Trend Analysis for Quarterly Insurance Time Series

Daniel Bortner, Cody Pulliam, Waiman Yam

Faculty Advisor: Michael Ludkovski

Department of Statistics and Applied Probability, University of California Santa Barbara

Abstract

Insurance companies commonly use **linear regression** to create predictive models by drawing a line of best fit through the data points. Here we are implementing the techniques of time series analysis to create a more accurate way to model quarterly data. Within the data, characteristics such as trend and seasonality can be utilized to improve upon basic **linear regression**. After comparing different models, the **ARIMA** model proves to be better at predicting the data than **linear regression**.

Objectives

➤ Explore whether time-series analysis is applicable to calculating future Pure Premium costs.
➤ Compare insurance firms' current **regression** forecasts to the time-series based methods.
➤ Identify which variables of the data provide more accurate predictions.
➤ Compare the precision of different predictive models.
➤ Establish an effective methodology for forecasting.

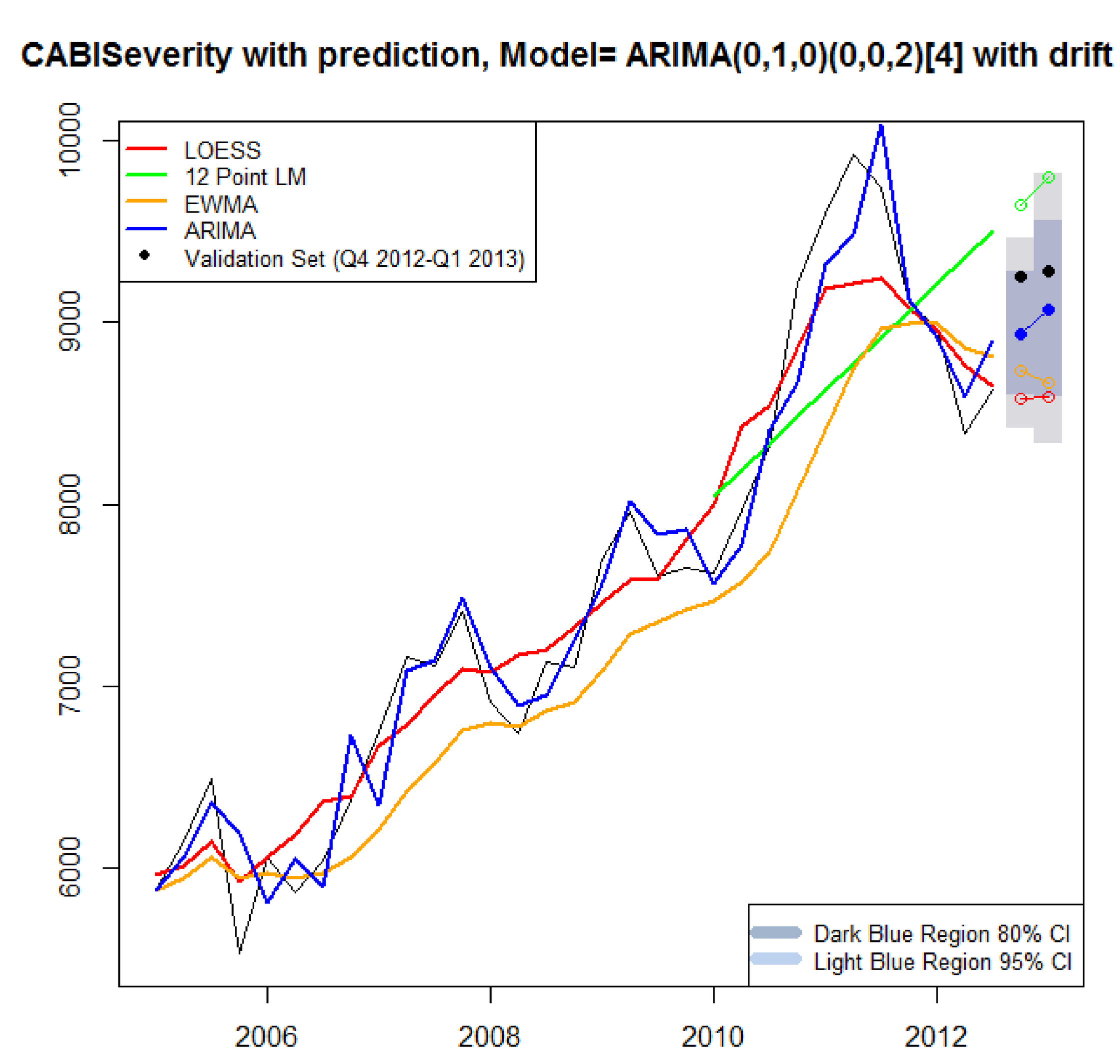
Data

➤ 18 States of Home Owners Policies (5 types of forms) and Auto Insurance Coverages (15 coverages) by quarter; Up to eight years ($2005Q1$ to $2013Q1$) worth of data and a maximum of 33 quarters per coverage/form.
