

ISSN 0280-5316  
ISRN LUTFD2/TFRT--5529--SE

# Prediktion av OMX-index med holografiska neurala nätverk

Marcus Nyström

Institutionen för Reglerteknik  
Lunds Tekniska Högskola  
April 1995

<b>Department of Automatic Control</b> <b>Lund Institute of Technology</b> P.O. Box 118 S-221 00 Lund Sweden		<i>Document name</i> MASTER THESIS	
		<i>Date of issue</i> April 1995	
		<i>Document Number</i> ISRN LUTFD2/TFRT--5529--SE	
<i>Author(s)</i> Marcus Nyström		<i>Supervisor</i> Rolf Johansson	
		<i>Sponsoring organisation</i>	
<i>Title and subtitle</i> Prediction of OMX-index with holographic neural networks. (Prediktion av OMX-index med holografiska neurala nätverk).			
<i>Abstract</i> <p>Holographic neural networks are quite different from the more widely used neural networks that are based on backpropagation. The thesis presents the mathematical framework for holographic neural networks. The networks are trained to make predictions of OMX-index (an index for some of the most traded stocks on the stock market in Stockholm). For comparison a backpropagation neural network is also tested with respect to predictions of OMX-index.</p>			
<i>Key words</i> Artificial Neural Networks, Holographic Neural Networks, Stock Market Prediction, Efficient Market Hypothesis, Technical Analysis of Stocks.			
<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> Swedish	<i>Number of pages</i> 49	<i>Recipient's notes</i>	
<i>Security classification</i>			

The report may be ordered from the Department of Automatic Control or borrowed through the University Library 2, Box 1010, S-221 03 Lund, Sweden, Fax +46 46 110019, Telex: 33248 lubbis lund.

## Innehållsförteckning

<b>1 Rapportens uppläggning.....</b>	<b>1</b>
<b>2 Teori för neurala nätverk.....</b>	<b>2</b>
2.1 Bakgrund till neurala nätverk .....	2
2.2 Konventionella neurala nätverk.....	2
2.3 Holografiska neurala nätverk.....	4
2.31 Information i den holografiska modellen.....	4
2.32 Inkodning (inläring).....	6
2.33 Avkodning (svarsavgivning).....	7
<b>3 Allmänt om de finansiella marknaderna .....</b>	<b>10</b>
3.1 Fundamental analys .....	10
3.2 Teknisk analys .....	10
3.3 Två sätt att se på börsen .....	12
<b>4 Prediktion av OMX-index med neurala nätverk .....</b>	<b>14</b>
4.1 Varför använda neurala nätverk?.....	14
4.2 Varför använda holografiska neurala nätverk?.....	14
4.3 Utförda undersökningar .....	15
4.31 Val av variabler.....	16
4.32 Tidshorisont och svarstyp .....	18
4.33 Träning av nätverket.....	23
4.34 Konventionellt nätverk .....	24
4.4 Utvärdering .....	26
4.5 Begränsningar i undersökningar och utvärderingar.....	29
4.6 Sammanfattning och slutsatser .....	29
<b>5 Referenser .....</b>	<b>31</b>
<b>Bilaga 1 .....</b>	<b>33</b>
Utförlig beskrivning av den holografiska metoden .....	33
Förbättrad inkodning .....	33
Illustration av metodens karakteristika .....	35
Vikten av symmetri.....	37
Dynamiskt minne.....	39
Högre ordningens system.....	40
<b>Bilaga 2 .....</b>	<b>43</b>
Undersökning av HNeT Discovery Package .....	43
HNeT:s parametrar .....	43
Stimulus/response conversion.....	43
Initial Statistics.....	44
Optimization .....	44
Learning Rate.....	44
Memory Profile.....	45
Thermalize .....	45
Resultat .....	45
Stimulus/response conversion.....	45
Initial Statistics.....	46
Learning Rate.....	46
Övriga parametrar .....	47
Brister.....	47

# 1 Rapportens uppläggning

Föreliggande examensarbete består av två delar:

- 1) Undersökning av möjligheten att använda holografiska artificiella neurala nätverk för att prediktera OMX-index.
- 2) Optimering av de användardefinierade parametrar som ingår i den använda programvaran. Programmet tillverkas av AND America Ltd. och heter "HNeT Discovery Package". En kort beskrivning av programmet samt resultatet av gjorda undersökningar av lämpliga inställningar presenteras i bilaga 2. Generalagenten för detta program, NovaCast AB i Ronneby, har vänligen bistått med såväl programvaran som råd i anslutning till arbetet.

Sedan slutet av 1980-talet har mycket resurser lagts ned på forskning kring artificiella neurala nätverk (ANN). Det finns en mängd områden där neurala nätverk redan idag används med framgång. Här kan nämnas processtyrning, mönsterigenkänning, expertsystem, prediktioner osv. De vanligaste ANN-systemen bygger på en teori som kallas för konnektionismen. Fortsättningsvis i rapporten benämnes sådana ANN för "konventionella nätverk". I detta arbete studeras och används ett neuralt nätverk som bygger på en annan princip, nämligen ett holografiskt neuralt nätverk. En undersökning med ett konventionellt nätverk finns dock med som jämförelse.

I kapitel 2 presenteras en allmän och enkel teori och bakgrund till såväl konventionella nätverk som holografiska. All teori för de holografiska nätverken bygger på uppgifter från [4, 5, 6, 7] i referenslistan.

I kapitel 3 presenteras en allmän betraktelse av de finansiella marknaderna. Avsikten är att ge en förståelse av den process som ska predikteras.

Kapitel 4 beskriver arbetet med att ta fram en modell för prediktion av OMX-index med neurala nätverk. Här presenteras även en utvärdering av resultatet av nedlagt arbete.

Bilaga 1 innehåller en fördjupad beskrivning av teorin bakom holografiska neurala nätverk.

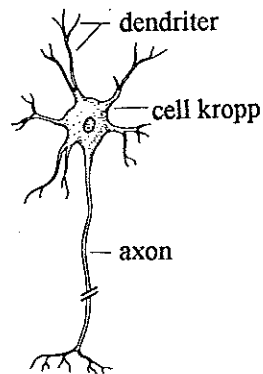
I bilaga 2 presenteras den använda programvaran samt resultatet av optimeringen av användardefinierade inställningar i detta program.

Avslutningsvis vill jag tacka min handledare, Rolf Johansson vid Institutionen för Reglerteknik, för att jag fick genomföra mitt examensarbete inom ett för mig mycket intressant område. Även Rudolf Sillén på NovaCast AB i Ronneby förtjänar ett stort tack för all hjälp med använd programvara.

## 2 Teori för neurala nätverk

### 2.1 Bakgrund till neurala nätverk

Alla kan nog vara överens om att den mänskliga hjärnan är en fantastisk skapelse! Mycket energi har lagts ner inom neurobiologin för att utforska och analysera hur hjärnan fungerar. Klart är att den består av ca  $10^{12}$  neuroner eller nervceller [1]. De är hjärnans elementära beräkningsenheter och kan liknas vid icke-linjära förstärkare. De har alla i stort sett samma struktur (se figur 1), men kan ha många olika uppgifter. Vissa neuron mottar signaler från receptorer och vidarebefordrar dessa till andra neuron medan andra neuron sänder signaler till organ och muskler. Varje neuron kopplar i genomsnitt till  $10^3$  andra neuron via förbindelser som kallas synapser. Kommunikationen via synapser sker med elektrokemiska impulser. Om en synaps aktiveras flitigt växer den i styrka och om det motsatta gäller försvagas den. Denna egenskap kallas för Hebbregeln [2] efter D. Hebb som är en av föregångarna inom ANN-teknik.



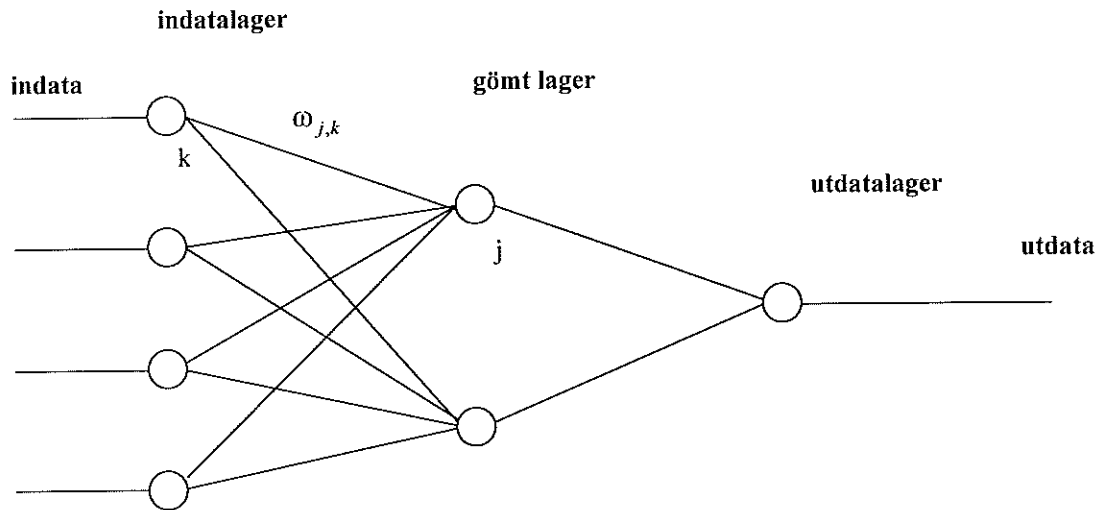
Figur 1 - Skiss av biologiskt neuron. Signaler i form av elektrokemiska impulser tas emot av dendriterna. Neuronen avfyrar en signal som propagerar genom axonen. Denna signal är en starkt olinjär funktion av de inkommande impulserna. Synapsens styrka bestämmer hur stor del av signalen som vidarebefordras till nästa neurons dendrit.

En dator har en signalhastighet som är 1000 gånger större än den mänskliga hjärnans. Trots detta är hjärnan överlägsen när det gäller att känna igen strukturer och göra associationer. Skälet till detta är hjärnans distribuerade struktur och dess förmåga att parallellt och icke-linjärt behandla information. Under 1940-talet och framåt har stora arbetsinsatser lagts ner för att skapa en artificiell motsvarighet till hjärnan som både utnyttjar den starkt framåtskridande datortekniken och principerna bakom den mänskliga hjärnans funktioner för mönsterigenkänning och associationsförmåga. Detta har lett fram till den moderna teorin för artificiella neurala nätverk.

### 2.2 Konventionella neurala nätverk

Ett neuralt nätverk enligt den mest använda modellen [1, 3] består av ett "nät" av ett stort antal neuron, som var och en är enkel vad gäller struktur och tilldelad uppgift. Första lagret består av inenheter och sista lagret av utenheter. Däremellan kan det finnas ett eller flera s.k. gömda lager (se figur 2). Varje neuron förbinder de i framtåtriktningen liggande neuronerna på ett sätt som efterliknar de biologiska neuronens synapser. Synapsernas

styrka motsvaras i modellen av vikter,  $\omega_{j,k}$ , där j står för mottagande neuron och k för avsändande neuron. Inläringen eller träningen går ut på att modifiera dessa vikter. Kunskapen ligger alltså inkapslad i uppsättningen av vikter.



Figur 2 - Ett konventionellt neuralt nätverk

Varje neuron,  $v_i$ , har ett värde eller ett tillstånd som vanligtvis stipuleras att ligga i intervallet  $[-1,1]$  eller  $[0,1]$ . Detta värde bestäms av tillstånden för alla de neuron,  $v_j$ , som via synapser kommunicerar med  $v_i$  (observera att i den här beskrivna modellen sker all kommunikation framåt; det finns andra modeller som tillåter att signaler även skickas bakåt):

$$v_i = F\left(\sum_j \omega_{i,j} v_j + \theta_i\right)$$

där  $F(x)$  kallas tröskelfunktionen. Denna representeras oftast av en sigmoid funktion.

$\theta_i$  är den sk tröskeltermen. Denna motsvaras i biologiska neuron av en membranpotential i mottagande neuron.

Träningen eller inläringen sker genom att låta nätverkets inenheter mottaga en uppsättning indata. Via vikterna (synapserna) genereras ett svar. Med hjälp av detta svar bildas en felfunktion som differensen mellan det riktiga svaret och det av nätverket avgivna. Målet är sedan att minimera denna funktion genom att feltermerna propageras bakåt i nätverket. Vid varje iteration bakåt av feltermerna används en gradientmetod för att uppdatera vikterna.

## 2.3 Holografiska neurala nätverk

Den grundläggande konceptuella skillnaden mellan holografiska neurala nätverk [4, 5, 6, 7] och konventionella ANN är att i den holografiska modellen åsätts varje enskilt neuron stor betydelse, medan det i konventionella ANN är ett stort antal enkla neuron och förbindelserna mellan dessa som är det centrala. Upphovsmannen till den holografiska neurala nätverksmodellen, John G. Sutherland, konstaterade att det inom neurobiologin framkommit resultat som pekade på att varje enskilt neuron i vissa avseenden borde tillmätas större betydelse och modelleras bättre än i konventionella ANN. I den holografiska neurala modellen kallas indata för stimulus och utdata (svar) kallas respons. Mellan varje uppsättning stimulus och respons råder ett visst samband (oftast icke-linjärt och ej helt deterministiskt). Detta samband kallas för avbildningen eller associationen mellan stimulus och respons. Karakteristiskt för metoden är att ett stort antal sådana stimulus-respons associationer kan hanteras av ett enda neuralt element. Detta sker genom att associationerna överförs eller transformeras till en korrelationsmatris. Elementen i denna korrelationsmatris motsvarar synaptiska förbindelser och representeras av komplexa tal. Individuella associationer inkodas i korrelationsmatrisen (eller lärs in) genom en enda icke-iterativ process. På så sätt innehåller korrelationsmatrisen all kunskap som erhållits vid inkodningen. Vid avkodningsprocessen presenteras ett icke tidigare visat stimulus för korrelationsmatrisen vilket resulterar i att en respons avges. Generaliseringsförmågan, dvs förmågan att avge vettiga responser för stimuli som ej inkodats, är härvidlag stor. Detta utgör en viktig skillnad mot andra tänkbara mönsterigenkänningsmetoder som t.ex. bygger på linjesökningsalgoritmer. Beteckningen holografiska neurala nätverk beror bl.a. på att en analogi med digital holografi [21] kan göras vad gäller de ekvationer som styr processerna i det neurala elementet.

### 2.31 Information i den holografiska modellen

I de flesta konventionella ANN-system används de reella talen för att representera information. I den holografiska modellen används i stället komplexa tal:  $\lambda e^{i\theta}$  där  $\lambda$  är absolutbeloppet och  $\theta$  är argumentet. Den ursprungliga uppsättningen av reella stimulus- och responsvärden omvandlas på ett sätt som beskrivs senare till en uppsättning komplexa tal. Argumentet motsvarar härvid värdet av den fysikaliska storheten, dvs argumentet är motsvarigheten till det som finns i stimulus-respons uppsättningen som den ser ut i obehandlad form. Absolutbeloppet representerar ett konfidensvärde. Idéen att använda komplexa tal på detta sätt stöds av karakteristik för intra-cellulär transmission av information i biologiska neuron.

