Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Reihe Ökonomie / Economics Series No. 18

Forecasting Austrian IPOs: An Application of Linear and Neural Network Error-Correction Models

Christian Haefke and Christian Helmenstein

Forecasting Austrian IPOs: An Application of Linear and Neural Network Error-Correction Models

Christian Haefke and Christian Helmenstein

Reihe Ökonomie / Economics Series No. 18

December 1995

Institut für Höhere Studien
Stumpergasse 56, A-1060 Wien
Fax: +43/1/599 91-163

Christian Häfke
Phone: +43/1/599 91-150
e-mail: chris@ihssv.wsr.ac.at
WWW: http://www.wsr.ac.at:80/ihs-html/economics/ipox/chris/

Christian Helmenstein
Phone: +43/1/599 91-152
e-mail: helmen@ihssv.wsr.ac.at
WWW: http://www.wsr.ac.at:80/ihs-html/economics/ipox/helmen/

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Abstract

In this paper we apply cointegration and Granger-causality analyses to construct linear and neural network error-correction models for an Austrian Initial Public Offerings IndeX ($IPOX_{ATX}$). We use the significant relationship between the $IPOX_{ATX}$ and the Austrian Stock Market Index *ATX* to forecast the $IPOX_{ATX}$. For prediction purposes we apply augmented feedforward neural networks whose architecture is determined by Sequential Network Construction with the Schwartz Information Criterion as an estimator for the prediction risk. Trading based on the forecasts yields results superior to Buy and Hold or Moving Average trading strategies in terms of mean-variance considerations.

Zusammenfassung

In dieser Arbeit verwenden wir Kointegrations- und Granger-Kausalitätsanalysen, um lineare und konnexionistische Fehlerkorrekturmodelle für den Österreichischen Erstemissionsindex ($IPOX_{ATX}$) zu spezifizieren. Wir verwenden die signifikanten Abhängigkeiten zwischen $IPOX_{ATX}$ und dem repräsentativen *ATX* (Austrian Traded Index), um den $IPOX_{ATX}$ zu prognostizieren. Wir ermitteln die optimale Architektur des feedforward Netzwerkes durch sequentielle Netzwerkkonstruktion wobei das Informationskriterium von Schwartz als Schätzer für das Prognoserisiko herangezogen wird. Eine Handelsstrategie auf Basis der Prognosen wirft unter Risiko-Varianz Gesichtspunkten höhere Renditen ab als eine Buy and Hold oder eine auf gleitenden Durchschnitten basierende Strategie.

Keywords

Initial public offerings, neural networks, stock market index, cointegration analysis.

JEL-Classifications

C53, C45, C43, G12

Remark

Previous versions of this paper were presented at:

IFAC/SEDC international conference "Computing in Economics and Finance", "Session on Neural Networks and Machine Learning in Economics and Finance", Amsterdam, The Netherlands, June 8-10, 1994.

"11th European Conference on Artificial Intelligence", "Workshop on Artificial Intelligence in Finance and Business", Amsterdam, The Netherlands, August 8-12, 1994.

"6th Workshop of the Austrian Working Group on Banking and Finance", Vienna University, November 15, 1994

"2nd International Workshop on Neural Networks in the Capital Markets", NNCM '94, California Institute of Technology and London Business School, Pasadena (CA.), November 17-18, 1994.

"21st Annual Convention of the Eastern Economic Association, Symposium on Non-Linear Dynamics and Econometrics", New York, March 17-19, 1995.

"IEEE/IAFE 1995 Conference on Computational Intelligence for Financial Engineering (CIFEr)", New York City, April 9-11, 1995.

1 Introduction

Four anomalies in the pricing of Initial Public Offerings (IPOs) have so far been established in the literature. These are the underpricing phenomenon, the "hot issue" market notion, the superior performance of governmental IPOs to non-governmental IPOs, and the long-run underperformance of IPOs.

The anomaly of long-run underperformance was revealed in Ritter (1991), who demonstrated that during the 1975-1984 period IPOs substantially underperformed a sample of matching firms from the closing price on the first day of public trading to the market price on their three-year anniversaries. The approach suggested in the present paper differs from earlier work with respect to both methodology and focus of attention. Ritter uses time-series data on *individual* stocks to investigate the long-run performance of IPOs. Contrary to his seminal paper, we construct an Initial Public Offerings IndeX ($IPOX$) to document the *aggregate* price dynamics of the market segment of Austrian IPOs in comparison with the blue chip market segment as measured by the Austrian Traded Index (ATX). The main subject of this paper consists of an inquiry into the statistical properties of these two market segments and their mutual relationship. In particular, we investigate the time series properties of the $IPOX$ in terms of its (auto)correlation, volatility, and causality patterns. Based on the findings of this analysis, we forecast the returns of Austrian IPOs using both linear and multilayer feedforward neural network error-correction models. As suggested by Leitch and Tanner (1991), we examine the profitability of various trading strategies in order to assess the economic value of our forecasts.

Before we turn to the construction of the $IPOX$, we briefly review the other three anomalies. In contrast to the long-run aftermarket performance of IPOs, short-run phenomena in the IPO market segment, such as the underpricing of initial public offerings, are well documented in the literature, including Ruud (1993), Carter and Manaster (1990), and Tiniç (1988). Building on the premise that any underpricing of IPOs can be expected to be eliminated after the first day of public trading due to an (almost) instantaneous price adjustment, Ibbotson, Sindelar and Ritter (1988) investigate a sample of 8,668 IPOs going public between 1960 and 1987.[1] These authors find an average return of 16.4 % from the offering price to the market price at the end of the first day of trading. The "hot issue market" phenomenon refers to the observation that during certain periods particular stock issues rise from their offering prices to higher than average premia in the aftermarket. In particular, statistical tests indicate serial correlation during the early months of aftermarket trading (Ibbotson, Sindelar and Ritter, 1988, Ibbotson and Jaffe, 1975). In an examination related to the market for Austrian IPOs, Helmenstein (1995) documents the excess perform-

[1]This argument is valid *a fortiori* in view of the observation that the shares of initial public offerings are usually subject to interbank trading (with prices available on request) before the first official listing on a stock exchange. Hence there is no necessity for investors to wait at least one trading session to gain an orientation about the market value of the IPO.

ance of governmental IPOs versus non-governmental IPOs. Drawing upon a set of $IPOX$-subindexes, the analysis reveals that the cumulative return of governmental IPOs exceeds that of the blue chip market segment by 65.8 % during a period of about two and a half years ending in May 1995. The opportunity costs of holding a portfolio of non-governmental IPOs, by contrast, cumulate to foregone earnings of 26.5 % relative to the ATX portfolio.

