RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Big Data as Audit Evidence: Utilizing Weather Indicators

Kyunghee Yoon and Alexander Kogan
Rutgers Business School

Why weather is matter?

- Weather is considered as one of factor affecting business performance (i.e, 50,704 times referred in 10-K filings from SeekiNF).

- Friendly's restaurants (2010)

"results for the year were negatively impacted" in part by

"unusually cool weather in the northeast, especially in the summer months."

- Nike Inc. (2015)

"weather events that impacted two of the three major holiday periods of the 2014/2015 ski season and adversely affected the ski industry in general."

Big data as audit evidence

- Thanks to the advanced technology auditors are likely to utilize a variety of nonfinancial information without high costs to enhance the accuracy of expectation (Trompeter and Wright 2010)

- Some parts of big data can provide more reliable, relevant, and sufficient audit evidence than traditional audit evidence, but there are limitations such as a way to integrate big data with traditional audit evidence and measure big data (Yoon et al. 2015).

- There are little studies to examine empirically to overcome those issues.

Contribution

- **This study:**

- Presents the value of big data as audit evidence by utilizing weather indicators for performing substantive analytical procedures.

- Provides a way to evaluate, utilize, and integrate big data with traditional audit evidence.

- Revises the value of store level of disaggregated data in analytical procedures.

The Values of Disaggregated Data in APs

- Analytical procedures (APs) are required at the planning and review phases of an audit (AICPA 1988), but more auditors adopt APs to use the substantive test of detail.
- Expectations developed at a detailed level generally have a greater chance of detecting misstatement of a given amount than do broad comparison (AICPA 1989).
- Allen et al. (1998) did not find store level data can provide better expectations but find peer stores have a significant predictive power.
- Peer stores might have similar economic environments (i.e. cities, rural areas).

The Values of Disaggregated data in APs (cont.)

- H1: Firm-wide sales expectation developed from disaggregated individual location produce more accurate and more precise expectation than firm-wide sales expectation derived from aggregated firm level data.

Nonfinancial Information and APs

- SAS No 56 (AICPA 1989) suggests that nonfinancial information should be considered when performing analytical procedures (AICPA 2002, 2007).
- That management can modify the amount of accounts to reach the expected trend, analyzing financial data tends to be ineffective for searching red flags (PCAOB 2004).
- Brazel et al.(2009) present that the inconsistency between financial information and non financial information is related to management fraudulent behaviors.
- Some research use macroeconomic indicators (i.e. GDP) (Lev 1980) or number of employees and production spaces (Brazel et al. 2012).

Nonfinancial Information and APs (cont.)

- However, these previous nonfinancial indicators have limitations and might not appropriate daily store level predictive models.
- Weather indicators such as temperatures and precipitations are updated in timely basis and locational basis.
- Starr-McCluer (2000) examines various kinds of retail firms and shows overall temperatures have relatively strong explanatory power for sales of retail firms. Especially, she argued that unfavorable weather conditions such as cold weather or precipitations can interrupt customers' store visit, thereby reducing the sales.

RUTGERS

Nonfinancial information and APs (cont.)

- H2: The model with weather indicators produces a more accurate and precise prediction than the model without weather indicators.

Data descriptions

- 12 variables (store address, date, daily total sales)
- 2011/1/31- 2/3/2013 (Total 735 days)

Description	# store	# obs
Original data over 54 states in the US (1,411 cities)	1,978	1,443,044
Less: stores closed/open during the period	8	3,406
Add: filing missing dates (i.e. Dec 25)	-	8,312
Less: stores without weather indicators	16	-11,760
Final data	1,954	1,436,190

