HOLT-WINTERS AND NEURAL-NETWORK METHODS FOR MEDIUM-TERM SALES FORECASTING

Apostolos Kotsialos, Markos Papageorgiou, and Antonios Poulimenos

Dynamic Systems and Simulation Laboratory, Technical University of Crete, 73100 Chania, Greece
Phone: +30-28210-37289 Fax: +30-28210- 37584 E-mail: markos@dssl.tuc.gr

Abstract: The problem of medium to long term sales forecasting raises a number of requirements that must be suitably addressed in the design of the employed forecasting methods. These include long forecasting horizons (up to 52 periods ahead), a high number of quantities to be forecasted, which limits the possibility of human intervention, as well as frequent introduction of new articles (for which no past sales are available for parameter calibration) and withdrawal of running articles. The problem has been tackled by use of a modified Holt-Winters method as well as Feedforward Multilayer Neural Networks (FMNN) applied to sales data from two German companies.

Keywords: Forecasts, feedforward neural networks, time-series analysis

1. INTRODUCTION

Sales forecasting over a sufficiently long future time horizon is an important prerequisite for efficient production planning and a solid basis for company policy decisions. A number of efficient forecasting algorithms with different levels of complexity have been developed and tested in the past. Most method assessments and comparisons, however, address the problem of short-term or even one-step-ahead forecasting (see, e.g. Makridakis *et al.*, 1982; Gaynor and Kirkpatrick, 1994; Lachtermacher and Fuller, 1995). A hybrid approach combining long-term with short-term forecasts employing the Holt-Winters method was investigated by Rajopadhye *et al.* (1999). The forecasting methods to be employed should be able to operate largely autonomously, because human supervision and intervention is hardly possible in case of thousands of article sales to be forecasted.

The well known Holt-Winters method (Chatfield, 1978) is suitably modified for long-term forecasting, and various Feedforward Multilayer Neural Network (FMNN) approaches are also proposed for the same problem. Both groups of methods as well as combinations thereof are applied to various kinds of articles, group sales, and total sales from two German companies (a total of 195 time-series are used) to assess and compare their forecasting performance and their suitability in view of the above requirements.

FMNN-based prediction has recently gained remarkable popularity due to the possibility to describe complex nonlinear interrelationships within a relatively convenient black-box approach (Chakraborty et al., 1992; Yao, 1999). FMNN methods have been increasingly applied to prediction problems during the last decade (Chakraborty et al., 1992; Cottrell, et al., 1995; Lachtermacher and Fuller, 1995; Bunn, 2000). Their basic advantage is that they may capture the unknown nonlinear structure of the process to be modelled. Their basic disadvantages are the high number of parameters to be calibrated and the black-box approach that renders the plausible interpretation of the modeling structure very difficult and excludes the possibility of ad-hoc parameter choice.

2. AVAILABLE DATA

A total of 195 time-series is used to test and compare the various versions of Holt-Winters and FMNN-based predictors. All time-series represent sales quantities from two industrial firms, namely the toy

Table 1 Average of seasonality correlation factor

Group	D1	D2	D3	D4
$\bar{r}_s$	0.29	0.36	0.06	0.016
max r_s	0.53	0.45	0.18	0.08
min r_s	-0.03	0.26	-0.07	0.0
Group	D5	D6	D7	D8
$\bar{r}_s$	0.25	0.48	-0.19	0.02
max r_s	0.56	0.48	0.27	0.18
min r_s	-0.13	0.48	-0.51	-0.08

producer sigikid (sk) and a wire-harness production factory for trucks owned by DaimlerChrysler (DC). All available time-series extend over the years 1997-1999 on a weekly base. For sk-data we have, for each time-series, 52 values for 1997, 52 values for 1998, and 40 values for 1999, while DC-data cover, for each time-series, 52 values for each of the three years. For a better analysis of results, the 195 time-series are assigned to the following 8 data groups:

D1: 60 time-series of sales numbers for 60 ordinary sk-articles with various average sales numbers .

D2: 4 time-series of sales values (in german DM) for 4 sk groups of articles with various average values.

D3: 10 time-series of sales numbers for sk-articles which were withdrawn from production during the period 1997-1999.

