A Machine Learning Analysis of Seasonal and Cyclical Sales in Weekly Scanner Data

Rishab Guha
Harvard University

Serena Ng
Columbia University and NBER

July 2018

1. Motivation

- Increasing availability of data from unconventional sources
- Data no longer cleaned and organized. Preprocessing data is responsibility of user.
- Case considered: **Nielsen Scanner data**, 2006-2014, collected weekly at major MSAs.
- Past uses of data:
  - IO and marketing research: selected products, aggregated over weeks.
  - Macro studies: aggregate data across products to monthly or quarterly frequency.

2. Statement of Problem

- Data potentially useful for monitoring economic conditions in 'real time' and for understanding spending behavior over business cycles at a local level.
- Goal is to extract the cyclical component in weekly data.
- In a demand system, shares are latent functions of prices/income common across goods.
- Large (N, T) allows us to consistently estimate the latent functions (or factors).
- Factors are found to be convolution of cyclical and seasonal variations.
- Three challenges in deseasonalizing weekly data:
  - weekly seasonal effects are quasi-periodic. Spikes from sales and special events.
  - only observe 9 years (seasonal observations) for each series
  - large volume and variety.
  - adjustment errors from series-by-series filtering become non-negligible upon aggregation in the presence of common seasonal effects.

3. Proposed Method: Pool data across counties.

- Sales in neighboring counties have similar seasonal patterns.
- Use algorithms (machine-learning methods) to choose predictors and functional form.

4. Preliminary Results

- On average, 60% of the variations in sales are predictable.
- find 3 factors in adjusted shares: trend, level, slope. The level and slope factors move closely with the level and 52 week change in consumer sentiment.
- analyze dispersion in cyclical sensitivity across product-groups. Food-in strongly cyclical.
- analyze sensitivity of counties to aggregate risk factors. Large counties more exposed.