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

Reihe Ökonomie / Economics Series No. 21

Forecasting Stock Market Averages to Enhance Profitable Trading Strategies

Christian Haefke and Christian Helmenstein

Forecasting Stock Market Averages to Enhance Profitable Trading Strategies

Christian Haefke and Christian Helmenstein

Reihe Ökonomie / Economics Series No. 21

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 design a simple trading strategy to exploit the hypothesized distinct informational content of the arithmetic and geometric mean. The rejection of cointegration between the two stock market indicators supports this conjecture. The profits generated by this cheaply replicable trading scheme cannot be expected to persist. Therefore we forecast the averages using autoregressive linear and neural network models to gain a competitive advantage relative to other investors. Refining the trading scheme using the forecasts further increases the mean return as compared to a buy and hold strategy.

Zusammenfassung

In der vorliegenden Arbeit formulieren wir ein einfaches Trading System, um den unterschiedlichen Informationsgehalt des arithmetischen und geometrischen Mittels auszuschöpfen. Kointegrationstests verwerfen die Existenz einer langfristigen Gleichgewichtsbeziehung zwischen den beiden Aktienindizes. Da dieses Handelssystem leicht replizierbar ist, kann eine Persistenz dieser Profite nicht erwartet werden. Aus diesem Grunde prognostizieren wir beide Kursdurchschnitte mit autoregressiven linearen und neuronalen Netzwerk Modellen, um gegenüber anderen Investoren einen kompetitiven Vorteil zu erlangen. Im Vergleich zu einer Buy and Hold Strategie erhöht diese Verbesserung des Handelssystems den durchschnittlichen Ertrag.

Keywords

Trading strategy, stock market index, neural networks, cointegration

JEL-Classifications

G14, C43, C45, C53

Remark

A previous version of this paper was presented at:

"3rd International Workshop on Neural Networks in the Capital Markets", London Business School, October 12-13, 1995.

1 Introduction

One of the most prominent mysteries of present day finance is the ample usage of such simple and dated concepts as the arithmetic and the geometric means as proxies for the aggregate price dynamics of leading international stock markets. While such undertakings may find their explanation, though not justification, in the inertia of the finance community to adopt more modern index concepts, it is even more astounding that during the last decade of the twentieth century some *newly* implemented stock market indexes are still constructed in the tradition of these principles.

It is known from theoretical analyses (Helmenstein and Haefke, 1995) that the arithmetic mean differs from the geometric mean in reflecting absolute and relative price changes of the index stocks. The two indexes may therefore offer distinct information to the investor. Building on this premise, we investigate whether the investor may profitably exploit trading signals which are solely due to different index construction principles whereas the underlying sample of index stocks is identical. If so, the choice of an index construction principle is by no means an insensitive issue, and our results have a substantial bearing for the validity of the efficient market hypothesis even in its weak form. According to a common criticism regarding the persistence of excess returns, it is a cheap and easy task to find promising trading rules and to exploit the buy and sell signals. Thus, it seems reasonable to expect that the profits will not be sustained and market efficiency in its weak form will be restored.

The set of alternatives how to exploit the information contained in the relationship between the two averages is richer, however. The efficienct market hypothesis, which has been the leading paradigm for at least two decades, finds itself on ever shakier grounds as the development of nonlinear forecasting techniques proceeds. Since White's (1988) paper numerous empirical economists have tried to find counterexamples to the efficient market hypothesis. The NNCM workshop series (Refenes, 1993, Abu-Mostafa, 1994, Refenes, 1995), and the CIFEr conference (1995) provide a plethora of papers which in one way or another refute the efficient market hypothesis. Using a neural network forecast of the arithmetic and the geometric average, in the present paper we demonstrate that simulated trading of the underlying stocks yields higher cumulated returns over the out-of-sample evaluation period than a simple buy-and-hold strategy.

The paper is organized as follows. Section 2 exposes the theoretical properties of the arithmetic and the geometric means. In section 3 we investigate the time trend properties of the averages and construct the models used for forecasting. Section 4 presents the results, and section 5 contains concluding remarks.

2 Properties of the Arithmetic and the Geometric Mean

Subsequently we discuss the question whether different index construction principles for the same set of underlying assets provide different information.

Table 1: Properties of the arithmetic and the geometric average

$ARI = \frac{1}{I}\sum_i^I p_i$	$GEO = c\sqrt[I]{\prod_i^I p_i}$
$\frac{\partial ARI}{\partial p_1} = \frac{1}{I}$	$\frac{\partial GEO}{\partial p_1} = \frac{1}{I}c\frac{\sqrt[I]{\prod_i^I p_i}}{p_i}$
$\eta_{ARI} = \frac{p_i}{\sum_i^I p_i}$	$\eta_{GEO} = \frac{1}{I}.$

p_i – price of stock i, I – number of stocks in the index sample, η – price elasticity of the respective average and c – constant scaling factor in order to obtain equal starting values for ARI and GEO, $c \geq 1$.