D4: 5 time-series of sales numbers for sk-articles which were first introduced within the period 1997-1999.

D5: 100 time-series as in D1 but for DC-articles.

D6: 1 time-series of sales numbers for a DC-group of articles.

D7: 10 time-series as in D3 but for DC-articles.

D8: 5 time-series as in D4 but for DC-articles.

The available data display some kind of more or less pronounced periodicity of the time-series over the time span of the year (i.e. over 52 weeks). This periodicity reflects the yearly seasonality of sales that may depend on annual fairs, end-of-year or summer holidays, etc. Assume a weekly time-series Y_t, t=1,…,52, over a year, with average value

$$L = \frac{1}{52}\sum_{t=1}^{52} Y_t . \qquad (1)$$

The seasonality factors S_t, t=1,…,52, for the weeks of that year are then defined as $S_t = Y_t/L$. For each time-series from D1-D8, the corresponding seasonality factors S_t^{1997} and S_t^{1998} , t=1,…,52, for the years 1997 and 1998, respectively, may be calculated. To avoid later difficulties, S_t = 0.1 is set whenever a value $Y_t \leq 0$ is encountered in the data (negative values correspond to very limited article returns). In order to calculate a practical index for the strength of seasonality, a simple correlation formula is applied to the 1997 and 1998 seasonality factors for each time-series as following

$$r_s = \frac{\sum_{t=1}^{52}\left(S_t^{1997}-1\right)\left(S_t^{1998}-1\right)}{\sqrt{\sum_{t=1}^{52}\left(S_t^{1997}-1\right)^2}\sqrt{\sum_{t=1}^{52}\left(S_t^{1998}-1\right)^2}} \qquad (2)$$

where r_s is the seasonality correlation factor. The higher r_s is for a particular time-series, the stronger the seasonality pattern. A value of close to zero indicates no seasonality while r_s<0 indicates negative seasonality (pathological cases).

Table 1 provides the averages of seasonality factors for each group D1-D8 as well as the corresponding maximum and minimum values. Most ordinary time-series, belonging to the groups D1, D2, D5, D6, have quite strong seasonality. Most time-series of articles that are either withdrawn or introduced within the 1997-1998 period (groups D3, D4, D7, D8) have rather low or even negative seasonality correlation factors.

3. FORECASTING METHODS USED

3.1 The Modified Holt-Winters Method

The Holt-Winters method (Makridakis and Wheelright, 1978) utilizes simple exponential smoothing to estimate the values of three basic components of a time-series: the level (average value), the trend, and the seasonality. The basic equations that are applied at each update period t = 1, 2, … are

level: $$L_t = \alpha\frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1}+b_{t-1}) \qquad (3)$$

trend: $$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \qquad (4)$$

seasonality: $$S_t = \gamma\frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \qquad (5)$$

forecast: $$F_{t+m} = (L_t + b_t m)S_{t-s+m}, \quad m = 1,..., M \qquad (6)$$

where s is the period of the seasonality (here s = 52), Y_t is the current value of the time-series, L_t denotes the level of the series (yearly average), b_t denotes the trend, S_t is the seasonal component, and F_{t+m} is the forecast for m=1,…,M, with M the forecasting horizon. In our case, we have M=52 for data groups D1-D4 and M=22 for D5-D8, in accordance with the needs of the corresponding companies.

In order to avoid unrealistic over-prediction, the linear-trend forecasting equation (6) may be modified as follows

$$F_{t+m} = (L_t + b_t\sqrt{m})S_{t+m} \qquad (7)$$

to forecast the future sales numbers. By square-rooting the m index, we obtain a sub-linear trend approach which leads to a more conservative trend extrapolation.

