Joseph Andrade

Fall 2015

Time Series Project:

Analysis of Seasonality Adjustments in Homeowners Loss Costs

Introduction:

I decided to dedicate my Time Series project to an analysis that will actually prove useful in my professional position: an analysis of the seasonality adjustments used in Loss Cost trends for Homeowners insurance. While Homeowners insurance does cover certain liabilities, a majority of losses historically have been related to the physical property of the home and its contents. Losses in this line of insurance business are therefore very susceptible to external influences such as weather, construction codes, availability of building materials, etc. Many of these factors, particularly weather, are notably seasonal. Thus, one can reasonably expect some seasonality in loss trends.

Often times, actuaries will attempt to smooth this kind of seasonality by using 4-quarter rolling data rather than adjusting the quarterly data for average quarterly deviations. 4-quarter rolling data means that the quarterly data is annualized by also including the three prior quarters (e.g. 2015Q1 includes data from 2015Q1, 2014Q4, 2014Q2, and 2014Q1). As each data point represents a full year of losses, the seasonality is eliminated. However, I have heard anecdotally that this adjustment can "over-smooth" the data, and can imply trends that aren't really there. For instance, if 2014Q1 was a terrible loss year, this will show up in 2014Q1, 2014Q2, 2014Q3, and 2014Q4. If I am trying to predict Pure Premiums for 2015Q1, I would see abnormally high losses for my past 4 quarters, which could look like more of an upward trend than simply an outlier in a single quarter.

There are many ways of smoothing or adjusting Pure Premiums before trying to use past trends to predict future losses (catastrophe loading, weather loading, large loss loading, etc.), but that is beyond the scope of this analysis. I simply wish to adjust for seasonality both ways, attempt to specify a model for each method, and see how different the results really are.

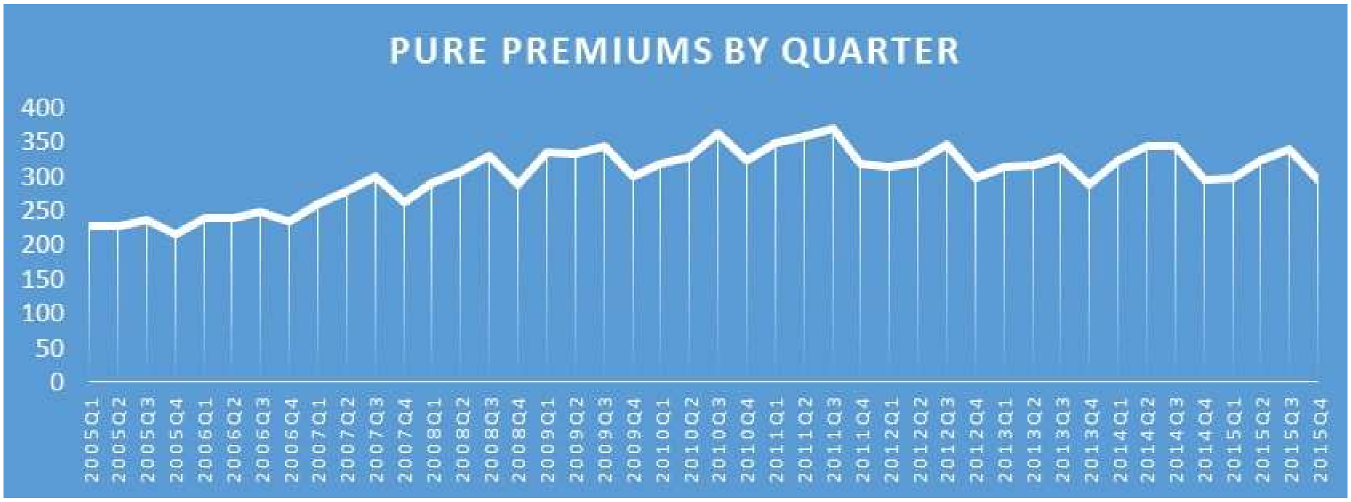
Data Source:

I analyzed 11 years of homeowners loss data by quarter, from 1st Quarter 2005 to 4th Quarter 2015. The data comes from Insurance Services Office (ISO), Homeowners "FAST TRACK DATA" circular reports:

https://www5.iso.com/circsearch/app/start.do

As you require a login to access this data, the downloaded datafiles will be attached.

Pure Premiums (otherwise known as Lost Costs) are calculated as average losses per home insured. The pure premiums over time are shown in the following graph:



All homeowners forms were included, and the data utilized was countrywide. This gives the largest amount of data, for a more credible analysis. Catastrophe losses were excluded from the analysis in order to minimize volatility.

