SOCIETY OF ACTUARIES

Article from:

ARCH 2014.1 Proceedings

July 31-August 3, 2013

Trend Analysis Algorithms and Applications to Health Rate Review

Ye (Zoe)Ye, Sarah M. Lin, Le Yin, Qiang Wu, and Don Hong

Actuarial Science Program
Department of Mathematical Sciences
Middle Tennessee State University
Murfreesboro, Tennessee

Outline

- Introduction
- Data Preprocessing
- Trend Analysis Algorithms and Package
- Application Results

TN Healthcare Rate Review Project

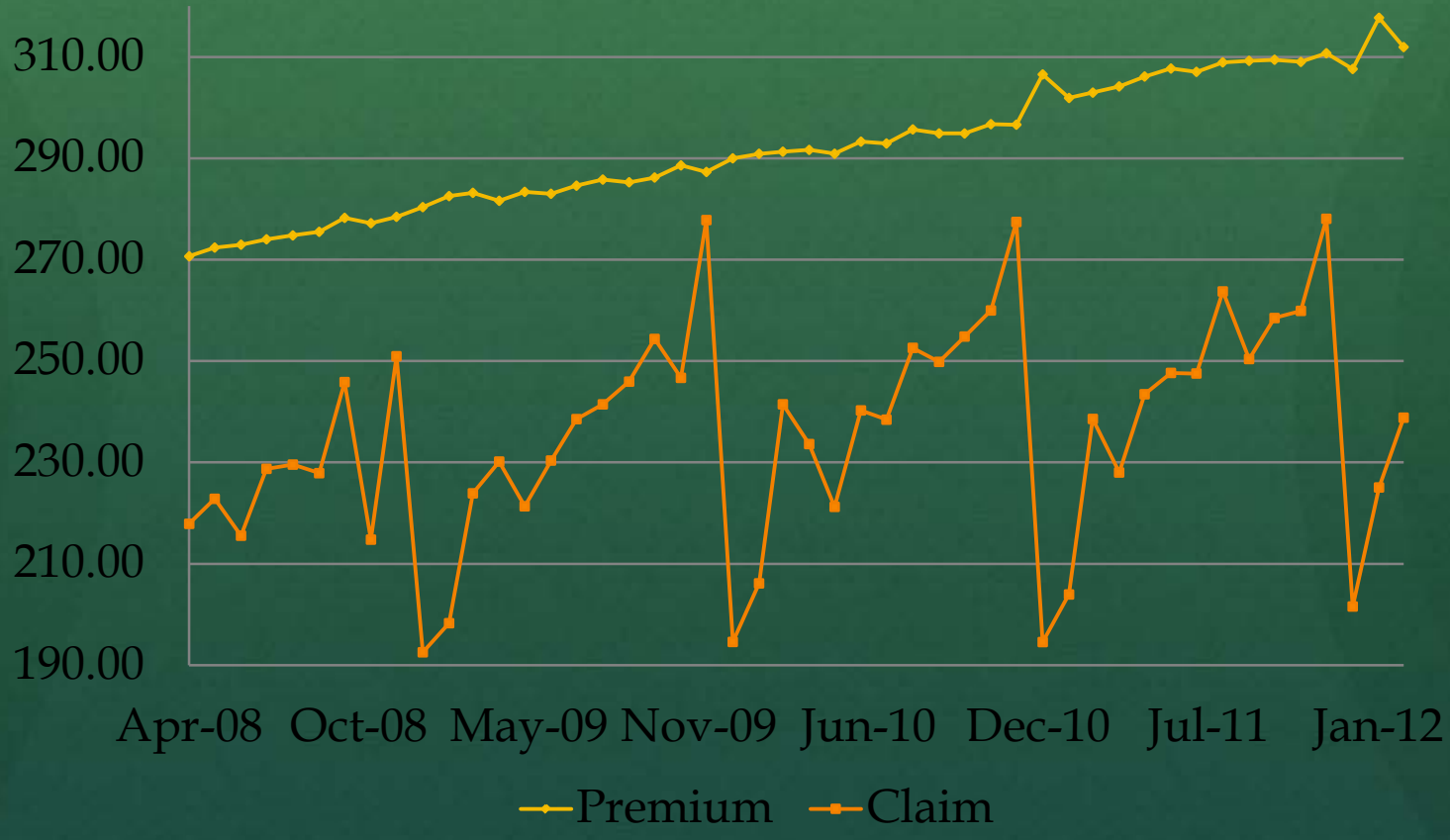
- MTSU's Actuarial Science Program was selected by the Tennessee Department of Commerce and Insurance(TDCI) to evaluate the rate review procedure. *(TN State received both Cycle I & Cycle II grants from the HHS)*
  - Cycle I: Actuaries' perspective on rate review process: evaluations, suggestions, improvements
  - Cycle II: Training courses and development for trend analysis.
- HHS released a final rule that addresses an assortment of issues with respect to the PPACA medical loss ratio (MLR) requirements.

Challenges

- There has a lot of factors which can be considered as effects on trend analysis:
- Trend analysis challenges:
  - Population Attributes
    - Aging / Morbidity/ Care management/ Selection by need
  - Accounting Practices
    - Cost shifting/ Billing and coding changes/ Inflation/ Benefit changes
  - Seasonality
  - Credibility
  - Deductible leveraging
  - MLR limitation
  - Projected period

Data Preprocessing

- Analysis on raw data



Data Preprocessing

- Needs of preprocessing from the raw data:
  - Data value among years can not be compared due to inflation rate
  - Data value are unstable
  - Data doesn't have other factors which may influence on the future trend.
- Adjustments:
  - Use individual incurred claims--per member per month data(PMPM)
  - Smooth data