Table 1 demonstrates that in the case of the *arithmetic average* (ARI) a given absolute change by ϵ in the price of a low-priced stock has the same effect on the index value as a change by ϵ in the price of a high-priced stock. By contrast, a given relative change by 1 % in the price of a high-priced stock entails a larger percentage change of the index value than a change by 1 % in the price of a low-priced stock.

In the case of the *geometric average* (GEO) relative price changes of individual stocks have the same influence on the index value regardless of the absolute level of the respective stock price. A 1% stock price change of any stock results in a $1/I\%$ change of the index value. Contrary to before, a given absolute change in the price of a low-priced stock has an over-proportional effect on the index value whereas an identical absolute change in the price of a high-priced stock has an under-proportional effect.

In preparation of the trading scheme we investigate which condition has to be fulfilled for the slope of the geometric average to exceed the slope of the arithmetic average,

$$\frac{1}{I} \leq \frac{1}{I}\frac{\left(\prod_i^I p_i\right)^{\frac{1}{I}}}{p_i}c \tag{1}$$

which implies

$$p_i \leq c\sqrt[I]{\prod_i^I p_i}. \tag{2}$$

The p_i has to lie between the geometric and the arithmetic average for the equality to hold. As for the elasticities, we find that p_i has to be less than the arithmetic mean if the price elasticity of the geometric average (η_{GEO}) is to be greater than that of the arithmetic mean (η_{ARI}).

How can this knowledge be exploited? It might be rewarding to develop a trading strategy based on the above properties of the two indexes. We formulate the *conservative* rule as follows. Do not invest at all in a downward trending market. If the market is bullish and the slope of the geometric average intersects the slope of the arithmetic average from below, we know that the price of the low-priced stocks is growing faster than the price of the high-priced stocks. Hence buy low-priced stocks. If the slope of the arithmetic mean intersects the slope of the geometric mean from below, buy the high-priced stocks and sell the low-priced ones. The *aggressive* rule resembles the conservative rule for positively sloping averages. In a bearish market we go short in low-priced stocks if the slope of the arithmetic average intersects the slope of the geometric average from below while we take a short position in high-priced stocks if the slope of the geometric average intersects the slope of the arithmetic average from below. Low-priced stocks are defined as stocks with a price lying below the geometric mean while high-priced stocks are those whose price is greater than the arithmetic average. By assumption, we buy an equal number of shares if we receive a buy signal.

3 Data and Models

A stock market index should not be influenced by stock price changes which are 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 such measures, *ARI* and *GEO* are adjusted using identical procedures. The stocks of the representative Austrian stock market index (*ATX*) constitute the index sample of both *ARI* and *GEO*. The number of stocks in the index increases from 16 to 19 during the period under consideration.

The data is of length 487, starts on November 2^{nd}, 1992, and ends on October 14^{th}, 1994. For estimating the parameters of our models we draw upon the first 387 observations. The remaining 100 observations are used to calculate the out-of-sample error measures and the out-of-sample trading profits.

After taking logarithms of *ARI* and *GEO*, we compute a set of descriptive statistics for their first differences (table 2). The sample autocorrelations as shown in table 3 are taken as a guideline towards the specification of the linear and the neural network model.

When forecasting two stock market indexes that are based on the same underlying assets, we may expect a certain comovement between them. In order to account for this possibility, we perform a cointegration analysis.

Table 2: Descriptive statistics of GEO and ARI returns

	ΔARI	ΔGEO
Sample mean	0.00048	0.00060
Standard deviation s	0.0094	0.0092
Standard error of sample mean $\frac{s}{\sqrt{n}}$	0.00048	0.00047
# of observations n	387	387

Table 3: Autocorrelations for observations 1 to 387

Variable	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
Δ ARI	0.234**	0.039	-0.064	0.063	0.138*	0.054	-0.025
Δ GEO	0.252**	0.060	-0.057	0.058	0.109*	0.049	-0.011

* Statistically significantly different from zero at the 0.05 significance level, here: $1.96/\sqrt{387} = 0.0996$.

