

# Prediction of White Noise Time Series using Artificial Neural Networks and Asymmetric Cost Functions

Sven F. Crone

Institute of Information Systems, University of Hamburg  
Von-Melle-Park 5, D-20146 Hamburg, Germany

**Abstract** - Artificial neural networks in time series prediction generally minimise a symmetric statistical error, such as the sum of squared errors, to learn relationships from the presented data. However, applications in business elucidate that real forecasting problems contain non-symmetric errors. The costs arising from suboptimal business decisions based on over-versus underprediction are dissimilar for errors of identical magnitude. To reflect this, a set of asymmetric cost functions is used as objective functions for neural network training, deriving superior forecasts even for white noise time series. Some experimental results are computed using a multilayer perceptron trained with various asymmetric cost functions, evaluating the performance in competition to conventional forecasting methods on a white noise time series extracted from the popular airline passenger data.

## I. INTRODUCTION

Artificial neural networks (ANN) have found increasing consideration in forecasting theory, leading to successful applications in time series and explanatory sales forecasting [5,19,22]. In management, forecasts are a prerequisite for all decisions based upon planning [2]. Therefore, the quality of a forecast must be evaluated considering its ability to enhance the quality of the management decision – consequently using the monetary costs arising from decisions based on incorrect forecasts [18]. These costs from over- and underprediction are typically not quadratic in form and frequently non-symmetric [12]. Although White noted the interdependence between the selection of the objective function and the underlying business decision in one of the first applications of neural networks for prediction [24], neural network theory and applications consistently focus on symmetric statistical error functions for training. Originating from modest research in non-quadratic error functions in neural network theory [5,16,21] and asymmetric costs for the ex post evaluation in prediction theory [1,12,23,9], a set of asymmetric cost functions as alternative objective functions for business applications was recently introduced to neural network training [11]. In this paper, we analyse the efficiency of a linear asymmetric cost function in minimizing the actual error of a forecast decision, training a multilayer perceptron to find a cost efficient business decision for a stationary white noise time series.

Following a brief introduction to the use of neural networks for time series prediction of white noise processes, section 3 assesses statistical error measures and asymmetric cost functions for neural network training. This is followed

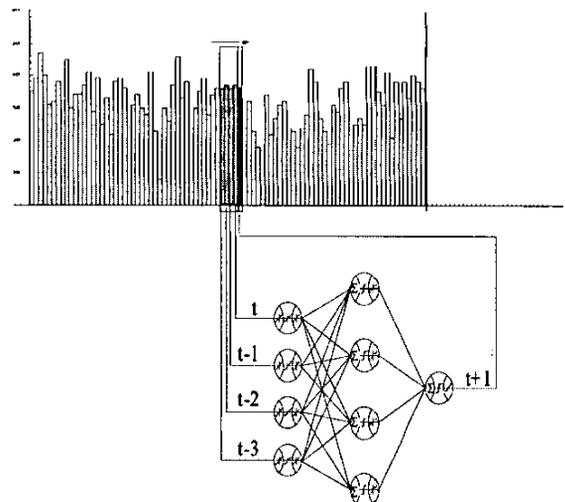
by an experimental evaluation of neural networks trained with asymmetric cost functions in competition with expert software-systems for time series prediction in section 4. Conclusions are given in section 5.

## II. NEURAL NETWORKS FOR WHITE-NOISE TIME SERIES PREDICTION

Forecasting time series with non-recurrent artificial neural networks (ANN) is generally based on modelling the network in analogy to a non-linear autoregressive AR( $n$ ) model [14]. At a point in time  $t$ , a one-step ahead forecast  $\hat{y}_{t+1}$  is computed using  $n$  observations  $y_t, y_{t-1}, \dots, y_{t-n+1}$  from  $n$  preceding points in time  $t, t-1, t-2, \dots, t-n+1$ , with  $n$  denoting the number of input units of the ANN. This models a time series prediction of the form

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-n+1}) \quad (1)$$

The architecture of a feed-forward multilayer perceptron (MLP) of arbitrary topology is displayed in figure 1. The task of the MLP is to model the underlying generator of the data during training, so that a valid forecast is made when the trained network is subsequently presented with a new value for the input vector [5]. Therefore the objective function used



**Fig. 1.** Neural Network application to time series forecasting with a (4-4-1)-MLP, using  $n=4$  input neurons for observations in  $t, t-1, t-2, t-3$ , four hidden units, one output neuron for time period  $t+1$  and two layers of 20 trainable weights [22] The bias node is not displayed.

for ANN training determines the resulting system behaviour and performance.

