Michael R. Larsen
Director & Actuary
May 19, 2015

Application of Time Series Analysis to Estimating Trend in Ratemaking

Agenda

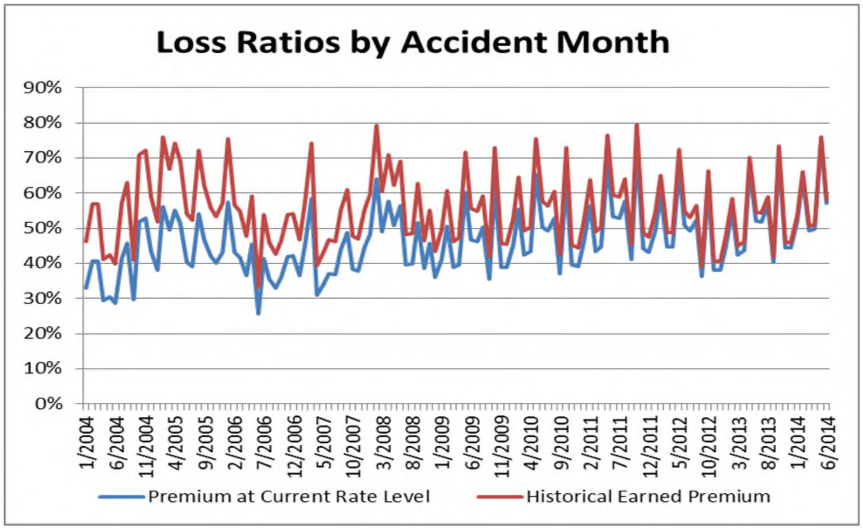
- Today's goal is to illustrate application of two time series techniques
  - ARIMA: Auto Regressive Integrated Moving Average
  - Regression with correction for serial correlation in residuals
  - High level summary of techniques
- Describe sample data set
- ARIMA model example
- Regression with serial correlation correction example
- Conclusion

Sample Data Set

Sample Data Set Construction

- Hypothetical data set built for this presentation with plausible assumptions
- Data available
  - No physical exposure units
  - Earned Premium
  - Earned Premium at Current Rate Level
  - Losses developed to ultimate (assume stable at six months)
- Monthly seasonality built into data set
- Long term loss cost trend of 3% annually
- Market trends for pricing affect observed loss ratios
- On average, the rate changes will match long term trend
- Annual rate changes earned on annual policies

Sample Data Set



ARIMA Example

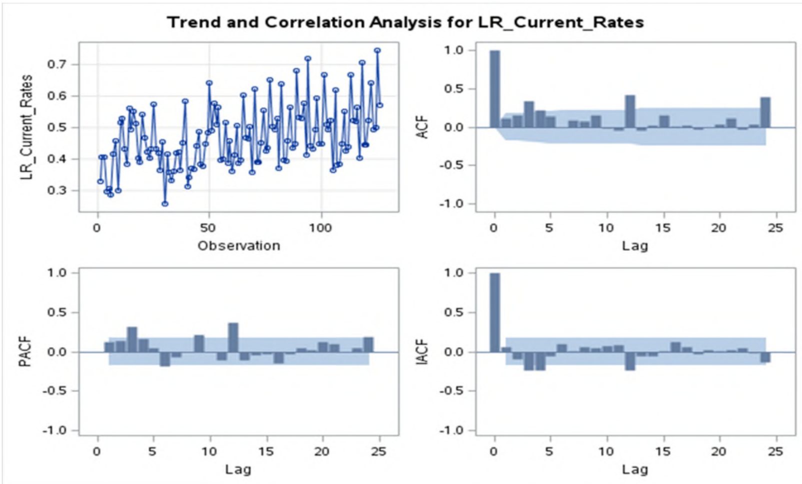
ARIMA Model Background

- Model for mean reverting time series
  - Common technique for analyzing financial time series
  - Sometimes, differencing the observations is necessary to obtain stationary time series
- ARIMA (Auto Regressive Integrated Moving Average) Sample Equation
  - Y_t - .6 Y_{t-1}= ε_t + .4 ε_{t-1}
  - Auto Regressive portion: .6 Y_{t-1}
  - Moving Average portion: .4 ε_{t-1}
  - Assumes mean of zero after differencing
  - Reverse differencing to forecast
- Use autocorrelation (ACF) and partial auto correlation (PACF) results to diagnose likely ARIMA form
- Iterative routines to solve for factors (Conditional Sum of Squares)