Seasonality Adjustment:

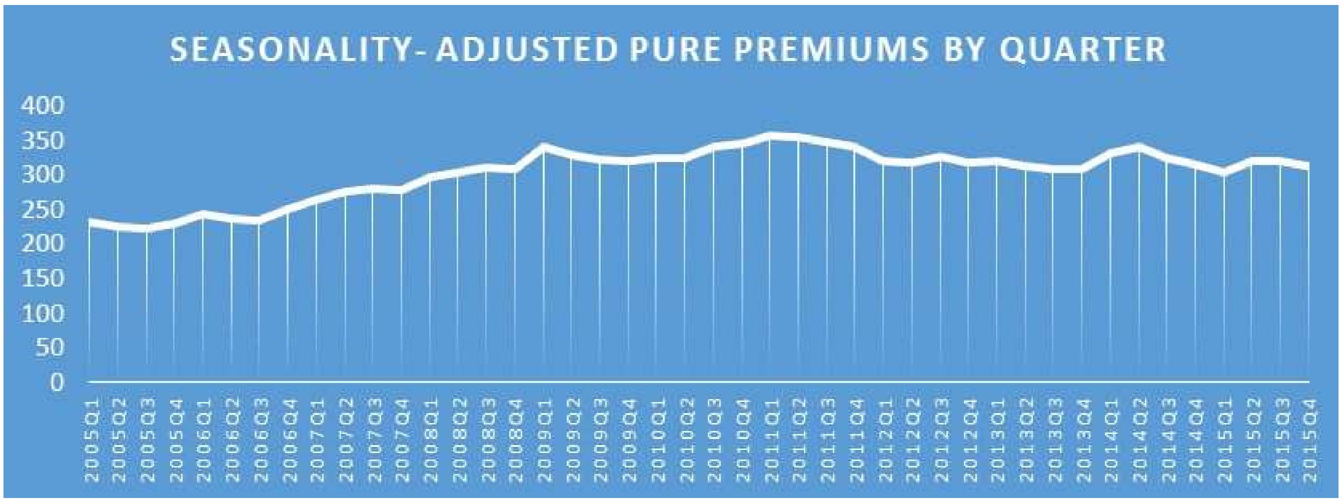
Losses in a heavily property-driven line such as homeowners are generally considered very seasonal. The above graph clearly supports this idea, with Q3 pure premiums in each of the eleven years being higher than the other three quarters in each year (with the exception of 2014, in which pure premiums hold almost perfectly steady from Q2 to Q3). The chart below of the raw data likewise illustrates this point:

Pure Premiums By Year and Quarter					4- Quarter
Year	Q1	Q2	Q3	Q4	Trendline
2005	226.96	226.4	235.67	215.97	
2006	239.33	238.73	248.48	233.83	
2007	259.9	278.51	298.5	261.23	
2008	290.63	305.72	330.06	288.52	
2009	333.95	331.96	343.24	300	
2010	318.1	327.55	362.48	323.05	
2011	349.39	358.57	369.9	319.39	
2012	314.36	321.47	346.33	297.76	
2013	312.46	316.9	328.03	288.93	
2014	324.05	344.98	344.08	296.05	
2015	298.31	322.11	339.55	294	

We therefore calculate average annual relativity factors for each quarter, which are shown below:

Quarter	Relativity Adjustment Factor
Q1	1.01849358
Q2	0.987357874
Q3	0.939129664
Q4	1.065244565

After adjusting our data for seasonality using these factors, we end up with the below graph:



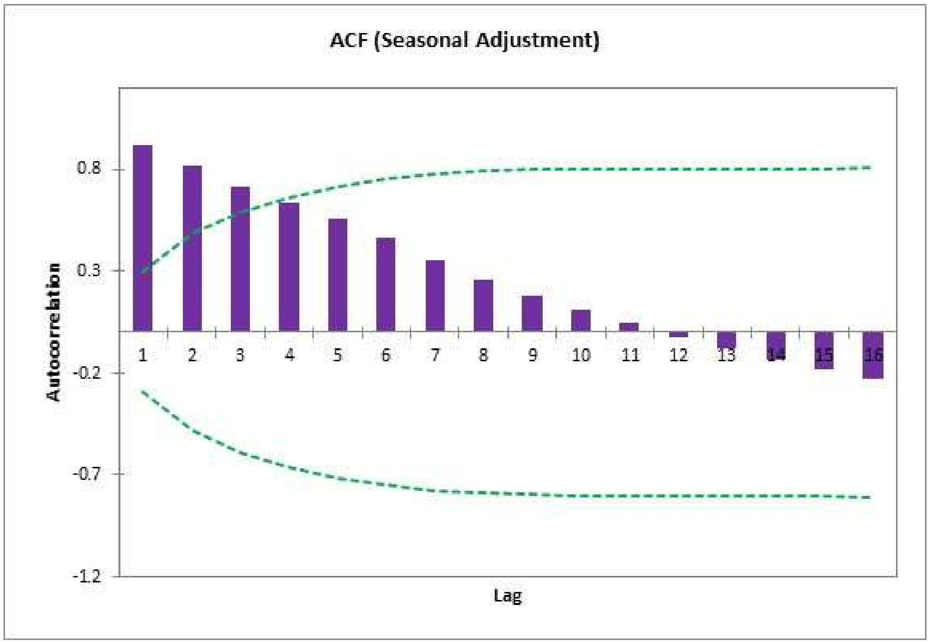
This graph is much smoother than the one before adjusting for seasonality, and a slight upward trend is much more apparent, particularly for years 2005 through 2011, after which the graph levels out somewhat.