The paper is organized as follows. The next section is dedicated to the construction of the $IPOX$ and the econometric analysis of the data. In section 3 we present the linear $IPOX$ and ATX models and in section 4 we introduce the neural network model. Section 5 provides a discussion of the error measures which are used for out-of-sample evaluation. A final section concludes the paper.

2 Data

We now introduce the concept of an Initial Public Offerings IndeX which is essential to the empirical analysis that follows.[2]

The $IPOX$ covers all IPOs in the official market segment of the Vienna Stock Exchange. Newly issued stock of companies whose stock other than the new category has been listed earlier is also included in the sample of index stocks. IPOs in the regulated and in the unregulated market segments are, by contrast, excluded from consideration.[3] In order to prevent the underpricing phenomenon from distorting the aftermarket performance analysis, each IPO enters the $IPOX$ with the first price in public trading and *not* with the offering price. As an aggregate measure to represent the blue chip market segment of the Vienna Stock Exchange we employ the ATX, which comprises about 70 % of total market capitalization. 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,

$$IPOX_t = IPOX_{t-1}\left[\frac{\sum_{i=1}^{I} P_{i,t}Q_{i,t-1}}{\sum_{i=1}^{I} P_{i,t-1}Q_{i,t-1}}\right], \quad ATX_t = ATX_{t-1}\left[\frac{\sum_{j=1}^{J} P_{j,t}Q_{j,t-1}}{\sum_{j=1}^{J} P_{j,t-1}Q_{j,t-1}}\right], \tag{1}$$

with $IPOX_t$ (ATX_t) as the $IPOX$ (ATX) value at day t, $P_{i,t}$ ($P_{j,t}$) the price of share i (j) at day t, $Q_{i,t-1}$ ($Q_{j,t-1}$) the number of shares of stock i (j) issued at day $t-1$, and I (J) the number of stocks in the $IPOX$ (ATX). Both the ATX and the $IPOX_{ATX}$ correspond to a capitalization-weighted stock price index. A stock market index should not be influenced by stock price changes

[2] For the composition of $IPOX_{ATX}$ see appendix A.

[3] *Bank Austria AG* and *Investitionskredit AG* have been removed from the index sample for the sake of analytical clarity. Due to numerous affiliates, the stock of *Bank Austria AG* represents a particular portfolio of Austrian companies itself, whereas for institutional reasons the risk structure of *Investitionskredit AG* shares is unlikely to be reproducible for any other Austrian company.

which are solely due to technical measures, e.g. the addition (deletion) of a stock to (from) the index sample, and rights issues. In order to compensate for the impact of these measures on the index, both indexes are adjusted using identical procedures.

A distinctive feature of IPOs are projections of future corporate earnings in the prospectus. As the underwriting bank(s) can, besides others, be held liable for wrong or misleading statements, the prospectus contains more comprehensive and reliable information than any other information source available to the outside investor. Considering the additional information as being a typical attribute of a stock 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 18 months on average. For this reason, one and a half years after the first listing on the stock exchange, a stock no longer qualifies as an IPO and is thus withdrawn from the index. Table 1 shows that the legal framework affects the statistical properties of the $IPOX_{ATX}$ returns[4] in terms of lower volatility relative to the ATX since the discounted value of future profits is less uncertain.

Table 1: Descriptive statistics of $IPOX_{ATX}$ and ATX returns

	$\Delta IPOX_{ATX}$	ΔATX
Sample mean	0.00042	0.00065
Standard deviation s	0.0072	0.0098
Standard error of sample mean $\frac{s}{\sqrt{n}}$	0.00032	0.00044
# of observations n	500	500

The data is of length 610, starts on December 15^{th}, 1992, and ends on May 31^{st}, 1995. For estimating the parameters of our models we draw upon the first 500 observations ending on December 19^{th}, 1994. The remaining 110 observations are used to calculate the out-of-sample error measures.

2.1 Autocorrelation and Cross-correlation Properties

We now turn to the time series properties of the indexes. Our analysis (table 2) reveals significant positive sample autocorrelation of order 1 for both the ATX and the $IPOX_{ATX}$ returns. Furthermore, for the ATX we also find significant negative autocorrelation for lag 3.

As ATX covers the most liquid shares traded on the Vienna Stock Exchange, it can be expected to reflect price changes due to new information most quickly. It may therefore qualify as an explaining variable for $IPOX_{ATX}$. Support for this hypothesis comes from the computation of cross-correlations between ATX

[4]Returns are defined as log changes.

Figure 1: $IPOX_{ATX}$ vs. ATX

Table 2: Sample autocorrelation for observations 1 to 500

Variable	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
$\Delta\ IPOX_{ATX}$	0.239**	0.075	0.034	0.071	0.014	-0.029	0.006
$\Delta\ ATX$	0.226**	0.027	-0.097*	-0.013	0.044	0.014	-0.041

* Statistically significantly different from zero at the 0.05 significance level, here: $1.96/\sqrt{500} = 0.0877$.

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

and $IPOX_{ATX}$. Table 3 displays a statistically significant cross-correlation

Table 3: Cross-correlations between the $IPOX_t$ and ATX_{t-lag} returns

Lags	lag=-4	lag=-3	lag=-2	lag=-1	lag=0
X - correlation	0.039	-0.049	0.049	0.070	0.492**
lags	lag=1	lag=2	lag=3	lag=4	lag=5
X - correlation	0.235**	0.087	0.020	0.069	0.069

* Statistically significantly different from zero at the 0.05 significance level, here: $1.96/\sqrt{500} = 0.0877$.

