

ISSN 0280-5316
ISRN LUTFD2/TFRT--5630--SE

System Identification for Modeling of Stock Market Index

Daniel Sandberg

Department of Automatic Control
Lund Institute of Technology
November 1999

Department of Automatic Control Lund Institute of Technology Box 118 SE-221 00 Lund Sweden	<i>Document name</i> Master Thesis	
	<i>Date of issue</i> November	
	<i>Document Number</i> ISRN LUTFD2/TFRT-5630--SE	
<i>Author(s)</i> Daniel Sandberg	<i>Supervisor</i> Rolf Johansson, LTH and Björn Hansson, LU	
	<i>Sponsoring organisation</i>	
<i>Title and subtitle</i> System Identification for Modeling of Stock Market Index. (Systemidentifering för modellering av börsindex - global variablers påverkan på den svenska börsen).		
<i>Abstract</i> <p>This MS Thesis is an attempt to evaluate how much of the movement, in a big Swedish stockindex, that can be attributed to global influences, such as other indices and different signals.</p> <p>First a number of input signals was found, based on availability, coherence and homoscedasticity. A large number of signals were evaluated and eventually seven input signals were chosen. Of these signals three is indices, two is indicators and two are more intangible signals.</p> <p>Next a number of methods were used to predict the index one step ahead. Three time-invariant methods and two time-variant methods. The methods were evaluated in three different ways and the result was that a subspace method called N4SID was the best method for this purpose.</p> <p>As a conclusion to how much of the movement is attributed to global influences a rough estimate is that at the opening of the market about 53-54% is dependent of global influences and at closing time about 70-73% is dependent of global influences.</p>		
<i>Key words</i>		
<i>Classification system and/or index terms (if any)</i>		
<i>Supplementary bibliographical information</i>		
<i>ISSN and key title</i> 0280-5316		<i>ISBN</i>
<i>Language</i> English	<i>Number of pages</i> 55	<i>Recipient's notes</i>
<i>Security classification</i>		

The report may be ordered from the Department of Automatic Control or borrowed through:
University Library 2, Box 3, SE-221 00 Lund, Sweden
Fax +46 46 222 44 22 E-mail ub2@ub2.lu.se

ABSTRACT	2
INTRODUCTION	3
EFFICIENT MARKET HYPOTHESIS	5
WEAK FORM EFFICIENCY.....	5
SEMI STRONG FORM EFFICIENCY	5
STRONG FORM EFFICIENCY	5
THIS THESIS	6
DATA DEFINITION	7
HETEROSCEDASTICITY	8
FACTOR ANALYSIS	9
OM-INDEX (OMX).....	10
XETRA DAX INDEX (DEUTSCHE BORSE AG).....	11
DOW JONES INDUSTRIAL AVERAGE	12
FTSE 100 INDEX	13
DOUBLE MOVING AVERAGE	14
SO10.....	15
U.S DOLLAR – SEK EXCHANGE RATE	16
PSYCHOLOGICAL VARIABLE	17
METHODS	18
TIME-INVARIANT METHODS	18
ARMAX.....	18
ARX.....	18
AR ORDER SELECTION.....	19
STATE-SPACE METHODS	20
N4SID	20
TIME-VARIANT MODELS	23
RESULTS	24
SIGN TEST.....	24
VAF	25
RESIDUAL TESTS	26
IMPLEMENTATION	28
CONCLUSIONS	29
FUTURE WORK	30
APPENDIX A	31
PSYK	32
ARX-ORDER.....	33
ARMAX-ORDER	33
SEGMENT1	34
ITERATIV	37
TEST.....	38
MONEY	38
MONEY3	39
MOVAVG	39
APPENDIX B	40
REFERENCE LIST	55

Introduction

To predict the future.

There was a time when an attempt to perform such a task would have been met with ridicule and laughter. But thanks to powerful mathematical tools this has become something thousands of people do in their everyday work at financial institutes and at statistical bureaus around the world.

The problem is not the actual predicting, but rather choosing input signals, time-spans and methods. And not even when you do all of this right, you can be sure to have an accurate prediction of tomorrow or next month, year or whatever time you choose.

The stockmarket is especially hard to model and predict, due to the many ways there are to influence it. Apart from the obvious variables, such as the economy of the country in which the market is situated and the surrounding markets, there is the power of mass psychology, which is not an easy thing to model.

However, in this thesis an attempt is being made to predict the Swedish stockmarket, namely the OMX-index. This is being done by choosing a number of input signals, putting them through tests which will determine if they have an influence on the OMX and thereafter model the OMX with five different models to see if there is one that is better than the other. The final step is the testing of the prediction to see how much money can be made (or lost) in a simulation that have taken into account transaction-costs.

There is a plethora of methods to make money on the study of graphs of stock prices, technical analysis, that is gradually becoming more and more accepted by the academic society. However the most part of these methods, are to say the least, doubtful as to whether or not they work as intended. Most of these methods are graphic methods where lines are being drawn and interpreted in an arbitrary way. A few of the methods used does seem to have a theoretical background, such as the Elliot Wave Theory and the Double Moving Average. The Double Moving Average has, in this thesis, been put to use as an input signal and interpreted as a measure of the index momentum.

This MS thesis can be viewed as just another attempt to develop a method for technical analysis, but the intention is more that of trying to model the influence of variables that have a global effect on the market, as opposed to the more direct influences, as the success or failure of a single company.

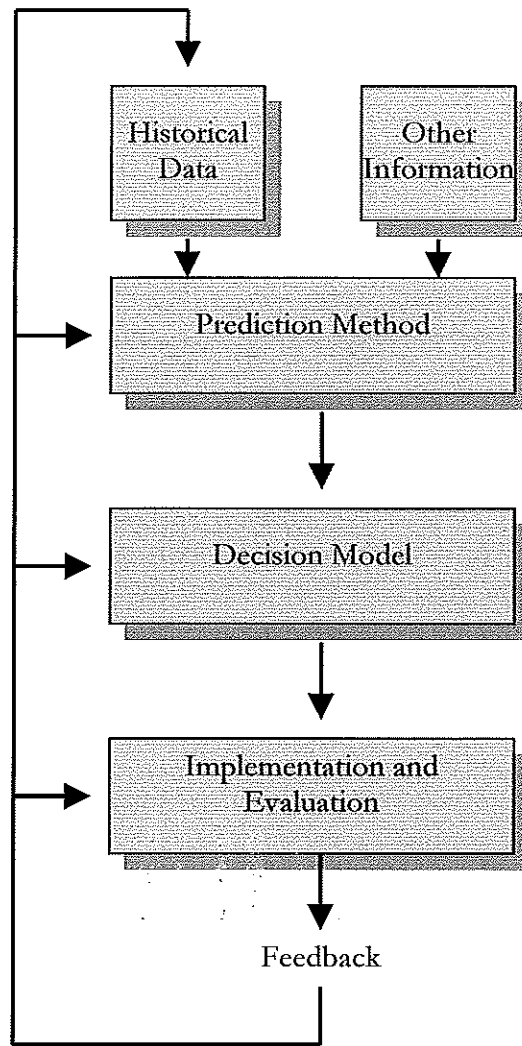


Fig. 1 General description of the prediction process

This figure is trying to show the different steps involved in the total process of using prediction to trade at the stockmarket. This thesis is primarily concerned with finding the best prediction method of those that are presented later. The decision model is a crude algorithm that doesn't take into account the magnitude of the prediction, only the sign. The transaction costs are taken into consideration, but the difference between buy- and sell-prices are disregarded since no such information is available. Also no risk-premium is subtracted from the result since the result is compared to a Buy and Hold strategy which should warrant the same risk-premium.

To be able to profit from a prediction, one has to have the prediction before 17.30 when the market closes for the day. However in this thesis one prediction will be made at 15.30, which is the opening of the N.Y. Stock Exchange, another at 22.00, which is the closing of the same and yet another one at 15.30 the same day as the predicted value. The last one is made because it is interesting what the influence of these global variables has been on the Swedish market.

Efficient Market Hypothesis

In this section I will discuss the strongest objection to making a model and trying to predict future results of the stock market, the Efficient Market Hypothesis.

The Efficient Market Hypothesis (EMH) postulates that a capital market is efficient if prices in this market "fully reflect" available information. This can be expressed with the following formula [2] and [6].

$$Z_{j,t+1} = r_{j,t+1} - E(r_{j,t+1} | \Phi_t) \quad j = 1, 2, \dots, N \text{ where } N \text{ is the total}$$

where $Z_{j,t+1}$ is the unexpected return in excess of the expected one for an asset j at the end of period $t+1$. $r_{j,t+1}$ is the actual return and $E(r_{j,t+1} | \Phi_t)$ the expected value of that return, based on the information set Φ obtained at time t . Further, $E(Z_{j,t+1}) = 0$ and $E(\rho z_{j,t}, z_{j,t+1}) = 0$, where ρ denotes correlation. If both of these requirements are met, the investor cannot earn profits at $t+1$ above those of the equilibrium return, basing his decisions on the information set Φ_t .

There are three forms of efficiency that occur at the stock markets, Weak, Semi strong and Strong form.

Weak Form Efficiency

Weak form efficiency implies that the current price of a stock fully reflects past price movements. This is the lowest form of information available on a stock, and is therefore the cheapest and easiest to obtain.

This means that if a model could be built only on past prices of a stock and used to outperform the market at a very low cost, inefficiency of the weak form could be proved.

Semi Strong Form Efficiency

Semi strong form of efficiency considers the adjustments of prices to public information. The market is considered to be efficient of the semi strong form if stock prices incorporate all available public news such that profit opportunities from the knowledge of this information is not possible.

This means that according to the EMH it should not be possible to make any predictions on future prices based on the news.

Strong Form Efficiency

If the market is efficient in the strong form, all information is reflected in the prices of stocks, including inside information. Inside information is such information known to a limited group of people only. According to the EMH, no information held by insiders could be used for economic gains. As soon as a member of these inner circles tries to act

upon such information, prices would immediately adjust to the new level dictated by the information content.

This Thesis

This thesis is dedicated to trying to build a model based on past prices information or variables based on such information. This implies that the form of EMH that can be applied in this case is the weak form. Hence should the model building succeed, inefficiency of the weak form would be proved.

Data definition

In this section, I will describe the data I have used as observed variable and input variables. Most data were provided by Update Finansiell Information. The observed variable (*OMX see below*) was collected for the period January 4 1994 to June 14 1999. It was thereafter carefully cleaned and checked with other sources. After cleaning there remained 1617 trading days which could be used.

The other data sets that are being used as input signals were thereafter imported and treated by the algorithm¹ that is described below

To solve the problem that holidays occur at different days in different countries, and causes discrepancies when there are trading in, for example, the U.S. on May 1:st, I made an algorithm that checked the time-series for such discrepancies. At days where the Stockholm Stock Exchange was open and the time-series that were being treated lacked that date, the algorithm simple copied in the value of the latest trading day in that space, but at days where Stockholm Stock Exchange were closed and the time series that were being treated had a value for that day, the algorithm removed the date and copied the value of that day onto the last open day for the time-series.

All data was transformed into percental changes per day as some kind of detrending. This was done by taking $[Index(k+1)/Index(k)]-1$.

The selection of data is based on availability and tests of coherence. Other relevant data is inflation data and predictions of the market situation

Input signals are:

Dax-Index
Dow Jones-Index
FTSE-Index
U.S.D-SEK Exchange rates
So10-interest rates
Double Moving Average
Psychological variable

These signals were arranged in the following input-output data matrix Z:

	<i>Omx</i>	<i>Dax</i>	<i>Dow</i>	<i>FTSE</i>	<i>Dollar</i>	<i>So10</i>	<i>MovAvg</i>	<i>Psyk</i>
	1	1	1	1	1	1	1	1
Z =	2	2	2	2	2	2	2	2
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>

Heteroscedasticity

An important assumption when dealing with least-squares identification models is that the error variance is constant. If this is not true the model is called heteroscedastic. When heteroscedasticity is present the least-squares estimation places more weight on observations with large error variance than on those with small error variance, causing a less efficient estimation of the model.

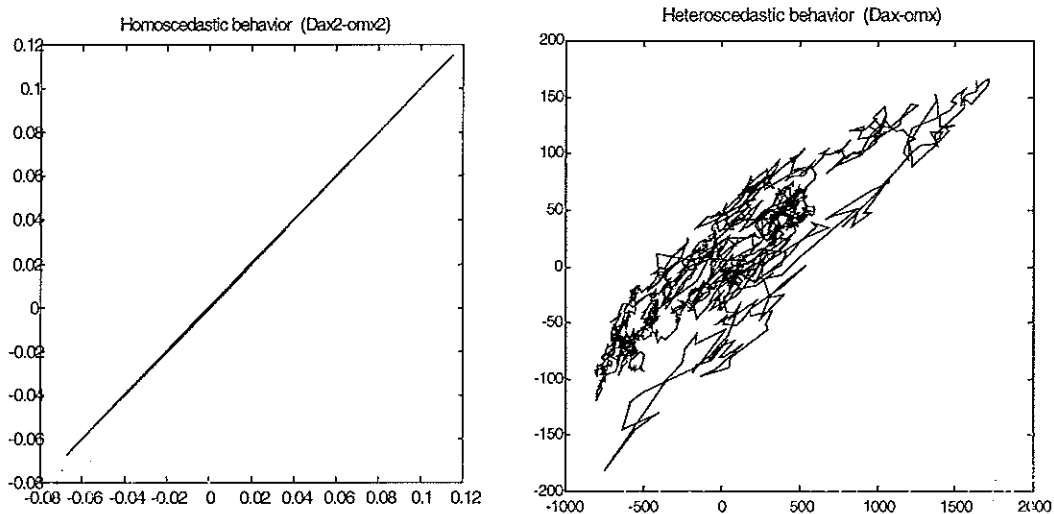


Fig. 2 Visualization of the difference between Homoscedasticity and Heteroscedasticity

From these plots it seems that the time series that hasn't been transformed into percental changes (right) exhibit a clear heteroscedastic behavior, but the transformed time series (left) does not. Since it is the transformed time series that are used for the identification a test of heteroscedasticity is performed on these. The test is called the Goldfeld-Quandt test and is described in detail in [5].

Dax:	$ESS_2/ESS_1=4.7432$	\therefore Reject Homoscedasticity
Signal	ESS_2/ESS_1	\therefore
Dax	0.1184	Homoscedastic at 95%
Dow	0.0169	Homoscedastic at 95%
Lond	0.0462	Homoscedastic at 95%
Dollar	0.1277	Homoscedastic at 95%
So10	0.4119	Homoscedastic at 95%
MovAvg	0.0342	Homoscedastic at 95%
Psyk	0.3646	Homoscedastic at 95%

The homoscedasticity hypothesis is rejected at values above approx. 1.2 in this F-test with 598 degrees of freedom in both the numerator and the denominator.

Factor analysis

In order to evaluate if all inputs to a system is independent, and therefore contributing, a factor analysis can be made.

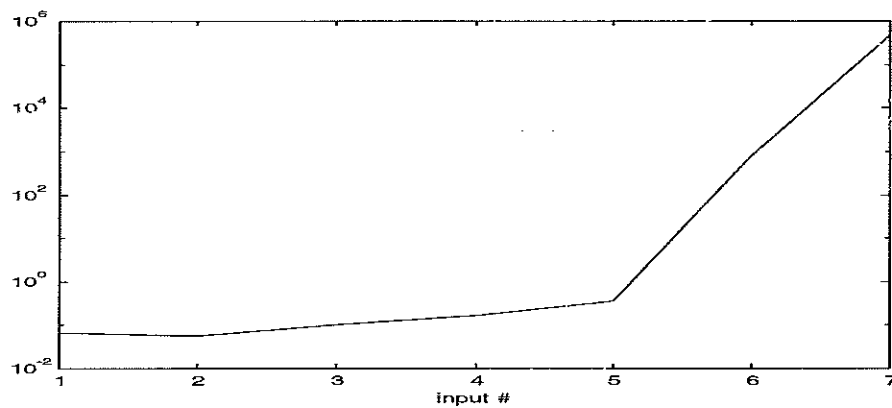
All input data is collected into a matrix:

$$U = \begin{matrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{N1} & \cdots & u_{Nn} \end{matrix}$$

After that an SVD is made:

$$SVD(U^T U)$$

If in the results of this SVD some of the singular values is significantly lower, the conclusion is made that they are not independent.



The only conclusion that can be made from this plot is that the input # 1-5 is not independent, these inputs are the different markets. This corresponds to the assumption that the different markets are not independent, but affect each other to a high degree.

OM-index (OMX)

I chose to try to predict the OM-index (OMX), rather than Generalindex (SGX), because nowadays the OMX can be traded as a whole through the OMX futures. This simplifies the trading and keeps down the magnitude of transactions that are made.

OMX [7] is a composite index that is comprised of the 30 most traded stocks on the Stockholm Stock Exchange and is calculated in the following way:

$$I_t = \frac{BV_t}{BV_{t-1} + J} \times I_{t-1}$$

t = Time at which the calculation is performed

Ka_t = The quote for Indexstock a at time t

Aa_t = Number of stocks per Indexstock a at time t

Va_t = Currencyfix for Indexstock a at time t

Bva_t = Stock Exchange Value for Indexstock a at time t, $Ka_t \times Aa_t \times Va_t$

BV_t = Stock Exchange Value for all Indexstocks at time t, $\sum BVa_t$

J = Amount for correction of basevalue at emissions, conversions and similar expressed in the Index presentation-currency.

Ja = Correction-amount for Indexstock a

I_t = Indexvalue at time t

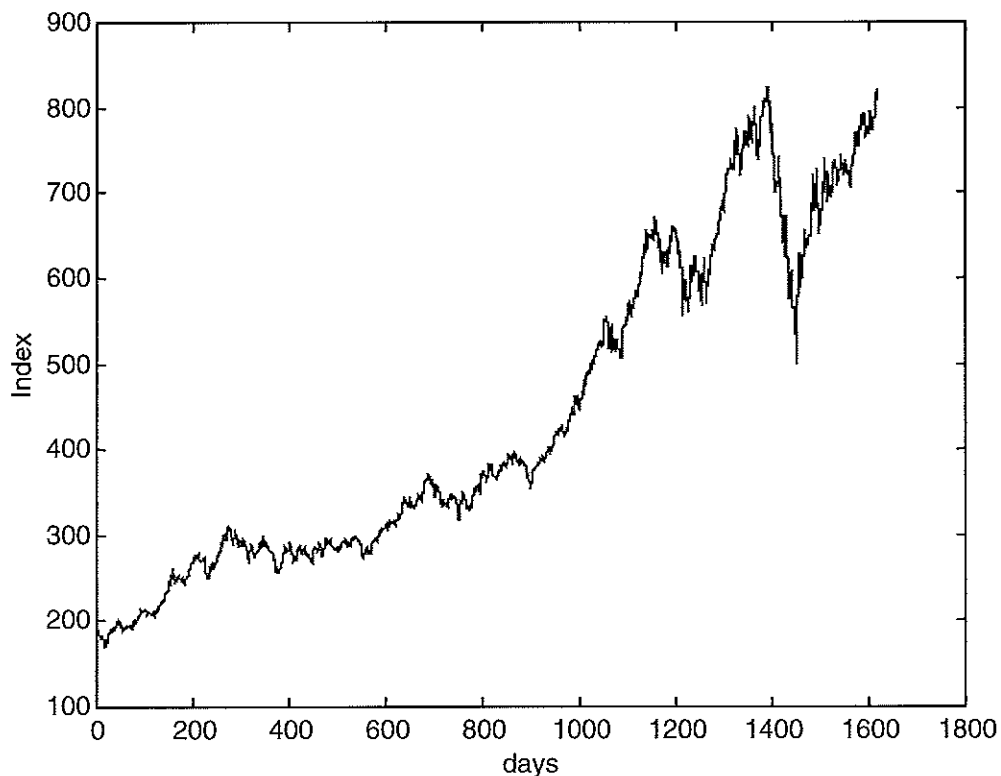


Fig. 3 OMX-index for 930104 - 990614

XETRA DAX INDEX (Deutsche Borse AG)

The German Stock Market Index (DAX) from the Frankfurt Stock Exchange. This is the counterpart of the Swedish OMX or General Index, that is, a composite index.

