

Comparison of Forecasting Methods for Influenza-like-Illness Surveillance

Marlena Gehret Plagianos, Nicholas Soulakis, Neil Calman



NEW YORK CITY DEPARTMENT of HEALTH and MENTAL HYGIENE



Background

- Prior method for influenza-like-illness (ILI) prediction and surveillance:
 - 7 (week)day moving average (MovAvg7)
- Strengths:
 - Simple; short baseline
 - Best method for detection of gradual outbreaks and good detection of spike outbreaks (Murphy & Burkom, 2008)
- Weakness:
 - No seasonal adjustments
 - Autocorrelation of residuals

Objective

- Could current method of ILI surveillance be improved with a time-series or regression model?
- Optimal model:
 - Good overall fit
 - Randomly distributed residuals
 - Assumption for statistical inference (signal detection)
 - Positive autocorrelation of residuals leads to underestimation of standard error

Data from Institute for Family Health

- 9 outpatient clinics in NYC
- July 2004 through May 2008 used to fit models
- Daily data; strong day of week variation
 - No Sunday visits, high variability on Saturdays
- ILI: % of clinical visits where the patient had a fever (measured temperature $>99.9^{\circ}\text{F}$ or reason for visit) plus a respiratory-related reason for visit or diagnosis.

Time Series Models Evaluated

Stepwise Autoregressive: combined trend and autoregressive model

$$ILI(t) = b_0 + b_1 t + u_t \quad u_t = a_1 u_{t-1} + a_2 u_{t-2} + \dots + a_m u_{t-m} + \varepsilon_t$$

Exponential Smoothing: more weight (w) given to more recent data

$$s_{t+1} = w_t (ILI_t) + (1-w_t) s_{t-1} \quad s_t = \text{smoothed value} = \text{forecasted ILI at } t+1$$

Winters: Exponential smoothing with seasonal parameters for day of week and month, $s(t)$

