Sentiment Data Analysis

State Space Models

Radha Pendyala

REFINITIV™

State Space Models

State Space Models

- For each asset, relevance score weighted sentiment probabilites are computed at a daily scale
- Net sentiment score is computed as follows:

$$Prob_{net} = (Prob_{positive} - Prob_{negative}) \times (1 - Prob_{neutral})$$

- Kalman Filter technique is used to find the state evolution of positive, negative and neutral sentiment.
- A *Local News Sentiment Level* state space model is constructed for the unobserved state(μ_t) and the observed signal (y_t)

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned}$$

- State space models of varying complexity can be fit to the data and then the relevant filtered data can be used to input to various Machine Learning models

Local Level Model

Local Level Model

- All the state space models have two main equations. One is the observed value equation and the other is the state equation.
- Key idea is to infer the distribution of state variable given the observed variable
- A *Local Level* state space model is constructed for the unobserved state(μ_t) and the observed signal (y_t)

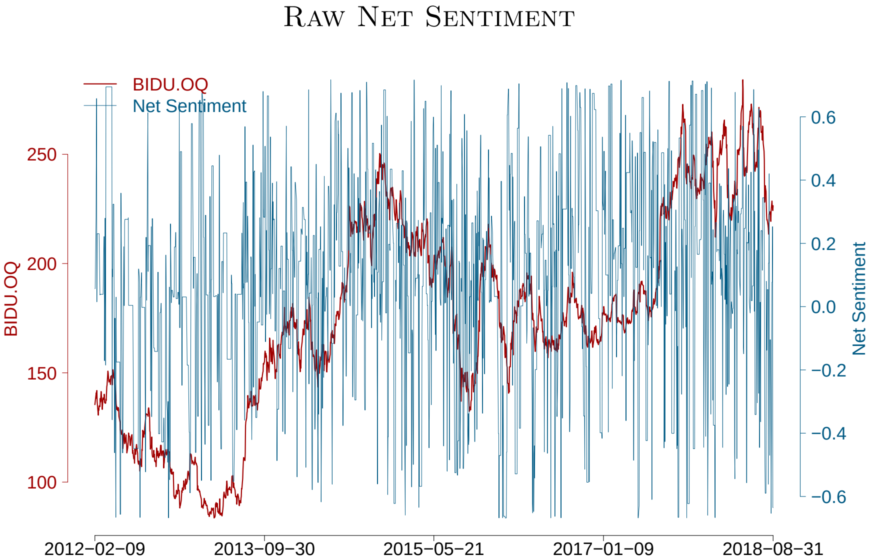
$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned}$$

- What are the main mathematical objects in the model ?
  Let Y_{t-1} be the vector of observations $(y_1, \ldots, y_{t-1})$

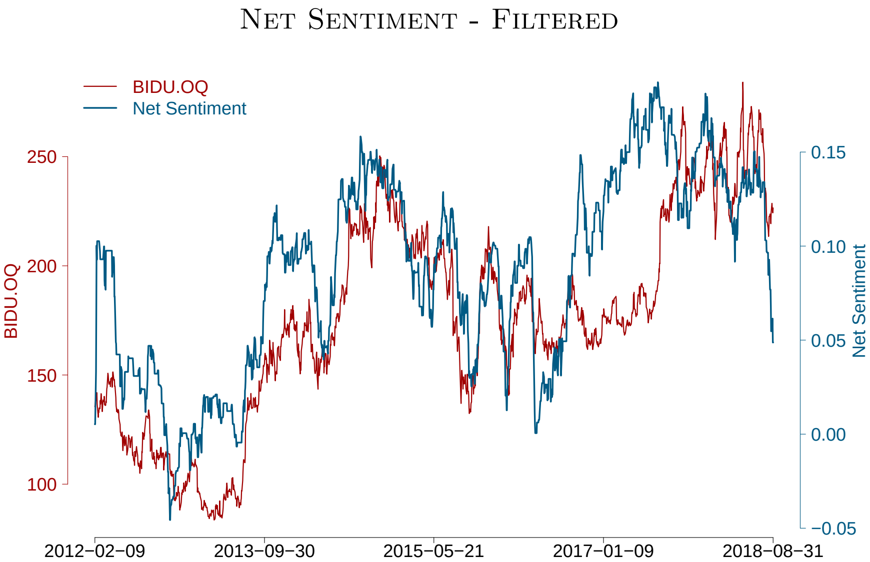
$$\begin{aligned} \mu_t | Y_{t-1} &\sim N(a_t, P_t) \\ \mu_t | Y_t &\sim N(a_{t|t}, P_{t|t}) \\ \mu_{t+1} | Y_t &\sim N(a_{t+1}, P_{t+1}) \\ v_t | Y_{t-1} &\sim N(0, P_t + \sigma_\epsilon^2) \\ v_t | \mu_t, Y_{t-1} &\sim N(\mu_t + \epsilon_t - a_t | \mu_t, Y_{t-1}) \end{aligned}$$

