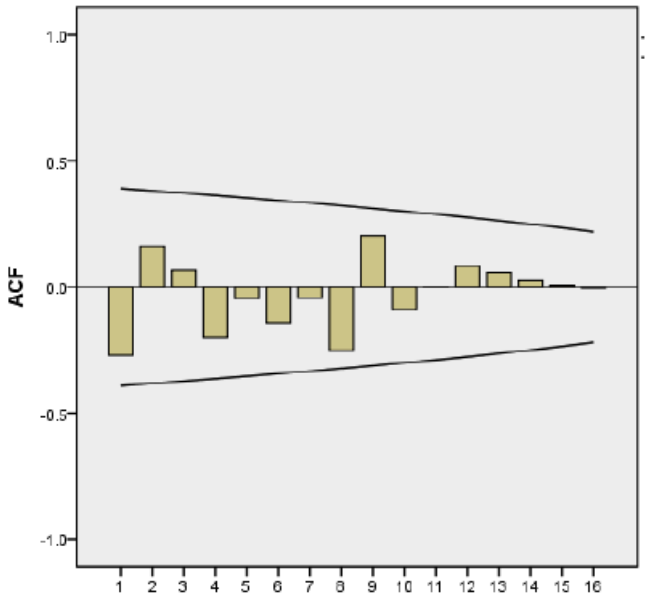

Fig. (3). Autocorrelation of series.

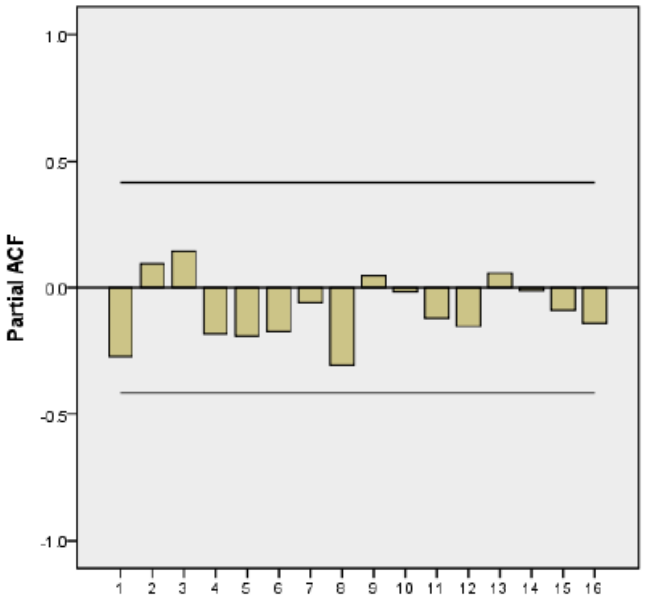

Fig. (4). Partial correlation of series.

Compare these four models with each other by SPSS , and the results as listed in Table **5**. As can be seen for this table, the smooth R-square represents the estimation value of the total variation explained by the model, and the larger the value, the better the fitting degree (the maximum value equals to 1). MAPE represents the average absolute percentage error. The smaller the value, the better it is. As one of the most commonly used evaluation criteria, the smaller the value of the standardized BIC, the better it is. Obviously, the final model will be different with a different evaluation. If you want to choose a smooth R-square as the largest model, it should be ARIMA (2,2,2). By contrast, if you want to choose the minimum value of MAPE, it should be ARIMA (1, 2,1), and if you want to choose the smallest standardized BIC, it should be ARIMA (1,2,1). Considering the model fitting degree of the history data, that is using the MAPE index [12] as the standard, based on the fact that MAPE index and standardized BIC index identify the same model, we choose the ARIMA (1,2,1) model.

3.3. Parameter Estimation

The parameters of ARIMA (1,2,1) model can be evaluated by SPSS software, which is shown in Table **6**. The model can be stated as follows:

$$Y_t = -0.94Y_{t-1} + 0.99e_{t-1} + e_t \quad (2)$$

3.4. Model Validation

It needs to test the residual series to determine whether the model has reached the optimum. If the residual series is a white noise series, which indicates that all residuals are random and that it cannot be used to make improvements to the model, then the model has reached its optimum and can be used to forecast. On the contrary, the model needs to be re-identified. Fig. (**5**) shows the autocorrelation of residual series in the ARIMA (1,2,1) model. As is shown in this figure, all correlation coefficients fall within the range randomly, and the residual series is a white noise series.

Table 5. Comparison of four models.

Model	Smooth R-Square	Mape	Standardized BIC
ARIMA(1,2,1)	0.488	1.219	32.917
ARIMA(1,2,2)	0.506	1.253	33.066
ARIMA(2,2,1)	0.488	1.225	33.11
ARIMA(2,2,2)	0.525	1.238	33.23

Table 6. ARIMA(1,2,1) model Parameter Estimation.