Search Peer stores

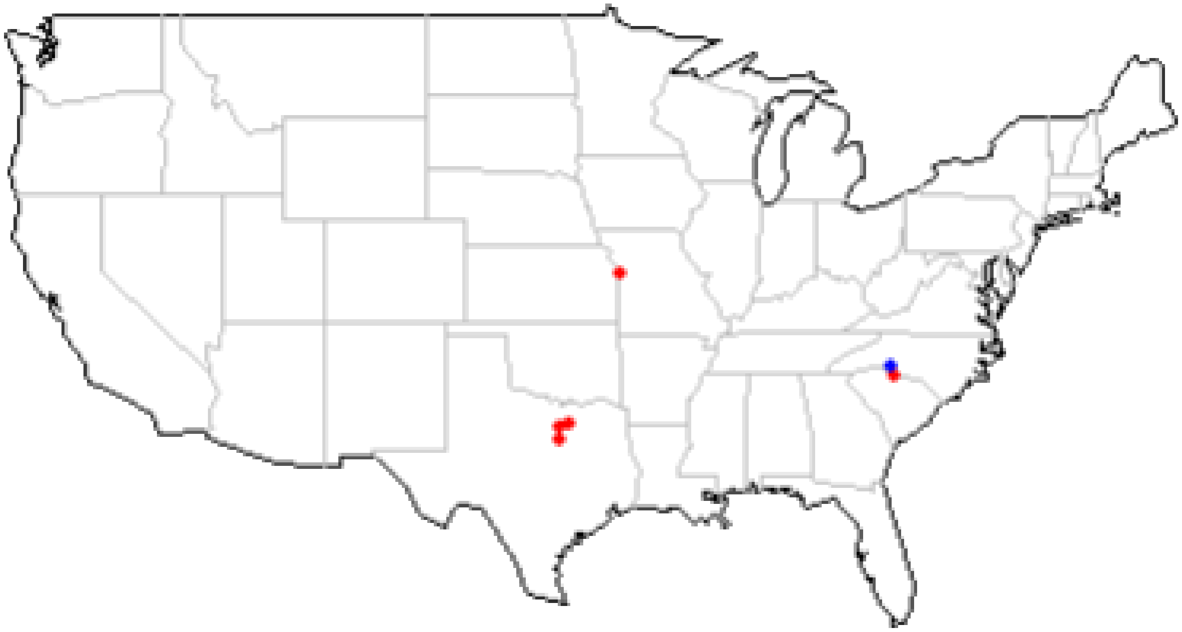
1) 40 highly correlated stores (daily sales)

2) Running stepwise regression with those variables

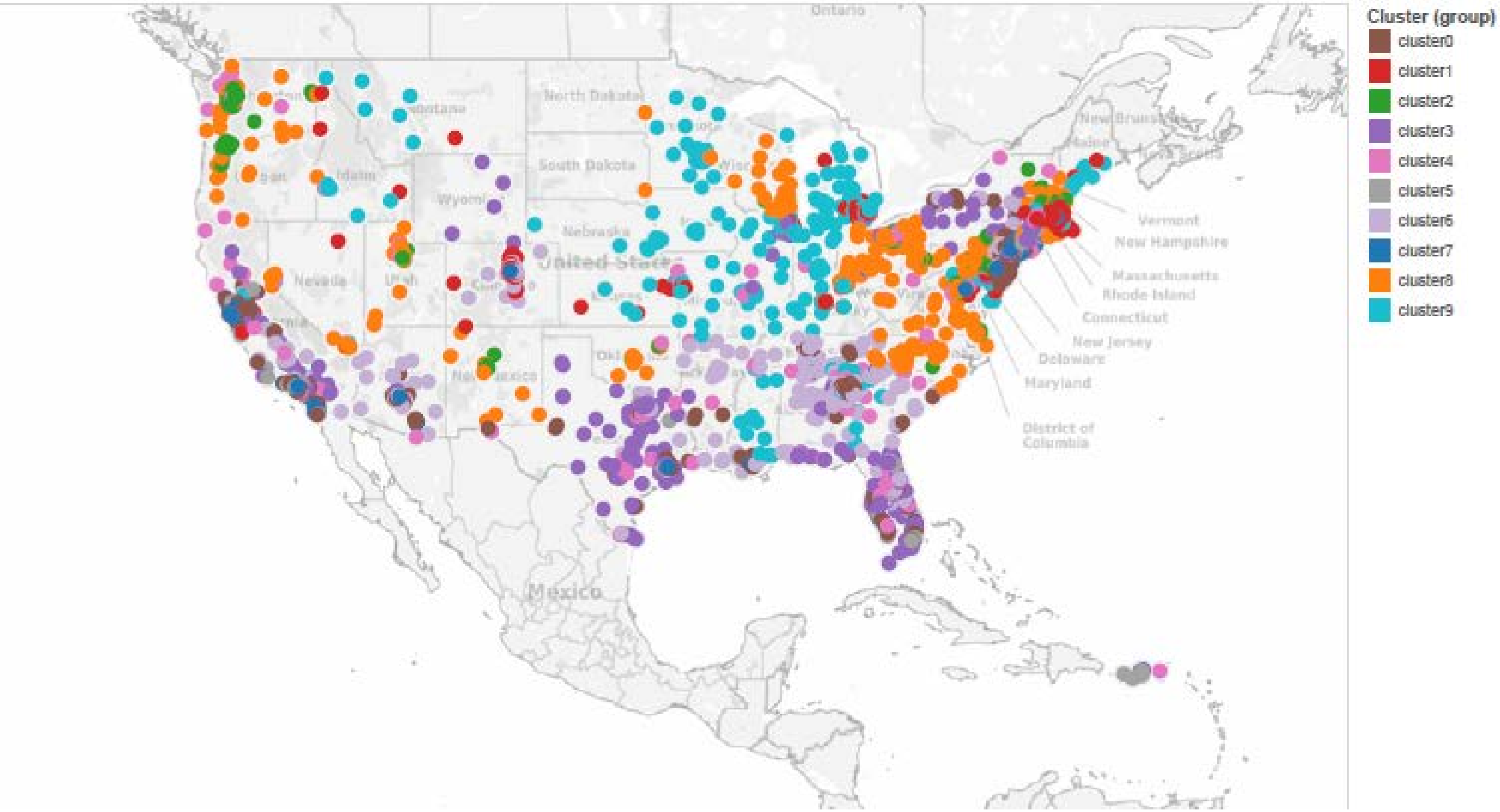
3) Getting a variable from peer stores following this ;

