

ePub^{WU} Institutional Repository

Christian Helmenstein and Christian Häfke

Neural Networks in the Capital Markets: An Application to Index Forecasting

Paper

Original Citation:

Helmenstein, Christian and Häfke, Christian (1995) Neural Networks in the Capital Markets: An Application to Index Forecasting. *Department of Economics Working Paper Series*, 32. Inst. für Volkswirtschaftstheorie und -politik, WU Vienna University of Economics and Business, Vienna.

This version is available at: <http://epub.wu.ac.at/6308/>

Available in ePub^{WU}: May 2018

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

***Neural Networks in the Capital Markets:
An Application to Index Forecasting***

*Christian Häfke
Christian Helmenstein*

*Working Paper No. 32
January 1995*

Address of authors:

Christian Häfke

***Department of Economics
Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria***

***voice: +43 1 59991 150
email: chris@ihssv.wsr.ac.at***

Christian Helmenstein

***Department of Economics
Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria***

***voice: +43 1 59991 254
email: helmen@ihssv.wsr.ac.at***



Abstract. In this article we construct an Index of Austrian Initial Public Offerings (IPOX) which is isomorph to the Austrian Traded Index (ATX). Conjecturing that the ATX qualifies as an explaining variable for the IPOX, we investigate the time trend properties of and the comovement between the two indices. We use the relationship to construct a neural network and a linear error-correction forecasting model for the IPOX and base a trading scheme on either forecast. The results suggest that trading based on the forecasts significantly increases an investor's return as compared to Buy and Hold or simple Moving Average trading strategies.

1. Introduction

Research in nonparametric, nonlinear methods has been extensive in recent years. Three popular examples of these approaches are projection pursuit regression (Friedman and Stuetzle, 1981), radial basis functions (Powell, 1987) and multilayer feedforward networks (Rumelhart *et al*, 1986). In this paper we shall focus on multilayer feedforward networks. The two most important areas of application of such networks are pattern classification and function approximation. Hornik, Stinchcombe and White (1989, 1990) show that neural networks are universal approximators and can learn arbitrary functions. A review of the theoretical findings is presented by Kuan and White (1994).

We apply a certain type of neural networks, an augmented single hidden layer feedforward network, to Initial Public Offerings which have gained rather little attention in financial econometrics so far. In this paper we construct an index of Initial Public Offerings (IPOX) which is isomorph to the Austrian Traded Index (ATX). We conjecture that the ATX constitutes a powerful explaining variable for the IPOX. First, we predict the ATX using linear and neural network models. In a second step we estimate the IPOX one day ahead based on observed ATX data. We compare the quality of this estimation to an IPOX forecast based on forecasted ATX values in a third step. For all predicting purposes in this paper we estimate linear models as well as neural networks, as the latter are becoming increasingly common in financial forecasting as well (for example (NNCM, 1994), (NNCM, 1993), (Trippi and Turban, 1993)). Section 2 analyses the data. Sections 3 and 4 present the models used. Section 5 discusses the error measures and the forecasting results, and section 6 contains concluding remarks.

2. Data

2.1. THE AUSTRIAN TRADED INDEX

At the foundation of the Austrian Futures and Options Exchange, ÖTOB, the Austrian Traded Index (ATX) was formed as a modern and reliable stock index for the Vienna Stock Exchange. The ATX serves as both a basis for futures and options contracts and as a market indicator which reflects a representative and liquid market segment of Austrian stocks with about 70% of total stock market activity (ÖTOB, 1994, 1991).

Three factors are essentially connected with the construction of a stock market index, viz. the selection of stocks, their weights, and the calculation method. The selection of stocks for the ATX follows the criteria of continuous trading, high market capitalisation, and a sufficiently high free float. The weight of a particular stock represents its equity market capitalisation. Thus, a title with a high market capitalisation (that is the number of stocks issued times the rate) has a larger impact on the index than a stock with a lower market capitalisation. The market capitalisation itself is corrected by a free float factor which ensures that the weight of a particular title in the ATX corresponds to the equity actually available for public trading at the stock exchange. The ATX is calculated according to the following formula

$$ATX_t = ATX_{t-1} \left[\frac{\sum_{i=1}^n P_{i,t} Q_{i,t-1}}{\sum_{i=1}^n P_{i,t-1} Q_{i,t-1}} \right], \quad (1)$$

with ATX_t as the ATX value at time t , $P_{i,t}$ the price of share i at time t , $Q_{i,t-1}$ the number of shares of stock i issued (corrected by the free float

factor), and n the number of stocks in the ATX. While the ATX reflects all price changes due to market fluctuations, technical price changes do not affect the index. For this reason the prices of the underlying shares are adjusted to changes in the dedicated capital or dividend payments. The basis of the ATX is 1000.00 as per January 2nd, 1991, the ATX value as of October 14th, 1994 is 1060.43. The data used for estimation is of length 487, starts on November 2nd, 1992 and ends on October 14th, 1994. The last 100 observations were used as test set for the out of sample error measures.

