Retail demand forecasting: what we know and how it is practised

Robert Fildes

Founding Director

Lancaster Centre for Marketing Analytics and Forecasting

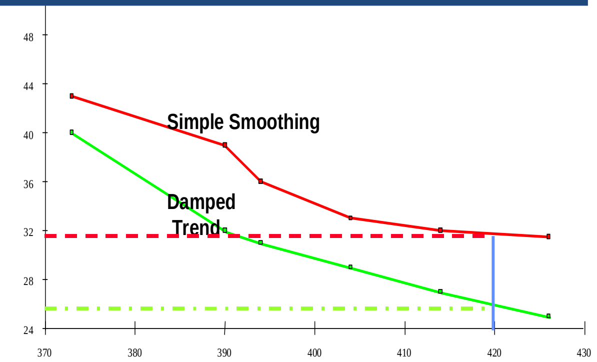
With Shaohui Ma, Nanjing Audit University, China

Stephan Kolassa, SAP Switzerland

Why is retail demand forecasting important & interesting?

- Chaos in retail
  - High street, out-of-town, on-line
- Logistics and environment
  - Packaging
  - Availability
- Service vs inventory: the trade-off
  - Poor forecasts, poor availability, excess stock: Costs
- Technical issues: 50K products x 400 stores, daily: 200K on-line offerings, human factors

Service - inventory tradeoff curves



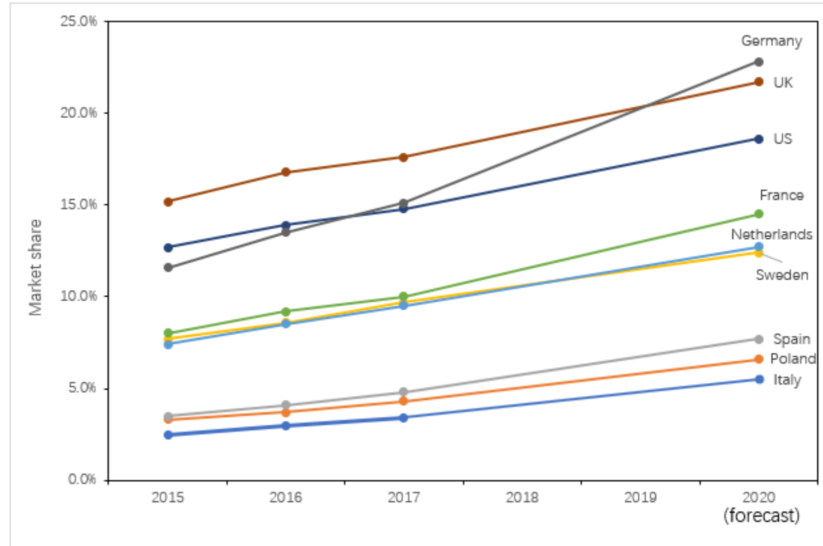
Outline

1. Challenges and decisions facing a retail chain
   - Forecast requirements
2. Aggregate forecasting
   - Strategic Store location
3. Product SKU level demand forecasting
   - Problem features
4. Many explanatory variables
   - Price optimization
   - Product SKU level forecasting Conclusions
5. New Products
6. Channels and Social Media Retail forecasting practice
7. Practical Challenges in Retail forecasting

Challenges in Retail Forecasting

- Strategic decisions
  - Rapidly changing competitive environment
    - **channels**
  - **Store locations**
  - On-line / in-town presence
  - CRM issues, e.g financing, loyalty cards
- Tactical
  - Categories and assortment
    - Brand forecasts
  - **Promotional plan**
  - **On-shelf availability and service level**
  - Distribution centre planning (space, fleet, staffing, service): volume forecasts by size and store
- Operational
  - **'Big data'**
    - SKU x store models for promotional planning and price optimization
  - Short life cycles/ new products/ intermittent demand
  - Rapid replenishment