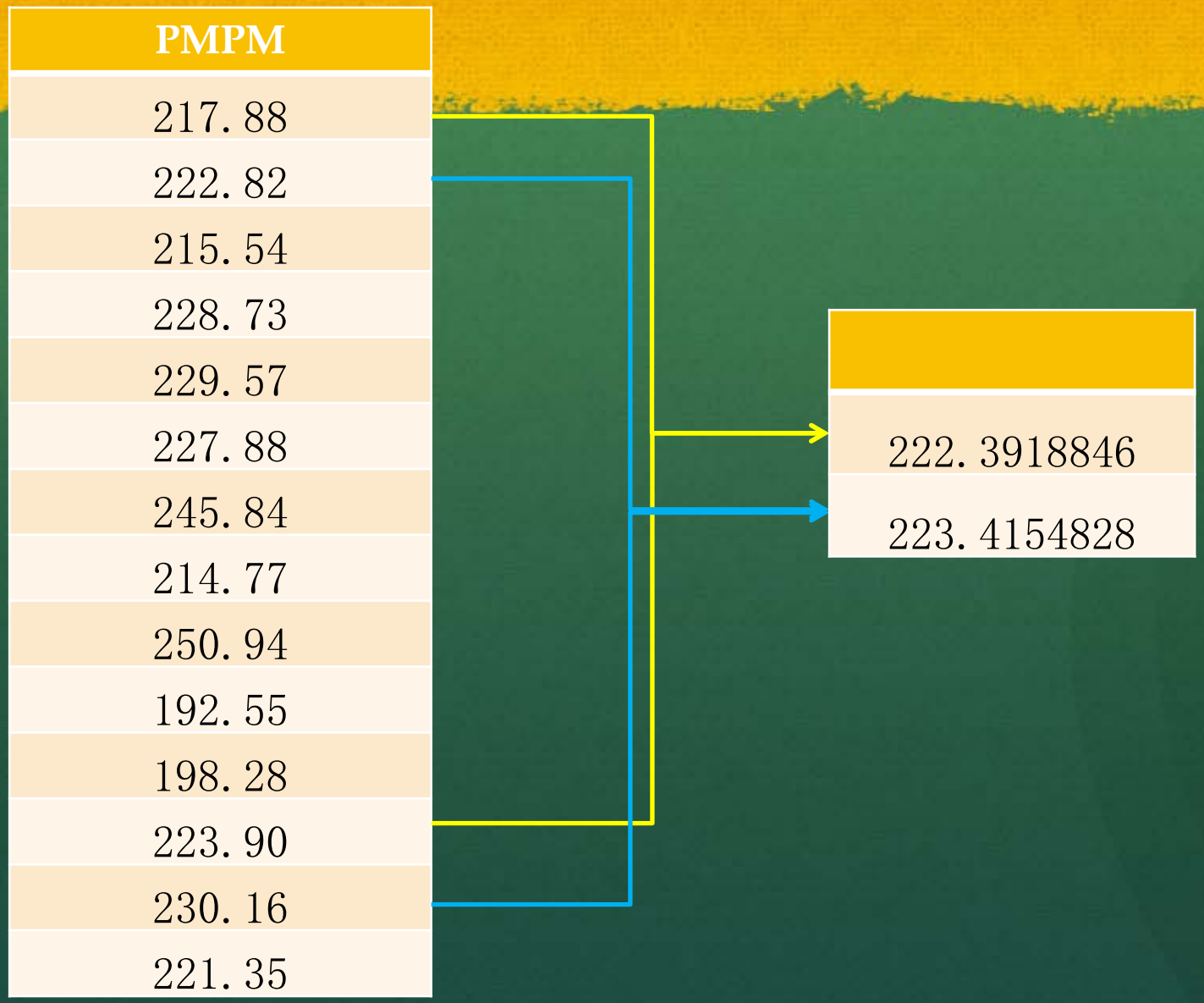
Data Preprocessing

- Get the PMPM:

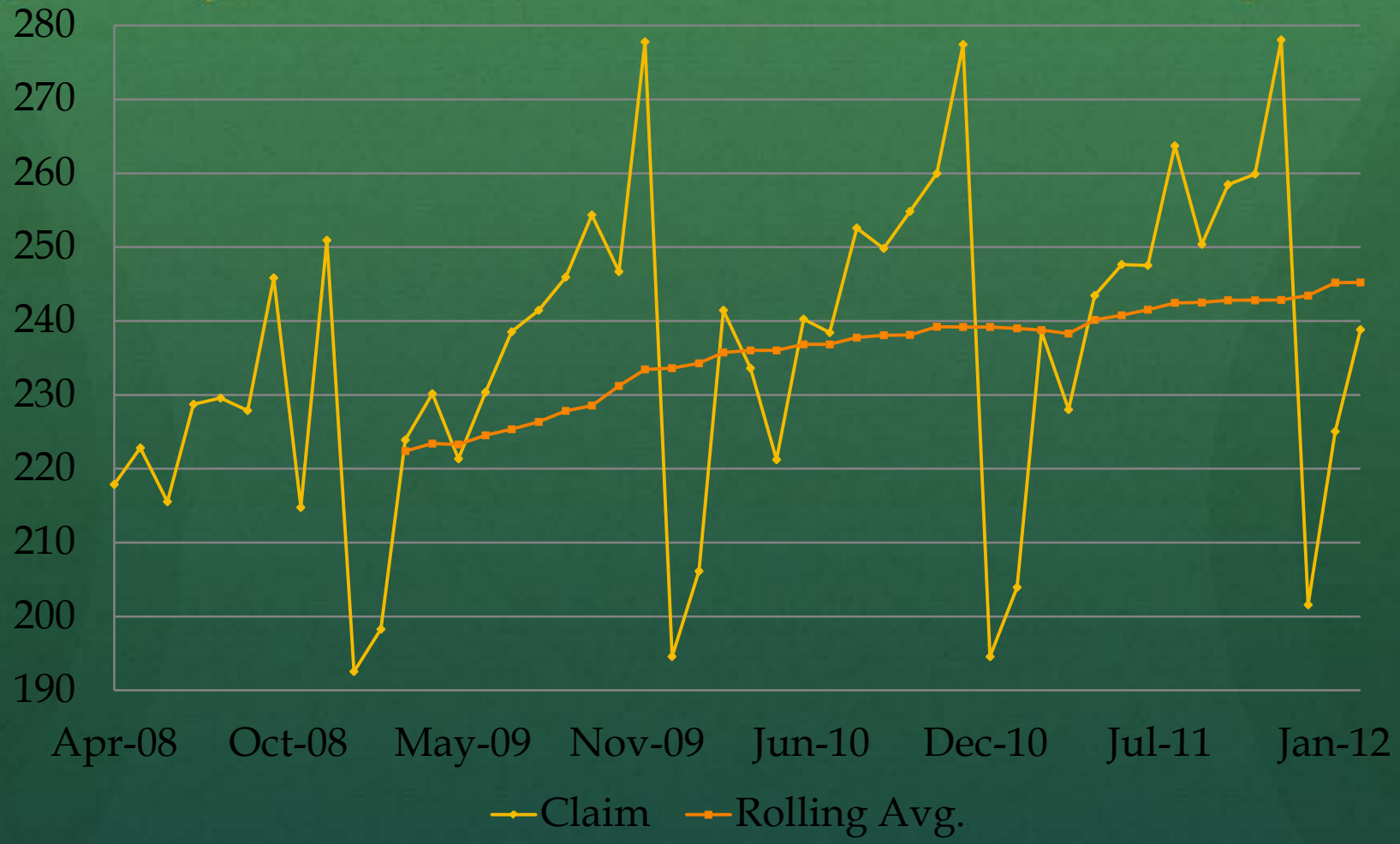
$$Individual\ Incurred\ Claims = \frac{Icurred\ Claims}{Members}$$