3.2 FMNN Predictors

A FMNN consists of several simple units, called neurons, that are organised into a number K of layers (Fig. 1) including an input layer, an output layer, and several hidden layers. Each hidden layer k contains a number N_k of neurons. The input and output layers have no neurons but they have as many entries as input and output variables, respectively, i.e. $N_1 = n$ and $N_K = p$. The number of hidden layers and the number of neurons for each hidden layer may be chosen

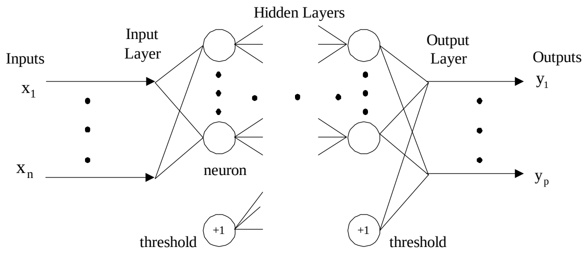

Fig. 1. The general FMNN structure.

freely. Each neuron i of each layer k is connected with each neuron j of layer k+1, except for the output layer. A corresponding weight w_{ij}^k is assigned to each connection between neurons of consecutive layers. Each layer, except for the input and output layers, has an extra neuron (threshold), which connects with each neuron of the next layer.

The weights related to these extra neurons are fixed in this paper such that, if $x_1=\ldots=x_n=0$, the outputs $y_1=\ldots=y_p=0$, whatever the values of the other weights. Each neuron's input is a linear combination of the previous-layer's neuron outputs, and each neuron's output is produced by the logistic function.

In the application of FMNN to forecasting, all quantities known at time t (i.e. the current and past time-series values Y_i, $i=t,t-1,\ldots,t-n+1$) up to a depth n are used as inputs to the FMNN while the future (forecasted) values F_{t+i}, $i=1,\ldots,M$, are the FMNN's outputs. In our approach, the seasonality factors S_t are estimated via a simple smoothing while the de-seasonalised current and past sales are used as FMNN-inputs to produce de-seasonalised future sales. The FMNN predictors are not designed to produce directly all M required forecasts but just 4 forecast values F_{t+m}, e.g. for m = 1, 4, 26, 52, while the rest of the forecasts are produced from these values via linear interpolation.

Note that in all FMNN-based prediction approaches to be presented in this section, a normalisation is applied to the corresponding FMNN inputs and outputs so that they don't exceed the range [0, 1]. It was found that a suitable normalisation parameter for each time-series equals 15 times the average weekly sales of 1997-1998.

Four distinct versions of this FMNN-based prediction are investigated:

Version N.1: There are 4 inputs to the FMNN, namely the four last (de-seasonalized, normalized) sales numbers Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}. The FMNN's hidden layer includes only one neuron, hence we have a 4-1-4 structure with 8 free weights.

Version N.2: Same as version N.1, but the inputs are now Y_t, Y_{t-2}, Y_{t-4}, Y_{t-6} , to cover a longer past period.

Version N.3: Same as version N.1 but with two neurons in the hidden layer, leading to a 4-2-4 structure with 16 free weights.

Version N.4: Same as version N.2 but with two neurons in the hidden layer, leading to a 4-2-4 structure with 16 free weights.

Because FMNN are nonlinear universal approximators, we may introduce a hybrid FMNN-based approach, whereby we keep the smoothing equations (3)-(5) of the Holt-Winters method (with ad-hoc parameters $\alpha=\beta=0.1$) to produce the current level L_t, trend b_t, and seasonality S_t, and apply subsequently a FMNN, instead of the extrapolation equation (6) to produce the forecasts F_{t+m} (m = 1,4,18,22 for data groups D5-D8). A total of 4 versions for this hybrid approach are considered:

Version H.1: There are 2 inputs to the FMNN, namely the (normalized) L_t, b_t; one hidden neuron; free threshold weights; leading to a 2-1-4 structure with 10 free weights.

Version H.2: Same as version H.1 but with fixed threshold weights leading to 6 free weights.

Version H.3: Same as version H.2 but with two hidden neurons, leading to a 2-2-4 structure with 12 free weights.

Version H.4: There is one input to the FMNN, namely the (normalized) b_t; one hidden neuron, leading to a 1-1-4 structure with 5 free weights.

4. CALIBRATION OF PARAMETERS

4.1 Parameter Calibration for the Holt-Winters Method

The calibration phase considers only the parameters α, β of equations (3), (4). Because the seasonality factors S_t, $t=1,\ldots,52$, are available only for two years, a veritable calibration of the parameter γ cannot be performed. For this reason, equation (5) and the parameter γ are not considered in the calibration.