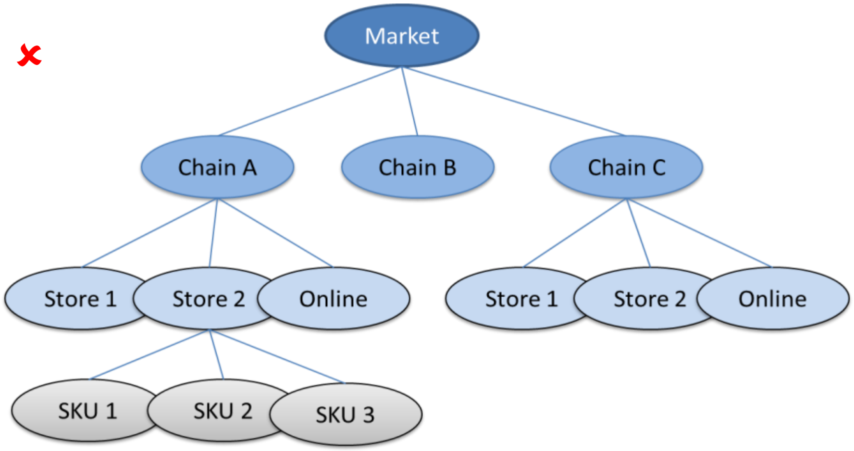
Online shares of Retail Trade



Aggregate forecasting

Total Retail sales in a market (at country or regional level)

- No models linking Retail sales to more aggregate economic variables (e.g. GDP)
  - Comparison with time series alternatives ✗
  - No single method performing best

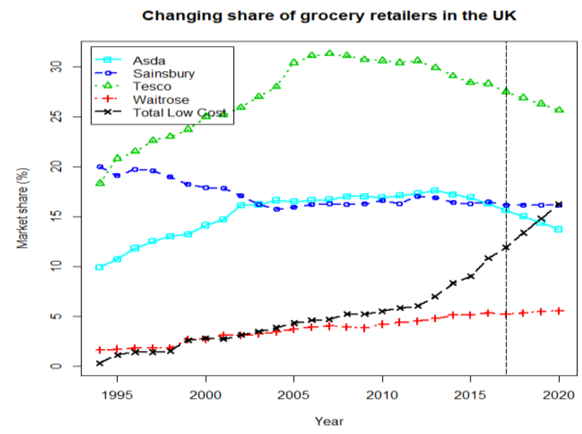


- Disaggregation by channel, by product category
  - Important as total sales masks changes in channel share

- By chain
  - Including info on the store mix, e.g. age mix of stores ✓
  - Financial variables ✓
  - ⇨ **Improved accuracy**

Forecasting Store Sales

- Rapid change in UK market
  - Shift away from out-of-town to convenience
  - Shift to on-line
  - Shift to low price
- New store location models
  - Variables: distance, location and image, services, competition: historical geographical set-up
  - Current Stores provide a biased sample
  - Decisions based on models + judgment
  - BUT changing purchasing behaviour and the shift to on-line?



Appraisal used for store closures

The problem

- *Current data on sales poor predictor*
- *Interaction with on-line*

The result

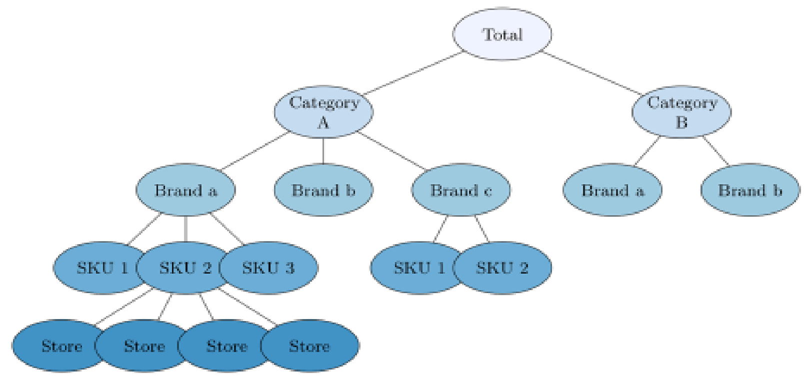
- *Reliance on judgment*

Tactical & Operational

Product level demand forecasting

Decisions:

- Category (tactical)
  - Brand, sku mix
  - Space allocation
- Brand
  - Promotional strategy (frequency)
  - Feature & display
- SKU (operational)
  - Revenue Optimisation
- SKU x Store
  - Segmented stores (e.g. in-town vs out-of-town)
- Distribution Centre: Store x volume
  - Logistics plan: DC volume

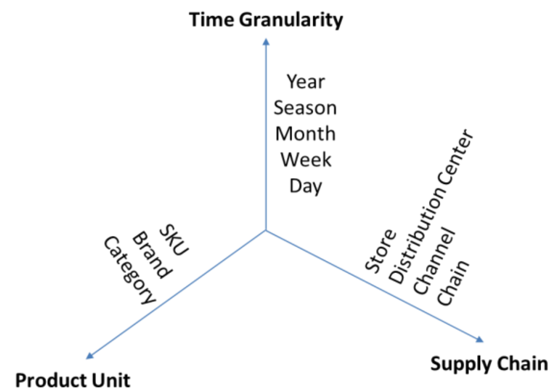


Aggregation approach?

No research on DC dependence on demand?

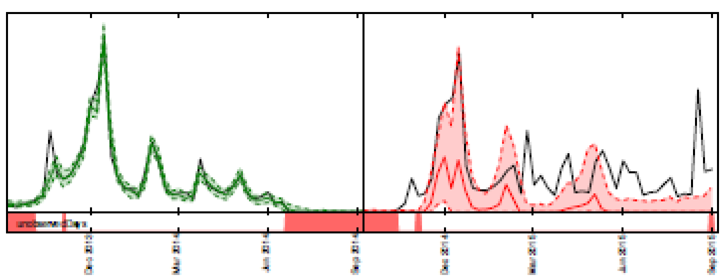
Product level features I

- Forecasts needed within different hierarchies
  - Time
    - Daily at store level for replenishment
    - Weekly at DC level for logistics (picks)
  - Product
  - Supply chain
    - Collaboration?
  - Consistency needed down each hierarchy

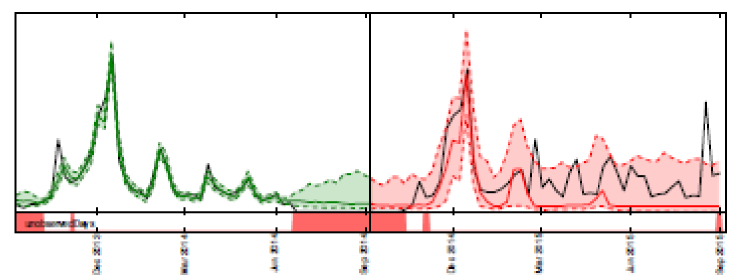


Multidimensional hierarchies

- Data characteristics
  - Stock-outs: demand vs sales
    - Limited data, new technologies (RFID), statistical models



