J. Korea Soc. Math. Educ. Ser. B: Pure Appl. Math.
Volume 11, Number 4 (November 2004), Pages 275–285

ISSN 1226-0657

COMPARATIVE ANALYSIS ON TIME SERIES MODELS FOR THE NUMBER OF REPORTED DEATH CLAIMS IN KOREAN COMPULSORY AUTOMOBILE INSURANCE

KANG SUP LEE AND YOUNG JA KIM

ABSTRACT. In this paper, the time series models for the number of reported death claims of compulsory automobile liability insurance in Korea are studied. We found that $\text{IMA}(0,1,1) \times (0,1,1)_{12}$ would the most appropriate model for the number of reported claims by the Box-Jenkins method.

1. INTRODUCTION

One of the most important roles in risk management of insurance industries is to forecast the number of claims and to prepare for alternation. An insurer must estimate the frequency and size of accidents to set up optimal premium about certain risks. Generally, Poisson-inverse Gaussian and negative binomial distribution were used as distribution for the number of claims, and Pareto, gamma, log-normal and Burr distribution were used as distribution for the size of claims (see Zi [7]). Especially, Lemaire & Zi [3] suggested Poisson distribution as a distribution for the number of claim using Taiwan data in the study of comparison and analysis for 30 Bonus-Malus System of 22 countries including Korea. Zi [7] also considered the negative binomial distribution for the frequency of claims in the automobile insurance using Korean data.

Although some discrete distribution models have nice properties, they aren't suitable for forecasting the number of accidents depending on the time. In this case, some time series models are better to forecast more appropriate loss reserves. As the model for loss reserves when the accident occurred, Lemaire [2] suggested AR model, and Verrall [5] considered similar model. Louis [4] found AR is an adequate

Received by the editors September 14, 2004 and, in revised form, November 15, 2004.
2000 *Mathematics Subject Classification.* Primary; 62-07, Secondary; 62P05.
Key words and phrases. Automobile insurance, claims, Box-Jenkins method.

model for the number of claims on Canadian data. However, there has been no time series model for Korean data.

Lee & Kim [1] suggested ARIMA$(1,1,1) \times (0,1,0)_{12}$ as the model for reported death claims in Korean automobile insurance. Lee & Kim [1] used the data from April 1996 to March 2002, in their study, they suggested two models as follows;

(1) IMA$(0,1,1) \times (0,1,1)_{12}$ model for the data from April 1996 to March 2003.
(2) AR(1) model for the data from April 1998 to March 2003.

2. Model For the Number of Reported Claims in Korea

Let us consider the ARIMA$(p,d,q) \times (P,D,Q)_s$ model as follows;

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_i = \Theta_Q(B^s)\theta_q(B)a_i,$$

where a_i is a white noise, and

$$\Phi_P(B^s) = 1 - \Phi_s B^s - \Phi_{2s}B^{2s} - \cdots - \Phi_{Ps}B^{Ps},$$
$$\Theta_Q(B^s) = 1 - \Theta_s B^s - \Theta_{2s}B^{2s} - \cdots - \Theta_{Qs}B^{Qs},$$

In this study, we checked the various values of p, d, q or P, D, Q to find a suitable model for the Korean data. The data, as shown in the following Table 1, is the number of reported death claims of compulsory automobile liability insurance in the observation period April 1996 to Jan 2004.

TABLE 1. The numbers of monthly claims

Month	96–97*	97–98	98–99	99–00	00–01	01–02	02–03	03–04
Apr.	946	751	629	620	632	516	468	472
May	952	810	690	619	676	574	460	475
Jun.	828	780	584	584	585	500	384	525
Jul.	879	705	648	638	638	455	408	458
Aug.	984	721	477	725	626	540	427	449
Sep.	876	700	706	651	625	547	506	454
Oct.	1,033	910	757	737	753	582	558	546
Nov.	1,041	879	737	779	703	611	535	531
Dec.	945	757	686	721	668	582	744	567
Jan.	887	632	550	684	541	533	477	431
Feb.	749	664	572	640	571	430	413	
Mar.	867	633	561	638	534	559	497	

* Accident year Apr. ~ Mar.
** From Korea Insurance Development Institute

1) The time series model of data for the past 6 years (1996. 4~2002. 3)

Let us consider the data from April 1996 to March 2002. The result is as following. The plot of the data (Figure 1. in the Appendix) and the figure of the SACF and SPACF (Figure 2. in the Appendix) suggest first difference, and also seasonal difference after first difference. Hence we did first difference and seasonal difference at seasonal period 12. As a result, we got the ARIMA$(1,1,1) \times (0,1,0)_{12}$ model (see Lee & Kim [1]).