With the world's third largest economy, Germany produced \$2.1 trillion worth of goods and services in 1998[9]. This and the fact that Germany is a neighbor to Sweden make a big influence on Swedish economy, a fact that is clear if you look at the coherence plot between DAX and OMX. In fact, it is not only a high coherence, it has been increasing these last few years.

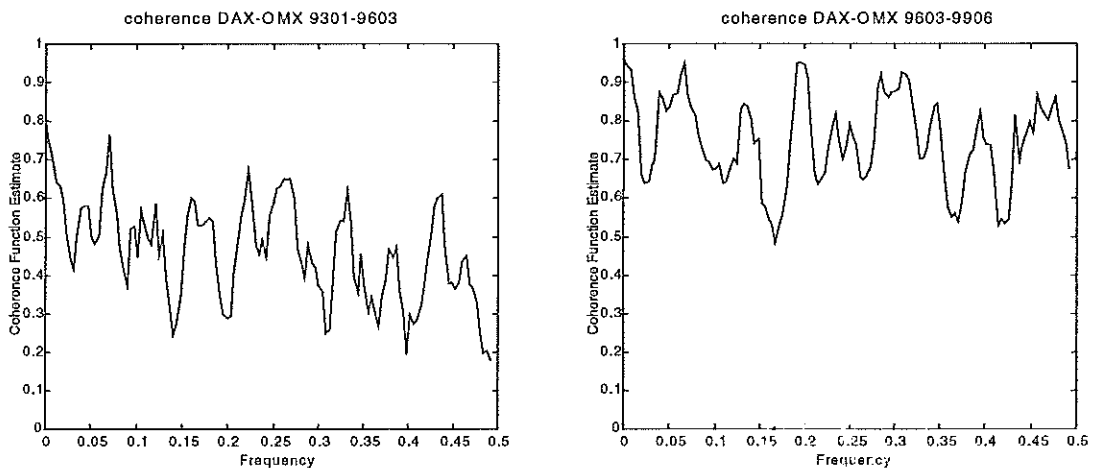


Fig. 4 Coherence plots DAX-OMX

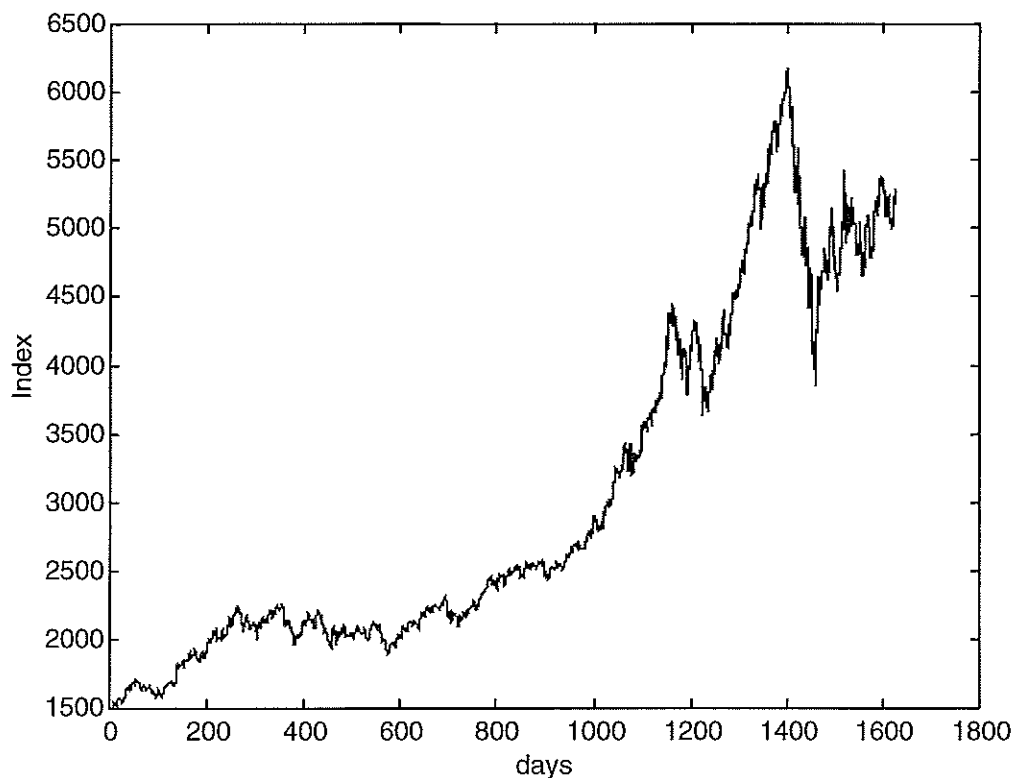


Fig. 5 DAX 930104-990614

Dow Jones Industrial Average

The Dow Jones Industrial Average (DJIA) is arguably the most widely followed and recognized stock index. The index is made up of a diversified portfolio of 30 top-quality blue chipⁱⁱ stocks. Some of the stocks that comprise the DJIA are IBM, Proctor & Gamble, Hewlett Packard, Coca-Cola, Johnson & Johnson, and other well recognized big name companies.

In later years there has been a greater emphasis on the U.S. market when it comes to influence on Swedish investors and companies. This makes the Dow Jones Index one of the major factors as to how the stock market react to changes in the global economy.

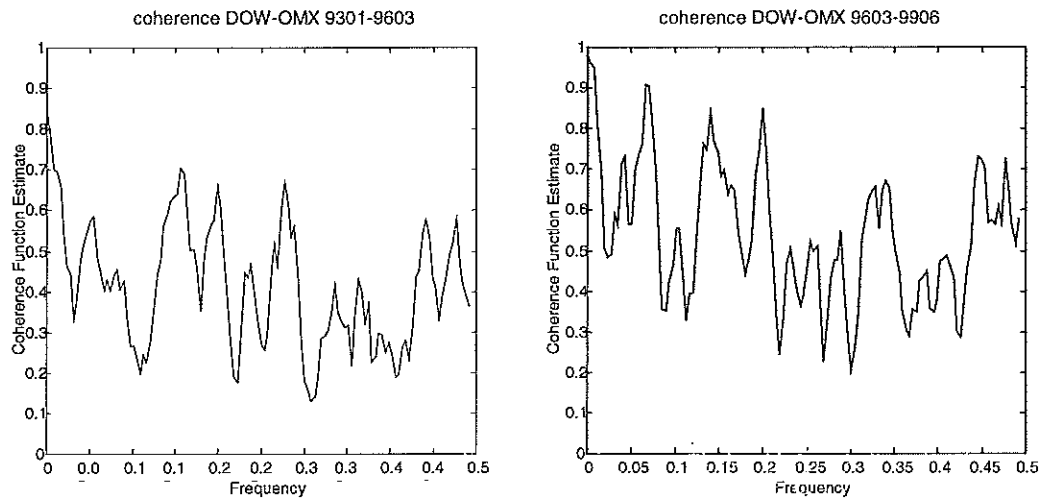


Fig. 6 Coherence plots Dow Jones-OMX

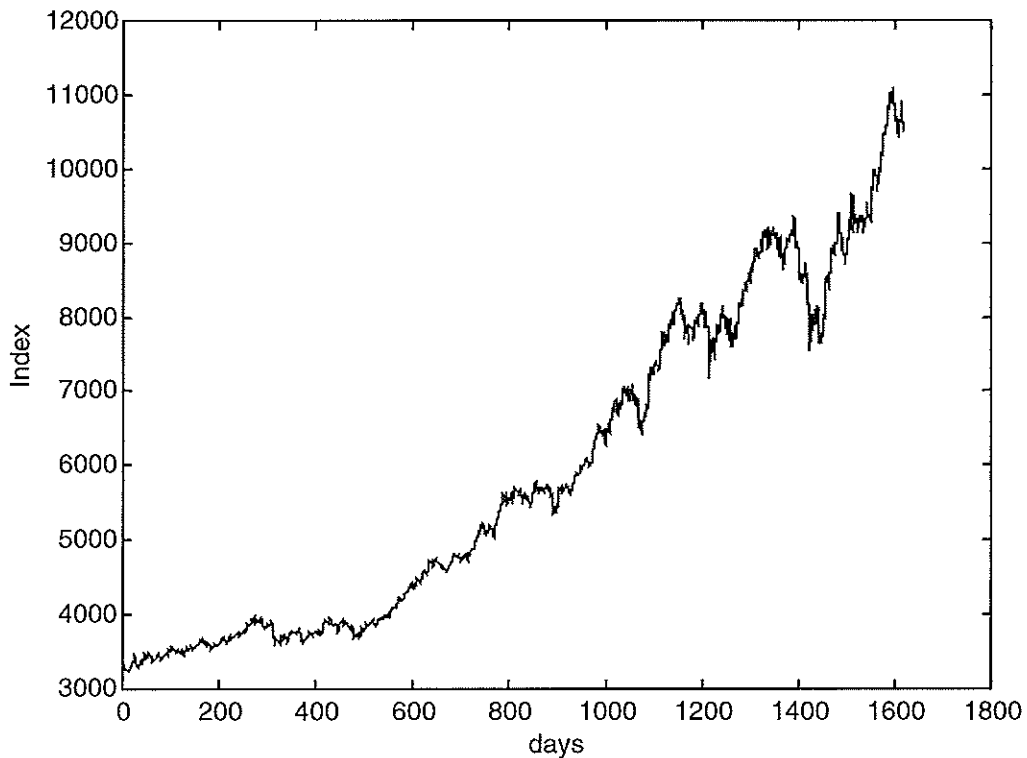


Fig. 7 Dow Jones 930104-990614

FTSE 100 INDEX

The most widely-quoted and 'popular' index for tracking the London stockmarket. The FTSE 100 contains the shares of the top 100 U.K. companies ranked by market capitalization[8]. It's jointly sponsored by the Financial Times, the London Stock Exchange, and the Institute and Faculty of Actuaries and marketed by a company called 'FT SE International'.

The 'FTSE' is a market capitalization-weighted index, re-weighted every day.

As in the case of Germany, the UK is a major economy in our vicinity and is thus bound to have a big influence on the Stockholm Stock Exchange

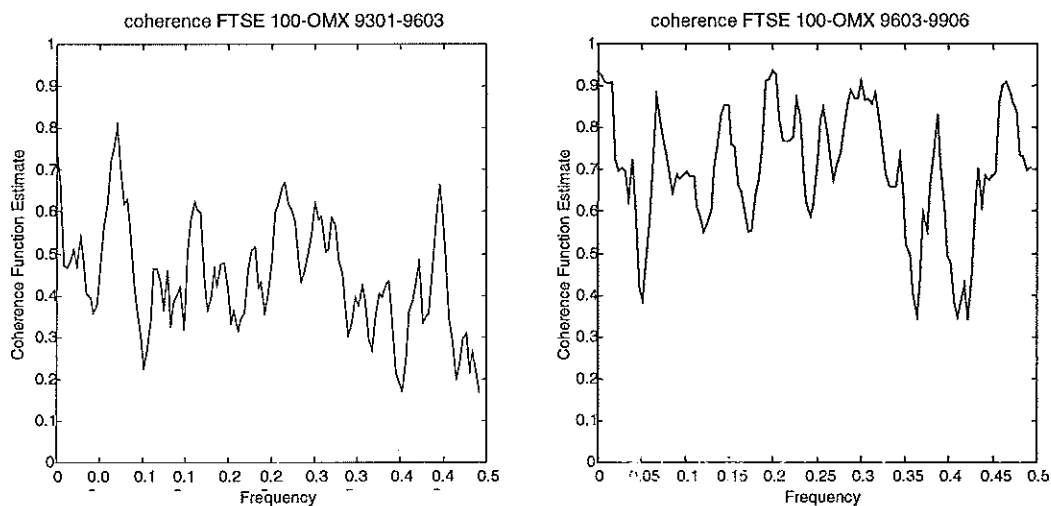


Fig. 8 Coherence plots FTSE 100-OMX

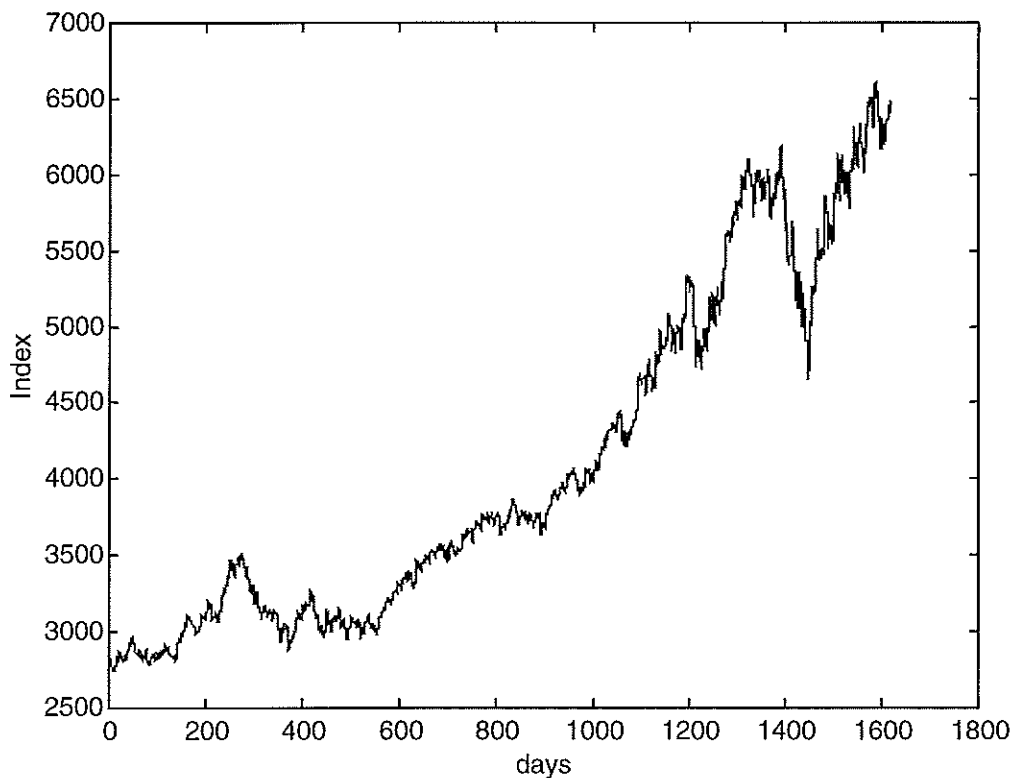


Fig. 9 FTSE 100 930104-990614

Double Moving Average

”Let the trend be your friend” is a common saying among investors. The Double Moving Average (DMA) trading technique[1] is a well-known use of this. As a brief description two moving averages are used, one calculated on a shorter number of days and one of a longer number. Whenever the shorter average crosses over the longer this technique tells the trader to go long (buy) in the stock and when the longer average crosses over the shorter tells the trader to go shortⁱⁱⁱ in the stock. A somewhat more aggressive approach can be made here since the signal is used as one of many and some of the false signals that usually would cancel some of the profits can be disregarded. Therefore I use the (1,21)DMA, where (1,21) stands for short average one day and long average 21 days.

In this paper I extend this technique to a more detailed indicator by taking the difference between the short and the long average and use it as an input signal for the identification process.

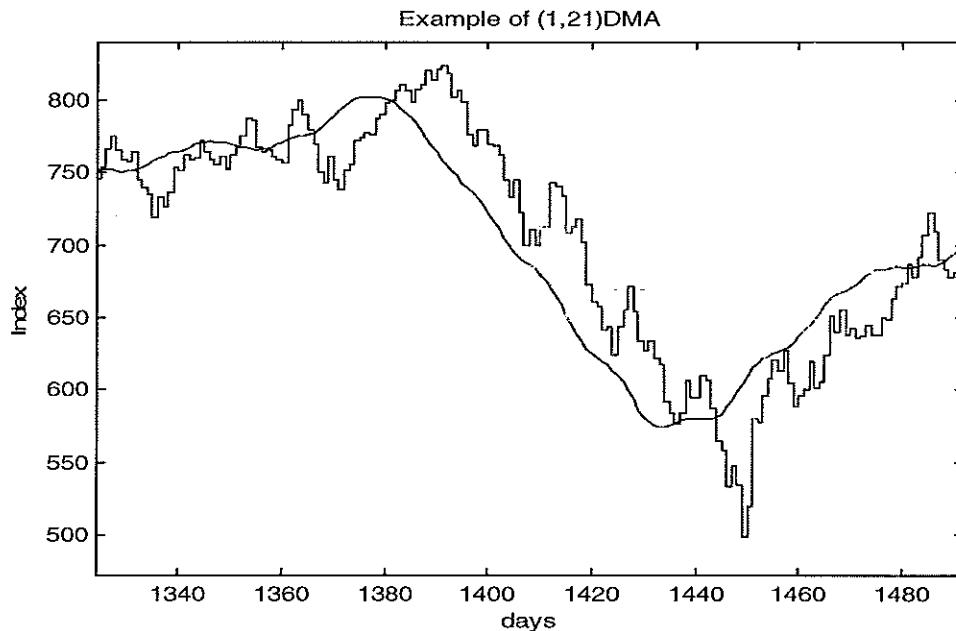


Fig. 11 Example of (1,21) DMA

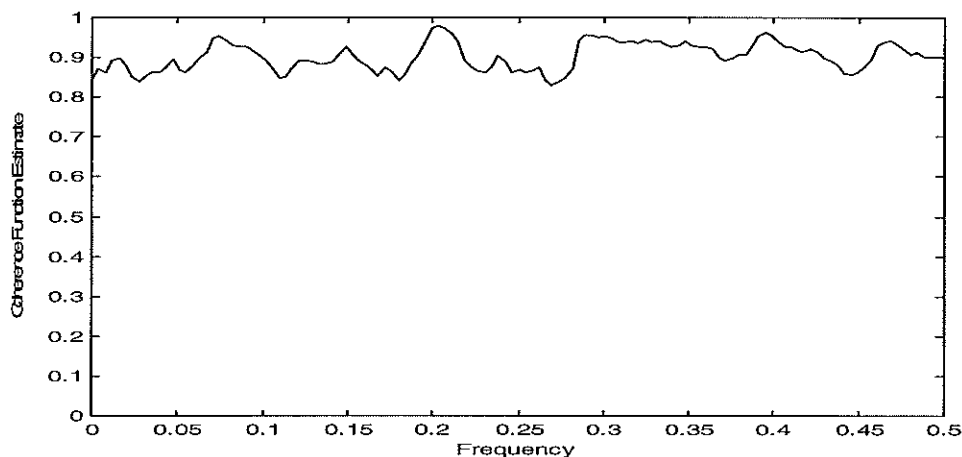


Fig. 10 Coherence DMA 9301-9906

SO10

The Swedish ten-year bond interest. The bond is an alternative investment to stocks and the interest will therefore have an impact on the general pricing on the stockmarket. If the bond interest goes up the bond will be a better investment. As a result the stock prices will fall.