2.2. THE INITIAL PUBLIC OFFERINGS INDEX

The projections of expected future profits by the company itself are a distinctive feature of Austrian IPOs. As the underwriting bank(s) can, beside others, be held liable for wrong or misleading statements, the prospectus contains relatively more comprehensive and reliable information than any other information source available to the outside investor. This feature of Austrian IPOs should affect their average performance in terms of lower volatility as compared to the market average, since the discounted value of future profits is less uncertain.

In order to render the ATX and the IPOX comparable to each other, that is, to exclude a systematic deviation of the IPOX from the ATX, the IPOX is constructed isomorphically to the ATX. The IPOX covers all initial public offerings in the official market segment, including newly issued stock of companies whose stock other than the new category has been listed earlier. Initial public offerings in the regulated and unregulated market segments are excluded from consideration.

We also disregard *Investitionskredit AG* and the three different categories of *Bank Austria AG* stock. As opposed to the other banks in the index, *Investitionskredit AG* has been founded to support investment in top priority projects as determined by economic policy (*Österreichische Investitionskredit AG*, 1994, p. 15). Being subject to governmental predilection the risk structure of its business and thus of its shares is unlikely to be reproducible for any other Austrian company and would therefore unduly distort the comparison between the ATX and the IPOX. Being concerned with the particular portfolio of Austrian IPOs we also exclude *Bank Austria AG* for the sake of analytical clarity. The reason is that this title represents a particular portfolio of Austrian companies itself due to numerous affiliates. This argument is valid a fortiori as *Bank Austria AG* stocks would account for 48.52% of the total capital covered by the IPOX at the beginning of the sample period, November 1992¹.

¹For the composition of the IPOX cf. Haefke and Helmenstein (1994)

Each individual initial public offering enters the IPOX with the first rate in public trading and *not* with the offering price. Considering the additional information as being a typical attribute of a share to be defined as an IPO, we have reason to expect that this status will vanish at the end of the forecasting horizon which is one and a half years on average. For this reason one and a half years after the first listing on the stock exchange a stock does no longer qualify as an IPO and it is withdrawn from the index. For estimating the parameters of our models we use the first 387 observations beginning on November 2nd, 1992. The remaining 100 observations are used to calculate the out of sample error measures.

2.3. AUTOCORRELATION AND CROSS-CORRELATION

After taking logarithms of the ATX and the IPOX we compute sample autocorrelations of their first differences. The results replicate the findings of Pichler (1993) that Austrian time series data of stock market indices exhibit significant sample autocorrelations of order 1 (Table 1). Our analysis confirms this outcome for another subsample of Austrian stocks, the initial public offerings.

TABLE 1. Autocorrelations for observations 1 to 487

Variable	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
IPOX	0.227	0.087	0.035	0.065	0.022	-0.005	0.012
ATX	0.240	0.033	-0.095	-0.002	0.061	0.007	-0.047

As the ATX covers the most liquid shares at the Vienna Stock Exchange, the ATX can be expected to reflect price changes due to new information most quickly. It might therefore qualify as an explaining variable for the IPOX. Support for this hypothesis comes from the computation of cross-correlations between the ATX and the IPOX.

Table 2 displays a statistically significant cross-correlation between the current IPOX value and the previous ATX value. These findings should subsequently be confirmed by an econometric model.

2.4. INTEGRATION AND COINTEGRATION

The usual asymptotic properties cannot be expected to apply if any of the variables in a regression model is generated by a nonstationary process. Using unit root tests we explore the time trend properties of the ATX and the IPOX series. If a series contains a stochastic trend, it is said to be

Figure 1. IPOX vs. ATX

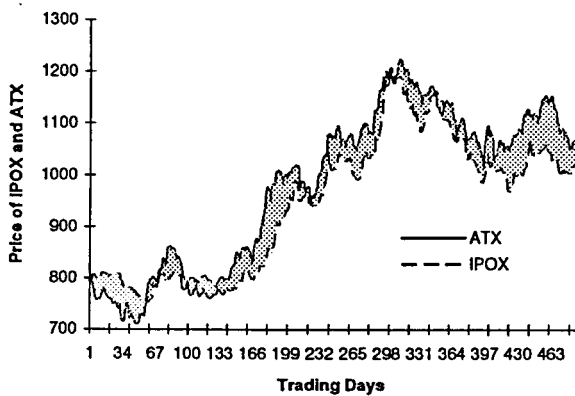


TABLE 2. Cross - correlations between the ATX and the IPOX

Lags	Lag -4	Lag -3	Lag -2	Lag -1	Lag 0
X - correlation	0.038	-0.045	0.039	0.075	0.458
Lags	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
X - correlation	0.225	0.105	0.020	0.075	0.069

integrated of order d , $I(d)$. Differencing d times then yields a stationary series.