- Smooth data – Rolling Average method:
  - Use 12-month rolling average
  - Undo rolling when calculate real predicted claims

Predicted Claims = Predicted 12-month Rolling Average Claims ×12 – the sum of previous eleven month Claims.


PMPM
217. 88
222. 82
215. 54
228. 73
229. 57
227. 88
245. 84
214. 77
250. 94
192. 55
198. 28
223. 90
230. 16
221. 35
222. 3918846
223. 4154828

Rolling Average Data



Trend Analysis Algorithms

- Rolling Average Method

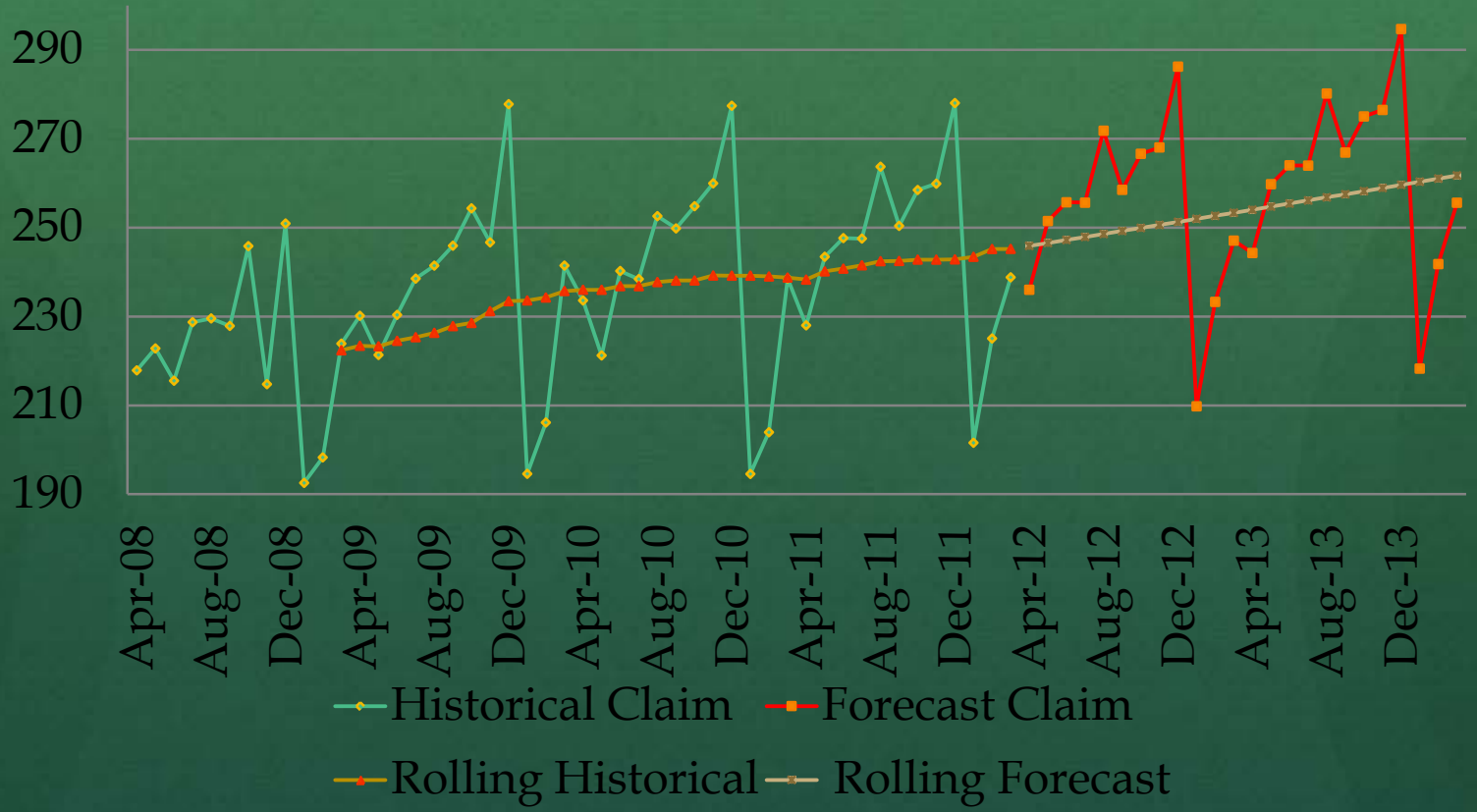
$$Increasing\ Rate_i = \frac{Claim_i - Claim_{i-1}}{Claim_{i-1}} * 100\%$$

$$Average\ Rate = \frac{\sum_{i=1}^{m}(Increased\ Rate\ i)}{m}$$

Rolling Average Data & Trend

Date	Claim Per Member	Claim Per Member (R)	% Change
Mar-09	223.90	222.39	(No Prior Month)
Apr-09	230.16	223.42	0.46%
May-09	221.35	223.29	-0.06%
Jun-09	230.37	224.53	0.55%
…	…	…	…
…	…	…	…
Dec-11	278.06	242.86	0.02%
Jan-12	201.58	243.44	0.24%
Feb-12	225.06	245.20	0.72%
Mar-12	238.83	245.22	0.01%
		Average %Δ	0.27%

