Advances in promotional modelling and analytics

High School of Economics
St. Petersburg

25 May 2016

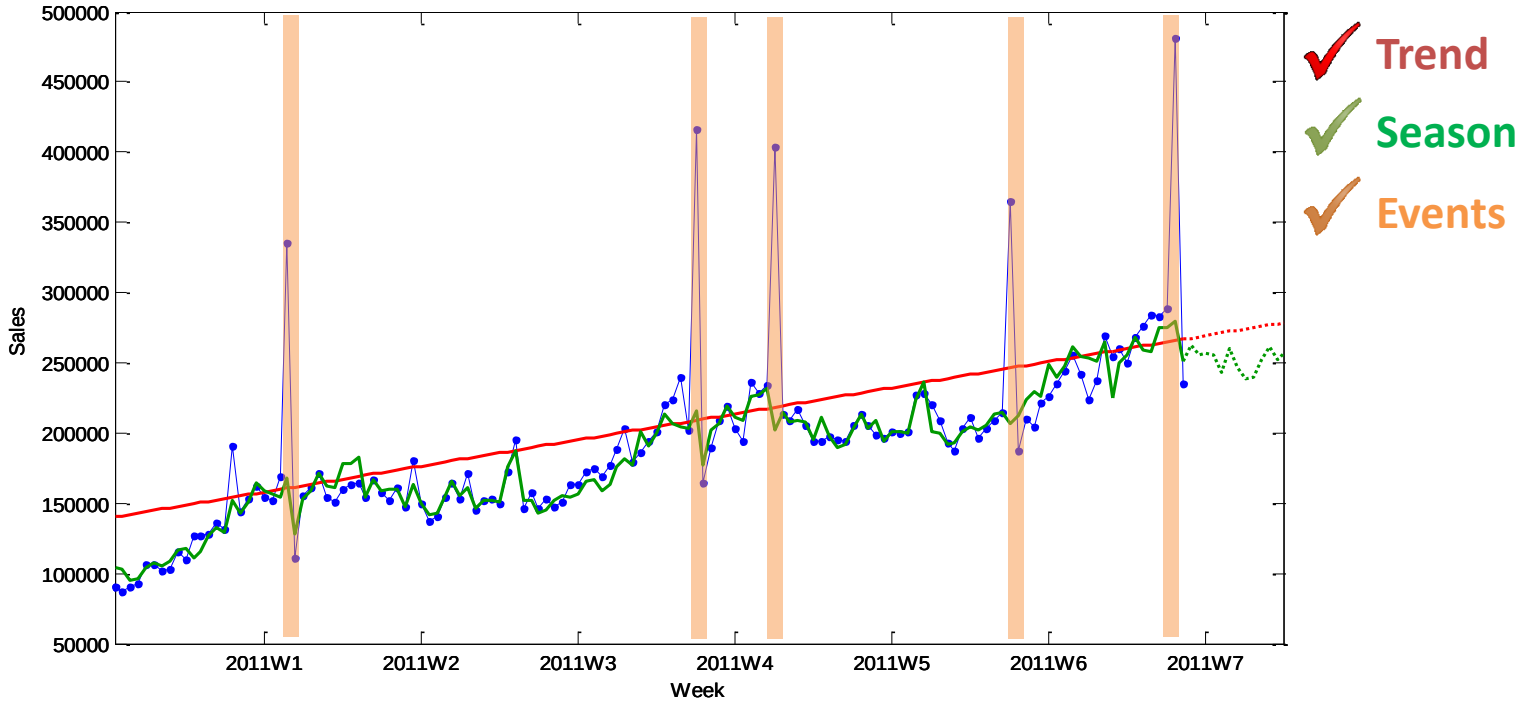
Nikolaos Kourentzes
n.kourentzes@lancaster.ac.uk

Lancaster Centre for Forecasting
Lancaster University Management School

Outline

1. What is forecasting?
2. Forecasting, Decisions & Uncertainty
3. Promotional forecasting
4. New sources of information

What is Forecasting?



What is forecasting? → A process perspective

1. **Identify systematic pattern** in the past (historical data)
2. Select (& parameterise) of an adequate model (set)
3. Execute forecasting model → **extrapolate structure into future**
4. Assess the accuracy of the model(s) in set

Decision making and forecasting

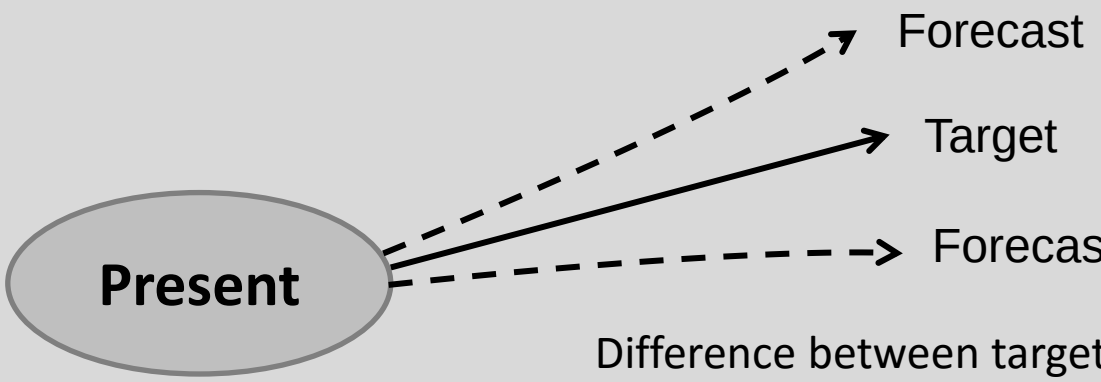
Decision making in organisations has at its core an element of forecasting

→ Accurate forecasts lead to reduced uncertainty → better decisions

→ Forecasts maybe implicit or explicit

Forecasts aims to provide information about the future, conditional on historical and current knowledge

Company targets and plans aim to provide direction towards a desirable future.



Difference between targets and forecasts, at different horizons, provide useful feedback

Decision making and forecasting

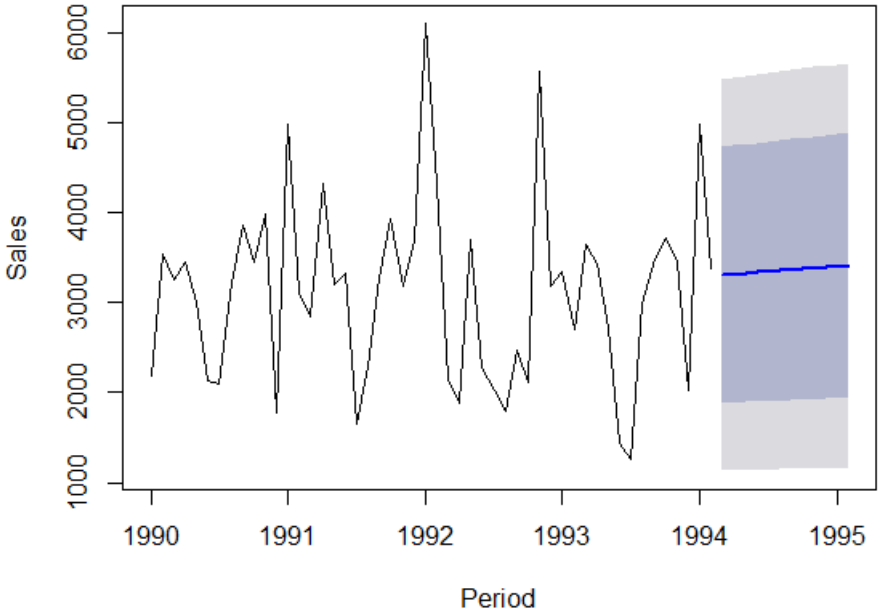
Forecasts are central in decisions relating to:

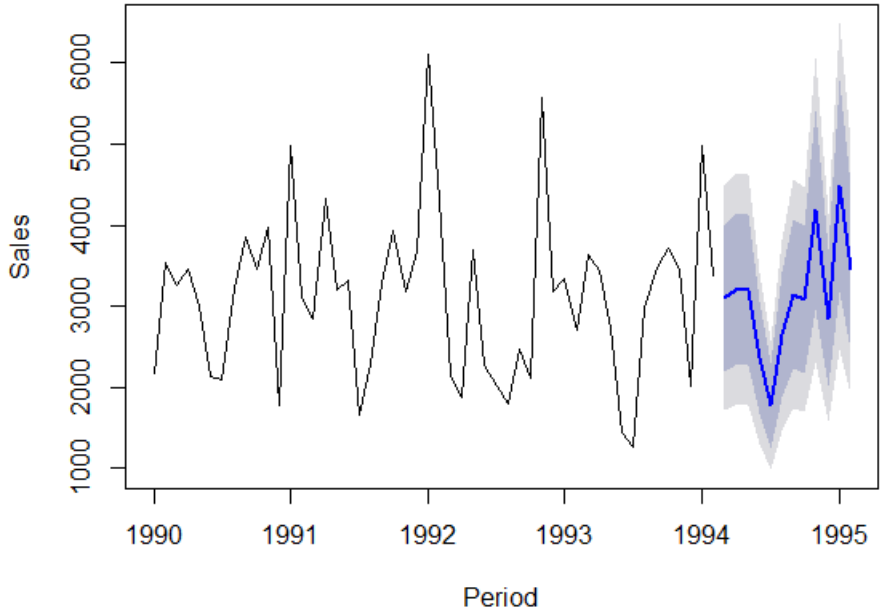
- Inventory management
- Promotional and marketing activities
- Logistics
- Human resource planning
- Purchasing and procurement
- Cash flow management
- Building new production/storage unit
- Entering new markets
- …

Accurate forecasts can support
- Decision making
- Identifying and capitalising on opportunities
- Cost saving

Forecasting and uncertainty

Statistical forecasts are (often) able to provide a forecast and a degree of uncertainty, reflected in the prediction intervals → Probability that true value will lie within bounds.





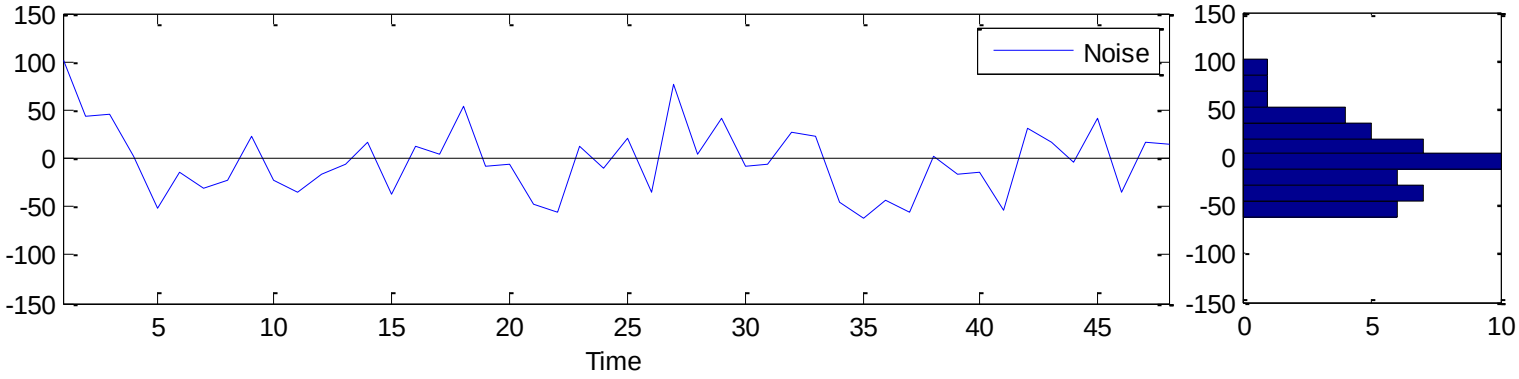
Additional useful information will lead to tighter prediction intervals

→ Less uncertainty about the future

Where does this uncertainty come from?

Noise (or randomness) is inherently unforecastable → it has no structure, otherwise it should be captured by the forecasting method used.

Noise in practice consists of all the information that is impossible to collect or observe, such as the actions of individuals or the mechanics of ill-understood systems. As such it follows an unknown distribution that has no structure.

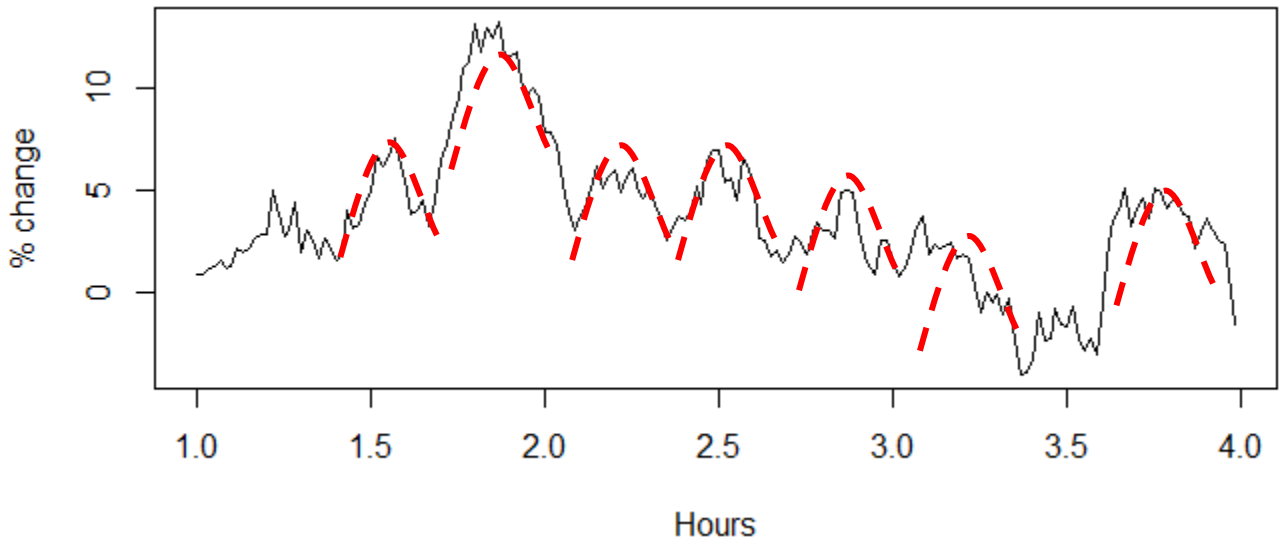


It is impossible to predict whether the noise time series will go up or down, as there is no structure. We may be fooled in seeing patterns in the noise, but there are none! Typically once more sample is collected this becomes clear → Noise cannot be forecasted!

Note that no structure means NO DYNAMICS, NO TREND and NO SEASONALITY in the noise → just randomly fluctuates around zero.

What is randomness?

Let us consider the following example. We collect a time series of stock prices every minute:



Is there a systematic pattern there? Sure, there is a repeating 45 minutes cycle

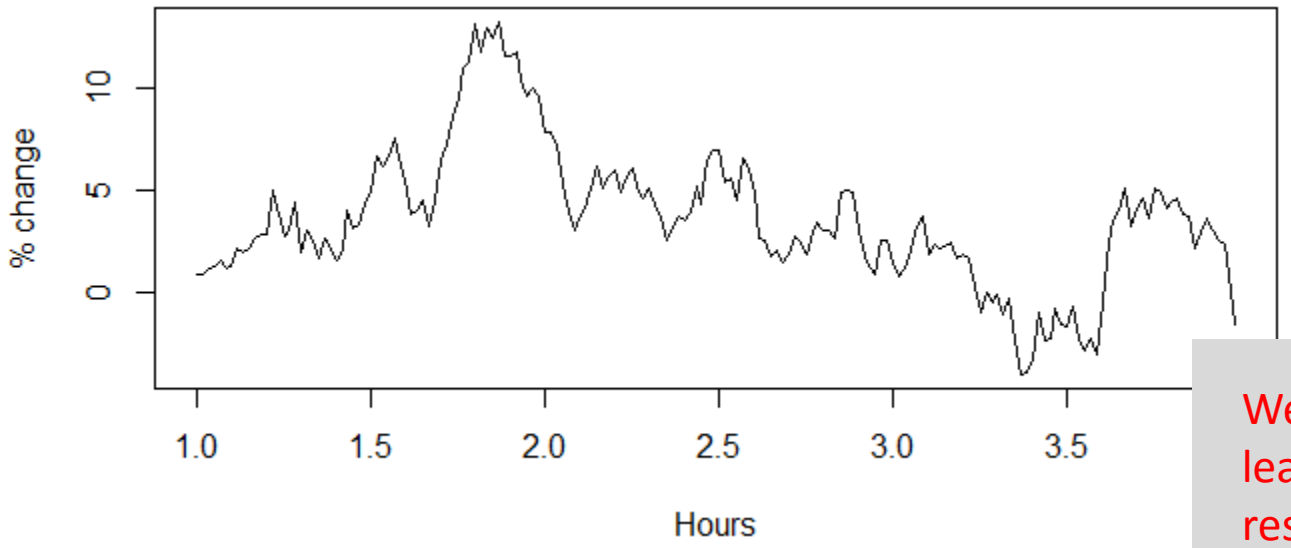
No! It's random

How do I know it **is** random?

a) I created it so!
b) I need to force the "pattern" to fit the data. Why am I changing its level? Why am I leaving gaps between cycles?
c) I focus on only what I can "explain" and I am not being objective.