Införandet av ett konfidensvärde ger en ytterligare möjlighet till kontroll av skeendet i neuronerna. Stimuli med ett lågt konfidensvärde har ett litet inflytande på de inkodade stimulus-respons avbildningarna. Genererade responser med lågt konfidensvärde indikerar liten samhörighet med stimuli (dvs svaret är "osäkert").

Stimulus-respons associationer bildas på ett enkelt sätt; de är helt enkelt differenserna mellan argumenten för stimulus och respons. Om t.ex. ett stimuluselement har tilldelats argumentet  $\theta_1$  och tillhörande respons har argumentet  $\theta_2$  så är storheten som representerar avbildningen eller associationen mellan stimulus och respons:

$$\theta_{diff} = \theta_1 - \theta_2 \quad (1)$$

Denna storhet ( $\theta_{diff}$ ) definierar ett kvanta av intelligent association i den holografiska modellen [6]. Modellen möjliggör att ett stort antal sådana kvanta superponeras eller inkapslas i en matris av komplexa tal. Denna matris är den tidigare nämnda korrelationsmatrisen.

Varje stimulusuppsättning består av ett antal reella tal vilket kan beskrivas med skrivsättet  $S\{s_1, s_2, s_3, \dots\}$  där  $s_1, s_2, \dots \in \mathbb{R}$ . Tillhörande respons beskrivs på motsvarande sätt. Dessa uppsättningar måste transformeras till komplexa tal, vilket åstadkoms med sigmoidala funktioner dvs  $\theta_j = F_s(s_j)$  där  $F_s(\cdot)$  är en sigmoidal funktion av nedanstående typ:

$$F_s(x) = 2\pi \left(1 + e^{\kappa \frac{(\mu-x)}{\sigma}}\right)^{-1} \quad (2)$$

$$F(x) = \arctan\left(\kappa \frac{(\mu-x)}{\sigma}\right)$$

där:  $\kappa$  är en proportionalitetsfaktor  
 $\mu$  är väntevärdet för fördelningen av  $x$   
 $\sigma$  är standardavvikelsen för fördelningen av  $x$

Absolutbeloppet sätts oftast till 1 för nya stimulus. Enligt ovanstående åstadkommes följande avbildning för vart och ett av stimulus elementen:

$$s_j \rightarrow \lambda_j e^{i\theta_j} \quad (3)$$

Vi kan nu beskriva stimulus- och responsuppsättningarna på följande allmänna sätt:

$$[S] = \begin{matrix} & \text{element index} \rightarrow \\ \begin{matrix} \downarrow \\ \text{tidsindex} \end{matrix} & \begin{bmatrix} \lambda_{1,t_1} e^{i\theta_{1,t_1}} & \lambda_{2,t_1} e^{i\theta_{2,t_1}} & \lambda_{3,t_1} e^{i\theta_{3,t_1}} & \dots \\ \lambda_{1,t_2} e^{i\theta_{1,t_2}} & \lambda_{2,t_2} e^{i\theta_{2,t_2}} & \lambda_{3,t_2} e^{i\theta_{3,t_2}} & \dots \\ \lambda_{1,t_3} e^{i\theta_{1,t_3}} & \lambda_{2,t_3} e^{i\theta_{2,t_3}} & \lambda_{3,t_3} e^{i\theta_{3,t_3}} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \end{matrix} \quad (4)$$

$$[R] = \begin{matrix} & \text{element index} \rightarrow \\ \begin{matrix} \downarrow \\ \text{tidsindex} \end{matrix} & \begin{bmatrix} \gamma_{1,t_1} e^{i\phi_{1,t_1}} & \gamma_{2,t_1} e^{i\phi_{2,t_1}} & \gamma_{3,t_1} e^{i\phi_{3,t_1}} & \dots \\ \gamma_{1,t_2} e^{i\phi_{1,t_2}} & \gamma_{2,t_2} e^{i\phi_{2,t_2}} & \gamma_{3,t_2} e^{i\phi_{3,t_2}} & \dots \\ \gamma_{1,t_3} e^{i\phi_{1,t_3}} & \gamma_{2,t_3} e^{i\phi_{2,t_3}} & \gamma_{3,t_3} e^{i\phi_{3,t_3}} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \end{matrix} \quad (5)$$



## 2.32 Inkodning (inlärning)

Med inkodning avses inlärning eller träning för att använda begrepp som används inom konventionell ANN-teori. Inlärningsprocessen uppvisar en holografisk effekt i förmågan att inkapsla ett stort antal stimulus-respons associationer i ett och samma medium, nämligen korrelationsmatrisen  $[X]$ . I sin enklaste form, vilken beskrivs här, består inkodningen av en komplexvärd matrismultiplikation enligt följande:

$$[X] = [\overline{S}]^T \cdot [R] \quad (6)$$

där  $[\overline{S}]^T$  är det komplexkonjugerade transponatet till  $[S]$ .

Med ovanstående operation inkapslas alla stimulus-respons associationer i korrelationsmatrisen  $[X]$ . För att åskådliggöra denna operation antar vi en respons per stimulusuppsättning ( $[R]$  är en kolonnvektor).  $[X]$  innehåller då element enligt nedan:

$$[X] = \begin{bmatrix} \sum_{T=t_1}^{t_N} \lambda_{1,T} \gamma_T e^{i(\phi_T - \theta_{1,T})} \\ \sum_{T=t_1}^{t_N} \lambda_{2,T} \gamma_T e^{i(\phi_T - \theta_{2,T})} \\ \dots \\ \dots \end{bmatrix} \quad (7)$$

Denna inkodningsprocess inkapslar eller kollapsar tidsparametern samtidigt som informationsinnehållet bevaras i den mening att om korrelationsmatrisen utsätts för ett inkodat stimulus så avges den associerade responsen (se vidare under rubriken "Avkodning"). Absolutbeloppen för elementen i korrelationsmatrisen motsvarar förklaringsvärdet eller betydelsen av motsvarande stimulus. Detta är en viktig aspekt som kommer att användas flitigt under det praktiska arbetet för att skapa en modell för prediktion av OMX-index (se vidare under rubriken "Prediktion av OMX-index"). Ovanstående utgör inkodningsmekanismen i dess mest elementära form. Det finns en mängd möjliga metoder att förstärka och förbättra den här beskrivna metoden. I bilaga 2 diskuteras hur det går att åstadkomma en inlärning som är beroende av tidigare inkodad information, vilket inte är fallet för den enkla form som beskrivits ovan. Vidare visas hur åtskilliga andra önskvärda karakteristika åstadkommes, som t.ex. möjlighet till kontroll av generaliseringsförmågan, ökade kapaciteter för lagring av information, stabilitet i neuronerna för en större mängd indatafördelningar och en högre noggrannhet i avgivna responser.

## 2.33 Avkodning (svarsavgivning)

Avkodning eller svarsavgivning innebär att ett stimulusfält transformeras till en associerad respons under påverkan av alla i korrelationsmatrisen inkapslade stimulus-respons avbildningar. Responsen avges i form av ett komplext tal, där som tidigare nämnts absolutbeloppet representerar graden av samstämmighet med tidigare inkodade stimulus-respons associationer och argumentet motsvarar det reella värdet på det önskade svaret. Avkodningstransformationen definieras som följande inre produkt:

$$[R] = \frac{1}{c} [S]^* \cdot [X] \quad (8)$$

Här är  $[S]^*$  den nya stimulusuppsättningen för vilken den associerade responsen ska framkallas:

$$[S]^* = [\lambda_1^* e^{i\theta_1}, \lambda_2^* e^{i\theta_2}, \dots] \quad (9)$$

Normaliseringskonstanten ( $c$ ) i (8) definieras som:

$$c = \sum_{j=1}^M \lambda_j^* \quad (10)$$

Med denna definition av ( $c$ ) erhålles ett absolutbelopp mellan 0 och 1, vilket är vad vi önskar.

Responsefelet (skillnaden mellan sökt respons och nätverkets respons) kan undersökas genom att inkoda helt slumpmässiga (okorrelerade) stimulus-respons associationer och sedan "testa" nätverket med ett sådant inkodat stimulus. Det noteras då att felet i avgiven respons ökar med antalet inkodade associationer. Felet är dock tillräckligt lågt för att en stor mängd stimulus-respons associationer ska kunna inkodas utan att felet får för stor menlig inverkan på resultatet.

Modellens förmåga att inkapsla information kan illustreras genom att titta närmare på ekvation (8). Då observeras att (8) representerar en summa av komplexa tal, där varje tal representerar en stimulustransformation under påverkan av en av de stimulus-respons associationer som tidigare inkodats. Genom att kombinera (7), (8) och (9) fås (under antagande av endast ett responsvärde för varje stimulusuppsättning):

$$r = \frac{1}{c} \sum_{j=1}^M \lambda_j^* e^{i\theta_j} \sum_{T=t_1}^{t_N} \lambda_{j,T} \gamma_T e^{i(\phi_T - \theta_{j,T})} \quad (11)$$

Detta kan skrivas om som:

$$r = \frac{1}{c} \sum_{T=t_1}^{t_N} \gamma_T e^{i\phi_T} \sum_{j=1}^M \lambda_j^* \lambda_{j,T} e^{i(\theta_j^* - \theta_{j,T})} \quad (12)$$

Ovanstående beskriver en sekvens av fristående komplexa tal summerade över tiden ( $T = t_1 \rightarrow t_N$ ). Vart och ett av dessa komplexa tal motsvarar en responskomponent, som har tillkommit genom en transformation av en

ny stimulusuppsättning under påverkan av en vid någon tid  $t$  inkodad association. Med detta i åtanke kan (12) skrivas på en mer kompakt form enligt nedanstående:

$$r = \frac{1}{c} \left[ \Lambda_{t_1} e^{i\phi_{t_1}^*} + \Lambda_{t_2} e^{i\phi_{t_2}^*} + \dots + \Lambda_{t_N} e^{i\phi_{t_N}^*} \right] \quad (13)$$

där  $\Lambda_j$  är konfidensvärdet för den del av responsen som motsvarar associationen inkodad vid tiden  $t$ . Genom identifiering av termer i (12) och (13) fås:

$$\Lambda_{t_1} e^{i\phi_{t_1}^*} = \gamma_{t_1} e^{i\phi_{t_1}} \sum_{j=1}^M \lambda_j^* \lambda_{j,t_1} e^{i(\theta_j^* - \theta_{j,t_1})} \quad (14)$$

Ur (14) följer direkt absolutbelopp och argument för termerna:

$$\Lambda_{t_1} = \gamma_{t_1} \left[ \left[ \sum_{j=1}^M \lambda_j^* \lambda_{j,t_1} \cos(\theta_j^* - \theta_{j,t_1}) \right]^2 + \left[ \sum_{j=1}^M \lambda_j^* \lambda_{j,t_1} \sin(\theta_j^* - \theta_{j,t_1}) \right]^2 \right]^{\frac{1}{2}} \quad (15)$$

$$\phi_{t_1}^* = \tan^{-1} \left[ \frac{\sum_{j=1}^M \lambda_j^* \lambda_{j,t_1} \sin(\theta_j^* - \theta_{j,t_1} + \phi_{t_1})}{\sum_{j=1}^M \lambda_j^* \lambda_{j,t_1} \cos(\theta_j^* - \theta_{j,t_1} + \phi_{t_1})} \right] \quad (16)$$

Varje term i HL av (13) har ett absolutbelopp eller konfidensvärde som avspeglar graden av överensstämmelse med stimulus-respons associationen inkodad vid en viss tidpunkt som vi kan kalla för  $t_1$ . Med andra ord så dominerar den av de tidigare inkodade associationerna som uppvisar störst likhet med den nya

stimulusuppsättningen  $[S]^*$ . Övriga termer i (13) (dvs de som motsvarar associationer inkodade vid  $T \neq t_1$ ) summeras ihop och lämnar ett bidrag till responsen. Kontentan blir att den slutliga responsen innehåller en felkomponent motsvarande summan av de icke associerade termerna i (13). Denna (vektoriella) summa utgör emellertid i stort sett en s.k. random walk i det komplexa talplanet under förutsättning att de komplexa talen har fördelats symmetriskt i det komplexa talplanet. Det är känt att absolutbeloppet efter  $N$  steg i en random walk blir  $\sqrt{N}$ . Härigenom erbjuds en metod att göra en approximativ beräkning av medelabsolutbeloppet för de icke-associerade responstermerna. De icke-associerade responstermerna i (13) får då ett medelbelopp enligt:

$$\left| r_{T \neq t_1} \right| = \left| \frac{1}{c} \sum_{T=t_2}^N \Lambda_T e^{i\phi^* T} \right| \approx \sqrt{\frac{N-1}{M}} \quad (17)$$

där:  $N$  är antalet inkodade stimulus-respons associationer

$M$  är antalet synapser in till cellen (dvs antalet kolonner i  $[S]$ )

$\gamma_t, \lambda_{j,t} \rightarrow 1$  för alla  $j$  och  $t$ .

På ett liknande sätt, vilket inte visas här, kan en övre gräns för felet i responsernas argument bestämmas som:

$$\phi_{error} \approx \frac{1}{\pi \sqrt{8}} \tan^{-1} \left( \sqrt{\frac{N}{M}} \right) \quad (18)$$

För att komma fram till detta uttryck har det antagits att stimulus-responsuppsättningarna är slumpmässigt genererade (dvs okorrelerade). Detta representerar så att säga "värsta fallet".

För att studera innebörden av (15) och (16) förs ett enkelt gränsvärdesresonemang enligt följande: låt elementen i  $[S]^*$  gå mot stimulusmönstret som inkodats vid tiden  $t_1$ . För att åstadkomma detta införs ett slumpmässigt fel  $\epsilon_r$  och följande samband mellan elementen i  $[S]^*$  och stimulusuppsättningen från tiden  $t_1$  etableras:

$$e^{i\theta^* j} = e^{i(\theta_{j,t_1} + \epsilon_r)} \quad j = 1, 2, \dots, N \quad (19)$$

Sätt in (19) i (15) och (16) och låt  $\epsilon_r \rightarrow 0$ . Då erhålles:

$$\Lambda_{t_1} \rightarrow \gamma_{t_1} \frac{\sum_{j=1}^N \lambda_j^* \lambda_{j,t_1}}{\sum_{j=1}^N \lambda_j^*} \quad (20)$$

Med konfidensvärdena satta till 1 erhålles:

$$\Lambda_{t_1} \rightarrow 1 \quad (21)$$

$$\phi_{t_1}^* \rightarrow \phi_{t_1} \quad (22)$$

Som tidigare nämnts är detta den enklaste modellen för holografiska neurala nätverk. Ett antal förstärkningar och förbättringar presenteras i bilaga 1.

