A multivariate state space model for IBNR reserve prediction

Daiane R. dos Santos * Cristiano Fernandes † Joel C. da Rosa ‡

Abstract

In this article we propose a multivariate extension of Atherino's model (2010), in which insurance claims are organized as "time series" by stacking the columns of the runoff triangle in which the IBNR (incurred but not reported) claims are grouped. Such "time series" will display periodic movements which can be duly captured by a "seasonal" component. In order to do so the structural time series model of Harvey (1989) is used. Our multivariate extension uses a SUTSE (seemingly unrelated time series equations) structure, in which each IBNR series has its own "seasonal" but the shocks are correlated. This approach provides a more parsimonious description of correlated IBNR reserves. We apply this model to a bivariate claim series of the Brazilian car insurance market. Our results show that the proposed model presents better results than its univariate counterpart.

Key Words: IBNR, SUTSE, runoff triangle, state space models.

1. Introduction

IBNR is a type of provision that a insurance company has to make for claims that have already happened but haven't yet been reported, given that, in practice, there is a delay between the time that a claim occurred and when the claim is reported. The accurate prediciton of these values is extremely important for the finance organization of a company.

Since the work published in Bornhuetter Fergunson (1972), a group of strategies for predicting IBNR has been proposed based on data organized in a special type of triangle known as the *runoff triangle*, as shown in Table 1. In this triangle the rows represent *accident years* or *years of origin* and the columns are the *development years*. Here we will consider the incremental form of this triangle and its cells are denoted by C_{wd}, $1 \leq w \leq J$ and $0 \leq d \leq J-1$.

Among the methods that take advantage of the runoff triangle framework, the chain-ladder (Mack, 1993) has been the most frequently used in the insurance industry. The successful application of this method is consequence of its simplicity and the fact that it is distribution-free, enhancing its flexibility. Because of its importance, we will use it as benchmark for comparisons.

Differently from the chain-ladder, here we follow the approach adopted by Atherino et al. (2010), in which the original double index of the runoff triangle is replaced by a single index which runs from top to bottom of the triangle, across its rows, as shown in Table 2. This we will refer as the row wise ordering of the run off triangle. As a result of this operation, the entries of the triangle are transformed into a time series with missing values. It is well known that the state space approach for time series deals very naturally with missing values, replacing the non observable values by estimates using a smoothing algorithm constructed via the Kalman filter (Durbin & Koopman, 2001). It is therefore a natural framework to estimate IBNR reserves when considering the row wise form of the runoff triangle. Atherino et al.(2010) apply a univariate state space model to estimate the IBNR reserve, which in this setup is obtained by the sum of estimated missing values (Equation 1).

*PUC-Rio, Marquês de São Vicente, Rio de Janeiro, RJ.
†PUC-Rio, Marquês de São Vicente, Rio de Janeiro, RJ.
‡Rockefeller University, 1230, York Avenue, New York, NY.

In this article we present a multivariate extension of Atherinos's model, namely, a state space model to jointly estimate the IBNR reserves associated with several run off triangles transformed into row wise time series of IBNR claims. Such framework may be useful when estimating IBNR reserves of a company which operates in several sectors which are expected to share common risk factors.

$$\text{IBNR} = \sum_{\text{t: } y_t \text{ is missing}} E\left[y_t | y_n\right] \tag{1}$$

When we model IBNR provisions series individually we discard possible association between IBNR provisions. In fact, when analyzing real row wise IBNR series obtained from different runoff triangles, we observe significant correlation coefficients between these series, indicating the plausibility of a multivariate model which will be dealt with in the next section. A different multivariate state space model for IBNR reserve has been recently proposed by De Jong (2011).

2. A multivariate state space model for IBNR estimation

2.1 SUTSE formulation with trends and seasonal

The row wise ordering of the run off triangle will result in a time series displaying at the start of any accident year, large claims, which will then be followed by decreasing values with the advance of the development years. This pattern will repeat itself across the resulting time series and can be duly captured by a model with a level and a periodic component. Our multivariate state space model will be specified using the SUTSE (seemingly unrelated time series equations) structure, as developed by Harvey (1989, pp. 463-464) and later used by Fernandez and Harvey (1990) and Jalles (2009), among others.The SUTSE structure is one in which each time series $y_{j,t}$, j=1,2 has its own level $\mu_{j,t}$, j=1,2 and periodic component $\gamma_{j,t}$, j=1,2, with each of these components evolving stochastically, as long as the variance of their shocks,η_t's and ω_t's, respectively, are non zero. Formally the SUTSE model is given defined by the following set of equations:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} + \begin{bmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{1,t} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} \gamma_{1,t+1} \\ \gamma_{2,t+1} \end{bmatrix} = \begin{bmatrix} -\Sigma_{j=1}^{n-1}\gamma_{1,t-j} \\ -\Sigma_{j=1}^{n-1}\gamma_{2,t-j} \end{bmatrix} + \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\varepsilon_1^2} & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_2^2} \end{bmatrix}\right], \tag{5}$$