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

3.1 Integration and Cointegration Properties

The usual asymptotic properties in time series analysis 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 ARI and GEO. 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 ARI and GEO 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 not stationary 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.[1]

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 the arithmetic and the geometric average. Two variables are said to be cointegrated of order one, $CI(1,1)$, if they are individually $I(1)$ and yet some linear

[1] 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
ARI	-0.72	-0.85	-0.79	-15.41**	-12.24**	-15.46**
GEO	-0.54	-0.70	-0.61	-15.11**	-11.95**	-15.16**

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

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 -1.34, an ADF test with one lag and with *GEO* as the independent variable does not reject the null of no cointegration at the 10% level.[2] 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 *ARI* as the independent variable. With a value of -1.26 the result confirms this requirement.

3.2 Linear Models

Based on the above findings we specify the linear models for both *ARI* and *GEO* as AR(1) processes. The coefficient estimates are presented in table 5.

3.3 Neural Network Model

Many different classes of neural network models are successfully applied to time series data with the simple single hidden layer network being one of them (White, 1988, Natter, Haefke, Soni and Otruba, 1994). However, it has frequently been noted that performance sometimes degrades after adding a hidden layer as compared to a simple perceptron. To avoid this shortcoming, we use an augmented single hidden layer feedforward neural network which combines a simple perceptron with a single hidden layer network. Therefore the output is calculated as follows:

$$f(\tilde{x}_t, \theta) = \tilde{x}_t'\alpha + \sum_{q=1}^{Q} G(\tilde{x}_t'\gamma_q)\beta_q \tag{3}$$

[2] Critical values for the ADF test are -3.34 and -3.04 at the 5% and 10% 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).

Table 5: Linear models

Independent variable	ΔGEO_t	ΔARI_t
Intercept	0.00042	0.00033
	(0.924)	(0.714)
ΔGEO_{t-1}	0.252**	–
	(5.100)	–
ΔARI_{t-1}	–	0.235**
	–	(4.734)
R^2	0.061	0.053
DW	1.979	1.984
Ljung-Box Q (36)	37.97	41.682
p-value of Q (36)	0.380	0.237

with $\tilde{x}_t$ denoting the input vector x_t augmented by a constant and θ representing a weight vector containing the weights α, β, γ, that is $\theta = (\alpha', \beta', \gamma')'$, $\beta = (\beta_1, \beta_2, \ldots, \beta_Q)'$, $\gamma = \left(\gamma_1', \ldots, \gamma_Q'\right)'$. Q is the number of hidden units and G is a nonlinear function, in this case $G(x) = \frac{2}{1+\exp^{-x}} - 1$. This architecture not only captures the nonlinearity in the data but also incorporates the well known linear regression approach and therefore ensures that the network will in sample perform at least as good as a linear model. If the input-output connections were dropped, this outcome could not be guaranteed.

Currently, the dominant approach in estimating neural networks involves early stopping or some variation of it. This means that it is not intended to approximate the unknown parameter vector θ as closely as possible but to some predefined level of accuracy. It is argued that longer training would result in fitting the noise. In our paper no early stopping is applied. We minimize the complexity of the network to avoid overfitting. Estimation takes place in two steps. First, the direct input-output connections α are estimated through OLS and fixed. In a second step the matrices β and γ are estimated to model the residuals of the linear regression. This approach generally improves the performance over OLS. We solve for

$$\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \left(y_t - f\left(\tilde{x}_t, \theta\right)\right)^2 \tag{4}$$

with α fixed. The programme used to estimate the feedforward networks is designed to find the optimal number of hidden units using the Schwartz information criterion (SIC) (Sawa, 1978, Schwartz, 1978). A number of networks are estimated, starting off with zero hidden units. Then a hidden unit is added and the weights are reestimated. This approach has been called *Sequential Network Construction* by Moody and Utans (1994). The in-sample errors generated from these nets are then used to determine SIC, which adds a penalty term to the number of parameters. SIC is calculated according to:

$$SIC = \ln MSE + \frac{w}{T} \ln T \tag{5}$$

where w denotes the number of parameters and T the number of available observations. Applying this procedure we receive an estimate for the out-of-sample performance which can be applied to linear as well as nonlinear and ARCH models (Granger, King and White, 1995).

4 Error Measures and Empirical Results

For the estimation we apply an autoregressive feedforward neural network model to forecast ARI and GEO, where we allow up to five lags and three hidden units. The quality of our results is evaluated using the following out-of-sample error measures:

- **Normalized mean squared error**

  $$NMSE = \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2} \tag{6}$$

  $NMSE$ was used by Weigend and Gershenfeld (1994) to evaluate entries into the Santa Fe Time Series Competition and normalizes the MSE by dividing it through the variance of the respective series;

- **Theil's coefficient of inequality**

  $$Theil = \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - y_{t-1})^2} \tag{7}$$

  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 in 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 the conservative trading scheme as described in section 2 without transaction costs. We start trading on the first day of the evaluation period;

- **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, Lakonishok and LeBaron, 1992)

$$t = \frac{\mu_t - \mu_b}{\sqrt{\frac{\sigma^2}{N_t} + \frac{\sigma^2}{N_b}}} \tag{8}$$

  with μ_t and μ_b 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_b the number of observations.