The resulting model is

$$(1 + 0.35524B)W12_t = (1 - 0.55263B)a_t$$

where $a_t \sim N(0, 5438.217)$, and

	Estimate	Standard error	t value	p value
$\hat{\theta}_1$	0.55263	0.14876	3.71	0.0002
$\hat{\phi}_1$	-0.35524	0.16491	-2.15	0.0312

2) The time series model of data for the past 7 years (1996. 4~2003. 3)

Based on the data from 1996 to 2003, we found that IMA$(0,1,1) \times (0,1,1)_{12}$ would be appropriate model. The plot of data (Figure 3. in the Appendix) shows decreasing trend in mean and constant variance among the observations. Its SACF and SPACF are in Figure 4. in the Appendix. For excluding the trend, we did first difference and there found seasonality in data. Therefore we did seasonal difference at the seasonal period 12.

Let $\hat{\theta}_1$ and $\hat{\Theta}_1$ be the MLE of θ_1 and Θ_1 then we have the following values.

The resulting model is

$$(1 - B)(1 - B^{12})W12_t = (1 - 0.68001B)(1 - 0.75225B^{12})a_t$$

where $a_t \sim N(0, 4007.188)$, and

	Estimate	Standard error	t value	p value
$\hat{\theta}_1$	0.68001	0.08682	7.83	$< .0001$
$\hat{\Theta}_1$	0.75225	0.23521	3.20	0.0014

3) The time series model of data for the past 5 years (1998. 4 ~ 2003. 3)

The Korean insurance industries just control and operate the data during the last five years only. Therefore we analyzed the data from April 1998 to March 2003.

Let Z_t $(t = 1, 2, \ldots, 60)$ be the numbers of reported claims with $t = 1$ means April 1998. The plot of Z_t (Figure 5. in the Appendix) against time shows neither trend in the mean nor constant variance, indicating that the stationarity assumption is adequate for the data. We observe that the SPACF (Figure 6. in the Appendix) goes to zero after lag 1, so it would be suggest that AR(1) process is an appropriate model.

Table 2 shows many other suitable models.

To select the appropriate model, let us find suitable values of p, q. Next, put into $p = 0, 1$, $q = 0, 1$ about orders p of AR and orders q of MA each from the types of the SACF and SPACF.

TABLE 2. Candidates for selected model

No.	Model	AIC / SBC*	Result
1	AR(1): constant	693.6908 / 697.8795	Estimated parameters and residuals are significant in $\alpha = 0.01$
2	AR(1): nonconstant	707.3641 / 709.4585	Estimated parameters are significant but residuals are not white noise
3	$ARMA(1,1)$: constant	694.1669 / 700.4500	Estimated parameters are not significant but residuals are white noise
4	$ARMA(1,1)$: nonconstant	700.9573 / 705.1460	Estimated parameters are significant but residuals are not white noise
5	IMA$(0,1,1)$: constant	685.2517 / 689.4067	Estimated parameters and residuals are not significant
6	IMA$(0,1,1)$: nonconstant	683.4930 / 685.5705	Estimated parameters are significant but residuals are not white noise

*AIC (Akaike Information Criterion), SBC (Schwartz Bayesian Criterion)

We performed significant test about the parameters and the residuals of selected models. The IMA$(0,1,1)$ model had the lowest values of both AIC and SBC, but the residuals of IMA$(0,1,1)$ model wasn't a white noise process. Hence, we selected AR(1) model as the best model.

For $\hat{\mu_1}$ and $\hat{\phi_1}$ be the MLE of μ_1 and ϕ_1 in the AR(1) model, we have the following values.

Parameter	Estimate	Standard error	t value	p value
$\hat{\mu}$	589.99763	25.82451	22.85	$< .0001$
$\hat{\phi}_1$	0.62783	0.10206	6.15	$< .0001$

Therefore the estimated model is as following.

$$(Z_t - 589.9976) = 0.6278(Z_{t-1} - 589.9976) + a_t \text{ where } a_t \sim N(0, 5898.883).$$

Also, we did portmanteau test (see Wei [6]) about residuals of selected model AR(1).