Additive: $ILI_{(t)} = (a_t) + s(t) + \varepsilon_t$

Multiplicative: $ILI_{(t)} = (a_t) s(t) + \varepsilon_t$

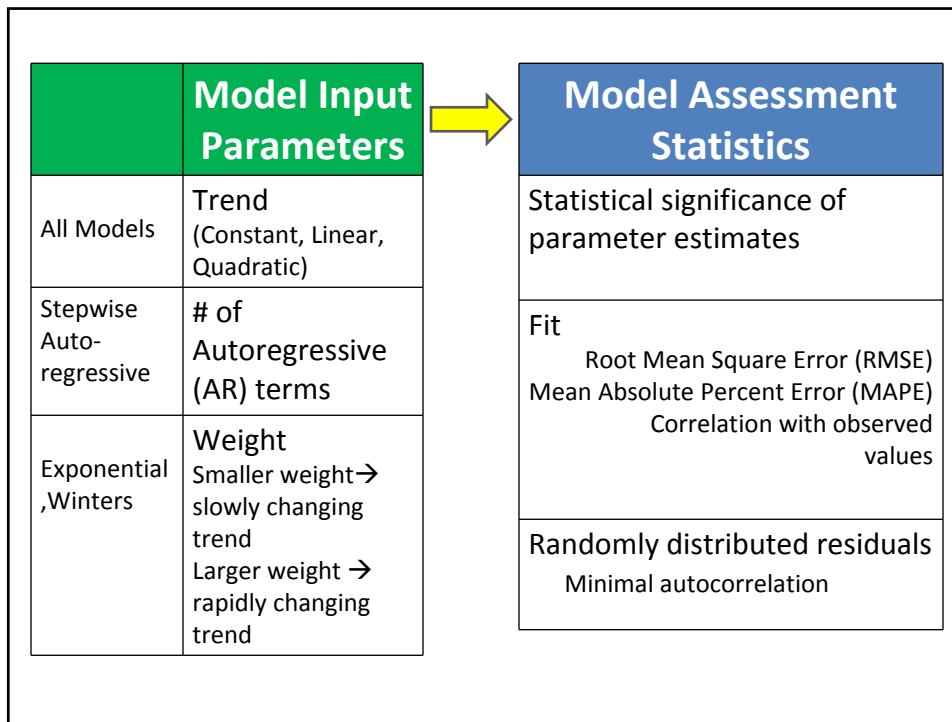
Regression Models Evaluated

Cyclical model: adjusting for day of week and week of year

$$ILI(t) = a + b_1 t + \beta_j (\text{day of week}=j) \\ + b_2 \sin(2\pi/52) \text{ week} + b_3 \cos(2\pi/52) \text{ week}$$

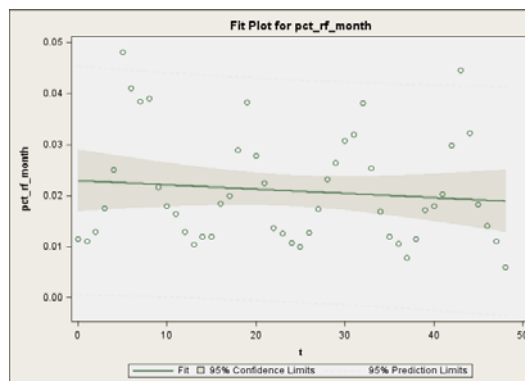
Cyclical model (AR7): adjusting for day of week and week of year, with 7 autoregressive (AR) components

$$ILI(t) = a + b_2 t + \beta_j (\text{day of week}=j) \\ + b_3 \sin(2\pi/52) \text{ week} + b_4 \cos(2\pi/52) \text{ week} \\ + b_1 ILI_{(t-1)} + b_2 ILI_{(t-2)} + \dots + b_7 ILI_{(t-7)}$$



What type of trend is appropriate?

- Linear?
monthly % ILI = $\alpha + \beta(t)$
 $\beta = -.000084$, $p = 0.445$
- Quadratic?
monthly % ILI = $\alpha + \beta(t^2)$
 $\beta = -0.000042$, $p = 0.445$



→ **Constant, No Trend**

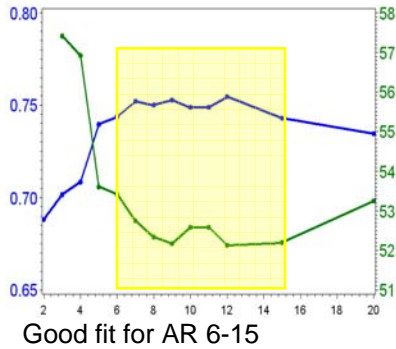
t = # of months since 7/2004

Stepwise Autoregressive

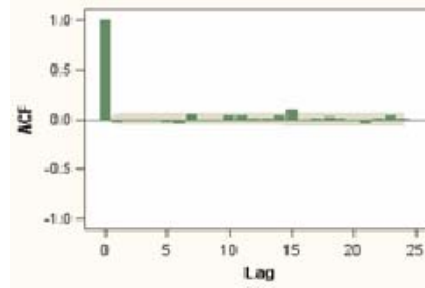
Selection of # of AR terms
Values tested: 2-15, 20, 24

Fit

Relationship between **Correlation**, **MAPE** and # of AR Terms



Autocorrelation of Residuals



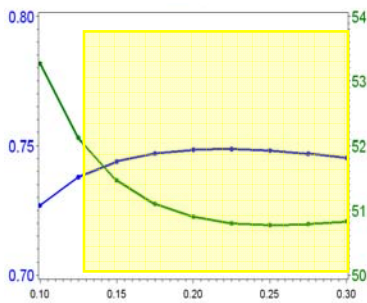
AR=7
autocorrelation decays and is not statistically significant up to 24 days

