Temporal aggregation and model selection: an empirical evaluation with promotional indicators

Matt Weller, Sven Crone, Robert Fildes

Lancaster University

International Symposium of Forecasting, Santander: 27th June, 2016

Lancaster University Management School | Lancaster Centre for Forecasting

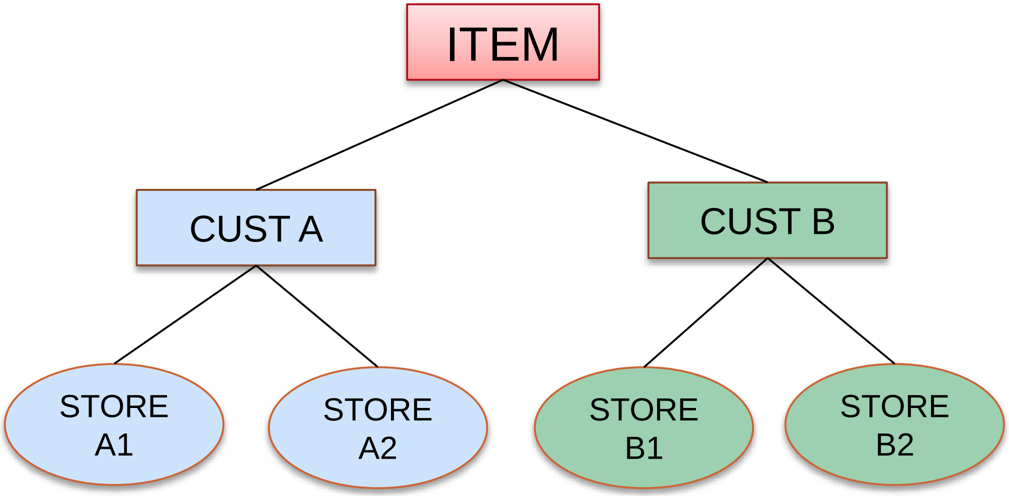
Outline of presentation

- Background to the research
- Empirical evaluation:
  - data, methods, experimental setup
- Results:
  - direct, top-down, bottom-up forecasting
- Implications for practice
- Limitations & opportunities for further research
- Questions

LCF

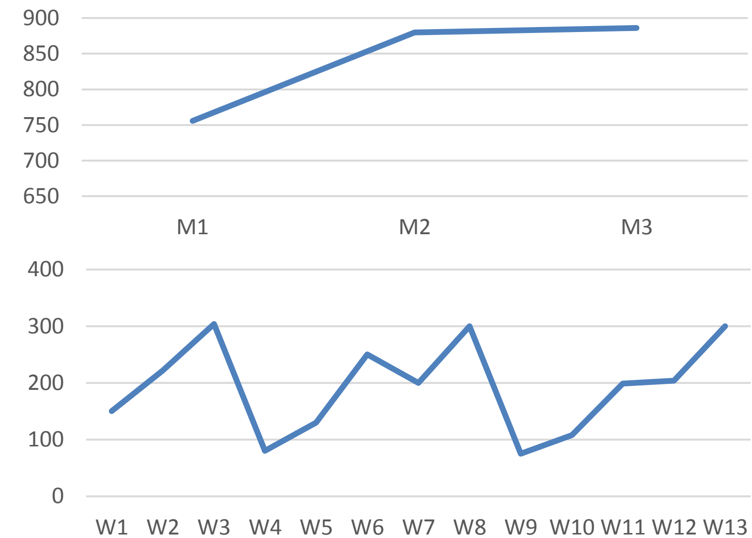
Background: Aggregation in Forecasting

Hierarchical Aggregation



- Customer and/or product hierarchy
- Top-down or bottom up? (*Zotteri et al., 2005*)
- Smoothing or information loss?
- Group seasonality inheritance (*Chen & Boylan, 2007*)
- Intermittent demand consolidation
- Importance of correlation among children

Temporal Aggregation



- ARMA processes (*Brewer, 1973*)
- For intermittent demand (*Nikopoulos et al., 2011*)
- Optimal Reconciliation of hierarchies (*Athanasopoulos et al, 2015*)
- With univariate methods on M3 data (*Spithourakis et al. 2013*)
- Only 3 studies with weekly data in supply chain
- Only 1 study with promotions (*Kourentzes & Petropoulos, 2016*)
- Typically, univariate methods considered!

