

# Flu Forecasting Using SARIMA/X

...

March 10, 2017

# CDC Influenza Surveillance Data Sets

The U.S. influenza surveillance system is a collaborative effort between CDC and its many partners in state, local, and territorial health departments.

Virologic Surveillance — Approximately 110 U.S. World Health Organization (WHO) Collaborating Laboratories and 240 National Respiratory and Enteric Virus Surveillance System (NREVSS) laboratories.

Outpatient Illness Surveillance — Information on patient visits to health care providers for influenza-like illness is collected through the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet). ILINet consists of more than 2,900 enrolled outpatient healthcare providers in all 50 states.

# Understanding the problem

## Item 1

Traditional  
surveillance  
methods lag 1-2  
weeks

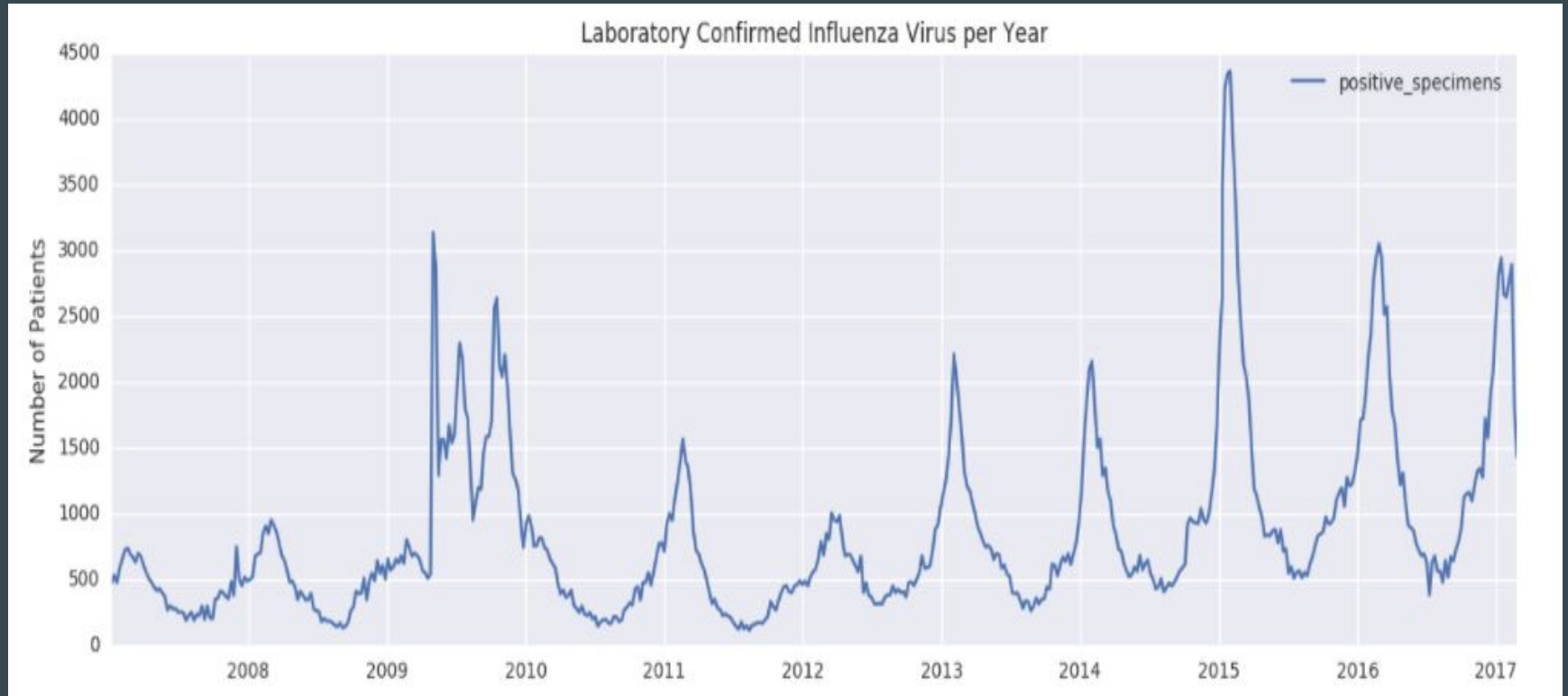
## Item 2

Reports  
represents a  
limited syndromic  
observation of  
people who seek  
medical attention

