Segmenting sales forecasting accuracy

An ABC-XYZ-analysis on the accuracy of Neural Networks, SAP APO-DP and Judgmental Forecasting at Beiersdorf

Michael Tramitzke
Dr. Sven Crone
The target of the study: product portfolio segmentation & benchmarking of forecasting systems per segment

- Identification of homogeneous subgroups of the assortment

- Benchmark the empirical accuracy of different forecasting methods for each subgroup

- Provide recommendations on what forecasting method to apply within each product segment
Agenda

- Empirical comparison of forecasting systems
- Study design
- Product portfolio segmentation
- Study results & recommendation
5 different forecasting systems

- Judgmental forecast derived by an S&OP process and the software SAP APO DP
- Automatic model selection procedure 2 of the software SAP APO DP (AMSP2)
- Artificial neural networks using the software Intelligent Forecaster
- Automatic model selection procedure of the software Forecast Pro
  → upper Benchmark
- Naive forecasting method (combination of naive 1 and naive 12 step ahead FC)
  → lower benchmark
Judgmental forecast derived by an S&OP process and SAP APO DP
Automatic model selection procedure 2 of SAP APO DP (AMSP2)

Univariate Forecast Profile

- Profile: 00_GB_MT
- Test: AUTOMODELLAUSWAHL 2
- Read Historical Data
  - Key Figure: Y_MKUNI
  - Version: 003
- Model Parameters
  - Forecast Strategy: Forecast with Automatic Model Select
  - Alpha: 0.20
  - Beta: Gamma
  - Sigma: 1.25
  - Periods: 12
  - Variation
- Alpha Start Value: 0.05
- Beta Start Value: 0.05
- Gamma Start Value: 0.40
- Weighting Profile
- Trend Damping Profile
- Hist. Val. Markings
- Diagnosis Group
- Control Parameters
  - Outlier Correction: None
  - Days in Period
- Forecast Errors
  - MAE
  - MSE
  - MAPE
  - Error Total
- Promotion
  - Key Figure
  - Select
  - Change Vals

Maintain Forecast Profile

- Planning Area: Y_DP_BASELINE
- Master Prl.: 00_GB_MT
- Description: PROFIL ANGELEGT DURCH /SAPAPO/SDP94
- Forecast Key Figure: Y_PR
- Forecast
- Additional Settings
  - Period Indicator: M
  - Fiscal Year Variant
  - Lifecycle Planning Active
- Forecast Horizon
  - From: 01.05.2007
  - To: 31.05.2007
  - Periods
  - Offset
- History Horizon
  - From: 01.06.2003
  - To: 30.04.2007
  - Periods
  - Offset
- Model Selection
  - Univariate Forecast: 00_GB_MT
  - Multiple Linear Regression
  - Composite Forecast
Automatic model selection procedure of the software Forecast Pro
Artificial neural networks using the Intelligent Forecaster
Agenda

- Empirical comparison of forecasting systems
- Study design
- Product portfolio segmentation
- Study results & recommendation
Design empirical comparison

- Multiple benchmark methods/systems
- Out-of-sample evaluation
  - Test period: 12 observations (one business cycle of 12 months)
  - Rolling origin – 12 origins
  - Optimization at every origin
- Forecasting horizon: t+3 (3 months ahead)
- Sample: 229 time series of length between 18 and 48 observations
- Multiple error measures
Assessment of results by means of four relative error measures

- **MAPE**

- **Percent Better:** 
  \[ PB_m = \frac{\sum_{i=1}^{n} x}{n} * 100 \]

  \( \{ x = 1 \ \forall i \mid FCE_{m_i} < FCE_{naive_i} \}, \{ x = 0 \ \forall i \mid FCE_{m_i} \geq FCE_{naive_i} \} \)

- **Rank on MAPE:** 
  \[ \text{Rank}_m = \frac{\sum_{i=1}^{n} \text{Rank}_i}{n} \]

- **Mean Improvement [%]:** 
  \[ \text{MIP}_m = \frac{\sum_{i=1}^{n} (FCE_{naive_i} - FCE_{m_i})}{n} \]

- **Mean Monetary Improvement [%]:** 
  \[ \text{MIM}_m = \frac{\sum_{i=1}^{n} (AE_{tp \ naive_i \ AvP_i}) - (AE_{tp \ m_i \ AvP_i})}{(\sum_{i=1}^{n} AE_{tp \ naive_i \ AvP_i})} \]

As FCE the Mean absolute percent error (MAPE) have been used
m=1;...;5, it denotes the observed method; n denotes the number of time series in the (sub)sample

Michael Tramnitzke, 23.06.08
### Average results of the 4 error measures for the 229 time series – Bias of the average by single high errors for MIM and MIP intentional