What is randomness?

In fact, this:



... has as much structure, as this:

... is a rabbit!

We have to learn to be resistant to mistaking randomness for structure

Controlling uncertainty

Then for forecasters the key question is:

"What additional information can we source to reduce the uncertainty?"

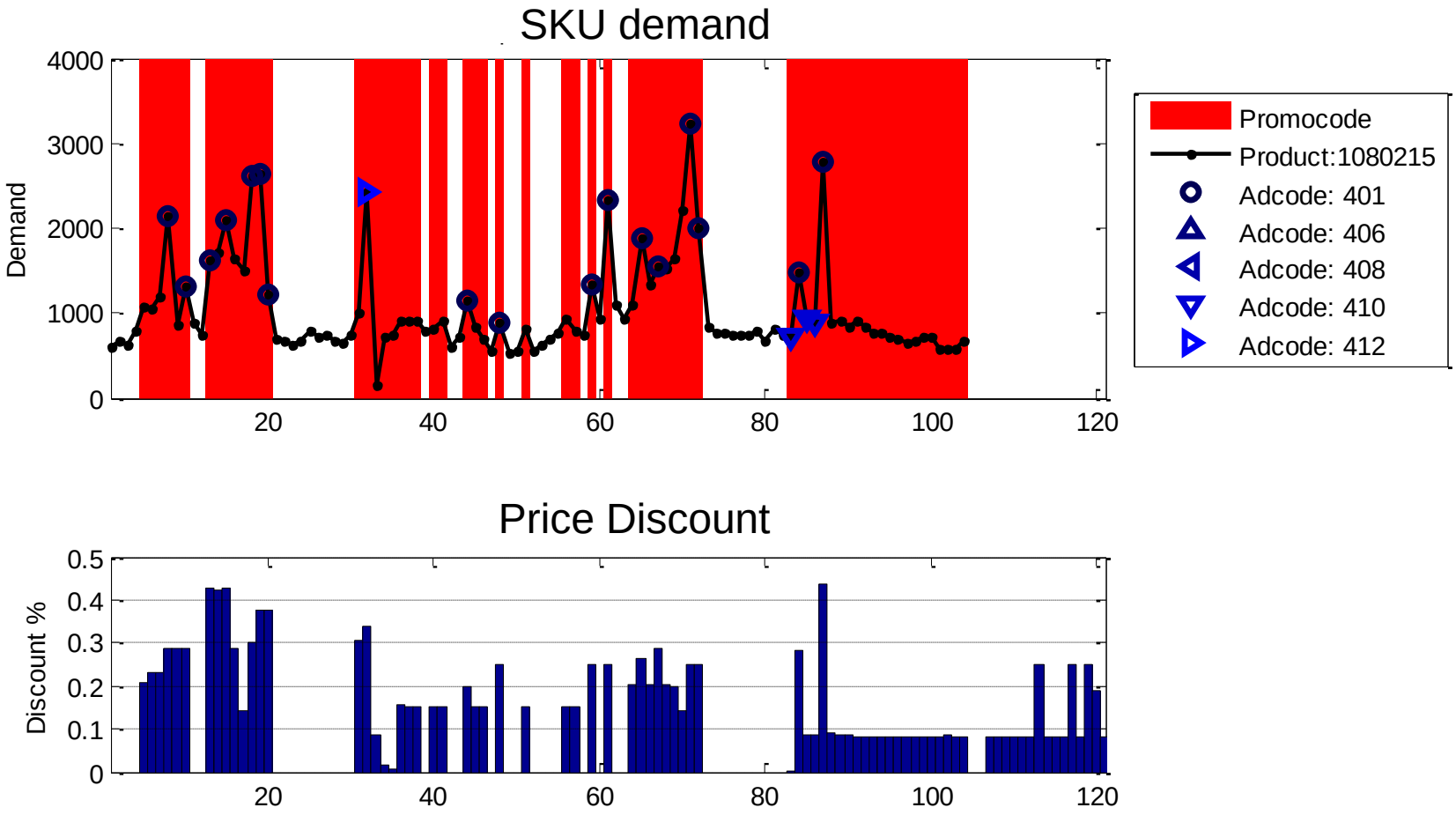
We will investigate this in the context of **promotional modelling**:

- Advances in promotional models
- New sources of information

Outline

1. ~~What is forecasting?~~
2. ~~Forecasting, Decisions & Uncertainty~~
3. Promotional forecasting
4. New sources of information

What is the promotional modelling problem?



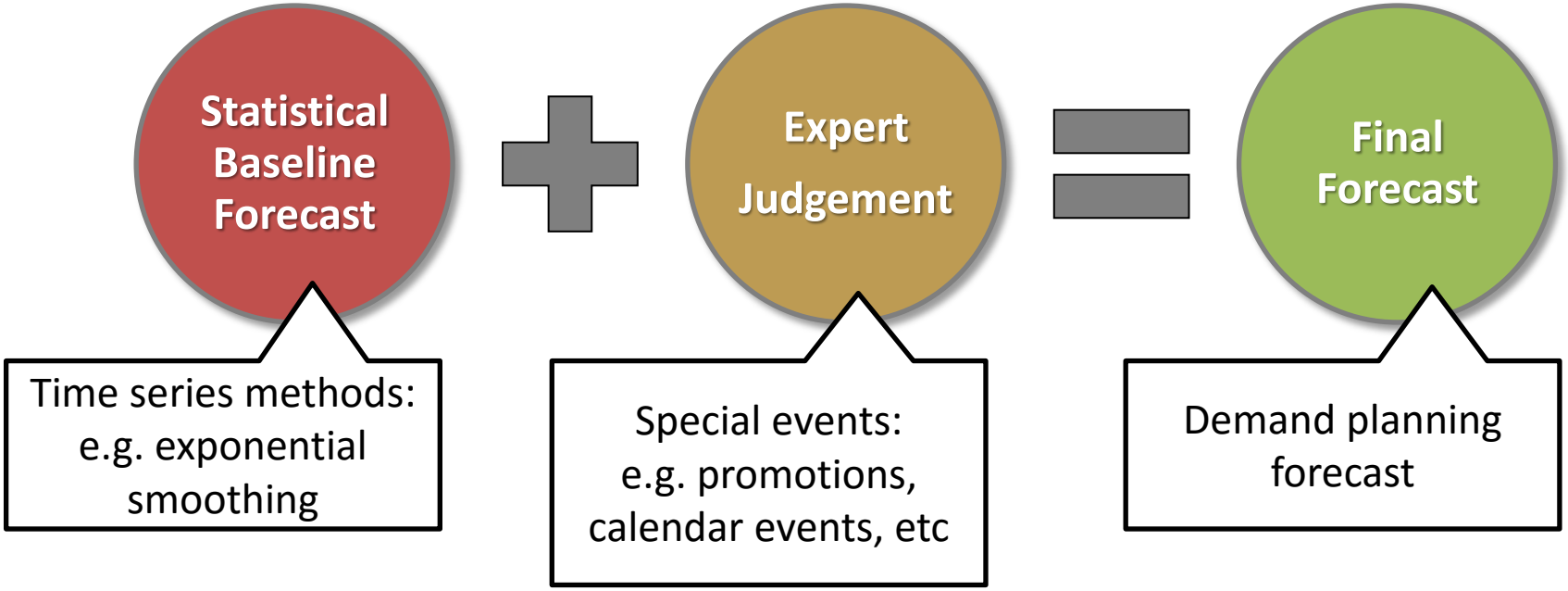
What is the future demand?

Normal demand?

With promotion(s)?

