# Judgmental Adjustments of Computer-based Forecasts
# Are they beneficial?
# Can they be improved?

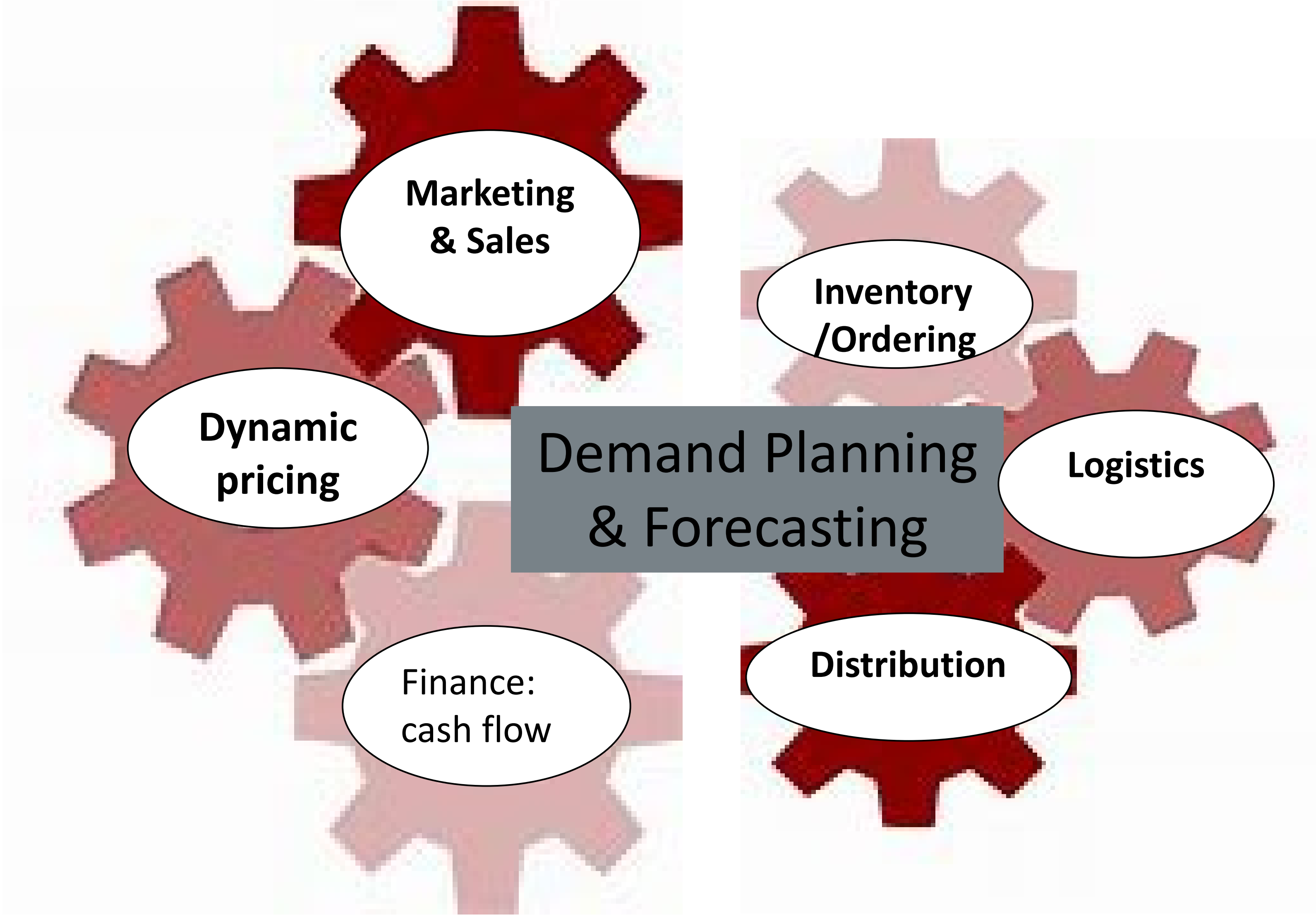
Robert Fildes & Paul Goodwin

with Shari DeBaets

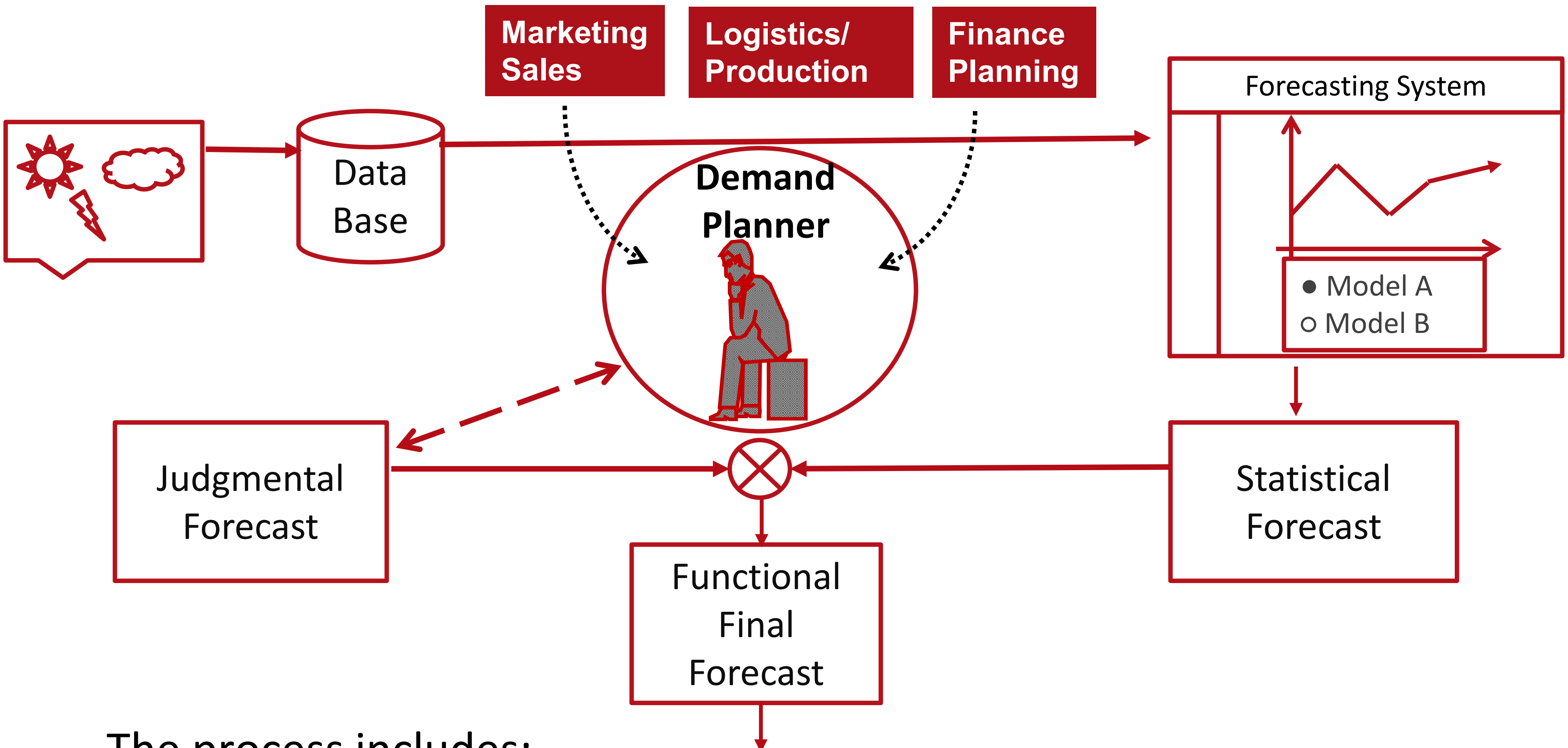
Lancaster University Centre for Marketing Analytics and Forecasting; University of Bath; OU Netherlands

Marketing Analytics & Forecasting | Lancaster University Management School

# The functional departments and demand forecasting



# Forecasting process in S&OP




The process includes:

- Selecting a model, parameterizing it to calculate the Statistical forecast
- Information cues from sales, market research, logistics, finance
- Incorporated into a final forecast

**Effective Judgment is a key component for integration**

# The Process of Adjustment:
## the use of information

- Information available: cues, e.g. promotion, weather
- Computer-based (statistical or machine learning) forecast

Demand planner’s tasks: to add value (Forecast Value Added)

- ✓ Chooses a statistical forecast or model
- ✓ Chooses (or not) to adjust the statistical forecast
- ✓ Selects and weights information cues
- ✓ Decides on the direction and size of the adjustment

# Key Questions for Demand Forecasters

1. How likely are adjustments to increase or reduce forecast accuracy?
2. How large are the resulting improvements or reductions in accuracy?
3. What factors are associated with decisions to adjust system forecasts?
4. In what ways can adjustments damage or improve accuracy?
5. Is it possible to improve accuracy by debiasing judgmental adjustments and/ or changing the forecasting process?