Following, we attempt to analyse the impact of alternative objective functions, i.e. asymmetric costs versus symmetric statistical errors, on the ex-post forecast performance. To control additional influences in the forecasting experiment we analyse a time series which is white noise. A white noise model represents a simple random model consisting of an overall level  $c$  and a random error component  $e_t$  which is uncorrelated from period to period: [14]

$$y_t = c + e_t \quad (2)$$

Considering the structure of a stationary white-noise model, lacking any systematic pattern in the residuals of  $e_t$ , it should prohibit one ANN to extract any underlying linear or nonlinear generator of the data and thus outperform competing ANNs or linear methods, ensuring an unbiased comparison of methods regardless of individual model performance. Experiencing random fluctuations, the ANN as other methods should predict the level  $c$  as the optimum predictor without overfitting to the training data.

### III. OBJECTIVE FUNCTIONS FOR NEURAL NETWORK TRAINING

#### A. Statistical Error Measures in Forecasting

Training a MLP is the task of adjusting the weights of the links  $w_{ij}$  between units  $j$  and adjusting their thresholds to minimize the error  $\delta_j$  between the actual and desired system behaviour [16] using various training algorithms for supervised online-training of a MLP. Gradient descent methods traditionally minimize a modified sum of squared errors (SSE),

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2 \quad (3)$$

ever since the popular description of the back-propagation algorithm by Rumelhart, Hinton and Williams [17]. The consistent use of the modified SSE in time series forecasting with ANN is motivated primarily by analytical simplicity [5,16] and the similarity of hetero associative neural network modelling to statistical regression problems, modelling the conditional distribution of the output variables [5].

In time series point prediction, the single network output  $o_p$  corresponds to the forecast  $\hat{y}_t$  of a network, the teaching input  $t_p$  represents the actual value of the observation  $y_t$  and the forecast error  $e_t$  represents the networks error  $\delta_j$  in the output-layer. The objective function in ANN training

$$E = f(t_j, o_j) \quad (4)$$

determining the size of the error  $\delta_j$  in the output-layer, may therefore be interpreted as measuring the significance of an overprediction or an underprediction for each point in time  $t$  or pattern  $p$ . For a business prediction, the objective function

should reflect the monetary significance in dependence of the underlying forecasting problem [16].

Although all statistical error measures produce a value of 0 for an optimal forecast and are symmetric about  $e_t = 0$ , each error measure implies a different weight for a deviation of the forecast value from the real value. Quadratic, cubic as other error measures of higher power penalize a forecast more for extreme deviations than for small ones [14]. In business applications, the errors arising from over- and underprediction are often considered to be not quadratic but linear in form, implying the use of absolute instead of quadratic error measures, which assign identical weight to every error regardless of scale [3,4]. Consequently, selected authors have applied absolute error-measures to neural network training in time series prediction. Thiesing [21] proposes the  $\ln(\cosh)$ -function as an approximation of the symmetric absolute error (AE) function. A comparison is given in Fig.(2).

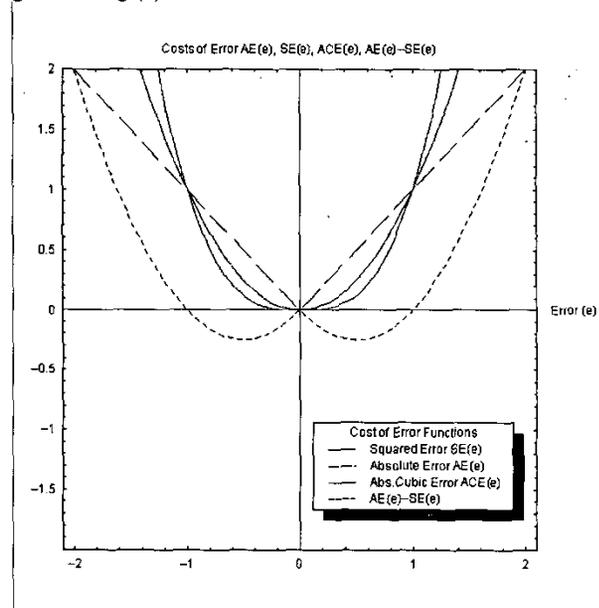


Fig. 2. Symmetric statistical error functions for ANN-training including the squared error  $SE(e)$ , absolute error  $AE(e)$ , absolute cubed error  $ACE(e)$  and the difference between the most two popular measures  $SE(e)$  and  $AE(e)$ .