## **Project objective:**

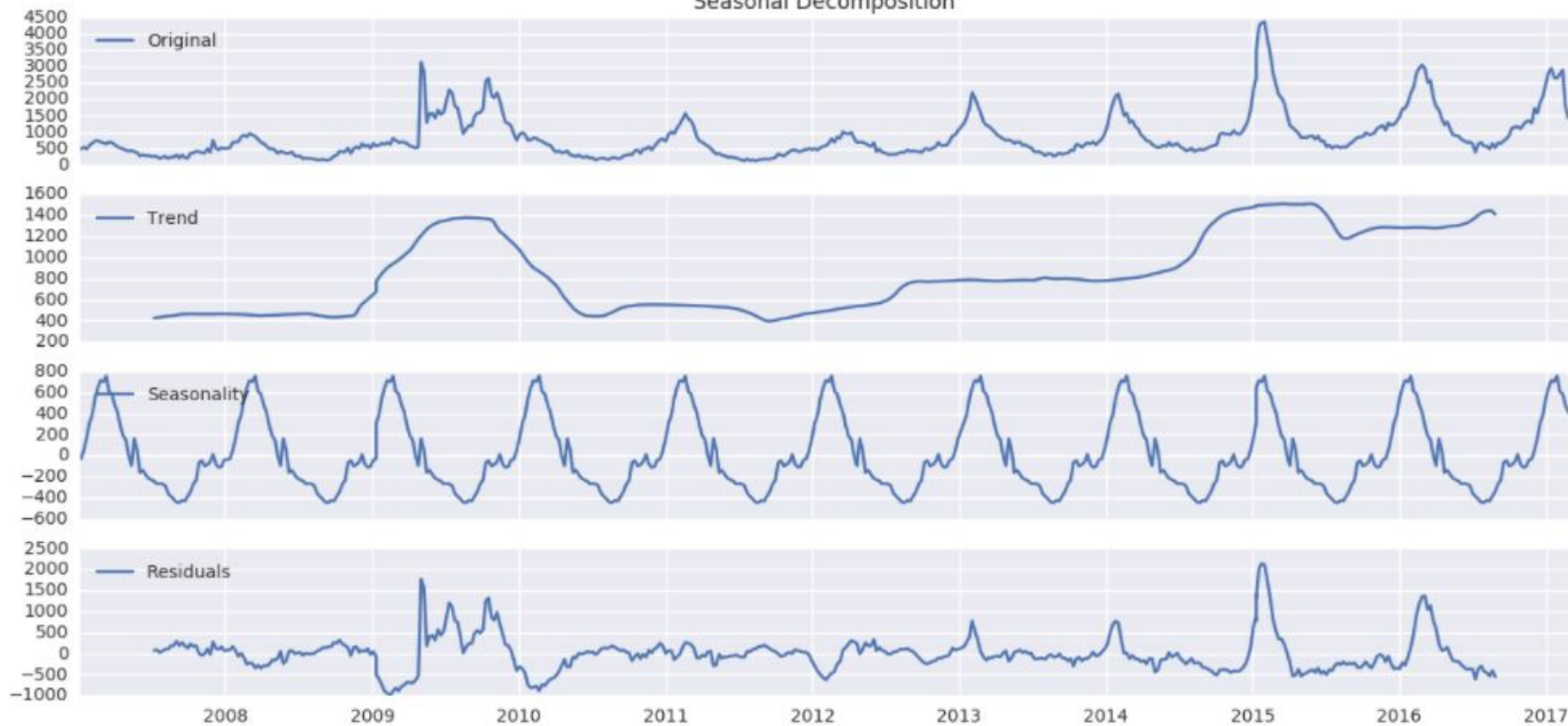
Build real time algorithm that combines flu sentiment, population density, incoming travelers, etc to help predict where, when and how severe the influenza outbreak will be.