<table>
<thead>
<tr>
<th>Rank on MAPE:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>2.63</td>
</tr>
<tr>
<td>FC Pro</td>
<td>2.65</td>
</tr>
<tr>
<td>AMSP2 – SAP APO</td>
<td>3.09</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>3.15</td>
</tr>
<tr>
<td>Naive FC</td>
<td>3.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent Better</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>67%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>67%</td>
</tr>
<tr>
<td>AMSP2 – SAP APO</td>
<td>59%</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>52%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIP</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>13%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>6%</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>3%</td>
</tr>
<tr>
<td>AMSP2 - SAP APO</td>
<td>-55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIM</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>20%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>5%</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>-17%</td>
</tr>
<tr>
<td>AMSP2 - SAP APO</td>
<td>-208%</td>
</tr>
</tbody>
</table>

Michael Tramitzke, 23.06.08
Agenda

- Empirical comparison of forecasting systems
- Study design
- Product portfolio segmentation
- Study results & recommendation
Importance - ABC analysis

- Classification Criterion:
  - A: 20% of the SKUs
  - B: 21-50% of the SKUs
  - C: 51-100% of SKUs

- Grochla (1978)
- Wildemann (1988)
- Ng (2007)
Forecasting complexity - XYZ analysis

**Classification Criterion:**
- **X:** MAPE of naïve method $\leq 30\%$
- **Y:** MAPE of naïve method $> 30\%$ and $\leq 60\%$
- **Z:** MAPE of naïve method $> 60\%$

- Grochla (1978) $\rightarrow$ RSU analysis
- Tempelmeier (2006) $\rightarrow$ RSU & XYZ synonymously
- Alicke (2005)
Forecasting complexity - XYZ analysis

Classification Criterion:

X: MAPE of naïve method <= 30%

Y: MAPE of naïve method > 30% and <= 60 %

Z: MAPE of naïve method > 60%

- Grochla (1978) → RSU analysis
- Tempelmeier (2006) → RSU & XYZ synonymously
- Alicke (2005)
Combination of ABC-XYZ results in 9 segments

<table>
<thead>
<tr>
<th>Filter</th>
<th>58 X</th>
<th>61 Y</th>
<th>90 Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>229 All</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46 A</td>
<td>19 AX</td>
<td>11 AY</td>
<td>16 AZ</td>
</tr>
<tr>
<td>69 B</td>
<td>27 BX</td>
<td>25 BY</td>
<td>17 BZ</td>
</tr>
<tr>
<td>114 C</td>
<td>12 CX</td>
<td>45 CY</td>
<td>37 CZ</td>
</tr>
</tbody>
</table>

- Exemplary Frequency of occurrence of the segments in the observed sample of 229 time series
Agenda

- Empirical comparison of forecasting systems
- Study design
- Product portfolio segmentation
- Study results & recommendation
Depending on the segment the performance of the methods differs

<table>
<thead>
<tr>
<th>MIP</th>
<th>average</th>
<th>MIP</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>7%</td>
<td>FC Pro</td>
<td>0%</td>
</tr>
<tr>
<td>for A</td>
<td></td>
<td>AMSP2 – SAP APO</td>
<td>-1%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>-3%</td>
<td>Judgmental FC</td>
<td>0%</td>
</tr>
<tr>
<td>for A</td>
<td></td>
<td>AMSP2 – SAP APO</td>
<td></td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>-15%</td>
<td>for X</td>
<td></td>
</tr>
<tr>
<td>AMSP2 – SAP APO</td>
<td>-141%</td>
<td>Intelligent Forecaster</td>
<td>-3%</td>
</tr>
<tr>
<td>Judgmental FC</td>
<td>11%</td>
<td>for Y</td>
<td>3%</td>
</tr>
<tr>
<td>for B</td>
<td></td>
<td>FC Pro</td>
<td>1%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>5%</td>
<td>Intelligent Forecaster</td>
<td>-9%</td>
</tr>
<tr>
<td>for B</td>
<td></td>
<td>AMSP2 – SAP APO</td>
<td>-81%</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>-3%</td>
<td>for C</td>
<td>33%</td>
</tr>
<tr>
<td>AMSP2 – SAP APO</td>
<td>-61%</td>
<td>Judgmental FC</td>
<td></td>
</tr>
<tr>
<td>Judgmental FC</td>
<td>16%</td>
<td>for C</td>
<td></td>
</tr>
<tr>
<td>for C</td>
<td></td>
<td>FC Pro</td>
<td></td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>15%</td>
<td>Intelligent Forecaster</td>
<td>21%</td>
</tr>
<tr>
<td>for C</td>
<td></td>
<td>AMSP2 – SAP APO</td>
<td></td>
</tr>
<tr>
<td>FC Pro</td>
<td>10%</td>
<td>for C</td>
<td>15%</td>
</tr>
<tr>
<td>AMSP2 – SAP APO</td>
<td>-18%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Michael Tramitzke, 23.06.08
Results of the empirical comparison – basis are the relative error measures