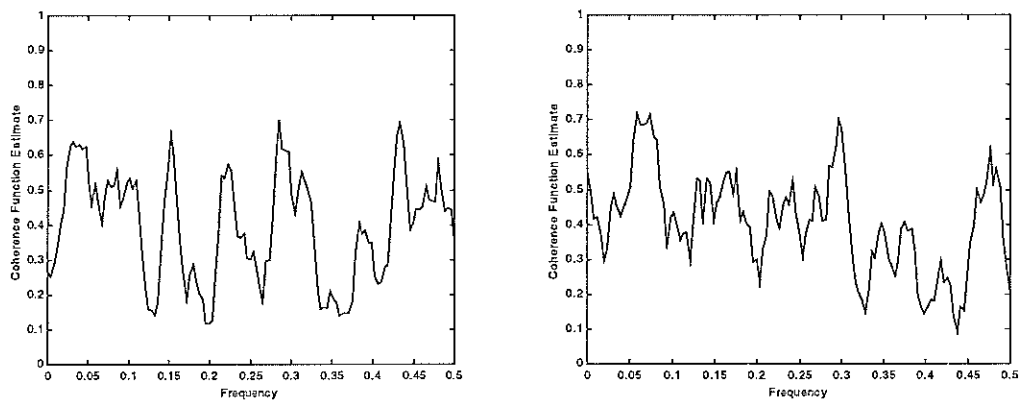


Fig. 13 Coherence plots SO10 - OMX

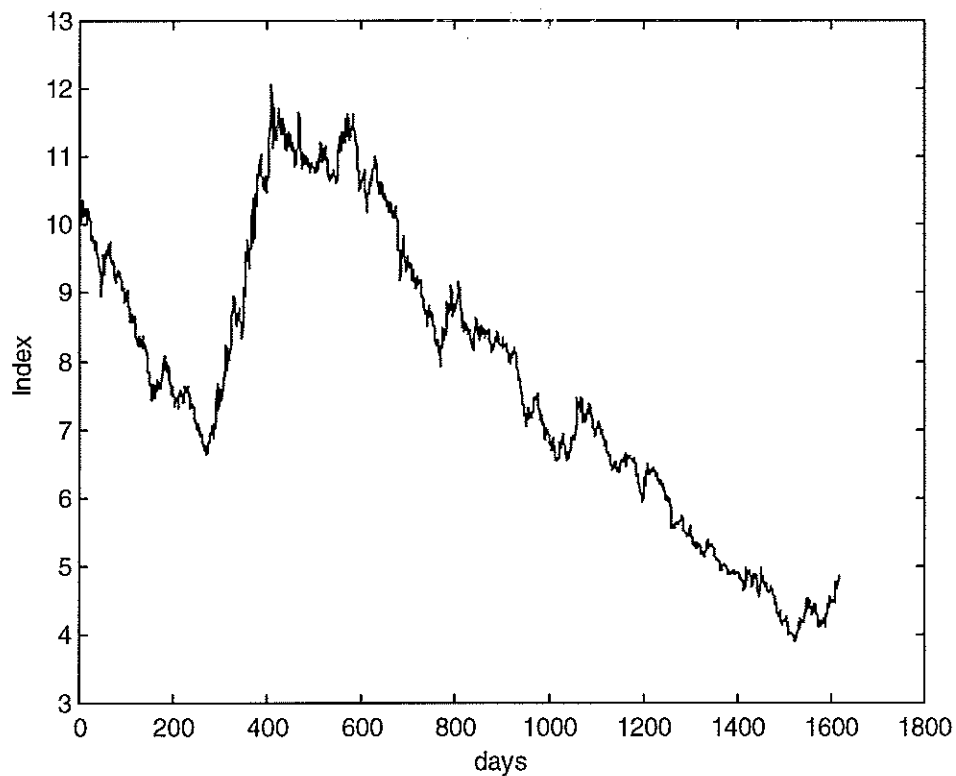


Fig. 12 SO10 930104 - 990614

U.S Dollar – SEK Exchange rate

The variations of the Dollar-SEK Exchange rate can be viewed as an indicator of the strength of the Swedish market versus the U.S market. Since it is a continuously changing variable it can, sampled at an appropriate moment, be used as an input signal for a prediction-model

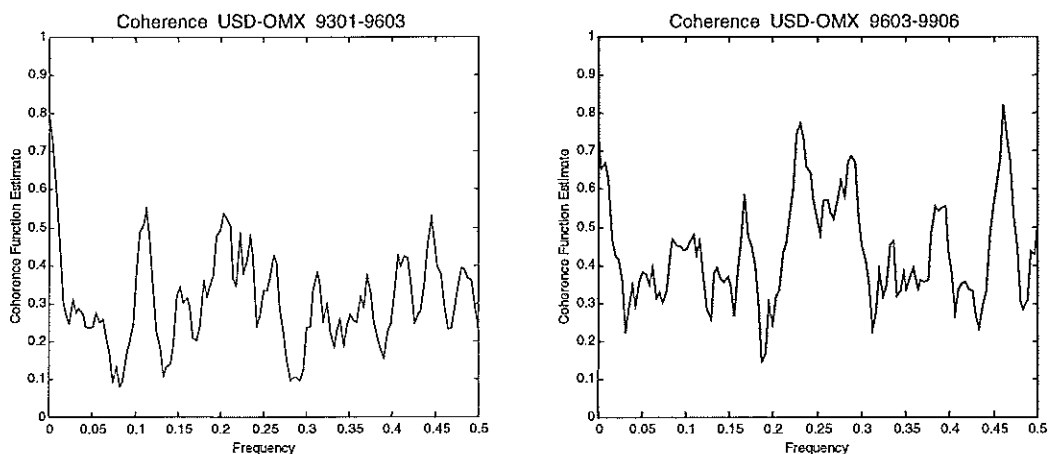


Fig. 14 Coherence plots USD-OMX

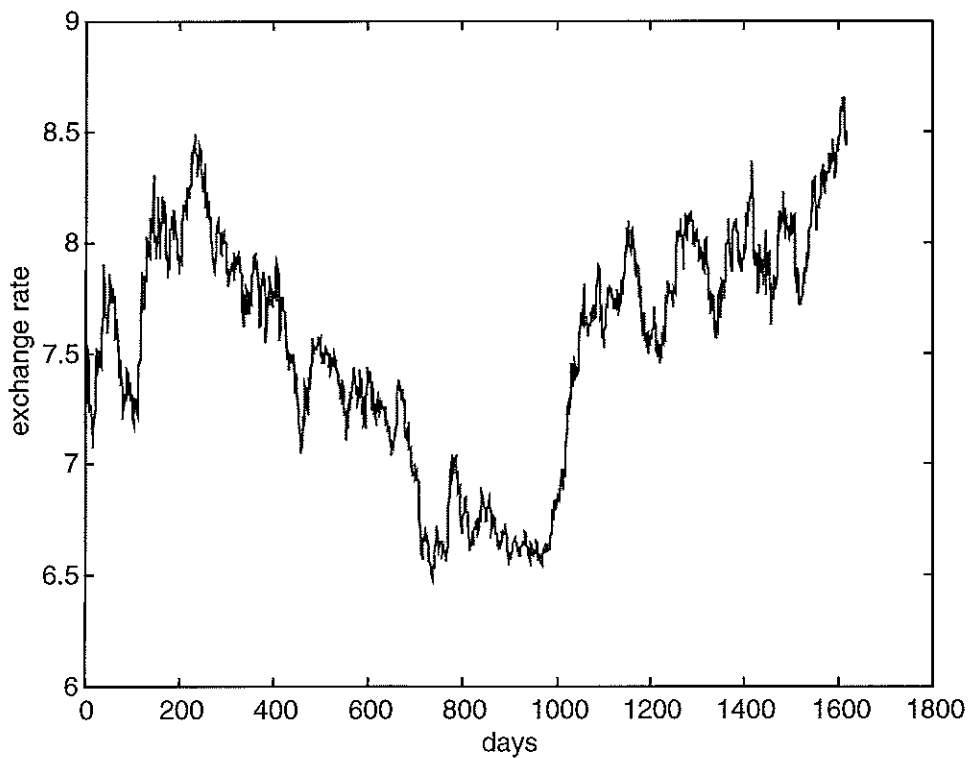


Fig. 15 SEK/USD 930104-990614

Psychological variable

The psychological variable is a constructed time series from a survey^{iv} on anticipations on the OMX considering the last few days' result. The answers from the survey were collected and plotted in a bar diagram to see if a significant coherence among the answers could be found.

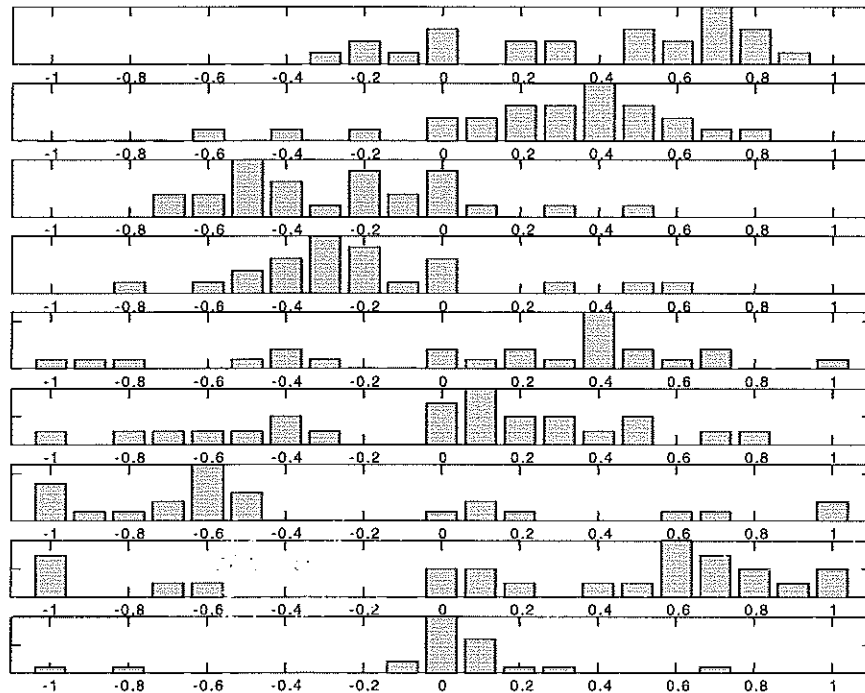


Fig. 16 Bar diagram of results from survey.

Although there are too few answers to the survey to have any statistical certainty, one can see that certain values are stronger than other. From the diagram, values were chosen and implemented in the algorithm that is used to calculate the variable psyk^v . This variable was tested for coherence with OMX-index and the result is plotted below.

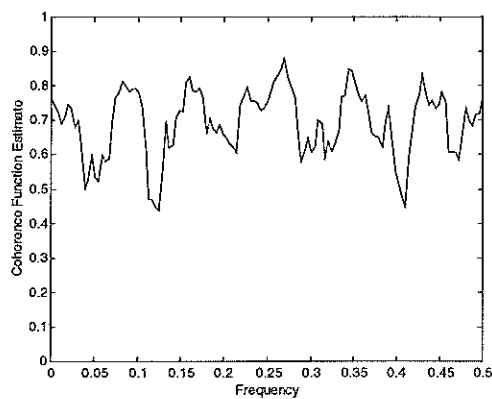


Fig. 15 Coherence Psyk-OMX 9301-9603

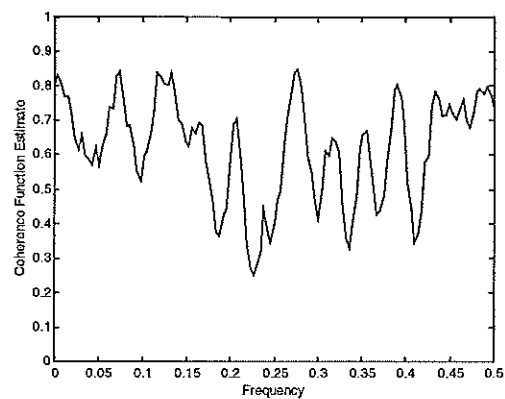


Fig. 16 Coherence Psyk-OMX 9603-9906

Methods

Time-invariant methods

The most intuitive way to describe a discrete model with inputs and outputs is the difference equation. The difference equation uses old inputs and outputs directly to form a new value of the output.

$$y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} = b_1 u_{k-1} + \dots + b_n u_{k-n}$$

The two following methods are difference equation methods, the ARMAX and the ARX, where the latter is a special case of the first.

ARMAX

A general form of difference equations is called ARMAX, autoregressive moving average with exogenous input. The ARMAX model has the form

$$A(z^{-1})y_k = z^{-d}B(z^{-1})u_k + C(z^{-1})w_k$$

The polynomial $A(z^{-1})$ contains information as to which and how much of the input parameters that affects the model, in the same way $B(z^{-1})$ controls the influence of the input and $C(z^{-1})$ the perturbations. The d is the order of delay between input and output. The model can be separated into a deterministic part and a stochastic part according to

$$\begin{aligned} A(z^{-1})x_k &= B(z^{-1})u_k \\ A(z^{-1})v_k &= C(z^{-1})w_k \end{aligned} \quad \Rightarrow y_k = x_k + v_k$$

ARX

The controlled autoregressive model (ARX)

$$A(z^{-1})y_k = z^{-d}B(z^{-1})u_k + w_k$$

is a special case of the ARMAX without the perturbation modeling and is therefore the deterministic part of the more general model.

Both of the algorithms uses ordinary least-squares to estimate the model and nothing of interest can really be said about them, except on the order selection which is covered below

AR Order Selection.

As the model order is not known beforehand, some error criterion must be computed for several different model orders.

Since the purpose of the model is to make a prediction we are not so much interested in finding the correct model order as we are interested in finding a good model for prediction.

Keeping prediction in mind we choose a model order optimization criteria which selects the model order that gives the best prediction, namely the Akaike final prediction error (FPE).

The FPE is calculated as

$$FPE(p) = \frac{N+p}{N-p} \cdot \frac{2}{N} \cdot V_N$$

where N is the number of observations we have, p is the model order and V_N is the loss function.

The choice of model order is then made from

$$\hat{p} = \arg \min FPE(p)$$

There is no straightforward way to select the orders on a multi-input model such as this. Calculating the FPE for every possible combination is not feasible because it would, in this case, mean 20^{16} ($6.55 \cdot 10^{20}$) different FPE:s

Instead of calculating every possible combination, the orders are chosen according to the following algorithm^{vi}:

1. nn (ordervector) is set to ones
2. The FPE:s is calculated from orders 1 to 20 for the first parameter in nn
 3. The order for which FPE is smallest is chosen
4. The FPE:s is calculated for the second parameter in nn
 5. This is continued to the last parameter
6. Steps 2-5 is repeated until nn doesn't change between two following calculations

This algorithm is used for the ARX, ARMAX, SEGMENT1 and ITERATIV models. The N4SID uses singular values to determine model order.

State-space Methods

The most part of systems identification is concerned with polynomial models, however these are known to give rise to numerically ill-conditioned mathematical problems, especially for Multi Input Multi Output systems

Consider the innovation model state-space system

$$\begin{aligned}x_{k+1} &= A x_k + B u_k + w_k \\ y_k &= C x_k + D u_k + e_k\end{aligned}$$

with input u_k , output y_k , state vector x_k , process noise w_k and measurement noise e_k .

The problem is, that from the input sequence $\{u_k\}_{k=1}^N$ and the output sequence $\{y_k\}_{k=1}^N$, consistently estimate the system matrices A,B,C and D.

N4SID

Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems.

The N4SID algorithm is somewhat complicated and demands some explaining.

We begin with forming the block Hankel matrices

$$U_{0|i-1} = \begin{pmatrix} u_0 & u_1 & u_2 & \cdots & u_{j-1} \\ u_1 & u_2 & u_3 & \cdots & u_j \\ \dots & \dots & \dots & \dots & \dots \\ u_{i-1} & u_1 & u_{i+1} & \cdots & u_{i+j-2} \end{pmatrix} \quad Y_{0|i-1} = \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{j-1} \\ y_1 & y_2 & y_3 & \cdots & y_j \\ \dots & \dots & \dots & \dots & \dots \\ y_{i-1} & y_i & y_{i+1} & \cdots & y_{i+j-2} \end{pmatrix}$$

Then the future outputs are projected onto the past and future inputs and the past outputs. The result is described as a function of the system matrices and the input-output block Hankel matrices.

$$Z_i = Y_{i|2i-1} / \left(\frac{U_{0|2i-1}}{Y_{0|i-1}} \right) \quad \text{and} \quad Z_{i+1} = Y_{i+1|2i-1} / \left(\frac{U_{0|2i-1}}{Y_{0|i}} \right)$$

These projections can be rewritten as follows

$$Z_i = \left(\begin{array}{c|c|c} \underbrace{L_i^1}_{l(i \times m)} & \underbrace{L_i^2}_{l(i \times m)} & \underbrace{L_i^3}_{l(i \times i)} \\ \hline \end{array} \right) \left(\begin{array}{c} U_{0|i-1} \\ U_{i|2i-1} \\ Y_{0|i-1} \end{array} \right) \quad \text{and} \quad Z_{i+1} = \left(\begin{array}{c|c|c} \underbrace{L_{i+1}^1}_{l(i-1 \times m(i+1))} & \underbrace{L_{i+1}^2}_{l(i-1 \times m(i-1))} & \underbrace{L_{i+1}^3}_{l(i-1 \times l(i+1))} \\ \hline \end{array} \right) \left(\begin{array}{c} U_{0|i} \\ U_{i+1|2i-1} \\ Y_{0|i} \end{array} \right)$$

with

$$\begin{aligned} L_i^1 &= \Gamma_i \left((A^i - Q_i \Gamma_i) S (R^{-1})_{11mi} + \Delta_i^d - Q_i H_i^d \right) \\ L_i^2 &= H_i^d + \Gamma_i \left[A^i - Q_i \Gamma_i \right] S (R^{-1})_{mi+112mi} \\ L_i^3 &= \Gamma_i Q_i \end{aligned}$$

where

$$\Gamma_i' = \text{extended observability matrix} = \begin{pmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{i-1} \end{pmatrix}$$

$$\Delta_i^d = \text{reversed extended controllability matrix} = \begin{pmatrix} A^{i-1}B & A^{i-2}B & \dots & AB & B \end{pmatrix}$$

$$H_i^d = \text{lower block triangular Toeplitz matrix} = \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix}$$

$$\text{and} \quad \mathbf{E} \left[\begin{pmatrix} w_k \\ e_k \end{pmatrix} \begin{pmatrix} w_l' & e_l' \end{pmatrix} \right] = \begin{pmatrix} Q^s & S^s \\ (S^s)' & R^s \end{pmatrix} \delta_{kl}$$

In an approach to determine the order of the system one notes that the column space of L_i^1 and L_i^3 coincides with the column space of Γ_i . This means that Γ_i and n can be determined from any one of these matrices.

