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 $IMA(0,1,1) \times (0,1,1)_{12}$ 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.
Key words and phrases. Automobile insurance, claims, Box-Jenkins method.

Kang Sup Lee and Young Ja Kim

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;

\[
\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D Z_t = \Theta_Q(B^s)\theta_q(B)a_t,
\]

where \(a_t\) is a white noise, and

\[
\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps},
\]

\[
\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs},
\]

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

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

* 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
\[
(1 + 0.35524B)W_{12t} = (1 − 0.55263B)a_t
\]
where \(a_t \sim N(0, 5438.217)\), and

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}_1)</td>
<td>0.55263</td>
<td>0.14876</td>
<td>3.71</td>
</tr>
<tr>
<td>(\hat{\phi}_1)</td>
<td>-0.35524</td>
<td>0.16491</td>
<td>-2.15</td>
</tr>
</tbody>
</table>

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 \(\theta_1\) and \(\Theta_1\) then we have the following values.

The resulting model is
\[
(1 − B)(1 − B^{12})W_{12t} = (1 − 0.68001B)(1 − 0.75225B^{12})a_t
\]
where \(a_t \sim N(0, 4007.188)\), and

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta_1)</td>
<td>0.68001</td>
<td>0.08682</td>
<td>7.83</td>
</tr>
<tr>
<td>(\Theta_1)</td>
<td>0.75225</td>
<td>0.23521</td>
<td>3.20</td>
</tr>
</tbody>
</table>

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

*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 $\mu_1$ and $\phi_1$ in the AR(1) model, we have the following values.
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 \(\alpha = 0.01\) )

<table>
<thead>
<tr>
<th>Lag</th>
<th>6</th>
<th>12</th>
<th>18</th>
<th>24</th>
<th>30</th>
<th>36</th>
<th>42</th>
<th>48</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q^*)</td>
<td>2.94</td>
<td>10.63</td>
<td>14.90</td>
<td>27.51</td>
<td>38.19</td>
<td>44.84</td>
<td>50.34</td>
<td>65.79</td>
<td>96.34</td>
</tr>
<tr>
<td>Freedom</td>
<td>5</td>
<td>11</td>
<td>17</td>
<td>23</td>
<td>29</td>
<td>35</td>
<td>41</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>Prob</td>
<td>0.7085</td>
<td>0.4745</td>
<td>0.6028</td>
<td>0.2347</td>
<td>0.1182</td>
<td>0.1232</td>
<td>0.1504</td>
<td>0.0364</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

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 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 IMA(0,1,1) \(\times\) (0,1,1)_{12} is suitable to Korean data.
IMA(0,1,1) × (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

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

**Table 5.** The comparison of forecasted models

<table>
<thead>
<tr>
<th></th>
<th>IMA(0,1,1) × (0,1,1)_{12}</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE*</td>
<td>24.6122</td>
<td>83.6588</td>
</tr>
<tr>
<td>MSE</td>
<td>1839.6435</td>
<td>8997.4778</td>
</tr>
<tr>
<td>MAPE</td>
<td>4.9238</td>
<td>17.9452</td>
</tr>
</tbody>
</table>

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

**REFERENCES**


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 $Z_t$

(K. S. Lee) Department of Mathematics Education, Dankook University, San-8, Han-mam-dong, Youngsan-gu, Seoul 140-714, Korea

Email address: leeks@dankook.ac.kr

(Y. J. Kim) Department of Mathematics Education, Dankook University, San-8, Han-mam-dong, Youngsan-gu, Seoul 140-714, Korea

Email address: yjakim@dankook.ac.kr