$$\begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\eta_1^2} & \sigma_{\eta_1\eta_2} \\ \sigma_{\eta_1\eta_2} & \sigma_{\eta_2^2} \end{bmatrix}\right], \tag{6}$$

$$\begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\omega_1^2} & \sigma_{\omega_1\omega_2} \\ \sigma_{\omega_1\omega_2} & \sigma_{\omega_2^2} \end{bmatrix} \right] \tag{7}$$

As it can be seen the shocks impacting both the level and periodic components, η_t's and ω_t's, respectively, are correlated.

2.2 State space form and the Kalman filtering

The bivariate SUTSE model previously presented can be cast into the state space form, and once this is done, estimation of the unobserved components, $\mu_{j,t}$ and $\gamma_{j,t}$, j=1,2, is accomplished by use of the Kalman filtering and related algorithms, such as smoothing and prediction error decomposition of the likelihood function. A *Gaussian linear state space form* consists of two equations. The first is the observations equation, which describes the evolution of a p-variate time series y_t, $t = 1, 2, \ldots$ in terms of the components that are encapsulated in the vector α_t. The second is the state equation specifying the way each of the components contained in α_t evolves stochastically. More specifically:

observation equation:

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t \qquad t = 1, 2, \ldots n \tag{8}$$

state equation:

$$\alpha_{t+1} = T_t \;\; \alpha_t + c_t + R_t\eta_t \tag{9}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathtt{NID} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right) \tag{10}$$

$$E[\varepsilon_t'\alpha_1] = E[\eta_t', \alpha_1] = 0, \forall t \tag{11}$$

$$\alpha_1 \sim N(a_1, p_1) \tag{12}$$

The system matrices, Z_t, d_t, c_t, T_t and R_t, are deterministic, and the errors are considered independent of each other and independent of the initial state. The fixed unknown elements contained in some of the system matrices are estimated together with the state vector. This is accomplished by use of the Kalman Filter, a set of recursive equations which gives estimated values for the mean and covariance matrices of the state vector at any time period. Full details on the Kalman filtering deduction and related algorithms can be found in Durbin & Koopman (2001).

3. Application

We carried out an application that fitted the described SUTSE bivariate model to time series containing IBNR claims from the Brazilian car insurance market. In this market, the policies are of three types, according to risk coverage: Casco (full insurance), RCFV (personal injury) and APP (private passenger auto).

The Casco insurance type covers general car damages, including thefts. The RCFV coverage, also known as third parties protection, is intended to cover injuries in individuals who are not associated with the insured vehicle. The APP coverage - Personal Accident Passenger - compensates damages arising from passengers personal accidents.

Our time series data was obtained after reshaping two runoff triangles that report payments from Casco and RCFV IBNR claims. This data was observed in the period between the first quarter of 2009 and the second quarter of 2013 in a total of 324 observations for each type of insurance coverage.

The Kalman Filter implementation to estimate the SUTSE model combined the exact likelihood and the EM algorithm and also uses the blocks method presented in Atherino et al. (2010). This approach enables the straightforward computation of the covariance between observations, including the missing observations, conditional to the whole set of observations Y . As a consequence, it results in an elegant and succinct expression for estimating the mean squared error for the IBNR provision.

Model estimation through maximum likelihood (via BFGS algorithm) resulted in numerical values for the fixed and unknown variances and covariances as given in equations (5), (6) and (7). Results are depicted in Table 3.

Given that all the variances estimates are negligible, one can conclude that both the level and periodic components driving each of the IBNR series are deterministic.

Figure 1 shows the smoothed series for each coverage.

As shown on the figure 1, the smoothed series fits to real data in both case.

Following the trend in the IBNR literature, in this article we will also compare our results to those obtained by fitting the chain-ladder method to our data set. For completeness we will also include the results of data fitting by Atherino's univariate model. For out of sample validation, we decided to leave out the the 17 observations of the last diagonal of the runoff triangle. Goodness of fit between the actual and predicted (through filtering) values was measured using MSE (mean squared error), MAPE (mean absolute percentage error) and R^2.

As it can be seen from the results on Tables 4 and 5, the best fit was produced by our multivariate SUTSE model. On Table 6 we present some diagnostics for our model (based on the modelss innovations) namely, Box-Ljung autocorrelation test, Jarque-Bera normality test, and ARCH effect through Box-Ljung applied to squared innovations. Only normality is rejected, but this is not a serious drawback for our predictions, since we can guarantee the best linear estimate for the expected values of the level and periodic components.