# CBC Lab Confirmed Influenza Virus per Year

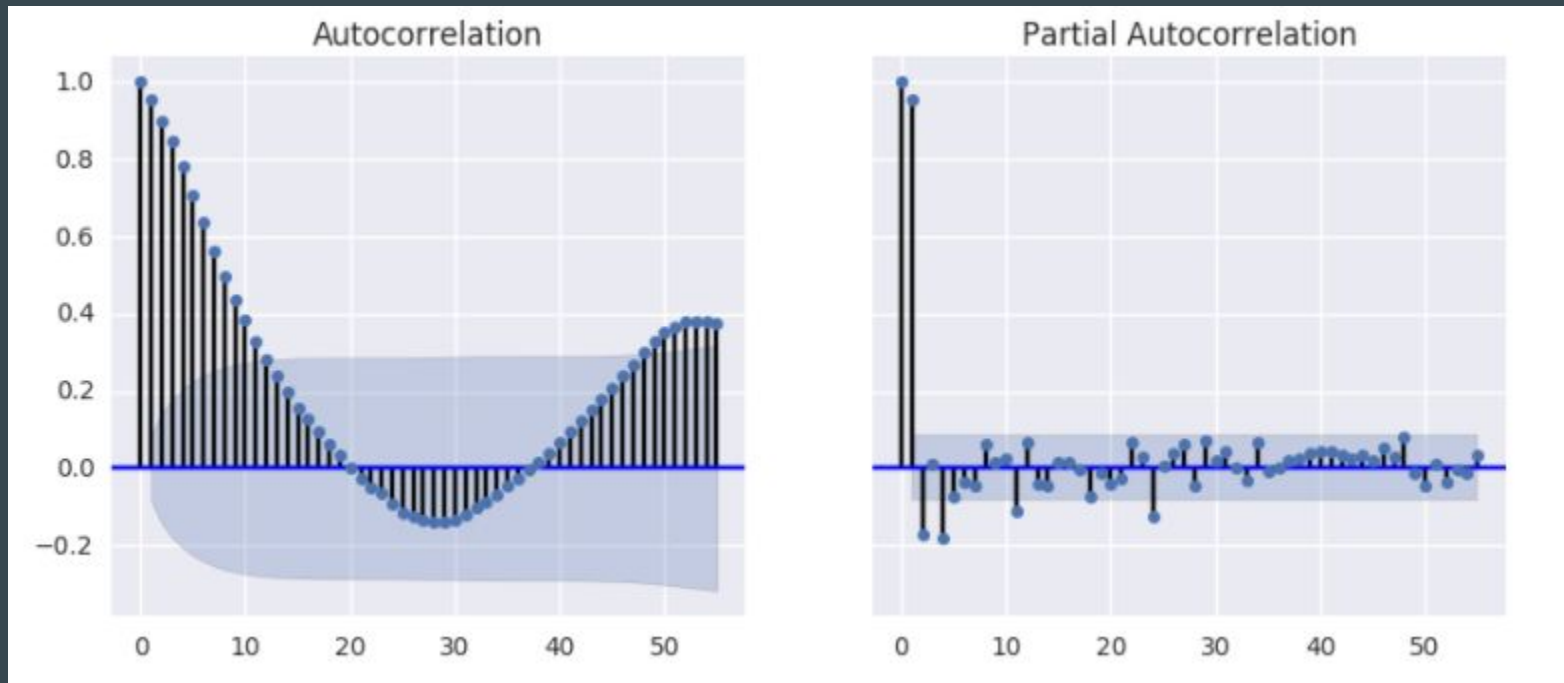


# Seasonal Decomposition

Seasonal Decomposition

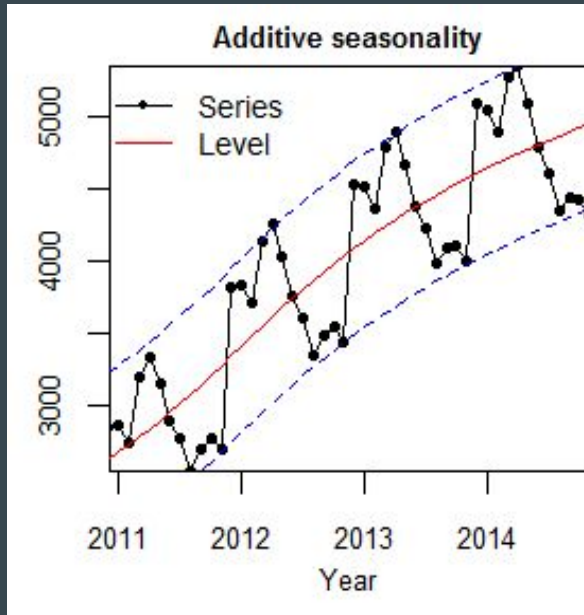


# Interpreting Seasonality with ACF and PACF

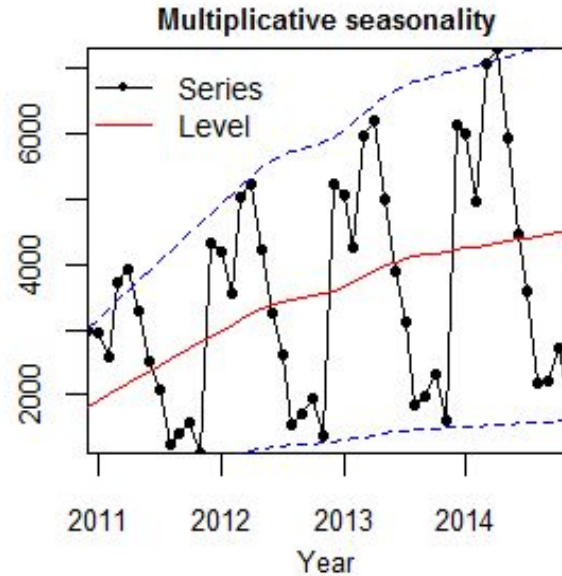


# Additive vs Multiplicative Seasonality

ARIMA(1,1,4)



ARIMA(2,1,0) x (1,1,0,12) model





# Additive SARIMAX Model

```
ar = 1
ma = (1,1,1,0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,0,0,0,1)
diff = sm.tsa.statespace.SARIMAX(df['total_specimens'].values, freq='W', order=(ar,1,ma))