A practical approach

- Use statistics for structured tasks that can be automated
- Use human expertise for events

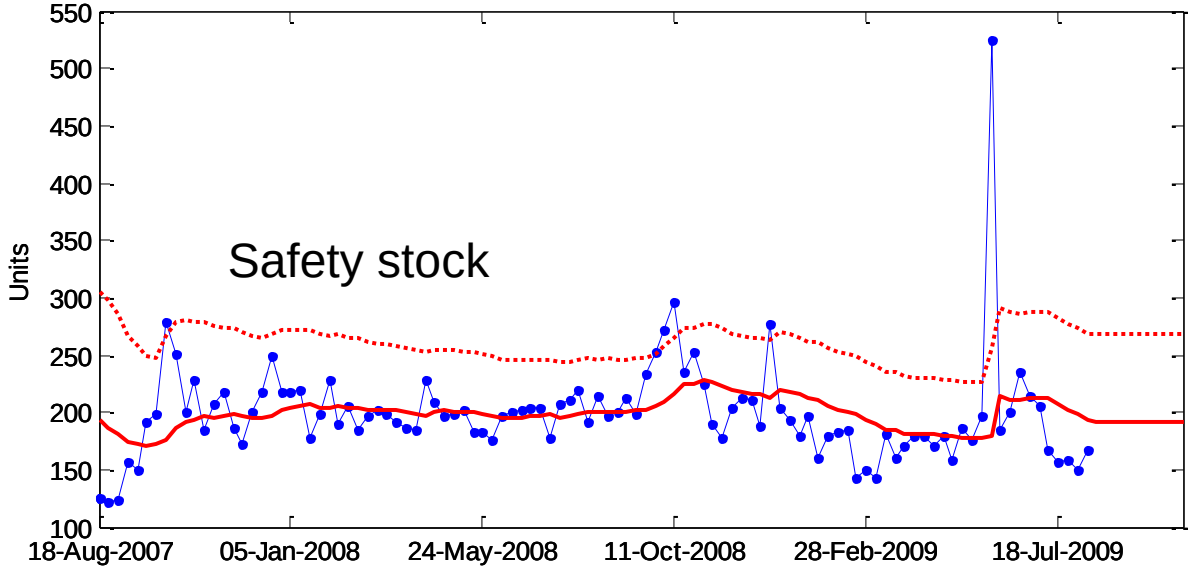


Promotional and advertising activity is one of the main reasons for adjusting statistical forecasts

Promotional Model Basics

What is the best forecast for this time series?

Out-of-stock: Promotion wasted



Overstocking erroneously

No additional information?

Use univariate modelling;
e.g. exponential smoothing

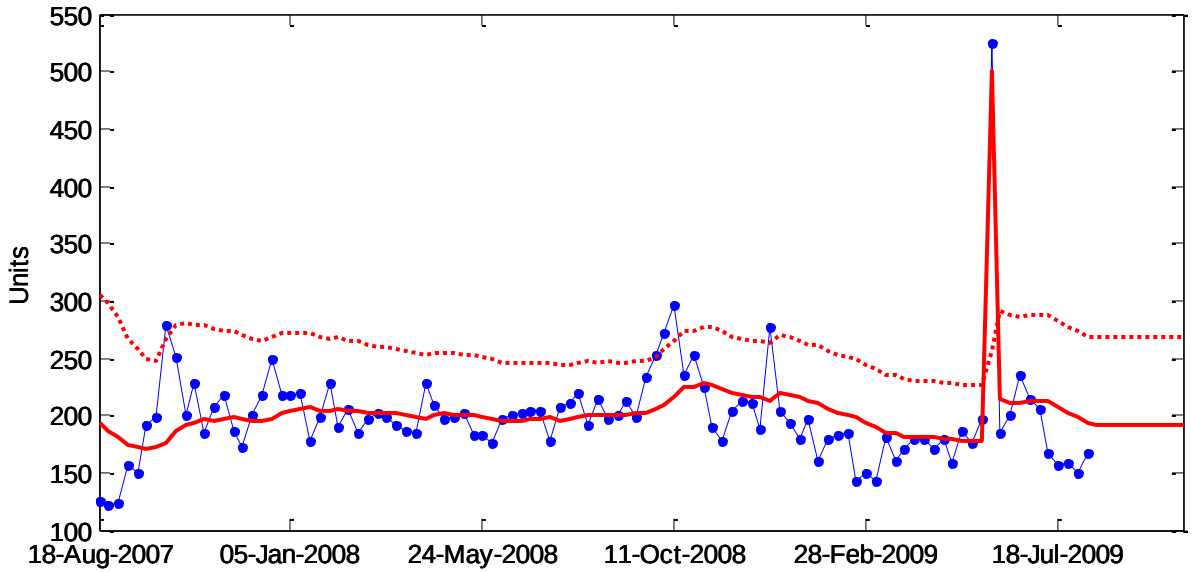
Lets assume a company goes on with this forecast
→ Take a look at the inventory side of things

(Also, how do you forecast the next promotion?)

Promotional Model Basics

What is the best forecast for this time series?

Enough stock, but...



Safety stock
Wrong!

Unstructured information
Use univariate modelling + judgemental adjustments (S&OP meetings = planners + marketers)

Promotional modelling requires building formal statistical models that capture promotions in their forecasts (systematically → elasticities = decision making) and therefore provide correct safety stocks

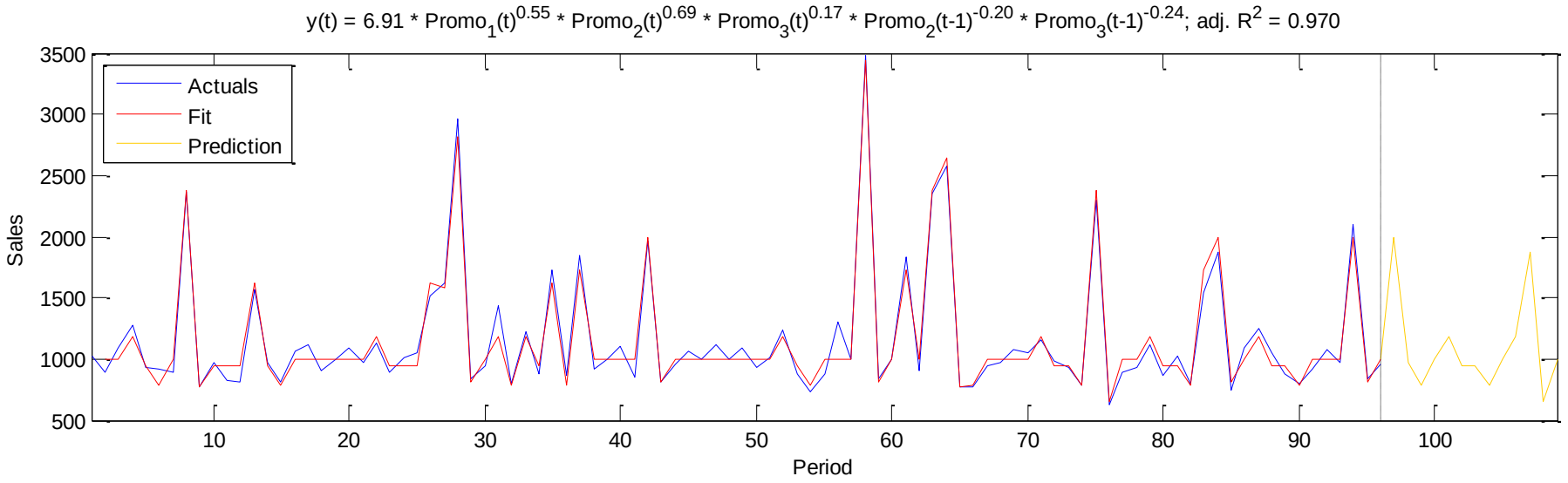
Promotional Model Basics

If the inventory does not account for the uncertainty (due to events/promotions) correctly, then…

A statistical approach

We can build a forecast using statistical models, enhanced by:

- Promotions
- Price
- Store
- Events
- Competition
- …



A statistical approach

What can go wrong? **A lot!**

- Effects across SKUs, categories, …
- Cannibalisation, how to identify it?
- Multiple effects → multicollinearity (i.e. things happening together and then it is difficult to identify which causes what)
- Modelling with limited data (or no data!)
- Scalability → run the model for multiple time series within reasonable times