The available calibration data of sales numbers cover two years (1997, 1998), i.e., a total of 104 weeks. The real available data are denoted Y_t, $t=1,\ldots,104$. Using (3), (4) and (6) or (7) at each time t, the forecasts F_{t+m}, $m=1,\ldots,52$, $t=1,\ldots,104-m$, may be produced for a given set of parameter values α, β. The optimisation problem formed, attempts to specify those values of α, β that minimize the following prediction accuracy criterion

$$J(\alpha,\beta)=\frac{\sum_{m=1}^{M} J_m(\alpha,\beta)}{M}=[\sum_{m=1}^{M}\frac{\sqrt{(104-m)\sum_{t=1}^{104-m}(F_{t+m}-Y_{t+m})^2}}{\sum_{t=1}^{104-m}Y_{t+m}}]/M \quad (8)$$

The overall criterion J is the average relative mean quadratic error (ARMQE). We will refer to this overall criterion as the prediction accuracy. The optimisation method used to minimise (8) with respect to α, β is exhaustive search in steps of 0.01 within the admissible range of [0, 1]. Table 2 provides the averages $\bar{J}$ of the achieved calibration accuracy (8) for each data group in the sub-linear trend case. Moreover, the best and worst achieved calibration accuracy within each data group is also displayed. It may be seen that results range from good to excellent for data groups D1, D2, D5, D6 of ordinary time-series while the calibration accuracy is rather mixed for groups D3, D4, D7, D8.

Table 2 Calibration accuracy results, sub-linear trend

Group	D1	D2	D3	D4
$\bar{J}$	0.61	0.53	2.85	2.25
min J	0.3	0.38	0.41	0.59
max J	1.07	0.9	5.74	3.13
Group	D5	D6	D7	D8
$\bar{J}$	0.23	0.21	0.56	1.02
min J	0.17	0.21	0.38	0.63
max J	0.81	0.21	1.5	1.26

Table 3 Calibration accuracy results, sub-linear trend, $\alpha=\beta=0.1$

Group	D1	D2	D3	D4
$\bar{J}$	0.74	0.56	4.01	2.31
min J	0.34	0.39	0.59	0.76
max J	2.02	0.94	12.8	3.19
Group	D5	D6	D7	D8
$\bar{J}$	0.27	0.25	0.91	1.1
min J	0.2	0.25	0.41	0.64
max J	0.83	0.25	4.35	1.36

Table 4 Calibration accuracy results, linear trend

Group	D1	D2	D3	D4
$\bar{J}$	0.61	0.53	2.8	2.3
min J	0.3	0.38	0.42	0.59
max J	1.07	0.91	5.63	3.17
Group	D5	D6	D7	D8
$\bar{J}$	0.23	0.21	0.56	1.04
min J	0.18	0.21	0.38	0.64
max J	0.86	0.21	1.5	1.31

Table 5 Overall calibration accuracy results

Appr.	Sub-linear trend	Sub-linear common par.	Linear trend.
$\bar{J}$	0.57	0.72	0.57
min J	0.17	0.2	0.18
max J	5.74	12.8	5.63

Furthermore, a reasonable prediction can be achieved by a common set of parameters α, β for all time series. Table 3 depicts the same information as Table 2, for this ad-hoc common-parameters approach. The calibration results for the linear-trend approach, i.e., when (6) is used instead of (7) for future predictions, are presented in Table 4. The overall results for each approach are presented in Table 5. The main difference between the linear and sub-linear approaches is that the latter provides much smoother (hence better) results with regard to the accuracy sensitivity.

4.2 Training of the FMNN Predictor

FMNNs are trained by use of known couples (examples) of inputs and outputs ($\tilde{\mathbf{x}}_t, \tilde{\mathbf{y}}_t$), t=1,…,T. If the FMNN is fed with $\tilde{\mathbf{x}}_t$, it should produce an output $\mathbf{y}_t = \mathbf{f}(\tilde{\mathbf{x}}_t;\mathbf{w})$ as close as possible to the given output $\tilde{\mathbf{y}}_t$. Thus, FMNN-training corresponds to solving the following non-linear optimisation problem

$$J(\mathbf{w}) = \frac{1}{2}\sum_{t=1}^{T}\left\|\tilde{\mathbf{y}}_t - \mathbf{f}(\tilde{\mathbf{x}}_t;\mathbf{w})\right\|^2 \to \underset{\mathbf{w}}{Min} \qquad (9)$$

subject to the bounds of **w**, the vector of all free weights of the FMNN. In our case, the weights' bounds are $-5 \le w_{ij}^k \le 5$. This optimisation problem is solved for each version of FMNN-based predictors and for each time-series by use of an efficient numerical optimisation algorithm employing conjugate-gradient search directions (Johansson et al. 1992).