Vocabulary

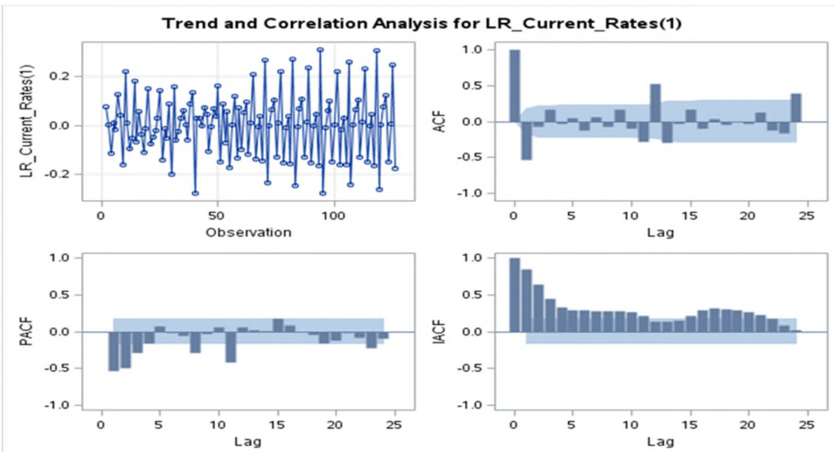
- Conditional Least Squares: Iterative calculation solving for ARIMA least squares parameter estimates assuming unknown prior values equal series mean.
- OLS: Ordinary Least Squares for normal distribution based regression
- Auto Correlation Function: Covariance between observations over time
- Partial Auto Correlation Function: Remaining correlation after removing effect of earlier lag correlation
- Inverse Auto Correlation Function: Calculate auto correlation after inverting the standard ARIMA formula
- Stationary Time Series: Condition for ARIMA model to converge. Time series (after transformation) is mean reverting with no long-term trend

ARIMA Identification Exhibits Before First Differencing



ACF: autocorrelation, PACF: partial autocorrelation, IACF: inverse auto correlation

ARIMA Identification Exhibits After Applying First Difference



ACF: autocorrelation, PACF: partial autocorrelation, IACF: inverse auto correlation

ARIMA Initial Model

THE HARTFORD

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	42.86	6	<.0001	-0.53	-0.075	0.166	-0.031	0.046	-0.127
12	98.96	12	<.0001	0.062	-0.072	0.172	-0.099	-0.278	0.525
18	117.05	18	<.0001	-0.291	-0.038	0.164	-0.102	0.04	-0.046
24	150.94	24	<.0001	0.001	-0.034	0.133	-0.121	-0.17	0.393

Autocorrelation pattern before Moving Average

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.86476	0.04649	18.6	<.0001	1

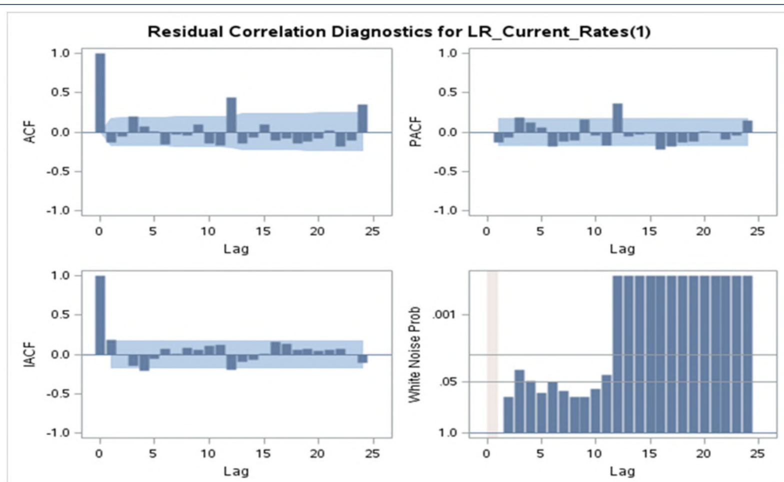
Parameter Estimate for Moving Average

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	10.59	5	0.0602	-0.106	-0.035	0.207	0.082	0.023	-0.135
12	45.15	11	<.0001	-0.012	-0.028	0.11	-0.129	-0.153	0.44
18	54.64	17	<.0001	-0.13	-0.057	0.105	-0.096	-0.074	-0.14
24	82.8	23	<.0001	-0.106	-0.076	0.023	-0.168	-0.095	0.355

