RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Predictive Audit: Improving Analytical Procedures Including Weather Indicators

Kyunghee Yoon and Alexander Kogan
Rutgers Business School

Background literature & Research Question

- Analytical procedures (APs) are required at the planning and review phases of an audit (AICPA 1988)
- APs have the power to recover misstatements and irregularities.
- Allen et al. (1998) found that peer stores have a significant predictive power.
- Weather indicators are related to retailers' sales (Starr-McCluer 2000)

RQ1) Can predictive models with peer stores generate more accuracy?

RQ2) Can predictive models with weather indicators generate more accuracy?

Why can peer stores improve predictive powers?

- Peer stores might have similar economic environments (i.e. cities, rural areas).

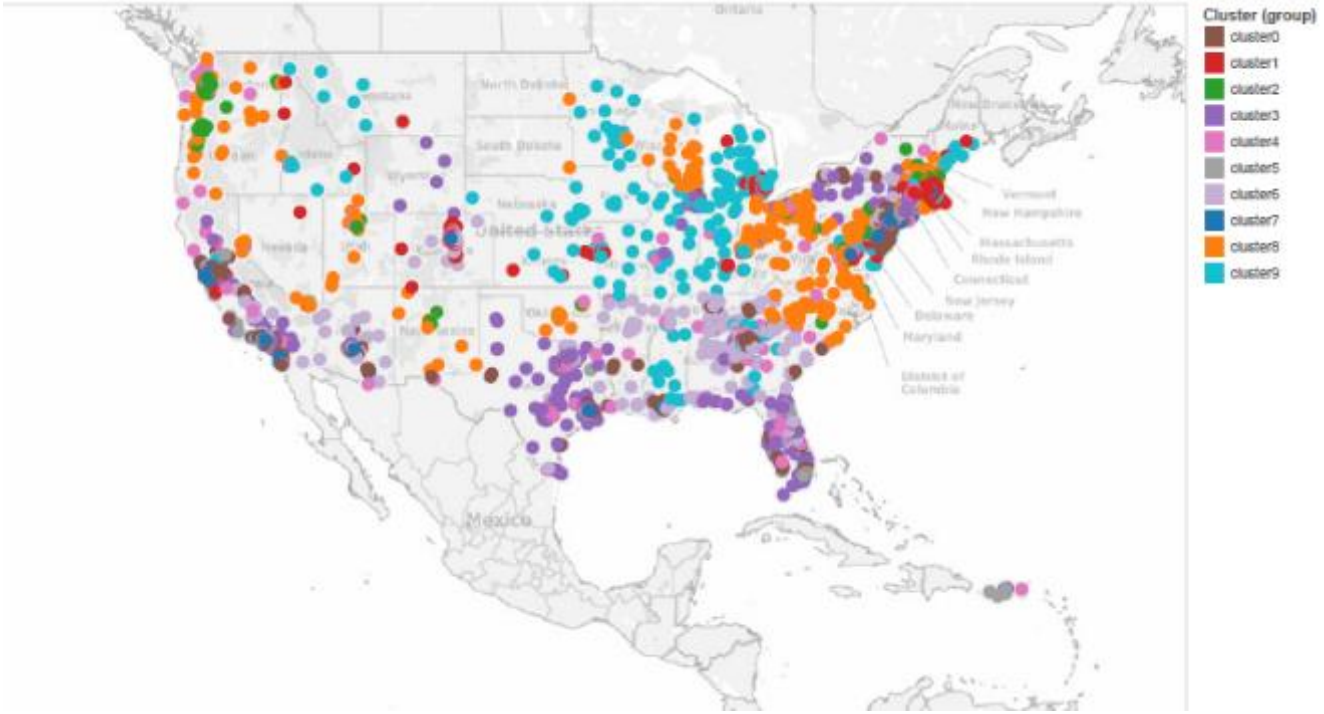
- Search Peer stores

1) 40 highly correlated stores
2) Running stepwise regression with those variables

   Getting a variable from peer stores following this ;

$$P_t = \frac{\sum_1^N p_{i,t}}{N}$$

Clustering using total store sale



Why can weather var. improve predictive powers?

- Weather can make shopping a more or less difficult experience.

- Certain goods complements activities related to weather.