Table 6 Overall training results for 8 FMNN-based predictors (number of free parameters in parenthesis)

FMNN	N.1 (8)	N.2 (8)	N.3 (16)	N.4 (16)
$\bar{J}$	0.175	0.189	0.159	0.171
FMNN	H.1 (10)	H.2 (6)	H.3 (12)	H.4 (5)
$\bar{J}$	0.125	0.185	0.173	0.211

Thus, optimal weight values are found for each particular time-series. Table 6 presents the summarized training results (i.e. the averages of the achieved training criterion (9) for all data groups) for the 8 FMNN- based predictor versions. Version H.1 with 10 free parameters delivers the best training results. Note that, for versions based on the same input-output concept and only differing in the number of hidden neurons (i.e. versions H.2 and H.3; N.1 and N.3; N.2 and N.4), the version with the higher number of free parameters (weights) has a strictly better training accuracy than its counterpart with a lower number of free parameters. A good (or better) training accuracy is no guarantee for a better approach, due to a possible over-parametrisation. Thus, the final comparison and selection can only be based on the generalisation results. The use of optimal weights for each individual time-series may have a number of disadvantages, the most important being that, for newly introduced articles, there are no past sales to allow for individual FMNN training. In the case of FMNN-based prediction, however, there is hardly any possibility to apply ad-hoc values for the weights. What can be done is to repeat the training procedure imposing common weights for all time-series (optimal common-parameters approach), in which case the forecasts for any newly introduced articles can be produced immediately.

The optimal common-parameters approach was pursued separately for the sigikid sales data (groups D1-D4) and for the DaimlerChrysler data (groups D5-D8). Only the articles included in the corresponding ordinary data groups D1 and D5 were used to specify the optimal common parameters for each FMNN-based predictor version. Table 7 presents the achieved training results, i.e. the achieved values of (9) for groups D1 and D5 and for each predictor version. The training accuracy deterioration due to the use of common (rather than individually optimised) weights is very small.

5. VERIFICATION RESULTS

5.1 Verification Results for the Holt-Winters Method

Verification is based on N values of the time-series from 1999 that were not included in the calibration exercise. We have N=42 for the time-series of D1-D4 and N=52 for those of D5-D8.

Table 7 Training results for the optimal common-parameters approach

FMNN	N.1	N.2	N.3	N.4
D1	0.246	0.246	0.240	0.259
D5	0.051	0.064	0.048	0.059
FMNN	H.1	H.2	H.3	H.4
D1	0.188	0.245	0.239	0.269
D5	0.043	0.068	0.063	0.070

Table 8 Verification accuracy results: optimal parameters.

Group	D1	D2	D3	D4
$\bar{J}$	0.82	1	303.6	1.03
min J	0.26	0.56	10.64	0.77
max J	0.99	2.17	1048	1.33
Group	D5	D6	D7	D8
$\bar{J}$	0.23	0.22	2607.4	0.54
min J	0.18	0.22	3.15	0.43
max J	0.3	0.22	10552	0.69

The verification phase was conducted only for the sub-linear trend (7) using two approaches, the optimal parameters approach and the ad-hoc common-parameters approach (α=β=0.1). For both approaches, the verification accuracy was calculated by use of equation (8), but considering only the weeks of 1999. More precisely, the verification accuracy is obtained by running every approach for t=1,…,104+N, but evaluating only the predictions for t=105,…,104+N via

$$J = \sum_{m=1}^{M} J_m / M = [\sum_{m=1}^{M} \frac{\sqrt{N \sum_{t=105-m}^{105+N-m} (F_{t+m} - Y_{t+m})^2}}{\sum_{t=105-m}^{105+N-m} Y_{t+m}}] / M \cdot \quad (10)$$