Amazon:Out of stock ignored



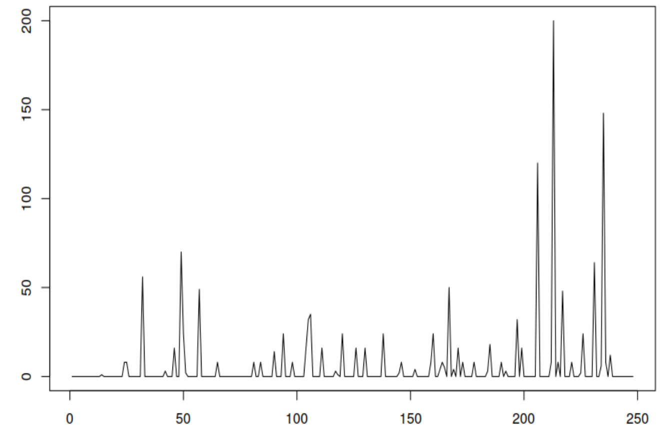
Out-of-stock treated as missing values

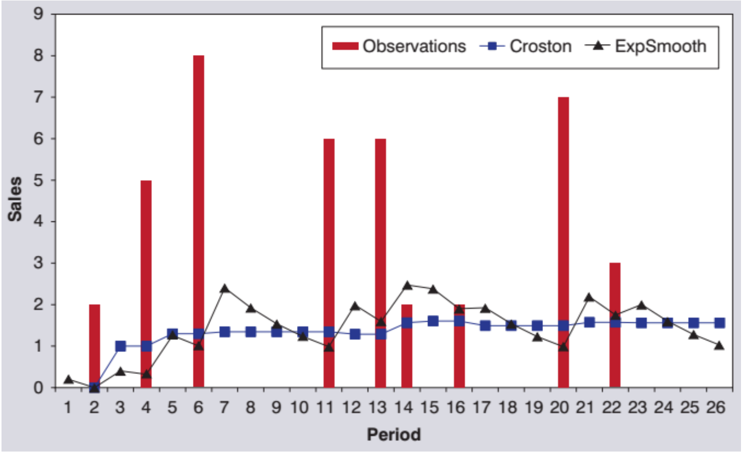
  - Intermittence (lots of it)

The forecasting accuracy punch line:
hierarchies, stock-outs, intermittence all matter

Intermittence – a neglected problem?

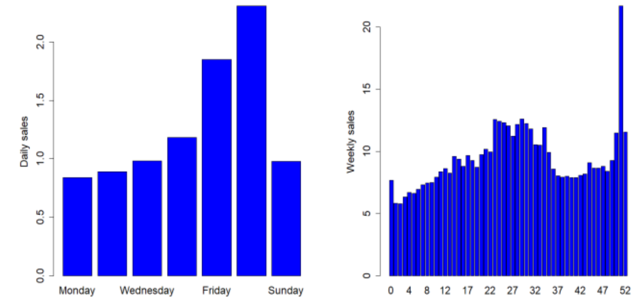
- Regular retailers (70% of SKUs intermittent per week)
- On-line (all?)
- Standard time series methods fail
- Croston the 1971 standard
- Recent activity creating new methods
- Measuring accuracy difficult
  - MAPE, MAE fail and distort
  - Zero forecast!
  - Stock measures?
  - Inventory policies based on normal distribution ✗
  - Model cumulative demand over the order period





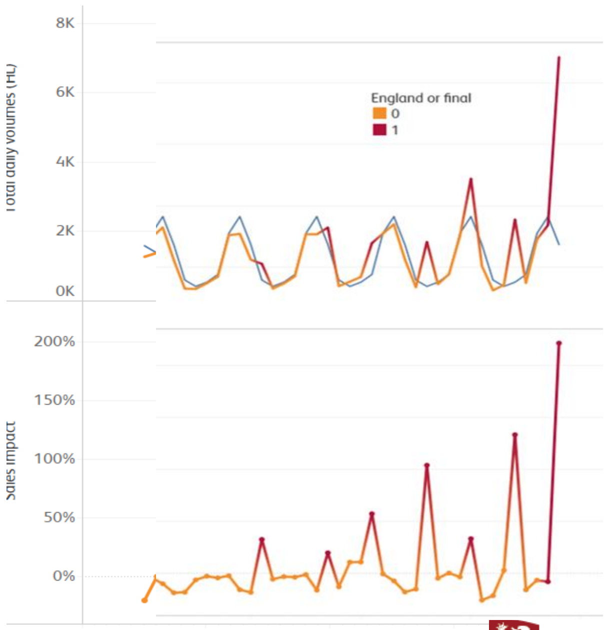
Product level features II

- Seasonality
  - Multiple seasonalities
  - Weekly and daily seasonals interact



Daily and weekly beer sales

- Weather impacts
  - Beer, ice-cream, barbecue
  - But forecasts: horizon, region?



