Forecasting, uncertainty and managing service level

Professor Robert Fildes
Distinguished Professor and Director,
Lancaster University Centre for Forecasting

With thanks to John Boylan, Xi Chen and Maulana Syuhada
1. Early beginnings
2. My Assumptions!
3. Demand uncertainty and forecasting models
4. What do we know about forecasting model accuracy?
5. Modelling the supply chain
6. Three examples
   i. Intermittent demand
   ii. Effects of uncertainty and forecasting error on supply chain
   iii. Integrating planning models with forecasting model choice – which matters when
7. Conclusions
• Forecasting must be considered in a ‘real’ problem context
  – Lead time, information, value (loss function), comparative evaluation

• Service level planning must be considered in a ‘real’ problem context
  – Supply chain features, demand uncertainty, forecast error

**Inventory planning and forecasting (OR generally) are defined by practice**
• Volatile
• Trend+ seasonal
• Step and level changes
• Outliers
• Intermittent?
• Promotional & Other effects

1. Changing volatility
2. Changing Trend+seasonal
3. Promotions
Researching forecasting - the developing research agenda

• Exponential smoothing (1950s)

• ARIMA (1970s) + intervention models

• State space models (1980s)
  – Single source of error (1990s)

• Computer intensive methods (1990s on)
  – Neural networks
  – Support vector machines
  – ‘fuzzy’ models
Comparative forecasting accuracy

- **The Forecasting Competitions** (Newbold and Granger, 1974; Makridakis and Hibon, 1979; M-competition, 1982; M-3 Competition, 2000)
  - Aim: to identify the ‘best’ forecasting methods
  - + the conditions in which one method outperforms another

- **Characteristics**
  - Many empirical time series
  - Alternative methods
  - Select a data series (from the relevant population)
  - Apply competitive forecasting methods
  - Produce forecasts for different horizons from methods
  - Calculate summary error statistics for each method for each series
  - Calculate error statistics across all the data series
  - Compare error statistics
  - Examine sub-set performance

Which methods win, which lose, when and why?
Conclusions

- Simple model specifications will often outperform complex alternatives
- Damped trend smoothing is on average the most accurate extrapolative forecasting method
- More general methods will not typically outperform constrained alternatives
- Combining leads to improved accuracy
- Problem/context specific methods will outperform general alternatives
- Causal methods will typically outperform extrapolative
- Loss functions matter!
Researching ‘forecasting and inventory’

- Keyword search in WoS
  - highly cited articles 20+ since 2004, 9+ since 2009

- Research priorities
  - Focus of research
  - Demand generation process
  - Forecasting techniques
  - Supply chain structure
  - Methodology
  - Validation
  - Loss functions
  - Conclusions
A good model?

• Fit for purpose
• Comprehensible to the user (Little, 1970)
• Complete on important issues
• Encompass other models
  – structure
  – results
• Black box (input-output validation)
• White box validation
  – micro structure supported by observation/empirical research
  – parameters supported by research
• Robust
  – model and results meaningful with plausible changes in parameters
Supply chain Structure

• Production system complexity
  – Echelons (Retailer(s), manufacturer(s), supplier(s))
  – lead times (fixed?)
  – interrelationships, bottlenecks
  – production reliability
  – back orders
  – frozen interval, planning horizon

• Demand & Forecasts
  – Demand generating process (grounded or not in empirical research)
  – Forecasts optimal (or not) for particular demand models
  – inter-related products
Supply chain Structure II

- Information sharing between supply chain levels
- Production and ordering inventory rules (lot sizing)
- Cost structure and Service Level
- Evaluation Criteria
  - service levels, cost,
  - Trade-off (Gardner, Man. Sci. 89)
  - bullwhip
    - Bullwhip not directly cost/ service relevant
Findings

• Focus of research
  – Bull-whip
  – Effect of forecasting accuracy (most important cause)
  – Evaluation of different forecasting methods
    • Computer intensive methods
  – Information sharing

• Demand generation process
  – usually unrealistic and unjustified
    • Independent demand, no outliers/ promotions, no structural change

• Forecasting techniques
  – Intermittent demand methods, CIS methods
  – Non-optimal for DGP
Findings II

• Methodology: math model + simulation
  – Rarely empirical

• Validation
  – Some parameter sensitivity testing
  – Need to ground the model in organisational reality
    (rarely done)
  – Test against observations (few examples, e.g.

• Loss functions: bullwhip amplification, inventory only
  – Little use of trade off curves

• Conclusions
  – Either obvious (information sharing is of potential value!)
  – Or obviously contingent (e.g. ‘bias improves performance’
    depends critically on assumed supply chain structure)
Bullwhip Effect (BWE)

Key problem in inventory management

BWE $\rightarrow$ Demand variability amplification when moving up the supply chain

\[
\text{Bullwhip Effect} = \frac{\text{var}(D_{\text{Supplier}})}{\text{Var}(D_{\text{ret}})}
\]

Claimed BWE leads to:
1. Excessive stock
2. Poor customer service
3. Increased costs

Note: with perfect supplier forecasting BWE has no adverse consequences

Style of Research:
- Strong assumptions
- Math Models
- Link between BW, forecasting and information sharing
- Ungrounded
- No empirical testing
Three examples of incorporating forecasting research into inventory/supply chain research

• Intermittent demand
  – The need for an appropriate loss function

• Manufacturing and lot sizing research
  – The need to incorporate demand uncertainty and forecasting error