	Estimate	SE	t	Sig.
Nature Log Constant	0.000	0.001	0.368	0.717
AR Lag 1	-0.94	0.255	-0.367	0.717
Difference	2			
MA Lag 1	0.99	2.613	0.379	0.709

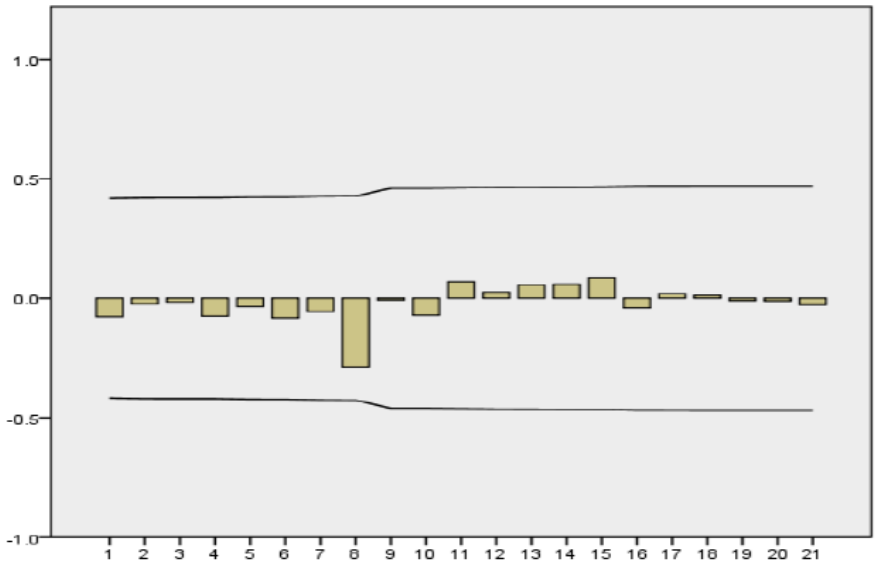

Fig. (5). Autocorrelation of residual series in ARIMA (1,2,1) model.

Table **7** lists the income forecasting of the data from 2013.1 to 2013.5 with the ARIMA (1,2,1) model. It can be seen from Fig. (**7**) that the maximum error is about 1.8% in March 2013, and the smallest error is about 0.4% in May 2013. So the average error is about 1% of the predicted five months, which means that the model can accurately forecast the telecom income data of a province.

3.5. Income Forecasting

After forecasting by the ARIMA (1,2,1) model, the income of June 2013 is 689,244,328 Yuan.

Currently all provinces have established a variety of BI analysis system to support the daily management and operation decision of the telecom. Statistical analysis techniques

Table 7. Income Forecasting in 2013.1-2013.5.

	2013.1	2013.2	2013.3	2013.4	2013.5
Actual value	624161218	624672805	661698665	654625562	680178948
Forecast value	618127030	619122322	649820067	656261014	677478953
Error	0.97%	0.89%	1.8%	1%	0.4%

and data mining technology can help top decision makers and business executives understand the business situation, monitor key production and operation targets and operational risks, identify potential law, and forecast future trends. Many provinces regard income forecasting as one of the key topics. ARIMA model used in this study does not involve huge data and complex system. Only based on simple data, methods, and tools, we can get a more accurate income forecasting value to help top managers in companies grasp income trends. The method proposed in this paper can be used as a reference method for other analysts.

CONCLUSION

The past estimation on telecom income mostly depends on empirical analysis or qualitative forecast. It is difficult to meet the precise needs although they can calculate out the income trend. In this paper, we construct time series for the telecom income from the historical data, use SPSS software to calculate the time series model, get optimal parameters, and fit income data in five month. The results show that the model simulates income data well and can be used to predict income in the future.

This study builds the model only based on data in 24 months, so the seasonal analysis of the data is not involved. The fitting degree of data in March is not very good due to seasonal factors. A more comprehensive ARIMA model can be built in the future based on more historical data including the trend, seasonal, and cyclical factors.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (No.61402350 and No. U1404622) and Fundamental Research Funds for the Central Universities (BDZ021430).

REFERENCES

[1] Y. Chun, "Research on operating income forecasting of telecommunication operators," *Technology and Standardization of Telecommunication Engineering*, vol. 11, pp. 14-17, 2008.

[2] Dongmeizhu, Y. Huang, J. Hu, "The Application of ARIMS time series model in income budget of telecom products," *Control Management*, vol. 22, no. 6, pp. 178-179, 2006.

[3] S. Wang, and Z. Chen, "Estimation of the order of ARIMA model by linear procedures," *Chinese Annuals Of Math*, vol. 6, pp. 53-70, 1985.

[4] P. Huang, and A. Xu, "Analysis of time-series forecast for development of marine economy in jiangsu province based on ARIMA model," *Jangsu Agricultural Science*, vol. 5, pp. 505-507, 2010.

[5] K. Xue, Z. Li, L. Liu, and S. Cheng-qian, "Network traffic prediction based on ARIMA model, " *Microelectronics & Computer*, vol. 21, no. 7, pp. 84-87, 2004.

[6] D. Huang, "Recursive method for ARIMA model estimation," *Acta Mathematicae Applicatae Sinica*, vol. 5, pp. 333-354, 1989.

[7] J. Cheng, "Estimation of flight delay using weighted spline combined with ARIMA model," In: *Proceedings of the 7th IEEE/International Conference on Advanced Infocomm Technology*(IEEE/ICAIT2014) , 2014.

[8] X. Hu, and M. Yu, "Based on ODE and ARIMA modeling for the population of China," In: *Proceedings of Quantitative Economics Conference(QEC 2013)*, 2013.

[9] A.Khosravi, S. Nahavandi, and D. Creighton, "Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 602-610, 2013.

[10] P. Xiong, "*Data Mining Algorithm and Clementine Practice*," Tsinghua University Press, Beijing, pp. 215-219, 2011.

[11] Y. Wang, and J. Rao, "Application and practice on the property insurance premium income of China forecast based on ARIMA model," vol. 6, pp. 39-44, 2010.

[12] W. Lu, "*SPSS Statistical Analysis*," Publishing House of Electronics Industry, Beijing, pp. 564-580, 2010.

Received: June 10, 2015 Revised: July 29, 2015 Accepted: August 15, 2015