Nonetheless, all error measures proposed apply symmetric error functions as an approximation of the true cost relationship. However, the cost arising in management forecasts are often not only non quadratic, but also non symmetric in form. Therefore a new set of error measures is introduced and applied directly to neural network training: asymmetric cost functions.

#### B. Asymmetric Cost Functions in Forecasting

In business management, all forecast are generated as a prerequisite of business decisions. Through decisions based on sub-optimal forecasts, costs arise to the decision maker in

choosing a inefficient alternative. Although the amount of costs will generally increase with the numerical magnitude of the errors, the sign of the error plays a significant role. Regarding business forecasts, the costs arising from over- and underprediction are frequently non symmetric and typically non quadratic in form [12,10]. In [11] we introduced a linear asymmetric cost function originally developed by Granger for forecasts in inventory management problems [13] to ANN training. The corresponding *LINLIN* cost function (*LLC*) yields:

$$LLC(y_{t+h}, \hat{y}_{t+h}) = \begin{cases} a|y_{t+h} - \hat{y}_{t+h}| & \text{for } y_{t+h} < \hat{y}_{t+h} \\ 0 & \text{for } y_{t+h} = \hat{y}_{t+h} \\ b|y_{t+h} - \hat{y}_{t+h}| & \text{for } y_{t+h} > \hat{y}_{t+h} \end{cases} \quad (5)$$

The *LLC* is linear to the left and right of 0. The parameters *a* and *b* give the slopes of the branches for each cost function and measure the costs of error for each stock keeping unit (SKU) difference between the forecast  $\hat{y}_{t+h}$  and the actual value  $y_{t+h}$ . The parameter *a* corresponds to an overprediction and the resulting stock-keeping costs while *b* relates to the costs of lost sales revenue for each underpredicted SKU. The shape of one asymmetric *LLC* as a valid linear approximation of a real cost function in inventory management is displayed in Fig. 3.

For  $a \neq b$  these cost functions are non-symmetric about 0 and are therefore called asymmetric cost functions. The degree of asymmetry is given by the ratio of *a* to *b* [9]. For  $a = b = 1$  the *LLC* equals the statistical error measure *AE*. The linear form of the ACF represents constant marginal costs arising from the business decision. This is consistent with the analysis of business decision based on linear marginal costs and profits.

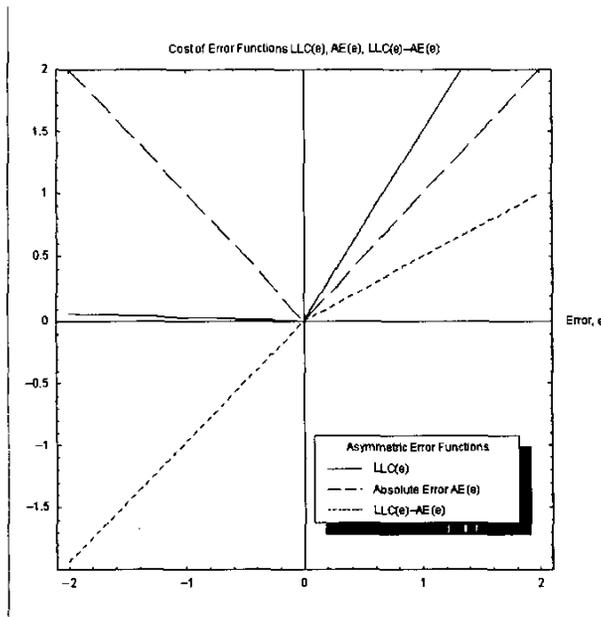


Fig. 3. Empirical Asymmetric Cost Function showing cost arising for over- and under-prediction, using  $a = \$0.01$  and  $b = \$1.50$ .