Table 8 provides the averages of the achieved verification accuracy (10) for each data group as well as the best and worst J within each group. Based on these results we may state the following:

- The forecasts are very good (comparable or even better than in the calibration) for the ordinary time-series of D1, D2, D5, D6.
- An explosion of the verification criterion is observed for the time-series with declining sales from D3, D7. This is because the corresponding articles have very low sales in 1999 (due to their withdrawal from production) rendering the denominator of (10) very small. The forecasting errors are temporarily large (before adapting to the zero sales) thus rendering J exorbitantly high in some cases. In conclusion, the partly exorbitant values of J observed are fully explicable and do not necessarily reflect an irrelevant forecasting behaviour.
- For the newly introduced articles of D4, D8, the warm-up period of the forecasting method is completed in 1999, hence the forecasts are clearly better than in the calibration. Table 9 provides the averages of the achieved verification accuracy (10) for each data group as well as the best and worst J within each group, when the common parameters are used. These results may be interpreted in the same way as for the optimal parameters above. We remark a slight

Table 9 Verification accuracy results: common parameters.

Group	D1	D2	D3	D4
$\bar{J}$	0.81	0.97	162.1	0.98
min J	0.16	0.54	5.59	0.79
max J	0.99	2.08	983	1.39
Group	D5	D6	D7	D8
$\bar{J}$	0.21	0.19	1392.4	0.52
min J	0.15	0.19	3.77	0.43
max J	0.35	0.19	6308.8	0.64

Table 10 Overall verification accuracy results

Appr.	Sub-linear trend	Sub-linear trend common par.
$\bar{J}$	136.35	73.02
min J	0.18	0.15
max J	10552	6308.8

amelioration for data groups D1, D2, D4, D5, D6, D8, while for D3and D7 the amelioration of J is more substantial (although the J-values are still exorbitantly high) due to α=0.1 which is in most cases higher than the optimal α-values resulting from calibration, thus leading to a quicker adaptation of the forecasts to the close-to-zero sales.

5.2 Generalization Results for the FMNN Predictors

The generalisation procedure for the various FMNN-based predictor versions follows the lines of the verification procedure of the previous subsection, and is based on the same 1999 article data and the same criterion (10). Thus the generalisation results are fully comparable across predictor versions, across individual time-series and, of course, comparable with the Holt-Winters verification results. Figure 2, depicts a representative sales time-series from data group D2 along with forecasted values for FMNN predictor H.4. Table 11 presents the summarized overall generalisation results for the 8 FMNN-based predictor versions. The explosion of the criterion value for the time-series with declining sales D3 and D7 is explained in precisely the same way as in the verification results. To circumvent the distortion in the average results, a second row has been added in Table 9 that does not include D3 and D7 results. The versions N.3, N.4, H.1, H.3, with a higher number of weights are seen to produce worse generalisation results than the versions with a lower number of parameters and are therefore discarded. Comparing the best FMNN-based predictor, H.4, with the common-parameters modified Holt-Winters predictor, we observe virtually equivalent results. The complex machinery of FMNN-based prediction and the higher number of parameters did not lead to considerably better results than the Holt-Winters method.

Table 12 presents the summarized overall generalisation results for the 8 FMNN-based predictor versions run with the common parameters that were calculated in the training phase. The versions N.3, N.4, H.1, H.3 with a higher number of weights are also here equivalent or worse than the simpler versions with

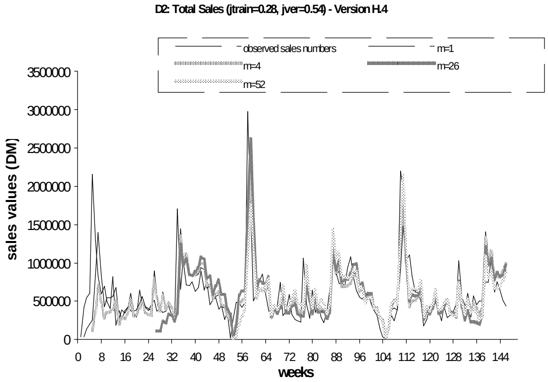

Fig. 2. Representative sales time-series from data group D2 along with forecasted values for FMNN predictor H.4.