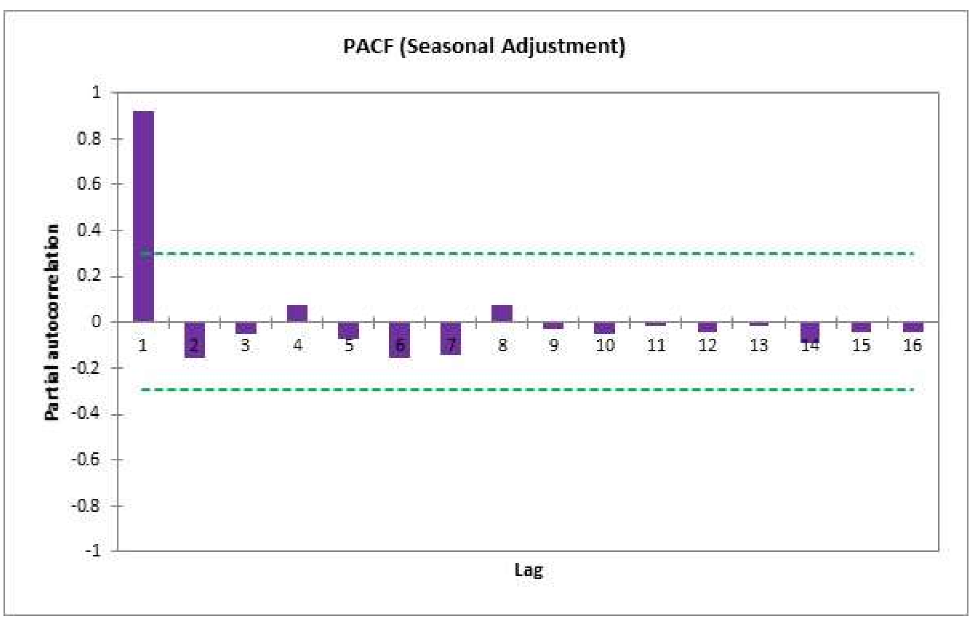
Alternatively, using the 4-quarter rolling seasonality adjustment, we get the following graph:

This graph is notably much smoother than the one that utilizes the average annual relativity adjustments, so at least the concept of “over-smoothing” using this methodology does seem readily apparent.

Specifying a Model (average annual relativity factors):

While the slight upward trend apparent from the seasonality-adjusted graph suggests that the model is non-stationary, we can further verify by plotting the ACF and PACF, as below:

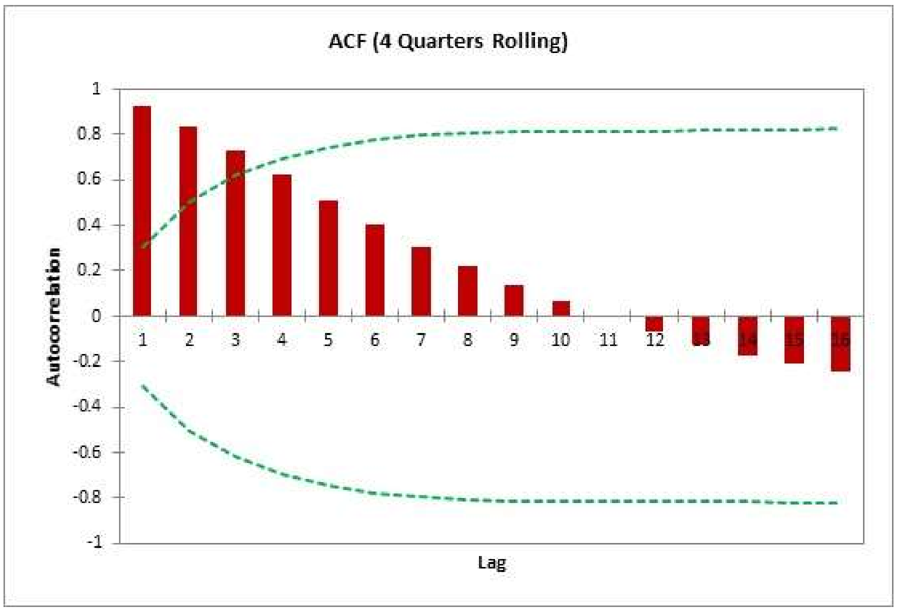


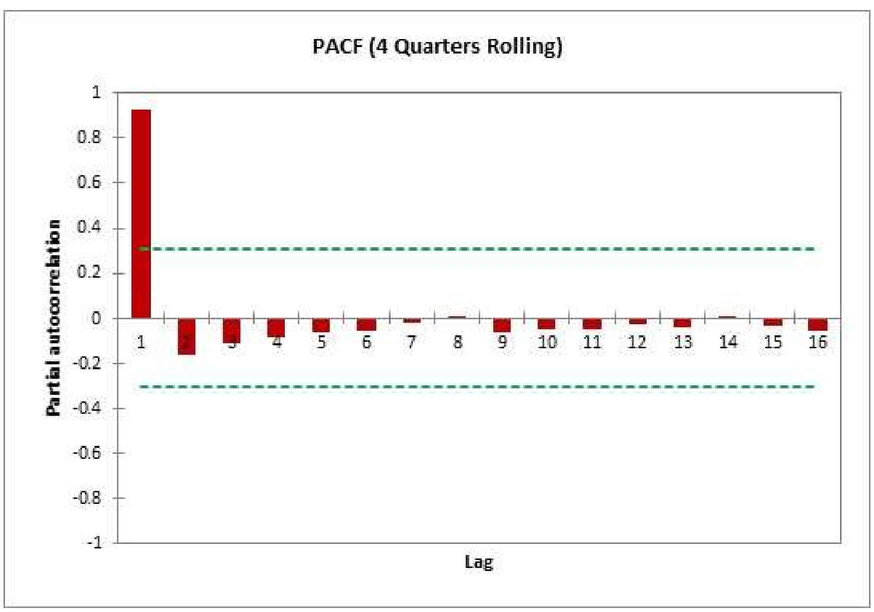


The Autocorrelation function tails off, rather than cutting off sharply to zero (or at least statistically insignificant) autocorrelation. On the other hand, the PACF clearly appears to cut off after 1 lag. Having an ACF that tails off and a PACF that cuts off after lag p is characteristic of an AR(p) model. By contrast, MA(q) models have an ACF that cuts off after lag q and a PACF that tails off, and ARMA(p,q) models have an ACF and PACF that both tail off. So an autoregressive model with p=1 seems most appropriate.

<u>Specifying a Model (4-Quarters Rolling):</u>

We can similarly plot the ACF and PACF for the 4-Quarter Rolling Model:

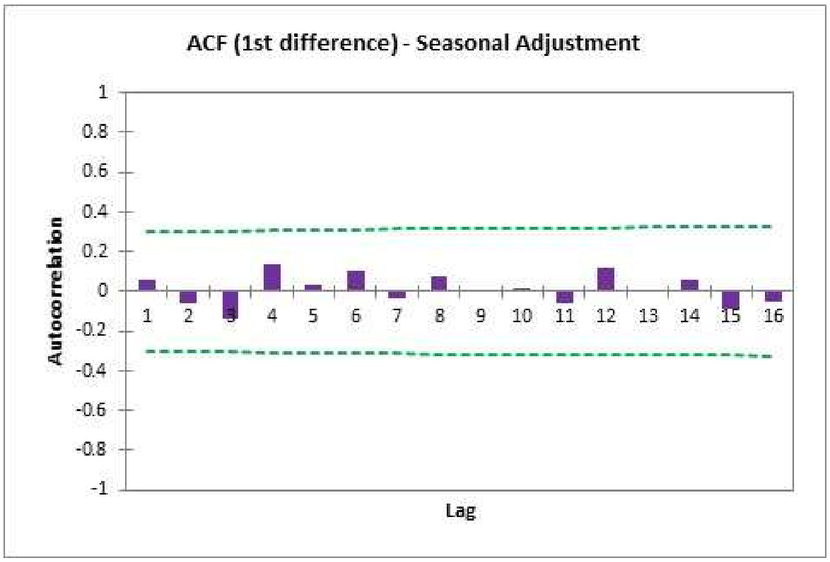


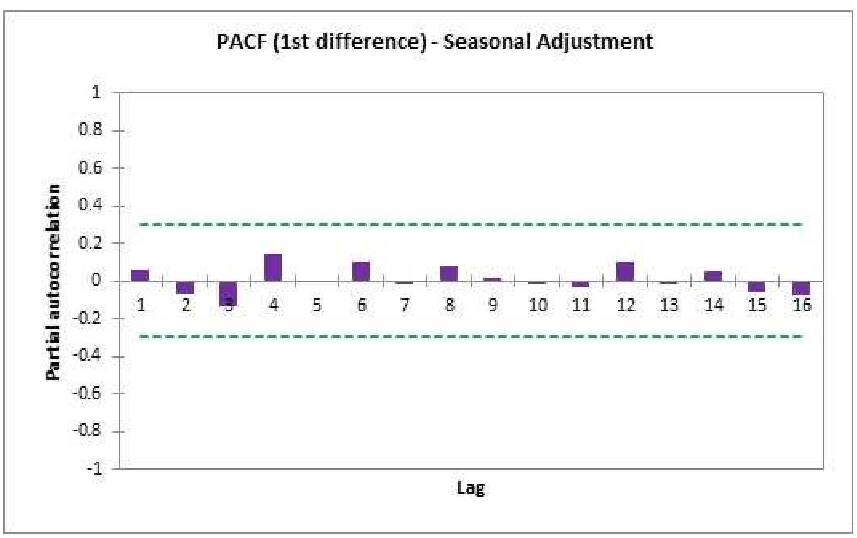


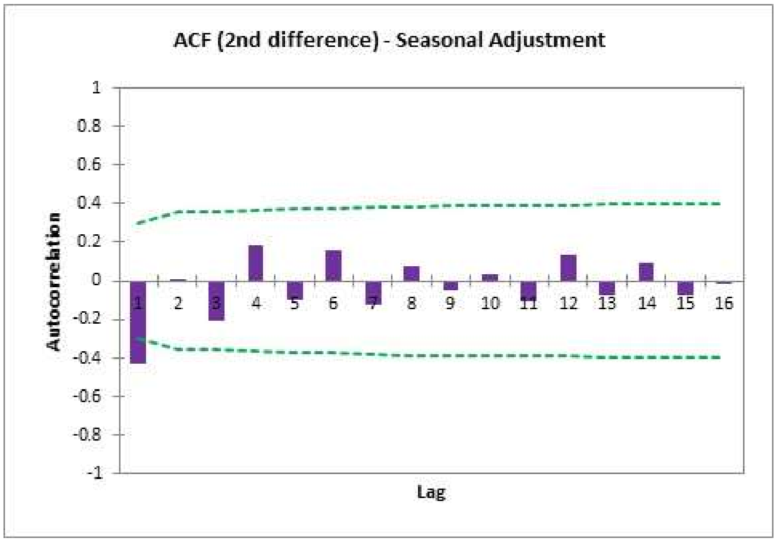
The shape of the ACF and PACF graphs for 4-quarter rolling data is very similar to the ACV graph adjusted using average annual relativity factors for seasonality. Thus, we can draw similar conclusions, and assume an autoregressive model with p=1 seems appropriate.

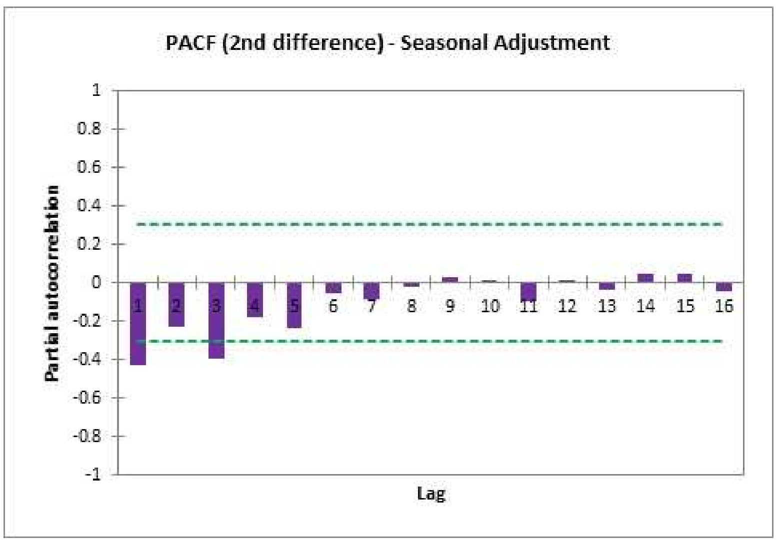
Differencing:

Since the series has positive autocorrelations out to a reasonably high number of lags, it likely requires differencing to correct non-stationarity. A model with one order of differencing assumes that the original series has a constant average trend (e.g. a random walk), whereas a model with two orders of total differencing assumes that the original series has a time-varying trend. It is hard to tell simply from the graph which better suits this particular dataset, so we graph the ACF and PACF functions for both the first-order and second-order differences. We start with the data that has been adjusted using average annual relativity factors. These graphs are shown below:



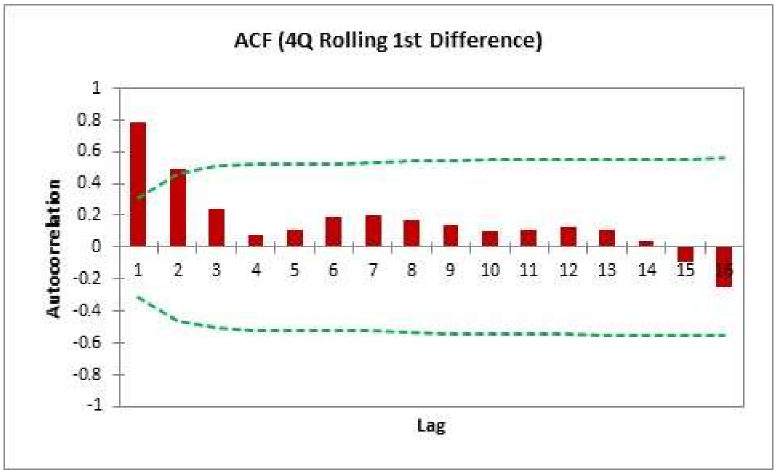


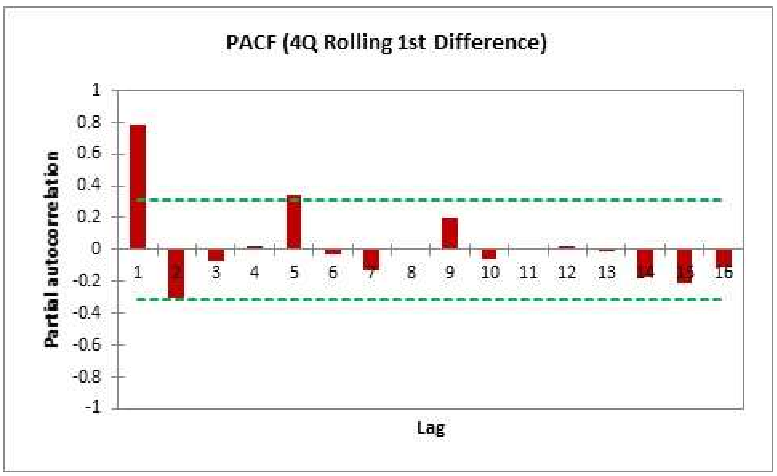


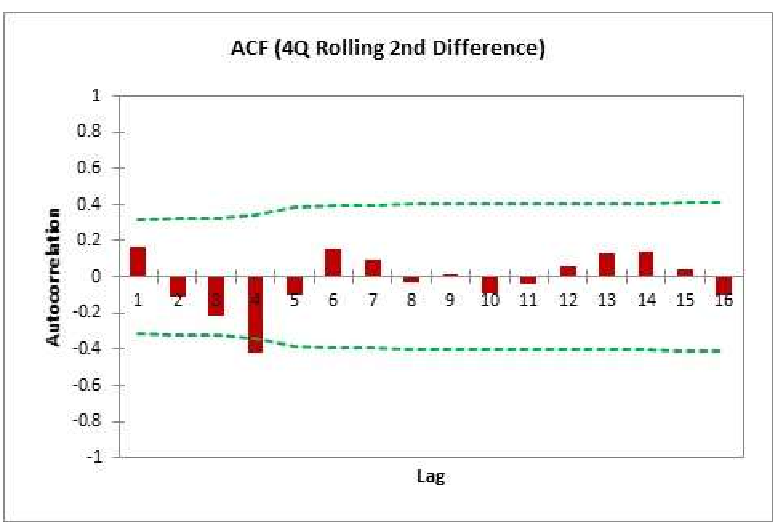


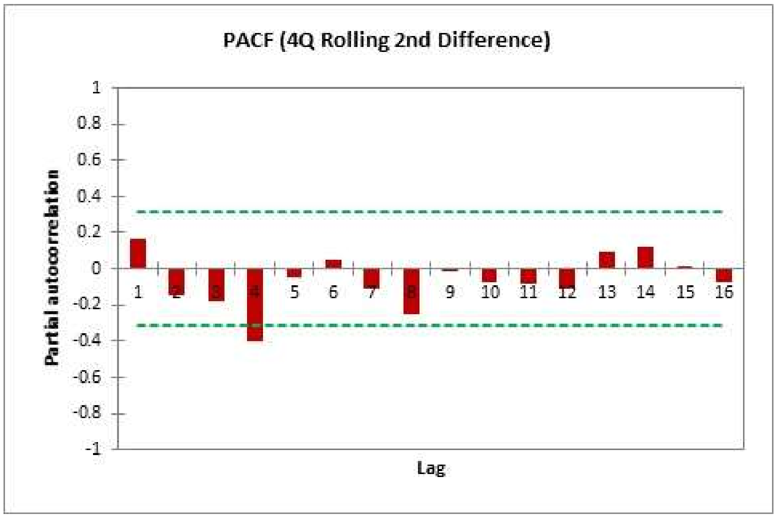
From these graphs it appears that the 2^{nd} difference adds very little value. In fact, more values are significantly different from 0 in the 2^{nd} difference ACF and PACF than in the first difference graphs.