Although a small number of linear, non-linear and mixed ACFs have been specified [1,15,9,8], most notably the *LINEX*-cost function by Varian [23] consisting of linear and an exponential branch, they do not relate directly to business decisions incorporating linear marginal costs and have not yet been applied to ANN-training. All cost-functions developed imply a fixed functional relationship between error and cost, disregarding dynamic developments in business decisions.

Yang, Chan and King introduce a systematic extension to objective function classification, introducing dynamic non-symmetric margins for support vector regression [25,26]. An adaptation towards objective functions in ANN training is given in Table 1, allowing the classification of all symmetric statistical error functions and asymmetric cost functions previously developed. Variable objective functions with dynamic variation of costs through time and dependent on other variables are yet to be developed for ANN-training.

Table 1. Classification of objective functions for neural network training

Variability	Symmetry	
	Symmetric objective functions	Non-symmetric objective functions
Fixed	<i>SE, AE, ACE</i> statistical error functions	<i>LINLIN, LINEX</i> asymmetric cost functions
Variable		

Asymmetric transformations of the error function alter the error surface significantly, resulting in changes of slope and creating different local and global minima. Therefore, using gradient descent algorithms, different solutions are found minimizing cost functions instead of symmetric error functions, finding a cost minimum prediction for the underlying problem. These asymmetric cost functions (ACF) may be applied in ANN training using a simple generalisation of the error-term of the back-propagation rule and its derivatives, amending only the error calculation for the weight adaptation in the output layer [11] using

$$E_p = C(t_{pj}, o_{pj}) \quad (6)$$

In case of non fully-differentiable cost functions, approximations of the actual cost function, the application of alternative training methods or global search methods may be required to allow network training [16,11].

#### IV. SIMULATION OF NEURAL NETWORKS USING ASYMMETRIC COST FUNCTIONS

##### A. Experimental Time Series and Business Scenario

Following, we conduct an experiment to evaluate the ability of a MLP to evolve a set of weights minimizing an *LLC* asymmetric cost function for a random, stationary time series. The experiment is computed using a white-noise time series extracted from the original monthly airline passenger data, introduced by Brown [7] and extended by Box and Jenkins [6]. In [14], the monthly values of the Brown data-set, dating from 01/1949 unto 12/1956, are de-trended and

deseasonalised using a X-12-ARIMA Census II decomposition, leaving only the residuals. An analysis of the autocorrelations and partial autocorrelations confirms a stationary white noise model of the form (2) with  $c=124900$  and no data-pattern in the residuals  $e_t$ .

In order to specify the underlying costs arising from the decision process we require a suitable objective function. As no information on market prices, actual versus satisfied demand etc. exists for the historic data on airflight-tickets we propose a simple business model to exploit the airline data. An airline carrier needs to allocate planes of different sizes to match passenger demand of flights. We assume that flying with an empty seat is less costly than not selling a ticket due to lacking seats. This equals a higher marginal-profit than marginal costs, a prerequisite for most business transactions. Consequently, the costs  $b$  of underpredicting customer demand through lost sales-revenue are higher than those of overprediction  $a$ ,  $b > a$ , disregarding fixed costs of the decision. This corresponds to a simple inventory model without backordering. We generate business forecasts based upon ordinary least-squares predictors and asymmetric cost predictors to decide the cost efficient amount of passenger seats provided for each month, assessing the ex-post performance. In order to generalise findings we calculate results for two ratios of asymmetry  $a/b$ , although ratios with  $b < a$  seem implausible from a strategic business perspective.