If Q^* is larger than $\chi^2(K-p)$, the residuals of selected model get white noise process. One side, if Q^* is less than $\chi^2(K-p)$, the residuals of selected model are not following to white noise process.

Table 3 is the result of portmanteau test in significant level $\alpha = 0.01$. In the result, the residuals of almost lag are followed to white noise process. The plot of residuals is the Figure 7. in the Appendix.

TABLE 3. Portmanteau test of AR(1) (in $\alpha = 0.01$)

Lag	6	12	18	24	30	36	42	48	54
Q^*	2.94	10.63	14.90	27.51	38.19	44.84	50.34	65.79	96.34
Freedom	5	11	17	23	29	35	41	47	53
Prob	0.7085	0.4745	0.6028	0.2347	0.1182	0.1232	0.1504	0.0364	0.0003

4) The comparison of forecasted models

Detail comparison on the models are presented in table 4 and Table 5. We showed actual values and forecast values of death claims of two models in Table 4.

Table 5 showed MAD, MSE, MAPE of two models respectively. In result of Table 5, we made clear that $\text{IMA}(0,1,1) \times (0,1,1)_{12}$ is better than AR(1) as model of the number of monthly death claims in Korean automobile insurance.

3. REMARKS

Korean models are significantly different in terms of the periods referred to construct the models. This dilemma is caused by unstableness of Korean data. In fact, the number of automobile accidents has ups and downs periods 1996 to 2001 in Korea by the various social and economical situations. And Korean data is just accumulated and controlled during the last five years, hence we have trouble fitting the model. As a result $\text{IMA}(0,1,1) \times (0,1,1)_{12}$ is suitable to Korean data.

IMA$(0,1,1) \times (0,1,1)_{12}$ is made of data for the past 7 years. Hence, Korean data have to be accumulated more than data during the last five years for the better model. We hope that the models in this paper will be a great help to calculate IBNR in Korean automobile insurance.

TABLE 4. The forecasted numbers of deaths claims

	IMA$(0,1,1)$ $\times(0,1,1)_{12}$	Std error	AR(1)	Std error	Actual value
Apr., 2003	448.2915	63.3024	531.6111	76.8042	472
May	476.0799	66.4643	553.3409	90.6865	475
Jun.	400.3059	69.4824	566.9835	95.6062	525
Jul.	416.3502	72.3748	575.5487	97.4772	458
Aug.	439.5384	75.1560	580.9262	98.2049	449
Sep.	464.6647	77.8379	584.3023	98.4902	454
Oct.	551.1552	80.4304	586.4219	98.6025	546
Nov.	544.3356	82.9419	587.7527	98.6467	531
Dec.	556.7959	85.3795	588.5882	98.6641	567
Jan., 2004	424.8311	87.7495	589.1127	98.6710	431
Feb.	381.0256	90.0571	589.4421	98.6737	
Mar.	428.0504	92.3071	589.6488	98.6747	

TABLE 5. The comparison of forecasted models

	IMA$(0,1,1) \times (0,1,1)_{12}$	AR(1)
MAE*	24.6122	83.6588
MSE	1839.6435	8997.4778
MAPE	4.9238	17.9452

*MAE(Mean Absolute Error)
MSE(Mean Square Error)
MAPE(Mean Absolute Percentage Error)

REFERENCES

1. K. S. Lee & Y. J. Kim: *Time Series Model for the Number of Reported Claims in Korean Automobile Insurance.* ENBIS3/ISIS3 conference 2003.
2. J. Lemaire: Claims Provisions in Liability Insurance. *Journal of Forecasting* **1** (1982), 303–318.
3. J. Lemaire & H. Zi: A Comparative Analysis of 30 Bonus-Malus Systems. *ASTIN Bulletin* **24** (1994), 287–309.
4. G. D. Louis: Constrained Forecasting of the Number of IBNR Claims. *Journal of Actuarial Practice* **4** (1996), no. 3, 287–305.

5. R. J. Verrall: Modelling claims runoff triangles with two-dimensional time series. *Scand. Actuar. J.* (1989), no. 3, 129–138. CMP1067098
6. W. S. Wei: *Time series analysis.* Univariate and multivariate methods. Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1990. MR **91c**:62084
7. H. Zi: An Analysis of Efficient of Bonus-Malus System in Korean Automobile Insurance. *The Journal of risk Management* **VI** (2000), no. 1, 3–38.