We now create the same 4 graphs for the 4-Quarters Rolling data:


ACF (4Q Rolling 1st Difference)
Autocorrelation
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16


PACF (4Q Rolling 1st Difference)
Partial autocorrelation
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16


ACF (4Q Rolling 2nd Difference)
Autocorrelation
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



In contrast to the data using average annual relativity factors to adjust for seasonality, the 4-quarters rolling data does, from the ACF and PACF graphs, appear to benefit from second differencing.

Validation:

To decide on a final model, I ran both the Seasonal-adjusted and 4-Quarter Rolling data through ARI(1,1) and ARI(1,2) regressions, to see which created the best fit. The results of these regressions are summarized in the below table:

Data	Model	R^2	Durbin-Watson Statistic
Adjusted with Seasonal Factors	ARI(1,1)	0.926	1.9900
	ARI(1,2)	0.897	2.1854
4-Quarters Rolling Data	ARI(1,1)	0.992	1.5439
	ARI(1,2)	0.991	1.9533

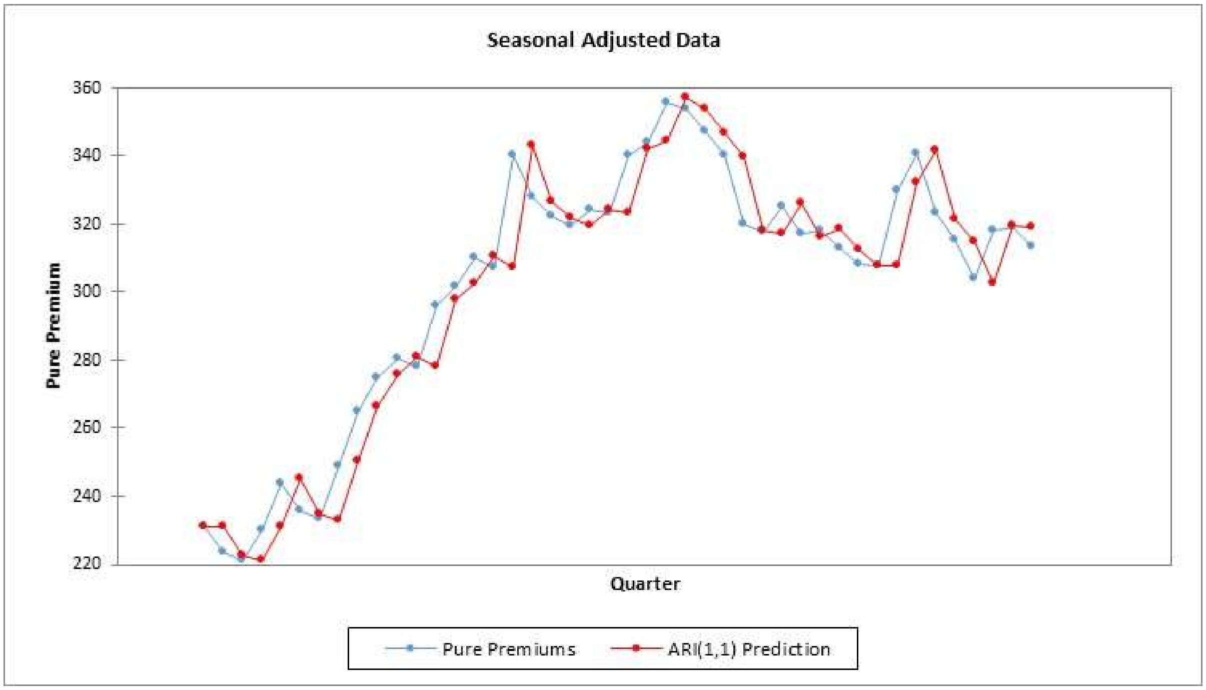
R^2 is a quick measure of goodness of fit between the model and the actual data. The Durbin Watson statistic is a number that tests for autocorrelation in the residuals from a statistical regression analysis. The Durbin-Watson statistic is always between 0 and 4, where a value of 2 means that there is no autocorrelation in the sample.