## B. Experimental Design of Forecasting Methods

A small sample of  $n=96$  observations is split into three consecutive datasets, using 72 observations for the training-, 12 for the validation- and 12 for the out-of sample test-set, resulting in 59, 12 and 12 predictable patterns in each set. All data was scaled from a range of 90000 to 110000 to [-1;1] for internal processing. We consider a fully connected MLP without shortcut connections as displayed in Fig. 1, with a topology of 13 input, 12 hidden, 1 output node. All processing units use a summation as an input-, the tanh as a semilinear activation- and the identity- as an output function. Additionally, 1 bias unit models the thresholds for all units in the hidden and output layer. Three sets of networks were trained. Set  $ANN_S$  was trained on minimizing the symmetric  $SE$ , set  $ANN_{C1}$  was trained minimizing an asymmetric cost

function with the parameters ( $a=\$0.001; b=\$1.00$ ) for  $LLC1$  and set  $ANN_{C2}$  ( $a=\$16.00; b=\$1.00$ ) for  $LLC2$  respectively (7). For better comparison of results, the right-hand side of our  $LLC$ -function was chosen to equal the absolute error function  $AE$ . Each MLP was initialised and trained for five times to account for [-0.3;0.3] randomised starting weights. Training consisted of a maximum of 10000 epochs with a validation after every epoch, applying early stopping if the validation error did not decrease for 1000 epochs. After training, the results for the best network, chosen on its performance on the validation set, as well as the average of all five networks are computed for all data-sets. Only the test-set-data is used to measure generalisation, applying a simple hold out method for cross-validation. No parameters were estimated using data from the test-set.

As a standard benchmark in time series prediction, the Naïve1 method, using last periods sales as a forecast is used:

$$\hat{y}_{t+1} = y_t \quad (7)$$

Additionally we compute the results using the software Forecast Pro for exponential smoothing and ARxMA and Autobox for ARIMA intervention models. Each software selects appropriate models based upon statistical testing and expert knowledge. Drawing upon the Tashman-Hoover tables [20] and the M-competition results [14] they are considered benchmarks in time series prediction. All ANNs were simulated using NeuralWorks Professional and distinct error function tables to bias the calculated standard error.

## C. Experimental Results on Asymmetric Costs

Table 2 displays the results using mean error measures computed on each data-set to allow comparison between data-sets of varying length. The results are given in the form (training-set / validation-set / test-set) to allow interpretation. A descriptive performance measures of the  $\beta$ -service-level notes the amount of suppressed sales per dataset in relation to all demand. An asymmetric ex-post performance measure is calculated, denoting the ex post mean  $LINLIN$  costs ( $MLLC$ ) resulting from a given forecast method. As the costs themselves used for training and ex-post evaluation differ, those nets trained on an  $LLC$ -cost function are evaluated only

Table 2. Experimental Results for Statistical Forecasting Methods and ANNs trained on minimizing linear Asymmetric Costs and Squared Error Measures

	Error Measures			Service Level	Rank on test-set		
	$MSE(e)$	$MLLC1(e)$	$MLLC2(e)$	Beta	$MSE$	$LLC1$	$LLC2$
Naïve 1(random walk)	18.4 / 10.26 / 0.82	1.57 / 1.53 / 0.32	27.65 / 23.5 / 6.32	98.75% / 98.78% / 99.74%	4	6	4
Freefore Professional ARIMA with pulse	1.80 / 2.95 / 0.58	0.53 / 0.58 / 0.25	8.00 / 12.57 / 6.12	99.58% / 99.54% / 99.80%	2	4	3
ForecastProXE ARIMA (0,0,0)	10.03 / 4.39 / 0.50	1.02 / 0.92 / 0.23	17.15 / 11.7 / 5.72	99.19% / 99.27% / 99.82%	1	3	2
$ANN_S$ trained on $SE$ , best net (no.45)	10.80 / 4.48 / 0.68	0.96 / 0.77 / 0.16	20.54 / 12.3 / 9.67	99.23% / 99.38% / 99.88%	3	5	6
$ANN_S$ trained on $SE$ , average of 5 nets	10.91 / 4.90 / 0.82	1.07 / 0.91 / 0.27	20.01 / 13.5 / 8.58	99.14% / 99.28% / 99.78%	4	3	5
$ANN_{C1}$ trained on $LLC1$ , best net (no.48)	18.30 / 9.05 / 8.47	0.40 / 0.17 / 0.003	-	99.68% / 99.86% / 100.0%	8	1	-
$ANN_{C1}$ trained on $LLC1$ , average of 5	15.23 / 8.87 / 6.36	0.45 / 0.17 / 0.003	-	99.65% / 99.81% / 100.0%	5	1	-
$ANN_{C2}$ trained on $LLC2$ , bestnet (no.54)	19.54 / 13.3 / 7.65	-	8.09 / 4.22 / 2.63	97.50% / 97.54% / 97.89%	6	-	1
$ANN_{C2}$ trained on $LLC2$ , average of 5 net	22.60 / 12.5 / 8.29	-	11.2 / 5.99 / 6.97	97.56% / 97.87% / 97.89%	7	-	3

upon the  $MLLC$  costs they were trained on. All other Methods are evaluated on both cost-measures  $LLC1$  and  $LLC2$ . The methods are ranked by performance for the mean  $SE$  ( $MSE$ ) and the  $MLLC$ .