diff_res = diff.fit(dispatch=False)
print(diff_res.summary())
```

## Statespace Model Results

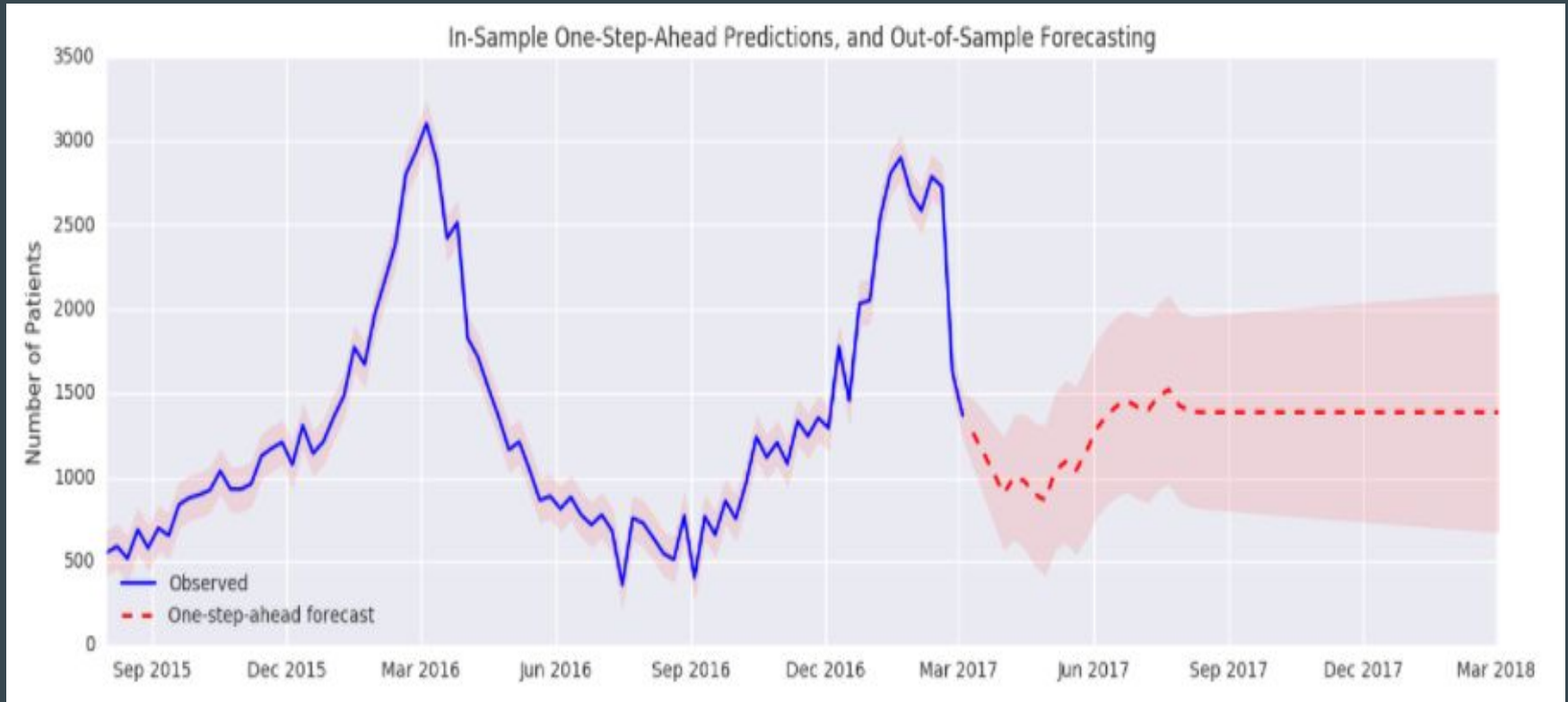
```
=====
Dep. Variable:          y      No. Observations:          534
Model:          SARIMAX(1, 1, (1, 2, 3, 7, 11, 14, 23))  Log Likelihood          -3561.523
Date:              Fri, 10 Mar 2017      AIC              7141.046
Time:              07:01:39              BIC              7179.570
Sample:            0                    HQIC             7156.120
                                     - 534
```

```
Covariance Type:          opg
```