Autocorrelation pattern after Moving Average Model

ARIMA Diagnostics After Application of Moving Average

THE HARTFORD



Combination of ACF & PACF indicate autoregressive at lag 12

ARIMA Model Results with Combined MA(1) AND AR(12)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.76574	0.05873	13.04	<.0001	1
AR1,1	0.54702	0.08411	6.5	<.0001	12

Both parameters significant

Autoregressive Factors	
Factor 1:	1 - 0.54702 B**(12)

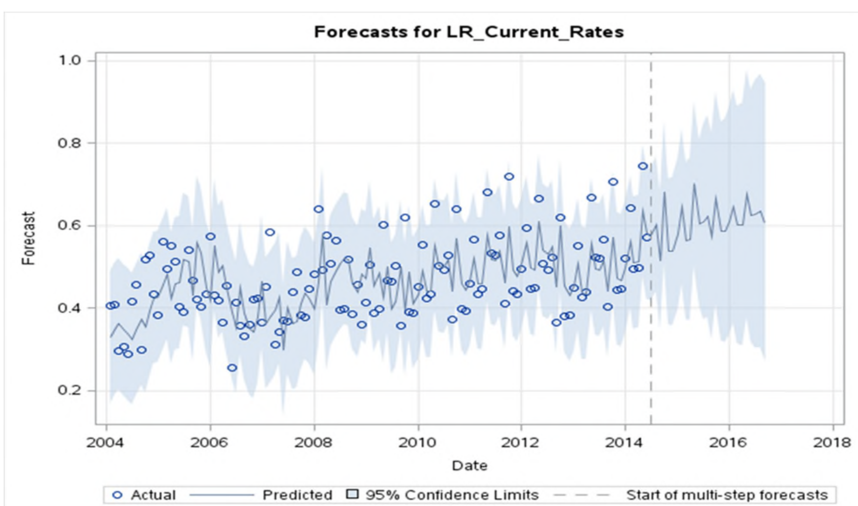
Moving Average Factors	
Factor 1:	1 - 0.76574 B**(1)

Parameters displayed in ARIMA equation format

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	7.47	4	0.113	-0.092	-0.024	0.114	0.149	0.083	-0.077
12	12.69	10	0.2416	-0.012	-0.043	0.045	-0.063	-0.102	-0.138
18	23.62	16	0.0982	-0.062	0.037	0.051	-0.235	-0.108	0.011
24	32.06	22	0.0764	-0.094	-0.093	0.001	-0.132	0.001	0.141

Model passes white noise test

ARIMA Model Actual & Projected



Regression with Serial Correlation Correction Example

Regression with Serial Correlation in Error Correction

- Method is Generalized Least Squares
  - Weights observations using correlation effect on variance
  - Calculates revised estimate of regression accuracy using adjusted variance
- Analysis Sequence
  - Run OLS on Data Set
  - Estimate significant correlation by lag time
  - Re-run with Generalized Least Squares

Regression Before Serial Error Correlation Correction

THE HARTFORD

Ordinary Least Squares Estimates			
SSE	4.626879	DFE	130
MSE	0.03559	Root MSE	0.18866
SBC	-57.956	AIC	-63.7216
MAE	0.153815	AICC	-63.6286
MAPE	23.19183	HQC	-61.3787
Durbin-Watson	2.05	Regress R-Square	0.1918
		Total R-Square	0.1918

Durbin-Watson indicates correlation in residuals.

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-0.9368	0.033	-28.36	<.0001	
Time	1	0.002393	0.000431	5.55	<.0001	Time

Time is monthly on log scale for loss ratio. Annual trend of about 2.9%.

Regression Serial Correlation in Residuals Analysis

THE HARTFORD

Backward Elimination of Autoregressive Terms			
Lag	Estimate	t Value	Pr > |t|
5	-0.00474	-0.05	0.9587
8	0.023218	0.28	0.7832
1	0.04065	0.47	0.637
9	-0.0554	-0.63	0.5273
10	0.084678	1.01	0.315
2	-0.11027	-1.39	0.1676
4	-0.12345	-1.51	0.1335
7	0.106178	1.33	0.1846
6	0.142359	1.73	0.0864

Data is monthly leading to question of lag over 12 months.

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
3	-0.21972	0.079946	-2.75
11	0.166534	0.079923	2.08
12	-0.32073	0.079935	-4.01

These results will be used in adjusting prediction.