## 3 Allmänt om de finansiella marknaderna

### 3.1 Fundamental analys

Vid fundamental analys studeras de faktorer som påverkar utbud och efterfrågan, som i sin tur styr priset. T.ex. så är efterfrågan på aktier en funktion av bl.a. kapitaltillgången (likviditeten), framtidsutsikter för ekonomin och ränteläget. Utbudet av t.ex. kaffe och andra naturtillgångar som handlas på råvarubörserna relateras till mängden som produceras, vilket i sin tur bestäms av t.ex. väderleksförhållanden. Efter att ha analyserat en mängd sådana här faktorer görs en prediktion av den framtida prisutvecklingen. Sammanfattningsvis kan man säga att fundamental analys går ut på att studera orsakerna till och effekterna av förändringar av utbud och efterfrågan och på grundval av detta göra en prediktion. Fundamental analys används oftast vid prognoser över utvecklingen på längre sikt.

### 3.2 Teknisk analys

Teknisk analys [8, 9] innebär studier av den historiska prisutvecklingen och omsättningen (volymen) för att därigenom kunna göra förutsägelser om framtiden. Den "tekniske analytikern" tittar på mönster i marknadens beteende och väntar sedan på att liknande mönster ska infinna sig. Den tekniska analysen utgår i från att prisförändringar inte enbart är en manifestation av ekonomiska data och nyheter, utan dessutom speglar

psykologin eller "humöret" hos de människor som handlar på marknaden. Ofta används den tekniska analysen för prognoser över den i tiden närmaste utvecklingen, dvs den används för att hitta rätt "timing".

Pris och volym studeras på en mängd olika sätt. En vanlig metod är att plotta priset (och/eller volymen) mot tiden. Ett sådant diagram kallas för ett barchart. I chartet identifieras och plottas trendlinjer, stödnivåer och motståndsnivåer. Vanligt är också att prisutvecklingen transformeras på något sätt för att bilda så kallade tekniska indikatorer. Dessa indikatorer indelas traditionellt i två grupper: oscillatorer och trendföljande indikatorer. Oscillatorer används för att hitta översålda och överköpta lägen (dvs tillfällen då priserna rört sig omotiverat mycket på för kort tid) och trendföljande indikatorer används som namnet antyder till att hitta begynnande trender. Gemensamt för alla indikatorer är att transformeringen ofta på något ställe innebär bildandet av glidande medeltal. Ett sådant kan vara av tre slag: enkelt, viktat och exponentiellt. Dessa definieras enligt nedanstående.

$P[0]$  = det aktuella priset, senaste betalkurs

$P[n]$  = priset  $n$  perioder tillbaka (med period avses vanligen en dag)

$$\text{Enkelt glidande medeltal} = \frac{\sum_{t=0}^{n-1} P[t]}{n} \quad (23)$$

$$\text{Viktat glidande medeltal} = \frac{\sum_{t=0}^{n-1} P[t] \cdot (n-t)}{\sum_{t=0}^{n-1} (n-t)} \quad (24)$$

$$\text{Exponentiellt glidande medeltal (EMA)} = P[0] \cdot \kappa + EMA[1] \cdot (1 - \kappa) \quad \text{där } \kappa = \frac{2}{(n+1)} \quad (25)$$

Som illustration definieras och förklaras två kända indikatorer nedan. Följande förberedande definitioner görs:

högst = högsta betalkurs under dagen

lägst = lägsta betalkurs under dagen

$nhögst(n) = \max(högst[0], högst[1], \dots, högst[n-1])$

$nlägst(n) = \min(lägst[0], lägst[1], \dots, lägst[n-1])$

MACD är en indikator som fungerar bäst i marknader som "trender" i endera riktning. Den består av 3 stycken glidande medeltal och definieras enligt följande:

- 1) Beräkna ett 12-dagars och ett 26-dagars exponentiellt glidande medeltal av senaste betalkurserna.
- 2) Beräkna differensen av 12-dagars och 26-dagars medeltalen. Resultatet kallas MACD-kurvan.
- 3) Beräkna ett 9-dagars exponentiellt glidande medeltal av MACD-kurvan. Detta ger den så kallade

signalkurvan.

4) Plotta MACD-kurvan och signalkurvan i samma diagram.

Indikatorn tolkas vanligen genom att definiera köpsignaler som tillfällen då MACD-kurvan bryter signalkurvan underifrån och säljsignaler som tillfällen då MACD-kurvan bryter signalkurvan ovanifrån.

Stochastic Oscillator använder sig förutom av senaste betalkurser även av högsta och lägsta noteringar. Indikatorn bygger på det faktum att marknader i uppåtgående ofta stänger nära högsta noteringen och fallande marknader nära lägsta noteringen. Indikatorn definieras enligt följande:

1) Beräkna %K-kurvan enligt  $\%K = 100 \frac{P[0] - n\text{lägst}[n]}{n\text{högst}[n] - n\text{lägst}[n]}$  ; vanligtvis används  $n=5$  (26)

2) Beräkna %D-kurvan. Denna beräknas genom att dividera summan av de tre senaste dagarnas värde på täljaren i (26) med summan av de tre senaste dagarnas värde på nämnaren i (26).

3) Plotta %K-kurvan och %D-kurvan i samma diagram.

Det finns många olika sätt att tolka denna indikator. Vanligt är att köpsignaler identifieras då %K-kurvan bryter %D-kurvan underifrån (eventuellt med tillägget att båda ska ha värden under 30%). En säljsignal identifieras då %K-kurvan bryter %D-kurvan ovanifrån (eventuellt med tillägget att båda ska ha värden över 70%). Indikatorn anses fungera bäst för marknader som oscillerar, dvs som inte "trender".

### 3.3 Två sätt att se på börsen

En förutsättning för att kunna arbeta med prediktioner med hjälp av neurala nätverk är att man har en god känsla för de karakteristika som beskriver den process man vill prediktera. Nedan anlägges två olika perspektiv på börsen; dels ett med utgångspunkt i de nationalekonomiska teorierna för de finansiella marknaderna och dels ett mer "ingenjörsmässigt", där börsen betraktas som ett icke-linjärt dynamiskt system.

Ett dominerande koncept i ekonomernas akademiska litteratur om de finansiella marknaderna [10, 11] har alltsedan 1960-talet varit teorin om den effektiva marknaden. Grovt sett säger denna teori att all tillgänglig information ska avspeglas (eller annorlunda uttryckt vara diskonterad) i priserna som sätts på de finansiella marknaderna. Detta är ett väldigt strängt och vittgående krav, varför man numera talar om tre grader eller styrkor av effektiv marknad:

*Svag form av effektiv marknad.* Kravet här är att all information som finns i den historiska prisutvecklingen ska vara diskonterad.

*Semi-stark form av effektiv marknad.* För att denna form av effektiv marknad ska anses råda ska all offentlig information vara diskonterad i priserna.

*Stark form av effektiv marknad.* Här gäller det att all information, dvs även icke-offentlig ska vara diskonterad i priserna.

Redan den svaga formen innebär att den tekniska analysen helt kan förkastas. Teorin innebär att priserna antas följa en random walk-modell<sup>1</sup> med stegvisa förändringar när ny information tillförs marknaden.

Det går att finna fenomen som talar mot den effektiva marknadsteorin, även om den förhärskande uppfattningen bland forskare i finansiell ekonomi är att teorin är riktig och oomtvistelig. Ett sådant är börskraschen 1987. Måndagen den 19 oktober gick Dow Jones-index ned med 23 %. Kraschen spred sig över hela världen med kursfall på världens alla börser som följd. Detta hände trots att ingen ny information hade tillförts marknaderna. Kritiker av effektiva marknadsteorin menar att om priserna dagen innan kraschen speglade all tillgänglig fundamental information, så borde inte index kunna falla med 23 % nästkommande dag. Generellt innehåller fördelningen för prisförändringar på de finansiella marknaderna alldeles för många stora värden än vad som borde vara fallet för en i stort sett Gaussisk fördelning enligt effektiva marknadsmodellen ( ”svansarna” är för stora). Kritikerna hävdar att det grundläggande felet är att effektiva marknadsmodellen utgår från att de finansiella marknaderna befinner sig i ett jämviktstillstånd vid avsaknad av påverkan från omvärlden (dvs ny ekonomisk data eller andra nyheter) [19]. Reaktionen på sådan påverkan från omvärlden sker enligt modellen omedelbart och på ett linjärt sätt. Det innebär att ingen ackumulering ledande till en fördröjd reaktion på en serie av händelser finns. Varje nyhet utövar sin påverkan direkt och är sedan ”förbrukad”. Detta antagande om jämvikt är fel enligt kritikerna. System i naturen befinner sig ofta i ett tillstånd av icke-jämvikt; en djurart i naturen måste för att överleva hela tiden anpassa sig och genomgå en evolution. Månen är ett system i perfekt jämvikt. Månen är en död planet. Samma förhållande råder i ekonomin. Försök att stabilisera eller tvinga fram jämvikt har misslyckats. Jämviktsmodellen innehåller inga faktorer som beskriver känslomässiga faktorer som girighet och fruktan. Det är dessa faktorer som ligger bakom den förmåga till utveckling och anpassning till nya förhållande som finns i en fri marknadsekonomi.

Om ett mer ingenjörsmässigt perspektiv anlägges vid studiet av börserna och tidsserierna som representerar prisförändringarna över tiden så ter det sig naturligt att betrakta börserna som en synnerligen icke-linjär och dynamisk process. Edgar E. Peters har studerat denna aspekt. I sin bok ”Chaos and Order in The Financial Markets” [19] kommer han fram till bl.a. följande för de finansiella marknaderna:

- 1) Tidsserierna som representerar prisförändringarna är fraktala. Sådana tidsserier karakteriseras av långtidsminne, cykler och trender och uppstår vid icke-linjära dynamiska system eller deterministiskt kaos.
- 2) Det finns en känslighet för begynnelsevärden (Lyapunovexponenten är positiv [20]).

En analogi med ett problem som påminner om prediktion av de finansiella marknaderna, nämligen förutsägelser av vädret, kan göras. I början av 1940-talet var bristen på datorkraft den begränsande faktorn vid prediktioner av vädret. Vädersystemen rörde sig helt enkelt fortare än vad datorerna räknade. Efterhand som datortekniken gick framåt erhöles den nödvändiga datorkraften. Den snabbaste datorn för prediktion av vädret finns idag i Redding, England (en Cray YMK, 800 Mflops) [14]. Med denna kan vädret för hela norra hemisfären en vecka framåt förutsägas efter bara en timmes CPU-tid. Efter 3-4 dagar blir emellertid felet i prognosen ohållbart stora. CPU-tid är alltså inte den begränsande faktorn som forskare trodde förr i tiden. Det är någonting annat i den icke-linjära strukturen hos problemet som sätter gränsen för hur lång tid framåt prediktioner kan sträcka sig.

Ovanstående resonemang vad gäller tidshorisonten för prediktioner belyses i detta arbete genom en undersökning av vilka tidshorisonten som ger bäst prediktion med det holografiska neurala nätverket som prediktionsverktyg. Nämnas kan också att tekniska indikatorer som bygger på kaosteori används i arbetet.

Avslutningsvis kan det konstateras att åtskilliga av de som utvecklat och empiriskt ”bevisat” effektiva marknadsmodellen blivit både erkända och berömda, men ingen av dem har blivit rik

---

<sup>1</sup> Vanligen definieras en random-walk modell med konstant väntevärde och varians. Detta antages dock inte i här.



## 4 Prediktion av OMX-index med neurala nätverk

Under 1980-talet tog forskningen inom Artificiell Intelligens fart. Ett flertal s.k. expertsystem utvecklades och testades med avseende på förmåga att bistå vid prediktioner av olika finansiella priser. Expertsystemen krävde dock som namnet antyder att ett antal "experter" tömdes på sin kunskap (systemen var alltså deduktiva till sin struktur). Särskilt inom ett område som "trading" kan det vara svårt att sätta fingret på och i ord förklara vilka förhållanden som i olika situationer leder fram till att en viss prediktion görs. Användningen av expertsystem inom det finansiella området beskrivs i artikeln "Artificial Intelligence and Market Analysis", *Technical Analysis of Stocks and Commodities*, mars 1991 [22].

Sedan slutet av 1980-talet har mycket forskning lagts ned på att undersöka och utveckla neurala nätverk för användning vid prediktioner. Det finns ett antal undersökningar [13, 14, 15, 16, 17] som visar att tradingsystem baserade på prediktioner gjorda av neurala nätverk presterar mycket bra resultat. Vissa uppgifter [17] gör gällande att upp emot 10 miljarder \$ redan i dag står under direkt kontroll av trading-system baserade på neurala nätverk (Fidelity Investment är exempel på en stor institution som i stor utsträckning använder sig av neurala nätverk i sin kapitalförvaltning. Här används dock nätverken till att välja ut enskilda aktier som bedöms ge bättre avkastning än motsvarande index). Nedan listas några skäl till varför neurala nätverk lämpar sig väl för prediktionsproblemet och varför holografiska nätverk valts i föreliggande arbete.

### 4.1 Varför använda neurala nätverk?

- Neurala nätverk finner och lär sig kraftigt icke-linjära och komplexa samband.
- Neurala nätverk extraherar själv den information och den kunskap som behövs (de utgör ett induktivt system). Detta skiljer de neurala nätverken från de förut använda expertsystemen.
- Nätverken kan effektivt hantera en stor mängd variabler, vilket är nödvändigt vid dessa komplexa problem.

### 4.2 Varför använda holografiska neurala nätverk?

- Som tidigare nämnts har det vid ett flertal tillfällen gjorts undersökningar av värdet av neurala nätverk vid prediktioner av utvecklingen på de finansiella marknaderna. Dessa har alla byggt på konventionella back-propagationnätverk. Det är nu dags att testa den holografiska metoden på detta svåra och komplexa problem.
- Inlärningsproceduren (inkodningen) och svarsavgivningen (avkodningen) tar mycket mindre tid med det holografiska nätverket. Det holografiska nätverket inkodar stora uppsättningar data på 10-30 minuter. Med ett backpropagationnätverk kan motsvarande procedur ta allt från flera timmar upp till flera dygn beroende på datamängdernas storlek.

- Det holografiska nätverket är lättare att använda. Vid användning av konventionella nätverk måste man empiriskt undersöka och bestämma det optimala värdet av en mängd storheter (t.ex. antal gömda lager, antal neuron i varje lager, överföringsfunktion, normalisering av data osv). Med tanke på de långa träningsstider som krävs innebär detta en klar begränsning.
- Konventionella nätverk är känsliga för överträning (generaliseringsförmågan ersätts av ren memorisering vid långvariga träningskörningar). De har också en tendens att fastna i lokala minima vid minimeringen av felfunktionen. Det holografiska nätverket uppvisar inga sådana egenskaper. Ofta används genetiska algoritmer vid minimeringen av felfunktionen i konventionella nätverk. Då holografiska nätverk används kan denna extra resurs i stället användas för att hitta den optimala uppsättningen variabler.