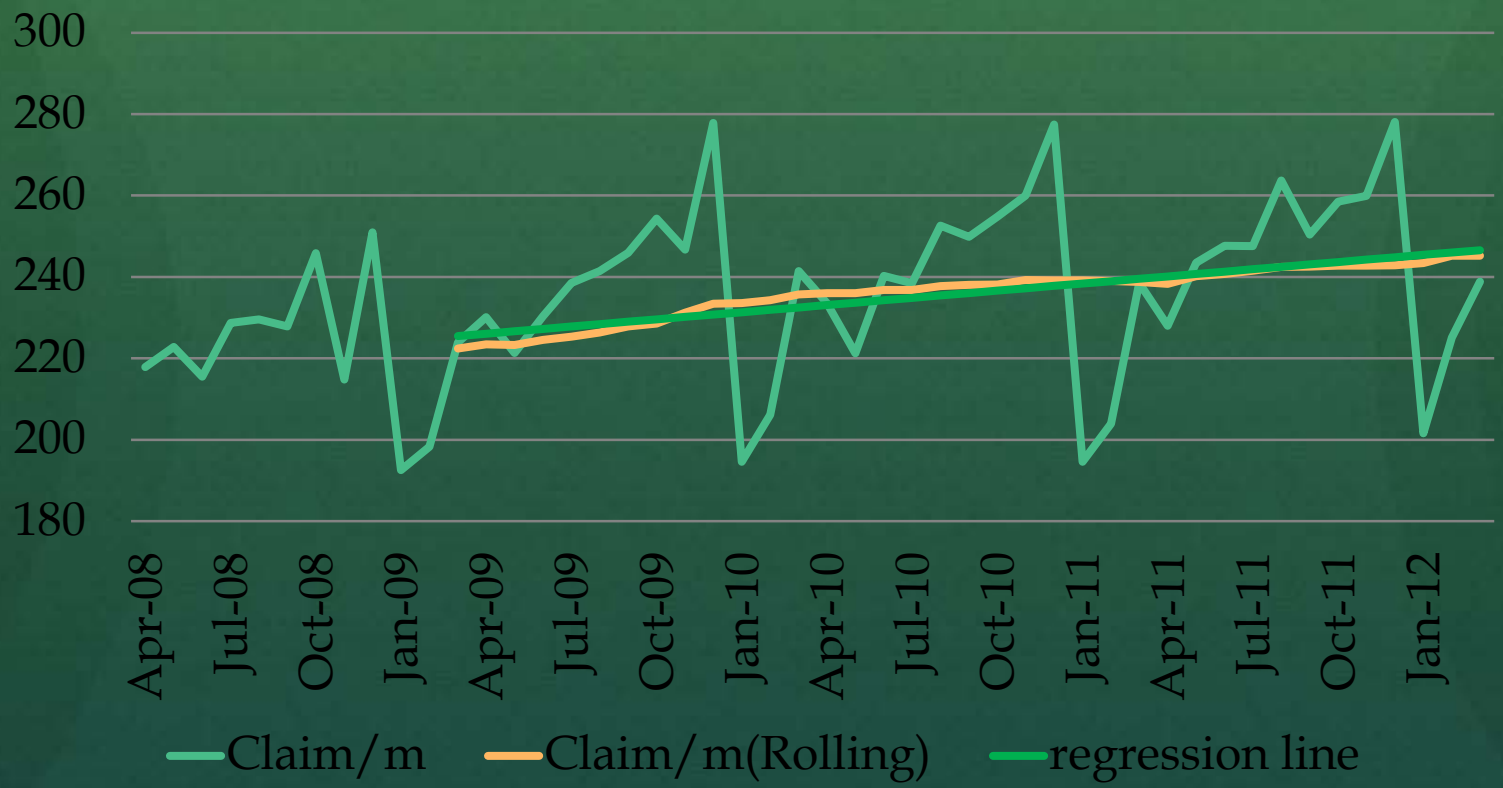
Rolling Prediction



Linear Regression

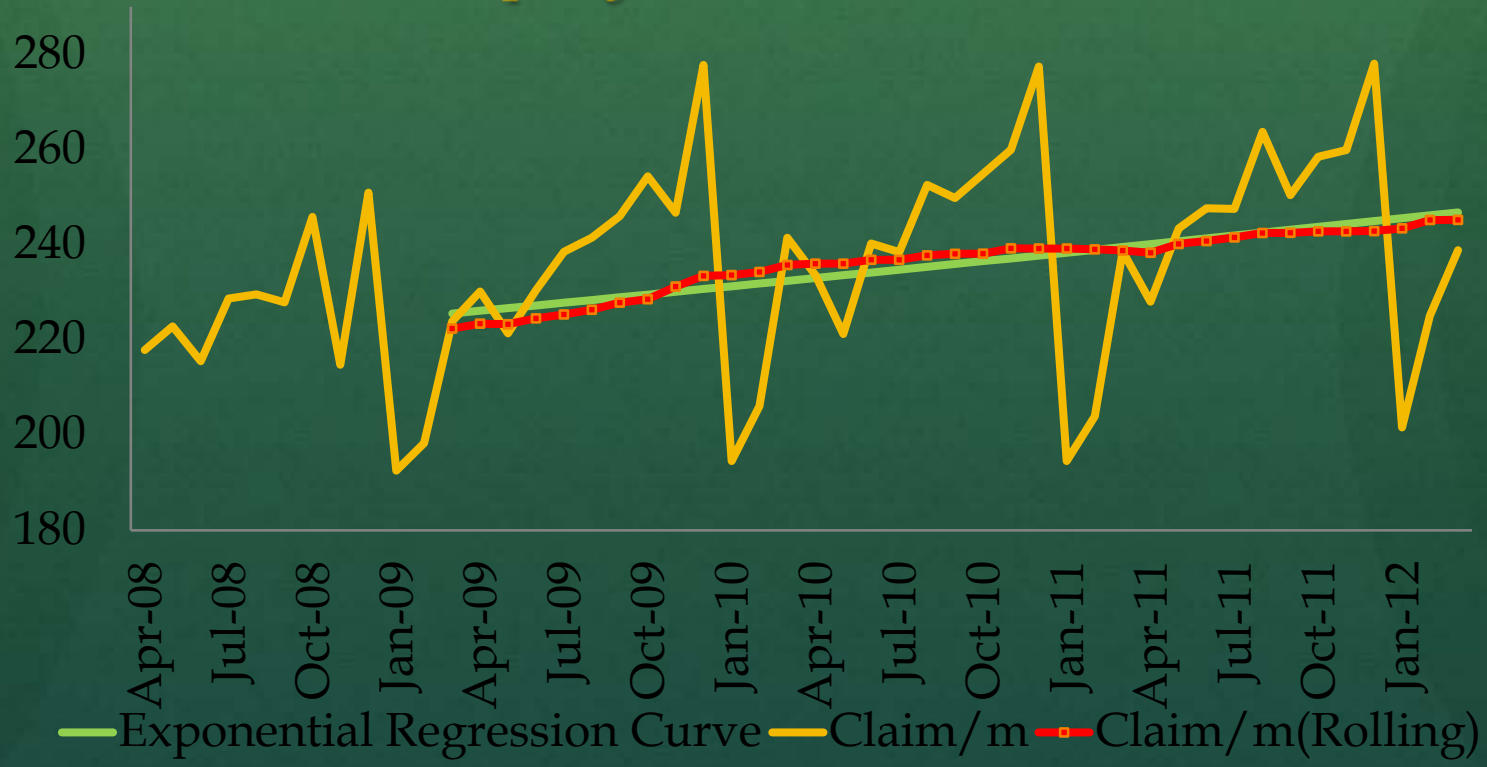
Y = Slope * X + Intercept

$y = 0.521462x + 222.7552$