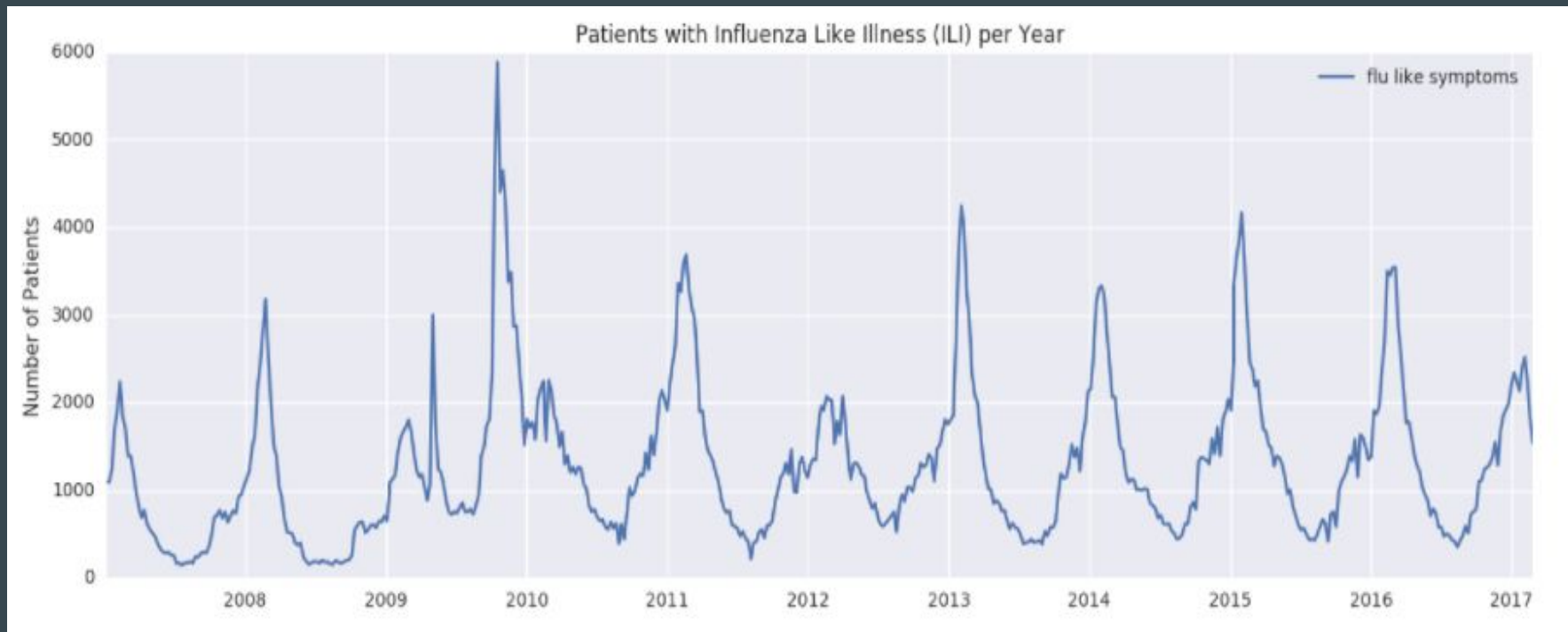
```
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          0.6267      0.096      6.542      0.000      0.439      0.814
ma.L1         -0.4874      0.097     -5.032      0.000     -0.677     -0.298
ma.L2         -0.1737      0.018     -9.815      0.000     -0.208     -0.139
ma.L3          0.1377      0.031      4.435      0.000      0.077      0.199
ma.L7         -0.1918      0.029     -6.653      0.000     -0.248     -0.135
ma.L11        -0.1638      0.046     -3.538      0.000     -0.254     -0.073
ma.L14        -0.1351      0.062     -2.181      0.029     -0.256     -0.014
ma.L23         0.0851      0.026      3.329      0.001      0.035      0.135
sigma2       3.713e+04    426.486     87.067      0.000    3.63e+04    3.8e+04
=====
```

```
=====
Ljung-Box (Q):          30.49      Jarque-Bera (JB):          95499.29
Prob(Q):              0.86      Prob(JB):              0.00
Heteroskedasticity (H): 0.61      Skew:              4.26
Prob(H) (two-sided):    0.00      Kurtosis:           68.02
=====
```