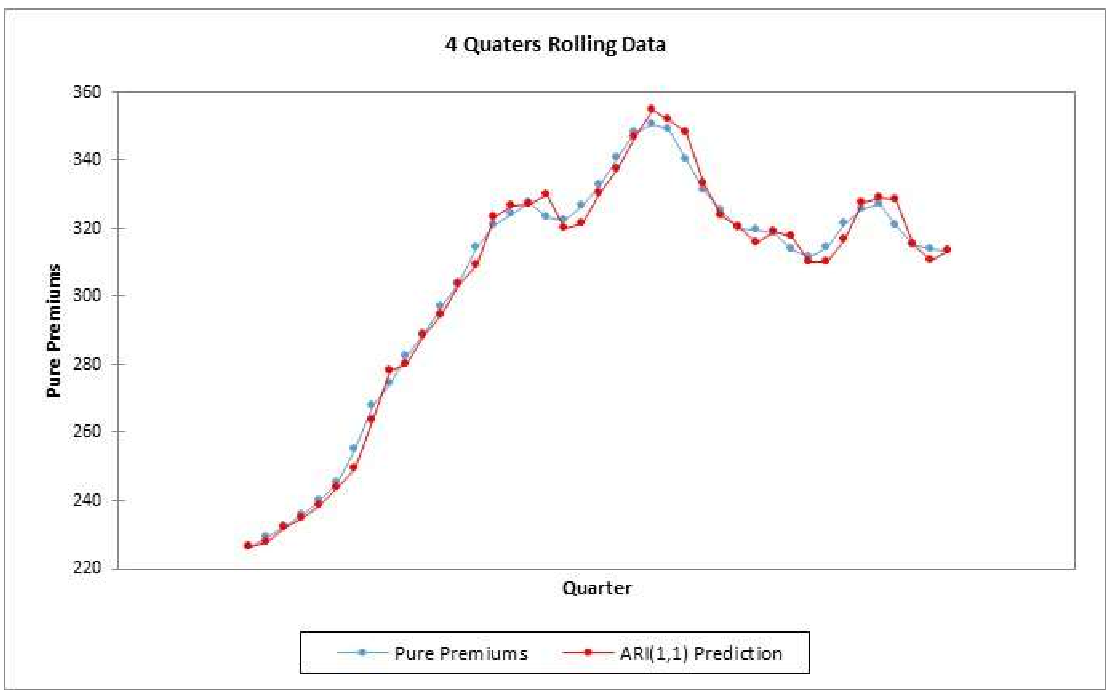
It is clear that, for the data adjusted with seasonal factors, ARI(1,1) is a better choice than ARI(1,2). Not only is the R^2 statistic higher, but the Durbin Watson statistic is also closer to 2, implying that there is less autocorrelation in the residuals for that model.

For the 4-Quarter Rolling data, the R^2 is very slightly better for the ARI(1,1) model, but the Durbin Watson statistic is so much closer to 2 for the ARI(1,2) model that I believe it is a very slightly better fit overall (since its residuals have less autocorrelation). However, for an easy apples-to-apples comparison with the data adjusted with seasonal factors, I will use the ARI(1,1) model for both datasets.

Results and Conclusion:

We can graph the predicted results of the ARI(1,1) models vs the actual data, for a quick view of how well they appear to fit:





As you can see, the fits in both cases appear reasonable. The fit of the 4-quarters rolling data is more precise, which was expected given its R^2 value shown earlier.

Below is a table of the resulting estimates for φ_1 using an ARI(1,1) model for both datasets:

	ϕ_1			
Data	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)
Adjusted with Seasonal Factors	0.095	0.151	-0.202	0.391
4-Quarters Rolling Data	0.818	0.087	0.648	0.988

As you can see, the φ_1 estimate for the 4-quarters rolling data is significantly higher than the data adjusted with seasonal factors.

So what is going on? When you use 4-quarters rolling data, one quarter and its subsequent quarter share ¾ of the underlying data. For instance, 2014Q4 uses data from 2014Q1, 2014Q2, 2014Q3, and 2014Q4, while 2015Q1 uses data from 2014Q2, 2014Q3, 2014Q4, and 2015Q1. Therefore, the correlation between one quarter and its subsequent quarter is going to extremely high. Even 2015Q3 and 2014Q4 share a full quarter of their data, so even at lag 3 you'd expect significantly more correlation than while adjusting individual quarters.

In addition, a curve fitted perfectly to 4 quarters rolling data is not made to predict the next quarter's pure premiums, but rather the next quarter's pure premiums combined with its prior 3 quarters. As such, you would expect significantly higher weight to be put on the prior quarter's data when predicting the next quarter, since they inherently share ¾ of the data to begin with. This translates to a much higher estimate for φ_1.

When using Pure Premiums to select loss trends, it is important to understand your data and what you are trying to predict. While the 4-quarters rolling data does help smooth more volatile loss trends, it gives an unnatural amount of weight to prior quarters that should more realistically be attributed to random fluctuations. This is the concept of "over-smoothing" that I have heard anecdotally in my profession, and is something to be very aware of when using this sort of data. For instance, just because you have a nice smooth trend of 4-quarter rolling pure premiums, you shouldn't be too confident in your prediction for the next month.