LCF

Motivation from Practice

Study into forecasting practice, 200+ demand planners in manufacturing organisations *(Weller and Crone, 2012)*

- Good availability of downstream data (e.g. CPFR, EDI, IRI/Neilsen) *(ibid)*
- Prefer monthly data with simple univariate methods *(ibid)*
- Automated forecasting software + Excel *(ibid)*
- Equal weights to split to weeks *(ibid)*
- Judgement is key element *(ibid)*
- Struggle to integrate downstream data *(Lapide, 2011)*
- Poor software support for temporal aggregation *(Rostami-Tabar, Babai, Syntetos, and Ducq, 2013)*

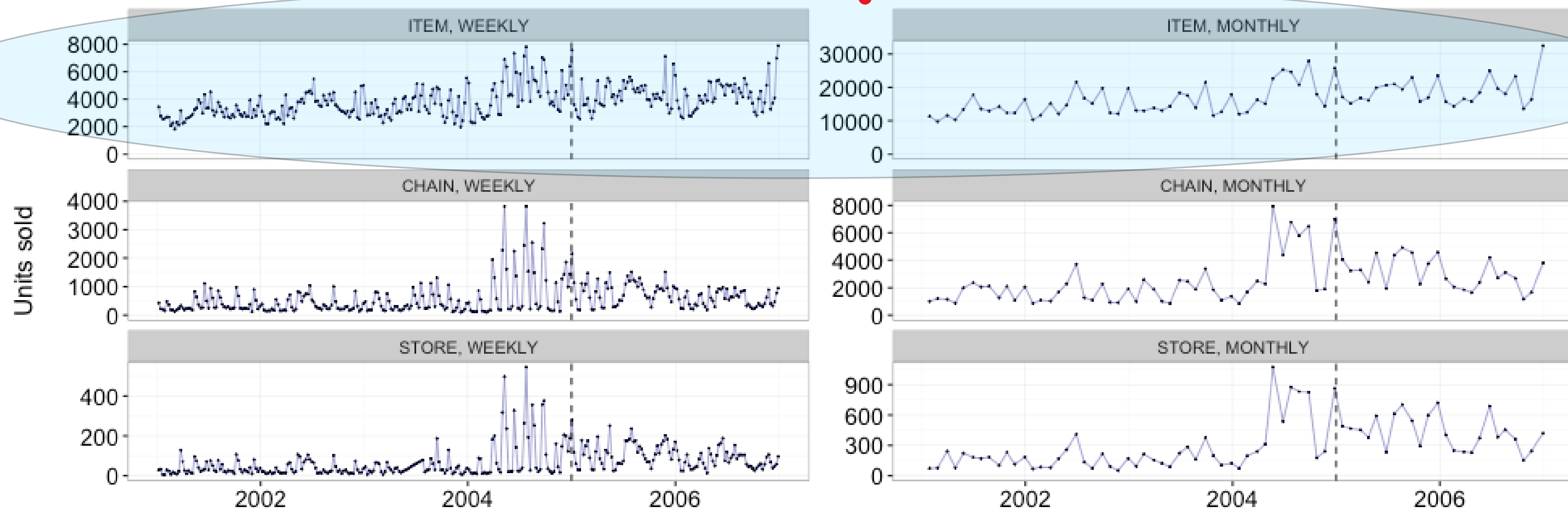
Why the disconnect? Key questions we address:

- Which methods perform best directly on promotional weekly/monthly data?
- Can indirect forecasting help to improve accuracy?
- Which data conditions are relevant in the choice of method/approach

LCF

The Dataset

Focus:
Item-Level



- IRI Academic Dataset *(Bronnenberg, Kruger, and Mela, 2008)*
- Multi-category, multi-manufacturer, multi-retailer
- 6 years weekly store-level data
- Promotional variables: FEATURE (x5), DISPLAY (x3), average selling price

- Aggregation of data (445 non-overlapping)
- Subset of items:
  - SKUs with 6 years' sales history
  - Stores with few zero periods
  - Chains with 4 or more stores
  - Category > 20 SKUs
- Our sample:20 categories, 1700 SKUs

‘Direct’ Forecast Methods

- For weekly and monthly data 8 forecast methods are used
- Univariate & multivariate routines implemented in R/3.3 *(R Core Team, 2016)*

--- Univariate Methods ---

- Naïve, seasonal Naïve benchmarks
- ‘Best fit’ exponential smoothing model selection
  - ETS *(Hyndman, 2016)* – limited to non-seasonal models for weekly data
  - ES *(Svetunkov, 2016)* – allows weekly seasonal models
- HW: HoltWinters function in R
- (S)ARIMA: auto.arima to pick best model *(Hyndman, 2016)*

--- Multivariate Methods ---

- REG-STEP: Stepwise (AIC) regression algorithm
  - Select number of harmonic/fourier terms for seasonality
  - Choose from price, promotional & holiday variables (with lead/lag)
- REG-ARIMA: (1) stepwise variable selection, (2) auto.arima with xreg

LCF

‘Indirect’ Forecasting Approaches: Top-down, Bottom-up

Indirect forecasting in a temporal aggregation context:

- *Transform direct forecasts from their original time frequency to a higher or lower frequency*

Top-down (months -> weeks)

- Forecast monthly with monthly data (8 direct methods)
- Split forecasts to weeks using suitable technique & calendar
  - Equal weights (chosen approach)
  - Historic weights
  - Forecast weights

Bottom-up (weeks -> months)

- Forecast in weeks using weekly data (n.b. 12 origins/year)
- Aggregate weekly buckets into months (445 calendar)