## 4.3 Utförda undersökningar

Allmänt kan man säga att det finns en i stort sett oändlig mängd olika alternativ vid utformningen av ett system för prediktion av OMX-index med ett neuralt nätverk. Några viktiga frågeställningar som måste beaktas är följande:

- Vilka variabler ska användas som indata vid träning av nätverket? Här finns en mängd olika alternativ: historisk prisutveckling, fundamentala indikatorer, aktuell tidsperiod (möjligen finns det mönster på börserna med avseende på veckodag eller månad). Det finns också gott om exempel på mer "exotiska" variabler som eventuellt kan användas. (Det finns bl.a. exempel på neurala nätverk som tränats med bredden på den amerikanske centralbankschefens leende vid hans morgonframträdande som indata.) Det gäller också att bestämma det optimala antalet variabler. För många oviktiga variabler tenderar att öka brusnivån i nätverket.
- Hur ska indata förbehandlas för att ge bästa resultat? Om t.ex. den historiska prisutvecklingen väljes som indata blir resultatet ofta useffektivt om priserna för de 10 senaste dagarna skickas in i nätverket och 11:e dagens pris är det önskade svaret. Bättre resultat erhålles om indata förbehandlas genom att t.ex. bilda tekniska indikatorer enligt tidigare avsnitt. Andra tänkbara variabler som kan bildas är: högsta/lägsta värdet de senaste 20 dagarna, antal dagar sedan detta värde erhöles, koefficienten för en regressionslinje över prisdata osv. Problemet kan illustreras genom att betrakta den yta som spänns upp i ett N-dimensionellt rum av de N variablerna i indata genom funktionen som avbildar variablerna på det önskade svaret. Det gäller nu att presentera denna yta på ett sätt så att nätverket kan se strukturen hos den (t ex. extrempunkter).
- Vilken tidshorisont ska nätverkets prediktioner sträcka sig över? Det gäller att empiriskt undersöka vilken tidshorisont som ger bäst prediktionsresultat.
- Vilket svar ska avkrävas av nätverket? Olika tänkbara alternativ är dels exakta svar i form av absoluta tal ("OMX-index i 1123 om 5 dagar"), och dels mindre pretentiösa svar av typen "upp/ned på 5 dagars sikt".

### 4.31 Val av variabler

Ett första problem av mer administrativ karaktär som måste lösas är införskaffandet av data. Det gäller här först och främst att hitta storheter som bedöms kunna ha ett prediktivt värde. Jag har skaffat tidsserier för OMX-index, totala omsättningen på börsen (volymen), Dow Jones-index, femåriga räntan och USD/SEK. Tidsserierna, som för OMX-index består av de dagliga värdena på sista, högsta och lägsta notering och för övriga serier av sista notering, sträcker sig bakåt i tiden till 1987. När väl tidsserierna är kompletta gäller det att med dessa som grund finna variabler som lämpar sig som indata till nätverket. Arbetet med att hitta rätt variabler är utan tvekan ett av de viktigaste och samtidigt mest arbetskrävande momenten. Det naturliga är att börja med att ta fram variabler ur obehandlad prisdata. T.ex. kan nätverket få de tre senaste värdena ur respektive tidsserie som indata. Oavsett vilket svar som nätverket avkrävs så fås i regel ett dåligt resultat. Däremot kan viss information om respektive tidsseries betydelse för prediktion av OMX-index erhållas. I diagram 1 visas vilken vikt nätverket har åsatt respektive variabel vid prediktion av upp- eller nedgång på 2, 5 och 10 dagars sikt. Två tekniska indikatorer av OMX-index har tagits med som jämförelse.

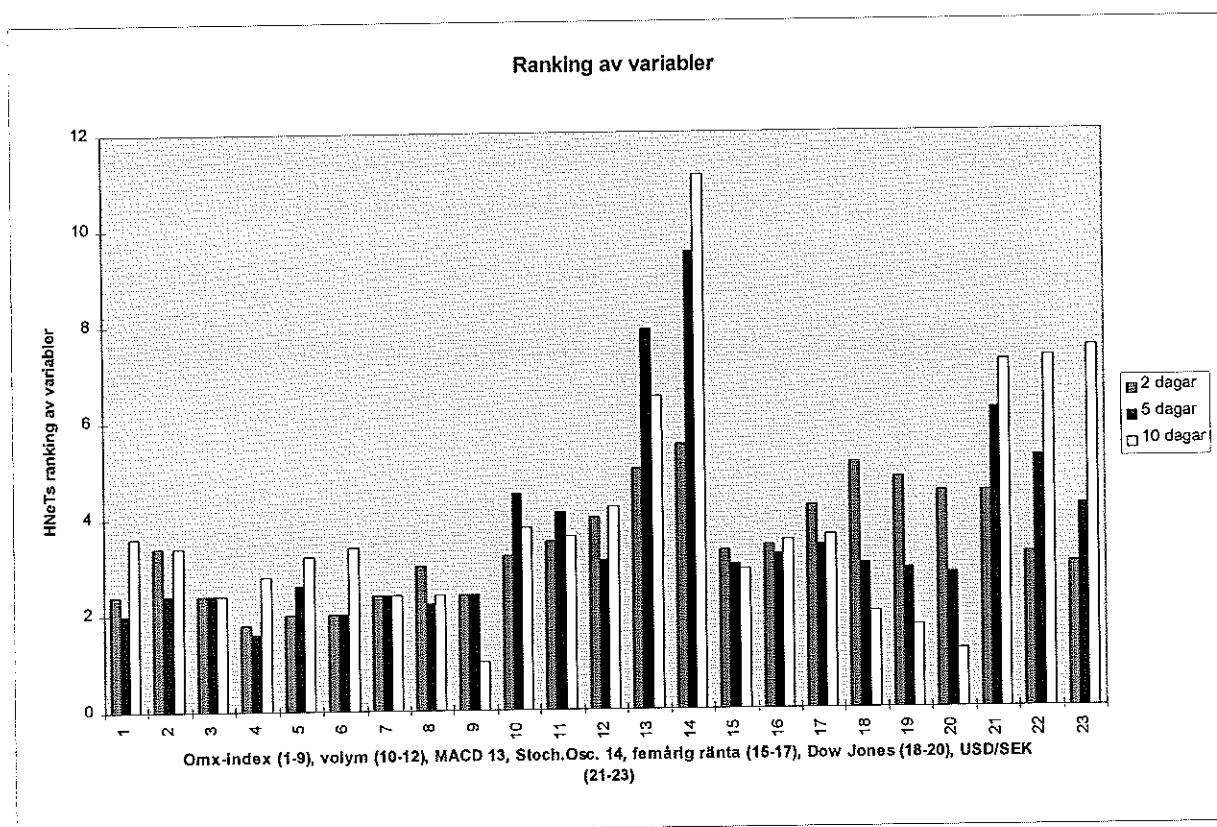


Diagram 1 - Nätverkets ranking av olika variabler. De tre senaste värdena i respektive tidsserie har använts som indata. Som jämförelse skickas dessutom två tekniska indikatorer (13-14) med som indata.

Vad kan man då dra för slutsatser av diagram 1. Tydligt är de tekniska indikatorerna "bra" variabler. Klart är att de har mycket högre värden än obehandlad OMX-data. Vidare noteras att New York-börsens utveckling endast har betydelse för den kortsiktiga prediktionen. USD/SEK är den fundamentala data som ges störst vikt. Som tidigare nämnts blir emellertid prediktionsresultatet med denna enkla uppsättning variabler dåligt, varför försök med transformering av tidsserierna har gjorts. Härvid noteras att högst rankingvärde erhålles med lämpliga

transformer av OMX-index. Dessutom tenderar obehandlad OMX-data att öka ganska kraftigt i betydelse då transformer läggs till medan de fundamentala variablernas betydelse minskar. Det verkar alltså som om just kombinationen av "rå" och transformerad OMX-data är viktig. USD/SEK samt dess transformer har dock fortfarande ett relativt högt värde.

Det är den stora mängden möjliga transformeringar som gör att det blir fråga om ett gigantiskt sökproblem. Antag att ovanstående 5 tidsserier används och att man för varje tidsserie kan hitta 15 möjliga variabler enligt nedanstående lista:

- 1) Värde idag
- 2) Förändring sedan föregående dag
- 3) Högsta värde de senaste 20 dagarna
- 4) Antal dagar sedan detta högsta värde antogs
- 5) Lägsta värde de senaste 20 dagarna
- 6) Antal dagar sedan detta lägsta värde antogs
- 7) Koefficient för en 30-dagars regressionslinje
- 8) 10 dagars glidande medeltal
- 9) 3 dagars glidande medeltal
- 10) Antal dagar med uppgång de senaste 10 dagarna
- 11-15) Tekniska indikatorer (t.ex MACD och Stochastic Oscillator)

På hur många olika sätt kan man välja 3 variabler ur varje tidsserie under förutsättning att de 15 ovanstående transformeringarna är möjliga? Kombinatoriken ger svaret 6825. Man skulle alltså behöva skapa nästan 7000 olika träningsfiler och träna nätverket på dessa för att hitta den bästa kombinationen. I verkligheten är antalet möjliga transformeringar enormt mycket större än de 15 som räknats upp ovan. För att på ett fullständigt eller i varje fall nöjaktigt sätt kunna gå igenom alla möjliga kombinationer krävs tillgång till genetiska algoritmer. Då detta inte har funnits har en mera "manuell" utprovning ägt rum, vilket har inneburit ett enormt tidskrävande arbete. I diagram 2 visas resultatet då nätverket fått göra en ranking av 42 st. variabler. De 19 första är en del av de transformeringar som jag till slut funnit ge bästa resultat och de övriga är samma uppsättning som i diagram 1.

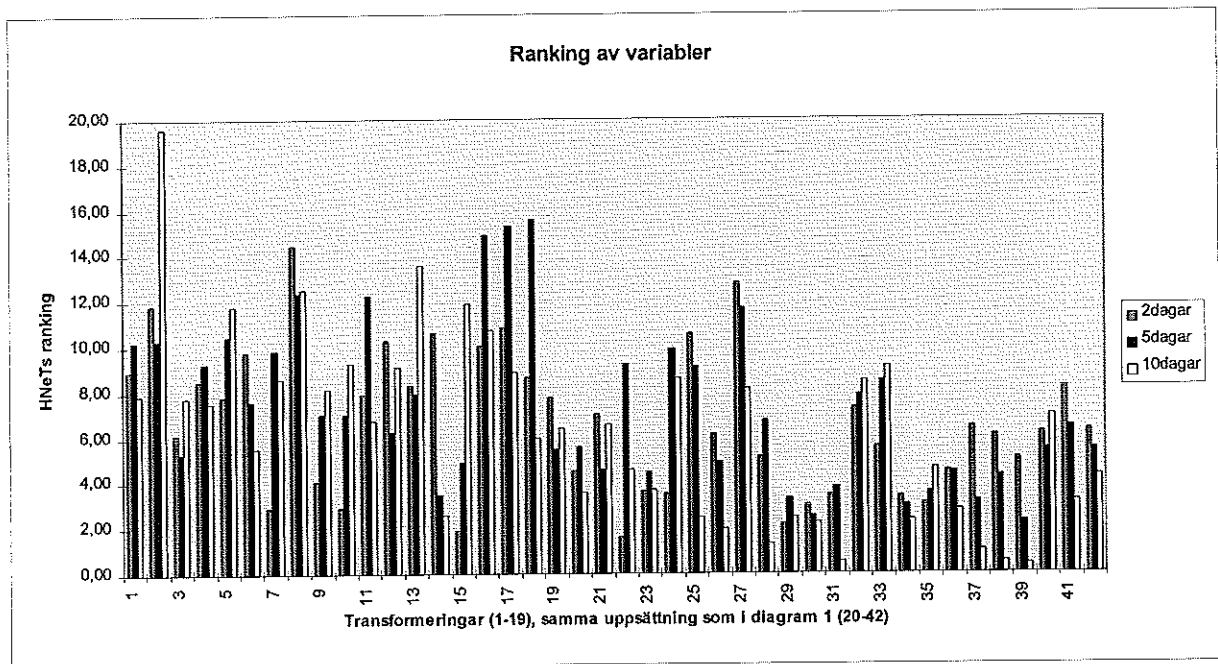


Diagram 2 - 19 av de bästa transformeringar som erhållits samt samma uppsättning som i diagram 1. Lägg märke till att den obehandlade OMX-datan ökat i värde och många av de fundamentala variablerna minskat i värde (t.ex. räntan 34 - 36). Volymen (29-31) har ett lågt värde medan USD (40-42) fortfarande är bra.

Hur många variabler ska träningsfilen innehålla? Svaret på denna fråga beror givetvis på hur "bra" variablerna är (dvs hur högt rankingvärde de får av HNeT). En massa dåliga variabler som inte tillför någon vettig information skapar bara onödigt brus i systemet. Jag har funnit att det holografiska nätverket fungerar bättre då ett relativt stort antal variabler används, närmare bestämt 30-40 stycken. Detta antal är betydligt större än det antal som använts i tidigare gjorda undersökningar med konventionella back-propagationnätverk [13, 14, 15, 16, 17].

#### 4.32 Tidshorisont och svarstyp

De olika alternativ som står till buds när det gäller typen av svar som man vill att nätverket ska avge är dels relativt precisa utsagor av typen "1123 om 5 dagar" eller "högsta 1148, lägsta 1107 de närmaste 10 dagarna" och dels mindre pretentiösa svar av typen "upp/ned på 5 dagars sikt". Det gäller också att bestämma den tidshorisont som ger bäst resultat. I diagram 3 och tabell 1 presenteras resultatet då nätverket fått träna på en uppsättning variabler som framtagits med undersökningar enligt föregående avsnitt. Tre olika tidshorisonter har använts - 2, 5 och 10 dagar. Svarstypen är "upp/ned", vilket i träningsfilen representeras av  $\pm 1$ . Resultatet för varje enskild

prediktion beräknas som OMX-index procentuella förändring multiplicerad med tecknet på HNeTs prediktion (dvs resultatet är  $\pm$ OMX-index förändring i procent under aktuell tidsperiod, där plus används då HNeT predikerat uppgång och minus då nedgång förutspåts).

Observera att testfilen består av records som nätverket aldrig "sett". Nätverket har inte heller tränat på records som tidsmässigt ligger efter de records som ingår i testfilen (se vidare under rubriken "Träning av nätverket").