  - $a_{t|t}$ is the filtered estimator
  - a_{t+1} is the one step ahead predictor of μ_{t+1}
  - v_t is the one-step ahead prediction error

Sentiment and Price overlay

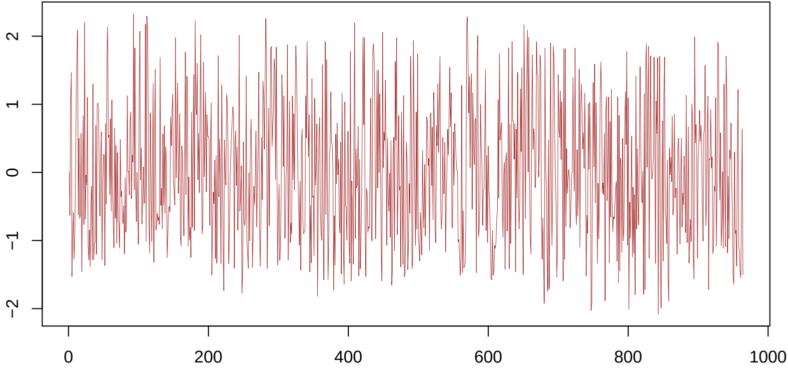


Local Level Model



Local Level Model

- Residuals



- AIC = 1990.5454107
- Box-Ljung Statistic = 0.2778486 for 10 lags
- Shapiro's test for normality(p value): 0

State Space Models - Varying complexity

State Space Models - Varying complexity

- A *Local Level Linear Trend* state space model (Deterministic Slope)

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_1 + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned}$$

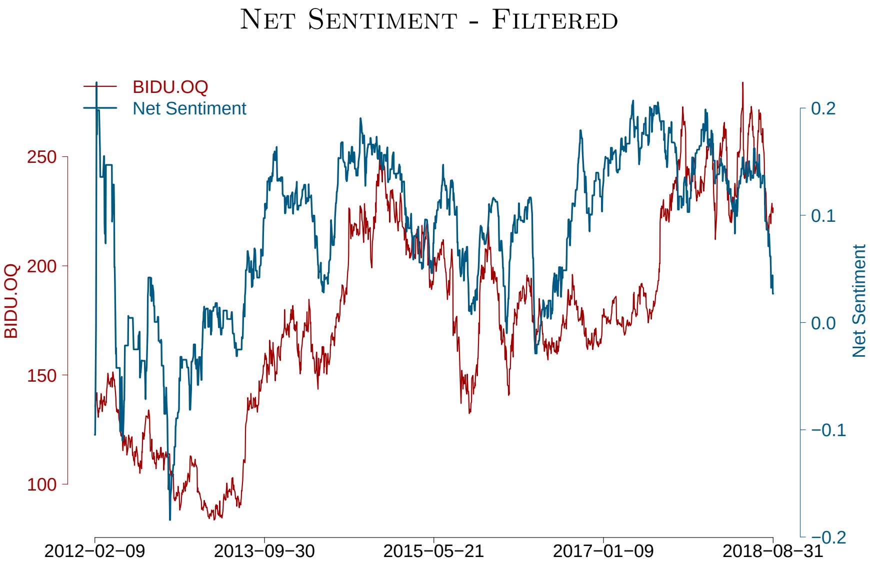
- A *Local Level Linear Trend* state space model (Stochastic Slope)

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \\ \nu_{t+1} &= \nu_t + \zeta_t, & \eta_t &\sim N(0, \sigma_\zeta^2) \end{aligned}$$