Exponential Smoothing Model

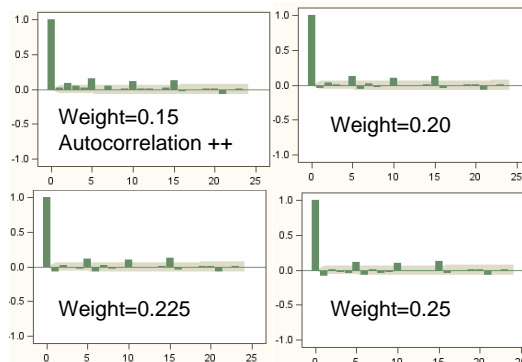
Selection of Weights
Values tested: 0.05 – 0.30

Fit

Relationship between **Correlation**, **MAPE** and Weighting Constant



Autocorrelation of Residuals



Autocorrelation statistically significant for all weights, but minimal when **weight=0.20**

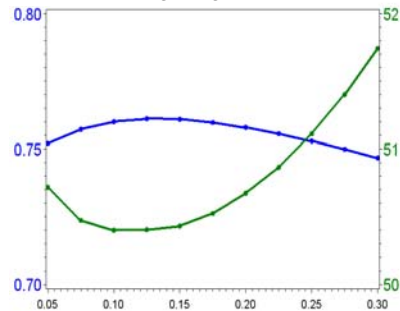
Additive Winters Model

Selection of Weights

Values tested: 0.05 – 0.30

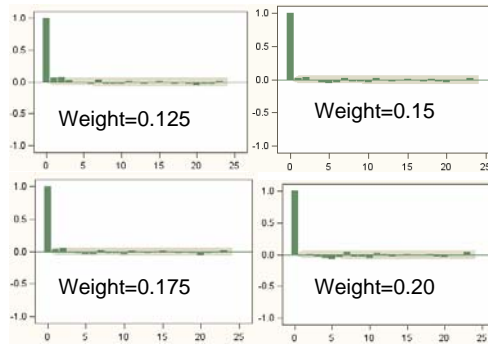
Fit

Relationship between Correlation, MAPE and Weighting Constant



Good fit for all weights

Autocorrelation of Residuals



Autocorrelation not statistically significant for weights=0.125-0.20 and minimal when **weight=0.175**