➤ The data provided had already been applied a 4-quarter moving average.
➤ A 'series', denoted as a state and coverage (California –Bodily Injury), consists of their Frequency, Severity, Pure Premium, and Earned Exposure variables by quarter.
✦ Frequency = $\frac{\text{number of claims}}{\text{number of exposures}}$, the rate at which claims occur.
✦ Severity = $\frac{\text{total losses}}{\text{number of claims}}$, average cost of claims.
✦ Pure Premium = $\frac{\text{total losses}}{\text{number of exposures}}$ = Severity × Frequency, the average loss per exposure, also known as the insurers expected risk per policy holder.
✦ Earned Exposure = Number of bookings for the quarter.

New Predictive Models

➤ **Loess Smoothing** - Non-parametric regression methods that fits simple regression models to localized subsets of the data to build a function, found in R-Package(stats), using **stl**.
➤ **EWMA** - Exponential Weighted Moving Averages Smoothing, found in R-Package(TTR), using **EMA**.
➤**ARIMA** - Compilation of Auto-regressive and moving average smoothers, found in R-Package(forecast), using **Auto.Arima**

Comparing Models



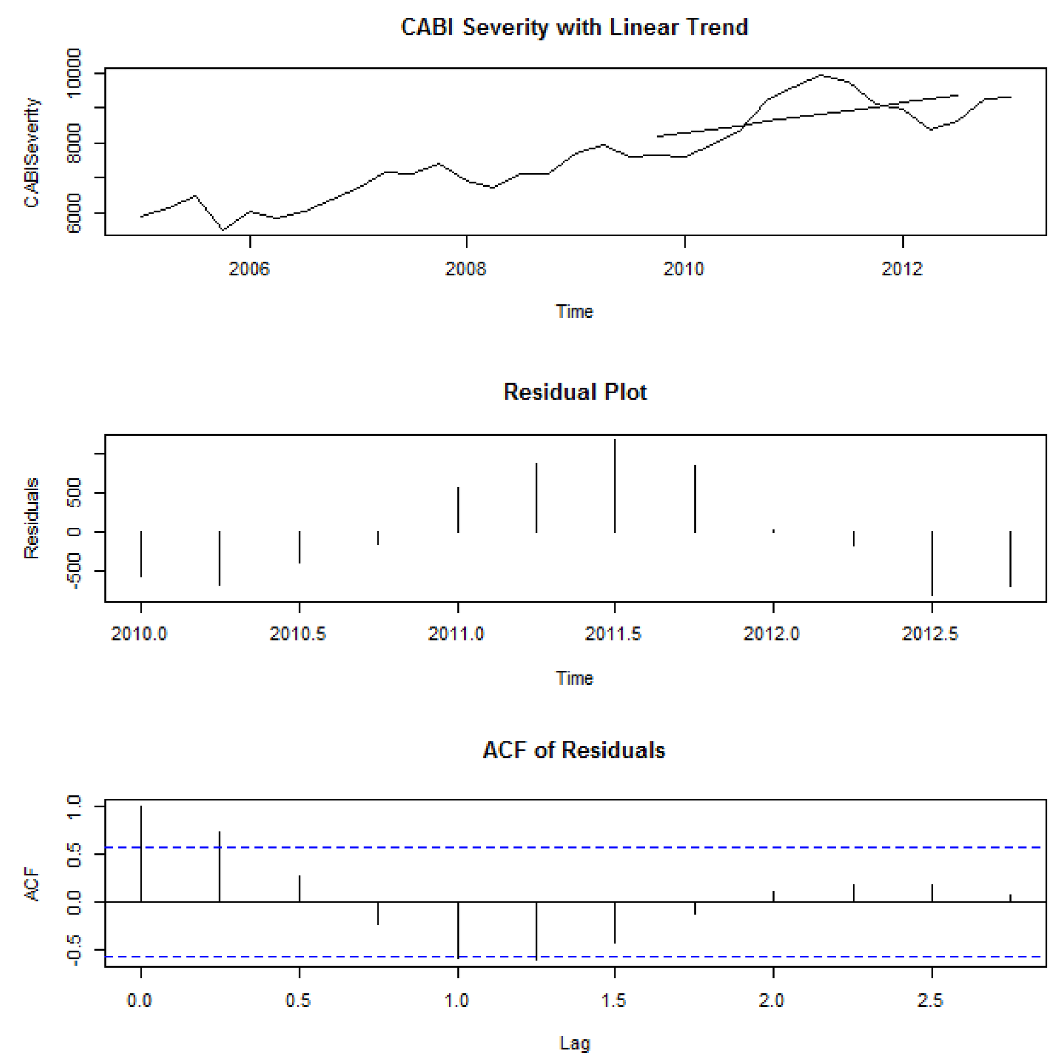
➤ The term "12 Point" refers to number of most recent quarters used.
➤ Forecasts for each model are compared to the last two quarters of the original data (the Validation Set in black).

