RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Big Data as Audit Evidence: Utilizing Weather Indicators

Kyunghee Yoon, Clark University
Alexander Kogan, Rutgers, The State University of New Jersey

Why Weather is Matter?

- Weather is often listed as one of the factors affecting business performance (e.g, times referred in 50,704 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."

Weather and Sales in Retail

- Even though certain sectors are completely weather sensitive, weather may differently affect the outcome of business activities in a variety of industries (Larsen 2006).

- In studying various types of retail firms, Starr-McCluer (2000) demonstrates that overall temperatures have a relatively strong explanatory power for sales. In specific, she argues that unfavorable weather conditions, such as cold weather or precipitation, can discourage customers' store visit, thereby reducing the sales.

Weather Variables as Audit Evidence

- Weather variables can be reliable and relevant audit evidence because

  - Weather variables are external independent variables which are not affected by management.

  - Weather components are updated in a timely and locational basis, thereby matching with timely and locational specific sales.

  - Weather information is contemporaneous information to the audit procedures.

Weather Variables as Audit Evidence (cont.)

- Prior research often uses macroeconomic indicators (i.e. GDP) (Lev 1980) or the number of employees and production spaces (Brazel et al. 2012).

- However, these variables have limitations and might not be appropriate for daily store level predictive models.

Contribution

- Weather variables have a potential to enhance the accuracy and precision of analytical procedures. Even though weather can impact the outcomes in many industries, accounting and audit domains have rarely shed light on its significance for business activity.

- This work contributes to reevaluating the value of disaggregated data by location.

Research Method- Case Study

- This study used a multi-location retail firm with homogeneous operations in the world, but focused only stores in the continental USA.
- It used data from the 2011 to the 2012 fiscal year
- It used 1,901 operating units.
- Even though daily and weekly balances were not audited separately,

  the external auditor expressed an unqualified opinion on the effectiveness of the case firm's internal controls.

Control Variable – Peer Stores

1. Highly correlated lagged macro economic indicators with sales are selected. Running Principal Component Analysis with Varimax Rotation- three independent factors are generated.

2. Rank factors; only stores having all three factors within n/7.

3. Calculating the average of peer stores' sales as follows;

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

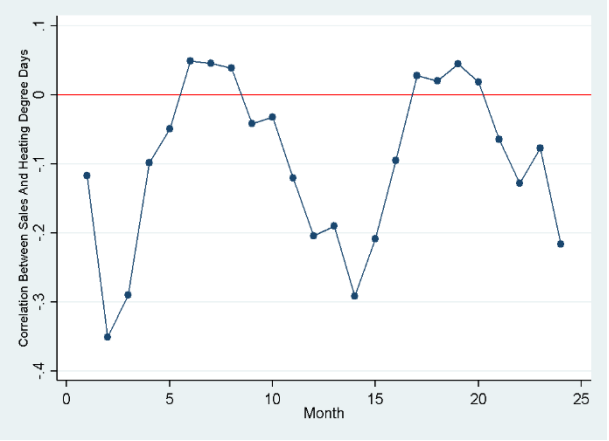
Search and Measure Weather Indicators

- Wunderground API
- Search indicators by using zip code – daily precipitation, daily mean temperature, wind speed, etc.

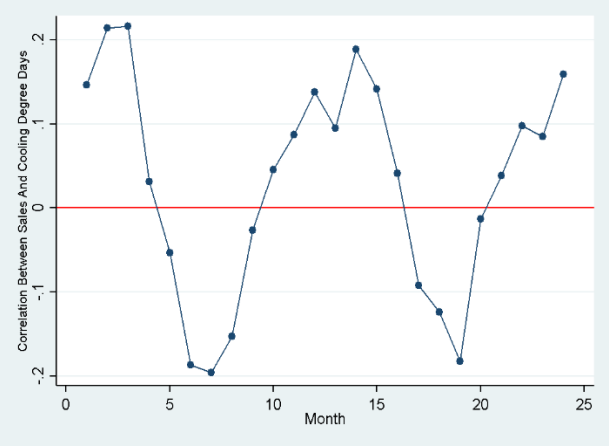
- Weather index:

1. “cooling degree days” (CDD) and “heating degree days” (HDD) (Larsen 2006; Starr-McCluer 2000).
2. Heat Index (HI) and Wind chill Index (WCI) are also used to evaluate unfavorable weather conditions (Feinberg and Genethliou 2005).