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

between the current $IPOX_{ATX}$ return and the previous ATX return. These findings should subsequently be confirmed by an econometric model.

2.2 Integration and Cointegration Properties

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 properties of the ATX and the $IPOX_{ATX}$ series. If a series contains a stochastic trend, it is said to be integrated of order d, $I(d)$. Differencing d times then yields a stationary series.

Table 4 reports the results of Dickey-Fuller tests (DF) (Dickey and Fuller, 1979), Augmented Dickey-Fuller tests (ADF), and Phillips-Perron tests (PP) (Phillips and Perron, 1988) that the ATX and the $IPOX_{ATX}$ series might have up to two unit roots. In no case is there significant evidence against the single unit root hypothesis. Thus the null hypothesis that both series are nonstationary in levels cannot be rejected. All test statistics for a second unit root, that is a unit root in the first differences of the series, are highly significant. We therefore adopt the alternative hypothesis that the series are stationary in first differences.[5]

Since both series contain 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 ATX and $IPOX_{ATX}$. 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

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

Table 4: Tests for integration

	Single Unit Root			Second Unit Root		
Series	DF	ADF	PP	DF	ADF	PP
$IPOX_{ATX}$	-1.25	-1.41	-1.26	-17.53**	-13.46**	-17.56**
ATX	-1.84	-1.78	-1.81	-17.70**	-14.23**	-17.75**

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

is nonstationary. Reporting a value of -3.70, an ADF test with one lag and with $IPOX_{ATX}$ as the independent variable rejects the null of no cointegration at the 2.5% level.[6] Since the cointegrating vector establishes an equilibrium relationship, the ADF test should not lead to a different conclusion if the cointegrating equation is estimated with ATX as the independent variable. With a value of -3.76 the result confirms this requirement.

Johansen (1988) proposed an alternative method of estimating and testing cointegrating relationships. Based on maximum likelihood, the idea is to analyze the canonical correlations between levels and first differences corrected for lagged differences and deterministic components (Kunst and Neusser, 1990). Under the assumption that no deterministic trend is present in the data, Johansen's test rejects the hypothesis of no cointegration at the 5 % significance level (table 5). The null hypothesis that at most one cointegrating relationship is present in the data against the alternative hypothesis that both $IPOX_{ATX}$ and ATX are stationary can, by contrast, not be rejected. We thus set the dimension of the cointegration space to 1, that is the two variables are considered to be $CI(1,1)$, which corroborates the results obtained before.

Table 5: Johansen cointegration tests

Hypothesized number of cointegrating relationships	Likelihood ratio	5 % critical value	1 % critical value
No cointegrating relationship	21.46	19.96	24.60
One cointegrating relationship	5.70	9.24	12.97

[6] Critical values for the ADF test are -3.59 and -3.34 at the 2.5% and 5% significance levels, respectively. These 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).

3 Linear $IPOX_{ATX}$ and ATX Models

Implementing the above findings, we base the $IPOX_{ATX}$ forecasts on a dynamic specification of a linear regression model with error-correction term

$$\begin{pmatrix} \Delta IPOX_t \\ \Delta ATX_t \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} \Delta IPOX_{t-1} \\ \Delta ATX_{t-1} \end{pmatrix} + \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} (IPOX - \nu ATX)_{t-1} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \tag{2}$$

The term $(IPOX - \nu ATX)_t$ represents the stationary long-run equilibrium relationship between the series. The adjustment coefficients χ_i describe the process of adjustment of the particular series to the long-run equilibrium while the lag polynomial $\psi_{ij}(L)$ represents additional short-run dynamics. As $\psi_{ij}(L)$ contains no unit roots, it does not affect the long-run behavior of the series.

The regression results (table 6, second column) provide evidence that the current value of $IPOX_{ATX}$ is positively related to the previous values of $IPOX_{ATX}$ and ATX. The highly significant value for the error-correction term $(IPOX - \nu ATX)_t$ with a lag of 1 reveals that deviations of $IPOX_{ATX}$ from ATX cause a strong pull back tendency towards ATX while the opposite does not hold. To check this finding, we use the same explaining variables as before to model ATX in first differences. Due to the insignificant values for the error-correction term

Table 6: ATX as explaining variable for $IPOX_{ATX}$

Independent variable	$\Delta IPOX_t$	ΔATX_t
Intercept	0.0003	0.0005
	(1.057)	(1.187)
$\Delta IPOX_{ATX,t-1}$	0.1609**	-0.0740
	(3.304)	(-1.079)
ΔATX_{t-1}	0.0889*	0.2537**
	(2.458)	(4.976)
$(IPOX_{ATX} - \nu ATX)_{t-1}$	-0.00005**	0.0000
where $\nu = 0.849$	(-3.873)	(0.084)
t-values in parentheses		
$\bar{R}^2$	0.098	0.048
DW	1.995	1.973
LM (Breusch-Godfrey, p-value)	0.6504	0.2926
Ljung-Box Q (36)	36.383	34.633
p-value of Q (36)	0.451	0.534

* Statistically significantly different from zero at the 0.05 significance level.
** Statistically significantly different from zero at the 0.01 significance level.

and the $IPOX_{ATX}$ term (table 6, third column), the result confirms the strong exogeneity of ATX.

In order to analyze the time-related interaction between ATX and $IPOX_{ATX}$, we perform Granger causality tests (Granger, 1969). Granger causality from ΔATX_t to $\Delta IPOX_t$ means that the conditional forecast for $\Delta IPOX_t$ can be improved by adding lagged ΔATX_t to the information set. Table 7 presents the results of three F-tests for mutual Granger causality constructed according to the sequential method of Toda and Philips (1993). The first test ($H_0 : \chi_i = 0$) examines the significance of the error-correction term while the second test ($H_0 : \psi_{ij}(L) = 0$) tests for the significance of the lag polynomial. The test for Granger causality ($H_0 : \chi_i = \psi_{ij}(L) = 0$) examines whether both the lag polynomial and the adjustment coefficient of the error-correction term are equal to zero.