Table 6: Results of out-of-sample ΔARI forecasts

Error measures	Linear model	ANN
NMSE	0.593	0.599
Theil	0.591	0.597
Confusion matrix	$\begin{bmatrix} 26 & 22 \\ 21 & 30 \end{bmatrix}$	$\begin{bmatrix} 27 & 23 \\ 20 & 29 \end{bmatrix}$
t-values	(1.32)	(1.32)

Table 7: Results of out-of-sample ΔGEO forecasts

Error measures	Linear model	ANN
NMSE	0.571	0.362
Theil	0.568	0.361
Confusion matrix	$\begin{bmatrix} 26 & 23 \\ 22 & 28 \end{bmatrix}$	$\begin{bmatrix} 27 & 23 \\ 21 & 28 \end{bmatrix}$
t-values	(0.91)	(1.11)

Tables 6 and 7 report the results for the ARI and GEO forecasts, respectively. Whereas we find no distinct advantage of the ANN over the linear model for the ARI, the ANN significantly boosts the forecast of the GEO. The SIC-best ANN model chosen uses three lagged values of the respective series and one hidden unit in both cases. In this application — unlike in Swanson and White (1995) — the SIC could be used as a computational shortcut towards the out-of-sample performance of the neural net models. The confusion matrices of all forecasts provide additional insights into the quality of the forecasts but nowhere can we reject the hypothesis that the up/down predictions are not correct in more than 50 % of the cases.

However, the quality of the forecasts becomes clear when we base the conservative trading rule of section 2 on them. The results are reported in table 8. The buy-and-hold strategy gives both the lowest cumulated as well as the lowest mean return of all approaches under consideration. The application of the trading scheme without the help of a forecasting model wins with regard to the annualized cumulated returns. Refining this trading scheme by the use of either linear or neural net forecasting models increases the mean return as compared to the unrefined approach. Whereas the standard deviation of the OLS forecast-based return series also increases, it remains virtually unchanged for the ANN at the expense of a higher number of trading days.

Table 8: Summary statistics for annualized returns of the conservative trading scheme

Estimation method	Cumulated returns	Number of transactions	t-value (vs. buy&hold)	Mean return	Std. dev. of return
Linear model	0.176	62	9.842	0.198	0.170
ANN model	0.151	79	12.182	0.212	0.156
No forecast	0.333	48	9.080	0.157	0.145
Buy and hold	-0.938	2	n.a.	-0.073	0.476

5 Conclusion

In this paper we introduce a trading strategy based on arithmetic and geometric averages. We find empirical evidence that the additional information contained in the relationship between these two indexes can be used to outperform a buy-and-hold strategy on the stock market.

Any investor should be able to take advantage of the described trading rule. In order to gain a competitive edge relative to the other market participants, we base the trading system on linear and neural network forecasts of the underlying indexes. The neural net forecast provides a higher mean return at the same level of risk — as measured by the standard deviation of the returns — than any other approach. Recall that the models — depending on the forecast — generate between 48 and 79 trading signals in just 100 days. Hence it remains to be inquired whether the profits can be sustained in an environment where transaction costs are taken into account.

6 Acknowledgements

We thank the participants of the 3^{rd} International Workshop on Neural Networks in the Capital Markets (*NNCM95*) for constructive comments. Partial financial support by the Austrian National Bank, Grant AEPC 7, is gratefully acknowledged. All neural network estimation was performed using GaussANN.

References

Abu-Mostafa, Y., *Neural Networks in the Capital Markets 1994,* 1994.

Brock, W., J. Lakonishok, and B. LeBaron, "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *Journal of Finance*, 1992, *47*, 1731–1764.

Computational Intelligence for Financial Engineering, New York: IEEE Press, 1995.

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

Helmenstein, C. and C. Haefke, "A Comparative Analysis of Stock Market Indexes," 1995. mimeo, Institute for Advanced Studies: Vienna.

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 *Neural Networks in the Capital Markets 1994* (Abu-Mostafa, 1994).

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

Refenes, A. N., *Neural Networks in the Capital Markets* 1993.

Refenes, A.N., *Neural Networks in the Capital Markets 1995*, London: World Scientific Publishers, 1995.

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.

Weigend, A.S. and N.A. Gershenfeld, *Time Series Prediction, Forecasting the Future and Understanding the Past*, Reading, MA: Addision Wesley, 1994.

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.

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