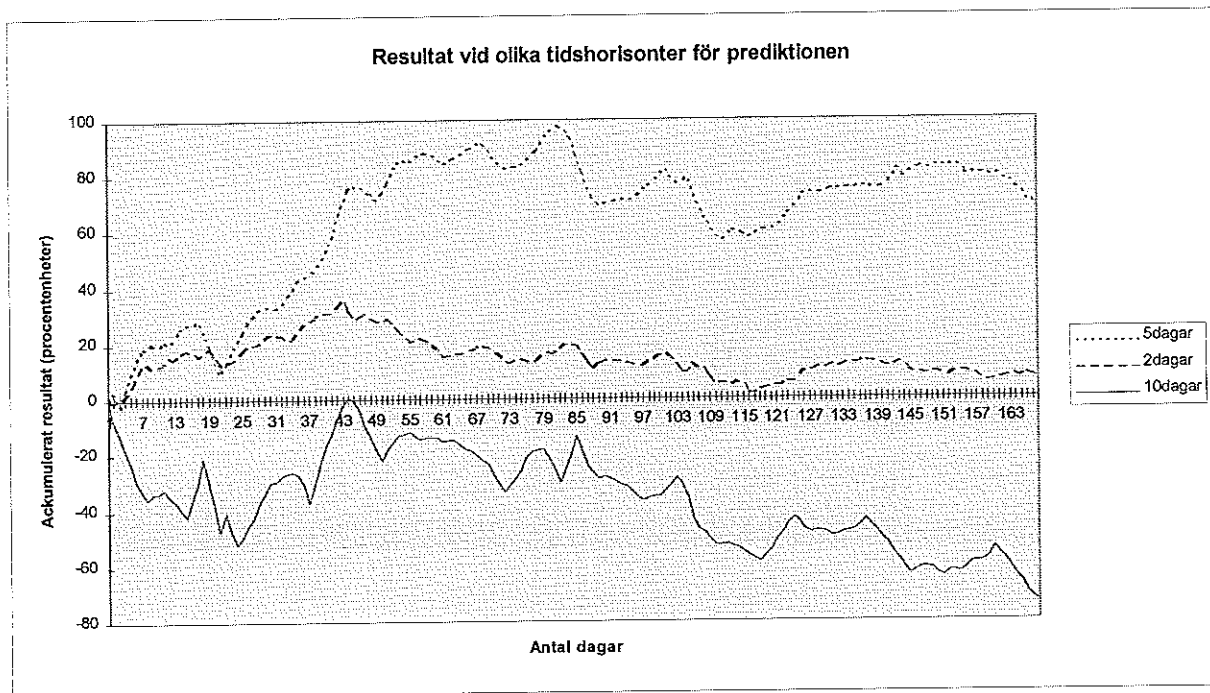


Diagram 3 - Resultat för olika längder på det tidsintervall framåt som nätverket ska prediktera. Svarstypen "upp/ned" används (dvs träningsfilen innehåller +1 om det riktiga svaret är "upp" och -1 om svaret är "ned"). Som synes ger 5 dagar överlägset bäst resultat.

	sanna köpsignaler	falska köpsignaler	sanna säljsignaler	falska säljsignaler
2 dagar	56	39	36	36
5 dagar	51	32	48	36
10 dagar	42	39	34	52

Tabell 1 - Fördelningen av prediktioner då svarstypen "upp/ned" använts på 2, 5 och 10 dagars sikt.

I diagram 4-6 presenteras motsvarande resultat då det exakta OMX-indexet efter aktuell tidsperiod har använts som svar. I diagrammen är prediktionen förskjuten framåt aktuellt antal dagar så att det riktiga OMX-indexet direkt kan jämföras med det predikerade.

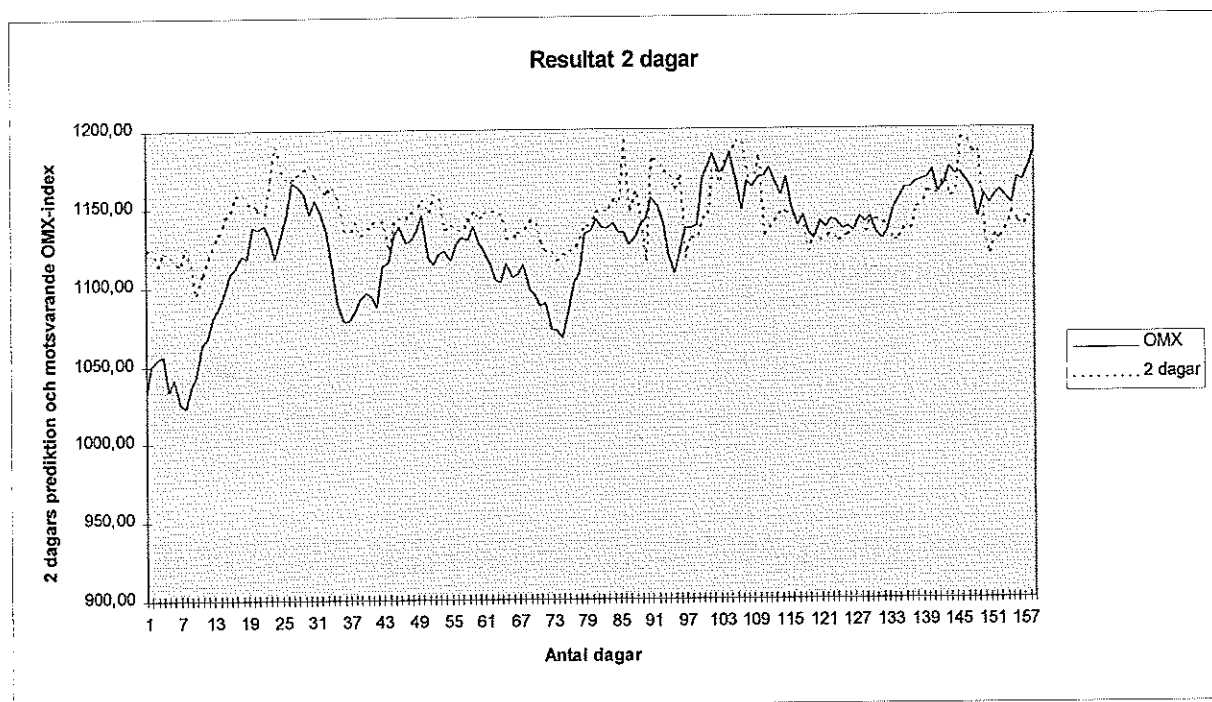


Diagram 4 - Prediktion med det exakta OMX-indexet om 2 dagar som svar. De predikerade värdena är förskjutna 2 dagar framåt så att en direkt jämförelse med det verkliga OMX-indexet kan göras.

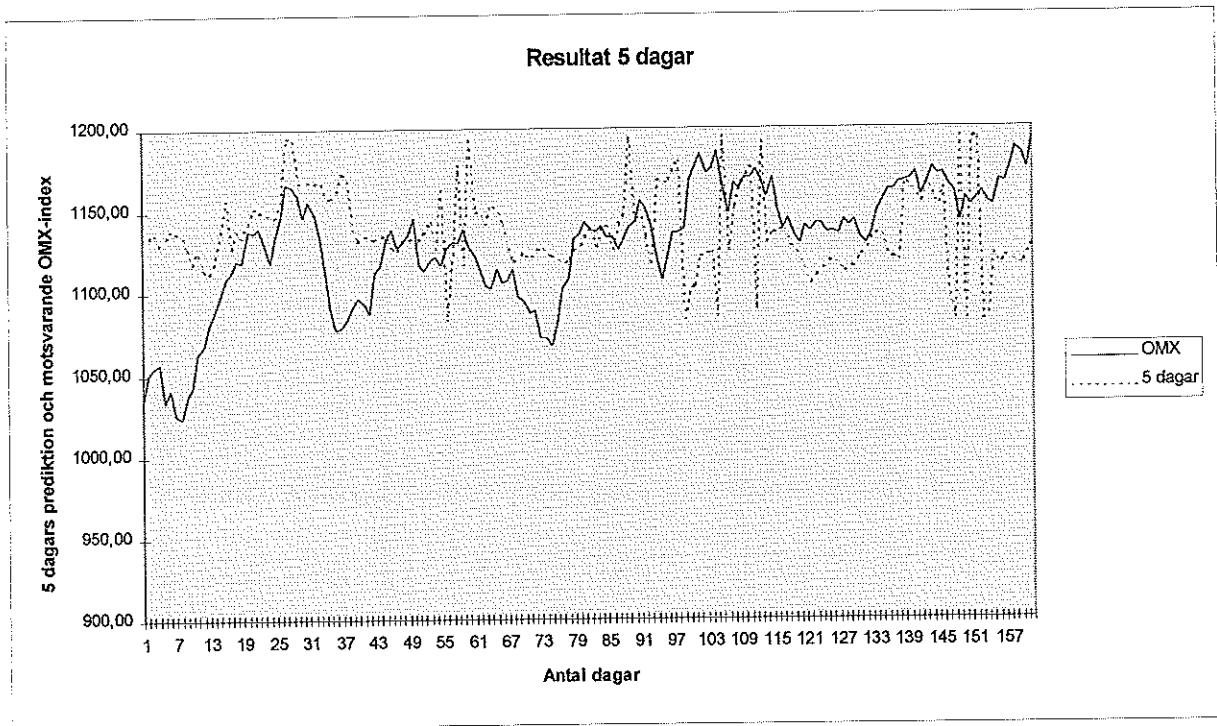


Diagram 5 - Prediktion med det exakta OMX-indexet om 5 dagar som svar. De predikterade värdena är förskjutna 5 dagar framåt så att en direkt jämförelse med det verkliga OMX-indexet kan göras.

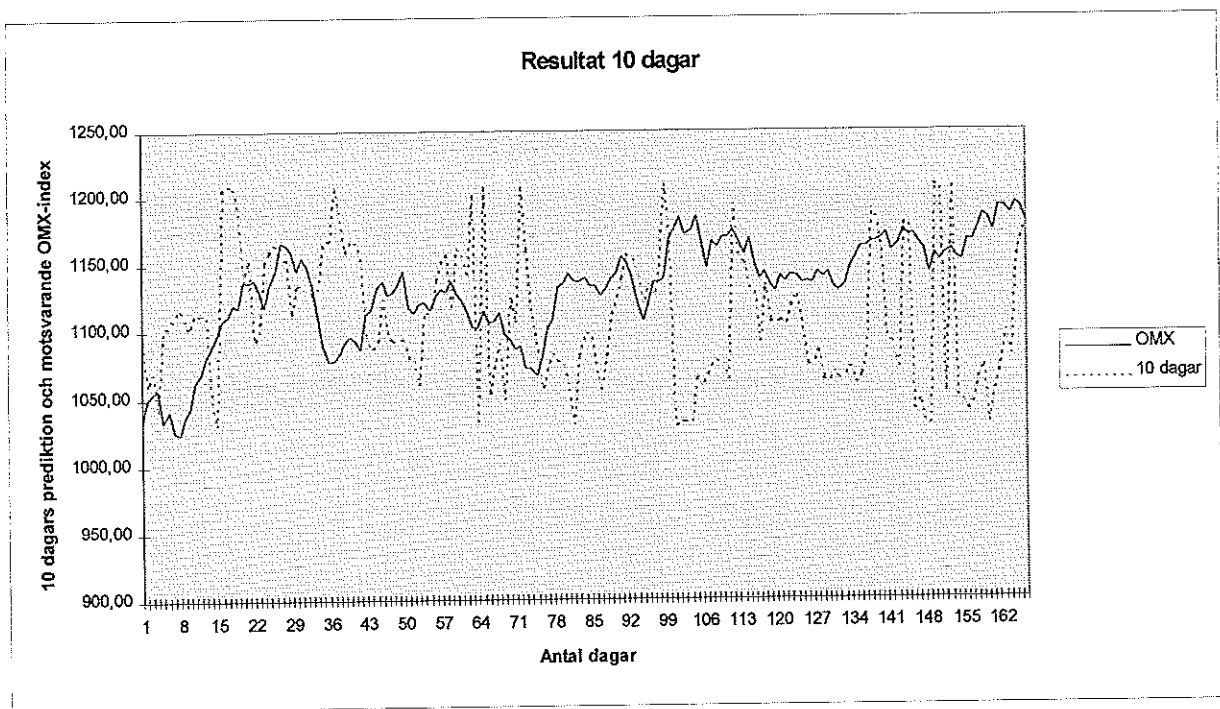


Diagram 6 - Prediktion med exakta OMX-indexet om 10 dagar som svar. De predikterade värdena är förskjutna 10 dagar framåt så att en direkt jämförelse med det verkliga OMX-indexet kan göras.



Som synes erhålles inte några direkt lysande resultat. Nätverket tenderar att bli "slängigt" såtillvida att det kastar sig mellan låga och höga värden på det predikterade OMX-indexet. Utseendet på resultatet gör att man misstänker att förståelsen för den underliggande processen inte är den bästa. Det finns emellertid ett annat sätt att utvärdera resultatet på som ger en intressant aspekt på det hela. En köpsignal kan definieras som tillfällen då nätverkets predikterade värde är större än det aktuella OMX-indexet och en säljsignal då det predikterade värdet är mindre än det aktuella. Då fås resultat enligt diagram 7 och tabell 2, där det fortfarande är den procentuella förändringen i OMX-index under aktuell tidsperiod som ligger till grund för resultatberäkningen.

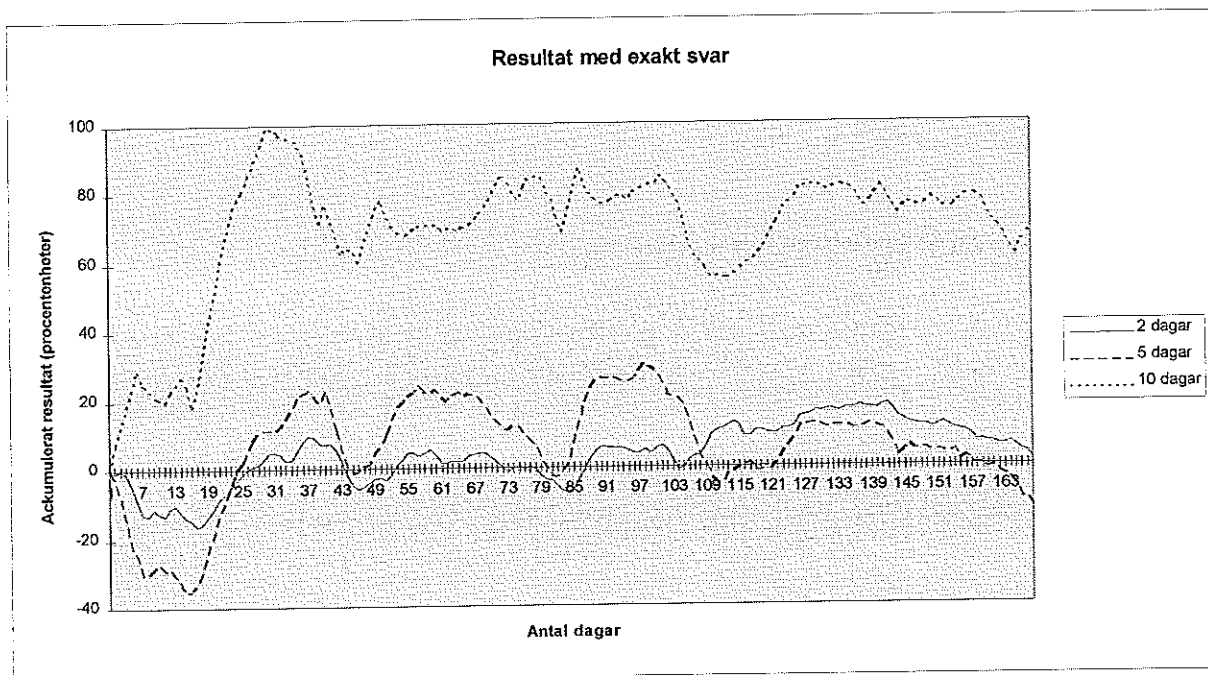


Diagram 7 - Resultat då köpsignal definierats som tillfällen då predikterat OMX-index är större än det aktuella och säljsignal definierats på motsvarande sätt. 10-dagarsprediktionen ger bäst resultat.

	sanna köpsignaler	falska köpsignaler	sanna säljsignaler	falska säljsignaler
2 dagar	66	56	19	26
5 dagar	54	50	30	33
10 dagar	43	25	48	51

Tabell 2 - Fördelningen av prediktioner då exakt svar använts.