Table 7: Granger causality tests (F-statistics)

Error-correction term $H_0 : \chi_i = 0$	Short-run dynamics $H_0 : \psi_{ij}(L) = 0$	Both $H_0 : \chi_i = \psi_{ij}(L) = 0$
Results for Granger-causality from ATX to $IPOX_{ATX}$		
14.77**	12.01**	12.43**
Results for Granger causality from $IPOX_{ATX}$ to ATX		
0.002	1.086	0.584

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

The tests reveal that ATX Granger causes $IPOX_{ATX}$ while the opposite does not hold. When $IPOX_{ATX}$ is the dependent variable, Granger causality is due to the high significance level of both the lag polynomial and the adjustment coefficient of the error-correction term.

While we choose an error-correction specification for the $IPOX_{ATX}$ forecast, we predict the one-day-ahead ATX by an autoregressive process of order 3, AR[3], that is

$$\Delta ATX_t = \phi_0 + \sum_{l=1}^{3} \phi_l \Delta ATX_{t-l} + \varepsilon_t \tag{3}$$

Table 8 presents the coefficient estimates.

Table 8: Estimated coefficients for the linear ATX model

Explaining variables	ΔATX
Intercept	0.0007
	(1.510)
ΔATX_{t-1}	0.2221**
	(4.979)
ΔATX_{t-2}	0.0013
	(0.028)
ΔATX_{t-3}	-0.1036*
	(-2.331)
R^2	0.054
DW	1.992
LM (Breusch Godfrey, p-value)	0.561
Ljung-Box Q (36)	32.824
p-value of Q (36)	0.620

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

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

4 Neural Network Models

In the recent literature a class of flexible nonlinear functions called artificial neural networks (ANNs) has been proposed. White (1988) uses them to test the efficient market hypothesis, Swanson and White (1995) investigate whether the forward interest rate contains information on the spot rate, Natter, Haefke, Soni and Otruba (1994) forecast industrial production, and Hutchinson, Lo and Poggio (1994) apply a special class of neural networks to option pricing. All this work is based on McCulloch and Pitts (1943), who constructed models to mimic brain functionality. Widrow and Hoff ((1960)) called the neurons that McCulloch and Pitts used, adaptive linear elements (ADALINEs) and described the output of one such ADALINE as:

$$f(\tilde{x}_t, \alpha) = G(\tilde{x}_t' \alpha), \tag{4}$$

with $\tilde{x}_t$ being an input vector x_t augmented by a constant and α representing a set of weights. For $G(u) = u$ 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 $G(u) = \frac{1}{1+\exp^{-u}}$, we obtain the binary logit model and for $G(u)$ being any normal cumulative distribution function, we get a binary probit. Hence even at the outset of neural network modelling, standard econometric models could easily be included as special cases. Later Werbos

(1974) and Rumelhart, Hinton and Williams (1986) combined the neurons into one function and called this a multilayer perceptron. This function can be represented as:

$$f\left(\tilde{x}_t, \beta, \gamma\right) = F\left(\sum_{q=1}^{Q} G\left(\tilde{x}_t'\gamma_q\right)\beta_q\right), \tag{5}$$

with $\beta = (\beta_1, \beta_2, \ldots, \beta_q)'$ and $\gamma = \left(\gamma_1', \ldots, \gamma_q'\right)'$.

In this paper we specify the functional form of the neural net as follows:

$$f\left(\tilde{x}_t, \theta\right) = \tilde{x}_t{}'\alpha + \sum_{q=1}^{Q} G\left(\tilde{x}_t'\gamma_q\right)\beta_q \tag{6}$$

where $G(u) = \frac{2}{1+\exp^{-u}} - 1$ and $\theta = (\alpha', \beta', \gamma')'$. Increasing Q is sensible, as long as $\sum_{q=1}^{Q} G\left(\tilde{x}_t{}'\gamma_q\right)\beta_q$ is capable of extracting important structure. Q is called the number of hidden units and equation 6 describes an augmented feedforward network.

In a first step we estimate a linear regression and fix α. In order to model the structure that is left in the residuals, Q is set to 1, thus introducing an additional regressor. This process is repeated until a maximum value for Q is reached which can be determined either heuristically or through formal testing. Such an iterative procedure of estimating the network has been termed Sequential Network Construction (SNC) by Moody and Utans (1994). Unlike Lee, White and Granger we do not test the signifcance of the single parameters using LM tests but take the Schwartz Information Criterion (SIC) (Sawa, 1978, Schwartz, 1978) as a measure of the out-of-sample performance of a model with a given Q.

The importance of parsimonious models becomes clear when we recall that Hornik, Stinchcombe and White (1989, 1990) showed, among others, that feedforward neural networks can approximate any Borel-measurable function arbitrarily well provided that Q is sufficiently large. The necessity for parsimony becomes even more obvious when we explicitly write down the number of elements in the parameter vector θ:[7] $W = Q(\Xi + 2) + 1$, where Ξ is the number of columns in the set of regressors $\tilde{x}$. The multiplicative relationship between Ξ and Q lets the number of weights shoot into dimensions beyond those which econometricians are used to deal with when designing their models. A regression with three explaining variables, one independent variable, and 3 hidden units results in 16 parameters as opposed to just 3 in the case of OLS.

In financial applications that involve forecasting, the main focus of interest is not the in-sample performance of any forecasting model but rather how well the model deals with previously unseen data. We denote this *prediction risk*

[7]Henceforth we will call the parameters "weights" or "connections", and W the total number of weights.