Table 3 reports the results of Dickey-Fuller tests (DF) (Dickey and Fuller, 1979) and Augmented Dickey-Fuller tests (ADF) that the ATX and the IPOX series might have up to two unit roots. In no case there is significant evidence against the single unit root hypothesis. Thus the null hypothesis that both series are not stationary in levels cannot be rejected. All test statistics for a second unit root, that is a unit root in the first difference of the series, are highly significant. We therefore adopt the alternative hypothesis that the series are stationary in first differences. Since both series con-

TABLE 3. Tests for Integration

Series ²	Single Unit Root		Second Unit Root	
	DF	ADF	DF	ADF
ATX	-1.09	-1.13	-17.22	-13.87
IPOX	-1.03	-1.09	-17.27	-13.11

tain a stochastic trend we proceed with investigating whether they share a common stochastic trend. This refers to testing for cointegration which is a way of testing for a long-run equilibrium relationship between the ATX and the IPOX. Two variables are said to be cointegrated of order one, $CI(1,1)$, if they are individually $I(1)$ and yet some linear combination of the two is $I(0)$ (Engle and Granger, 1987). Under the assumption that a first order model is correct, we test whether the estimated residual of the cointegrating regression is stationary. Specifically, we perform ADF tests in order to test the null hypothesis that the residual series of the cointegrating regression is nonstationary. Reporting a value of -3.28 an ADF test with

²Critical values for 500 observations at the 1% and 5% significance level, respectively, are -3.44 and -2.87.

one lag and with the IPOX as the independent variable rejects the null of no cointegration at the 10% level (critical values for the ADF test are -3.34 and -3.04 at the 5% and 10% significance levels, respectively)³. Since the cointegrating vector establishes an equilibrium relationship, the ADF test should not lead to a different conclusion if the cointegrating equation is estimated invertedly, that is with the ATX as the independent variable. With a value of -3.30 the result indeed confirms this requirement.

3. Linear ATX and IPOX Models

Implementing the above findings we base the IPOX forecasts on a dynamic specification of a linear regression model with error-correction term.

$$dIPOX_t = \alpha_0 + \alpha_1 dIPOX_{t-1} + \alpha_2 dATX_{t-1} + \alpha_3 (IPOX - ATX)_{t-1} + \epsilon_t \quad (2)$$

The regression results (table 4, second column) give evidence that the current value of the IPOX is positively related to the previous values of

TABLE 4. The ATX as explaining variable for the IPOX

Independent variable	$dIPOX_t$	$dATX_t$
Intercept	-0.000158 (-0.3779)	0.000438 (0.8089)
$dIPOX_{t-1}$	0.1426** (2.6293)	-0.0780 (-1.1132)
$dATX_{t-1}$	0.0893* (2.0509)	0.2616** (4.6468)
$IPOX - ATX_{t-1}$	-0.0343** (-3.7765)	-0.00709 (-0.6047)
\bar{R}^2	0.0943	0.0546
DW	2.0141	1.9632
LM (p-value)	0.6422	0.1402
Ljung-Box Q	43.35	35.51
p-value of Q	0.1865	0.4916

the IPOX and the ATX. The highly significant value for the error-correction

³These critical values differ from those used above as the asymptotic distributions of residual-based cointegration test statistics are not the same as those of ordinary unit root test statistics (cf. Davidson and MacKinnon (1993), p. 720).

*Statistically significantly different from zero at the 0.05 significance level.

**Statistically significantly different from zero at the 0.01 significance level.

term $(IPOX - ATX)_t$ with a lag of order 1 reveals that deviations of the IPOX from the ATX cause a strong pull back tendency towards the ATX while the opposite does not hold. In order to check this finding, we use the same explaining variables as before to model the ATX in first differences,

$$dATX_t = \beta_0 + \beta_1 dIPOX_{t-1} + \beta_2 dATX_{t-1} + \beta_3 (IPOX - ATX)_{t-1} + \eta_t \quad (3)$$

Due to the insignificant values for the error-correction term and for the IPOX term (table 4, third column), the above result finds support.

In order to forecast the ATX one day ahead we choose an autoregressive process of order 1, AR[1], that is

$$dATX_t = \gamma_0 + \gamma_1 dATX_{t-1} + \varepsilon_t \quad (4)$$

Table 5 presents the coefficient estimates.