Let T be any rank deficient matrix whose column space coincides with that of Γ_i

Perform a SVD

$$T = (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} V'$$

Since T is of rank n the number of singular values different from zero will be the order of the system. In reality it is often not unambiguous to determine when a singular value is non-zero as the following plot shows

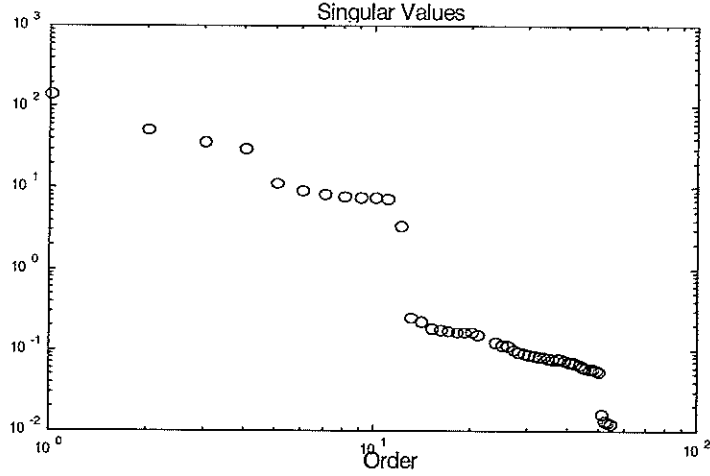


Fig. 17 Singular values used for order determination

Furthermore Γ_i can be put equal to $U_1 \Sigma_1^{1/2}$.

According to theorem 1 in [?] the estimated states can be written as

$$\begin{aligned}\hat{X}_i &= \Gamma_i^+ (Z_i - H_i^d U_{i|2i-1}) \\ \hat{X}_{i+1} &= \Gamma_{i+1}^+ (Z_{i+1} - H_{i+1}^d U_{i+1|2i-1})\end{aligned}$$

and as a conclusion from this same theorem

$$\begin{pmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \hat{X}_i + \begin{pmatrix} B \\ D \end{pmatrix} U_{i|i} + \begin{pmatrix} U_{0|2i-1} \\ Z_i \\ \hat{X}_i \end{pmatrix}^\perp$$

If we now combine these last two equations we get

$$\begin{pmatrix} \Gamma_{i+1}^+ Z_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \Gamma_i^+ Z_i + \begin{pmatrix} \Lambda_{12} \\ \Lambda_{22} \end{pmatrix} U_{i|2i-1} + \begin{pmatrix} U_{0|2i-1} \\ Z_i \\ \hat{X}_i \end{pmatrix}^\perp$$

which from its first term gives us the system matrices A and C. The matrices B and D appear linearly in Λ_{12} and Λ_{22} and can therefore be found there through

$$\begin{pmatrix} \Lambda_{12} \\ \Lambda_{22} \end{pmatrix} = \begin{pmatrix} B - A\Gamma_i^+ \begin{pmatrix} D \\ \Gamma_{i-1} B \end{pmatrix} & \Gamma_{i-1}^+ H_{i-1}^d - A\Gamma_i^+ \begin{pmatrix} 0 \\ H_{i-1}^d \end{pmatrix} \\ D - C\Gamma_i^+ \begin{pmatrix} D \\ \Gamma_{i-1} B \end{pmatrix} & -C\Gamma_i^+ \begin{pmatrix} 0 \\ H_{i-1}^d \end{pmatrix} \end{pmatrix}$$

This is of course not a complete account for the algorithm, but at least a comprehensive start and a basic understanding.

Time-variant Models

Two different algorithms have been used to compute time-variant models for prediction. One method where the model order and the model are estimated for every sample and one method where the Matlab algorithm segment have been used and altered.

The first method^{vii} estimates the model order using FPE for samples $k-400$ to k , then estimates the model parameters using ARX for the same samples and finally predicts the result for sample $k+1$. This method returns a vector of the predicted values $[yp]$.

The second method^{viii} uses the Segment algorithm in Matlab and the model order estimation algorithm for samples $k-400$ to k and then predicts the result for sample $k+1$. This method returns the vector $[yp]$, and the parameters of the model used at every sample $[segm]$

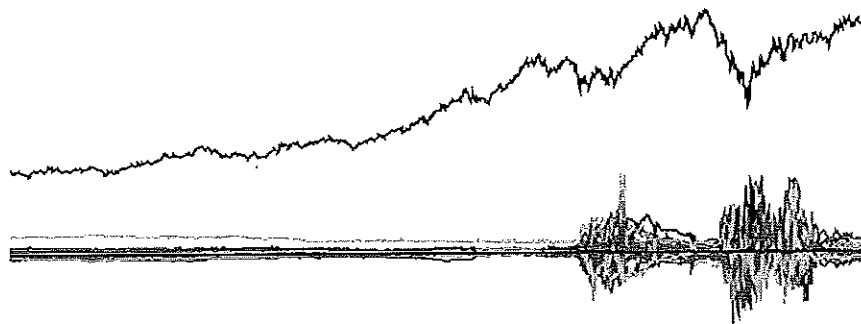


Fig. 18 OMX and parameters from segment

As can be seen from the plot above, the magnitude of the parameters changes when the stockmarket undergoes considerable changes, such as last years (1998) “crash”. With a suitable choice of variance (in Segment1) these parameters could be used to determine at an early stage if a major correction is at hand or not.

Results

The outcome of the prediction algorithms is evaluated to see which identification is most suited for predicting this kind of financial process. Several tests are needed to come to a valid conclusion.

Three different tests are performed against the true outcome of the OMX-index.

The comparing of signs which tells you if the prediction was right in the direction of the market.

$$Sign = \frac{\sum_{k=1}^N (sign(omx2) == sign(yp))}{N}$$

Next is the Variance-Accounted-For, which is calculated as

$$VAF = \frac{1 - \text{variance}(omx2 - yp)}{\text{variance}(omx2)} \times 100\%$$

This value indicates how close the original signal and its estimate resembles each other. If they are equal the value is 100%. If not, the value is lower, down to minus infinity if the error is bigger than the signal.

The last test is the residual test, which looks at the correlation of the residuals and the signals.

Sign test

The Sign-values for every prediction is calculated and plotted against the time when it can be calculated, for example, the “open” value is dependent on the opening prices of the Dow Jones and can therefore be calculated at 15.30 Central European time.

The pattern that presents itself hints that the correctness of the prediction might be linear to the amount of information provided. At least within the relevant range an attempt to interpolate to the time of 17.30 can be made. This is useful because in the data available to this work only the opening- and closing-prices are given and if you would like to do the prediction as accurate as possible, you need the prices for 17.30. Of course, if this work would be used for the purpose of trading, the data needed could be collected.

In the plot one sees that the N4SID- and Segment1-methods give the best results for the opening values.

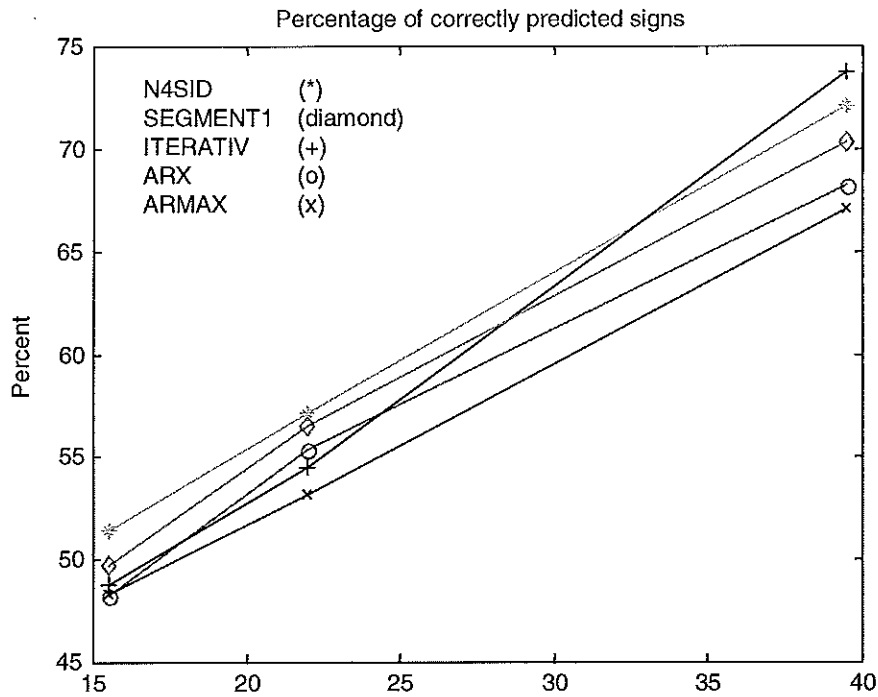


Fig. 19 The Sign-value plotted against time. (15.30 day one to 15.30 day two)

VAF

The Variance-Accounted-For test gives results that are mostly negative, which could be expected given the high noise-ratio in financial data, which in turn will give large errors. Other than that observation nothing else is deduced from the VAF-test. The N4SID- and Iterativ-methods gives the best results for the opening values.

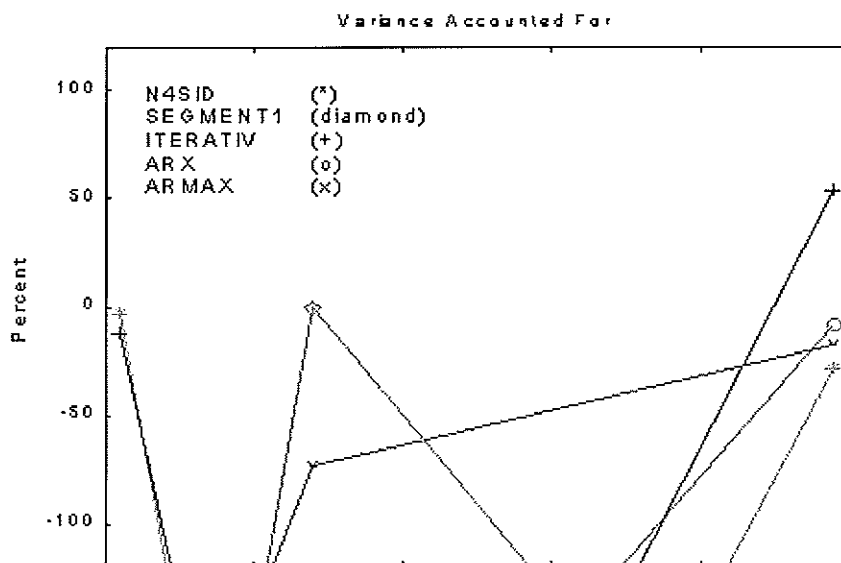


Fig. 20 VAF-values plotted against time. (15.30 day one to 15.30 day two)

Residual tests

One way of determining if a model is correct is examining if there is a structure in the residuals obtained. If all information in the inputs and the outputs is used in the model, the residuals should be structureless and uncorrelated to every other signal.

A few tests are made on the residuals of the prediction:

$$\varepsilon_k = \text{omx}2_k - y_{k,\text{predicted}}$$

First, an autocorrelation test, and second a crosscorrelation test of the inputs and the residuals.

The autocovariance-function of the residuals

$$\hat{C}_{\varepsilon\varepsilon} = \frac{1}{N - \tau} \sum_{k=\tau+1}^N \varepsilon_k \varepsilon_{k-\tau}$$

are put together and normalized in a vector

$$r_{\varepsilon\varepsilon} = \frac{1}{\hat{C}_{\varepsilon\varepsilon}(0)} \left(\hat{C}_{\varepsilon\varepsilon}(1) \quad \dots \quad \hat{C}_{\varepsilon\varepsilon}(m) \right).$$

To fulfill the requirement of structurelessness and uncorrelation this vector needs to be asymptotically distributed as

$$\sqrt{N} r_{\varepsilon\varepsilon} \xrightarrow{\text{dist}} \text{N}(0, I_{m \times m})$$

The test-variable we are using in the autocorrelation test is

$$\tau_{\varepsilon\varepsilon} = N r_{\varepsilon\varepsilon}^T r_{\varepsilon\varepsilon} \xrightarrow{\text{dist}} \chi^2(m)$$

which is tested with a standard ANOVA. If $\tau_{\varepsilon\varepsilon} < \chi_{\alpha}^2(m)$, then H_0 is accepted.

The cross-covariance test is performed in much the same way with

$$\hat{C}_{u\varepsilon} = \frac{1}{N - \tau} \sum_{k=\tau+1}^N \varepsilon_k u_{k-\tau}$$

and

$$r_{u\varepsilon} = \frac{1}{\sqrt{\hat{C}_{\varepsilon\varepsilon}(0)}} \left(\hat{C}_{u\varepsilon}(\tau+1) \quad \dots \quad \hat{C}_{u\varepsilon}(\tau+m) \right)$$

and the matrix

$$\hat{R}_{uu}(m) = \frac{1}{N - m} \sum_{k=m+1}^N \begin{pmatrix} u_{k-1} \\ \vdots \\ u_{k-m} \end{pmatrix} \begin{pmatrix} u_{k-1} & \dots & u_{k-m} \end{pmatrix}$$

The asymptotic distribution is

$$\sqrt{N} r_{ue} \xrightarrow{dist} N(0, \hat{R}_{uu})$$

and the test variable which is ANOVA tested is

$$\tau_{ue}(m) = N r_{ue}^T \hat{R}_{uu}^{-1} r_{ue} \xrightarrow{dist} \chi^2(m)$$

The results of the residual testing can be visualized by plotting the residual vector with the 99%-confidence interval. The null hypothesis H_0 is accepted if the residuals are within the confidence interval.

See f ex. the plots below. The rest of the residual plots can be found in Appendix B.

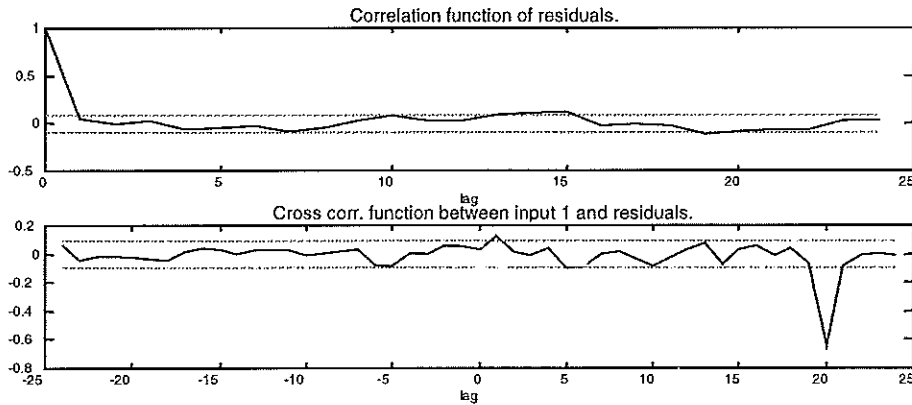


Fig. 21 Example of residual-test plots

From the plots in Appendix B one can see that there in most cases are a negative spike at lag #20 of the cross-correlation between the input and the residuals, this can probably be explained by the fact that the AR-order determination algorithm limits the order to 20. This is corroborated by the fact that the N4SID that doesn't use the algorithm, doesn't display this singularity.

From the plots one can see that the N4SID- and Segment-methods give the best results.

The residual tests has been given a major roll in the development of the models since they have been used continually during the work to improve the selection of model order and the model values.

Implementation

Just for the fun of it, a crude algorithm was written to see just how much money one would earn, or loose, if the methods here were used in trading. Several important facts are disregarded to keep the calculation simple. Only information to 15.30 is used, that is 2 hours worth of information lost. The difference between buy- and sell-prices is disregarded because no such information is available. A transaction fee of 0.14% is used whatever the amount traded for is, or how many different companies involved. In reality the best commodity would probably be options or futures.

The results are compared to a simple Buy and Hold Strategy, which follows index and would give a result of 222 680 SEK.

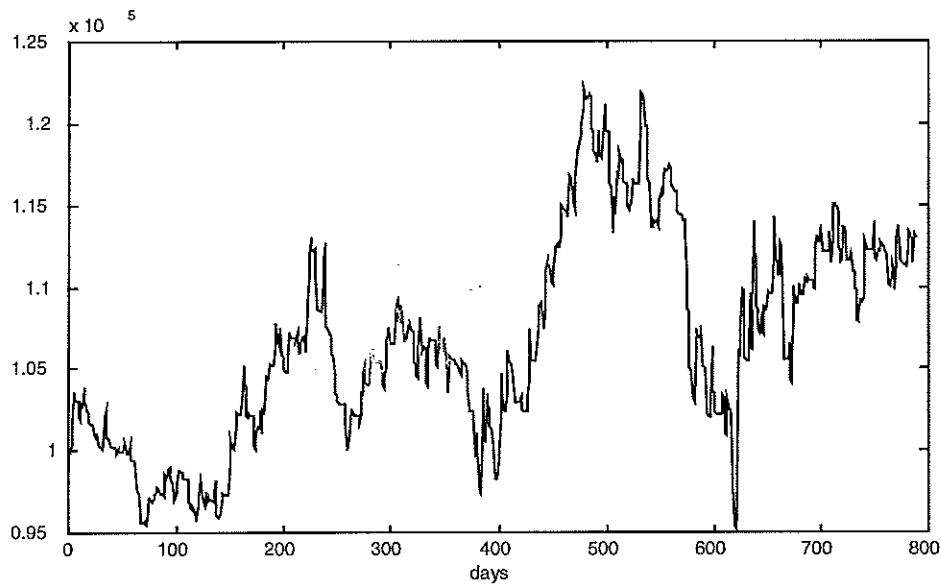


Fig. 22 Cash-result after 796 days from N4SID_Open 113,010 SEK

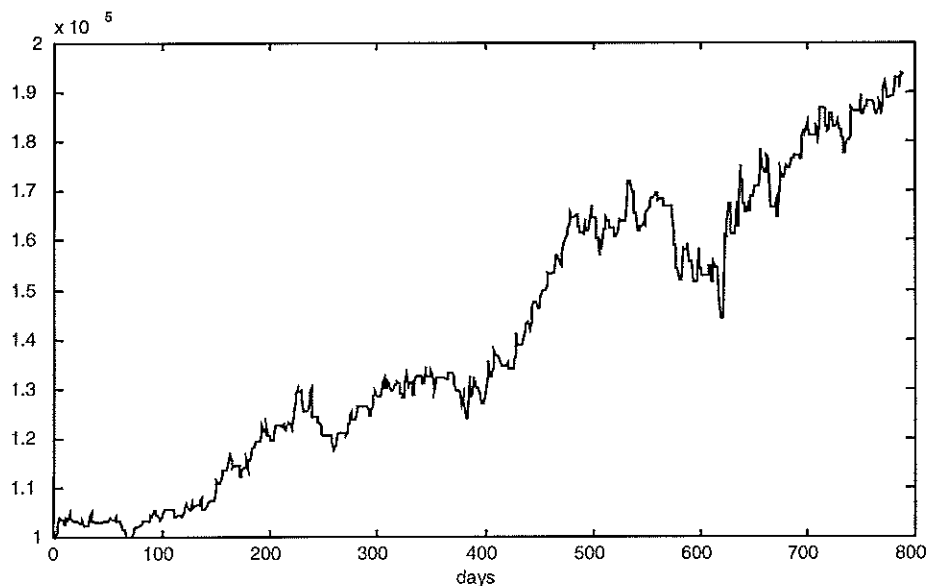


Fig. 23 Cash-result after 796 days from N4SID_Open, no transaction costs 193,540 SEK

If the interpolation mentioned in connection to the sign-test is made, a result of approximately 53.5% can be reached, which would give a result like this

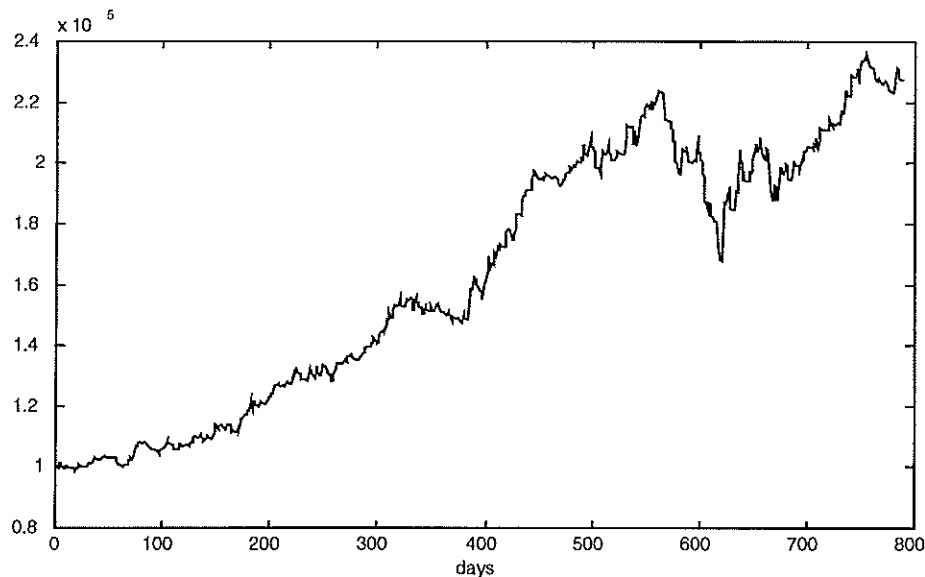


Fig. 24 Cash-result after 796 days from N4SID_Open, interpolated to 17.30 228,310 SEK

Conclusions

The approach to this identification-/prediction- problem, global variables working on a single index without the influences of result-reports or announcements from the Federal Reserve Bank, for example, gives predictions which are true if nothing happens in anything which can influence the predicted index. Thus every result-report, newspaper article or even rumour are perturbations to this system, and since there are thousands of such perturbations every day they will have a great effect on the predicted index.