- Judgmental forecasting fails to increase forecasting accuracy on simple 'X'-products

- Judgmental forecasting outperforms automatic methods on:
  - Complex ‘Z’ products
  - Short and medium time series
  - Time series driven by causal effects

- Neural Networks cannot outperform other forecasting methods on average but were almost as good as the human planner for forecasting 'Z'-products

- Segmenting product portfolios is highly reasonable

→ recommendations have been developed on the basis of the ABC-XYZ classification
Recommendation for the forecasting process on the basis of the ABC XYZ classification

A
review regularly with high intensity

B
review on occasion with medium intensity

C
review irregularly (spot checks)

X
partly automated forecasting based on statistical models and interactive adjustments

Y
Automated forecasting based on present statistical models

Z
Manual / interactive forecasting without usage of staticitical models

importance of market intelligence

degree of manual control
Recommendation for the forecasting process on the basis of the ABC XYZ classification

- Problems occur for the CZ products due to manual planning

- ANNs could be an alternative for Z products

- Training ANNs for single segments should be further investigated

- Time savings should be invested into integrated sales forecasting

<table>
<thead>
<tr>
<th>MIP – CZ segment</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental FC</td>
<td>32%</td>
</tr>
<tr>
<td>Intelligent Forecaster</td>
<td>31%</td>
</tr>
<tr>
<td>FC Pro</td>
<td>18%</td>
</tr>
<tr>
<td>AMSP2 - SAP APO</td>
<td>-36%</td>
</tr>
</tbody>
</table>
Thanks for your attention!

Questions?

Links Backup:
- Distribution of products across segments
- Characteristics of the FMCGI
- Definition of the Naive method
- Additional segmentation strategies
- Empirical comparison
- Artificial neural networks
References

Distribution of products across segments
## Distribution of the products among the segments ABC XYZ

<table>
<thead>
<tr>
<th>Segments</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19</td>
<td>17</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>A XYZ percent</td>
<td>42.2%</td>
<td>37.8%</td>
<td>20.0%</td>
<td></td>
</tr>
<tr>
<td>X ABC percent</td>
<td>33.9%</td>
<td>19.5%</td>
<td>10.7%</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>28</td>
<td>19</td>
<td>69</td>
</tr>
<tr>
<td>B XYZ percent</td>
<td>31.9%</td>
<td>40.6%</td>
<td>27.5%</td>
<td></td>
</tr>
<tr>
<td>Y ABC percent</td>
<td>39.3%</td>
<td>32.2%</td>
<td>22.6%</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>15</td>
<td>42</td>
<td>56</td>
<td>113</td>
</tr>
<tr>
<td>C XYZ percent</td>
<td>13.3%</td>
<td>37.2%</td>
<td>49.6%</td>
<td></td>
</tr>
<tr>
<td>Z ABC percent</td>
<td>26.8%</td>
<td>48.3%</td>
<td>66.7%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>87</td>
<td>84</td>
<td>227</td>
</tr>
</tbody>
</table>
Characteristics of the FMCGI
Characteristics Fast Moving Consumer Goods Industry (FMCGI)

- High number of heterogeneous products
- Different sales volumes
- Various levels of randomness of the time series
- Different time series lengths
- Varying time series patterns
- Numerous causal impacts
Definition of the Naive method
Naïve forecast defines XYZ classification criterion & lower benchmark

- Literature: Naïve 1
- Forecast horizon at BDF is 3 step ahead → use naïve 3 step ahead
- Problems arise for seasonal products
Naïve forecast defines XYZ classification criterion & lower benchmark

- Literature: Naïve 1
- Forecast horizon at BDF is 3 step ahead → use naïve 3 step ahead
- Problems arise for seasonal products
- Alternative use naïve 12 step ahead
- Problems arise in case of trends, level shifts or short time series
- Decision: combination of both methods
Times series sections for determination of the naïve forecast

- Determine naïve 3 step ahead forecast for the selection sample
- Determine naïve 12 step ahead forecast for the selection sample
- Select the superior method
- Calculate with the selected method the forecast for the hold out sample
- Calculate the FCE for the hold out sample

![Diagram showing the timeline for naïve forecasts and selection and hold out samples](image-url)
Additional segmentation strategies
Additional segmentation criteria of the study