TABLE 5. Estimated coefficients for the linear ATX model

Explaining variables	dATX
Intercept	0.0005 (1.0782)
dATX(t-1)	0.2415** (4.8673)

4. Neural Network Models

The idea of modelling brain functionality goes back to McCulloch and Pitts (1943) who first introduced units that were constructed analogously to a neuron in the brain. By combining these neurons we arrive at a linear perceptron, or ADALINE as it was first called by Widrow and Hoff, (1960).

$$f(\tilde{x}_t, \alpha) = G(\tilde{x}_t' \alpha), \quad (5)$$

with \tilde{x}_t being the input vector augmented by a constant and α a set of weights. By taking a closer look at the formula, we see that for $G(x) = x$ we arrive at the simple linear model which is a standard paradigm in economic and econometric modelling. Kuan and White (1994) point out, that for $G(x) = \frac{1}{1+\exp^{-x}}$ we arrive at the binary logit model and for $G(x)$ being any

**Statistically significantly different from zero at the 0.01 significance level.

normal cumulative distribution function we obtain a binary probit. Hence even at the outset of neural network modelling, standard econometric models could easily be included as special cases. With these possibilities, the perceptrons' popularity kept increasing until Minsky and Papert 1969 published their book *Perceptrons* in which they pointed out that perceptrons were only capable of solving linearly separable problems. A way to overcome the separability problem was obtained by looking at nature again. It is very rare that signals directly flow from the sending to the receiving cell. They usually pass a number of intermediate layers. These intermediate layers were also adopted by the neural network community (Rosenblatt, 1958). However, until Werbos (1974) and Rumelhart *et al.* (1986) there was no way of actually estimating such networks. The output produced by such a multilayer perceptron is given by:

$$f(\tilde{x}_t, \beta, \gamma) = F \left(\sum_{j=1}^q G(\tilde{x}_t' \gamma) \beta \right). \quad (6)$$

Standard neural network learning algorithms use incremental updates of the form

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \eta \nabla f(\tilde{x}_t, \hat{\theta}_t) (y_t - f(\tilde{x}_t, \hat{\theta}_t)) \quad (7)$$

with \tilde{x} denoting the input vector x augmented by a constant, and θ denoting a weight vector. White (1987) pointed out that this is actually a form of stochastic approximation (Robbins and Monro, 1951) where η is fixed over time instead of being dependent on t . For further discussion see Kuan and White (1994). There are now a number of approaches that explicitly allow for a time varying learning rate η . It has been useful to start with a high η and slowly decrease it which incorporates a special form of simulated annealing. Many different forms of neural networks are successfully applied to time series data with the simple single hidden layer network being one of them (White, 1988), (Lee and Park, 1992). However, it has frequently been noted that performance sometimes degrades after adding a hidden layer as compared to a simple perceptron. To avoid these shortcomings we use an augmented single hidden layer feedforward neural network as proposed for example by Swanson and White (1992) which covers all the models discussed so far and thus constitutes a very flexible model for econometric tasks. This structure incorporates a simple perceptron and a simple single hidden layer network. Therefore the output is calculated as follows:

$$f(\tilde{x}_t, \theta) = \tilde{x}_t' \alpha + \sum_{j=1}^q G(\tilde{x}_t' \gamma) \beta \quad (8)$$

with \tilde{x} denoting the input vector x augmented by a constant, and θ denoting a weight vector containing the weights α, β, γ , that is $\theta = (\alpha', \beta', \gamma)'$,

$\beta = (\beta_1, \beta_2, \dots, \beta_q)'$, $\gamma = (\gamma_1', \dots, \gamma_q')$. q is the number of hidden units and G is a nonlinear function, in this case

$$G(x) = \frac{2}{1 + \exp^{-x}} - 1, \quad (9)$$

mapping x into the $]-1; +1[$ interval. This architecture not only allows to capture the nonlinearity in the data but also makes use of the well known linear regression approach and therefore ensures that the ANN will in sample perform at least as good as a linear model. If the input - output connections were dropped, this could not be guaranteed. Training takes place in two steps. First, the direct input-output connections α are estimated through OLS and fixed. In a first step we estimate

$$f(\tilde{x}_t, \alpha) = \tilde{x}_t' \alpha + \epsilon_t \quad (10)$$

with \tilde{x} being the input vector x augmented by a constant, α the corresponding weight vector, and ϵ_t the vector of the residuals. Matrices β and γ are estimated to model the residuals of the linear regression with any nonlinear optimisation technique. This approach generally improves the performance over OLS. In a second step we solve the problem:

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^n (y_t - f(\tilde{x}_t, \theta))^2 \quad (11)$$

with α fixed. Our programme is designed to find the optimal number of hidden units itself, using SIC (Schwartz, 1978) as suggested in Swanson and White (1992). This approach helps increasing the generalisation performance of the net, which is a topic of keen interest (for example (Moody, 1992), (Teräsvirta and Lin, 1993), (Moody and Utans, 1994)).