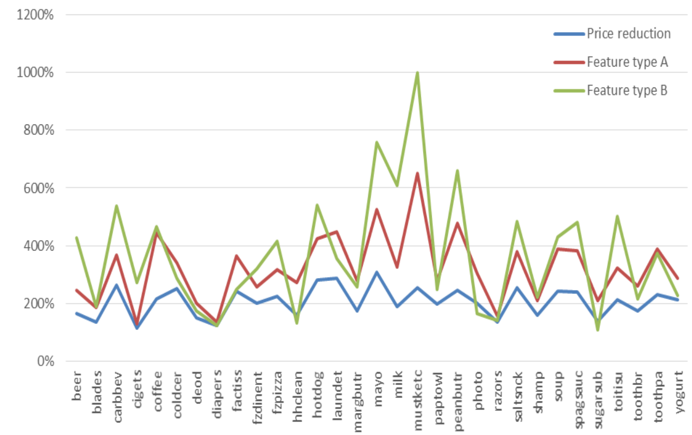
World cup effects on beer – win or lose

- Events

Improved model forecast accuracy - but in a model?

Product level features III

- Promotions
  - Promotional type
  - Category
  - Lagged effects
    - Black Friday stealing sales from Xmas

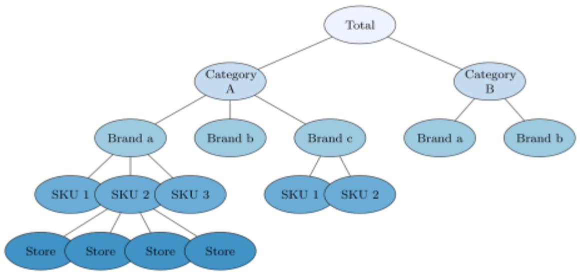


Promotional effects: price, feature and display across categories

- On-line reviews and social media

Research issues and solutions in SKU level forecasting

- Aggregation and consistency
  - Top down vs bottom-up vs middle out
  - Aim for consistency
    - But no consistent best performer



- Disaggregation and explanatory variable effects
  - Disaggregate models needed for heterogeneous effects
    - Store level
    - Category SKUs
  - Many variables
    - But which ones matter?
- Price-promotional optimization

Evaluation

Key issue: relate to decision problem and lead time

- Mean Absolute error

$$MAE = \sum_{i=1}^{m} |Y_{t+i} - F_{t+i}| / m = \sum_{i=1}^{m} |e_{t+i}| / m.$$

- MAPE most often used

$$MAPE = \frac{100}{m} \sum_{i=1}^{m} \frac{|Y_{t+i} - F_{t+i}|}{Y_{t+i}} = \frac{100}{m} \sum_{i=1}^{m} \frac{|e_{t+i}|}{Y_{t+i}}.$$

- Define Relative Mean Absolute Error (compared to benchmark method *B*):

$$\mathrm{RelMAE}_i = {MAE_{Ai}}\big/{MAE_{Bi}}$$

- Summarize over series (for fixed lead time):

$$MAPE = \underset{i}{Mean}(MAPE_i)$$

$$\mathrm{RelMAE} = Geometric\ \underset{i}{Mean}(\mathrm{RelMAE}_i)$$

  - Error < 1 method better than benchmark
  - Error > 1 method worse than benchmark

Evaluation

Key issue: relate to decision problem and lead time

- Mean A
- MAPE r
- Define ... od *B*):
- Summarize over series (for fixed lead time):

The issue:

- Company KPIs poorly define
- No link to decision problem
- Software poorly configured

Consequences:

- Service/inventory tradeoff
- Inappropiate choice of forecasting method

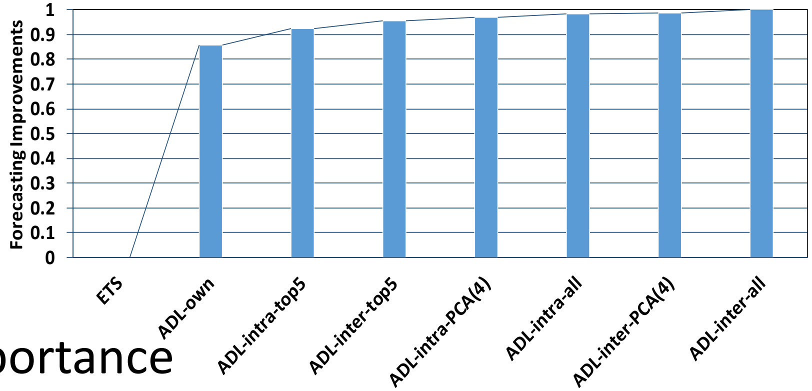
$$MAPE = \underset{i}{Mean}(MAPE_i)$$

$$\mathrm{Rel}MAE = Geometric\ \underset{i}{Mean}(\mathrm{Rel}MAE_i)$$

- Error < 1 method better than benchmark
- Error > 1 method worse than benchmark

Conclusions from SKU modelling of regular products

- Base models using last promotional uplift wholly inadequate
- Pooling data and models across SKUs and Stores improves estimation and forecast accuracy
- Increasingly complex models deliver value
  - Using focal SKU
  - Using core competitive SKUs
  - Using all SKUs in category



- Non-linearities?
  - Software companies emphasizing its importance

Practical issues:

- *Best 'simple' methods?*
- *Are non-linear effects valuable?*
- *Use of software*
  - *Judgment?*

New Products I

Defined as products with less than 2 seasons data history

- Decision context
  - Initial stocking
  - Short Life cycle (fashion goods: electronics)
    - Buying ahead: re-order?
  - The assortment decision: adding a new SKU to a category
  - Distributional consequences of new SKU
- How prevalent?
  - In UK non-food hardware, homeware and garden
    - 50% in data base have less than 2 years history

High variability?

- Retailers as manufacturers
  - Same techniques: market testing, choice models, diffusion
- Fashion forecasting as new product forecasting
  - Literature on non-linear methods unconvincing
  - New methods based on clustering new products based on features
    - colour, price, segment, + click data
    - Forecasting models for clusters

New Products II
New product forecasting methods for retail

- Continuity of data with past SKUs
- Judgment
- Structured judgment
  - Analogous products
  - Interactions with manufacturers (& their forecasts)
- Attribute models of similar products (Vaidyanathan, 2011)
- Bayesian methods based on analogous products
  - Clustering (see Goodwin et al.)
  - Clustering+regression within clusters

No/ little modelling and evaluation

Practical impact: high

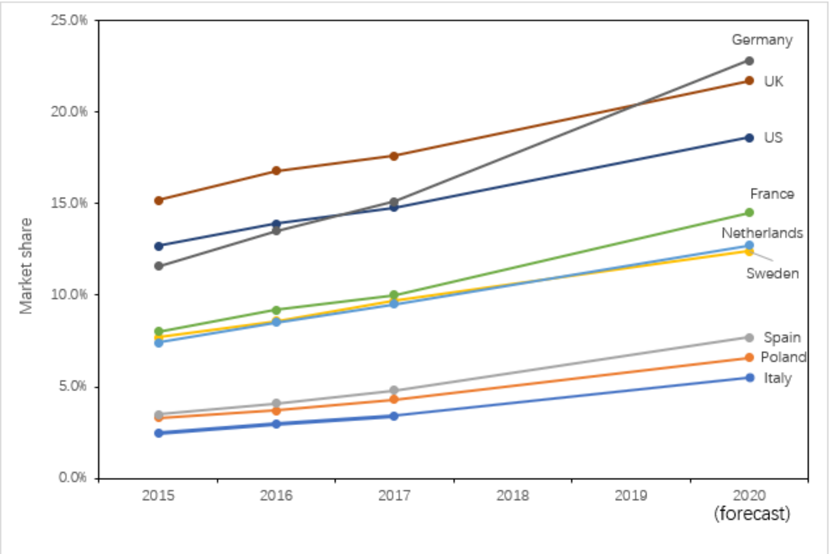
- Major application possibilities in fashion forecasting but…;