Linear Regression

$$\hat{Y}_i = b_0 + b_1 x_i + \epsilon_i \quad (1)$$

➤ Minimizes the amount of error between a best fit line and the actual data. Assumes residual component (ϵ_i) demonstrates random noise.
➤ Displays the overall trend of the data set.
➤**Regression** tends not to work well with volatile data.

Time Series Diagnostics



➤ **Autocorrelation Function (ACF):**
✦ Residual component above does not reflect a random process, hence there is unexplained dependence, possibly a seasonal component.
✦ The ACF calculates the correlation at different lag intervals to help identify any dependence within the data (i.e. Lag 1 = One year) .
✦ If lags in the ACF exceed the confidence interval (blue dotted line), the process is non-stationary, and thus is used in the **ARIMA** model to capture the dependent lag with high autocorrelation.
✦ Many of the series also resulted in a strong autocorrelation at lag 1, indicating a seasonal component at one year.

➤ **Forecast Error:**
✦ Forecast error is the distance between the predictions produced by the model and the last two quarters of the original data.
✦ Identifies which predictive model is closest to the Validation set.
✦ From the auto and home data sets, the forecast error is calculated for the Frequency, Severity, and Pure Premium variables in each series.
✦ Confidence intervals narrow when using ARIMA models versus the insurance firms' regression methods.

Results

Model	Reg.	Exponential Reg.	ARIMA	EWMA	Loess
Accuracy	0.0%	13.9%	38.9%	30.6%	16.7%

➤ Table shows percentage of how often a certain model has the lowest forecast error.
➤ Finding a seasonal component through autocorrelation proves Time Series analysis works well with data.
➤ **ARIMA** modeling best forecasts the insurance's Pure Premium.
➤ The Pure Premium is best estimated when forecasting the Frequency and Severity variables separate.

Citations, and Acknowledgments

[1] Neter, John, William Wasserman, and Michael H. Kutner. Applied Linear Statistical Models. Homewood, IL: Irwin, 1990. Print.
[2] Brockwell, P. J. and R. A. Davis. Introduction to Time Series and Forecasting. 2nd ed. New York, NY: Springer, 2002.
[3] Joshua Ulrich (2013). TTR: Technical Trading Rules. R package version 0.22-0. http://CRAN.Rproject. org/package=TTR
[4] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

★ We would like to thank CSAA and our faculty advisor Mike Ludkovski for their assistance, guidance, and enthusiasm for this research.