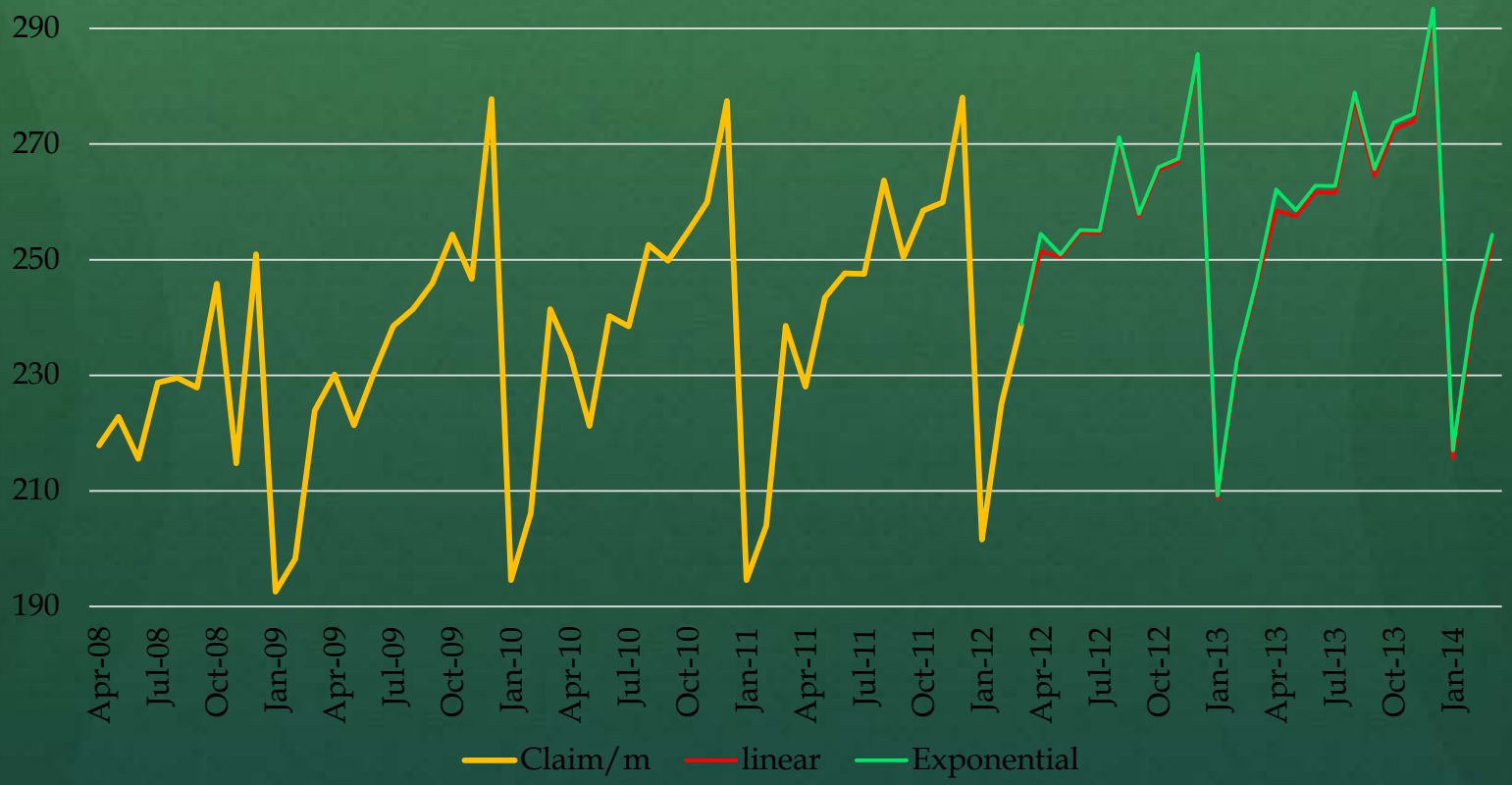
Exponential Regression

- Ln(Y) = Slope * X + Intercept
- $Y = e^{0.002459\,x+5.412866}$



Linear vs. Exponential Regression

- Short term and long term forecasting.