# In-Sample + Out-of-Sample Flu Forecasting



# Explanatory Variable: Patients with Flu Like Symptoms



# SARIMAX Model with Positive Specimen + ILI

```
# Fit the model with exogenous data
mod = sm.tsa.statespace.SARIMAX(endog, exog, order=(1,1,4))
res = mod.fit(dispatch=False)
print(res.summary())
```

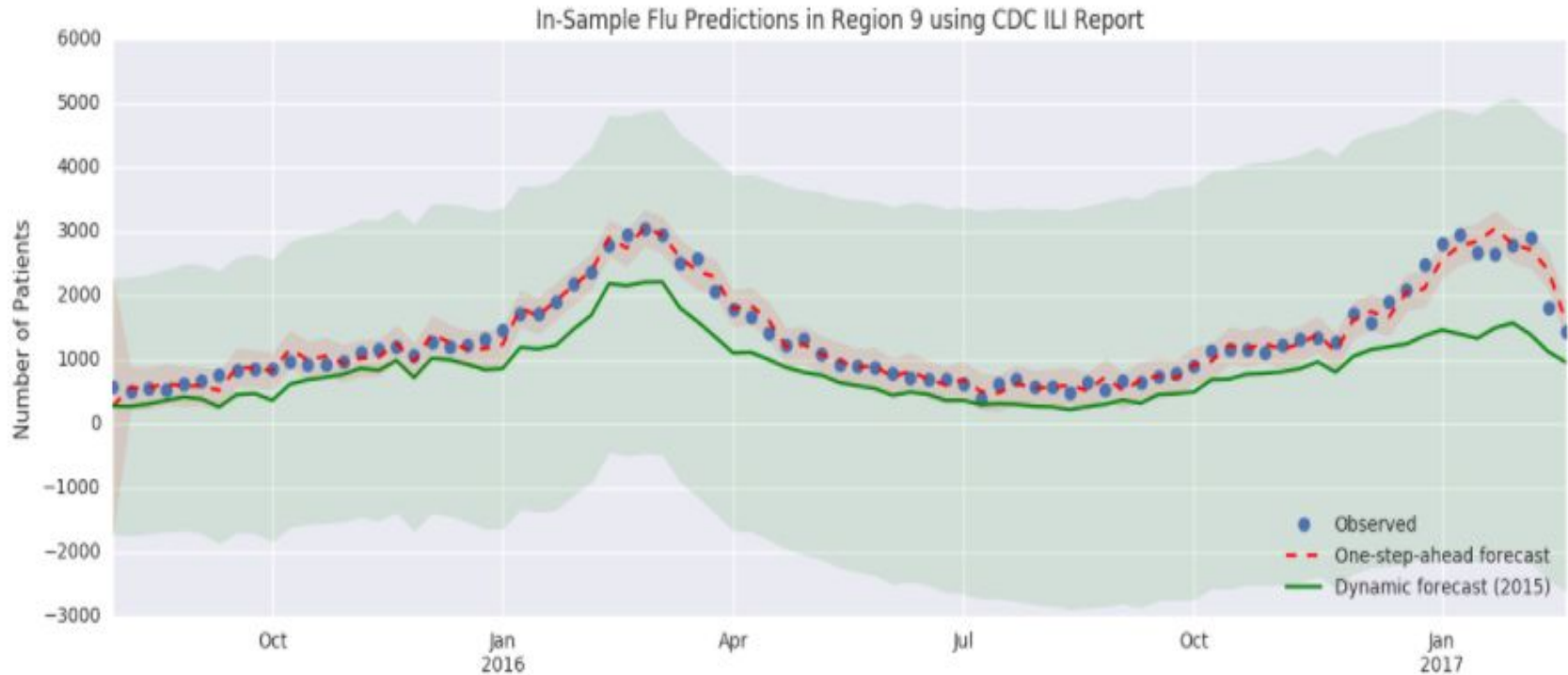
## Statespace Model Results

```
=====
Dep. Variable:          total_specimens      No. Observations:          83
Model:                  SARIMAX(1, 1, 4)      Log Likelihood              -524.472
Date:                   Fri, 10 Mar 2017      AIC                        1064.943
Time:                   02:11:24              BIC                        1084.294
Sample:                 08-01-2015            HQIC                       1072.717
                        - 02-25-2017
Covariance Type:                opg
=====
```