5. Empirical Results

The quality of our results is evaluated using the following out of sample error measures:

- out of sample MSE
- out of sample R^2
- Theil's coefficient of inequality

$$Theil = \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - y_{t-1})^2} \quad (12)$$

This measure constitutes a simple sanity check of our forecasts against a no-change forecast which performs better for $Theil > 1$ (Theil, 1966);

– Normalised Mean Squared Error

$$NMSE = \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2} \quad (13)$$

$NMSE$ is a second sanity check against the out of sample mean of the dependent variable. This measure is used by Weigend and Gershenfeld (1994) to evaluate entries into the Santa Fe Time Series Competition;

– Confusion Matrix

The forecasts, obtained through a feedforward pass through the network, are then evaluated, and the up and down signals of the net are used to compute a confusion matrix as in Swanson and White (1992). We find the number of correct classifications in the main diagonal and the errors off the diagonal. The rows contain the actual ups and downs, while the columns contain the forecasts. The confusion rate is calculated as the sum of the off diagonal elements divided by the total number of elements. A binomial test is performed to check if the number of correct classifications differs significantly from 50%;

– Trading Scheme

We apply a very simple and conservative trading scheme without transaction costs. We start out on the first day of the evaluation period. If the forecast for the following day indicates a rise in prices and we do not yet hold the IPOX - portfolio, we buy. In case that we already hold it we do not buy again but just keep still. In case of falling prices we sell if we hold but never go short. Returns are annualised and compared to a Buy and Hold strategy and to a case of perfect foresight which represents the maximum return achievable with this strategy;

– Moving Average Trading Rule

We also compare our returns against the returns generated by a 2-50 MA-Trading Rule. If the short MA intersects the long MA from below we receive a buy signal and keep the portfolio until the two moving averages intersect again and vice versa.

– t-values for returns of the Trading Scheme

In order to test whether the returns generated through the trading scheme are significantly different from the Buy and Hold strategy, t-values are computed according to the following formula (Brock *et al*, 1991):

$$t = \frac{\mu_t - \mu_b}{\sqrt{\frac{\sigma^2}{N_t} + \frac{\sigma^2}{N}}} \quad (14)$$

with μ being the mean returns of the two series, σ^2 the estimated variance for the entire sample, N_t the number of days a stock is held under the trading scheme, and N the number of observations.

Table 6 presents the results for the ATX forecast. We see that the

TABLE 6. Results of out of sample ATX forecasts

Error measures	Linear model	ANN
MSE	95.291	96.759
R^2	0.921	0.919
Theil	0.953	0.967
NMSE	0.079	0.081
Confusion Matrix	$\begin{bmatrix} 28 & 21 \\ 21 & 30 \end{bmatrix}$	$\begin{bmatrix} 29 & 21 \\ 20 & 30 \end{bmatrix}$
t-values	(1.62)	(1.83)

nonlinear model with 1 lag and 1 hidden unit performs as well as the AR[1] with regard to the MSE criteria. By looking at the Theil measure, we still detect a great potential for improving the forecast which will be left for further work.

TABLE 7. Results of out of sample IPOX forecasts with a linear model

Error measures	Linear ATX model	ANN ATX	Observed ATX
MSE	57.646	57.698	54.879
R^2	0.888	0.888	0.894
Theil	1.008	1.009	0.960
NMSE	0.119	0.119	0.113
Confusion Matrix	$\begin{bmatrix} 30 & 23 \\ 20 & 27 \end{bmatrix}$	$\begin{bmatrix} 30 & 23 \\ 20 & 27 \end{bmatrix}$	$\begin{bmatrix} 30 & 22 \\ 20 & 28 \end{bmatrix}$
t-values	(1.41)	(1.41)	(1.62)

We compare the findings of the linear model as reported in table 7 to the results obtained through an estimation of the IPOX with an artificial neural network with 1 hidden unit in table 8. It turns out that, given the ATX forecasting method, the IPOX forecast based on an ANN model outperforms the linear IPOX forecast except for the linear ATX. By contrast, given the IPOX forecasting method, the ANN ATX forecast matches or beats the linear ATX forecast. We see that for the linear IPOX the ANN does not boost the performance of the IPOX forecasts as all three

TABLE 8. Results of out of sample IPOX forecasts with ANN model

Error measures	Linear ATX model	ANN ATX	Observed ATX
MSE	79.263	51.226	50.783
R^2	0.836	0.896	0.896
Theil	1.386	0.896	0.888
NMSE	0.164	0.106	0.105
Confusion Matrix	$\begin{bmatrix} 27 & 24 \\ 23 & 26 \end{bmatrix}$	$\begin{bmatrix} 29 & 21 \\ 21 & 29 \end{bmatrix}$	$\begin{bmatrix} 30 & 22 \\ 20 & 28 \end{bmatrix}$
t-values	(0.60)	(1.62)	(1.62)

models are of approximately equal quality with the one based on the original ATX series slightly outperforming chance as well as the other models. From the ANN IPOX forecast we conclude that the ANN ATX forecasts seem to contain some nonlinear information absent in the linear ATX forecasts. Therefore the ANN IPOX estimation outperforms the linear IPOX estimation in all cases except for the linear ATX forecast where this information is missing. This result gives evidence for some nonlinearity in both the IPOX and the ATX data. The analysis of the trading schemes