Regression with Serial Errors in Residual Correction Results

THE HARTFORD

Yule-Walker Estimates			
SSE	3.632187	DFE	127
MSE	0.0286	Root MSE	0.16912
SBC	-73.2619	AIC	-87.6759
MAE	0.131619	AICC	-87.1997
MAPE	19.07665	HQC	-81.8187
Durbin-Watson	1.941	Regress R-Square	0.1313
		Total R-Square	0.3655

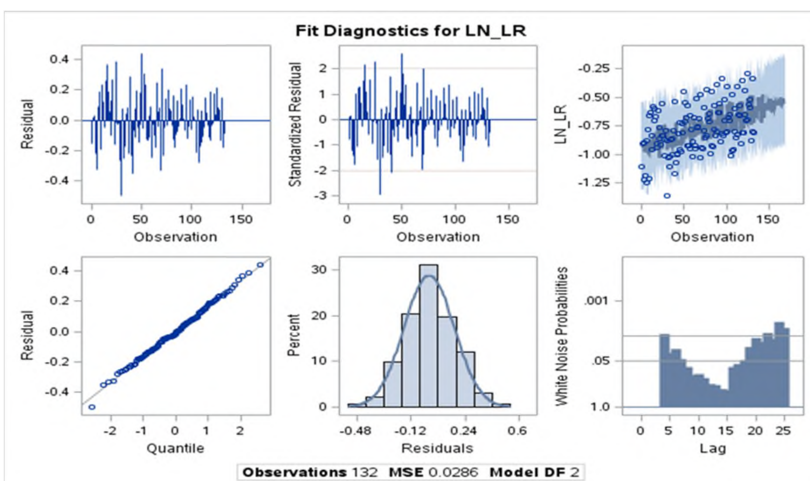
Note that estimation changed from OLS to Yule-Walker. Tot R-squared improves with recognizing correlation in residuals.

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-0.9422	0.0438	-21.52	<.0001	
Time	1	0.002463	0.000562	4.38	<.0001	Time

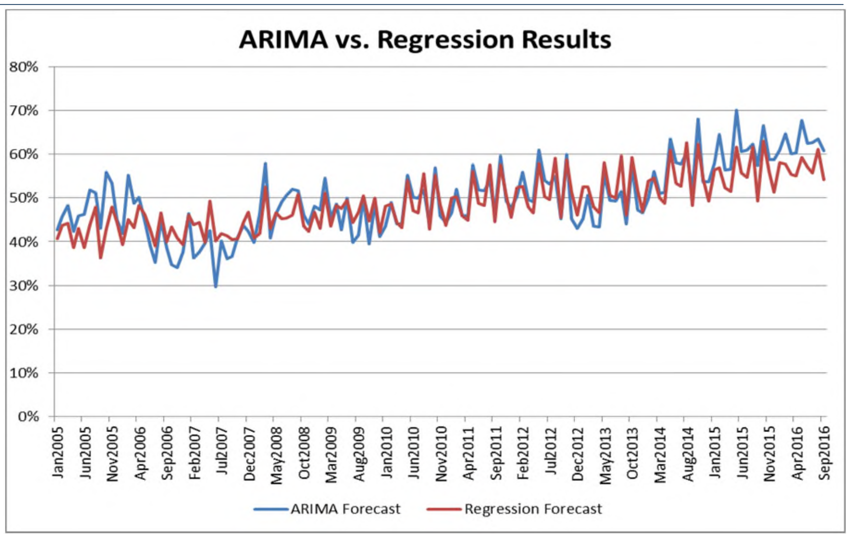
Regression with Serial Errors in Residual Correction Diagnostics

THE HARTFORD



Conclusion

Results from Modeling Loss Ratios at Current Rate Level



 22

Conclusion

- Trend Data is Time Series Data
  - Statistics collected on regularly spaced intervals
  - Limited explanatory variables
  - Seasonality and correlation are common in trend data
- Time series models use patterns in correlation
- Models shown are a small sample of time series models
- Advances in software make time series modeling accessible
  - Numerous packages in R
  - GUI front end in SAS for basic ARIMA
- Time series methods offer improvement in ratemaking trend analysis

References

- Brocklebank, J. C.; Dickey, D. A., *SAS for Forecasting Time Series (2nd Edition)*, 2003, SAS Institute
- Cowpertwait, P. S.; Metcalfe, A. V., *Introductory Time Series with R*, 2009, Springer
- Chatfield, C., *The Analysis of Time Series An Introduction (6th Edition)*, 2004, Chapman & Hall/CRC