LCF

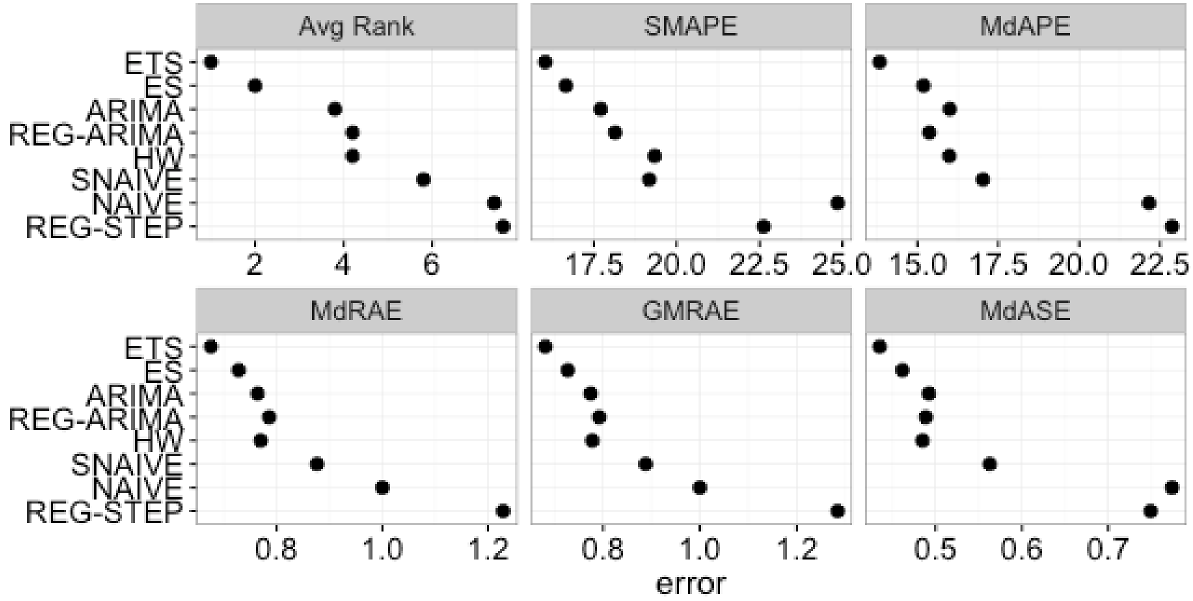
Experimental Design

- Data split: 4 years train & 2 years test
- Rolling origin: 24 months, 104 weeks
- Forecast using each method at each origin
- Persist initial model -> re-estimate parameters at each origin
- Horizon = 3 months, 13 weeks
- Multiple Error Measures:
  - RMSE, MAE
  - sMAPE, MdAPE, MAPE
  - MASE, MdASE, GMRAE, MRAE, MdRAE (against naïve & seasonal naïve)

LCF

Results: Monthly Direct Accuracy

	Avg Rank	SMAPE	MdAPE	MdRAE	GMRAE	MdASE
ETS	**1.00**	**16.04**	**13.83**	**0.68**	**0.68**	**0.44**
ES	2.00	16.66	15.19	0.73	0.73	0.46
ARIMA	3.80	17.71	16.00	0.76	0.77	0.49
REG-ARIMA	4.20	18.15	15.37	0.79	0.79	0.49
HW	4.20	19.33	15.99	0.77	0.78	0.49
SNAIVE	5.80	19.18	17.03	0.88	0.89	0.56
NAIVE	7.40	24.85	22.16	1.00	1.00	0.77
REG-STEP	7.60	22.62	22.87	1.23	1.28	0.75