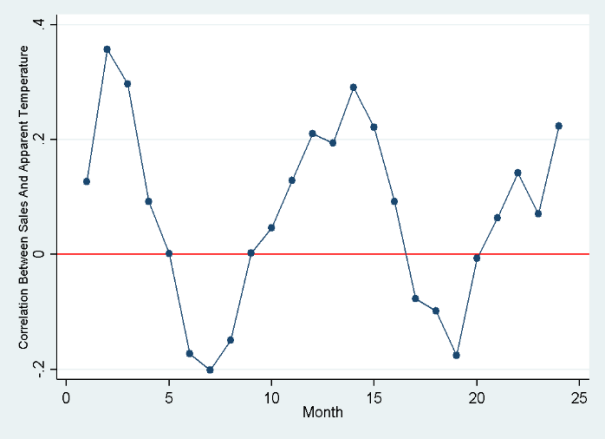
Correlation between sales and HDD



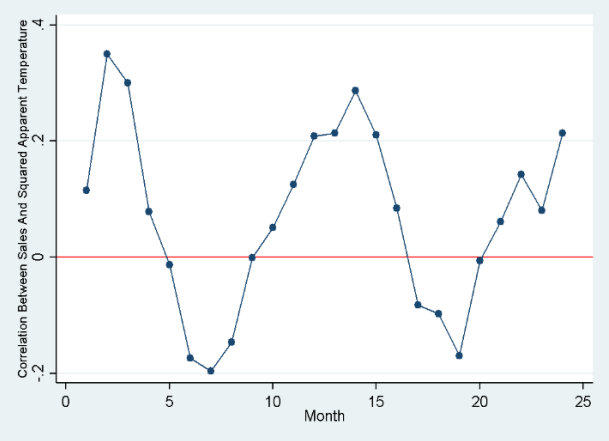
Correlation between sales and AT



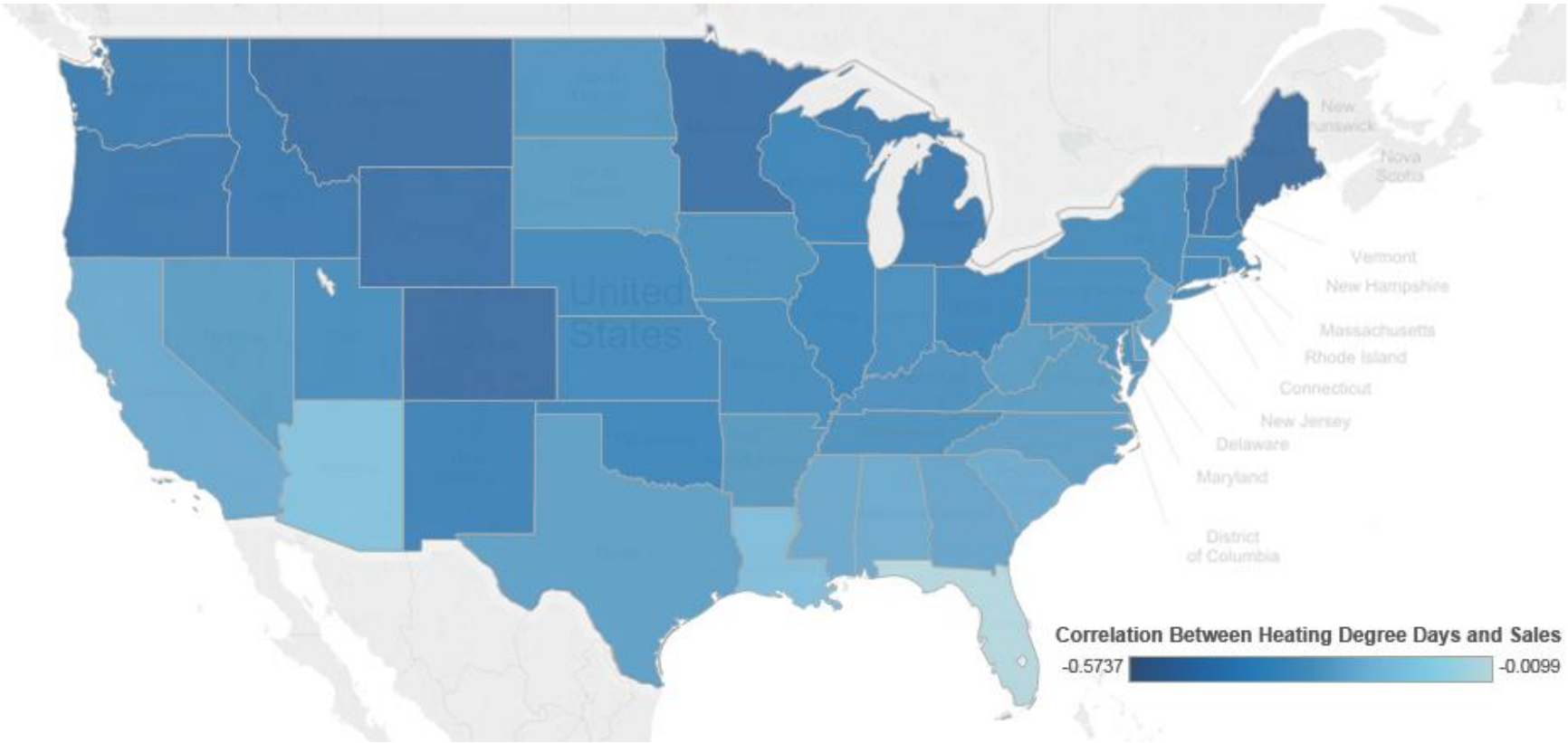
Correlation between sales and CDD



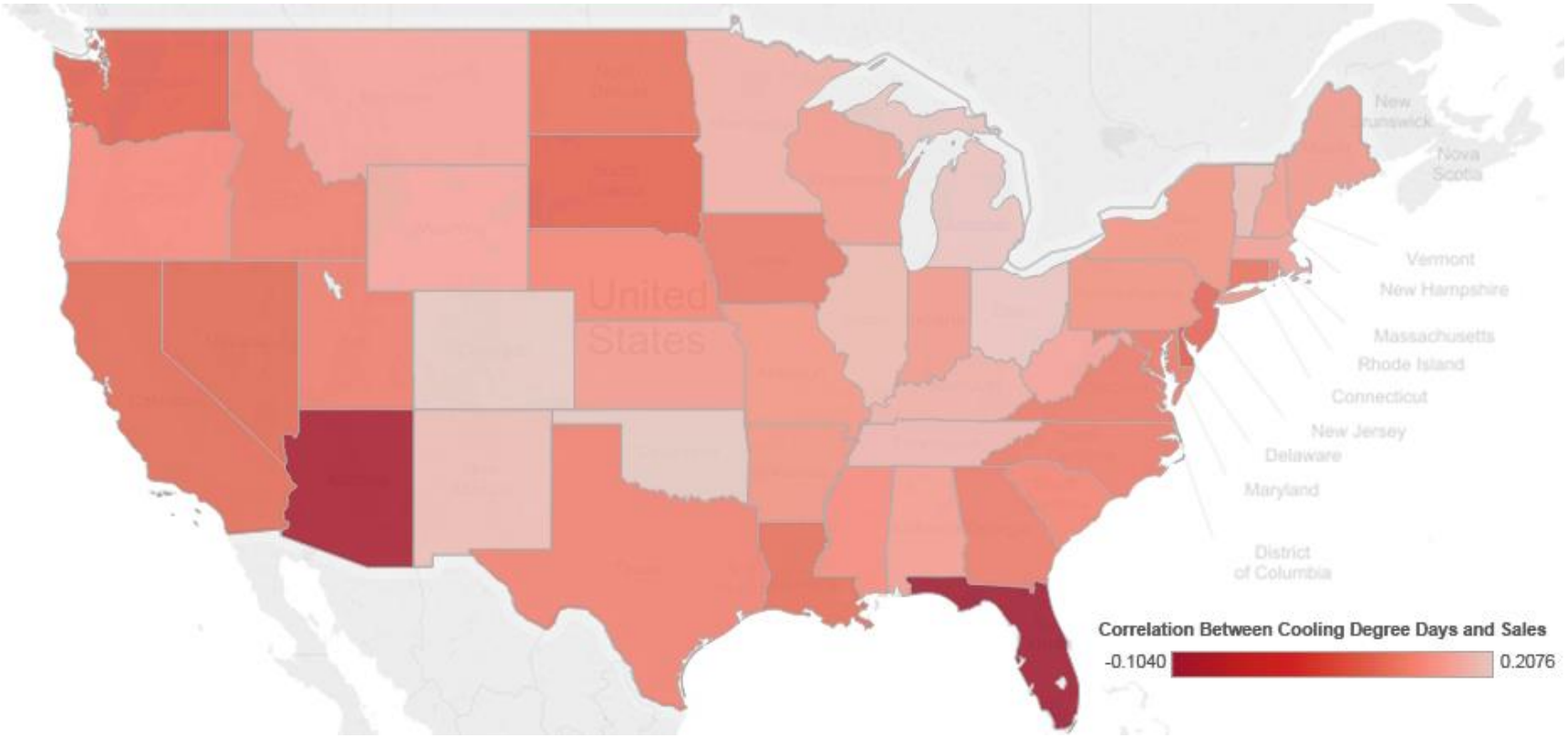
Correlation between sales and SAT



The impact of weather on sales by region - HDD



The impact of weather on sales by region - CDD



Models

- Multivariate regression model

$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 W_{i,t})$

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

- Time series model with/without an exogenous variable

