A Machine Learning Analysis of Seasonal and Cyclical Sales in Weekly Scanner Data
Rishab Guha (Harvard University) and Serena Ng (Columbia University and NBER

1. Motivation and Questions

- 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.
- Data potentially useful for monitoring economic conditions and understanding spending behavior at a local level and at the weekly frequency. Data can also be used to construct cost-of-living index. But the data are dominated by seasonal effects.
- Three challenges in seasonal adjustment of the Nielsen 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 (almost sure) presence of common seasonal effects.

2. Framework and Methodology

- Target: the broad-based, cyclical component in the data.
- Economic framework: demand system
  - shares are latent functions of prices and income that are common across goods.
  - demand systems are typically estimated for a small group of products. Proxy used in place ot theoretical price index to by-pass non-linearity.
- We have $T = 469$ weeks and $N_g = 108$ product groups. Large (N_g, T) allows consistent estimation of the latent functions (or factors) from shares data alone.
- A panel data, machine learning approach to seasonal adjustment
  - Why panel data? sales in neighboring counties have similar seasonal patterns.
  - Why machine-learning? multi-level heterogeneity. Let algorithms choose relevant predictors and choose functional form.

4. Preliminary Results

- Strong seasonal effects. On average, 60% of the variations in sales are predictable.
- 3 factors in demand system of seasonally adjusted shares: trend, level, slope. The level and slope factors move closely with the level and 52 week change in consumer sentiment.
- significant dispersion in cyclical sensitivity of product groups. A basket relating to 'Home Cooking' is particularly counter-cyclical.
- Strong dispersion in sensitivity of counties to aggregate risk factors. Large counties are more exposed.