Various results may be drawn from the experiment. As expected, the  $ANN_S$  trained using the standard  $SE$  gives forecasts close to the white noise level  $c$ , displayed in Fig. 4.

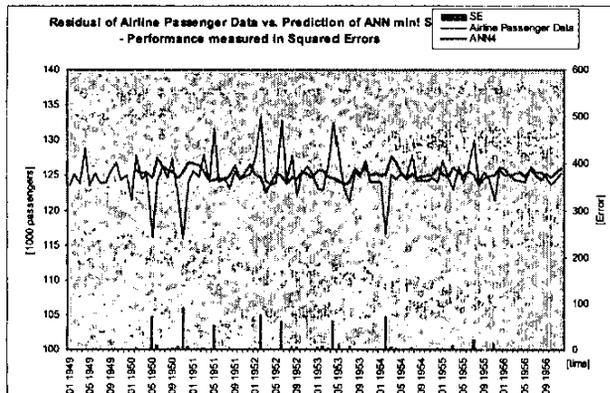


Fig. 4. ANN trained on minimizing the symmetric sum of squared error ( $SSE$ ) to forecast monthly airline passengers, showing the actual ticket sales, the ANN forecast and the ex-post forecast error measured by the  $SE(e)$ .

The best  $ANN_{C1}$  trained with the asymmetric  $LINLIN$  cost function  $LLC1$ , gives a superior forecast regarding the business objective, achieving the lowest mean costs on the test-data with 0.03. It exceeds all methods and clearly outperforms forecasts of ANN trained with the  $SSE$  criteria and the software expert systems. Analysing the behaviour of the forecast based upon the asymmetry of the costs function, the neural network  $ANN_{C1}$  in Fig.5 raises its predictions in comparison to  $ANN_S$  in Fig.4, to achieve a cost efficient forecast and a cost efficient inventory level, accounting for higher costs of underprediction versus overprediction and therefore avoiding costly stockouts. This is also evident in an increased  $\beta$ -service-level of 100.00%.

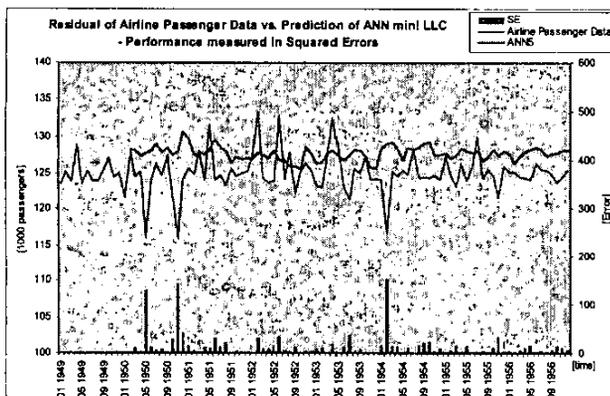


Fig. 5. ANN trained on minimizing the asymmetric cost function  $LLC1$  to forecast monthly airline passengers, showing the actual ticket sales, the ANN forecast and the ex-post forecast error measured by the  $SE(e)$ .

Alternatively, the  $ANN_{C2}$  trained on minimizing an asymmetric cost function with a ratio of  $b > a$  lowers its forecasts as seen in Fig.6, resulting in less overpredictions and more stock-outs, also corresponding to the asymmetric distribution of decision costs.