R, which can either be determined through one of the various kinds of cross-validation and bootstrapping techniques or through criteria that are based on the in-sample error. Information criteria such as the AIC (Akaike, 1974) or SIC represent an explicit formulation of the principle of parsimony. Both were derived as an approximately unbiased estimate for the Kullback-Leibler Information Criterion, which measures the minimum possible distance between a model and the true distribution. However, asymptotically the AIC selects too large a model even for AR processes, and hence we will rely on the SIC, which puts more emphasis on the parsimony of the models. Whereas the AIC assumes, for a given class of nested alternative models, that for each model the estimate of the variance is nearly true in the sense that the difference from the true variance tends to zero as the number of observations n tends to infinity, the SIC for each model is evaluated assuming that the most complex model within the class would nearly be true but the rest not necessarily so. Taking this difference in the underlying assumptions into account, Sawa (1978) arrives at the SIC

$$SIC\,(W) = n \log MSE + W \log n. \tag{7}$$

Rissanen (1987) introduced another criterion that can be used with nonlinear and ARCH models (Granger, King and White, 1995) but which again leads to the same criterion as in equation 7.

In order to select the appropriate neural network architecture, we first determine α and then increase Q until the SIC of the model under consideration rises again. The SIC-minimum model is used to generate the out-of-sample forecasts. By construction, this approach not only captures the nonlinearity in the data but also uses linear regression and hence guarantees that the ANN will in sample perform at least as well as a linear model. White (1990) showed that ANNs can be used to perform nonparametric regression, consistently estimating any unknown square integrable function.

For the estimation we apply an autoregressive ANN model to forecast ATX, where we allow up to three lags for the sake of comparability with the linear model. Subsequently we use an ANN to estimate the error correction model for $IPOX_{ATX}$, again only permitting the same inputs as in the linear approach.

5 Out-of-Sample Error Measures and Model Evaluation

The selected error measures are chosen in such a way that they focus on different dimensions of goodness of fit. Mathews and Diamantopoulos (1994) analyze the most widely used error measures and identify four factors that together provide a more comprehensive assessment of a forecast than any of them on its own. The first and most important factor in their analysis is a ratio-type accuracy measure such as the adjusted mean absolute percentage error, $AMAPE$. The second factor can be described by the mean absolute error, MAE, and the mean squared error, MSE. Factor three, as measured through the mean error, ME, accounts for the bias in a forecast, whereas factor four, R^2, constitutes a measure of fit. Hence the information it provides can be interpreted as a pattern-matching indicator rather than a pure distance metric.

In addition to the error measures suggested by Mathews and Diamantopoulos, we also calculate Theil's measure of inequality and a confusion matrix. Furthermore, we compute results for hypothetical trading and thus obtain a well defined loss function for the comparison of our models. In detail the error measures are:

- **Adjusted mean absolute percentage error**

$$AMAPE = \frac{1}{M} \sum_{m=1}^{M} \left| \frac{y_m - \hat{y}_m}{y_m + \hat{y}_m} \right|; \tag{8}$$

- **Mean absolute error**

$$MAE = \frac{1}{M} \sum_{m=1}^{M} |y_m - \hat{y}_m|; \tag{9}$$

- **Mean error**

$$ME = \frac{1}{M} \sum_{m=1}^{M} y_m - \hat{y}_m; \tag{10}$$

- **Coefficient of determination**

$$R^2 = 1 - \frac{\sum_{m=1}^{M} (y_m - \hat{y}_m)^2}{\sum_{m=1}^{M} (y_m - \bar{y})^2}; \tag{11}$$

- **Theil's coefficient of inequality**

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

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

- **Confusion matrix**
  The up and down signals of the forecasts are used to compute a confusion matrix. We find the number of correct classifications on the main diagonal and the errors off the diagonal. The columns contain the actual ups and downs while the rows contain the forecasts. As Swanson and White (1995) note, this is simply a 2x2 contingency table, and the hypothesis that a

given model is of no value in forecasting the sign of the price movement can be expressed as the hypothesis of independence between the actual and predicted directions. A binomial test is performed to check if the confusion rate – this is the sum of the off diagonal elements over the total number of elements – differs significantly from 50 %;

- **Trading scheme**
We apply a very simple and conservative trading scheme with 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 index portfolio, then we buy. If we already hold it, we do not buy again. In the case of falling prices, we sell if we hold but never go short. Returns are annualized and compared to a Buy and Hold strategy. Transaction costs are assumed to be 0.1% of each transaction which is the amount usually faced by large scale investors on the Vienna Stock Exchange.

- **Moving Average Trading Rule**
In addition, we compare our returns to the returns generated by a 2-50 MA-Trading Rule. If the short MA(2) intersects the long MA(50) from below, we receive a buy signal and keep the portfolio until the two moving averages intersect again, and vice versa.

Table 9 presents the results of the ATX forecast. We see that the nonlinear

Table 9: Out-of-sample ΔATX forecasts

Error measures	Linear model	ANN
AMAPE	2.523	2.607
MAE	0.006	0.006
ME	-0.001	-0.003
R^2	-0.269	-0.229
Theil	1.281	1.240
Confusion Matrix	$\begin{bmatrix} 23 & 27 \\ 27 & 33 \end{bmatrix}$	$\begin{bmatrix} 25 & 29 \\ 25 & 31 \end{bmatrix}$
t-values	(0.191)	(0.191)

model with 3 lags and 1 hidden unit performs as well as the AR[3] with regard to statistical criteria. By looking at the Theil measure, however, we still detect a considerable potential for improving the forecast.

We compare the findings of the linear model (table 10) to the results obtained through forecasting $IPOX_{ATX}$ with an artificial neural network with $Q = 1$ (table 11). It turns out that for the five statistical error measures, the linear $IPOX_{ATX}$ models rank better than their neural network counterparts.