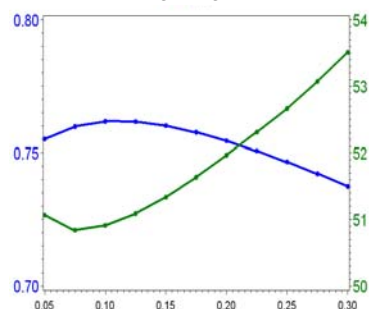
Multiplicative Winters Model

Selection of Weights

Values tested: 0.05 – 0.30

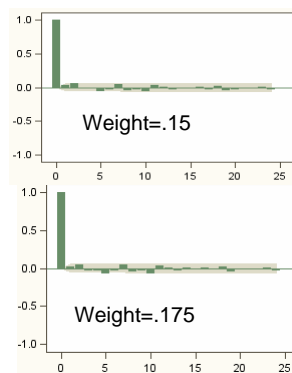
Fit

Relationship between Correlation, MAPE and Weighting Constant



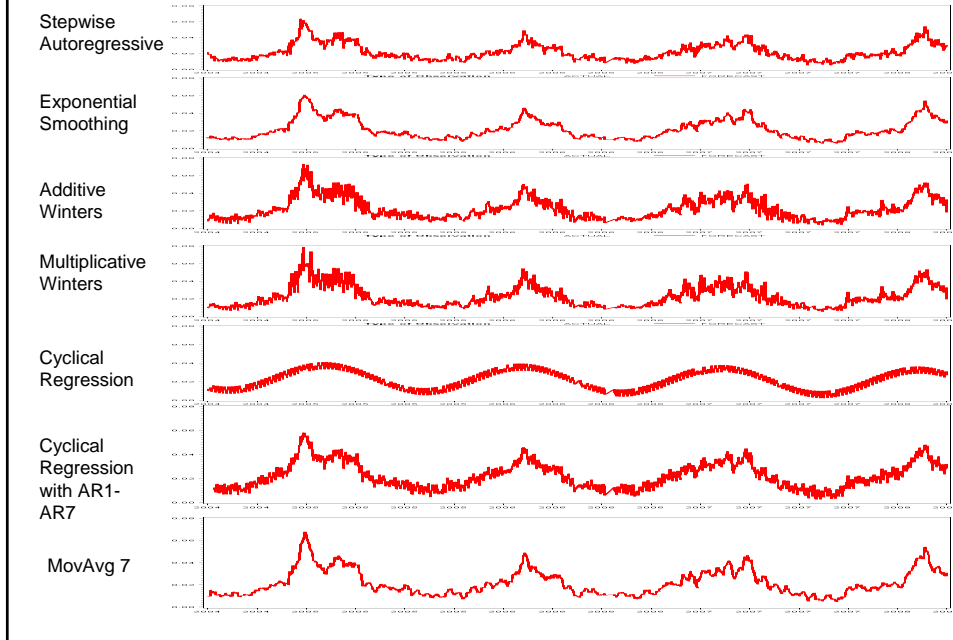
Good fit for all weights

Autocorrelation of Residuals

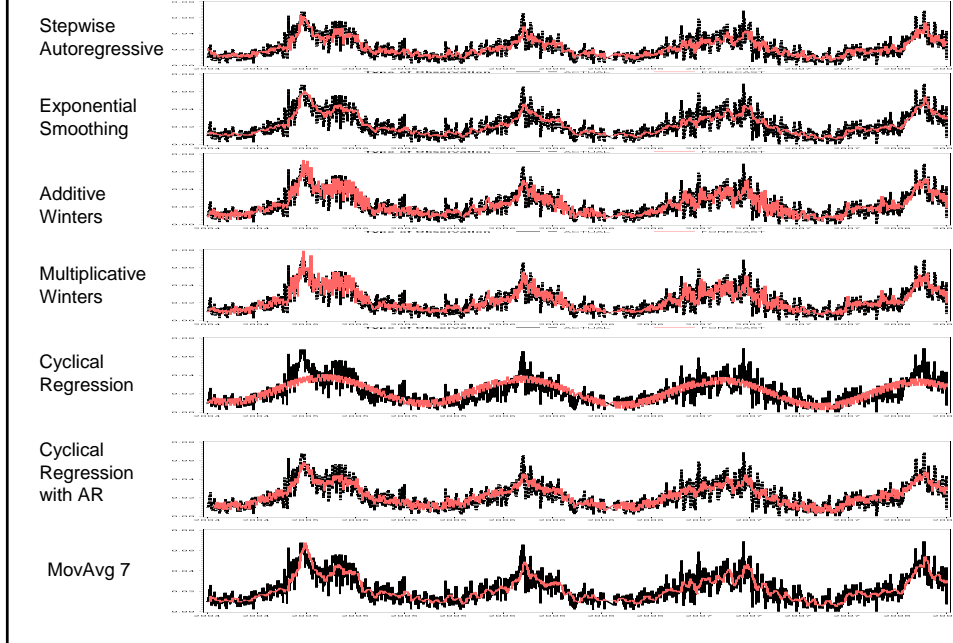


Autocorrelation not statistically significant for weights=0.15, 0.175; and minimal when **weight=0.15**