Resultatet enligt diagram 7 och tabell 2 är kanske lite förvånande och motsägelsefullt. Ska det tolkas som att nätverket trots allt har lärt sig en del samband men "tar i" för mycket då ett värde ska avges? Varför är det 10-dagarsprediktionen som ger bäst svar här när motsatsen gäller i diagram 3 där "upp/ned" används som svar? Hur som helst förefaller inte detta 10-dagarssystem vara nöjaktigt då en fördjupad undersökning av signalernas

fördelning och snittavkastningen görs. Under rubriken "Utvärdering" studeras dessa aspekter för det system som erhålles med svarstypen "upp/ned" på 5 dagars sikt.

### 4.33 Träning av nätverket

Det praktiska arbetet består i att träna och sedan testa sitt nätverk på datauppsättningar som har tagits fram med ovan nämnda undersökningar. Det gäller här att bestämma hur mycket träningsdata som skall användas (dvs hur långt tillbaka i tiden använda tidsserier ska sträcka sig). En tumregel som härvidlag brukar användas är att antalet träningsfakta (eller records som det brukar kallas i ANN-teori) bör vara minst tio gånger fler än antalet variabler. Undersökningar av resultatet då tidsserier av olika längd använts har gjorts. Jag har funnit att för långa tidsserier ger sämre resultat troligen beroende på att förhållandena på börsen ändras med tiden. Samband som gällde 1987 gäller kanske inte 1995. För korta tidsserier innebär att nätverket inte får se alla tänkbara situationer. Vill man t.ex. att nätverket ska kunna förutsäga ett rejält börsras, "en krasch", så måste träningsdata innehålla exempel på sådana "krascher".

Det erhållna resultatet kan förbättras ytterligare något genom att i görligaste mån plocka bort s.k. "outliers" ur träningsdatauppsättningen. "Outliers" i vårt fall kan vara perioder då börsen påverkats av naturkatastrofer, krigsutbrott eller liknande. Det kan också vara perioder då börsen bedöms ha uppträtt ovanligt "kaotiskt". Genom att på detta sätt plocka bort records som eventuellt utgör motstridig och förvirrande information för nätverket kan brusnivån sänkas. Det ska dock nämnas att det om det överhuvudtaget låter sig göras är en mycket arbetssam procedur.

Valet av testset är inte helt självklart. Ofta rekommenderas att 70 % av antalet records används för att träna nätverket och 30 % används för att testa det erhållna färdiga nätverket. Om dessa 30 % väljs slumpmässigt fås ofta effekten att nätverket ger mycket bra resultat för testdata men dåligt resultat då det testas med helt "ny" data (dvs data som t.ex. hämtas dag för dag från börsen). Detta fenomen beror på att det för varje faktauppsättning (record) i testsetet finns två mycket snarlika uppsättningar i träningsfilen, nämligen "dagen före" och "dagen efter". Detta gör att nätverket inte behöver ha någon generaliseringsförmåga för att klara prediktionsuppgiften. Den här effekten blir givetvis större ju långsiktigare prediktionen är.

En annan aspekt vad gäller träning är om och i så fall när omträning ska ske. Resultatet för prediktionen på 5 dagars sikt i diagram 3 uppvisar sämre resultat under den senare delen av tidsperioden. Möjligen skulle en omträning kunna förbättra resultatet under denna period. För att undersöka detta har jag tränat om nätverket efter varje 30-dagarsperiod. Jag tränar alltså nätverket och låter det prediktera de följande 30 dagarna. Därefter inlemmas dessa 30 dagar i träningsfilen och en ny träningskörning sker och nästa 30-dagars period predikteras. Resultatet presenteras i diagram 8 och tabell 3. Som synes gör det i stort sett vare sig till eller i fråga om omträning sker, vilket förkastar möjligheten att det sämre resultatet under den senare perioden förklaras av att nätverket inte har blivit "uppdaterat"

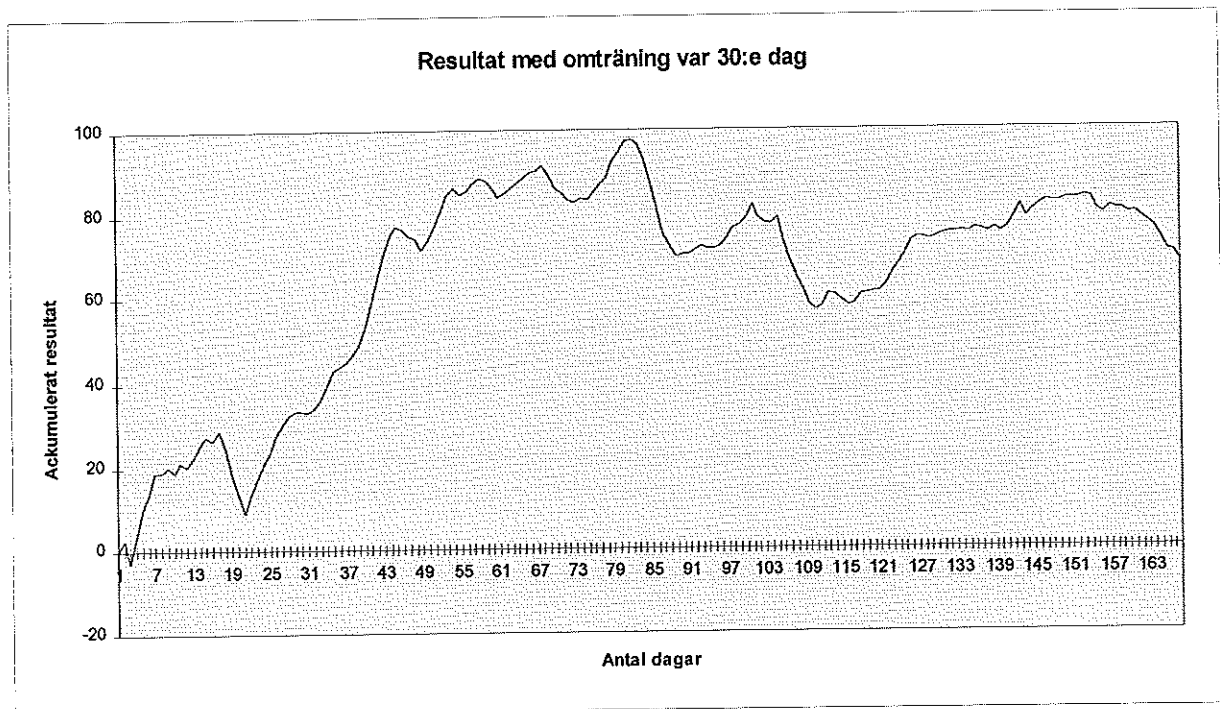


Diagram 8 - Samma konfiguration som i 5-dagarsfallet enligt diagram 3. Här har dessutom en omträning skett var 30:e dag. De data som nätverket gjort en prediktion för inlemmas i träningsfilen till nästa träningskörning.

sanna köpsignaler	falska köpsignaler	sanna säljsignaler	falska säljsignaler
51	32	48	36

Tabell 3 - Fördelningen av prediktioner då omträning gjorts. Ingen skillnad märks gentemot tabell 1.

Givetvis är det vid träningen också av största vikt att ha rätt interna inställningar i använd programvara. En undersökning av optimala värden på de parametrar som ingår i HNeT Discovery Package presenteras i bilaga 2.

#### 4.34 Konventionellt nätverk

Som en jämförelse har ett konventionellt nätverk (Brainmaker version 2.0; California Scientific Software) använts med samma uppsättning variabler som i diagram 3. Träning har endast skett med 5-dagarsalternativet. I diagram 9 och tabell 4 visas resultatet för den bästa konfigurationen av antalet neuroner. 5 olika värden på antalet gömda neuroner i ett gömt lager har provats. Det visar sig att resultatet blir bäst med 41 neuroner. Med färre neuroner blir resultatet markant sämre. Nätverket är dock inte särskilt känsligt för om antalet neuroner ökas.

En mängd olika konfigurationer med två gömda lager har provats. Ingen har varit lyckad. Ofta fås i det här fallet ett nätverk som konstant ger säljsignal.

Träningen tar i det här fallet flera timmar, speciellt då flera gömda lager används.

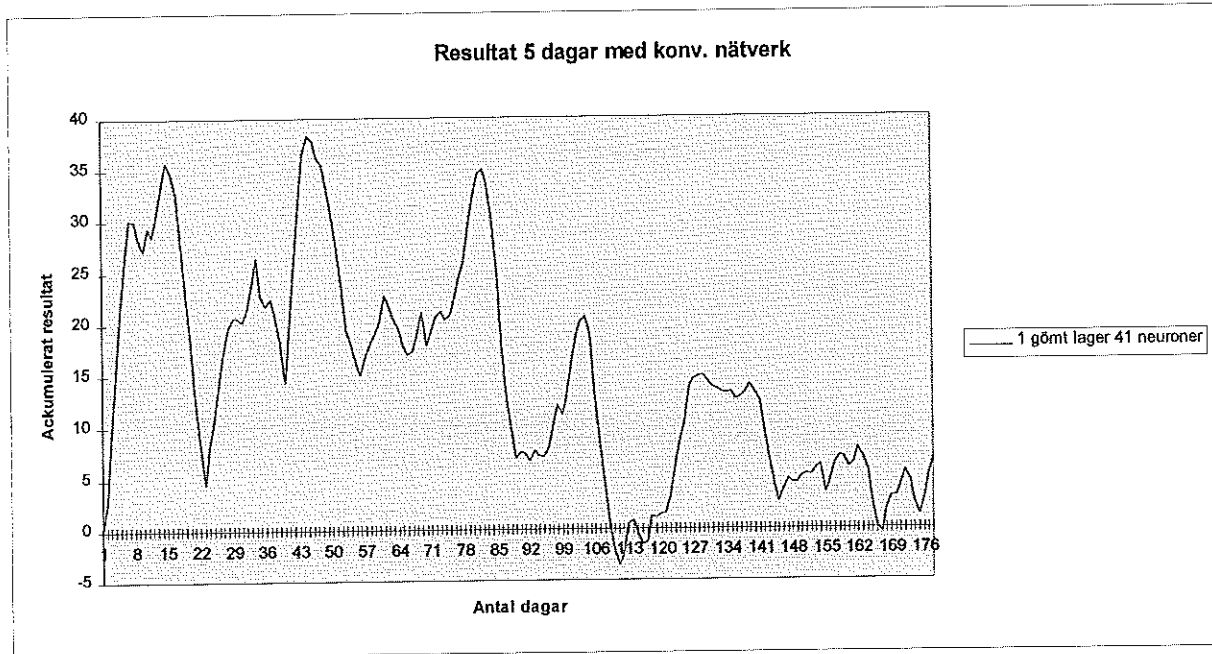


Diagram 9 - Bästa erhållna resultat då ett konventionellt nätverk tränats med samma träningsfil som använts i 5-dagarsfallet i diagram 3. Ett gömt lager med 41 neuroner har använts.

sanna köpsignaler	falska köpsignaler	sanna säljsignaler	falska säljsignaler
26	22	65	64

Tabell 4 - Fördelningen av prediktioner då ett konventionellt nätverk med 1 gömt lager med 41 neuroner tränats med samma träningsfil som i 5-dagarsfallet i tabell 1.

Avslutningsvis bör det betonas att det huvudsakliga arbetet har skett med HNeT Discovery Package, varför det med ökade arbetsinsatser borde gå att erhålla bättre resultat även med back-propagationnätverk. Det är t.ex. inte säkert att det är samma uppsättning variabler som ska användas vid de olika typerna av nätverk. Som tidigare nämnts har en mängd undersökningar gett bra resultat även med konventionella nätverk [14, 15, 16, 17]. I dessa undersökningar har oftast ett mindre antal variabler använts än vad som är fallet här.

## 4.4 Utvärdering

Nedan visas och utvärderas resultaten för två tänkbara system som bildas med nätverkets prediktioner som grund. Nätverket tränas på att förutsäga om OMX-index går upp eller ned på 5 dagars sikt. Träningsfilen innehåller ett relativt stort antal variabler framtagna med undersökningar enligt tidigare avsnitt. Det riktiga svaret är +1 om index gått upp och -1 om index gått ned. Ett positivt värde på nätverkets svar innebär således en köpsignal och ett negativt på motsvarande sätt en säljsignal. Det ena systemet "handlar" på alla signaler och det andra enbart på signaler vars absolutbelopp är större än 0.8. Dessa sistnämnda signaler kallas fortsättningsvis för starka signaler. Tanken är att signaler som ligger närmare  $\pm 1$  är säkrare än signaler som ligger kring nollstrecket (se bilaga 2, "Real Respons Conversion").

Det första som måste klarläggas vid utvärderingen av ett mekaniserat tradingsystem är huruvida systemet slår en "buy and hold strategi" (dvs indexets utveckling under den aktuella tidsperioden). Enligt diagram 10 klaras detta första test galant. OMX-index rör sig 5.68 % uppåt medan systemet ger ett resultat på 66.7 % om alla signaler följes och 16.0 % om endast de starka signalerna beaktas..

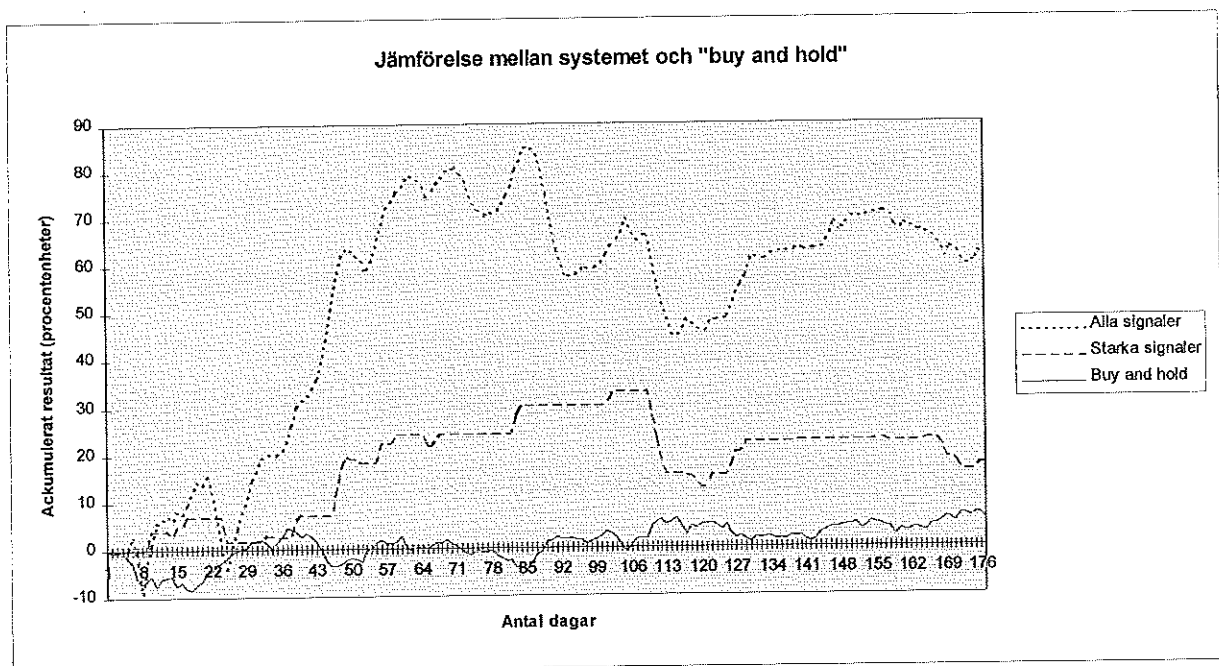


Diagram 10 - Resultat om alla signaler följes samt motsvarande om endast de s.k. starka signalerna följes. Som jämförelse har OMX-index utveckling lagts in.