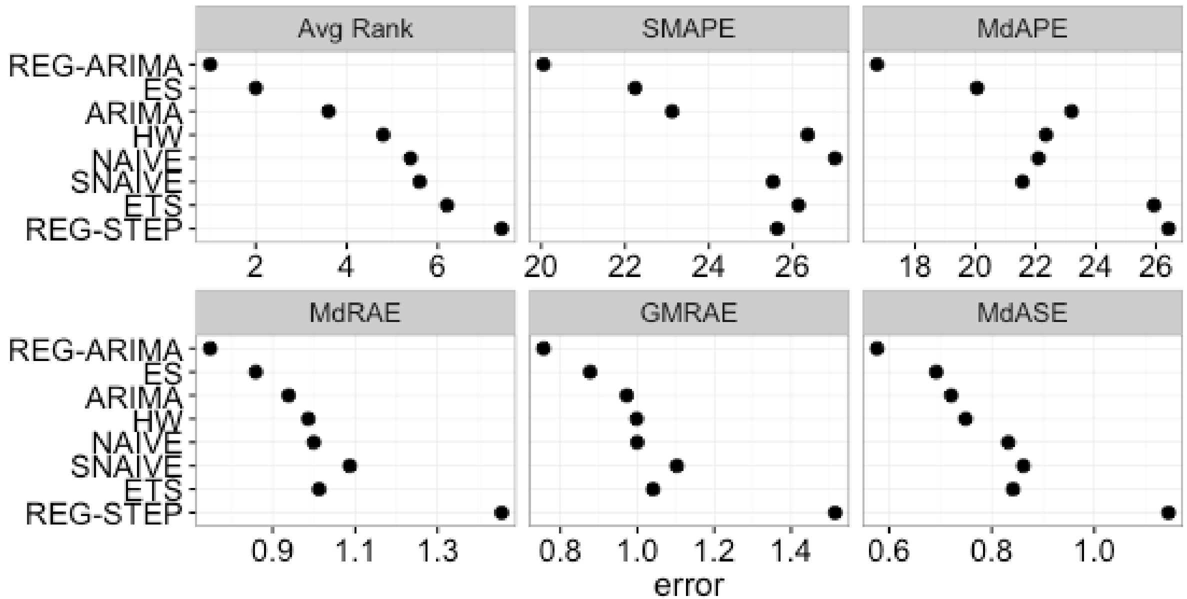
Observations

- Univariate methods: strong performance
- Simpler methods outperform complex ones
- ETS/ES state-space frameworks best
- Holt-Winters alone does not compete
- Explanatory variable do not add value (e.g. REG-ARIMA)

LCF

Results: Weekly Direct Accuracy

	Avg Rank	SMAPE	MdAPE	MdRAE	GMRAE	MdASE
REG-ARIMA	**1.00**	**20.06**	**16.74**	**0.75**	**0.76**	**0.58**
ES	2.00	22.24	20.05	0.86	0.88	0.69
ARIMA	3.60	23.12	23.20	0.94	0.97	0.72
HW	4.80	26.36	22.34	0.99	1.00	0.75
NAIVE	5.40	27.01	22.10	1.00	1.00	0.83
SNAIVE	5.60	25.53	21.56	1.09	1.10	0.86
ETS	6.20	26.14	25.93	1.01	1.04	0.84
REG-STEP	7.40	25.63	26.41	1.46	1.51	1.14



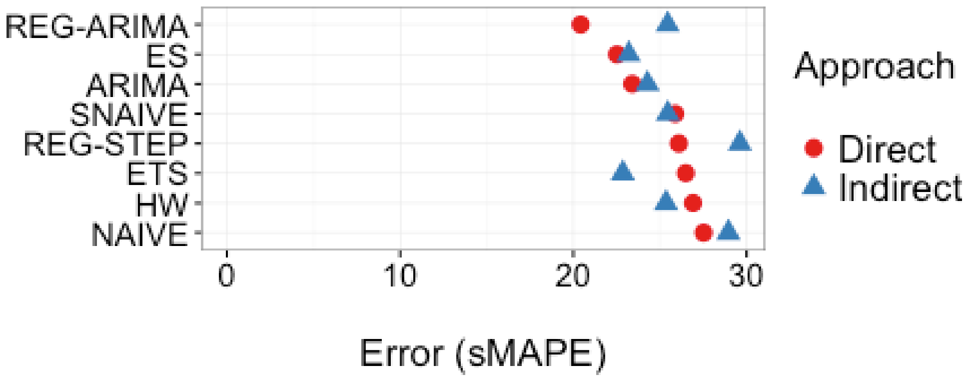
Observations

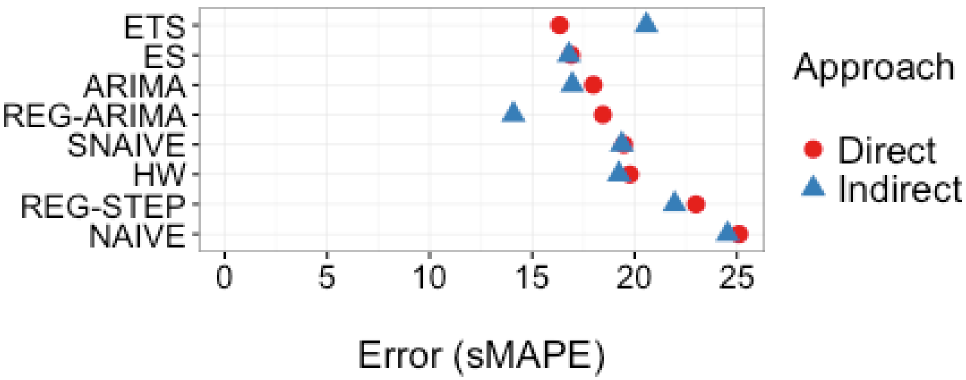
- ETS restrictions make it uncompetitive
- Exogenous variables with weekly granularity are valuable
- REG-ARIMA best of all
- Little improvement for stepwise
- ES now best univariate (includes seasonal in best fit)