The result of the prediction should therefore be viewed as a guide only, to the general trend of the market. If in using this method one could filter away big perturbations, the model would be drastically improved.

There are a few results in this work which are worth to be pointed out:

The first thing is the high coherence between different markets, despite the fact that very few of the stocks traded are the same at two markets.

Second is the change in coherence between these same markets, during the last six years. In every coherence test, the coherence was higher 96-99, than 93-96. This may be a tangible sign of the ongoing globalization.

The third finding is the parameter values in the Segment1-method, which seems to be changing early when the market is undergoing corrections. This has not been tested further but would it prove to hold it would be a valuable tool.

Future work

The things that show greatest potential to improve the results of this work is:

1. Other input signals
2. Other State-Space methods
3. More relevant data (see above)

Apart from the above mentioned there are a few more things that has shown themselves to have potential, but will remain unmentioned for obvious reasons.

Appendix A

In Appendix A, all MATLAB[®] code that has been written for the work is collected.

Appendix B

In Appendix B, plots of the residual tests have been compiled.

Sort

```
%makro used to sort data

% If there is a date in the new variable that doesn't exist in the old
data, then that day
% is removed and it's value copied onto the previous date.
% If a date is missing in the new variable that exists in the old data,
then that day is
% created and the value of the previous day is copied onto it.
E=data;
F=B;      % B : New variable
k=0;

while k<12000;
    k=k+1;
    C=E(k,1)-F(k,1);

    if C < 0;
        F=[F(1:k-1,:); E(k,1) F(k-1,2); F(k:length(F),:)]];
    end
    if C > 0;
        F=[F(1:k-2,:); F(k-1,1) F(k,2); F(k+1:length(F),:)]];
        k=k-1;
    end
end;
end;
```

Psyk

```
function [y] = psyk(index)

%PSYK function used to make the Psyk-variable for a whole time-series

for k=1:5
    psyk=0;
end

for k=6:length(index)
if sign(index(k-1))==1
    psyk(k)=0.7;
    if sign(index(k-2))==1
        psyk(k)=0.4;
        if sign(index(k-3))==1
            psyk(k)=0.4;
            if sign(index(k-4))==1
                psyk(k)=-0.6;
            end
        end
    end
end
else
    psyk(k)=-0.5;
    if sign(index(k-2))==-1
        psyk(k)=-0.3;
        if sign(index(k-3))==-1
            psyk(k)=0.1;
            if sign(index(k-4))==-1
                psyk(k)=0.6;
            end
        end
    end
end
end
y = psyk;
```

ARX-Order

```
function [nn]=arxorder(Z);
%ARXORDER Select model order for ARX, according to Final Prediction
Error
%
% [nn]=arxorder(Z)
% Z : input-output data
% nn : order vector

nn=[ones(1,2*(size(Z,2)-1)-1) 0 0]; %sets all n to 1 except nk-psyk and
mov
figure, pause(0.5)
for f=1:5;
for p=1:(length(nn)-2);
fpe=[];
for k=1:20;
nn(1,p)=k;
TH = ARX(Z(1:800,:),nn);
fpe=[fpe TH(2,1)]; %final prediction error
end;
bar(1:20,fpe(1,:));
axis([0 25 (min(fpe)-0.2*(max(fpe)-min(fpe)))
(max(fpe)+0.04*(max(fpe)-min(fpe))))]);
pause(0.5)
[d,i]=min(fpe); % selects minimum FPE
nn(1,p)=i; % sets nn to minimum FPE
end;
nn=nn
end
```

ARMAX-Order

```
function [nn]=armaxorder(Z);
%ARMAXORDER Select model order for ARMAX, according to Final Prediction
%Error
%
% [nn]=armaxorder(Z)
% Z : input-output data
% nn : order vector

nn=[ones(1,2*(size(Z,2)-1)) 0 0]; %sets all n to 1 except nk-psyk and mov
figure
pause(1)
for f=1:3;
for p=1:(length(nn));
fpe=[];
for k=1:20;
TH = ARMAX(Z(1:800,:),[nn(1,1:size(Z,2)) nn(1,size(Z,2)+1)-1
nn(1,size(Z,2)+2:2*size(Z,2))]);
fpe=[fpe TH(2,1)];
end;
bar(1:20,fpe(1,:));
title(num2str(p));
axis([0 25 (min(fpe)-0.2*(max(fpe)-min(fpe)))
(max(fpe)+0.04*(max(fpe)-min(fpe))))]);
pause(1)
[d,i]=min(fpe); % selects minimum FPE
nn(1,p)=i; % sets nn to minimum FPE
end;
end;
end;
```

Segment1

```
function [segm,V,yp]=segment1(z,nn,r2,q,r1,M,th0,p0,lifelength,mu)

%SEGMENT1 Segments, tracks abruptly changing systems and predicts one
step ahead.
% [segm,V,yp] = SEGMENT(z,nn,r2,q)
%
% z : The output-input data vector z = [y u].
% nn: ARX or ARMAX model orders nn = [na nb nk] or nn = [na nb nc nk].
% See also ARX or ARMAX. The algorithm handles multi-input systems.
%
% r2: The equation error variance(Default:estimated, but better to
guess)
% q : The probability that the system jumps at each sample.(Default
0.01)
% segm: The parameters of the segmented data. Row k is for sample # k
% The parameters are given in "alphabetical order".
% V: The loss function corresponding to segm
% yp: The predicted time-series
% The time-varying estimates th, and the estimated values of r2 are
given
% by [segm,V,th,r2] = SEGMENT(z,nn)
%
% Original code by:
% L. Ljung 10-1-89
% Copyright (c) 1986-98 by The MathWorks, Inc.
% $Revision: 2.3 $ $Date: 1997/12/02 03:44:05 $
echo off
if nargin < 2
    disp('Usage: SEGM = SEGMENT(DATA,ORDERS)')
    disp([' [SEGM,LOSS] = '
SEGMENT(DATA,ORDERS,NOISE_VAR,JUMP_PROB,',...
'JUMP_SIZE,No_OF_MODELS,TH0,P0,LL,MU)'])
    return
end

[Ncap,nz]=size(z);
if Ncap<nz,disp('Warning: Have you entered the data as column
vectors?'),end
nu=nz-1; nl=length(nn);
if nu==0,na=nn(1);nb=0;nc=0;nk=1;if nl>1,nc=nn(2);end
if nl>2,error('For a time series, nn = na or nn = [na nc]!'),end
else
    na=nn(1);nb=nn(2:nu+1);
    if nl==2*nu+1,nk=nn(nu+2:2*nu+1);nc=0;
    else if nl==2*nu+2,nc=nn(nu+2);nk=nn(nu+3:2*nu+2);
    else error('Incorrect number of orders specified: nn=[na nb nk] or
nn=[na nb nc nk]'),end,end
end
d=na+sum(nb)+nc; % Number of parameters
if nargin<10, mu=[];end
if nargin<9, lifelength=[];end
if nargin<8 p0=[];end
if nargin<7,th0=[];end
if nargin<6 M=[];end
if nargin<5 r1=[];end
if nargin<4 q=[];end
if nargin<3 r2=[];end
if isempty(q),q=0.01;end
if isempty(r1),r1=eye(d);end
if isempty(M),M=5;end
if isempty(th0),th0=zeros(1,d);end
if isempty(p0),p0=10*eye(d);end
if isempty(lifelength),lifelength=1;end
if isempty(mu),mu=0.97;end
```

```

if isempty(r2), r2dum=1; R2v=ones(1,M); r2=r2dum; estr2=1; else estr2=0; end;
alfa=i/M*ones(1,M); % [1, zeros(1,M-1)];

seg2=[]; kp=[]; ko=[]; yp=[];
thm=zeros(Ncap,d);
Phicap=zeros(Ncap,d);
th=th0'*ones(1,M); age=zeros(1,M); hist=zeros(Ncap,M); r2e=zeros(Ncap,1);
P=[]; for i=1:M, P=[P p0/r2]; end;

if nc>0, nu=nu+1; nk(nu)=1; nb(nu)=nc; z(1:nc, nu+1)=zeros(nc,1); end
for i=max([na+1, max(nb+nk), nc+1]):Ncap-1
    phi=[-z(i-1:-1:i-na,1)];
    for ku=1:nu,
        phi=[phi; z(i-nk(ku):-1:i-nk(ku)-nb(ku)+1,1+ku)];
    end
    y=z(i,1);
    for j=1:M,
        Pj=P(:, d*(j-1)+1:d*(j-1)+d);
        den(j)=(r2+phi'*Pj*phi);

        epsi(j)=y-th(:,j)'*phi;

        alfabar(j)=alfa(j)/sqrt(den(j))*exp(-
        epsi(j)^2/(2*den(j)));
        th(:,j)=th(:,j)+1/den(j)*P(:, d*(j-1)+1:d*(j-
        1)+d)*phi*epsi(j);
        P(:, d*(j-1)+1:d*(j-1)+d)=Pj-(Pj*phi*phi'*Pj)/den(j);
    end;

    aind=find((age(1:M)>lifelength)); if length(aind)>0,
        [dummy, jmin]=min(alfabar(aind)); jmin=aind(jmin);
        [dummy, jmax]=max(alfabar);

        P(:, d*(jmin-1)+1:d*(jmin-1)+d)=P(:, d*(jmax-1)+1:d*(jmax-1)+d)+r1;
        th(:, jmin)=th(:, jmax);
        alfabar(jmin)=q*alfabar(jmax); age(jmin)=0;
        hist(:, jmin)=hist(:, jmax); hist(i, jmin)=1;
        else jmax=1; end
        age=age+ones(1,M);
        alfa=1/sum(alfabar)*alfabar;

        theta=th*alfa';
        epsilon=y-theta'*phi; if nc>0, z(i, nu+1)=epsilon; end
%
% estimate R2
%
if estr2
    r2dum=r2dum+(1-mu)*(epsilon^2-r2dum);
    agedum=max(age,2)-1;
    R2v= R2v +max(ones(1,M)./agedum, (1-mu)).*(epsi.^2./den-R2v);
    Rr=R2v(find(age>d));
    if length(Rr)>0, r2=min(Rr); else r2=r2dum; end
end

%
%
thm(i,:)=theta';
seg2=[seg2; theta'];

r2e(i)=r2; Phicap(i,:)=phi';
%

hh=hist(:, jmax);
kk=find(hh==1);
ko=[ko kk'];
kp=[1 ko Ncap];

```

```

np=length(kp);
thm(kp(np),:)=thm(i,:);
for k=2:np
segm(kp(k-1):kp(k),:)=ones(kp(k)-kp(k-1)+1,1)*thm(kp(k),:);
end
figure(1);plot(segm); % Plots model-parameters
title(num2str(i));
pause(0.1);

%-----
-
A=[1 segm(end,nn(1))];

B1=[zeros(1,nn(size(z,2)+1)) segm(end,nn(1)+1:nn(2)+1)];p=nn(1)+nn(2);
B2=[zeros(1,nn(size(z,2)+2)) segm(end,p+1:p+nn(3))];p=p+nn(3);
B3=[zeros(1,nn(size(z,2)+3)) segm(end,p+1:p+nn(4))];p=p+nn(4);
B4=[zeros(1,nn(size(z,2)+4)) segm(end,p+1:p+nn(5))];p=p+nn(5);
B5=[zeros(1,nn(size(z,2)+5)) segm(end,p+1:p+nn(6))];p=p+nn(6);
B6=[zeros(1,nn(size(z,2)+6)) segm(end,p+1:p+nn(7))];p=p+nn(7);
B7=[zeros(1,nn(size(z,2)+7)) segm(end,p+1:p+nn(8))];

B=[B1 zeros(1,14-length(B1));
B2 zeros(1,14-length(B2));
B3 zeros(1,14-length(B3));
B4 zeros(1,14-length(B4)); % Makes A and B polynomials for ARX
B5 zeros(1,14-length(B5));
B6 zeros(1,14-length(B6));
B7 zeros(1,14-length(B7))];

%-----
-----
theta=poly2th(A,B,1); % Creates theta-model from A and B
yv=predict(z(1:i+1,:),theta); % Predicts i+1
yp=[yp yv(end)];
figure(2);stem(yp);hold;stem(z(15:i+1,1),'r');hold; % Plots yp and omx in
the same plot
%
end
e=z(:,1)-sum(segm'.*Phicap)';
V=e'*e/Ncap; % Loss-function
Vaf=vaf(z(15:1596,1),yp); %"Variance Accounted For" (See SMI)
Sign=sum(sign(yp)==sign(z(15:1596,1)')) % Percentage of correct signs

```

Iterativ

```
function [yp]=iterativ(Z);

%ITERATIV Timevariant ARX-identification and prediction one step ahead.
% [yp] = ITERATIV(z)
%
% z : The output-input data vector z = [y u].
% yp: The predicted time-series

yp=[];

for s=801:1596
s
nn=[ones(1,2*(size(Z,2)-1)-1) 0 0]; %sets all n to 1 except nk-psyk and
mov
for f=1:3;
for p=1:(length(nn)-2);
fpe=[];
for k=1:20;
nn(1,p)=k;
TH = ARX(Z(s-800:s-1,:),nn);
fpe=[fpe TH(2,1)]; %final prediction error
end;
%bar(1:20,fpe(1,:));
%axis([0 25 (min(fpe)-0.2*(max(fpe)-min(fpe)))
(max(fpe)+0.04*(max(fpe)-min(fpe)))]);
%pause(1)
[d,i]=min(fpe); %selects minimum FPE
nn(1,p)=i; %sets nn till minimum FPE
%pause;
end;
end
bar(1:length(nn),nn);
pause(0.3)

th=arx(Z(s-800:s-1,:),nn); % Makes ARX-model for the time to s-1
y=predict(Z(s-400:s,:),TH); % Predicts to the time s
yp=[yp y(end)];
end;
```

Test

```
function [y]=test(index);

%TEST function that calculates the outcome of 100,000 SEK put on the
%stockmarket and left there.

for k=1:9
test(k)=100000-99;
end
for k=10:1570;
    test(k)=test(k-1)*(1+(index(k)));
end
y=test;
```

Money

```
function [x,c]=pengar(index2,pred);
%PENGAR Simulates the outcome of investments
% based on the predicted variable yp
%
% x=result of investment
% c=number of transactions made
% index2=real time-series
% pred=predicted time-series
c=0;
pengar=100000;
vektor(1)=100000;
h=1;
for k=10:796;
    h=h+1;
    if sign(pred(k))==1
        if sign(pred(k-1))==-1
            pengar=pengar*0.9986;
            c=c+1;
        end
        pengar=pengar*(1+(index2(k)));
    else
        if sign(pred(k-1))==1
            pengar=pengar*0.9986;
            c=c+1;
        end
        pengar;
    end
    vektor(h)=pengar;
end
plot([1:1:length(vektor)],vektor,'-');
x=pengar
```

Money3

```
function [x,c]=pengar3(index2,pred);

%PENGAR Simulates the outcome of investments
% based on the predicted variable yp
% Short selling is used when yp indicates a drop
% x=result of investment
% c=number of transactions made
% index2=real time-series
% pred=predicted time-series

c=0;
pengar=100000;
vektor(1)=100000;
h=1;
for k=10:796;
    h=h+1;
    if sign(pred(k))==1
        if sign(pred(k-1))==-1
            pengar=pengar*0.9986;
            c=c+1;
        end
        pengar=pengar*(1+(index2(k)));
    else
        if sign(pred(k-1))==1
            pengar=pengar*0.9986;
            c=c+1;
        end
        pengar=pengar*(1-(index2(k)));
    end
    vektor(h)=pengar;
end
plot([1:1:length(vektor)],vektor,'-');
x=pengar
```