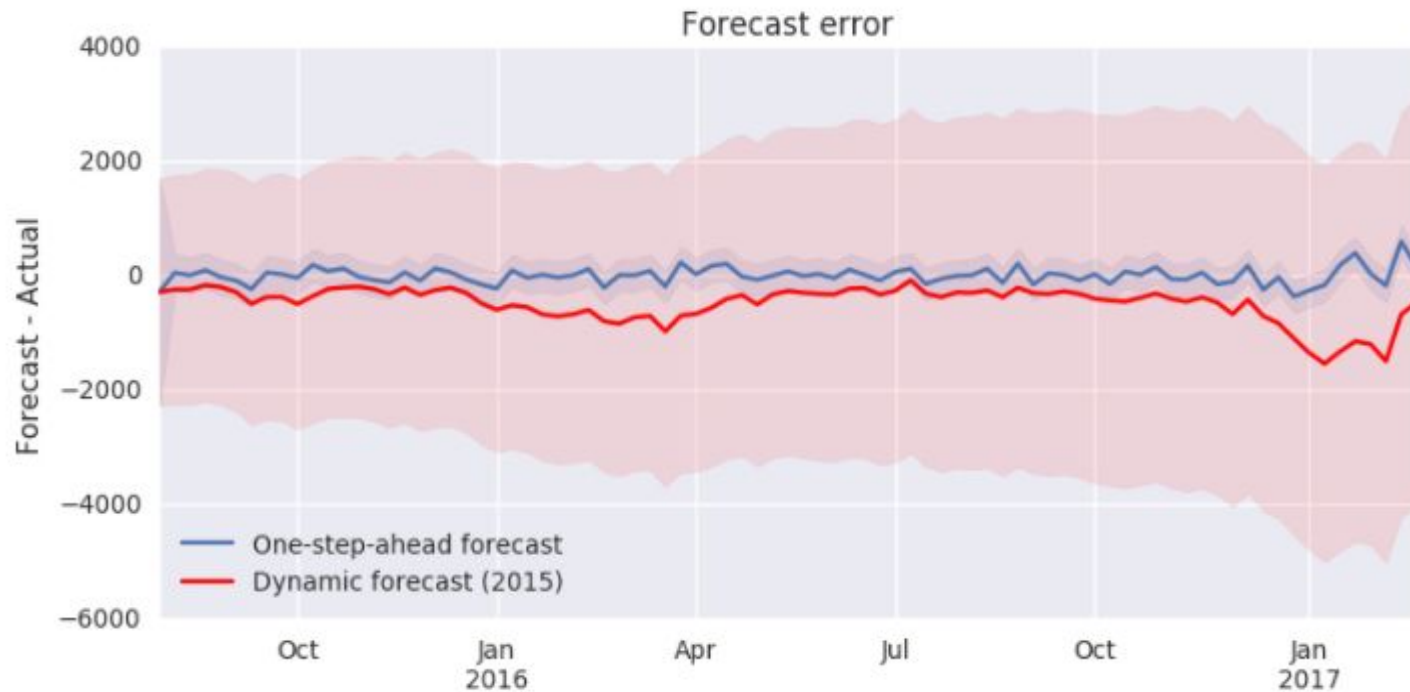
|          | coef      | std err  | z         | P> z  | [0.025    | 0.975]   |
|----------|-----------|----------|-----------|-------|-----------|----------|
| const    | -0.0068   | 2.29e+04 | -2.98e-07 | 1.000 | -4.48e+04 | 4.48e+04 |
| ilitotal | 0.6234    | 0.078    | 7.965     | 0.000 | 0.470     | 0.777    |
| ar.L1    | -0.7157   | 0.153    | -4.689    | 0.000 | -1.015    | -0.417   |
| ma.L1    | 0.6484    | 0.194    | 3.340     | 0.001 | 0.268     | 1.029    |
| ma.L2    | -0.2803   | 0.145    | -1.930    | 0.054 | -0.565    | 0.004    |
| ma.L3    | 0.1255    | 0.104    | 1.209     | 0.227 | -0.078    | 0.329    |
| ma.L4    | 0.5063    | 0.076    | 6.694     | 0.000 | 0.358     | 0.655    |
| sigma2   | 2.065e+04 | 3249.891 | 6.354     | 0.000 | 1.43e+04  | 2.7e+04  |

```
=====
Ljung-Box (Q):          16.83      Jarque-Bera (JB):          33.99
Prob(Q):                1.00      Prob(JB):                  0.00
Heteroskedasticity (H):  4.10      Skew:                      -0.72
Prob(H) (two-sided):    0.00      Kurtosis:                   5.80
=====
```