The Tables 7 and 8 shows the IBNR reserves estimated by the three methods: chain ladder, Atherino's univariate state space model and our proposed multivariate state space model (SUTSE).

As we can see, for both coverages, the total IBNR reserves estimated by the proposed model are slightly larger than those estimated by Atherino's model. However, there was a significant decrease in the estimated coefficient of variation of our model as compared to Atherino's, for the majority of the quarters.

4. Conclusion

In this article we developed a bivariate version of Atherino's model for IBNR estimation. In order to do so we use the SUTSE framework for multivariate state space models, which is applied to a IBNR bivariate time series obtained from two related runoff triangles. The results were satisfactory when compared to both the chain ladder and Atherino's univariate model. Our results show that the periodicity of the two IBNR series are deterministic and both have a very smooth level. We have left out 17 observations from each series for model validation. According to the goodness of fit measures used for these observations (MSE, MAPE and R2) our proposed model produced, overall, the best fit. The in sample residuals are uncorrelated, homoscedastic, but non normal. Our estimated INBR reserves were slightly larger than those estimated by Atherino's univariate model. However, most of the coefficients of variation obtained from our model were lower .

REFERENCES

ATHERINO, R. S., PIZZINGA, A. AND FERNANDES, C.; A Row-wise Stacking of The Run-off Triangle: State Space Alternatives for IBNR Reserve Prediction. ASTIN Bulletin, v. 40, p. 917-946. 2010.

BORNHUETTER, R. L. and FERGUSON, R. E.; The Actuary and IBNR. Proceedings of the Casualty Actuarial Society LIX. Casualty Actuarial Society. Arlington, Virginia. 1972.

DE JONG, P.; Modeling dependence between loss triangles. North American Actuarial Journal, 16(1), 74 - 86. 2011.

DURBIN, J. AND KOOPMAN, S., Time Series Analysis by Space State Methods, Oxford University Press. 2001.

FERNANDEZ, F. J. AND HARVEY A. C., Seemingly unrelated time series equations and a test for homogeneity,Journal of Business & Economic Statistics 8, 71-81. 1990.

HARVEY, A. C., Forecasting, Structural Time Series Models and the Kalman Filter. first edition - Cambridge Univ. Press, Cambridge. 1989.

JALLES, J. T., Structural Time Series Models and the Kalman Filter: a concise review,FEUNL Working Paper Series wp 541. 2009.

MACK, T., Distribution-free Calculation of The Standard Error of Chain Ladder Reserve Estimates, ASTIN Bulletin, 23(2): 213-225. 1993.

SCHIMIDT, K. D. and ZOCHER, M.,The Bornhuetter-Ferguson Principle. Casualty Actuarial Society. Volume 2/Issue 1. 2008.

Accident Year w	Development d				
	0	1	2	$\dots$	$J-1$
1	C_{10}	C_{11}	C_{12}	$\dots$	C_{1J-1}
2	C_{20}	C_{21}	$\dots$	C_{1J-2}	
3	C_{30}	$\vdots$			
$\vdots$	$\vdots$	$\vdots$			
$\vdots$	$\vdots$	$C_{J-1\,1}$			
J	C_{J0}				

Table 1: *Runoff* triangle.

Accident Year y	Development year				
	0	1	2	$\cdots$	n-1
1	Y_1	Y_2	Y_3	$\cdots$	Y_n
2	Y_{n+1}	Y_{n+2}	$\cdots$	Y_{2n-1}	Y_{2n}
3	Y_{2n+1}	Y_{2n+2}	$\cdots$	$\cdots$	Y_{3n}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	Y_{3n-1}	...
n	$Y_{(n-1)n+1}$	$Y_{(n-1)n+2}$	$\cdots$	Y_{n^2-1}	Y_{n^2}

Table 2: Row wise of *runoff* triangle.

Parameters	Values estimated
$\sigma^2_{\varepsilon_1}$	8.48E-02
$\sigma^2_{\varepsilon_2}$	6.60E-02
$\sigma_{\varepsilon_1\varepsilon_2}$	2.25E-02
$\sigma^2_{\eta_1}$	1.94E-04
$\sigma^2_{\eta_2}$	1.06E-04
$\sigma_{\eta_1\eta_2}$	1.49E-04
σ^2_{ω}	1.33E-19
$\sigma^2_{\omega_2}$	2.04E-19
$\sigma_{\omega_1\omega_2}$	-1.04E-19

Table 3: Estimated parameters - Proposed model.