• Service level operations
  – The importance of including a range of plausible policies
Case I - Intermittent demand
- researching forecasting method selection

• Range of alternative methods (Kourentzes, IJPE)
  – Croston, exponential smoothing, SBA, neural nets (non-linear)
• Loss functions
  – Forecast error based, vs trade-off losses
• Methodology
  – Empirical
• Validation
  – Empirical
• Conclusions
  – Forecast Accuracy metrics inadequate to capture inventory performance
  – Neural nets lead to improved performance (but no improvement in accuracy)

Context specific loss functions critical to evaluation
# Models of the supply chain

<table>
<thead>
<tr>
<th>Building blocks</th>
<th>Typical Assumptions</th>
<th>Realistic Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Structure</td>
<td>• Two-echelon</td>
<td>• Multiple-echelon</td>
</tr>
<tr>
<td>• Demand generation process and parameters</td>
<td>• Known (I.I.D. or AR(1))</td>
<td>• Unknown (real demand, AR(1) or other ARIMA)</td>
</tr>
<tr>
<td>• Forecast function</td>
<td>• MMSE</td>
<td>• Exponential Smoothing or ARIMA</td>
</tr>
<tr>
<td>• Ordering and delivery rules</td>
<td>• Order-up-to (backorder)</td>
<td>• Order-up-to (backorder or lost sale)</td>
</tr>
<tr>
<td>• Lead time (LT)</td>
<td>• Fixed</td>
<td>• Stochastic, estimated based on realised LT</td>
</tr>
<tr>
<td>• Performance measures</td>
<td>• BWE ratio</td>
<td>• Inventory level, service level, or total cost</td>
</tr>
<tr>
<td>• References</td>
<td>E.g. Chen et. al. (2000), Lee et. al. (2000)</td>
<td>E.g. Ali and Boylan (2010), Syuhada (2014)</td>
</tr>
</tbody>
</table>
Case II - Manufacturing and lot sizing research - uncertainty and forecasting
(Fildes and Kingsman, JORS, 2010)

• Motivation: Lee and Adam (Man Sci, 1986)
  – Transparently incorrect!

• Loss functions
  – trade-off between service level and inventory

• Methodology
  – Simulation

• Forecasting methods
  – Optimal + Sub-optimal specification error (exponential smoothing)

• Lot sizing rules including EOQ and ‘optimal’

• Validation
  – Demand and Forecasting models based on realistic characterisation
    • But no ‘outliers’, promotions etc
  – Realistic specification of uncertainty and forecast error

Different LSRs
- perfect info

Cost

Service
Two levels – range of lot sizing rules

Top level demand uncertainty (ARMA(1,1))

\[ D_t = \delta + \phi D_{t-1} + e_t - \theta e_{t-1} \]

Forecast error

- Optimal forecast
  \[ F_{t-1,t}^{opt} = \delta + \rho D_{t-1} - \theta e_{t-1} \]
  - Assumes parameters and past error \( e_{t-1} \) known,
- Forecast can be non-optimal
  - Actual forecast = Optimal forecast + Noise
  \[ F_{t-1,t+k} = F_{t-1,t+k}^{opt} + \nu_t \]
- Observed error = demand uncertainty + forecast error
- Includes bias, specification error, demand uncertainty

Additional realism

- DGP unknown
- Grounded error distribution
- Forecast error can be evaluated
Conclusions
Supply chain models and uncertainty

• Models that exclude demand uncertainty and forecast error have (very) limited value in ranking policy effectiveness
  – Lot-sizing doesn’t really matter! EOQ robust (Glock et al., IJPE, 2014)

• Specifying realistic levels of uncertainty affects the conclusions
  – Size matters – not significance

• Confusing demand uncertainty, optimal forecasting and mis-specification mislead
  – In their policy recommendations
  – In their quantitative estimates of benefits.

• **For high variance, major improvements from reduced error**
  - But do these conclusions apply?
    - For more complex supply chains
    - For realistic demand generation and forecasting including information sharing
Case III - Service level operations – Call Centre Planning

Call centre planning process
- **Forecasting**
  - Models & accuracy
- **Staffing requirement**
  - Queuing models
- **Rostering**
  - Integer programming
  - HR constraints

**Call arrivals assumptions**
- Multiple Seasonality + Autocorrelations
- Different levels of forecast error/demand uncertainty in addition to Poisson randomness

**Staffing rules assumptions**
- Steady-state queue models (simple) vs Time-dependent models (complex)
- Decision constraints
  - Service level
  - Staff Cost
  - Availability etc...

**Queue performance** may disrupt forecasting inputs
- Balking, Abandonment, Retrial
Conclusions

**Target Service Level vs Intraday Variance**

### High target service level regime, e.g. TSF(0s) at 90%,
- System has low congestion level
- Service level degradation reduced at low level intraday variance
- System performance can be improved by using better forecast methods

### Low target service level regime, e.g. TSF(0s) at 10%,
- System has high congestion level, transient behavior of the queue is significant
- Implicit asymmetric cost function for staffing
- Intraday variance causes “knock-on” effects from queue backlogs
- Queueing model is the dominant effect overshadow forecast methods

Forecast methods matters

Planning methods matters

(queueuing model for staffing)
(Most) Research on inventory/service level offers limited guidance in practice

Why?

- Forecasting error poorly integrated into the planning models
- Validation ungrounded in practicalities

Most supply chain forecasting research offers limited guidance

Why?

- Inappropriate loss functions
- Poor empirical validation

Need to integrate models and methods from the two perspectives
Questions?