APPENDIX

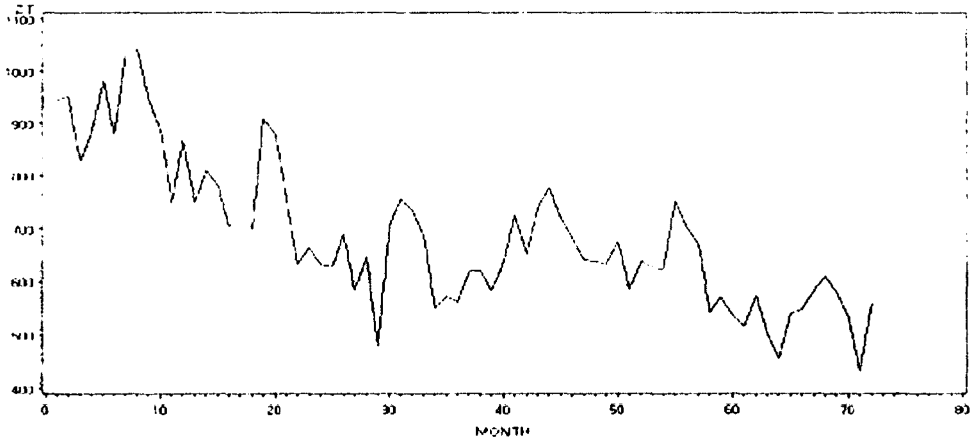

FIGURE 1. The Plot of Zt

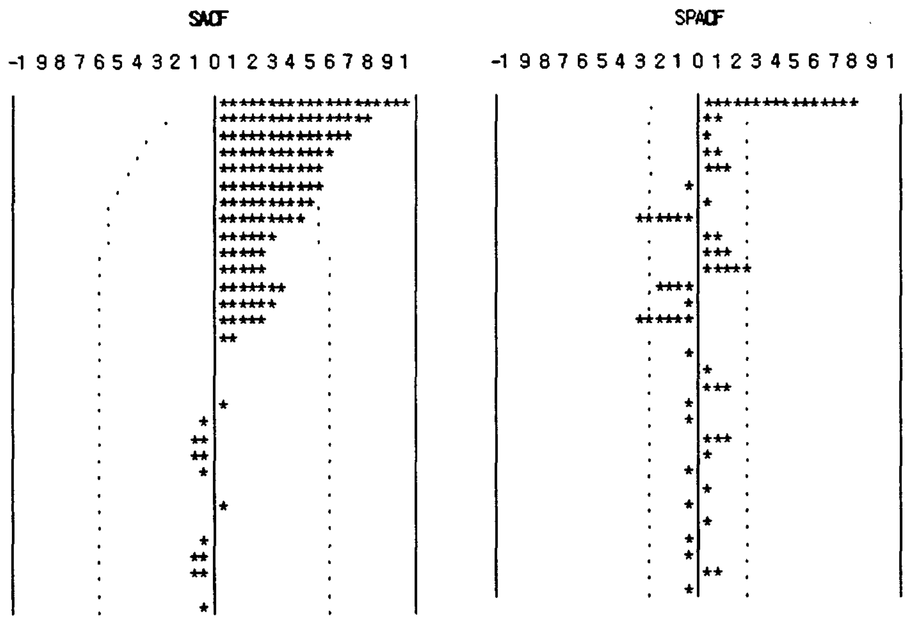

FIGURE 2. The SACF and SPACF of Zt

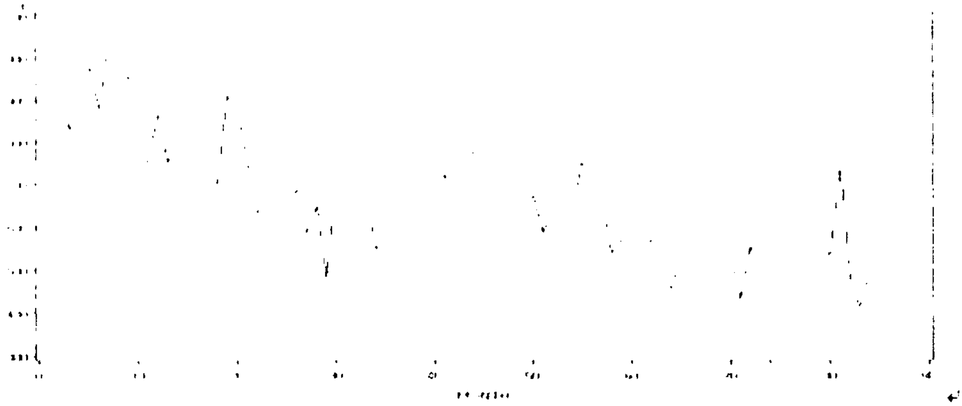

FIGURE 3. The Plot of Zt

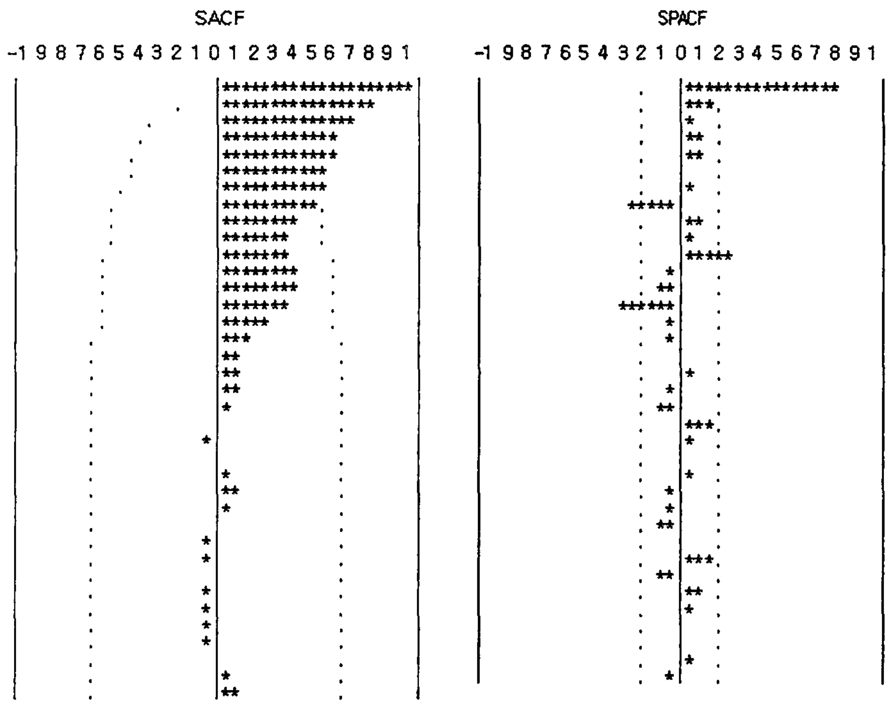