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

Search Peer stores – an example



Clustering using total store sale



Search and Measure Weather Indicators

1) Wunderground API

2) Search indicators by using zip code – daily precipitation, daily mean temperature, wind speed, etc.

3) Weather index:

$$\alpha * \text{Standardized Heat Index} + \beta * \text{Standardized Wind chill}$$

Where $\alpha = 1$ if mean temperature > 50

$\beta = 1$ if mean temperature $<= 50$

4) precipitations, temperature…

RUTGERS

Models- H1

1. Firm level sales prediction derived from aggregated data

$$\widehat{Y_t} = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_7 Y_{t-7}$$

2. Firm level sales prediction derived from disaggregated data

$$\widehat{Y_t} = \sum_{i=1}^{j} (\beta_0 + \beta_1 y_{i,t-1} + \cdots + \beta_7 y_{i,t-7})$$

Where
Y_t = a daily firm level account balance series under audit
y_t = daily store level account balance

Models- H2

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

$$\widehat{Y}_t = \sum_{i=1}^{j}(\beta_0 + \beta_1 P_{i,t})$$

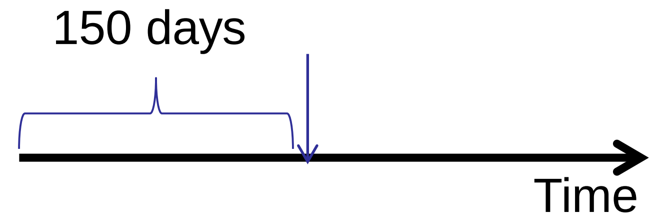
$$\widehat{Y}_t = \sum_{i=1}^{j}(\beta_0 + \beta_1 P_{i,t} + \beta_2 W_{i,t})$$

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

$$\widehat{Y}_t = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \cdots + \beta_7 y_{i,t-7} + \beta_8 P_{i,t} + \beta_9 W_{i,t})$$

Let Y_t be a daily firm level account balance series under audit, where t is day and where $P$is the average values calculated by peer stores' accounts for each store and W_i is the daily weather index

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.

Results – H1

	Firm Level (Aggregate Data)				Store Level (Disaggregate Data)			
Model	MAPE	Std.Dev	Min.	Max.	MAPE	Std.Dev	Min.	Max.
AR (1-7)	**0.1097**	0.1659	0.0010	1.5557	**0.1008**	0.1148	0.0000	0.9592

Results – H2

	With peer stores				With peer stores and weather indicators			
Model	MAPE	Std.Dev	Min.	Max.	MAPE	Std.Dev	Min.	Max.
Regression	**0.0594**	0.2810	0.0000	21.121	**0.0195**	0.2771	0.0000	6.6932
AR(1-7)	**0.0724**	0.4069	0.0000	8.4677	**0.1603**	0.3955	0.0438	6.6964

RUTGERS

Next Step

- Modifying strategy to select peer store
- Using economic data by geographical locations (i.e. total taxable income, populations, real estate taxes, etc.)

Thanks!!