TABLE 9. Summary statistics for returns of various forecasts

Estimation method	Cumulated returns	Number of transactions	t-value (vs. Buy&Hold)	t-value (vs. 2-50 MA)
Linear IPOX, linear ATX	0.277	44	9.650	5.869
Linear IPOX, ANN ATX	0.277	44	9.650	5.869
Linear IPOX, orig. ATX	0.283	43	8.900	5.280
ANN IPOX, linear ATX	0.242	31	9.650	5.783
ANN IPOX, ANN ATX	0.254	45	9.650	5.318
ANN IPOX, orig. ATX	0.252	43	8.900	5.374
MA 2-50	0.011	26	5.227	n.a.
Buy & Hold	-0.002	2	n.a.	-5.227

yields the results presented in table 9. As benchmarks we use a simple Buy and Hold strategy, the 2-50 Moving Average and - in order to determine the highest possible return - a strategy based on perfect foresight of the next period's IPOX value. Considering the linear approach to forecast the IPOX, we compare three datasets. Two of the underlying ATX series are

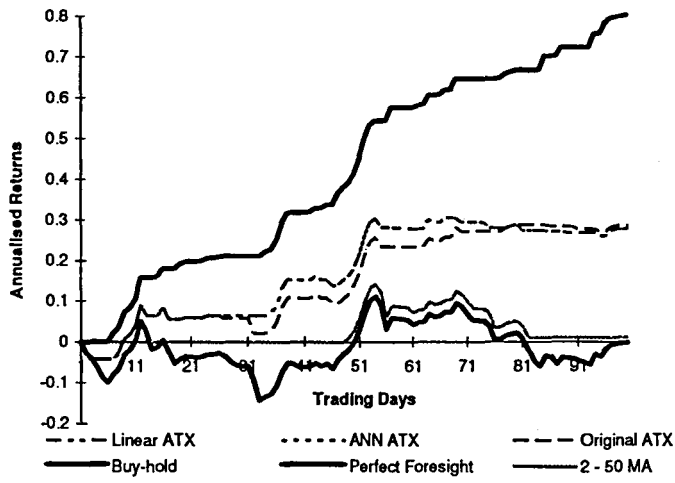
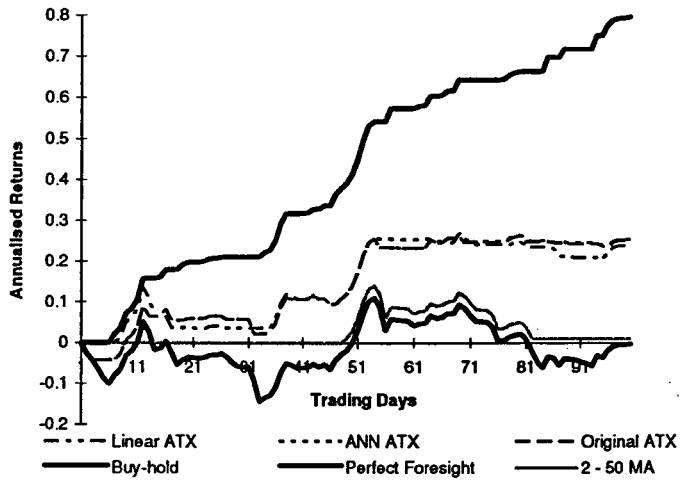
Figure 2. Annualised Cumulated Returns for the Linear IPOX

Figure 3. Annualised Cumulated Returns for the ANN IPOX



generated by a linear model and a neural network. The third one is made up by the original ATX series. All of the datasets are very close to each other and all of them significantly beat the Buy and Hold strategy as well as the Moving Average. Second, modelling the IPOX with an ANN, we achieve similar results. All estimates are quite close again and far better than the returns based on Buy & Hold or the Moving Average strategy. Compared to the first case the profitability of all forecasts is lower.

6. Conclusions

Both the linear as well as the ANN ATX forecasts yield better results than a no-change forecast. Note that as far as the ATX is concerned, the ANN does not outperform the linear model. A multivariate ATX forecast might, however, improve the significance level of the correct classifications. Any market participant capable of forecasting the ATX one day ahead should be able to take advantage of the property of the ATX to qualify as an explaining variable for the IPOX. The inspection of the results of a simple trading scheme reveals that nonzero profits can be expected by applying any of the IPOX models. Recall that the models - depending on the forecast - generate between 31 and 45 trading signals in just 100 days. Thus it remains to be inquired whether these profits sustain in an environment where transaction costs are accounted for.

It is a well-established result that models with low *Theil* and *NMSE* are not the best ones for forecasting the sign of the index movement whereas those with good confusion rates are not optimal with respect to the accurate approximation of the index. Hence the choice of an adequate error function strongly depends on the objective of the forecast. As in this case we try to approximate the (co)movement of the ATX and IPOX as good as possible, the MSE is the appropriate error function. For other purposes different error functions may be considered, such as minimising the confusion rate or maximising profits obtained through transactions based on the above trading rules.

Acknowledgements

The authors gratefully acknowledge constructive comments provided by Raimund Alt, Bernhard Felderer, Michael Hauser, Robert Kunst, and Andreas Wörgötter. We are indebted to Franz Köstl from Austrian Kontrollbank AG for supplying the stock market data and to Silvia Aigner for preparing spreadsheet macros.