FIGURE 4. The SACF and SPACF of Zt

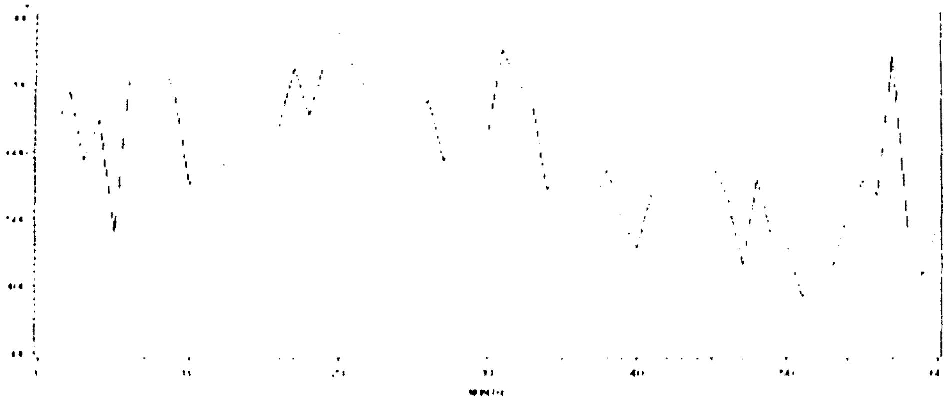

FIGURE 5. The Plot of Zt

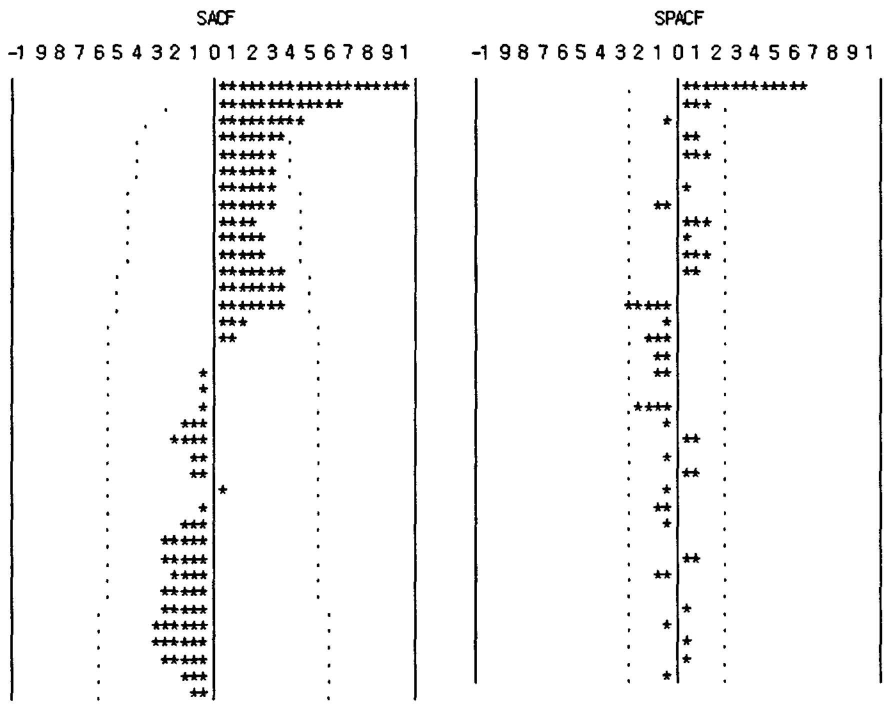

FIGURE 6. The SACF and SPACF of Zt

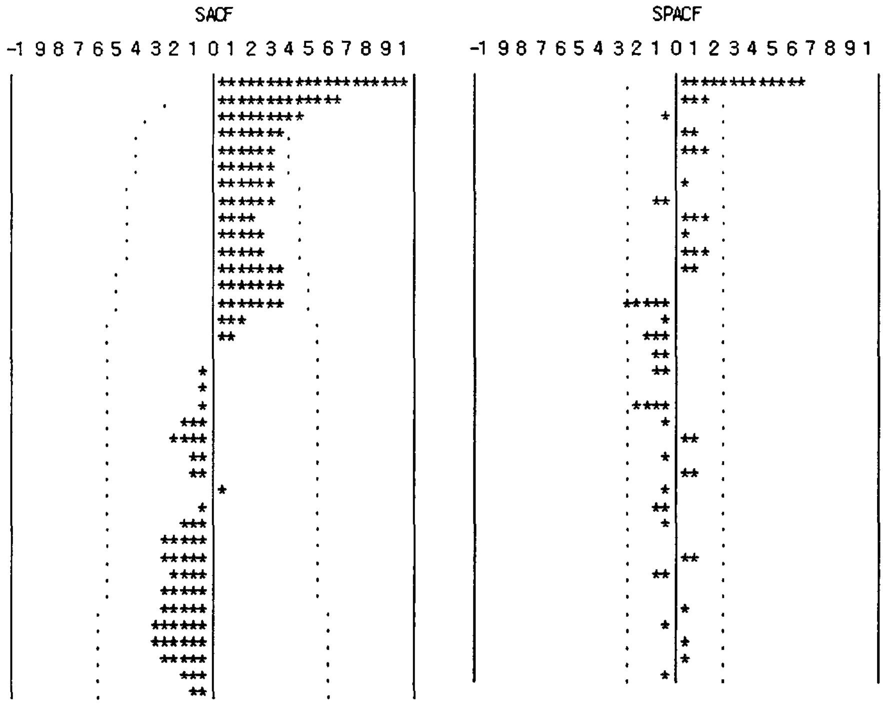

FIGURE 7. The Plot of Zt

(K. S. LEE) DEPARTMENT OF MATHEMATICS EDUCATION, DANKOOK UNIVERSITY, SAN-8, HANMAM-DONG, YOUNGSAN-GU, SEOUL 140-714, KOREA
Email address: leeks@dankook.ac.kr

(Y. J. KIM) DEPARTMENT OF MATHEMATICS EDUCATION, DANKOOK UNIVERSITY, SAN-8, HANMAM-DONG, YOUNGSAN-GU, SEOUL 140-714, KOREA
Email address: yjakim@dankook.ac.kr