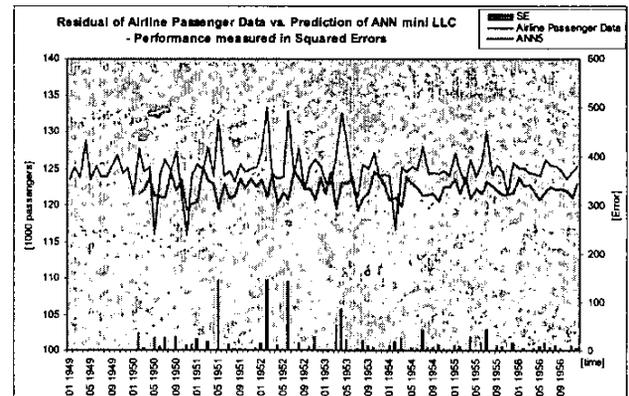


Fig. 6. ANN trained on minimizing the asymmetric cost function  $LLC2$  to forecast monthly airline passengers, showing the actual ticket sales, the ANN forecast and the ex-post forecast error measured by the  $SE(e)$ .

Consequently, the neural network no longer predicts the expected mean  $c$  of the white noise function but instead produces a biased optimum predictor, as proposed by Grangers original work [12]. This may be interpreted as finding a point on the conditional distribution of the optimal predictor depending on the distributions standard-deviation. For an inventory management problem, the network finds cost efficient inventory level without the separate calculation of safety stocks. This reduces the complexity of the overall management process of stock control, calculating a cost efficient stock level directly from the forecasting method.

Networks trained with cost functions show a significantly high  $MSE$  and a low  $MLLC$ , and vice versa. Performance on the cost measures is negatively correlated with the performance on squared error measures. Therefore, a MLP may be used successfully to minimize an arbitrary cost function distinct from the squared error paradigm. Additionally, all five networks in each set minimize the  $LLC$  or  $SE$  robustly, shown by the homogenous low performance measures over all nets in the set.

The experiment further validates, that choosing an ANN or any alternative forecasting method through an ex-post comparison of the  $MSE$  or other symmetric error measures may lead to the selection of an inferior forecast method, regarding the actual cost situation and, consequently, suboptimal decisions. Therefore, cost of error measures should be applied to the ex post evaluation of real-world forecasting applications, regardless of the training objective.

## V. CONCLUSION

We have examined symmetric and asymmetric error functions as performance measures for neural network training. The restriction on using squared error measures in neural network training may be motivated by analytical simplicity, but it leads to biased results regarding the final performance of forecasting methods. Asymmetric cost functions can capture the actual problem structure and allow the robust minimization of relevant costs using standard multilayer perceptrons and training methods, similar to minimizing an arbitrary statistical error function. Our approach to train neural networks with asymmetric cost functions has a number of advantages. Minimizing an asymmetric cost function allows the neural network to learn directly from actual cost functions, taking the model building process closer towards business reality. For instance, considerations of optimum service levels in inventory management may be incorporated within the forecasting process, leading directly to the forecast of a cost minimum stock level without further computations.

However, the limitations and promises of using asymmetric cost functions with neural networks require systematic analysis. Future research may incorporate the modelling of dynamic carry-over-, spill-over-, threshold- and saturation-effects for exact asymmetric cost functions where applicable. In particular, verification on multiple time series, other network topologies and architectures is required, in order to evaluate current research results.