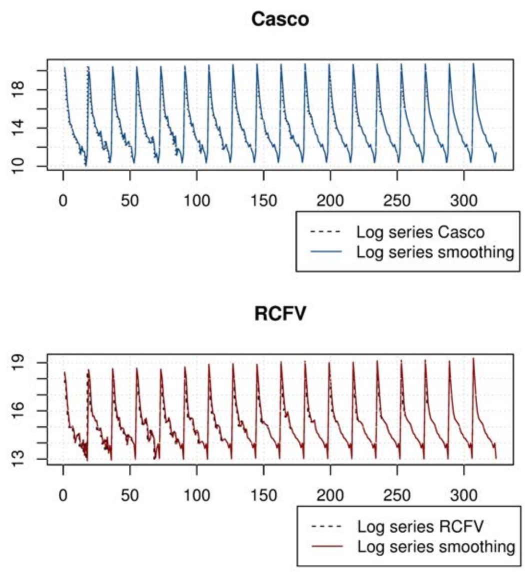

Figure 1: Logarithm of the series and the logarithm of the smoothed series.

	Chain-ladder	**Atherino uni.**	**Proposed model**
MSE($+10^{13}$)	20.41	5.30	1.59
Pseudo R^2 (%)	99.44	99.61	99.75
MAPE (%)	35.03	34.75	33.64

Table 4: **Casco** - Model comparison statistics (out of sample).

	Chain-ladder	**Atherino uni.**	**Proposed model**
MSE ($+10^{13}$)	11.54	1.37	1.08
Pseudo R^2 (%)	98.63	99.39	99.69
MAPE (%)	31.49	23.30	19.76

Table 5: **RCFV** - Model comparison statistics (out of sample).

Testing and diagnostics	Proposed model - p-values	
	Casco	RCFV
Heteroscedasticity	0.999	0.998
Box-Pierce	0.798	0.987
Jarque-Bera	0.003	0.000

Table 6: Testing and diagnostics applied to standardized innovations.

Quarter	Chain Ladder		Atherino Univ.		Proposed model	
	Provisions	CV (%)	Provisions	CV (%)	Provisions	CV (%)
2	80	36.0	88	43.3	90	43.3
3	111	27.3	122	33.4	128	33.5
4	203	21.0	212	24.1	218	24.2
5	341	20.7	334	19.6	340	19.6
6	570	17.4	557	17.6	562	17.6
7	795	26.1	735	15.7	752	15.6
8	1,141	22.9	988	14.5	1,029	14.5
9	1,490	20.0	1,262	13.6	1,321	13.6
10	2,093	17.4	1,828	13.9	1,927	13.8
11	2,930	17.4	2,492	13.5	2,615	13.4
12	3,851	14.2	3,344	13.4	3,458	13.3
13	5,260	11.4	4,761	13.8	4,844	13.7
14	7,983	13.8	7,219	14.6	7,323	14.4
15	12,342	11.4	11,231	15.3	11,355	15.1
16	19,892	188.1	18,682	16.4	18,951	16.3
17	46,563	133.0	45,107	20.4	45,877	20.4
18	248,937	32.9	222,766	26.2	230,240	26.3
Total	**354,580**		**321,726**		**331,028**	

Table 7: IBNR provisions in thousands of Brazilian Reais for Casco coverage.

Quarter	Chain Ladder		Atherino Univ.		Proposed model	
	Provisions	CV (%)	Provisions	CV (%)	Provisions	CV (%)
2	455	77.2	441	34.7	443	36.1
3	1,700	26.0	1,653	22.8	1,635	23.4
4	2,825	22.0	2,567	18.0	2,541	18.2
5	4,221	22.0	3,748	15.3	3,752	15.2
6	5,906	17.8	5,147	13.7	5,128	13.4
7	7,835	15.0	6,870	12.6	6,862	12.2
8	10,674	14.2	8,900	11.8	8,668	11.2
9	12,170	12.8	11,370	11.2	10,728	10.6
10	16,997	11.8	13,929	10.8	13,099	10.1
11	20,108	10.1	16,900	10.6	16,158	9.8
12	24,147	9.3	20,605	10.5	19,889	9.7
13	28,410	8.9	24,548	10.5	24,032	9.5
14	35,984	8.6	29,212	10.5	29,006	9.4
15	43,138	7.9	34,832	10.7	34,738	9.4
16	52,973	30.2	43,090	11.2	43,177	9.7
17	69,285	33.2	61,211	12.6	61,435	11.2
18	177,752	21.5	122,517	16.6	126,641	15.4
Total	**514,581**		**407,543**		**407,932**	

Table 8: IBNR provisions in thousands of Brazilian Reais for RCFV coverage.