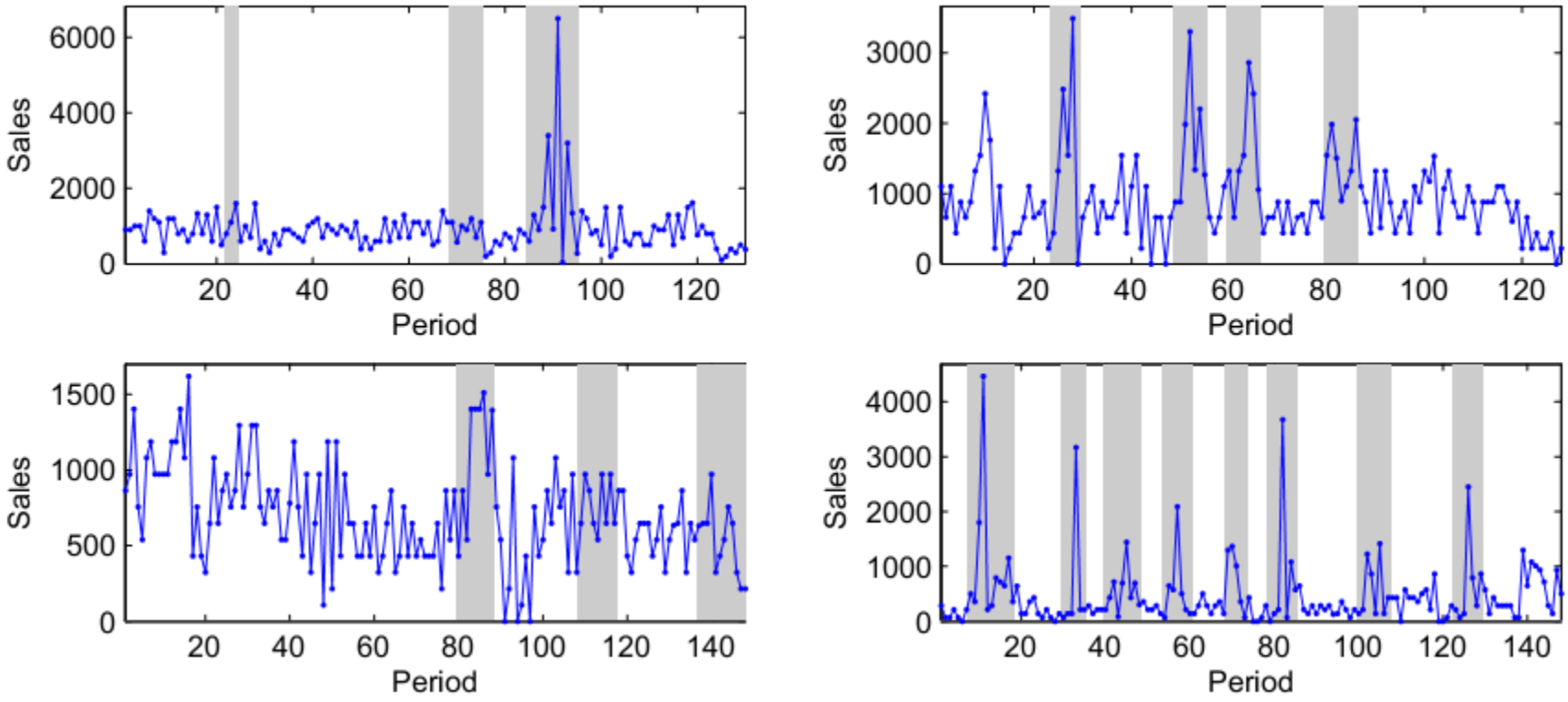
Promotional Forecasting: Case Study

Used a subset of 60 SKUs:

- Sales
- One-step-ahead system forecasts (SF)
- One-step-ahead adjusted or final forecasts (FF)
- Promotional information:
  1. Price cuts
  2. Feature advertising
  3. Shelf display
  4. Product category (22 categories)
  5. Customer (2 main customers)
  6. Days promoted in each week

Data withheld for out-of-sample forecast evaluation
Some products do not contain promotions in the parameterisation (historical) sample

Promotional Forecasting: Case Study



Periods highlighted in grey have some type of promotion

Promotional Forecasting: Case Study

The advanced promotional model contains the following elements:

- **Principal component analysis of promotional inputs**: The 26 promotional inputs are combined a new composite inputs that are no longer multicollinear → Only a few new inputs are needed now, simplifying the model.

- **Pooled regression**: When an SKU does not have promotions in each history, we pool together information from other SKUs (the product category information is available to the model) to identify an average promotional effect. When enough promotional history for an SKU is available we do not pool information from other products

- **Dynamics of promotions**: Carry-over effects of promotions, after they are finished, are captured

- **Dynamics of the time series**: Demand dynamics are modelled for both promotional and non-promotional periods.

Some interesting examples

Let's assume that we have 3 different products with promotions and a 4th that is new, without any promotional history. All products have promotion type A and some promotion type B.

Product I	Promo A	Promo B
144	0	0
1032	1	1
1046	1	1
1128	1	1
997	1	1
111	0	0
118	0	0
156	0	0
123	0	0
159	0	0
138	0	0
131	0	0
	1	1
	1	1
	0	0

Multicollinearity

Product II	Promo A	Promo B
134	0	0
184	0	0
120	0	0
161	0	0
771	0	1
495	1	0
711	0	1
522	1	0
507	1	0
482	1	0
180	0	0
175	0	0
	0	1
	0	1
	0	0

Fine

Product III	Promo A
159	0
420	1
511	1
577	1
530	1
145	0
129	0
112	0
192	0
113	0
128	0
197	0
	0
	0
	0

A prolonged promotion seen only once

Product IV	Promo A	Promo B
155	0	0
153	0	0
180	0	0
	0	1
	1	1
	1	1

No promotional history

Modelling tricks

With pool regression, instead of fitting four models, we pool the datasets and fit only a single one.

Sales	Promo A	Promo B
144	0	0
1032	1	1
1046	1	1
…	…	…
134	0	0
184	0	0
120	0	0
…	…	…
159	0	0
420	1	0
511	1	0
…	…	…
155	0	0
153	0	0
180	0	0

Pooled estimation lifts multicollinearity

Product I	Promo A	Promo B
144	0	0
1032	1	1
1046	1	1
1128	0	0
…	…	…

Product II	Promo A	Promo B
134	0	0
184	0	0
120	0	0
161	1	0
…	…	…

Product III	Promo A
159	0
420	1
511	1
577	0
…	…

Product IV	Promo A	Promo B
155	0	0
153	0	0
…	…	…

and provides estimates for products with no history

Now we estimate the model on the pooled data. Sometimes it is helpful to include indicator dummies for each different product.