Multiple Linear Regression

- Because of seasonality of the monthly data, rolling average is also suitable during the data preparation.
- If we have the detailed data of every trend components, such as age, policies, inflation, cost shifting and so on, we can use multiple linear regression, $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \cdots$, where x_i represents the quantity of the i^{th} trend factor.

Autoregressive Model (AR): Time Series

- **Autoregressive(AR) Model** explains data in a time series by lagged values.
- The AR(p) model
  - Assumes the value Y_t linearly depends on its p lagged values:

    $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t$$

  - The AR(1) is a special case with p=1
  - It implies each value is determined by one period ago.

    $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$

Choosing The Correct "p"

- Correct "p" indicates the number of appropriate lags
- For each $AR(p)$ model find the $RSS(p)$

$$RSS(p) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

  - Where $\hat{Y}_i$ is the value ***estimated*** by $AR(p)$ at period i
  - $RSS(p)$ is the residual sum of squares
- $RSS(p)$ used to compute $\hat{p}$, the estimator of correct p
  - Correct "p" minimizes $BIC(p)$ among
    - $p = 0, 1, \ldots, p_{max}$, where
      - p_{max} is the largest value of p considered.

Choosing The Correct "p"

- Bayes information criterion (BIC):
  - Chosen p minimizes the Bayes information

$$BIC(p) = \ln\left(\frac{RSS(p)}{n}\right) + (\mathrm{p}+1)\frac{\ln n}{n}$$

- Akaike information criterion (AIC):
  - Chosen p minimizes Akaike information

$$AIC(p) = \ln\left(\frac{RSS(p)}{n}\right) + (\mathrm{p}+1)\frac{2}{n}$$

- AIC may slightly over estimate the optimal p.

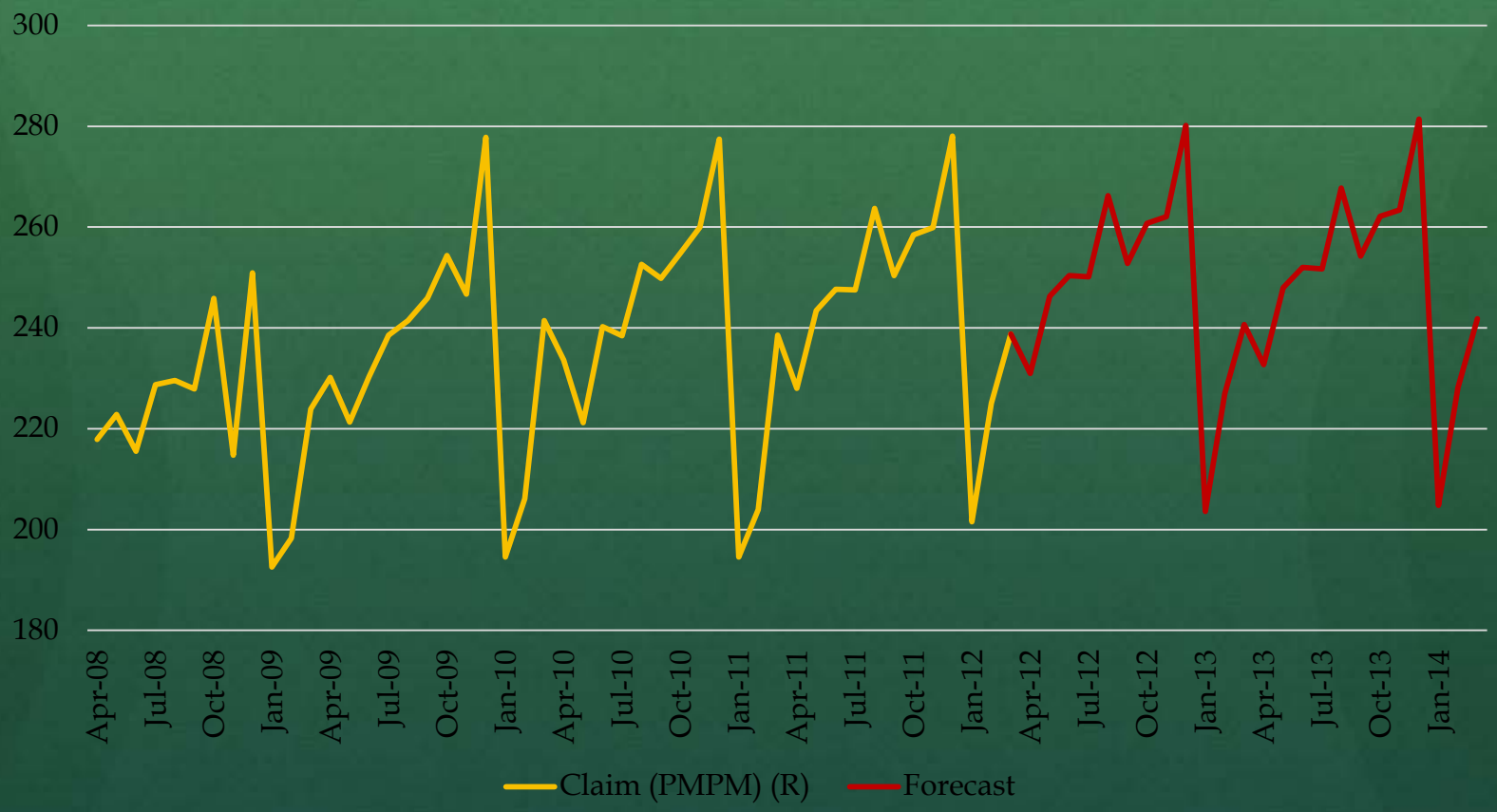
Optimal p for AR(p)

p	BIC	AIC
1	-0.56801	-0.65598
2	-0.43986	-0.57317
3	-0.42983	-0.6094
4	-0.31786	-0.5446
5	-0.26237	-0.5372

AR(1): Rolling Average Data

Date	Claim (PMPM) (R) Y_t Variable	1st Lag X_1 Variable
Mar-09	222.39	
Apr-09	223.42	222.39
May-09	223.29	223.42
Jun-09	224.53	223.29
Jul-09	225.35	224.53
Aug-09	226.34	225.35
Sep-09	227.84	226.34
…	…	…
Nov-11	242.80	242.81
Dec-11	242.86	242.80
Jan-12	243.44	242.86
Feb-12	245.20	243.44
Mar-12	245.22	245.20