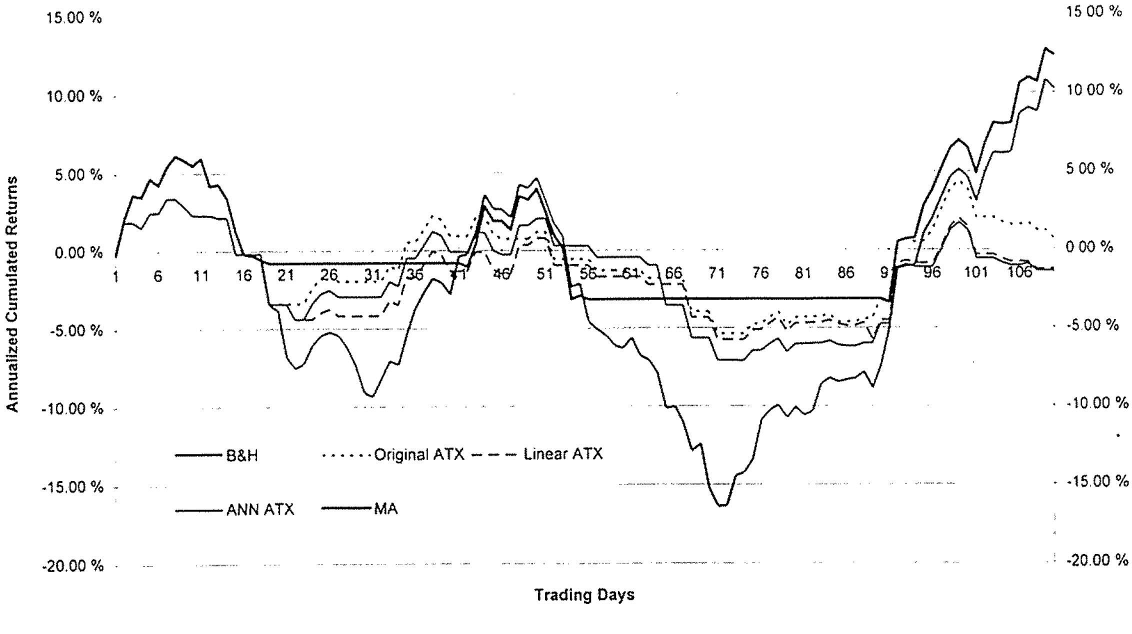
Table 10: Out-of-sample $\Delta IPOX_{ATX}$ forecasts with a linear model

Error measures	Linear ATX model	ANN ATX	Observed ATX
AMAPE	2.075	7.890	2.265
MAE	0.004	0.005	0.004
ME	-0.001	-0.001	-0.000
R^2	0.016	0.017	0.043
Theil	0.988	0.986	0.961
Confusion Matrix	$\begin{bmatrix} 32 & 24 \\ 26 & 28 \end{bmatrix}$	$\begin{bmatrix} 30 & 24 \\ 28 & 28 \end{bmatrix}$	$\begin{bmatrix} 32 & 23 \\ 26 & 29 \end{bmatrix}$
t-values	(0.957)	(0.573)	(1.151)

Table 11: Out-of-sample $\Delta IPOX_{ATX}$ forecasts with an ANN model

Error measures	Linear ATX model	ANN ATX	Observed ATX
AMAPE	1.922	5.410	2.385
MAE	0.005	0.005	0.005
ME	-0.001	-0.001	-0.002
R^2	-0.002	0.008	-0.095
Theil	1.005	0.996	1.099
Confusion Matrix	$\begin{bmatrix} 28 & 17 \\ 30 & 35 \end{bmatrix}$	$\begin{bmatrix} 26 & 17 \\ 32 & 35 \end{bmatrix}$	$\begin{bmatrix} 29 & 18 \\ 29 & 34 \end{bmatrix}$
t-values	(1.542)	(1.151)	(1.542)

Figure 2: Annualized Cumulated Returns for the Linear $IPOX_{ATX}$

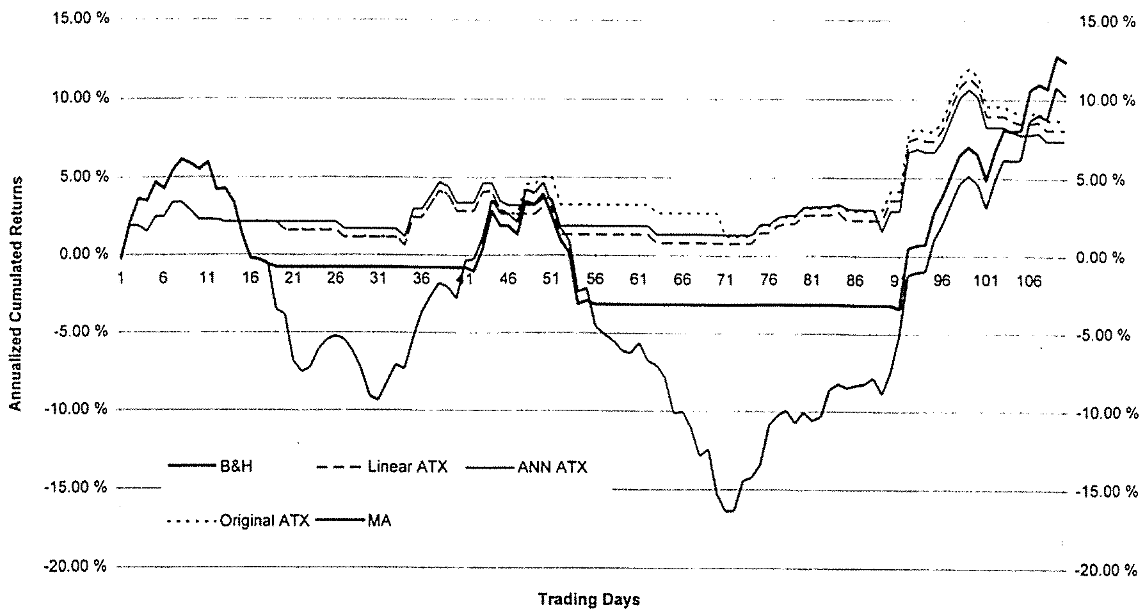


Concerning the forecasting of the ATX, we conclude from tables 9–11 that the neural net forecast does slightly better than the linear model while an inspection of the confusion matrices suggests an advantage for the linear models. However, for both the statistical measures and the confusion matrix we have to bear in mind that they need not be good indicators of the performance of a trading scheme; while the MSE does not consider the sign, the confusion matrix neglects the magnitude of the price change. A good performance in both of these measures is highly desirable for trading.