- **Time series pattern**
  - Stationary, seasonal, trend, trend seasonal

- **Time series length**
  - Short (6-12 months), medium (13-24 months), long (25-36 months)

- **Promotion activity**
  - Promotionally driven (2 for 1, 50% price of, TV campaign)
  - Not promotionally driven (no promo activity, 3 for 2, 25% price off, 2 for x €)
Empirical comparison
Empirical comparison per method

FCA 64%

average % FC error per selection 49.9%

average FCA per selection 64.3%

average % importance improvement 14.1%

<table>
<thead>
<tr>
<th>Sales Value</th>
<th>sum Inv</th>
<th>sum Inv FC</th>
<th>% FC err</th>
<th>max 300</th>
<th>sum abs N3 FC Error</th>
<th>sum abs N3 FC Error*</th>
<th>sum abs N3 FC Error* (enr. price)</th>
<th>FC per product</th>
<th>Park % FC Error</th>
<th>Park Inv</th>
<th>Park sales value</th>
<th>XYZ</th>
<th>ABC</th>
<th>time series length</th>
</tr>
</thead>
<tbody>
<tr>
<td>£4,418,241</td>
<td>119,742</td>
<td>119,742</td>
<td>16.19%</td>
<td>16.19%</td>
<td>776,572</td>
<td>776,572</td>
<td>461640</td>
<td>64.82%</td>
<td>12</td>
<td>Y</td>
<td>A</td>
<td>25</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£5,255,644</td>
<td>299,236</td>
<td>115,702</td>
<td>30.30%</td>
<td>26.42%</td>
<td>1,063,404</td>
<td>1,063,404</td>
<td>1,063,404</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£2,025,701</td>
<td>150,923</td>
<td>30,938</td>
<td>26.94%</td>
<td>26.94%</td>
<td>700,635</td>
<td>700,635</td>
<td>1,025,701</td>
<td>64.91%</td>
<td>12</td>
<td>Y</td>
<td>A</td>
<td>25</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£2,545,701</td>
<td>637,542</td>
<td>147,075</td>
<td>22.26%</td>
<td>22.26%</td>
<td>1,041,533</td>
<td>1,041,533</td>
<td>1,041,533</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£2,725,800</td>
<td>285,344</td>
<td>105,667</td>
<td>39.33%</td>
<td>39.33%</td>
<td>1,652,997</td>
<td>1,652,997</td>
<td>1,652,997</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£2,041,779</td>
<td>562,222</td>
<td>185,786</td>
<td>33.04%</td>
<td>33.04%</td>
<td>1,783,584</td>
<td>1,783,584</td>
<td>1,783,584</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£1,045,700</td>
<td>100,237</td>
<td>200,000</td>
<td>10.00%</td>
<td>10.00%</td>
<td>1,946,492</td>
<td>1,946,492</td>
<td>1,946,492</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>£2,000,700</td>
<td>150,226</td>
<td>275,000</td>
<td>11.00%</td>
<td>11.00%</td>
<td>1,574,000</td>
<td>1,574,000</td>
<td>1,574,000</td>
<td>2,125,642</td>
<td>26</td>
<td>Y</td>
<td>A</td>
<td>26</td>
<td>36</td>
<td>36</td>
</tr>
</tbody>
</table>
Artificial neural networks
Sensitivity analysis: parameters of the artificial neural network have been determined by means of forecasting generic time series

- Adjust ANN parameters to the time series before the forecast

- Parameter selection by means of generic time series not original sample

- Normally 36 generic time series \(\rightarrow\) 12 patterns combined with 3 noise levels

- The 12 most occurring patterns of the 36 possible patterns have been selected

- For the 12 patterns subsequently one parameter have been varied
Starting with an basis MLP 7 parameters have been varied

- Basic MLP selected by means of literature recommendations
- Following parameters have been varied:
  - Number of input neurons
  - Number of hidden neurons
  - Initialization range of starting weights
  - Number of initializations (twice)
  - Learn rate
  - Momentum term
  - Activation function of input and hidden layer
Used MLP of the study

- **Topology:**
  - One, two and ten input neurons
  - One output neuron
  - One hidden layer with five hidden neurons
  - Activation function of the hidden layer: Sigmoid
  - Activation function of the output layer: Identity
  - Target function Identity

- **Learning:**
  - Input data have been scaled to the interval \([-0.6;0.6]\)
  - Choosing with repetition, 1000 epochs
  - Learning algorithm Backpropagation; learn rate 0.3; momentum term 0.4; cooling rate 0.99, frequency 1 epoch
  - Learning with early stopping
  - 20 initializations; init interval \([-0.7;0.7]\)
Screenshot of the analysis graphic during the trainings process
Screenshot of the error evaluation for the input neuron variation
End…