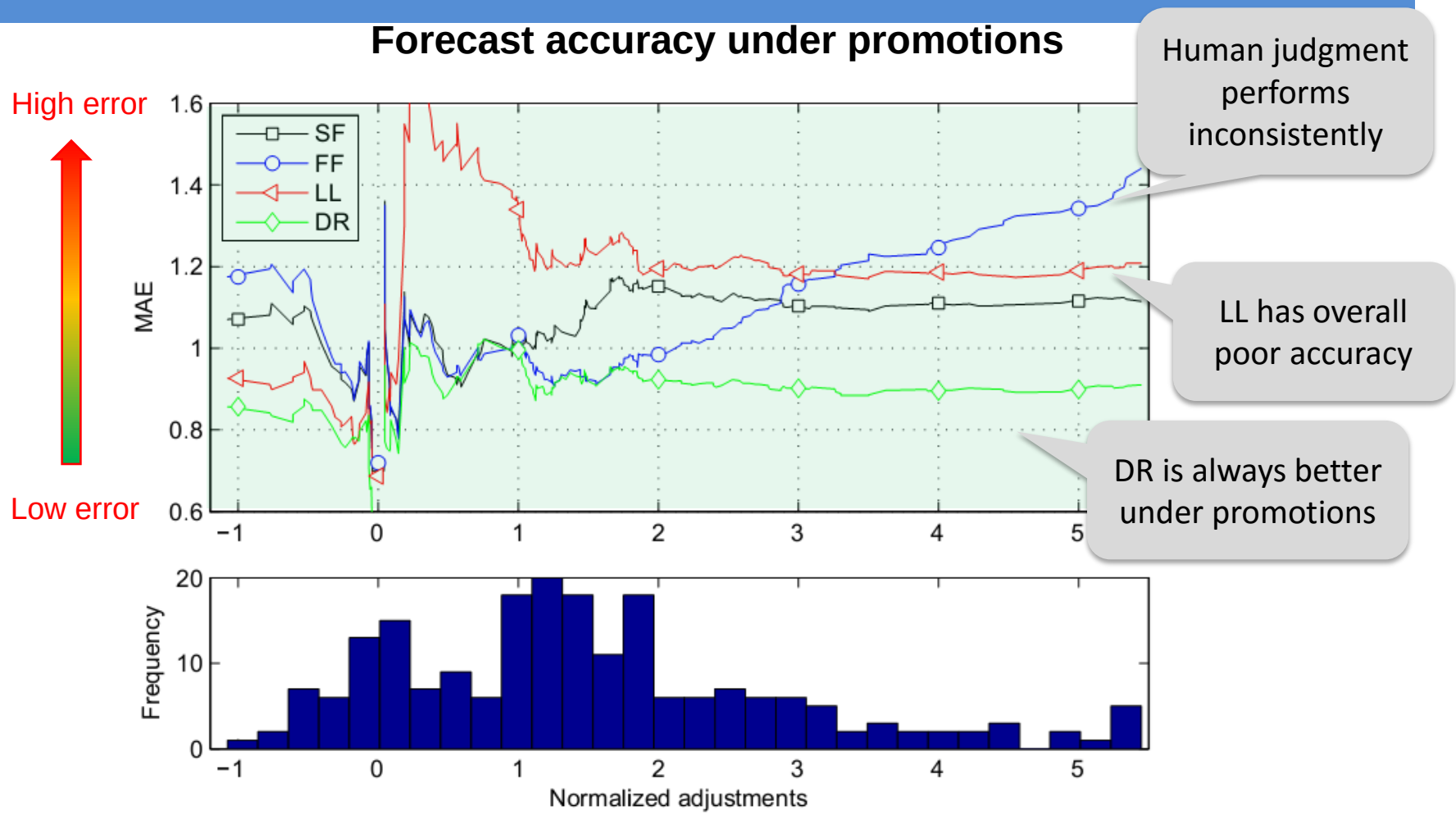
Promotional Forecasting: Case Study

A series of benchmark models are used to assess the performance of the proposed promotional model

- System Forecast (**SF**): The baseline forecast of the case study company
- Final Forecast (**FF**): This is the SF adjusted by human experts in the company
- **Naïve**: A simple benchmark that assumes no structure in the demand data
- Exponential Smoothing (**SES**): An established benchmark that has been shown to perform well in supply chain forecasting problems. It cannot capture promotions
- Last Like promotion (**LL**): This is a SES forecast adjusted by the last observed promotion. When a promotion occurs the forecast is adjusted by the impact of the last observed promotion

Note that only FF and LL can capture promotions from our benchmarks

Promotional Forecasting Case Study

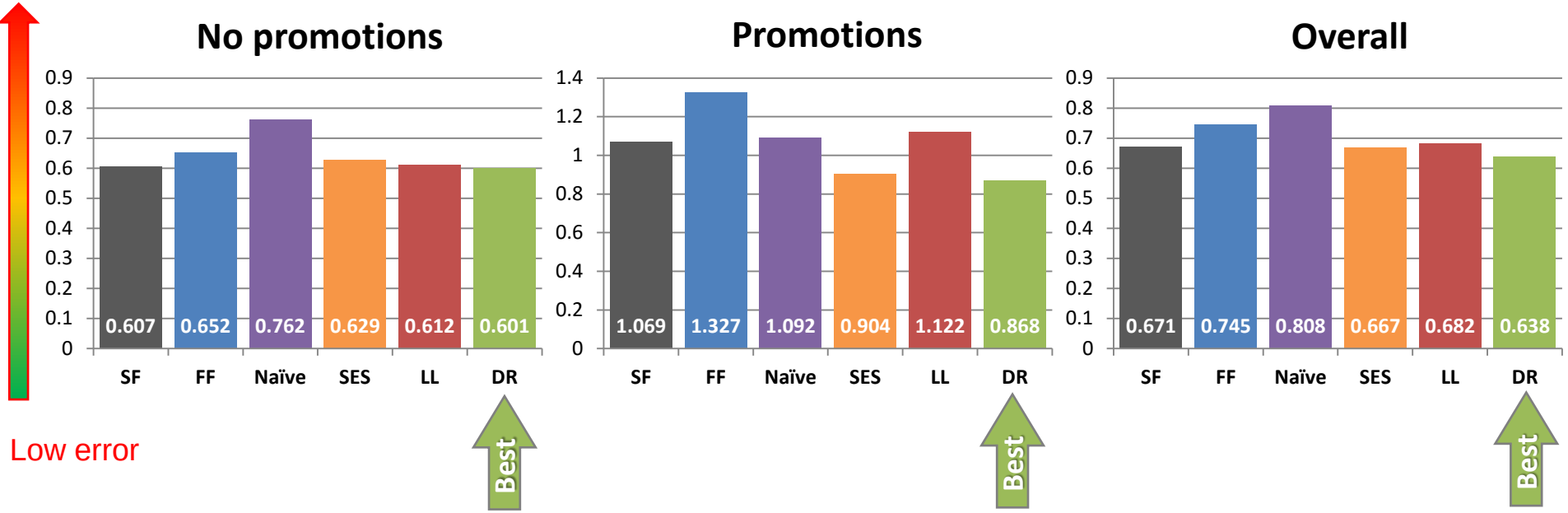


DR is more accurate than human experts and statistical benchmarks in promotional periods

Promotional Forecasting Case Study

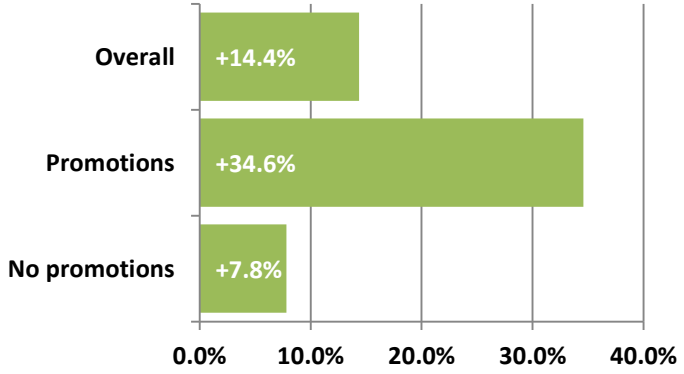
High error

Low error



- Advanced promotional model is consistently the most accurate
- Substantially outperforms current case study company practice (FF)
- Major improvements during promotional periods