- Search weather indicators

1) Wunderground API
2) Search indicators – daily precipitation and daily mean temperature

Correlation Matrix

	sales	peer	meantemp	precm
sales	1.000			
peer	0.702	1.000		
meantemp	0.002	0.002	1.000	
precm	-0.016	-0.019	0.000	1.000

Models

1. Multivariate regression model with/without the peer store indicator and weather indicators

$$Y_t = \beta_0 + \beta_1 P_t + \varepsilon_t$$
$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1t} + \varepsilon_t$$
$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1t} + \beta_2 W_{2t} + \varepsilon_t$$

2. AR(7) with/without the peer store indicator and weather indicators

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_7 Y_{t-7} + \beta_8 P_t + \varepsilon_t$$
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_7 Y_{t-7} + \beta_8 P_t + \beta_9 W_{1t} + \varepsilon_t$$
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_7 Y_{t-7} + \beta_8 P_t + \beta_9 W_{1t} + \beta_2 W_{2t} + \varepsilon_t$$

Example

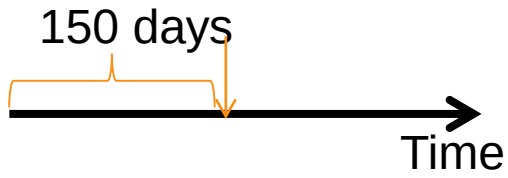
```
R-sq:  within  = 0.5577                         Obs per group: min =       646
       between = 0.9463                                        avg =     713.0
       overall = 0.8127                                        max =       727

                                                Wald chi2(10)      =  6.05e+06
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

totalsales	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
totalsales						
L1.	.1341271	.0006513	205.94	0.000	.1328506	.1354036
L2.	.0877506	.0006564	133.68	0.000	.086464	.0890372
L3.	.0837113	.0006577	127.27	0.000	.0824222	.0850004
L4.	.0644155	.0006594	97.69	0.000	.0631231	.0657079
L5.	.0777139	.000658	118.11	0.000	.0764243	.0790035
L6.	.0917585	.0006564	139.79	0.000	.090472	.093045
L7.	.2205388	.000668	330.15	0.000	.2192296	.221848
ind	.4680301	.0006068	771.26	0.000	.4668407	.4692194
precm	-21.93735	.7361784	-29.80	0.000	-23.38024	-20.49447
meantempm	-.0503913	.0192044	-2.62	0.009	-.0880313	-.0127514
_cons	-24138.89	54.53102	-442.66	0.000	-24245.77	-24032.01
sigma_u	0					
sigma_e	14723.565					
rho	0	(fraction of variance due to u_i)				

Evaluation

- One step ahead prediction



- Recurring rolling regression
- from 1 to Nth observation are used to predict (N+1) th observation

- MAPE (Mean Absolute Percentage Error)

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

Where

At= Actual value;

Ft= Predicted Value.

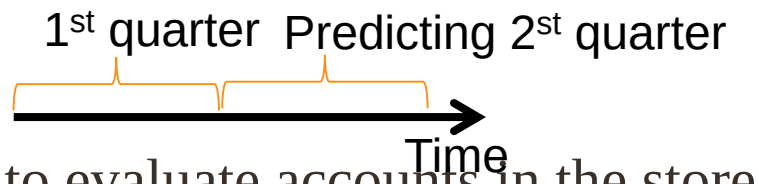
Preliminary Results

	Firm Level (Aggregate Model)				Store Level (Disaggregate Model)			
Model	MAPE	Std.Dev	Min.	Max.	MAPE	Std.Dev	Min.	Max.
AR (7)	0.1097	0.1659	0.0010	1.5557	0.1008	0.1148	0.0000	0.9592

	MAPE	Std.Dev	Min.	Max
Regression with peer stores	0.0594	0.2810	0.0000	21.121
Regression with peer stores and precipitation	0.0195	0.2771	0.0000	6.6932
Regression with peer stores and temperature	0.0289	0.3606	0.0000	7.0191
AR (7) with peer stores	0.0724	0.4069	0.0000	8.4677
AR(7) with peer stores, precipitation and temperature	0.1603	0.3955	0.0438	6.6964
AR(7) with peer stores and precipitation	0.1611	0.3513	0.0444	7.1721

Implications

- By using peer store data audit efforts can be reduced for predicting the next period.



- Especially, it can motivate to evaluate accounts in the store level.

- Contemporaneous weather indicators can improve understanding of possible outliers.

Future research

- **Using Current Dataset**

1. Evaluating weather indicator differently (i.e. temperature–humidity index (THI))
2. Adding other accounts (i.e. account receivables, inventories, the total hours of works)
3. Adding other external information (i.e. social media)

- **Using Other Datasets**

1. The association between social media (Twitter or Yelp.com) and sales of service firms
2. The association between RFID data and inventory account
3. Survival analysis and allowance for loan losses (From Tim)
4. The influence of macro economic indicators (With Lucas)