MovAvg

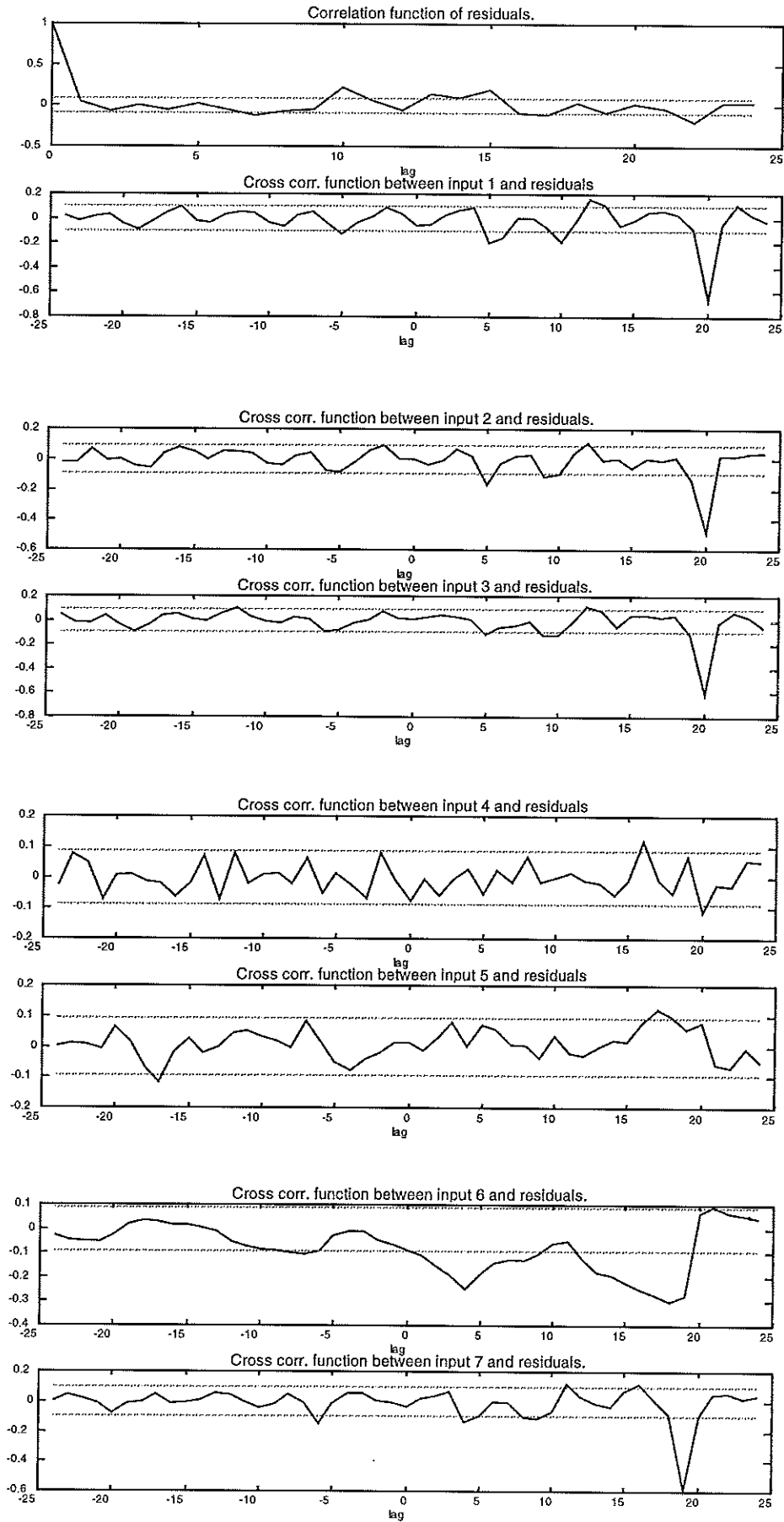
```
function [mov]=movavg(index,days);

%MOVAVG Calculates the Double Moving Average variable for whole time-
series

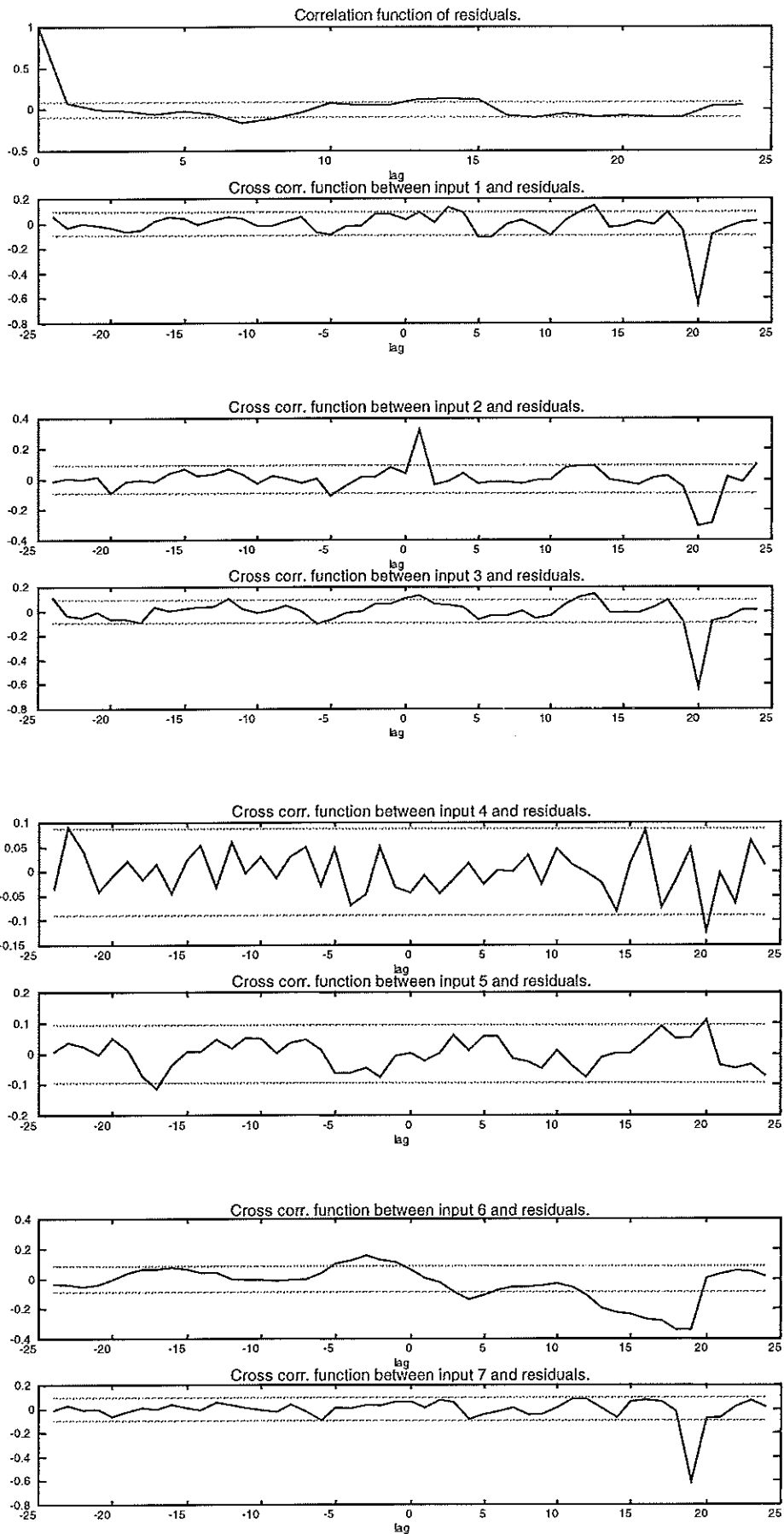
mov=[];
avg=[];

for k=days:length(index);
    avg=[avg sum(index(k-days+1:k))/days];
    mov=[mov; index(k)-avg(end)];
end;
plot([1:1597],avg)
end
```