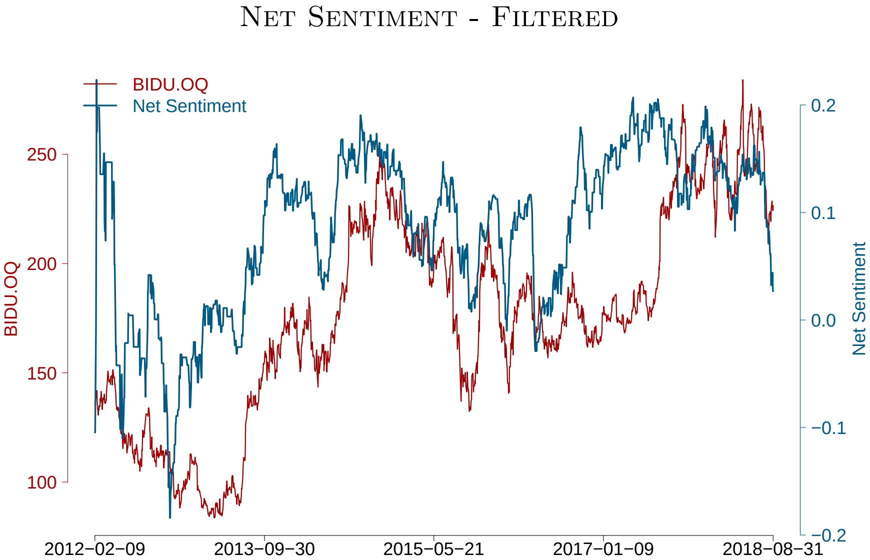
- A *Local Level* and a seasonal model

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \\ \gamma_{1,t+1} &= -\gamma_{1,t} - \gamma_{2,t} - \gamma_{3,t} + \omega_t, & \omega_t &\sim N(0, \sigma_\omega^2) \\ \gamma_{2,t+1} &= \gamma_{1,t} \\ \gamma_{3,t+1} &= \gamma_{2,t} \end{aligned}$$

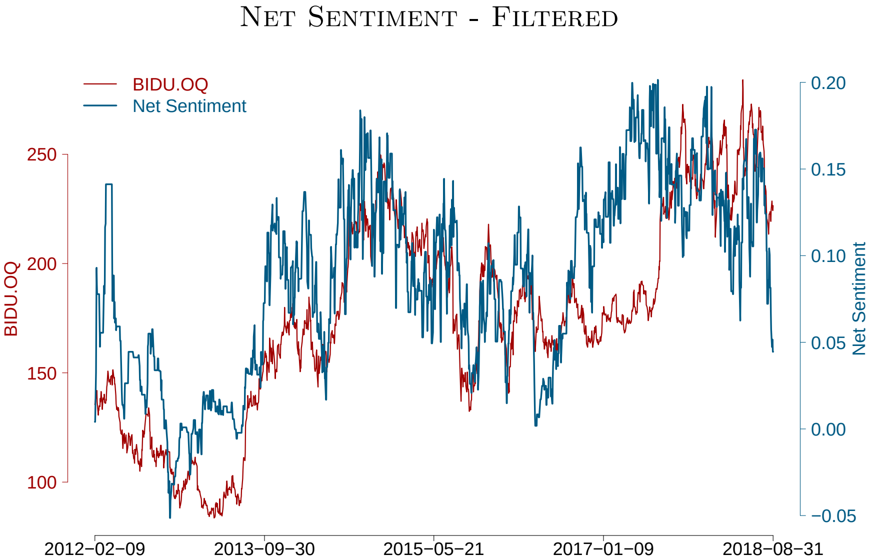
Local Level Linear Trend Model (Deterministic Slope)



Local Level Linear Trend Model (Stochastic Slope)



Local Level Seasonal Model



Diagnostics across Models

Sentiment Type	State Space Model	AIC	Shapiro	Box-Ljung
Positive Probability	Local Level	1990.55	0.00	0.28
Negative Probability	Local Level	1957.49	0.00	0.37
Neutral Probability	Local Level	1988.34	0.00	0.37
Positive Probability	Local Linear Deterministic Slope	1735.69	0.00	0.32
Negative Probability	Local Linear Deterministic Slope	1990.55	0.00	0.28
Neutral Probability	Local Linear Deterministic Slope	1957.49	0.00	0.37
Positive Probability	Local Linear Stochastic slope	1988.34	0.00	0.37
Negative Probability	Local Linear Stochastic slope	1735.69	0.00	0.32
Neutral Probability	Local Linear Stochastic slope	1990.55	0.00	0.28
Positive Probability	Local Level Seasonal	1957.49	0.00	0.37
Negative Probability	Local Level Seasonal	1988.34	0.00	0.37
Neutral Probability	Local Level Seasonal	1735.69	0.00	0.32

Modeling choices to make

- Should one consider additional variables in the state space model ?
- Should there be seasonality consideration for the state space model ?
- Should one fit separate models to various sentiment probabilities ?