Improvement of DR over FF

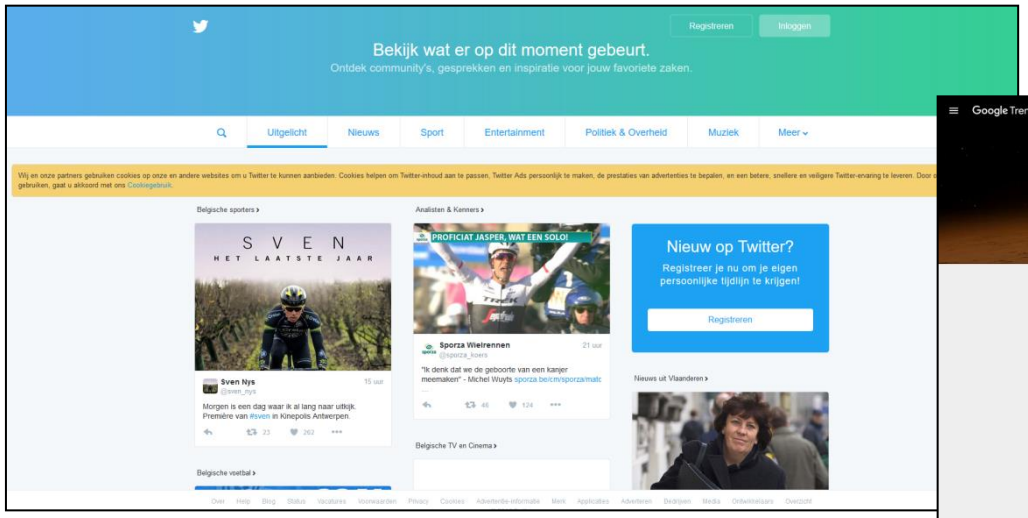


Outline

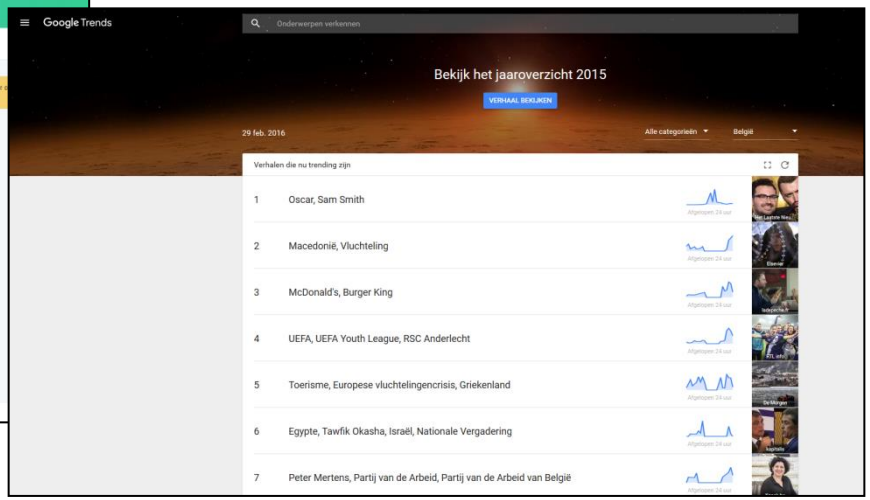
1. ~~What is forecasting?~~
2. ~~Forecasting, Decisions & Uncertainty~~
3. ~~Promotional forecasting~~
4. New sources of information

New sources of information

Twitter



Google trends



Online stores and news outlets

Going one step further: consumer sentiment

Tracking market and economic development under financial uncertainty (or crisis) is notoriously difficult. Let us try a difficult task:

- Can we predict better the movements of the Greek economy looking at online information?

The focus for **nowcasting** is to `predict' the current state of a variable, often the GDP, when it is currently unknown, due to substantial publication lags of itself or its covariates. Also, its publication frequency may be lower than our decision making.

For a company such nowcasts can be important for strategic decision making.

Creation of variables

It is fairly easy to either download online text, or access online using a crawler. Let us assume that we have a database containing online textual content.

We can mine the text for keywords of interest, however just the frequency of word occurrence is not useful, as it is out of context.

Instead we can look for pairs of words. The hypothesis is:

`If two keywords are close by in a text (not necessarily in the same sentence) then the author implicitly or explicitly places some semantic association on them'

1.Define Θ keywords

These can be identified using experts or mined tag clouds

2.Find all keyword pairs

There are $p = \binom{\Theta}{2}$ combinations, but maybe not all are meaningful

3.Find keyword distances & count frequency per period

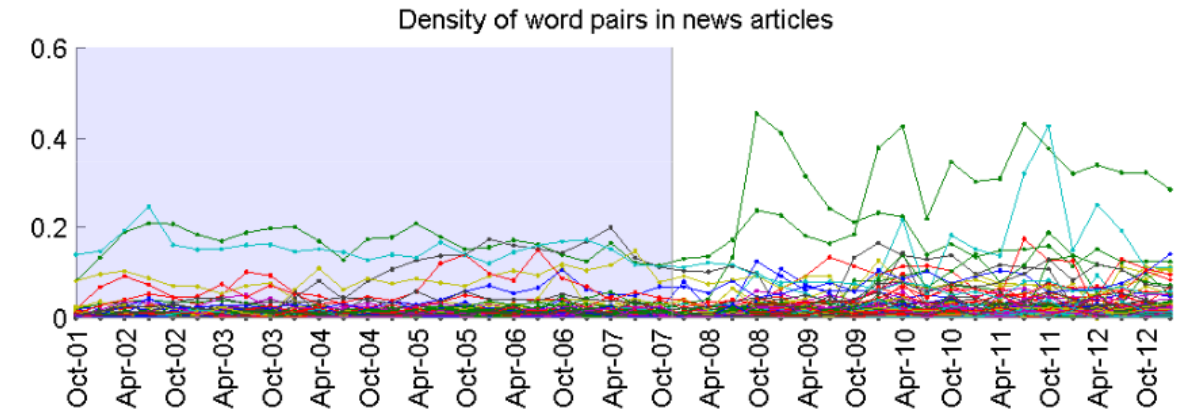
Find location of keywords and then intra-pair distances

...

Creation of variables

Keywords for case study

austerity	drachma	income	recovery	treasury
budget	economy	memorandum	stock exchange	unemployment
CDS	euro	parliament	surplus	
crisis	GDP	prime minister	tax	
deficit	IMF	recession	troika	



Top correlations with GDP growth

Keywords	Correlation
economy - treasury	-0.757
recession - deficit	-0.745
euro - stock exchange	0.730
IMF - Troika	-0.702
euro - treasury	-0.697

Selection of variables

A key problem for building statistical models using such variables is the identification and selection of the appropriate ones:

- Large number of variables causes the `fat regression' problem → more variables than datapoints
- Spurious correlations and multicollinearity → weak forecasting models
- Selection a small subset of variables → better estimation/performance, more transparency → less chance of overfitting

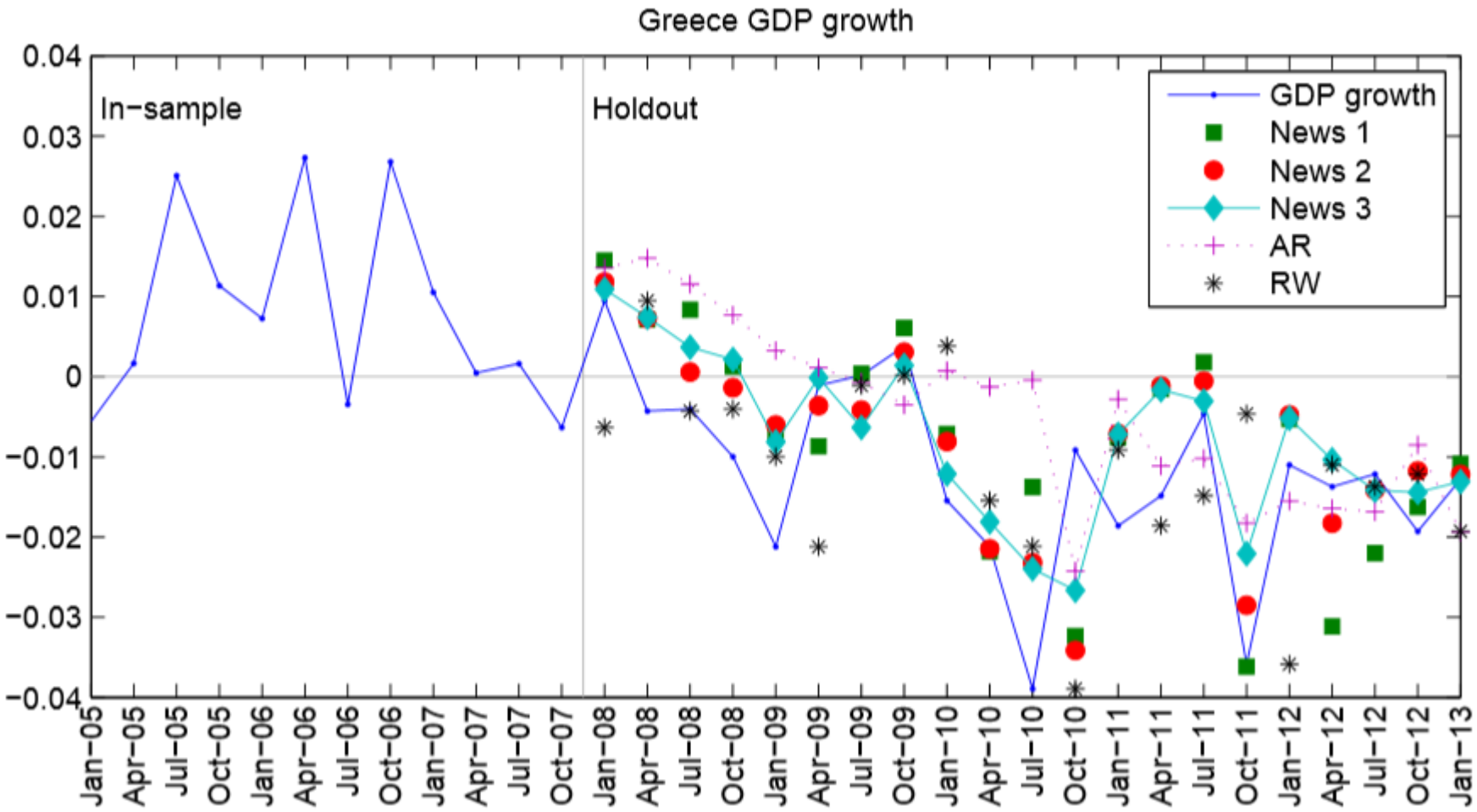
Two popular approaches in resolving this issue is regularisation (e.g. LASSO regression) and Principal Components Analysis.