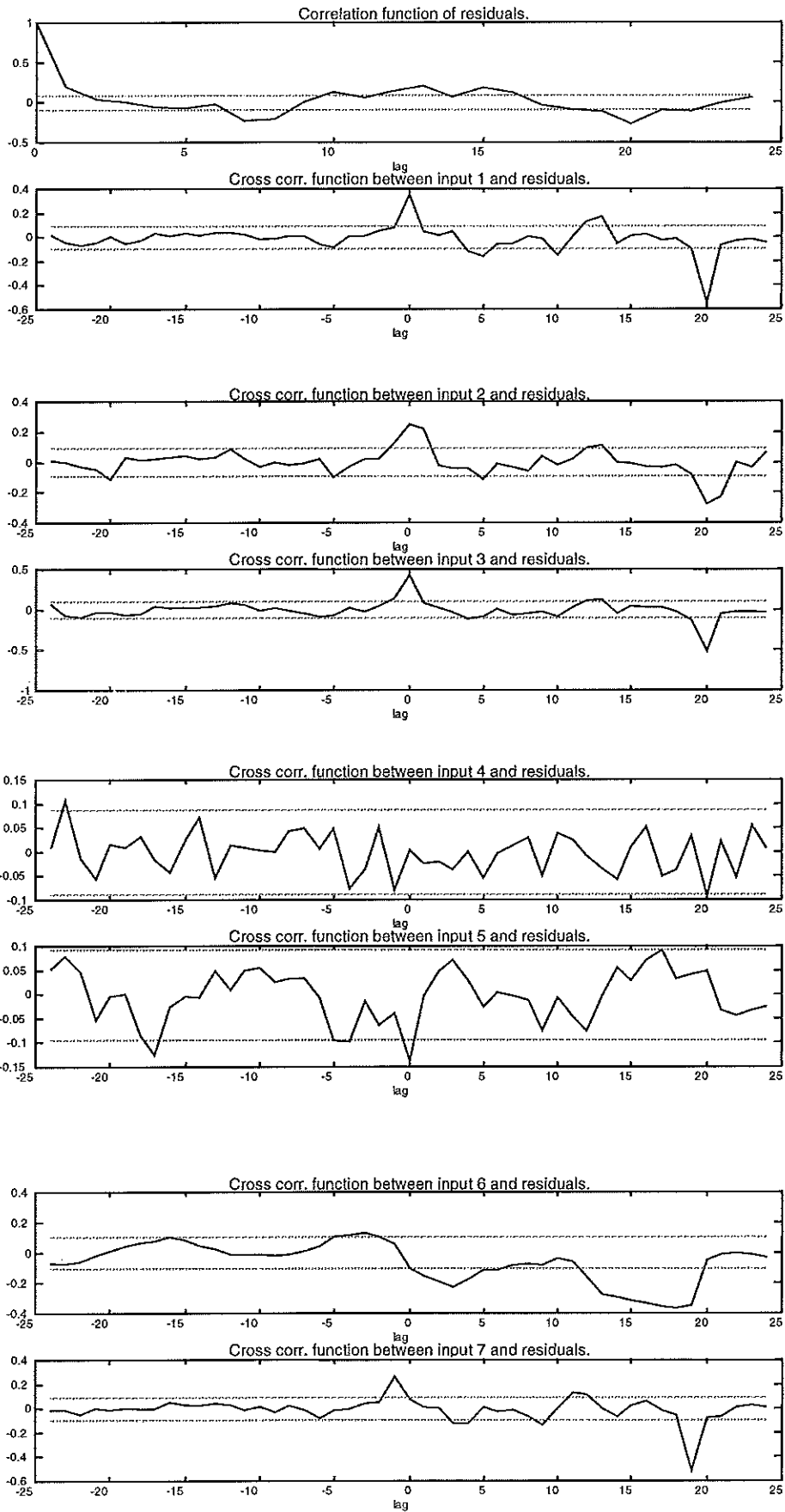

ARX_Open



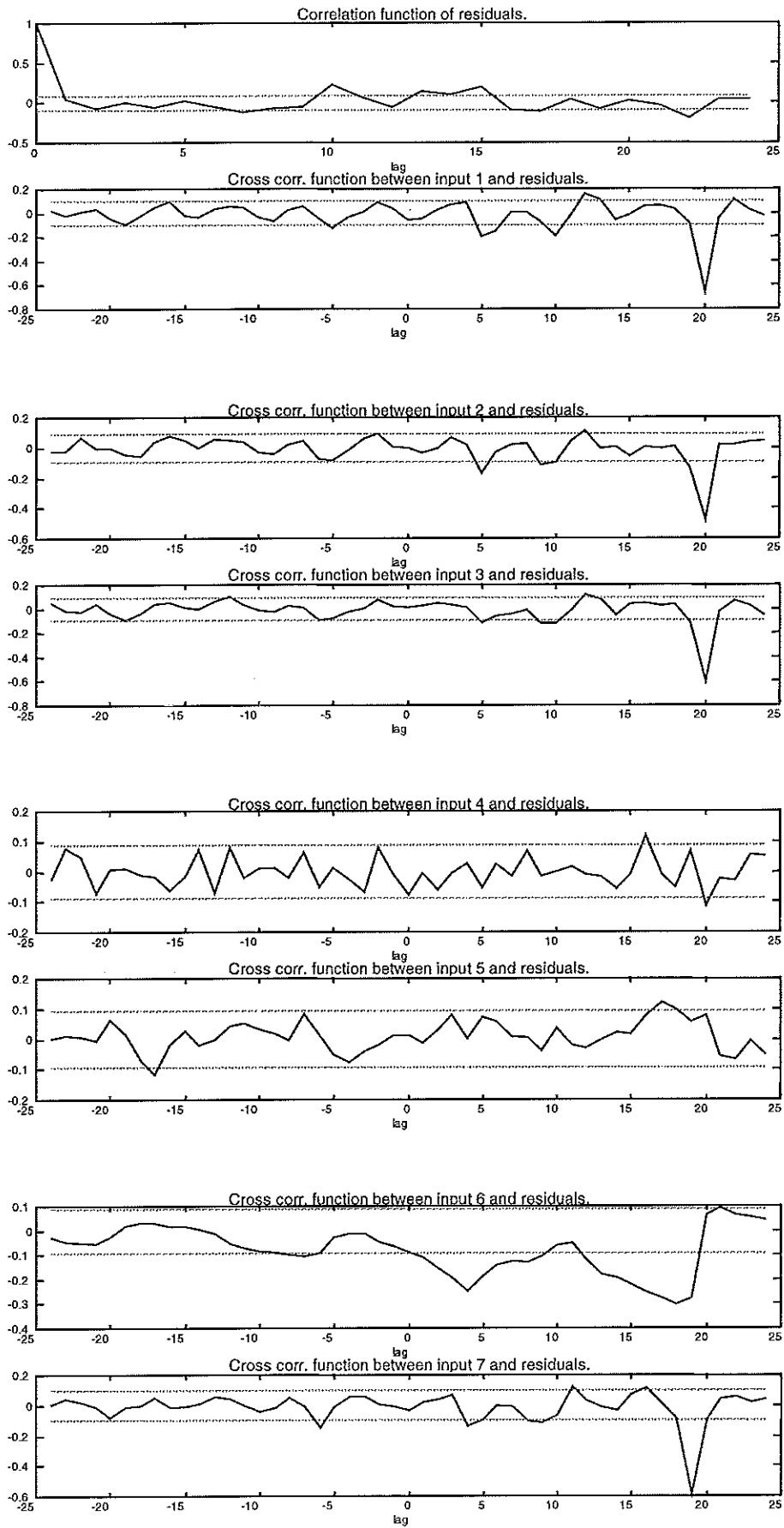
ARX_Close



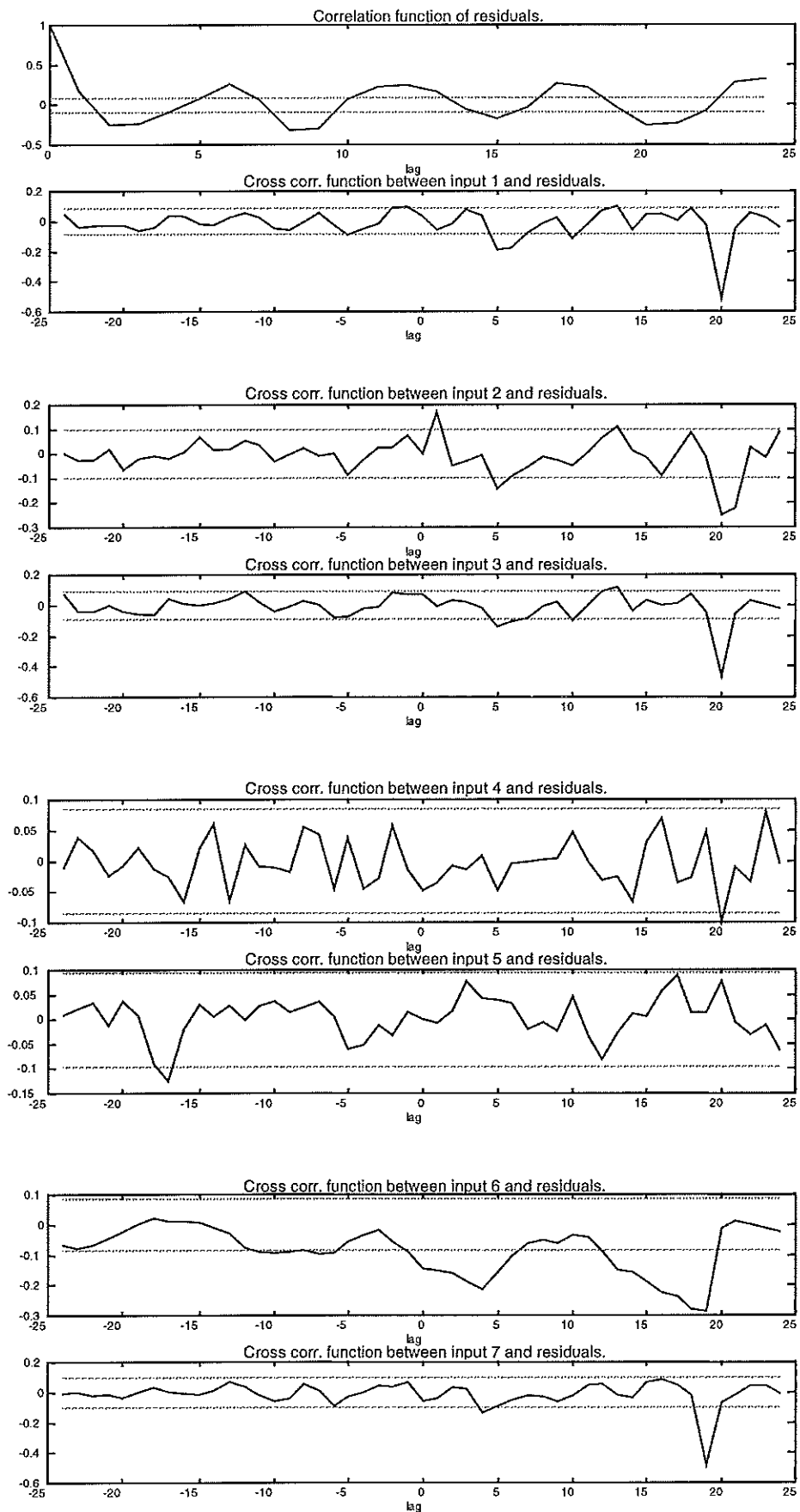
ARX_Direct



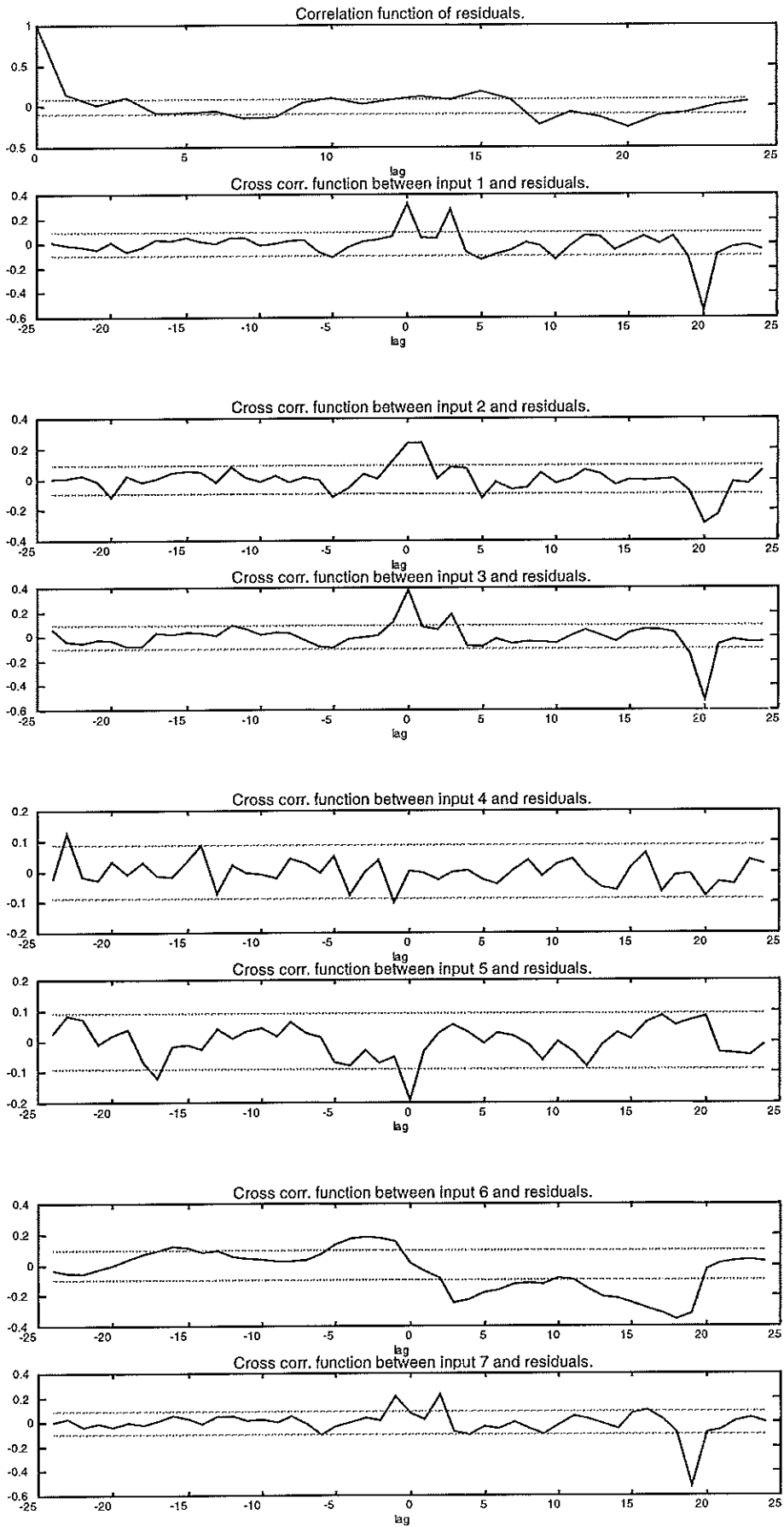
ARMAX_Open



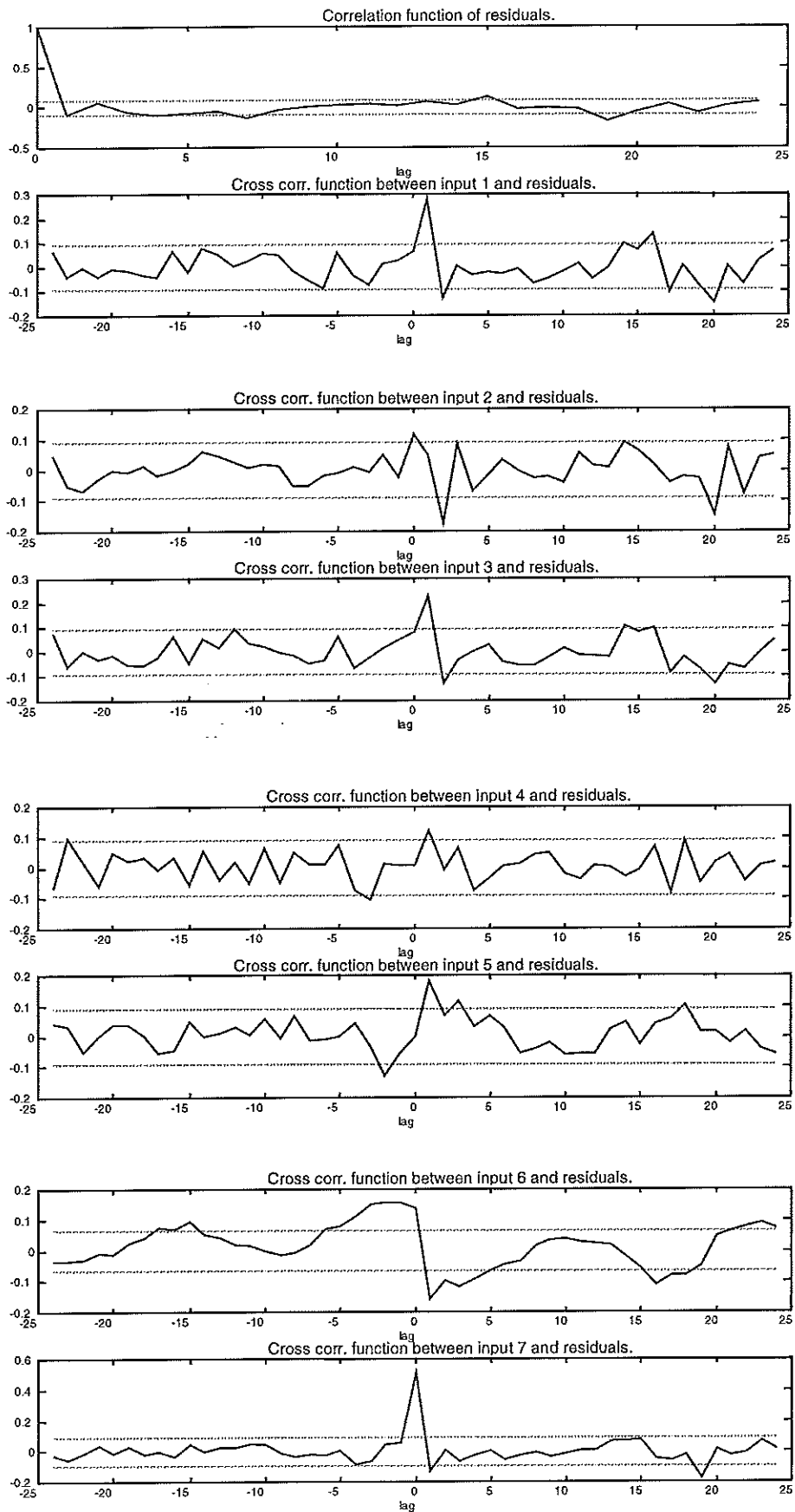
ARMAX_Close



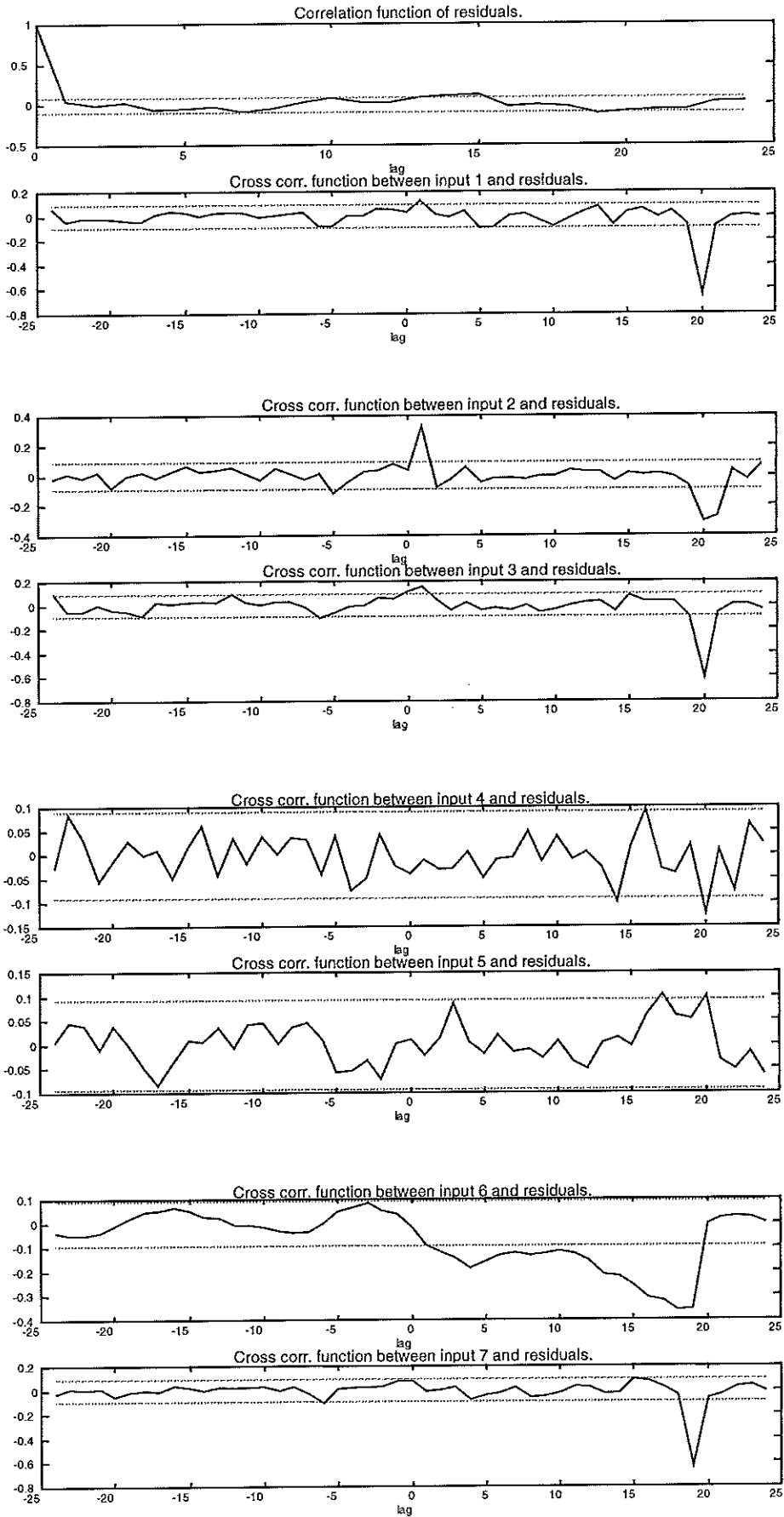
ARMAX_Direct



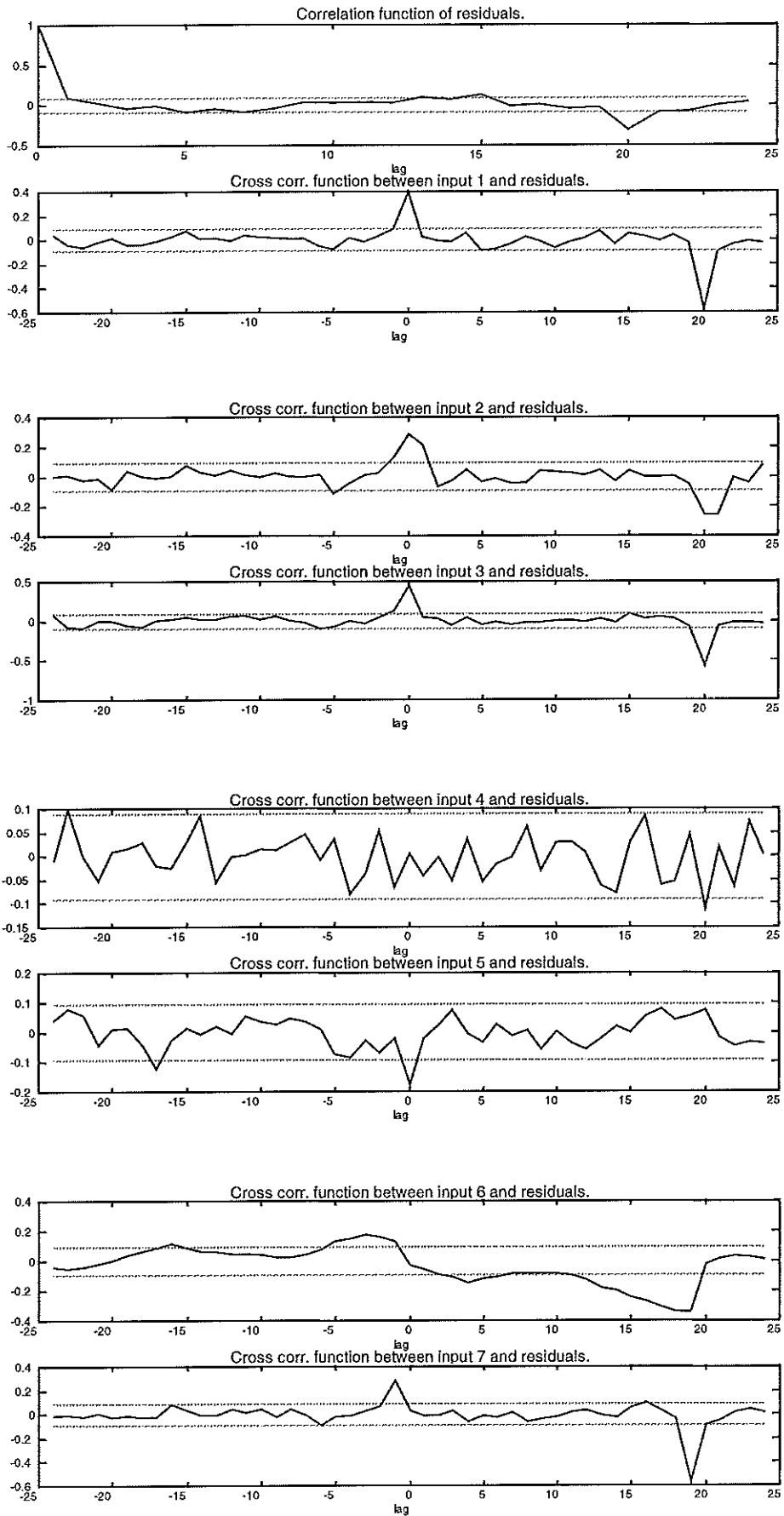
N4SID_Open



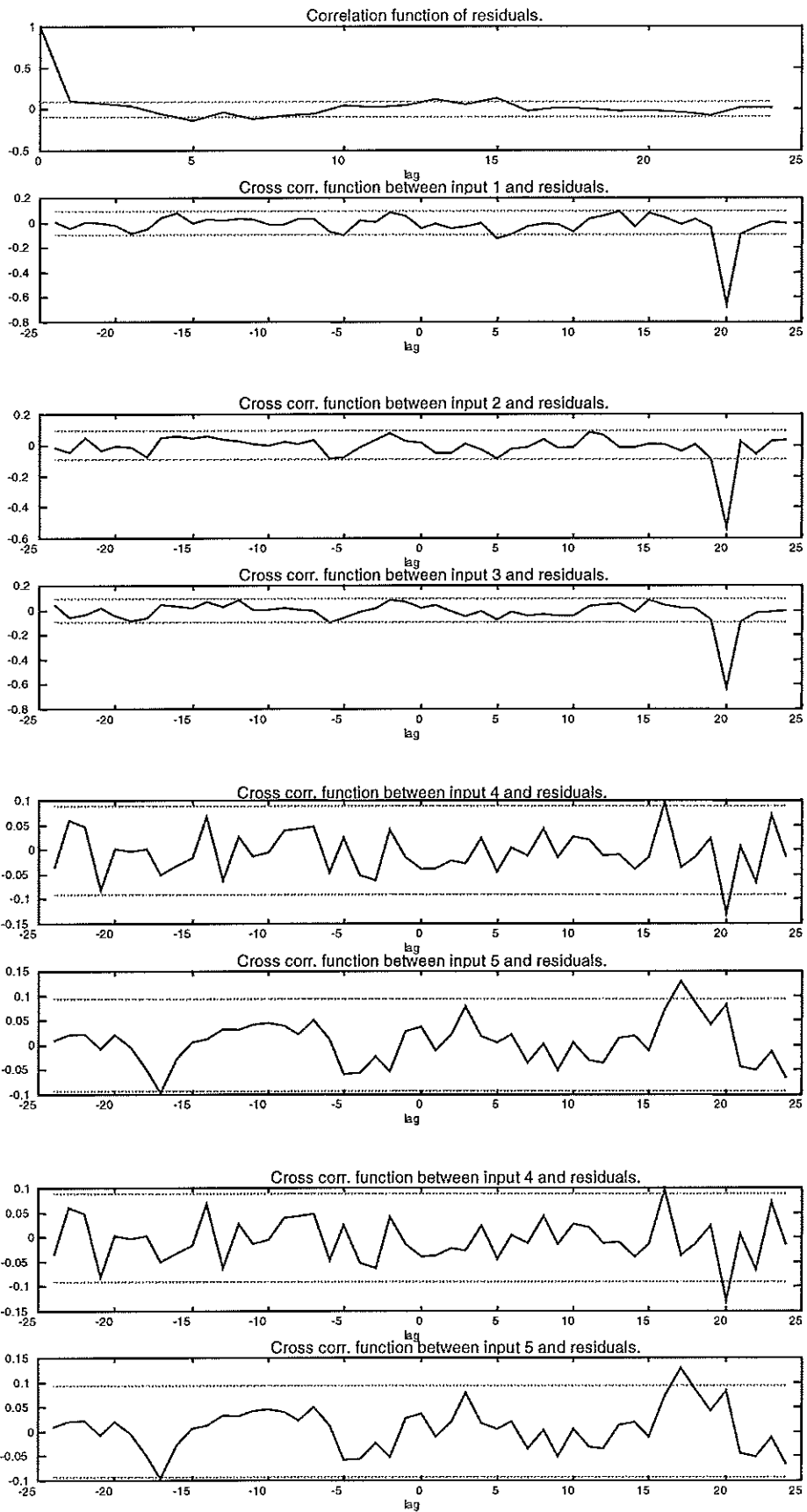
N4SID_Close



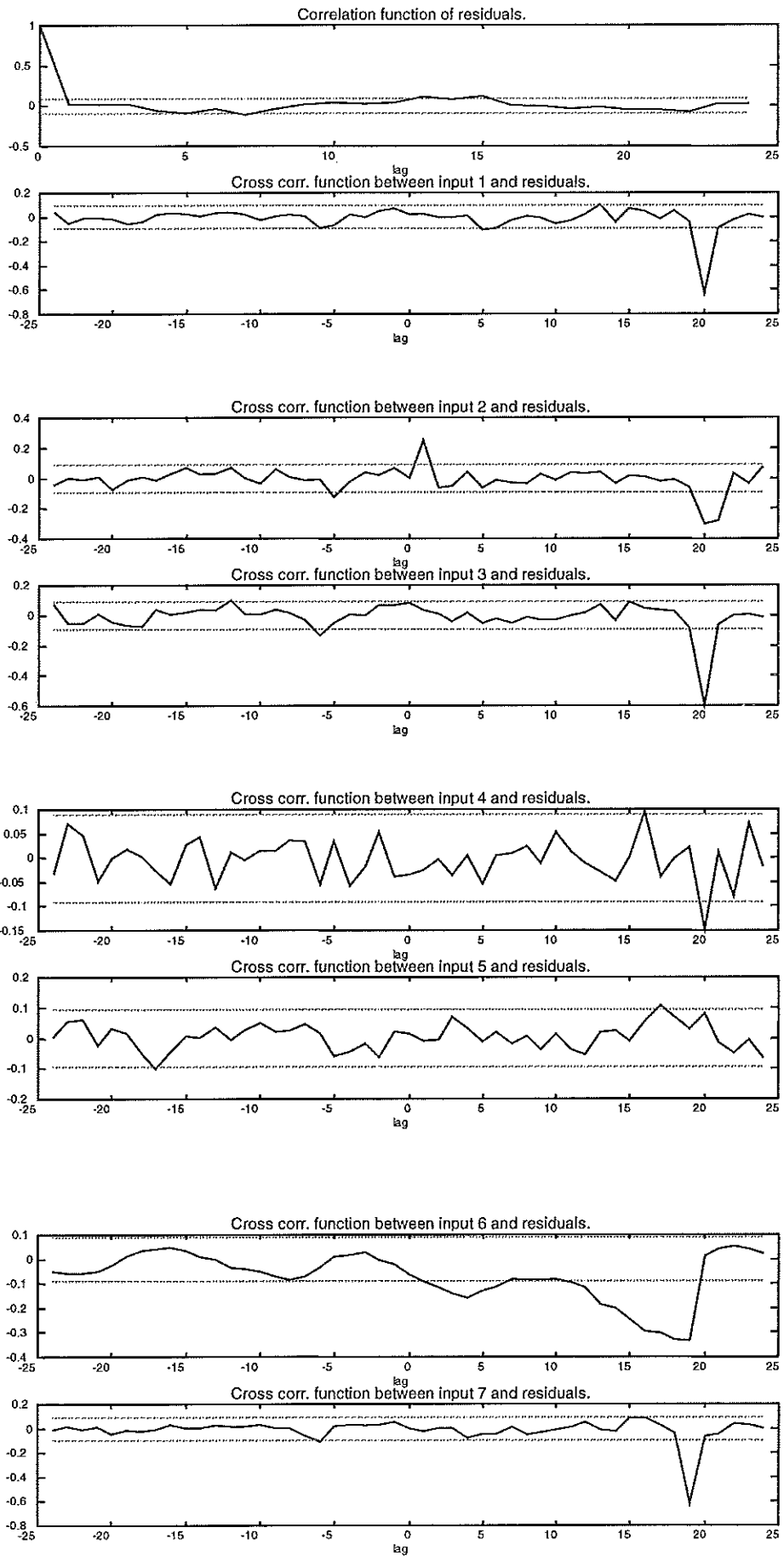
N4SID_Direct



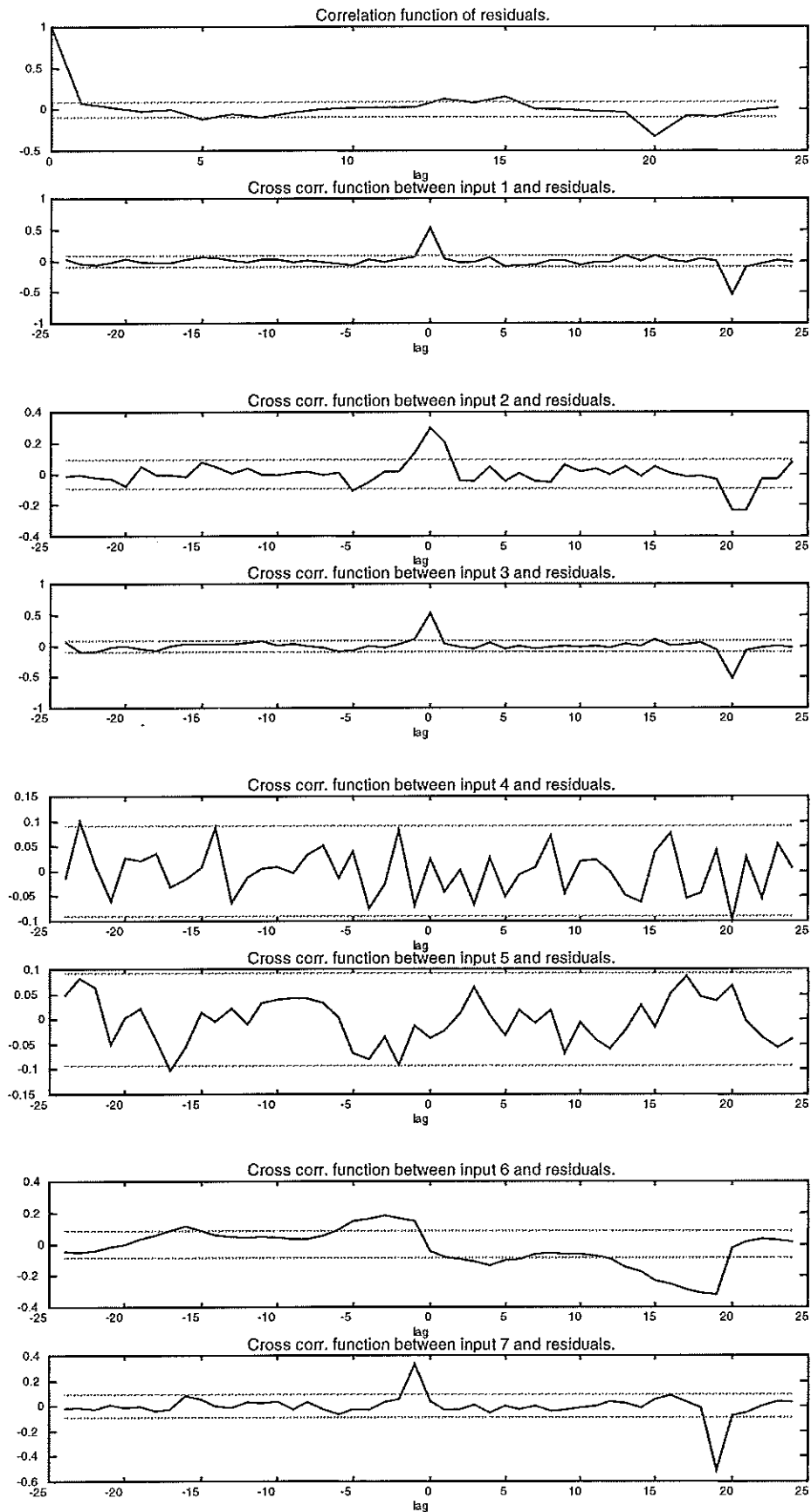
ITERATIV_Open



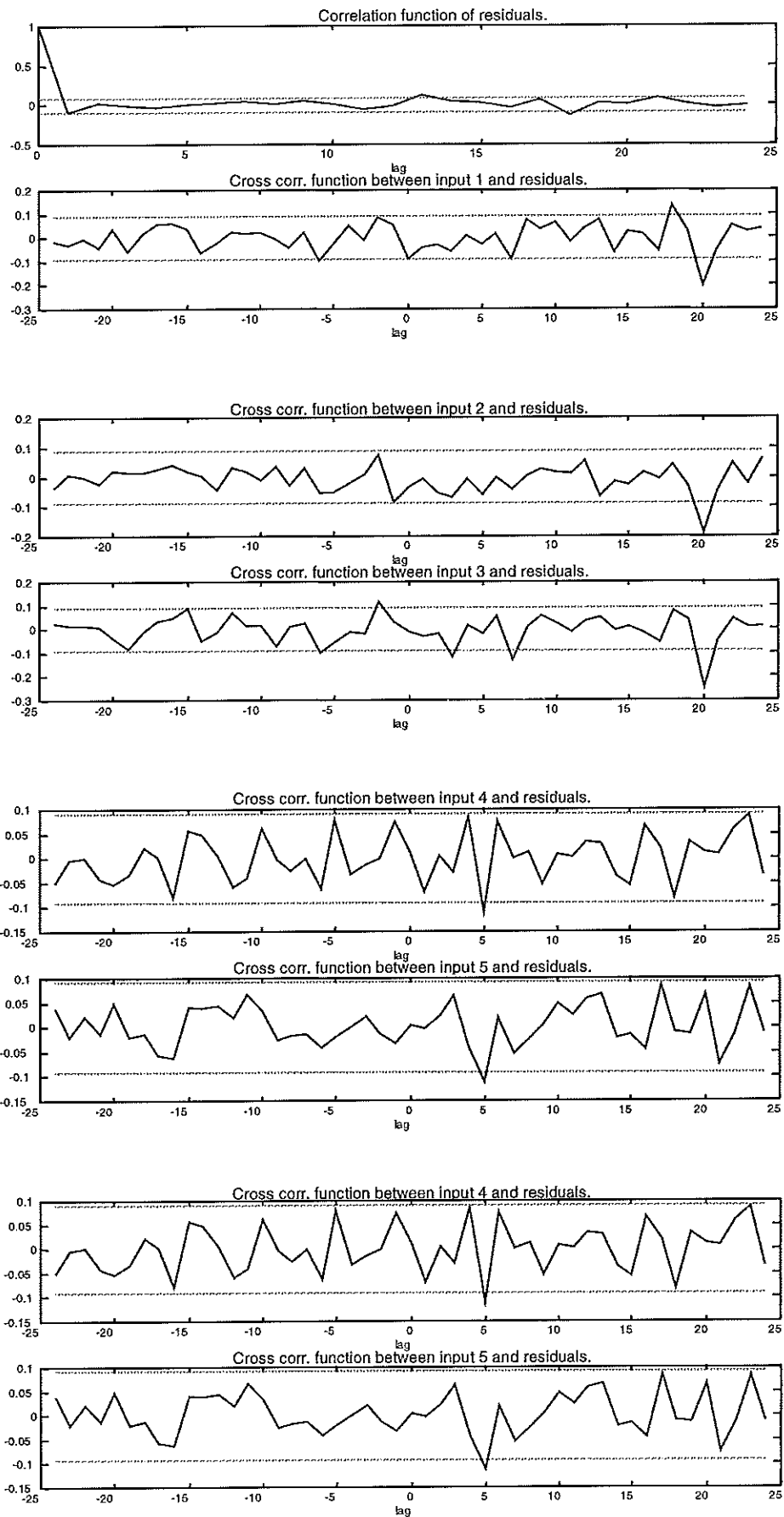
ITERATIV_Close



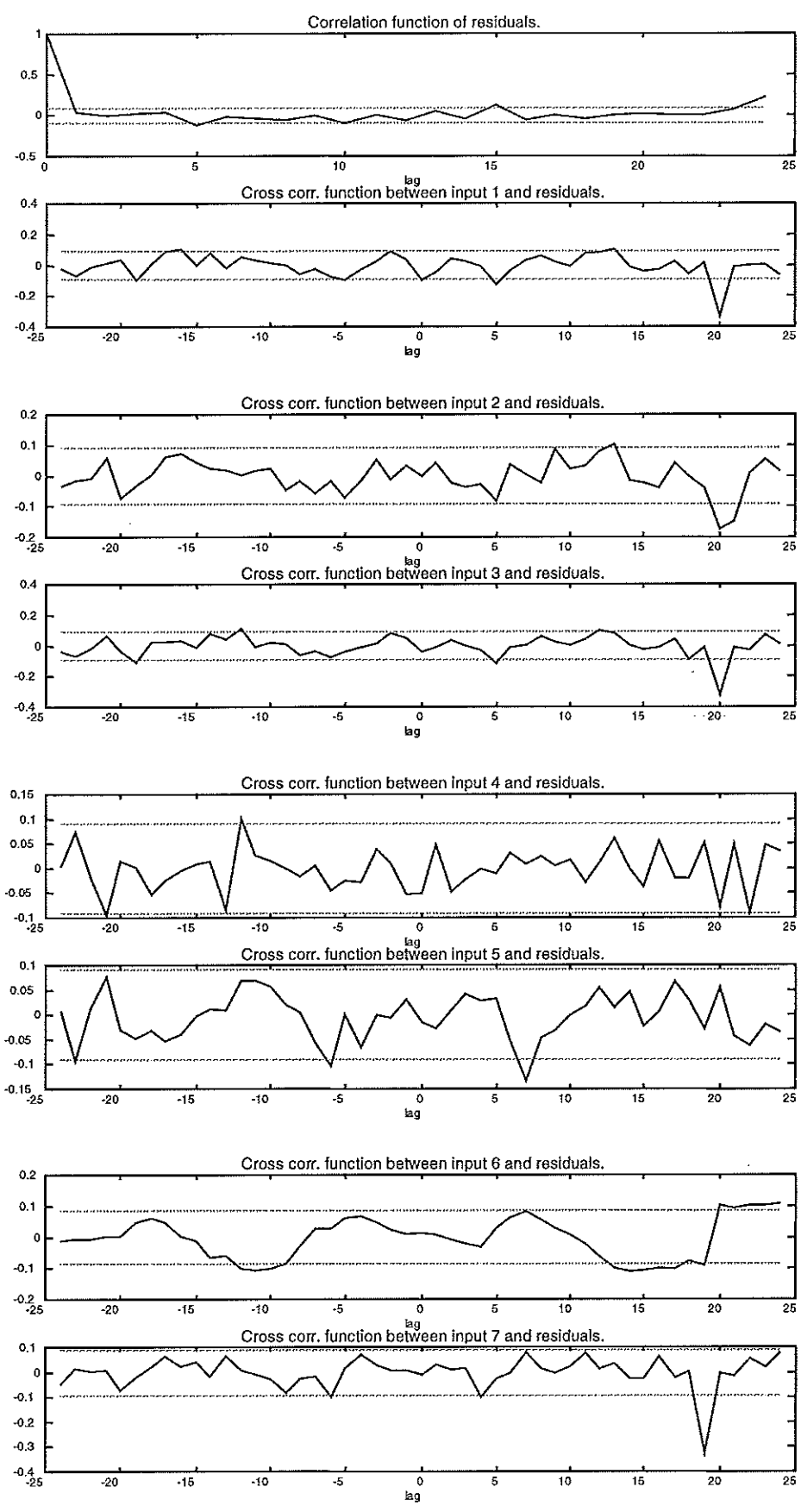
ITERATIV_Direct



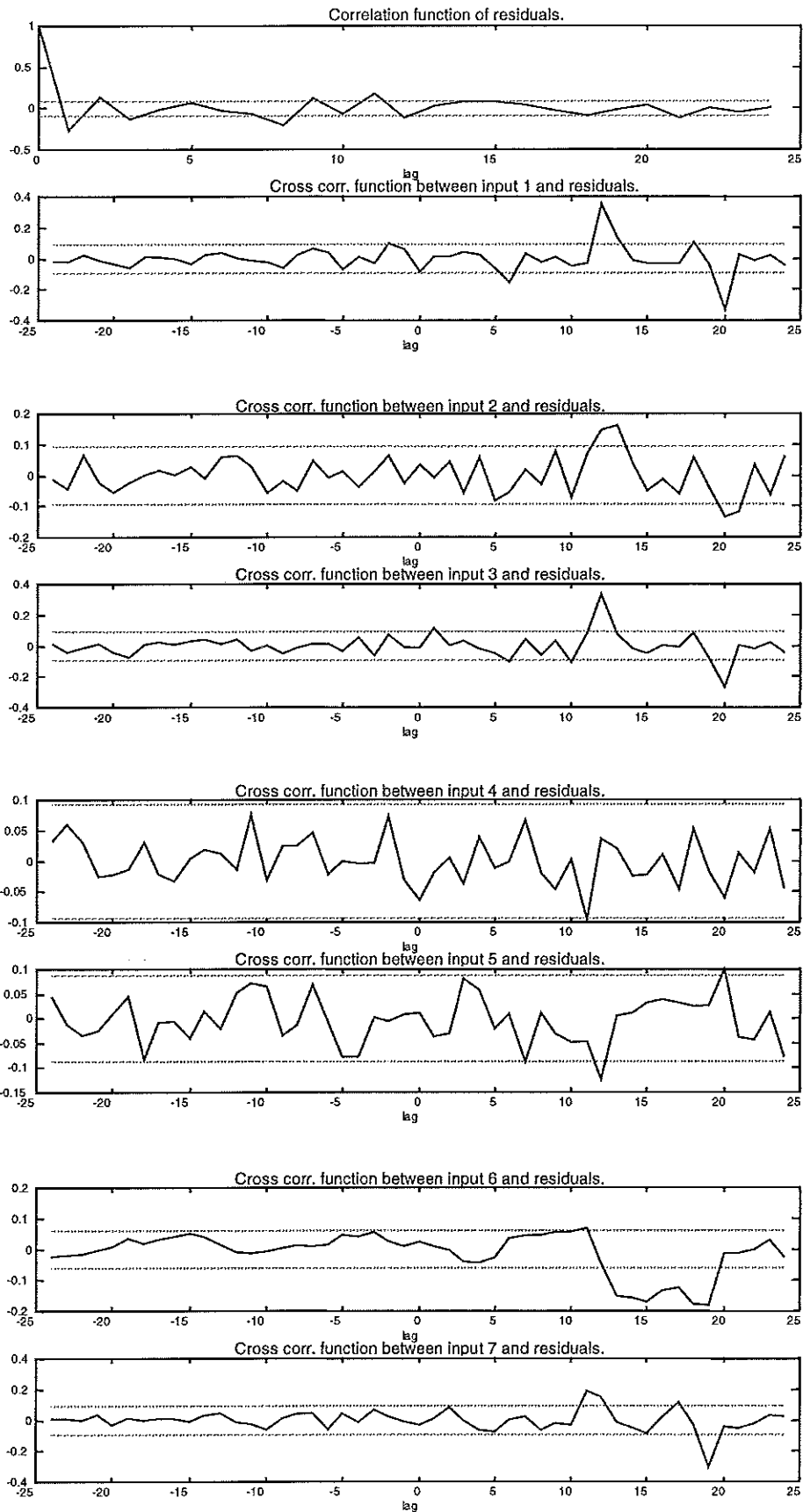
SEGMENT1_Open



SEGMENT1_Close



SEGMENT1_Direct



Reference list

- [1] Brock, W., J. Lakonishok, and B. LeBaron (1992). *Simple Technical Trading Rules and the Stochastic Properties of Stock Returns*. Journal of Finance, pp. 1731-1764.
- [2] Claesson, K. (1987). *Effektiviteten på Stockholms Fondbörs* (Doctoral dissertation). Stockholm:Stockholms Handelshögskola
- [3] Fama, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical work*. Journal of Finance, Vol. 25, pp. 383-417.
- [4] Fama, E. F. (1991). *Efficient Capital Markets II*. Journal of Finance, Vol. 46, pp. 1575-1617.
- [5] Goldfeld, S. M. and Quandt, R. E. *Some tests for homoscedasticity*, Journal of the American Statistical Society, Vol. 60, pp. 539-547.
- [6] Grehn, H.(1999). *Investment Strategies for the Small Investor: Is Technical analysis Workable?*, School of Economics and Commercial Law - Göteborg University (Working Paper)