Furthermore, note that all reported $IPOX_{ATX}$ forecasts are based on true two-step ahead forecasts, and therefore our trading strategy can indeed be implemented. An analysis of the trading schemes yields the results presented in table 12. As benchmarks we use Buy and Hold and the 2–50 Moving Average. The results are very close to each other. All return series generated by the forecasting models fall short of the results for both the Buy and Hold strategy and the Moving Average strategy at the end of the 110 day period. From figure 2 and columns three and four of table 10, however, we may obtain a different picture. All linear $IPOX_{ATX}$ forecasts yield more efficient results in a risk-return setting than the Buy and Hold strategy. As compared to the 2–50 MA, the return is lower but so is the risk.

The outcome for the neural net error correction model is even more appealing

Figure 3: Annualized Cumulated Returns for the ANN $IPOX_{ATX}$



to an investor. All models μ-σ dominate both the Buy and Hold and the 2–50 Moving Average (table 12). Figure 3 plots the annualized cumulated return series generated by the ANN forecasts, which are characterized by a considerably lower draw-down than in the linear case. A comparison between the models shows an advantage for the ANN ATX forecast compared to the linear ATX as the mean return is slightly higher at a lower risk. Through trading on the basis of the forecasted time series we arrive at a well defined loss function. If we accept this approach, the neural net models outperform the linear models for this application, which is in line with previous findings for the Austrian stock market (Haefke and Helmenstein, 1996).

Table 12: Summary statistics for returns of various forecasts

Estimation method	Cumulated returns	Number of transactions	Mean return μ	Std. Dev. of returns σ
Linear $IPOX_{ATX}$, linear ATX	-0.013	58	-0.018	0.025
Linear $IPOX_{ATX}$, ANN ATX	-0.014	62	-0.016	0.029
Linear $IPOX_{ATX}$, orig. ATX	0.006	57	-0.007	0.026
ANN $IPOX_{ATX}$, linear ATX	0.081	48	0.031	0.027
ANN $IPOX_{ATX}$, ANN ATX	0.074	46	0.033	0.023
ANN $IPOX_{ATX}$, orig. ATX	0.087	52	0.038	0.028
MA 2-50	0.124	6	0.006	0.041
Buy & Hold	0.102	2	-0.027	0.067

6 Conclusions

In the case of Austria the market segment of initial public offerings provides opportunities to generate trading profits on the basis of two day ahead neural net forecasts. The cointegrating relationship between the Initial Public Offerings IndeX ($IPOX_{ATX}$) and the representative market index (ATX) can thus be successfully exploited.

A simple trading scheme yields higher mean returns than a Buy and Hold strategy irrespective of the underlying $IPOX_{ATX}$ model. The use of artificial neural networks further boosts the performance of the trading scheme in terms of mean-variance considerations. The associated cumulated returns series are less volatile and have a higher mean than profits from a 2–50 Moving Average even if transaction costs are taken into account. With regard to profitability, neural network models outperform linear models in this application. Caution has to prevail, however, since an overproportional number of illiquid stocks and thin trading are largely characteristic of the Austrian stock market. These specific phenomena, in our view, aid the predictability of the index returns. It needs to be investigated whether this conjecture holds for different market segments as well as other thin markets.

From the methodological point of view, open questions remain. The role of the SIC as a reliable guide towards the out-of-sample trading performance of a neural network model has to be studied more closely. So far no general inferences are possible.

An interesting extension of this paper would be the determination of the cointegrating relationship ν by means of a neural network which we leave for further work.

Acknowledgements

The authors gratefully acknowledge constructive comments provided by Bernhard Felderer, Michael Hauser, Robert Kunst, Terence Mills, Gerhard Rünstler, Norman Swanson, Andreas Wörgötter, and one anonymous referee. We are indebted to Franz Köstl from Austrian Kontrollbank AG for supplying the stock market data. Spreadsheet macros were excellently programmed by Silvia Aigner, and all neural network estimation was performed in GaussANN.

References

Akaike, H., "A New Look at Statistical Model Identification," *IEEE Transactions on Automated Circuits*, 1974, *19*, 716–723.

Carter, R. and S. Manaster, "Initial Public Offerings and Underwriter Reputation," *Journal of Finance*, 1990, *45*, 1045-1067.

Davidson, R. and J.G. MacKinnon, *Estimation and Inference in Econometrics*, New York and Oxford: Oxford University Press, 1993.

Dickey, D.A. and W.A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 1979, *74*, 427–431.

Engle, R.F. and C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 1987, *55*, 251–276.

Granger, C.W.J., "Investigating Causal Relations by Econometric Models and Cross- Spectral Methods," *Econometrica*, 1969, *37*, 424–438.

—— , **M. King, and H. White**, "Comments on Testing Economic Theories and the Use of Model Selection Criteria," *Journal of Econometrics*, 1995, *67*, 173–187.

Haefke, C. and C. Helmenstein, "Neural Networks in the Capital Markets: An Application to Index Forecasting," *Computational Economics*, 1996. forthcoming.

Helmenstein, C., "The Austrian Privatization Experience: The Aftermarket Performance of Governmental versus Non-Governmental Initial Public Offerings," 1995. Institute for Advanced Studies, Economics Series, 19.

Hornik, K., M. Stinchcombe, and H. White, "Multi-Layer Feedforward Networks Are Universal Approximators," *Neural Networks*, 1989, *2*, 359–366.

——, ——, and ——, "Universal Approximation of an Unknown Mapping and Its Derivatives," *Neural Networks*, 1990, *3*, 551–560.