The first tries to build parsimonious models by penalising heavily the inclusion of new variables

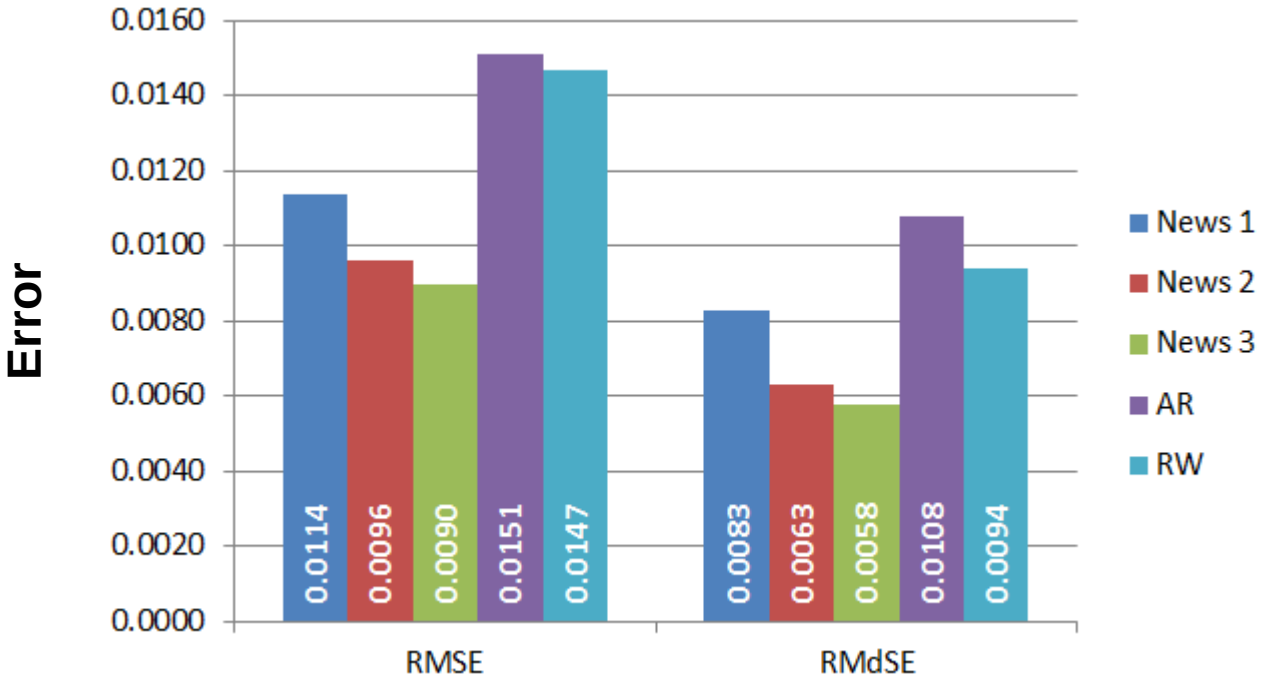
The second tries to combine the variables in a way that the maximum amount of variance is explained by only a handful of composite variables.

Case study

Having identified and selected variables we can create statistical now-/forecasting models.



Case study



Using online information we were able to nowcast more accurately the movements of the economy, and crucially predict the existence and timing of the observed triple dip in the growth of GDP, which had been notoriously difficult.

Similar techniques can be applied to increase the information base for various business forecasts, such as market share, evolution of segments, demand forecasting, etc.

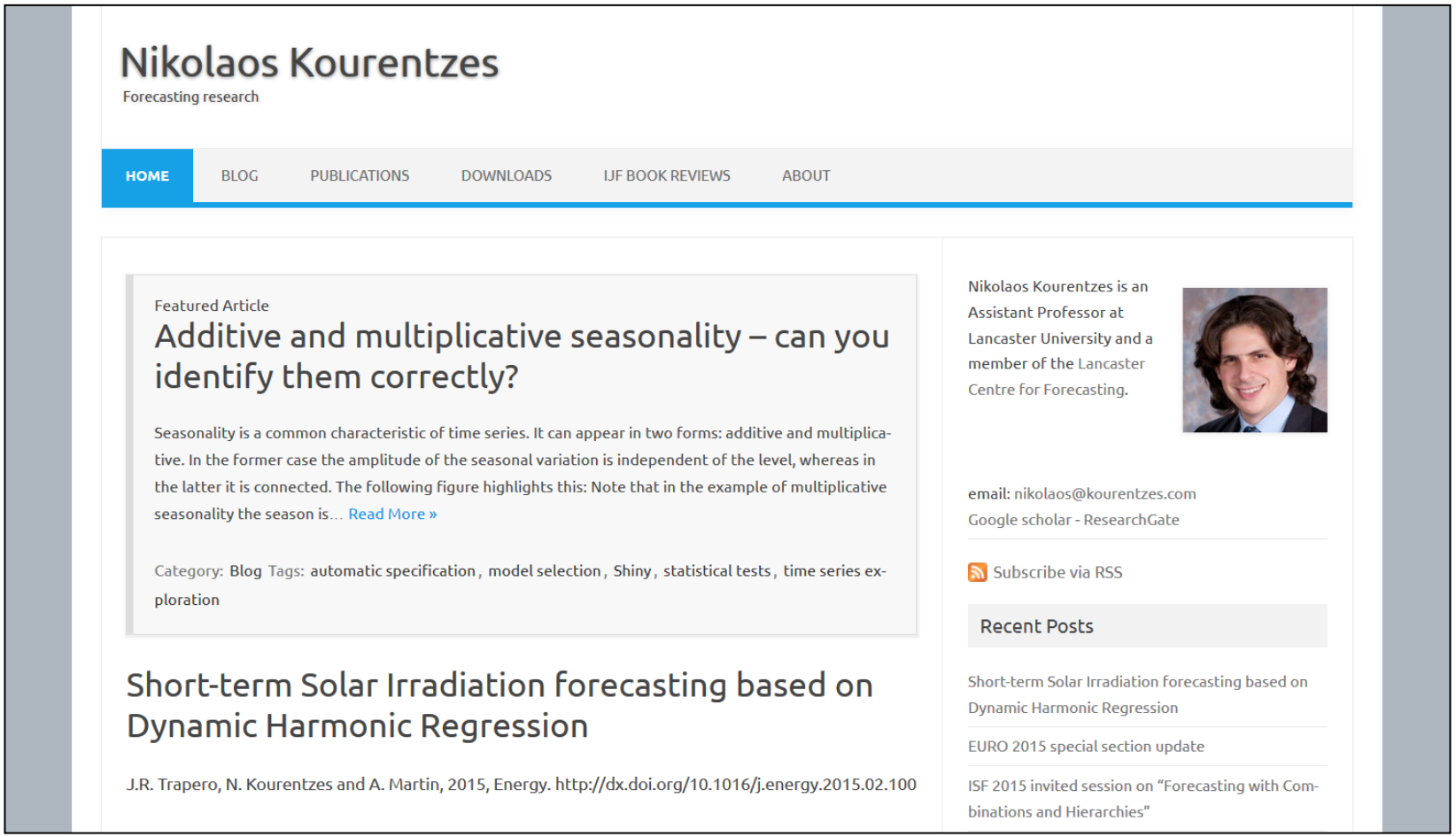
Take-aways

- Forecasting vs target: keep them separate and use them!
- Model and manage uncertainty: you improve your forecast accuracy → so what?
- It is not the point forecast that matters, but the uncertainty around it.
- Forecasting for a retailer, many complex problems
  - Noisy & limited data
  - Multiple lines and categories
  - Statistical complications

  → Recent advances in modelling can address effectively many of these.
- Use of innovative sources of information: listen to your customers

Forecasting Blog

http://nikolaos.kourentzes.com



Thank you for your attention!

Nikolaos Kourentzes
Lancaster University Management School
Lancaster Centre for Forecasting - Lancaster, LA1 4YX
email: n.kourentzes@lancaster.ac.uk
URL: http://nikolaos.kourentzes.com

www.forecasting-centre.com/

Lancaster Centre for Forecasting | Lancaster University Management School