References

- BROCK, W., LAKONISHOK, J., LeBARON, B. (1991): Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *Journal of Finance*, vol 47, pp 1731-1764.
- DAVIDSON, R. and MacKinnon, J.G. (1993): Estimation and Inference in Econometrics, Oxford University Press: New York and Oxford.
- DICKEY, D.A. and FULLER, W.A. (1979): Distribution of the Estimators for Autoregressive Time Series with a Unit Root *Journal of the American Statistical Association* vol 74, pp 427-431.
- ENGLE, R.F. and GRANGER, C.W.J. (1987): Co-integration and Error Correction: Representation, Estimation and Testing *Econometrica* vol 55, pp 251-276.
- FRIEDMAN, J.H. and STUETZLE, W. (1981): Projection Pursuit Regression, *Journal of the American Statistical Association*, vol 76, pp 817-823.
- HAEFKE, C., HELMENSTEIN, C. (1994): IPOX - An Initial Public Offerings Index, *Working Paper N 351, Institute for Advanced Studies: Vienna*.
- HORNIK, K., STINCHCOMBE, M., and WHITE, H. (1989): Multi-Layer Feedforward Networks Are Universal Approximators, *Neural Networks*, vol 2, pp 359-366.
- HORNIK, K., STINCHCOMBE, M., and WHITE, H. (1990): Universal Approximation of an Unknown Mapping and Its Derivatives, *Neural Networks*, vol 3, pp 551-560.
- KUAN, C., and WHITE, H. (1994): Artificial Neural Networks: An econometric perspective, *Econometric Reviews*, vol 13, N 1, pp 1-91.
- LEE, K., and PARK, J. (1992): Short-term Load Forecasting Using an Artificial Neural Network, *IEEE Transactions on Power Systems*, vol 7, N 1, pp 124-132.
- McCULLOCH, W.S., and PITTS, W. (1943): A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, vol 5, pp 115-133.
- MINSKY, M., PAPERT, S. (1969): Perceptrons: An Introduction to Computational Geometry, MIT Press: Cambridge, MA.
- MOODY, J. (1992): The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems, *Advances in Neural Information Processing Systems 4*, Morgan Kaufmann Publishers: San Mateo, CA, 1992.
- MOODY, J., UTANS, J. (1994): Architecture Selection Strategies for Neural Networks: Application to Corporate Bond Rating Prediction, in: REFENES, A.N.: *Neural Networks in the Capital Market*, John Wiley & Sons: New York, forthcoming.
- NNCM93 (1993): Neural Networks in the Capital Markets, Conference at the London Business School, November 1993.
- NNCM94 (1994): Neural Networks in the Capital Markets, Conference at the California Institute of Technology, Pasadena, CA, November 1994.
- Österreichische Investitionskredit AG (1994): Annual Report 1993, Österreichische Investitionskredit AG: Vienna.
- ÖTOB (1994, 1991): ATX Austrian Traded Index, ÖTOB Clearing Bank AG: Vienna.
- PICHLER, S. (1993): Size Effect and Settlement Effect am österreichischen Aktienmarkt (Size Effect and Settlement Effect at the Austrian Stock Market), *Österreichisches BankArchiv*, vol 3, pp 195-201.
- POWELL, M.J.D. (1987): Radial Basis Functions for Multivariable Interpolation: A Review, in: J.C. MASON and M.G. COX (eds.), *Algorithms for Approximation*, Clarendon Press: Oxford, pp 143-167.
- ROBBINS, H. and MONRO, S. (1951): A Stochastic Approximation Method, *Annals of Mathematical Statistics*, vol 22, pp 400-407.
- ROSENBLATT, R. (1958): The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Reviews*, vol 62, pp 386-408.
- RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J. (1986): Learning Internal Representations by Error Propagation, in: D.E. Rumelhart and J.L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol 1, MIT Press: Cambridge, MA, pp 318-362.