Forecasts by Model Type 2004-2008



Observed and Forecasts 2004-2008

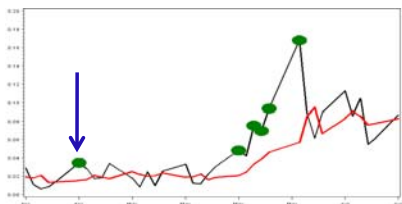


Results

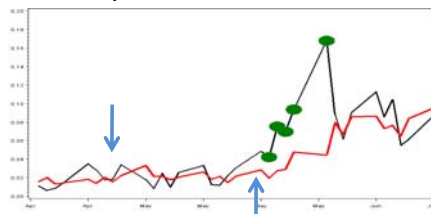
	Model	RMSE	MAPE	Correlation	Autocorrelation of residuals
↑	Regression with AR1-AR7	0.0086	48.3	0.79	Not significant
↑	Winters (Additive)	0.0091	50.5	0.76	Not significant
↑	Winters (Multiplicative)	0.0091	51.3	0.76	Not significant
↑	Stepwise Autoregressive	0.0091	52.8	0.75	Not significant
↔	MovAvg7	0.0832	51.5	0.75	Significant, decay, + weekly
↔	Exponential	0.0092	50.9	0.75	Significant, decay, + weekly
↓	Regression	0.0096	51.9	0.72	Significant +++

Observed, Expected and Signals H1N1 Outbreak, Spring 2009

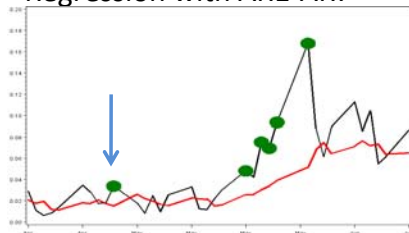
Stepwise Autoregressive



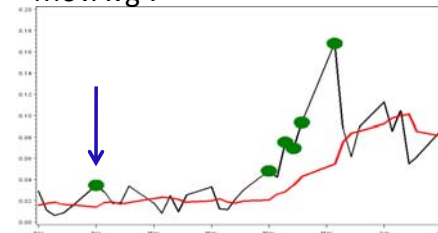
Multiplicative Winters



Regression with AR1-AR7

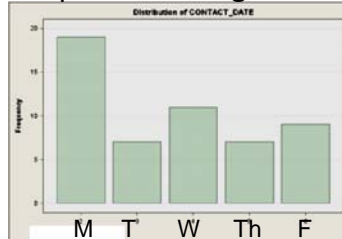


MovAvg 7

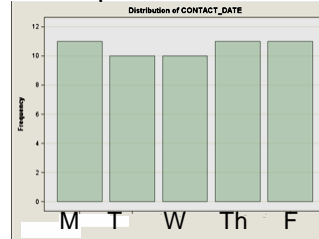


Distribution of Signals by Day of Week

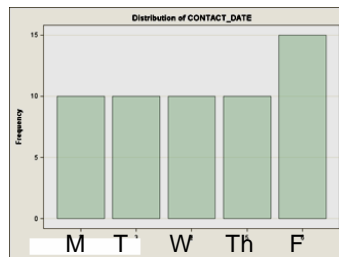
Stepwise Autoregressive



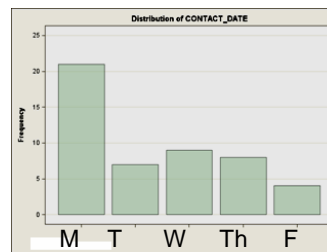
Multiplicative Winters



Regression with AR1-AR7



MovAvg 7



Summary

Model	Overall Fit	Baseline	Adjust for day of week?	Conclusions
Stepwise Autoregressive	👍	Short (>7 days)	No	Best with short baseline
Winters (Multiplicative)	👍	Long (>1 year)	Yes	Best if long baseline available. Reassess weighting constants with simulated outbreaks
MovAvg7	👍	Short (>7 days)	No	Signals affected by no day of week adjustment and autocorrelation of residuals. Reassess sens/spec post simulation.
Regression with AR1-AR7	👍	Long (>1 year)	Yes	May be over-fit to training data, does not adjust to new trends