Table 11 Overall generalisation results for 8 FMNN-based predictors (number of free parameters in parenthesis)

FMNN	N.1 (8)	N.2 (8)	N.3 (16)	N.4 (16)
$\bar{J}$	59.99	72.46	63.42	79.30
Without D3 & D7	0.50	0.47	0.50	0.51
FMNN	H.1 (10)	H.2 (6)	H.3 (12)	H.4 (5)
$\bar{J}$	549.84	49.77	59.00	99.00
Without D3 & D7	0.54	0.47	0.48	0.46

less parameters. The achieved generalisation results with optimal common parameters are seen to be slightly better than those achieved with individually optimised parameters. This indicates that the necessary number of parameters required to produce reliable and robust sales forecasts does not need to be high. Compared with the modified Holt-Winters results obtained for the same time-series only a slight amelioration has been achieved.

6. CONCLUSIONS

The problem of medium-term sales forecasting has been tackled by use of the Holt-Winters method which does not require individual analysis of each time-series to produce forecasts, and by use of 8 distinct versions of FMNN. A modification of the Holt-Winters method (sub-linear extrapolation) reduces substantially the sensitivity of the forecasting results with respect to the smoothing parameters thus allowing for the calibration phase to be circumvented via use of standard (common) parameter values for all time-series. This way, reliable forecasts may be produced immediately even for newly introduced articles for which no past sales are available for calibration. The achieved forecasting results are excellent for ordinary time-series and quite satisfactory for special types of time-series. FMNN-based forecasts were only slightly better than their counterparts that were produced with the much simpler modified Holt- Winters method. Although the slightly increased forecasting accuracy does not appear to justify the application of complex FMNN-based predictors for the investigated data, bigger differences might occur with other sales time-series.

Table 12 Overall generalisation results for 8 FMNN-based predictors with common parameters (without D3 and D7)

FMNN	N.1	N.2	N.3	N.4
$\bar{J}$	0.47	0.48	0.48	0.48
FMNN	H.1	H.2	H.3	H.4
$\bar{J}$	0.56	0.44	0.46	0.45

ACKNOWLEDGEMENT

This research was conducted with partial support of the European Commission Programme ESPRIT in the frame of the project DELPHI. The authors would like to thank sigikid and DaimlerChrysler for providing the data and for helpful support. All contents and statements in the paper are under the sole responsibility of the authors.

REFERENCES

Bunn, D.W. (2000) Forecasting loads and prices in competitive power markets, *Proceedings of the IEEE*, **88**, 163-169.

Chatfield, C. (1978) The Holt-Winters forecasting procedure, *Applied Statistics*, **27**, 264-279.

Chakraborty, K., K. Mehrotra, C.K. Mohan and S. Ranka (1992) Forecasting the behavior of multivariate time series using neural networks, *Neural Networks*, **5**, 961-970.

Cottrell, M., B. Girard, Y. Girard, M. Mangeas, and C. Muller (1995) Neural modeling for time series: A statistical stepwise method for weight elimination, *IEEE Trans. Neural Networks*, **6**, 1355-1364.

Gaynor, P.E. and R.C. Kirkpatrick (1994) *Introduction to Time-Series Modeling and Forecasting in Business and Economics*, McGraw-Hill International, Singapore.

Johansson, E.M., F.U. Dowla, and D.M. Goodman (1992) Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method, *International Journal of Neural Systems*, **2**, 291-301.

Lactermacher, G. and J.D. Fuller (1995) Backpropagation in time-series forecasting, *Journal of Forecasting*, **14**, 381-393.

Makridakis, S. and S.C Wheelwright (1978) *Forecasting: Methods and Applications*, New York, Wiley/Hamilton.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler (1982) The accuracy of extrapolation (time-series) methods: Results of a forecasting competition, *Journal of Forecasting* **1**, 111-153.

Rajopadhye, M., M. Ben Chalia, P.P. Wang,, T. Baker and C.V. Eister (1999) Forecasting uncertain hotel room demand, In: *Proc. 1999 American Control Conference*, San Diego, California, 1925-1929.

Yao, X. (1999) Evolving artificial neural networks, *Proceedings of the IEEE*, **87**, 1423-1447.