CMAF

# Evidence base for findings

- **Surveys of demand planners**

- **Case studies** **-**observing forecasting processes and interviewing participants

- **Experiments** – to test forecasters’ decisions under controlled conditions

- **Field studies –** data collected from companies on computer-based forecasts, adjustments & outcomes.

# Evidence from surveys

Table 1: Survey studies of methods used in practice

| Method | Study | | | | Average |
|---|---|---|---|---|---|
| | A | B | C | D | |
| Judgment alone | 30% | 25% | 24% | 14% | 23% |
| Statistical methods exclusively | 29% | 25% | 32% | 30% | 29% |
| Average statistical & judgment | 41% | 17% | – | 19% | 18% |
| Adjusted statistical forecast | | 33% | 44% | 37% | 38% |
| Sample size | 240 | 149 | 59 | 42 | |

A: Sanders and Manrodt (2003); B: Fildes and Goodwin (2007); C: Weller and Crone (2012); D: Fildes and Petropoulos (2015).

***Even for retailers, adjustments are common***

# Evidence from Case Study

- Pharmaceutical company.
- Commercial software sold on basis of accuracy of both its forecasting algorithms and the ease which adjustments could be made.
- Many adjustments made within algorithm (e.g., changing parameters)
  - ✓ judgmental adjustment by the ‘back door’.
- So-called computer-based forecasts then presented at one of 17 monthly review meetings for possible adjustments.

A time consuming expensive process – so why?

✗ No improvement in accuracy!

✓ But managers welcomed having control over final forecasts and they could demonstrate their expertise on their markets at review meetings.

# Evidence from Experiments: Fildes et al. (2015)

- Participants told sales promotion will take place next month and presented with:


- product details;
- statistical forecast that excludes promotion effects;
- past demand history including effect of past promotions and past forecasts;
- told typical promotions increase demand by 50%
- qualitative information e.g., weather forecasts, promotion expenditure, rumours, hype.

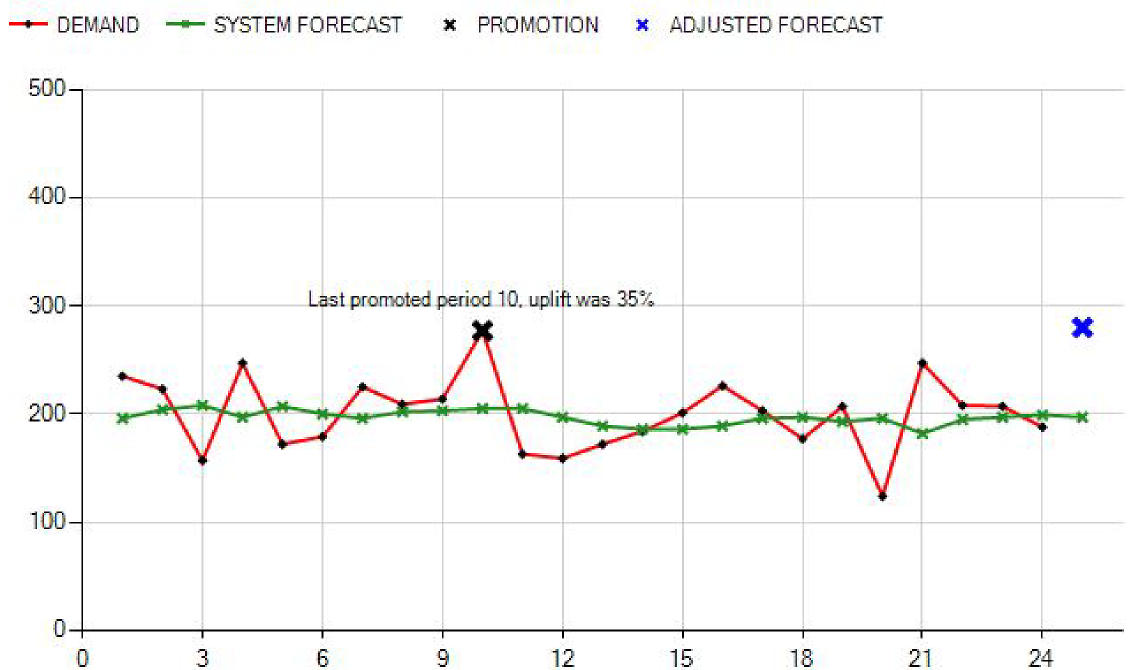
# Evidence from Experiments: Fildes et al. (2015)