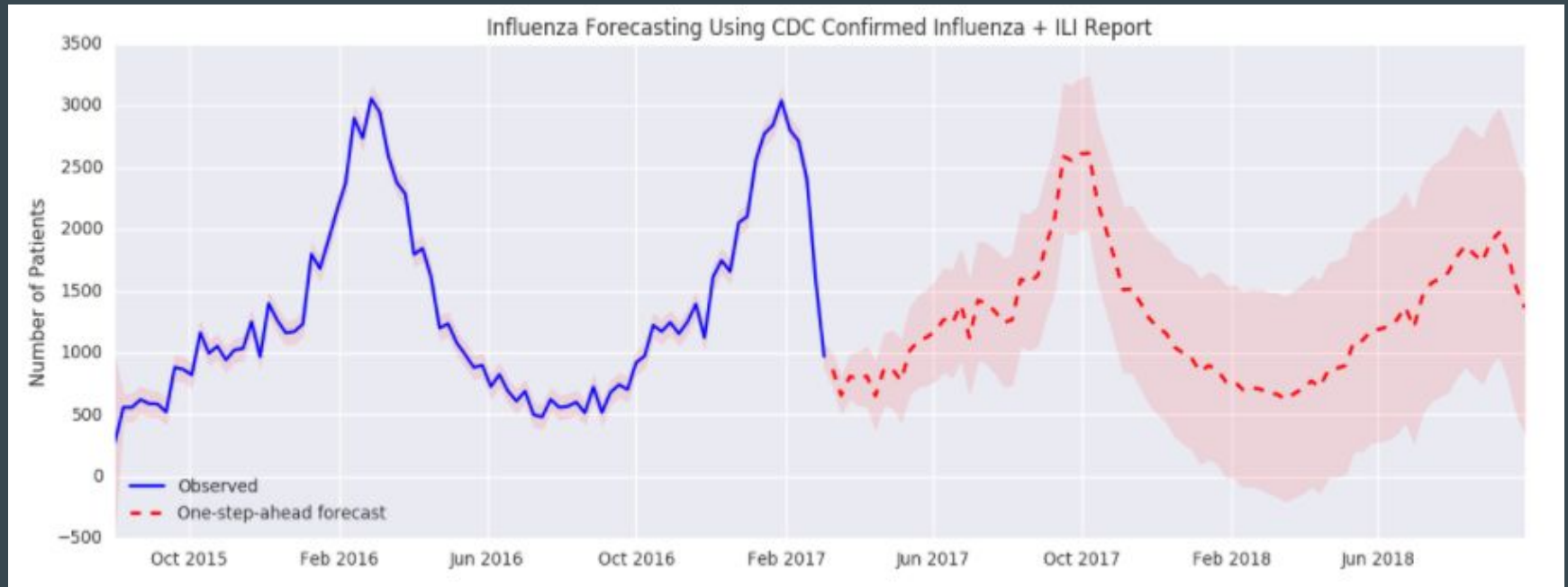
# In-Sample Flu Prediction in Region 9



# Forecast Error



# Out-of-Sample Flu Prediction with ILI Data



# What's Next

Real Time Tweet

- Real Time Tweet Analysis
- 

GeoSpacial

- 
- 

Time Series w sklearn

- 
- 

Flask/Exposed API

- 
-