Indirect Forecasting Results

	Weekly Direct	Top-down	Gain %
REG-ARIMA	**<u>20.42</u>**	25.44	-24.55
ES	22.51	23.19	-3.03
ARIMA	23.40	24.27	-3.75
SNAIVE	25.87	25.44	1.66
REG-STEP	26.08	29.61	-13.51
ETS	26.48	**<u>22.85</u>**	**<u>13.73</u>**
HW	26.91	25.34	5.82
NAIVE	27.52	28.94	-5.18

	Monthly Direct	Bottom-up	Gain %
ETS	**<u>16.35</u>**	20.58	-25.90
ES	16.91	16.80	0.66
ARIMA	17.99	16.97	5.68
REG-ARIMA	18.45	**<u>14.09</u>**	**<u>23.64</u>**
SNAIVE	19.48	19.38	0.50
HW	19.77	19.24	2.66
REG-STEP	23.00	21.96	4.53
NAIVE	25.10	24.54	2.23





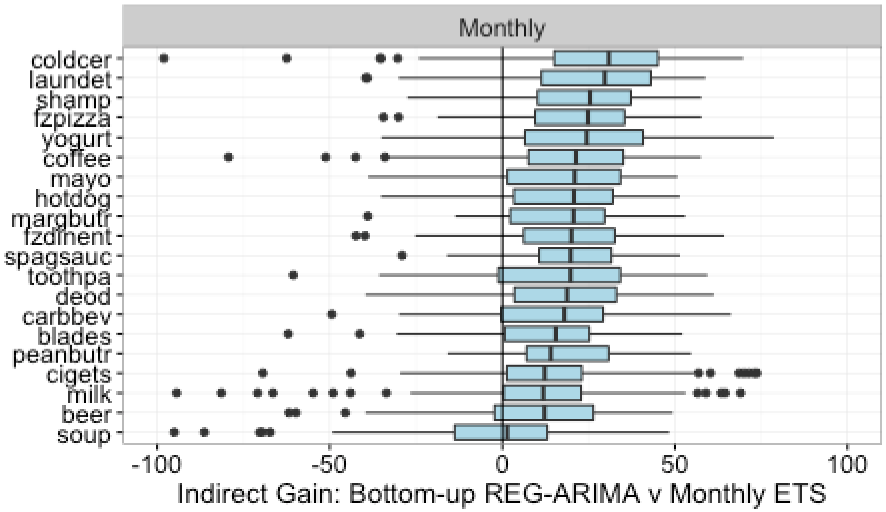
- Top-down univariate does not improve accuracy
- ETS offers best top-down performance

- Bottom-up -> slight improvements for most methods
- REG-ARIMA major gains -> best monthly accuracy