**Experimental Forecasting System** Logout

| | | % of System Forecast |
|---|---|---|
| System Forecast: | 197 | |
| Current Forecast | 280 | 42 |
| Adjusted Forecast | 280 | 42 |

Next Series

**LIVE EXPERIMENT SERIES**




Series number: 8 of 14

**PLEASE TICK RELEVANT INFORMATION ABOUT PERIOD 25 PROMOTION**

☐ Product Details

Health care product which has been long established; sold into many supermarkets, chemists

☐ Weather

Weather conditions in the Midlands where this product is popular, should help to boost sales substantially

# Evidence from Experiments: Fildes et al. (2015)

- Participants gave too much weight to effect of previous promotion -a single outcome –ignoring the typical (base rate) uplift of 50%.
- They were distracted by extraneous qualitative information like hype and rumours .
- Similar result found in  experiments in Sroginis et al. (2023).

# Evidence from Field Study

- Earlier evidence used disparate methods so hard to compare

*We aimed to summarize and extend these studies with a robust methodology*

- Data obtained from 6 studies comprising over 146,000 computer-based forecasts, adjustments and outcomes from 15 supply-chain companies/business units.
- Industries: pharmaceutical products, retail distribution, household cleaning products , industrial coating, food manufacturing, beer production.

# Measuring Forecast Value Added (FVA)

- Measuring accuracy improvement

$$\text{Geometric mean of } \frac{\text{Mean Absolute Error after Adjustment}}{\text{Mean Absolute Error of Computer Forecast}}$$

If <1 then adjustments have typically improved accuracy
If >1 then adjustments have typically worsened accuracy

E.g. 0.8 = 20% improvement in accuracy
1.4 = 40% worsening in accuracy

- Similar measure used to detect improvements in bias (i.e., a tendency to forecast too high or too low).

# Evidence from Field Study

- How likely are adjustments to increase or reduce forecast accuracy?

| Dataset | No. of skus | % SKUs where bias improved | % SKUs where FVA improved |
|---|---|---|---|
| **Set 1** | 582 | 40.4 | 62.4 |
| **Set 2** | 759 | 77.5 | 16.5 |
| **Set 3** | 1100 | 58.3 | 49.8 |
| **Set 4** | 1217 | 44.4 | 55.1 |
| **Set 5** | 45 | 48.9 | 53.3 |
| **Set 6** | 32 | 37.5 | 84.4 |
| **Overall** | **3735** | **53.0** | **52.6** |

# Evidence from Field Study

- How large were changes in accuracy and bias?

Overall results:

| Adjustment Direction | No. of observations | % | Relative Bias | Relative Accuracy |
|---|---|---|---|---|
| Positive | 46151 | 53.7 | 1.002 | 1.088 |
| Negative | 39802 | 46.3 | 0.623 | 0.845 |
| Overall | 85953 | 60.1 | 0.940 | 0.982 |
| Unadjusted | 57117 | 39.9 | n.r. | n.r. |

# Evidence from Field Study

What factors are associated with decisions to adjust system forecasts?

- ✓ Previous direction of adjustment
- ✓ Change predicted by computer-based forecast
- ✓ Size of computer-based forecast
- ✓ Last error of computer-based forecast
- ✓ Last error of adjusted forecast
- ✓ SKU and forecaster characteristics

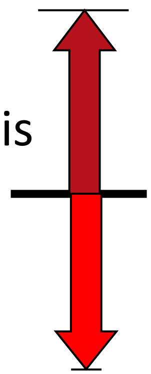
But **extra** information

available potentially only to the forecaster had a relatively weak effect.