Hutchinson, J., A. Lo, and T. Poggio, "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks," *Journal of Finance*, 1994, *49*, 851–889.

Ibbotson, R.G. and J.F. Jaffe, "'Hot Issue' Markets," *Journal of Finance*, 1975, *30*, 1027-1042.

——, **J.L. Sindelar, and J.R. Ritter**, "Initial Public Offerings," *Journal of Applied Corporate Finance*, 1988, *1*, 37–45.

Johansen, S., "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 1988, *12*, 231-254.

Kuan, C.M. and H. White, "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 1994, *13*, 1–91.

Kunst, R. and K. Neusser, "Cointegration in a Macroeconomic System," *Journal of Applied Econometrics*, 1990, *5*, 351–365.

Lee, T., H. White, and C.W.J. Granger, "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 1993, *56*, 269–290.

Leitch, G. and J.E. Tanner, "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, 1991, *81*, 580–590.

Mathews, B.P. and A. Diamantopoulos, "Towards a Taxonomy of Forecast Error Measures," *Journal of Forecasting*, 1994, *13*, 409–416.

McCulloch, W.S. and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 1943, *5*, 115–133.

Moody, J. and J. Utans, "Architecture Selection Strategies for Neural Networks: Application to Corporate Bond Rating Prediction," in "Neural Networks in the Capital Market," New York: John Wiley & Sons, 1994, pp. 277–300.

Natter, M., C. Haefke, T. Soni, and H. Otruba, "Macroeconomic Forecasting Using Neural Networks," in Y. Abu-Mostafa, ed., *Neural Networks in the Capital Markets 1994*, 1994.

Phillips, P.C.B. and P. Perron, "Testing for a Unit Root in Time Series Regression," *Biometrika*, 1988, *75*, 335–346.

Rissanen, J., "Stochastic Complexity and the MDL Principle," *Econometric Reviews*, 1987, *6*, 85–102.

Ritter, J.R., "The Long-run Performance of Initial Public Offerings," *Journal of Finance*, 1991, *46*, 3-27.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, *Learning Internal Representations by Error Propagation*, Vol. 1, Cambridge, MA: MIT Press, 1986.

Ruud, J.S., "Underwriter Price Support and the IPO Underpricing Puzzle," *Journal of Financial Economics*, 1993, *34*, 135–151.

Sawa, T., "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 1978, *46*, 1273-1291.

Schwartz, G., "Estimating the Dimension of a Model," *Annals of Statistics*, 1978, *6*, 461–464.

Swanson, N. and H. White, "A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Network Models," *Journal of Business and Economic Statistics*, 1995, *13*, 265–275.

Theil, H., *Applied Economic Forecasting*, Amsterdam: North Holland Publishing Company, 1966.

Tiniç, S.M., "Anatomy of Initial Public Offerings of Common Stocks," *Journal of Finance*, 1988, *43*, 789–822.

Toda, H.J. and P.C.B. Phillips, "Vector Autoregression and Causality," *Econometrica*, 1993, *61*, 1367–1393.

Werbos, P., "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences." PhD dissertation, Harvard University, Department of Applied Mathematics, 1974.

White, H., "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," in "Proceedings of the Second Annual IEEE Conference in Neural Networks," New York: IEEE Press, 1988, pp. 451–458.

——, "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings," *Neural Networks*, 1990, *3*, 535–549.

Widrow, B. and M.E. Hoff, "Adaptive Switching Circuits," *Institute of Radio Engineers WESCON Convention Record*, 1960, *4*, 96–104.

Appendix A

Table 13: The historical index sample of the $IPOX_{ATX}$

Company	Date of inclusion in the index	Weight at the day of the first listing
Agrana Pref.	Dec. 15, 1992	1.53 %
Austria Mikro Systeme Com.	Jul. 12, 1993	9.68 %
Bank für Tirol & Vorarlberg Pref.	Nov. 29, 1993	0.33 %
Bau Holding Com.	Oct. 18, 1993	32.78 %
Billareal Com.	Jun. 10, 1994	20.24 %
Binder Com.	Dec. 15, 1992	1.43 %
BKS Pref.	Dec. 15, 1992	0.17 %
Böhler-Uddeholm Com.	Apr. 10, 1995	11.79 %
BWT Com.	Dec. 15, 1992	4.11 %
Constantia ISO Holding Com.	May 22, 1995	3.44 %
Constantia Verpackungen Com.	May 22, 1995	3.24 %
Erste Österreichische Sparkasse Pref.	Nov. 22, 1993	17.12 %
Flughafen Wien Com.	Dec. 15, 1992	8.79 %
Kapital & Wert Com.	Dec. 15, 1992	3.90 %
Kies-Union Com. res.	Dec. 15, 1992	4.88 %
Mayr-Melnhof Com.	Apr. 22, 1994	55.64 % *
Oberbank Pref.	Dec. 15, 1992	0.48 %
ÖMAG Com.	Dec. 15, 1992 **	4.9 %
Ottakringer Com.	Nov. 10, 1994	1.49 %
Pengg Kabel Com.	Dec. 21, 1992	2.13 %
Rosenbauer Com.	Sep. 27, 1994	2.84 %
UBM Pref.	Dec. 15, 1992	0.57 %
VA Eisenbahnsysteme Com.	Dec. 15, 1992	3.82 %
VA Technologie Com.	May 25, 1994	34.85 %
Viso Data Com.	Dec. 15, 1992	5.07 %
Vogel & Noot Wärmetechnik Com.	Jul. 13, 1995	2.43 %
Wiener Städtische Pref.	Oct. 17, 1994	2.76 %
Wienerberger Immobilien Com.	Dec. 15, 1992 **	60.36 %
Wolford Com.	Feb. 14, 1995	2.26 %

* Weight on July 22, 1994: 28.62 %.
** Date of withdrawal from the sample: February 03, 1993.

Source: Institute for Advanced Studies, 1995.

Institut für Höhere Studien
Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria

Phone: +43-1-599 91-145
Fax: +43-1-599 91-163