AR(1) Forecast: Rolling Average Data



Software Package

- Use previous four methods to predict next year's/month's cost trend
- Annual data:

$$Increasing\ Rate = \left(\frac{Data_{i+1}}{Data_i} - 1\right) * 100\%$$

- Monthly data:

$$Increasing\ Rate\ of\ i + 1^{th}\ month = \left(\frac{M_{i+1}}{M_i} - 1\right) * 100\%$$

Cost Trend Software

This software is used for project the future Annual or Monthly cost trend. The data we need is "Year" and "PMPM" (Per Month Per Member). If the data you get are not PMPM, you need to calculate this first. The use of the software is as follows:

First click the buttom "ENTER" in the corner;Then

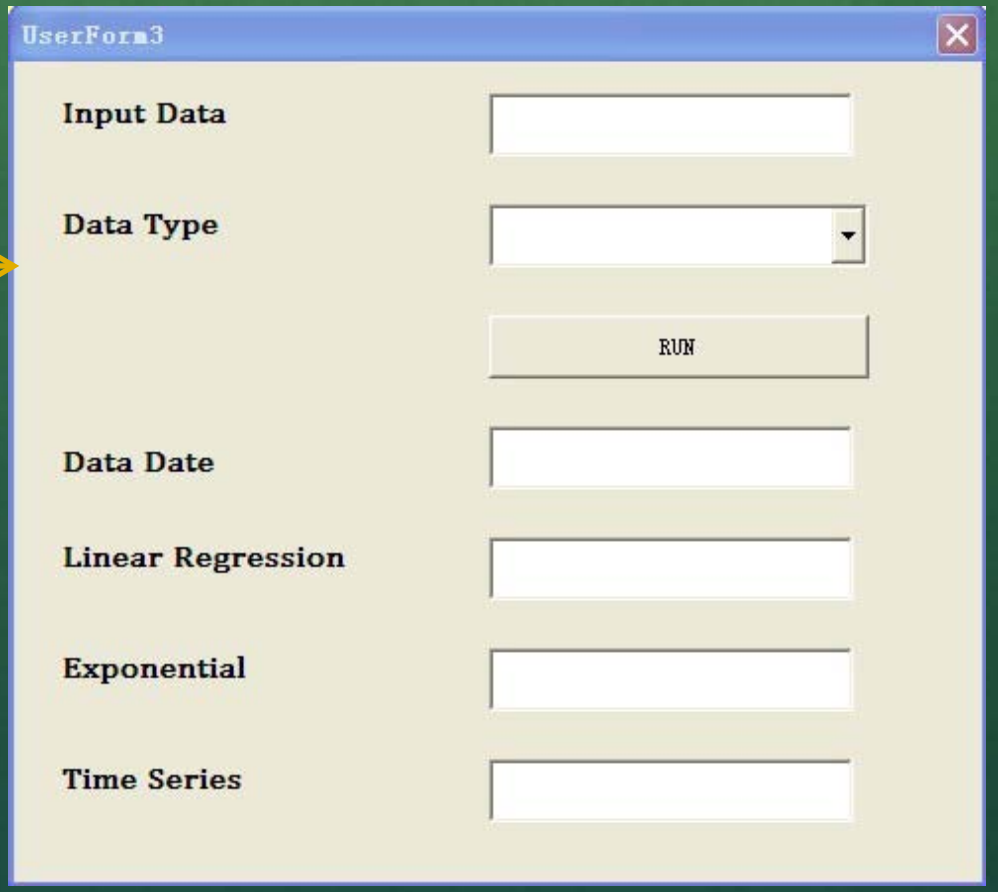
Input Data: B*:B* (the cell location should be "Capital" letter)

Then choose Data Type: Annual or Monthly

Then click "RUN"