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

$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_2 W_{i,t})$

$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_8 P_{i,t})$

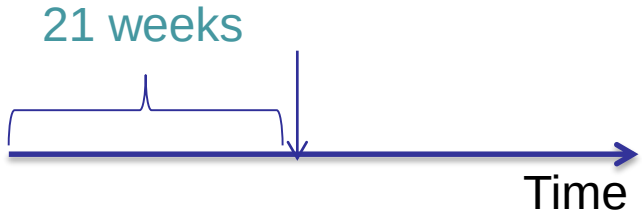
$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_2 P_{i,t} + \beta_3 W_{i,t})$

Let Y_t be a weekly 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

Approaches to Set Expectations and Evaluate Expectations

- Setting an expectation

1. One step ahead prediction: Recurring rolling regression
   - from 1 to Nth observation are used to predict (N+1) th observation



2. Stepwise model selection
   - A group of stores selected by manager has a certain model for a month.

- Evaluation

1. MAPE (Mean Absolute Percentage Error)
2. False positives/false negatives

Conclusion

	Evaluation	Best model	Variables
Prediction accuracy	MAPE	Multivariate regression model	Peer
Error detection	False positive	Multivariate regression model	Squared apparent temperature
	False negative (worse scenario)	Multivariate regression model	Peer Heating degree days Cooling degree days
	False negative (best scenario)	Multivariate regression model	Lagged sales Peer Heating degree days Cooling degree days

Thanks!!