- [7]<http://domino.omgroup.com/www/omrulesandregulations.nsf/19f5ab73f0a94a2f412567130054bba2/8477262005429a3341256753004f5548?OpenDocument>
- [8]<http://www.moneyworld.co.uk/glossary/gl00188.html>
- [9]http://www.germany-info.org/f_gic/content_business_economy.html
- [10] Johansson, R. (1993). *System Modeling and Identification*
Prentice Hall, Englewood Cliffs, NJ, ISBN 0-13-482308-7
- [11] Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*
Springer-Verlag, Berlin, ISBN 0-387-53194-7
- [12] Van Overschee P., De Moor B., (1994). *N4SID : Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems* Automatica, Special Issue on Statistical Signal Processing and Control, Vol. 30, No. 1, 1994, pp. 75-93.

ⁱ In appendix referenced as Sort.

ⁱⁱ Blue-chip stocks are the companies known nationally for the quality and wide acceptance of their products and services and their ability to make money and pay dividends. These blue-chip stocks are typically traded on the New York Stock Exchange. Some examples would be Coca-Cola, General Motors, and IBM.

ⁱⁱⁱ Short selling is when commodities are borrowed and sold in hope of falling prices, the commodity is then re-bought and returned to the owner. Should the price go up, a loss is made.

^{iv} In appendix referenced as enkät.

^v In appendix referenced as psyk.

^{vi} In appendix A referenced as ARX-Order and ARMAX-Order.

^{vii} In appendix referenced as seg.

^{viii} In appendix referenced as Iterativ.

^{ix} For a full account of H_0 see [5].