Därefter beaktas signalernas fördelning. Systemen måste prestera lika bra resultat både på köp- och säljsignalerna. Enligt tabell 5 och 6 förefaller signalerna vara jämnt fördelade både vad gäller kvalitet och kvantitet. Noteras kan att då endast de starka signalerna beaktas fås ett sämre förhållande mellan sanna och

falska säljsignaler, vilket inte är bra. Även avkastningarna och deras fördelningar måste studeras. Helst vill man här se ett förhållande mellan medelavkastningen för sanna signaler och falska signaler på uppemot 2. Tyvärr uppfyller inget av systemen detta krav. Mera glädjande är värdet på snittavkastningen. Konfidensintervallet för avkastningen för alla signaler säkerställer statistiskt att snittavkastningen är positiv. Tyvärr släpar även här de starka signalerna efter.

Avkastningsdata		Signalernas fördelning			
högsta avkastning	6.75	vinstgivande signaler	105	sanna köpsignaler	55
lägsta avkastning	-6.58			sanna säljsignaler	50
snittavkastning	0.38	förlustbringande signaler	72	falska köpsignaler	37
median avkastning	0.64			falska säljsignaler	35
standardavvikelse avkastning	2.49				
medel för vinstgivande signaler	1.96				
medel för förlustgivande signaler	-1.94				
95% konf.intervall för avkastningen	[0.009 , 0.74]				
riktningskoeff. för regr.linje i diagram 11	0.20				
80 % konf.intervall för regr.koeff	[-0.17 , 0.56]				

Tabell 5 - Resultat för det system som erhålles om alla signaler följes. Alla avkastningsdata avser resultat i procentenheter på 5 dagars sikt. En riktningkoefficient för en regressionslinje i Diagram 11 har beräknats samt motsvarande konfidensintervall.

Avkastningsdata		Signalernas fördelning			
högsta avkastning	6.02	vinstgivande signaler	27	sanna köpsignaler	15
lägsta avkastning	-5.53			sanna säljsignaler	12
snittavkastning	0.25	förlustbringande signaler	22	falska köpsignaler	10
median avkastning	0.32			falska säljsignaler	12
standardavvikelse avkastning	2.54				
medel för vinstgivande signaler	2.03				
medel för förlustgivande signaler	-1.92				
95% konf.intervall för avkastningen	[-0.33 , 0.91]				
riktningskoeff. för regr.linje i diagram 12	0.20				
80 % konf.intervall för regr.koeff	[-0.24 , 0.67]				

Tabell 6 - Motsvarar tabell 5 fast med data från endast de starka signalerna. Tråkigt nog erhålles sämre resultat än då alla signaler följes. Tanken var att bättre snittavkastning och bättre förhållande mellan sanna och falska signaler skulle erhållas när endast de "säkra" signalerna följdes.

I diagram 11 och 12 har HNeTs prediktion plottats mot den verkliga procentuella förändringen av OMX-index efter 5 dagar. Helst vill man här kunna skönja en tendens till att värdena ligger längs en linje med positiv riktningkoefficient. I tabell 5 och 6 visas riktningkoefficienten för en regressionslinje i diagrammet samt ett 80 % konfidensintervall för denna koefficient.

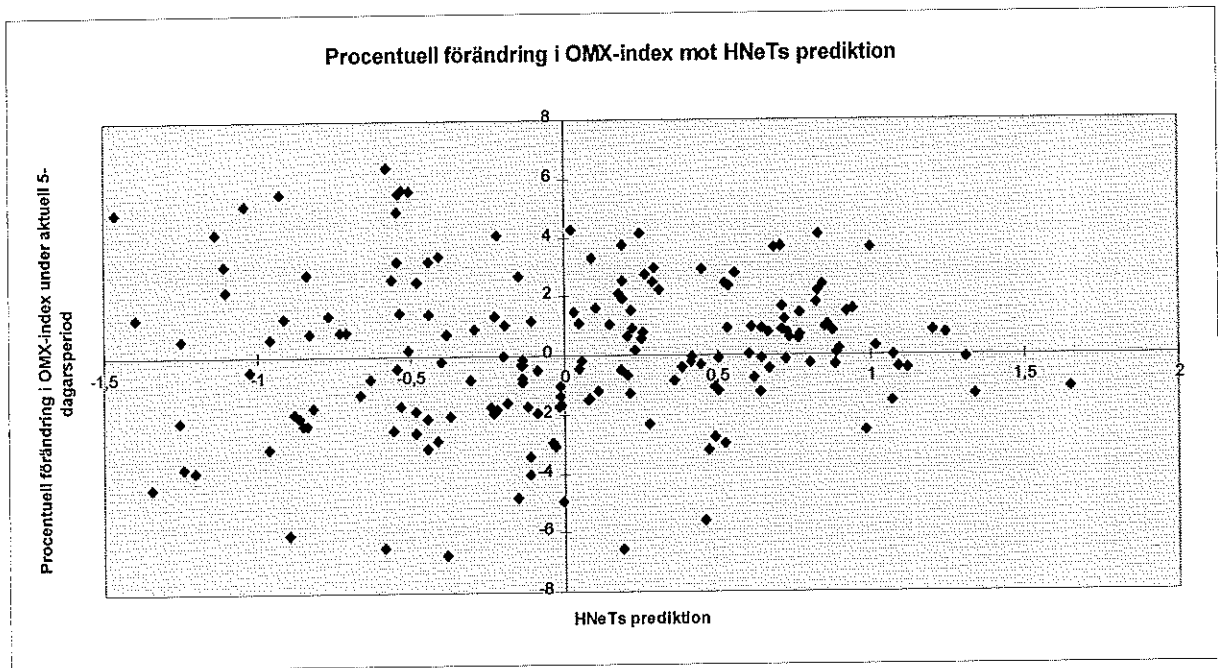


Diagram 11 - HNeTs prediktion på horisontella axeln och OMX-index förändring på 5 dagars sikt på vertikala axeln. Helst ska ett rätlinjigt samband med positiv riktningskoefficient kunna skönjas. En beräknad regressionslinjes riktningskoefficient samt ett konfidensintervall för densamma redovisas i tabell 5.

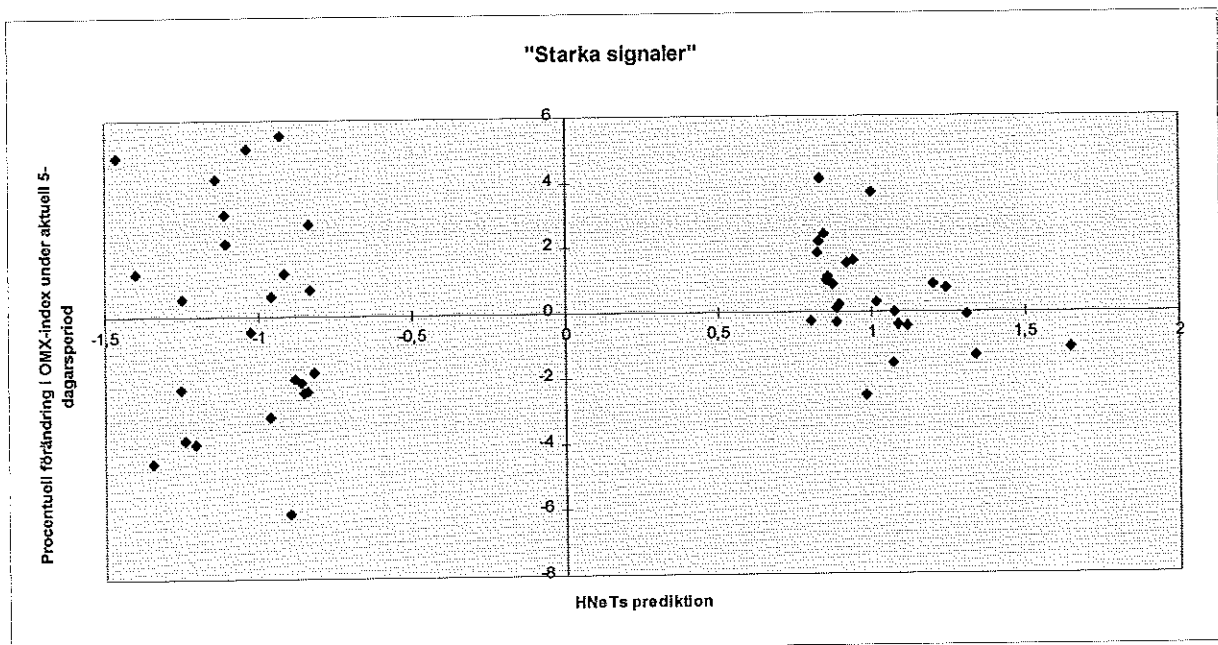


Diagram 12 - Samma som diagram 11 fast med endast de starka signalerna och motsvarande förändring i OMX-index.

## 4.5 Begränsningar i undersökningar och utvärderingar

I föreliggande arbete har tidsserier för OMX-index, omsättningen på börsen, Dow Jones-index, femåriga marknadsräntan och USD/SEK studerats med avseende på möjligheten att finna lämpliga variabler till ett neuralt nätverk för prediktion av OMX-index. Det finns säkert en hel del andra tänkbara kandidater som är värda att studera (t.ex. DAX-index, FTSE-index, kronindex och data från den till OMX-index knutna derivatmarknaden). Även då endast de använda tidsserierna beaktas kan inga anspråk göras på att fullständigt ha behandlat och studerat de möjliga transformeringar som kan göras. Som tidigare nämnts hade genetiska algoritmer varit ett superbt hjälpmedel i detta avseende.

Jämförelsen med konventionella nätverk blir möjligen något orättvis eftersom den största energin lagts på modellerandet av det holografiska nätverket. Variablerna har t.ex. valts ut med hjälp av det holografiska neurala nätverket och det är troligt att ett något bättre resultat kunnat erhållas med det konventionella nätverket om ytterligare tid lagts på att hitta "rätt" uppsättning variabler och rätt konfiguration för det konventionella nätverket.

Vid utvärderingen beaktas endast de procentuella förändringar som OMX-index rörelse ger upphov till. Inga resultat i form av antal punkters vinst per terminsaffär eller liknande har beräknats. Härigenom har ingen hänsyn tagits till diverse avgifter (typ courtage) som minskar en eventuell vinst räknad i kronor och ören. För att göra en fullständig beräkning fordras tillgång till motsvarande tidsserie för terminspriserna. Det finns dessutom en hel del att fundera på vid utformningen av ett "tradingssystem" baserat på nätverkets prediktioner. Det är inte säkert att metoden med ett system som "alltid ligger i marknaden" och som alltid fullföljer affärerna under hela 5-dagarsperioden är det optimala sättet att utnyttja prediktionerna på.

## 4.6 Sammanfattning och slutsatser

Utförda undersökningar visar att det är av yttersta vikt att hitta lämpliga transformeringar av datan i berörda tidsserier. Kombinationen av obehandlad och transformerad OMX-data har visat sig ge bra förklaringsvärde. Av de övriga tidsserierna framstår dollarkursen och dess transformer som klart bäst. Dow Jones-index ges ett högt förklaringsvärde endast vid kortsiktig prediktion. Det har visat sig att nätverket ger bäst resultat då ett enkelt svar av typen "upp/ned" avkrävs. Tidshorizonten för prediktionen bör sättas till 5 dagar. Man kan fundera lite kring varför just 5 dagar ger bäst resultat. Begränsningen uppåt i tid kan förklaras genom att göra en jämförelse med det tidigare diskuterade problemet att förutsäga vädret. På 10 dagars sikt tar helt enkelt de kaotiska egenskaperna överhanden. Det faktum att det med all sannolikhet finns ytterligare storheter med påverkan på börsen som inte tas med i den här använda modellen bidrar också till det sämre resultatet. Dessutom ökar ju sannolikheten att ny och fundamentalt omdanande information ska nå marknaden. Hur förklarar man begränsningen nedåt i tid då? En möjlig förklaring kan vara att endast indata med tre värden per dag (sista, högsta och lägsta) används. Nätverket får aldrig se den finare struktur som behövs för att göra kortsiktigare prediktioner. Hade tillgång till data insamlad t.ex. varje kvart funnits hade troligtvis bättre prognoser av utvecklingen på en eller två dagars sikt kunnat göras.

Hur skall man tolka resultatet av föreliggande arbete? För en akademiskt inriktad nationalekonom är den naturliga utgångspunkten den effektiva marknadsmodellen. Teorin för denna modell är allmänt betraktad som en grund för hur de finansiella marknaderna anses fungera. Klart är att det enligt denna modell inte ska finnas någon prediktiv information att hämta i t.ex. den historiska prisutvecklingen. Ett flertal undersökningar har gjorts för att visa på meningslösheten med teknisk analys [10, 11]. Föreliggande arbete måste emellertid betraktas som ett argument för att prediktiv information finns i den historiska prisutvecklingen. Systemet uppvisar dels en snittavkastning som med 95 % konfidens är större än noll och dels en uppenbarligen god separation av sanna och falska signaler. Även tidigare gjorda undersökningar där konventionella back-



propagationnätverk använts visar på en klar effekt av "överavkastning" (dvs en avkastning som med bred marginal slår en "buy and hold"-strategi) [13, 14, 15, 16, 17].

Vissa frågetecken finns dock när det gäller det erhållna resultatet. De s.k. starka signalerna borde vara mycket bättre än vad fallet är. Detta är den allvarligaste bristen hos systemet, eftersom den absoluta storleken på svaret ska spegla hur "säkert" svaret är. Här faller en viktig möjlighet till bedömning av tillförlitligheten i avgivna prediktioner bort.

Det kan konstateras att arbetet med att hitta rätt form för det neurala systemet (variabler, svarstyp, tidshorisont osv.) är ett stort "trial and error"-problem. Antalet möjliga varianter är i stort sett oändligt. Tillgången till det snabba holografiska nätverket har härvidlag underlättat arbetet mycket. Med ett konventionellt nätverk hade endast en bråkdel av de nu gjorda undersökningarna kunnat göras.