- SCHWARTZ, G. (1978): Estimating the Dimension of a Model, *Annals of Statistics*, vol 6, pp 461-464.
- SWANSON, N., and WHITE, H. (1992): A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks, *Research Group for Econometric Analysis, Department of Economics, University of California, San Diego*.
- TERÄSVIRTA, T., and LIN, C. (1993): Determining the number of hidden units in a single hidden-layer neural network model, *presented at the NBER Time Series Seminar, October 21, Institute for Advanced Studies, Vienna*.
- THEIL, H. (1966): *Applied Economic Forecasting*. North Holland Publishing Company: Amsterdam.
- TRIPPI, R., and TURBAN, E. (eds.) (1993): *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, Probus Pub. Co: Chicago, IL.
- WEIGEND, A.S., and GERSHENFELD, N.A. (eds.) (1994): *Time Series Prediction, Forecasting the Future and Understanding the Past, Proceedings in the SFI studies of complexity*, Addison-Wesley, Reading: MA.
- WERBOS, P. (1974): *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, PhD thesis, Harvard University, Department of Applied Mathematics.
- WHITE, H. (1987): Some Asymptotic Results for Back-Propagation, *Proceedings of the IEEE First International Conference on Neural Networks*, San Diego, vol 3, IEEE Press: New York, NY, pp 261-266.
- WHITE, H. (1988): Economic prediction using neural networks: the case of IBM daily stock returns, *Proceedings of the Second Annual IEEE Conference in Neural Networks*, San Diego, vol 2, IEEE Press: New York, pp 451-458.
- WHITE, H. (1989): An Additional Hidden Unit Test for Neglected Nonlinearity, *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., vol 2, IEEE Press: New York, pp 451-455.
- WHITE, H. (1990): Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings, *Neural Networks*, vol 3, pp 535-549.
- WIDROW, B. and HOFF, M.E. (1960): Adaptive Switching Circuits, *Institute of Radio Engineers WESCON Convention Record, Part 4*, pp 96-104.

Bisher sind in dieser Reihe erschienen:

- Eigl R., Experimentielle Methoden in der Mikroökonomik, No. 1, Mai 1991.
- Dockner E., Long N.V., International Pollution Control: Cooperative versus Non-Cooperative Strategies, No. 2, September 1991.
- Andraea C.A., Eigl R., Der öffentliche Sektor aus ordnungspolitischer Sicht, No. 3, Oktober 1991.
- Dockner E., A Dynamic Theory of Conjectural Variations, No. 4, Oktober 1991.
- Feichtinger G., Dockner E., Cyclical Consumption Pattern and Rational Addictions, No. 5, Oktober 1991.
- Marterbauer M., Die Rolle der Fiskalpolitik im Schwedischen Wohlfahrtsstaat, No. 6, Dezember 1991.
- Pichler E., Cost-Sharing of General and Specific Training with Depreciation of Human Capital, No. 7, Dezember 1991.
- Pichler E., Union Wage Bargaining and Status, No. 8, Dezember 1991.
- Pichler E., Costs of Negotiations and the Structure of Bargaining - a Note, No. 9, Dezember 1991.
- Nowotny E., The Austrian Social Partnership and Democracy, No. 10, Dezember 1991.
- Pichler E., Walther H., The Economics of Sabbath, No. 11, April 1992.
- Klatzer E., Unger B., Will Internationalization Lead to a Convergence of National Economic Policies?, No. 12, June 1992.
- Bellak C., Towards a Flexible Concept of Competitiveness, No. 13, May 1992.
- Koren St., Stiassny A., The Temporal Causality between Government Taxes and Spending, No. 14, August 1992.
- Altzinger W., Ost-West-Migration ohne Steuerungsmöglichkeiten?, No. 15, September 1992.
- Bellack Ch., Outsiders' Response to Europe 1992, Case of Austria, No. 16, December 1992.
- Guger A., Marterbauer M., Europäische Währungsunion und Konsequenzen für die Kollektivvertragspolitik, No. 17, January 1993.
- Unger B., van Waarden F., Characteristics, Governance, Performance and Future Perspectives, No. 18, January 1993.
- Scharmer F., The Validity Issue in Applied General Equilibrium Tax Models, No. 19, May 1993.
- Ragacs Ch., Minimum Wages in Austria: Estimation of Employment Functions, No. 20, June 1993.
- Ragacs Ch., Employment, Productivity, Output and Minimum Wages in Austria: A Time Series Analysis, No. 21, September 1993.
- Stiassny A., TVP - Ein Programm zur Schätzung von Modellen mit zeitvariierenden Parametern, No. 22, December 1993.
- Gstach D., Scale Efficiency: Where Data Envelopment Analysis Outperforms Stochastic Production Function Estimation, No. 23, December 1993.

Gstach D., Comparing Structural Efficiency of Unbalanced Subsamples: A Resampling Adaptation of Data Envelopment Analysis, No. 24, December 1993.

Klausinger H., Die Klassische Ökonomie und die Keynesianische Alternative. Revision ein Mythos?, No. 25, December 1993.

Grandner T., Gewerkschaften in einem Cournot-Duopol. Sequentielle versus simultane Lohnverhandlungen, No. 26, April 1994.

Stiassny A., A Note on Frequency Domain Properties of Estimated VARs, No. 27, June 1994.

Koren St., Stiassny A., Tax and Spend or Spend and Tax ? An International Study, No. 28, August 1994.

Gstach D., Data Envelopment Analysis in a Stochastic Setting: The right answer form the wrong model?, No. 29, August 1994.

Cantwell J., Bellak Ch., Measuring the Importance of International Production: The Re-Estimation of Foreign Direct Investment at Current Values, No. 30, January 1995.

Klausinger H., Pigou's Macroeconomics of Unemployment (1933). A Simple Model, No. 31, February 1995.