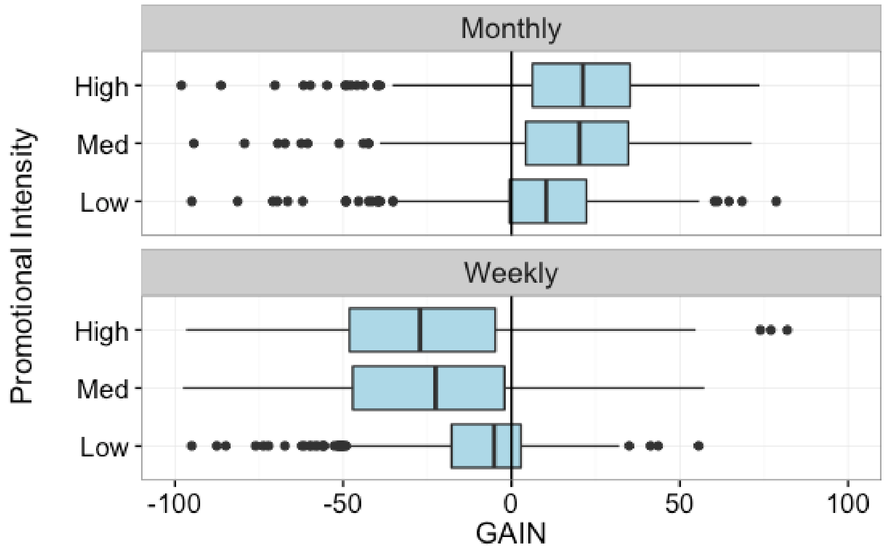
LCF

Factors in the Results

Breakdown by Category



Promotional Intensity split



- Gain from bottom-up REG-ARIMA consistent across categories
- Negative gains from top-down univariate (common in practice) also consistent
- Heavily promoted items show greater gains
- Results are consistent across horizons (M1-M3)

LCF

Implications for Practice

- Results show strong gains for bottom-up forecasting with explanatory variables
- Manufacturers should examine benefits of weekly forecasting:
  - Where promotional intensity is high
  - Impact of promotions is high
  - Public holidays have significant impact (e.g. beer)
- Software vendors should provide functionality:
  - Functionality for top-down/bottom-up comparisons
  - To integrate POS data into forecasting
  - Wider range of models to utilise explanatory variables

Limitations & Further Research

Limitations

- Sample restricted to long history, stable network
  - Neglects new products & listings/de-listings
  - Excludes intermittent demand stores
- Aggregates promo variables for whole network of stores: realistic?
- Stepwise variable selection based on AIC not best practice
- Only considers the 3-month horizon
- Does not include order data, only sales (POS) data
- No interaction effects (e.g. DISPLAY+FEATURE+HOLIDAY)

Further Research

- Temporal & hierarchical aggregation approaches combined
- Consider a wider range of techniques:
  - LASSO/Least Angle Regression
  - Dynamic regression
  - ESX – exponential smoothing with regressors
- Use in-sample "best fit" for method selection

LCF

Thank you for your attention!

Q&A?!

Matt Weller, Sven Crone, Robert Fildes,

Lancaster University

matt.weller@lancaster.ac.uk

Lancaster University Management School | Lancaster Centre for Forecasting

References

[1] B. J. Bronnenberg, M. W. Kruger and C. F. Mela. "Database Paper-The IRI Marketing Data Set". In: Marketing Science 27.4 (Jul. 2008), pp. 745-748. ISSN: 0732-2399. DOI: 10.1287/mksc.1080.0450. (visited on 03/06/2015).

[2] R. J. Hyndman. forecast: Forecasting functions for time series and linear models. R package version 7.1. 2016. .

[3] N. Kourentzes and F. Petropoulos. "Forecasting with multivariate temporal aggregation: The case of promotional modelling". In: International Journal of Production Economics (2015). ISSN: 0925-5273. DOI: 10.1016/j.ijpe.2015.09.011. (visited on 01/20/2016).

[4] L. Lapide. "Integrated Demand Signals Update". In: Journal of Business Forecasting 30.1 (2011), pp. 29-32. ISSN: 1930126X. (visited on 02/08/2014).

[5] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2016. .

[6] B. Rostami-Tabar, M. Z. Babai, A. Syntetos, et al. "Demand forecasting by temporal aggregation". En. In: Naval Research Logistics (NRL) 60.6 (2013), pp. 479-498. ISSN: 1520-6750. DOI: 10.1002/nav.21546. (visited on 01/26/2014).

[7] I. Svetunkov. smooth: Smoothing functions. R package version 0.7. Lancaster Centre for Forecasting. 2016. .

[8] M. Weller and S. Crone. Supply Chain Forecasting: Best Practices & Benchmarking Study. Technical Paper. Lancaster Centre for Forecasting, 2012, pp. 1-42.

LCF