  M&S's views

Channels
On-line, catalogue vs Bricks & Mortar

- Rapid growth (in some categories) in on-line
- Competition, cannibalization and complementarity between channels (strategic/ tactical)
  - Generic
  - Niche
  - Search
- On-line shopping (Operational)
  - Web-site design and effects on sales
  - Individual Customer Models
    - Recommender systems (If you like that you'll like this)
    - Returns (and profitability)

Online shares of Retail Trade



Channels: internet sources (social media) and big-data: What we know

- Customer behavioural data
  - Useful for short-term sales generation
  - Potential
    - At SKU level
    - Promotional ‘customer centric’ targeting (Kolassa)
- Social media data
  - Some value for short-term forecasting of ‘instant’ impulse products, e.g. games, music
  - Weak signals (Kolassa, 2017)
    - Do they help?

Tactical & Operational

Variety of practices: bricks-and-mortar and on-line

Retail Forecasting in practice

SAP F&R forecasts: causal model
By SKU and Store: 40K in 400 stores, 25K 'regular'
All skus held at two distribution centres
Focal Horizon ? 26 weeks, 1-6 days for DC
10 promotional types: promotions across chain
Focal horizon for orders: 13 weeks

Promotional info
Events

Sales & Marketing
Rolling two year promotional plan store level, FMCG, DIY, : :
Nalysts interventions based on:
Weather, advertising
New products

Orders to vendors monthly
Stock control
Order to stores

Forecasts: **SKU x store x day**
Daily forecasts automatically prorated?

Distribution Centre:
Translates into picks
Updated daily: depends on current stock levels
Forecast horizon? 1-10 days
Staff Scheduling

Issues in practice

Interviews + presentations from 10 international companies: Household, groceries, fashion, convenience stores

- Commercial software includes ‘demand sensing’ causal capabilities and non-linear methods.
- Few companies have routinized the use of these more advanced procedures; promotional modelling remains simplistic.
- New product forecasting remains heavily judgmental and informal.
- Intermittent demand is a key problem where current ‘best practice’ research has not been adopted.
- KPIs and accuracy measurement is typically not given sufficient attention.
- Lead time issues linked to the supply chain are rarely considered.
- The area of demand planning in retailing is manpower intensive where staff may have overly limited technical expertise.
  - Some companies have a ‘data science’ team to support the core forecasting activity.
- Judgmental intervention superimposed on model based forecasts remains a significant element in retail forecasting.

More tentatively, the diffusion of best practice modelling remains slow.

What do we (not) know?

- Advanced causal methods on sku x store data offer (substantially) improved accuracy

- Advanced new product methods promising
  - Clustering on attributes

- Machine learning methods have potential
  - But not yet well validated on a range of applications

- Social media and search data
  - Probably not valuable for aggregate retail forecasting
  - Delivers for individual customer behaviour (A ‘Kolassa’ priority – the customer of one)

- Big data from customers, IoT and in-store unproven
  - Within day valuable

- On-line and bricks-and-mortar interaction?

Issues of practice
- what gets forgotten?

- By practitioners
- By researchers
- By software designers

- Messy inadequate data
  - Incomplete short histories; new product introductions; intermittent demand; out-of-stock
  - ⇨ Routine algorithms fail to manage exceptions
  - Event history
  - ⇨ Better methods lack data on which they rely

- KPIs
  - The need to link to decisions
  - Forecast error history

- Value added of judgmental interventions
  - How much should organizations rely on their software?
  - How can interventions be made more effective?

Questions and Comments?

Fildes, R., Ma, S., & Kolassa, S. (2018). Retail forecasting: Research and practice. *Working Paper 2018:4*. Lancaster University.

Kolassa, S. (2017). Commentary: Big data or big hype? *Foresight: The International Journal of Applied Forecasting*, 22-23.

Schaer, O., Kourentzes, N., & Fildes, R. (2019). Demand forecasting with user-generated online information. *International Journal of Forecasting*, 197-212.

Ma, S., & Fildes, R. (2017). A retail store SKU promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research, 260*, 680-692