*It is this unaccessed information that holds the key to success.*

# In what ways can adjustments damage accuracy?

- Two ways
  - adjusting in the wrong directions.
    - e.g., adjusting upwards when a downward adjustment is needed

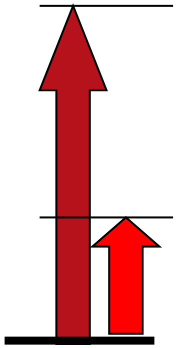



  - an excessive adjustment in the right direction.

e.g. Computer forecast = 100

Actual Outcome = 150

Adjusted forecast = 201



# Evidence from Field Study

| Adjustment Direction | Problem | % of All Adjustments | Relative Accuracy | Relative bias |
|---|---|---|---|---|
| Positive | Excessive | 9.18 | 3.82 | 3.40 |
| | Wrong direction | 22.51 | 2.51 | 2.12 |
| | | | | |
| Negative | Excessive | 4.43 | 3.62 | 3.49 |
| | Wrong direction | 13.55 | 2.10 | 1.76 |

# Evidence from Field Study

- Are larger adjustments better or worse?

| | | Adjustment size % | | | | | |
|---|---|---|---|---|---|---|---|
| Direction | | <10 | 10 to <50 | 50 to <100 | 100 to <250 | 250 | Overall |
| Positive | % that improved accuracy | 45.5 | 44.9 | 46.6 | 44.1 | 48.6 | 45.3 |
| | % where bias reduced | 45.4 | 45.8 | 48.8 | 50.5 | 62.4 | |
| Negative | % that improved accuracy | 53.8 | 59.5 | 67.4 | 47.8 | | 54.7 |
| | % where bias reduced | 56.9 | 62.8 | 69.4 | 33.3 | | |

✓ *For negative adjustments, improvements with size*

- ✗ *Apart from those making the final forecast zero*

# Evidence from Field Study

- Are larger adjustments better or worse?

| | | | Adjustment size % | | | |
|---|---|---|---|---|---|---|
| Direction | | <10 | 10 to 50 | 50 -100 | 100-250 | >250 |
| Positive | Relative accuracy | 1.01 | 1.09 | 1.09 | 1.22 | 1.02 |
| | Relative bias | 1.05 | 1.13 | 1.05 | 1.17 | 0.80 |
| Negative | Relative accuracy | 0.95 | 0.80 | 0.49 | 1.39 | |
| | Relative bias | 0.91 | 0.62 | 0.41 | 1.68 | |

# Evidence from Field Study

Is it possible to debias adjustments

- ✓ To improve accuracy?
- ✓ To identify how to make improvements?

Three approaches:

1. Predicted forecast error = f(computer forecast, adjustment)

   [Optimal weight model]

2. Predicted forecast error =f(computer forecast, previous adjustment, Change, Previous error of adjusted forecast)

   [Full model]

3. 50:50 average of computer forecast & adjusted forecast

# Evidence from Field Study

- Debiasing adjusted forecasts:
  - Out-of-sample FVA compared to System forecast

Overall, adjustment damages for positive adjustments
improves for negative adjustments

✓ But individual business units/ companies make adjustment work consistently

Too much weight given to 'expert adjustments' –debiasing works
Full model debiasing delivers ≈ 20% improvement.

# Implications I

- Possession of important information not available to the computer forecast is the key justification  for judgmental adjustment.
- Yet our research suggests that this plays a relatively small role in motivating decisions to adjust.
- Forecasters appear to make adjustments based on information with low predictive value
  - previous adjustments, previous errors (which may be a result of noise), rumours and hype.

# Implications II

- Upwards adjustments fare badly suggesting forecasters are over optimistic or are responding to asymmetric loss
  - Forecasts as targets?
- Debiasing models can have a useful role in mitigating this bias.
  - But demand planners game-play any constraints
- Need for the forecasting process (S&OP) and software to focus on the significant cues relating to future events

# Final thoughts

- Models robust across time, company processes and software
  - ✓ Applicable to other decisions in Operations, e.g scheduling
- No theoretical models available to explain adjustment
  - ✗ Irrational expectations!
  - ✗ Organizational or Psychological models (Kahneman and Tversky: biases, framing of the demand planning task)
- ***Mixed methods*** vital to understand when and why adjustments are valuable
- But few company case studies/ action research
  - A failure of the interdisciplinary research paradigm

# Research opportunities

- Case observation/ software design/ action research to discover how to overcome the consistent mis-understanding of S&OP information
- The link to effective decisions (e.g. ordering)
  - ✓ When does the decision context matter?
- The many other organizational decisions where model based advice is responded to:
  - ✓ Inefficient processing of cues including model recommendations
  - ✓ When is a model trusted?
  - ✓ And does AI/ ML change the equation?

# Thank you for your attention!

# Q&A?!

Robert and Paul