## REFERENCES

- [1] G. Armingier, N. Götz, Asymmetric Loss Functions for Evaluating the Quality of Forecasts in Time Series for Goods Management Systems - Technical Report 22, Dortmund, 1999
- [2] J.S. Armstrong, Strategic Planning and Forecasting Fundamentals, in K. Albert (eds.), The Strategic Management Handbook, McGraw-Hill, New York, 1983, 2, 2-1 - 2-32
- [3] J.S. Armstrong, Error Measures for Generalizing About Forecasting Methods - Empirical Comparisons, International Journal of Forecasting, 1992, 8, 69-80
- [4] J.S. Armstrong, Evaluating Forecasting Methods, in J. S. Armstrong (eds.), Principles of Forecasting - A Handbook for Researchers and Practitioners, Kluwer Academic, Boston, 2001, 443-472
- [5] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, 1995
- [6] G. Box, G. Jenkins, Time Series Analysis - Forecasting and Control, Holden-Day, San Francisco, 1970
- [7] R.G. Brown, Statistical Forecasting for Inventory Control, McGraw-Hill, New York, 1959
- [8] P.F. Christoffersen, F.X. Diebold, Further Results on Forecasting and Model Selection under Asymmetric Loss, Journal of applied econometrics, 11, 1996, 5, 561-572
- [9] P.F. Christoffersen, F.X. Diebold, Optimal Prediction under General Asymmetric Loss, Econometric Theory, 13, 1997, 808-817, 1997
- [10] M.P. Clements, D. F. Hendry (eds.), Forecasting economic time series, Cambridge University Press, Cambridge, 1998
- [11] S.F. Crone, Training Artificial Neural Networks for Time Series Prediction using Asymmetric Cost Functions, in L. Wang, C. Rajapakse, K. Fukushima, S.-Y. Lee, X. Yao (eds.), Computational Intelligence for the E-Age - Proceedings of the 9th International Conference on Neural Information Processing ICONIP'02, Nov. 18-22 2002, Singapore, 2002, III, III-1301-III-1308
- [12] C.W.J. Granger, Prediction with a Generalized Cost of Error Function, Operational Research Society (Hrsg.), Operational Research Quarterly, 20, 1969, S. 199-207
- [13] C.W.J. Granger, Forecasting in Business and Economics, Academic Press, New York, 1980
- [14] S. Makridakis, S. C. Wheelwright, R. J. Hyndman, Forecasting - Methods and Applications, Wiley, New York, 1998
- [15] Parsian, N. S. Farsipour, Estimation of the mean of the selected population under asymmetric loss function, O. Anderson, et al. (Hrsg.), Metrika, International Journal for theoretical and applied statistics, 50, 1999, 2, 89 - 107
- [16] R.D. Reed, R.J. Marks II, Neural Smoothing - Supervised Learning in Feedforward Artificial Neural Networks, MIT Press, Cambridge, MA, 1998
- [17] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors - (reprint), in J.A. Anderson, E. Rosenfeld (eds.), Neurocomputing - Foundations of Research, MIT Press, Cambridge, MA, 1988, I, 696-699
- [18] J. Schwarze, Statistische Kenngrößen zur Ex-Post-Beurteilung von Prognosen - (Prognosefehlermaße), in J. Schwarze (eds.), Angewandte Prognoseverfahren, Neue Wirtschafts-Briefe, Herne, 1980, 317-344
- [19] Z. Tang, P.A. Fishwick, Feed-forward Neural Nets as Models for Time Series Forecasting - Technical Report TR91-008 Computer and Information Sciences, University Florida, 1991
- [20] L.J. Tashman, J. Hoover, Diffusion of Principles through Software, in J. S. Armstrong (eds.), Principles of Forecasting - A Handbook for Researchers and Practitioners, Kluwer Academic, Boston, 2001, 651-676
- [21] F.M. Thiesing, Analyse und Prognose von Zeitreihen mit Neuronalen Netzen, Shaker, Aachen, 1998
- [22] F.M. Thiesing, O. Vornberger, Sales Forecasting using Neural Networks, in IEEE (eds.), Proceedings ICNN'97 - Houston, Texas, 9-12 June 1997, 1997, 4, 2125-2128
- [23] H. White, Economic Prediction using neural networks - the case of IBM daily stock returns (reprint), in J. A. Anderson, E. Rosenfeld (eds.), Neurocomputing - Foundations of Research, MIT Press, Cambridge, MA, 1988, II, II451-II-459
- [24] H.R. Varian, A Bayesian approach to real estate assessment, in S.E. Fienberg, A. Zellner (eds.), Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage, North-Holland Publishing, Amsterdam, 1975, 195-208
- [25] H. Yang, I. King, L. Chan, Non-fixed Asymmetrical Margin Approach to Stock Market Prediction using Support Vector Regression, in L. Wang, C. Rajapakse, K. Fukushima, S.-Y. Lee, X. Yao (eds.), Computational Intelligence for the E-Age - Proceedings of the 9th International Conference on Neural Information Processing ICONIP'02, Nov. 18-22 2002, Singapore, 2002, 1398-1403
- [26] H. Yang, I. King, L. Chan, Support Vector Machine Regression for Volatile Stock Market Prediction, in H. Yin (eds.), IDEAL 2002, Springer Verlag, Heidelberg, 2002, 391-396,