ENTER

Click → ENTER →



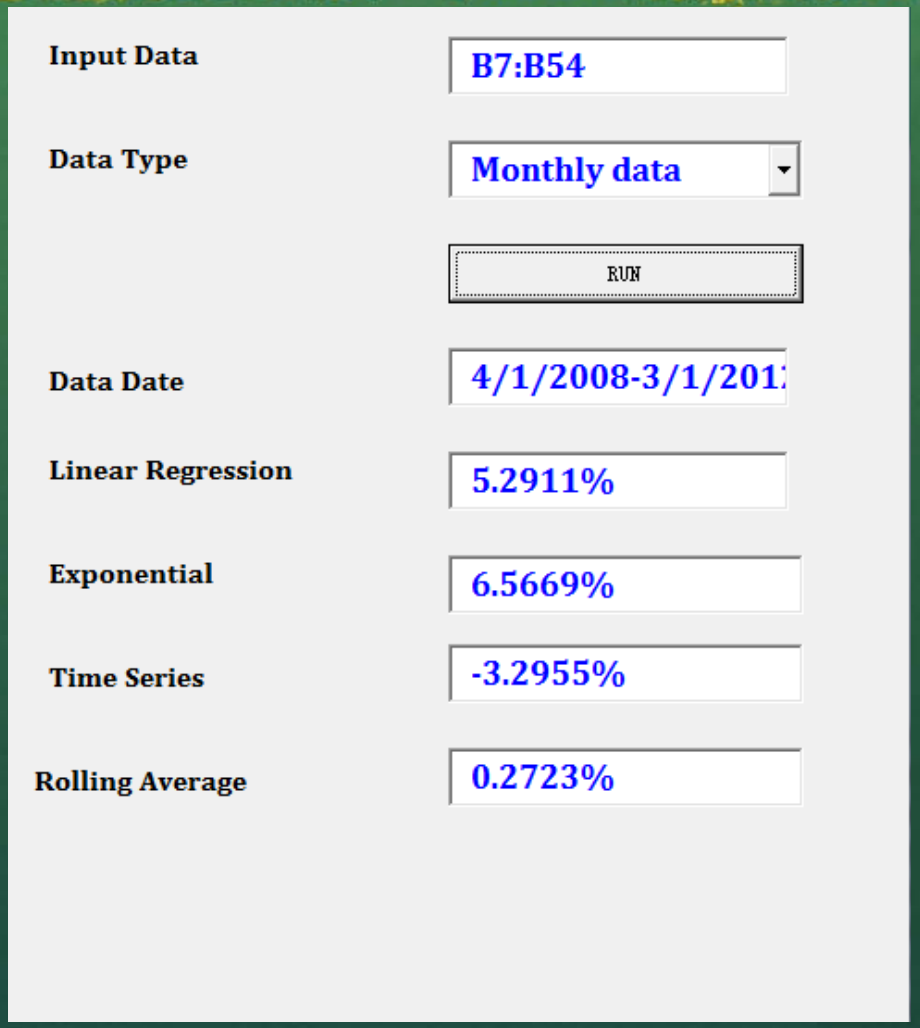
Using Monthly Data

Here, we give an example consisting of one company's data from Tennessee.

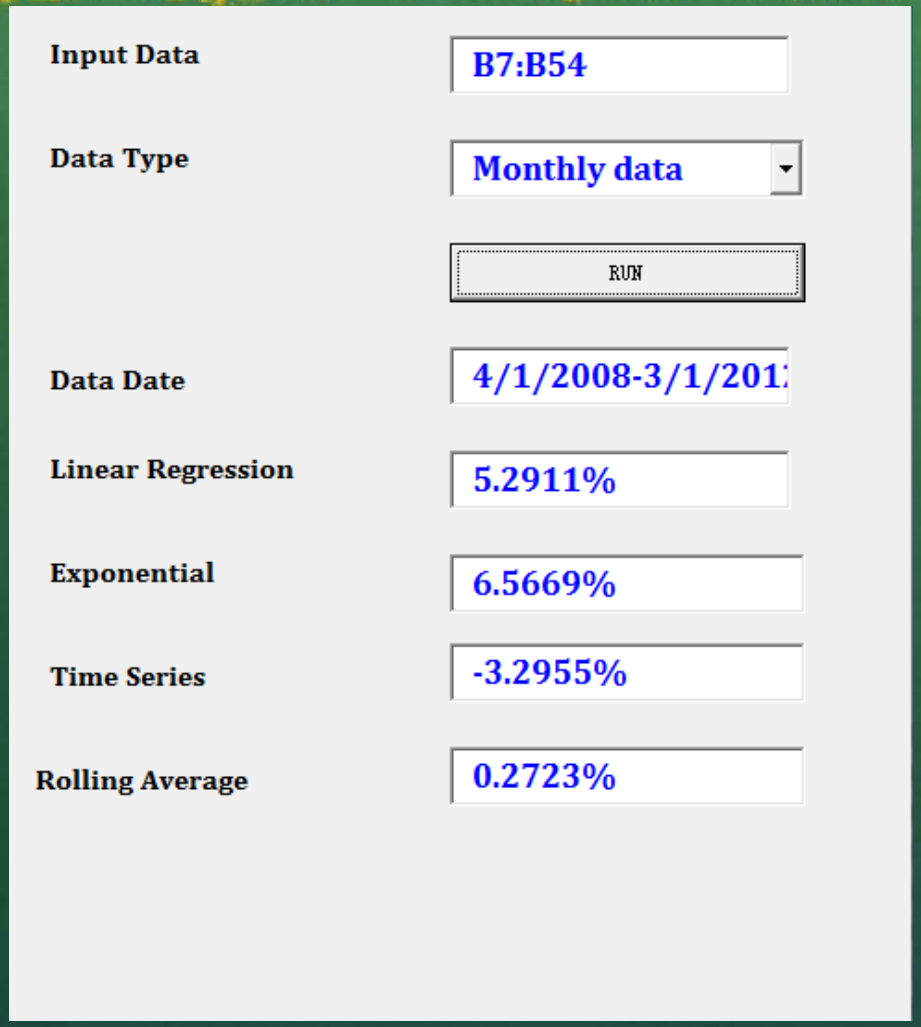
Apr-08	217.88
May-08	222.82
Jun-08	215.54
Jul-08	228.73
Aug-08	229.57
Sep-08	227.88
Oct-08	245.84
Nov-08	214.77
Dec-08	250.94
Jan-09	192.55
Feb-09	198.28
Mar-09	223.90
Apr-09	230.16
May-09	221.35
Jun-09	230.37
Jul-09	238.54
Aug-09	241.46
Sep-09	245.94
Oct-09	254.36
Nov-09	246.69
Dec-09	277.80
Jan-10	194.58
Feb-10	206.14
Mar-10	241.46
Apr-10	233.62
May-10	221.23
Jun-10	240.26
Jul-10	238.43
Aug-10	252.59
Sep-10	249.83
Oct-10	254.83
Nov-10	259.98
Dec-10	277.44
Jan-11	194.56
Feb-11	203.95
Mar-11	238.57
Apr-11	228.01
May-11	243.45
Jun-11	247.65
Jul-11	247.54
Aug-11	263.72
Sep-11	250.39
Oct-11	258.48
Nov-11	259.89
Dec-11	278.06
Jan-12	201.58
Feb-12	225.06
Mar-12	238.83

Monthly Data

Year	PMPM
Apr-08	217.88
May-08	222.82
Jun-08	215.54
Jul-08	228.73
Aug-08	229.57
Sep-08	227.88
Oct-08	245.84
Nov-08	214.77
Dec-08	250.94
Jan-09	192.55
Feb-09	198.28
Mar-09	223.90
Apr-09	230.16
May-09	221.35
Jun-09	230.37



Comparison



Linear Regression	5.29%
Exponential Regression	6.57%
Time Series	-3.30%
Rolling Average	0.27%

Any Questions?