## 5 Referenser

- [1] Carsten Peterson & Thorsteinn Rögnvaldsson, "Artificiella neuronnätverk", Institutionen för Teoretisk Fysik, Lunds Universitet, även publicerad i *KOSMOS* 1992, Lund 20 januari 1992.
- [2] D. Hebb, "The first stage of Perception; Growth of the Assembly", *The Organisation of behavior*, John Wiley & Sons Inc., pp60-78, 1949
- [3] Adam Blum, *Neural Networks in C++ - An Objectoriented Framework for Building Connectionist Systems*, John Wiley & Sons Inc., 1992
- [4] AND America Ltd., Manual till HNeT Discovery Package, 1993
- [5] John G. Sutherland, "The Holographic Cell- A Quantum Perspective", *Frontier Decision Support Concepts*, John Wiley & Sons Inc., 1994
- [6] John G. Sutherland, "A Holographic Model of Memory, Learning and Expression", *International Journal of Neural System Vol 1 No 3 pp 256-267*, 1990
- [7] ITANIS International Inc., "What is holographic Neural Technology", 1995
- [8] Martin J. Pring, *Technical Analysis Explained*, McGraw - Hill Inc, 1991
- [9] Robert P. Rotella, *The Elements of Successful Trading*, New York Institute of Finance, Simon & Schuster, 1992
- [10] Richard A. Brealey & Stewart C. Meyers, *Principles of Corporate Finance*, McGraw - Hill Inc., 1991
- [11] Edwin J. Elton & Martin J. Gruber, *Modern Portfolio Theory and Investment Analysis*, John Wiley & Sons Inc. 1991
- [12] G. E. D. Box & G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976
- [13] Robert R. Trippi & Duane DeSiento, "Trading Equity Index Futures With a Neural Network", *The Journal of Portfolio Management pp 27-32*, våren 1992
- [14] W. E. Bosarge, Jr, "Adaptive Processes To Exploit The Nonlinear Structure of Financial Markets" presenterad vid Santa Fe Institute of Complexity Conference: Neural Networks and Pattern Recognition in Forecasting Financial Markets, 1991
- [15] T. Kimoto, K. Asakuma, M. Yoda, M. Takeoda, "Stock Market Prediction System With Modular Neural Networks", *Proceedings of the IEEE International Joint Conferenc on Neural Networks*, pp 11-16, San Diego, 1990
- [16] Shaun-inn Wu & Ruey-Pyang Lu, "Combining Artificial Neural Networks and Statistics for Stock Market Forecasting", ACM, 1993
- [17] John & Tod Loofbourrow, "Traders Using Nets Beat Those Who Don't", *The magazine of Artificial Intelligence in Finance No 1*, 1994
- [18] NovaCast AB, "HNeT - Användartips från NovaCast AB", NovaCast AB, Ronneby, Maj 1994
- [19] Edgar E. Peters, *Chaos and Order in The Capital Markets*, John Wiley & Sons Inc., 1991

- [20] Gunnar Ohlén & Sven Åberg, *Kaos Inom Naturvetenskap och Teknik*, Institutionen för Matematisk Fysik  
Lunds Universitet, Augusti 1993
- [21] L. Yaroslavski, *Methods of Digital Holography*, Plenum Publishing Corporation, New York, 1980
- [22] M. Fishman, D. Barr & W Loick, "Artificial Intelligence and Market Analysis", *Technical Analysis of Stocks and Commodities*, Mars 1991

# Bilaga 1

## Utförlig beskrivning av den holografiska metoden

Nedan följer en redovisning av de olika möjligeter som finns för att förstärka och förbättra den holografiska metoden som den beskrivits tidigare i detta arbete.

### Förbättrad inkodning

Kraftigt förbättrade egenskaper uppnås i HNeT-metoden genom användning av en inkodningsprocedur som tar hänsyn till tidigare inkodade associationer (dvs det minne eller den information som finns inkapslad i korrelationsmatrisen). Denna procedur innebär att hela rummet av inputstimuli kan avbildas näst intill exakt på de tillhörande responserna. Antalet distinkta stimulus-respons associationer som kan inkodas utökas genom användning av indata av högre ordning (detta beskrivs senare).

I den enkla inkodningsproceduren inkapslas stimulus-respons associationer i korrelationsmatrisen. Varje association inkodas med ett förutbestämt konfidensvärde. Inläring sker oberoende av den kunskap som redan förvärvats. Denna enkla procedur är inte optimal i det avseende att många likartade stimulus-respons associationer tenderar att förvränga eller påverka alla övriga associationer som inkodats. Asymmetrier kan uppstå i dessa inkapslade associationer vilket leder till att alla återgivna responser tenderar att bli kraftigt förstärkta. Dessa icke önskvärda egenskaper eller begränsningar elimineras med den förbättrade inkodningsprocedur som används i HNeT. Fortfarande gäller att konfidensvärdena (absolutbeloppen) hos de genererade responserna ligger i intervallet  $[0,1]$ , vilket ger numerisk stabilitet för proceduren.

Den förbättrade inkodningsproceduren ger dessutom bättre egenskaper vad gäller förmågan att distinkt separera (eller separera "topologiskt") avbildningar för närliggande stimulusregioner eller stimulusuppsättningar i närheten av extrempunkter i "stimulusrummet". Responsfelet för den förbättrade inkodningsproceduren är kraftigt reducerat jämfört med den enkla inkodningsproceduren.

Genom att låta inkodningen vara beroende av det ackumulerade minnet fås automatiskt kontroll över graden av inläring av stimulus-respons associationerna. Endast helt nya associationer får stort inflytande på korrelationsmatrisen. Motsatt så ger ny inkodning av tidigare inkodade associationer endast en marginell påverkan av korrelationsmatrisen.

Den förbättrade inkodningsproceduren kan indelas i tre steg. Först sker en avkodning där den nya stimulusvektorn,  $[S]$ , transformeras genom korrelationsmatrisen till den avgivna responsen. I andra steget beräknas differensen mellan den önskade responsen och den avgivna. I sista steget utförs en inkodning där den nya stimulusvektorn avbildas på denna differens. Stegen beskrivs nedan, där det som vanligt antages en respons per stimulusuppsättning.

1) Genomför en avkodning av den nya stimulusvektorn:

$$R' = \frac{1}{c} [S] \cdot [X] \quad (27)$$

2) Beräkna differensen mellan den önskade responsen och den ovan avgivna:

$$R_{diff} = R - R' \quad (28)$$

3) Inkoda avbildningen från stimulusvektorn till differensen:

$$[X] = [\overline{S}]^T \cdot R_{diff} \quad (29)$$

Denna procedur utgör en inkodningsmetod där den nya stimulusvektorn avbildas näst intill exakt på önskad respons efter endast en inkodning. Proceduren kan sammanfattas enligt:

$$[X] = [\overline{S}]^T \cdot \left( R - \frac{1}{c} [S] \cdot [X] \right) \quad (30)$$

eller alternativt:

$$[X] = [\overline{S}]^T \cdot R - [\mathbf{H}] \cdot [X] \quad (31)$$

enkel            förbättrad  
inkodning      inkodning

där  $[\mathbf{H}]$  representerar den hermitska matrisen enligt:

$$[\mathbf{H}] = \frac{1}{c} [\overline{S}]^T \cdot [S] \quad (32)$$

Den optimala avbildningen uppnås då  $[X]$  konvergerar mot en stabil minpunkt enligt:

$$[X] = [\mathbf{H}]^{-1} \cdot [X]_{enkel} \quad (33)$$

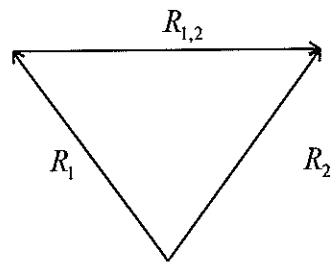
där  $[X]_{enkel}$  är korrelationsmatrisen enligt den enkla inkodningsproceduren (6). Beräkningen av inversen till den hermitska matrisen enligt (32) kräver emellertid att alla uppsättningar av stimuli är närvarande och är numeriskt

besvärlig då den i genomsnitt kräver ca  $M^3$  beräkningar vid eliminering enligt Gauss-Jordans metod. Detta gör att (32) inte används i praktiken.

## Illustration av metodens karakteristika

Sättet att använda differensen mellan avgiven och önskad respons (28) påminner lite grann om sättet att iterera feltermen bakåt (back-propagation) i konventionella nätverk.

Processen kan illustreras på följande sätt: Betrakta två olika stimulusmönster som är nästan identiska. Vid inkodning associeras dessa två nästan identiska stimuli med två olika responser. Responserna tillhörande stimulus 1 definieras av  $R_1 = e^{i\phi_1}$  och på motsvarande sätt representeras responserna tillhörande stimulus 2 av  $R_2 = e^{i\phi_2}$ . Här gäller alltså  $\phi_1 \neq \phi_2$ . Efter en inkodning av stimulus-respons association 1 ger en avkodning med stimulus 2 en respons som ligger nära  $R_1$ . Approximativt motsvarar differensen  $R_{diff}$ , som inkodas i nästa steg (dvs vid inkodning av mönster nr 2), differensen mellan  $R_1$  och  $R_2$ . Illustrerat med vektorer:



där alltså  $R_{diff} \approx R_{1,2} = \beta e^{i\phi_{1,2}}$

$$\text{och } \beta = 2 \sin\left(\frac{\phi_1 - \phi_2}{2}\right)$$

Om ordningen skiftas så att mönster 2 inkodas först och en efterföljande avkodning av mönster 1 görs så genereras motsvarande vektor ( $R_{2,1}$ ). Nedan beskrivs inkapslingen av avbildning 1 i korrelationselementet  $x_j$  givet att avbildning 2 inkapslats tidigare:

$$x_j = \overline{s_j} \cdot R_{diff} \quad (34)$$

eller ekvivalent

$$x_j = e^{-i\theta_{j,1}} \beta e^{i\phi_{1,2}} \quad (35)$$

På samma sätt då 2 inkodas givet att 1 inkodats tidigare:

$$x_j = e^{-i\theta_{j,2}} \beta e^{i(\phi_{1,2} + \pi)} \quad (36)$$

där:  $e^{i\theta_{j,t}}$  är element j i stimulusvektorn för mönster nr t  
 $\phi_{1,2}$  är argumentet för differensen  $R_{1,2}$

Detta innebär att för gemensamma regioner i de två stimulusmönstren tar bidragen till elementen,  $x_j$ , i korrelationsmatrisen i stort sett ut varandra:

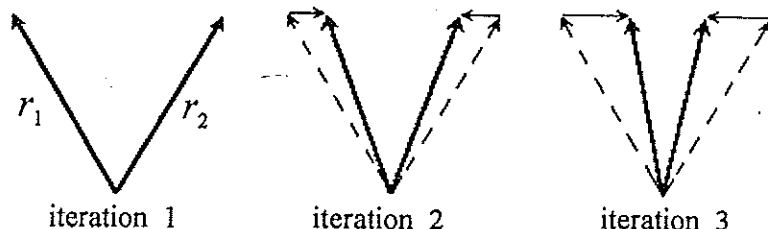
$$\left| \sum_{t=1,2} x_{j,t} \right| \rightarrow 0 \quad \text{för gemensamma regioner, dvs då } \theta_{j,1} \approx \theta_{j,2} \quad (37)$$

där  $x_{j,t}$  är korrelationsvärdet för element j och stimulusmönster t.

Kontentan blir att vid upprepad inkodning av stimulusvektor 1 och 2 konvergerar avbildningen av de gemensamma regionerna till en interpolation mellan  $R_1$  och  $R_2$ . Detta illustreras nedan där  $r_1$  och  $r_2$  betecknar responskomponenterna som erhållits vid avkodning över de gemensamma regionerna.

$$r_1 = x_j \cdot s_{j,1} \quad (38)$$

$$r_2 = x_j \cdot s_{j,2} \quad (39)$$

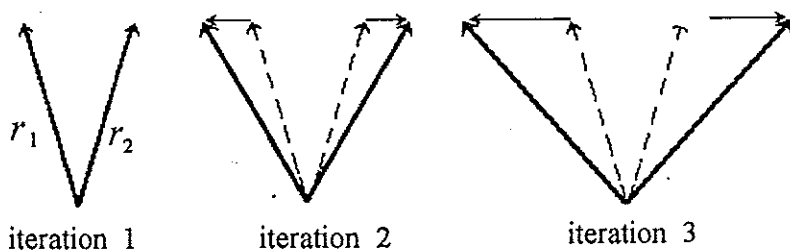


Figur 3 - Konvergens för gemensamma regioner ( $s_{j,1} \approx s_{j,2}$ )

Däremot gäller för regioner som inte är gemensamma att differenserna förstärker de associerade avbildningarna i korrelationsmatrisen. För icke-gemensamma regioner i stimulusvektorena (dvs  $\theta_{j,1} \neq \theta_{j,2}$ ) ger en inkodningskörning ett tillskott till korrelationsvärdet,  $x_j$ , av storleksordningen:

$$\left| \sum_{t=1,2} x_{j,t} \right| = 2\beta \sin\left(\frac{\theta_{j,1} - \theta_{j,2}}{2}\right) \quad (40)$$

Vid upprepad inkodning fås en motsvarighet till figur 3 enligt nedan:



Figur 4 - Konvergens för icke-gemensamma regioner ( $s_{j,1} \neq s_{j,2}$ )

Sammanfattningsvis så noteras att den förbättrade inkodningsproceduren dämpar avbildningar i korrelationsmatrisen som svarar mot gemensamma områden i stimulusvektorerna men som har olika tillhörande responsvärden. För sådan gemensamma regioner erhålles en interpolation mellan de önskade responserna. I det motsatta fallet, för icke gemensamma regioner, blir motsvarande element i korrelationsmatrisen förstärkta enligt figur 4. Härigenom kan systemet extrahera och förstärka karakteristika i stimuli genom upprepad inkodning. Detta ger en avbildning på korrelationsmatrisen som konvergerar stabilt för de flesta klasser av stimulus-respons associationer och som har förmåga att förstärka de drag som särskiljer olika stimulusmönster.

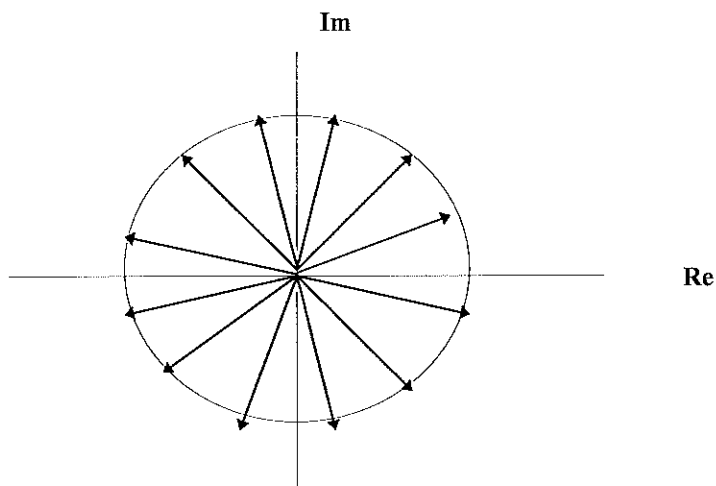
Noggrannheten i avbildningarna som genereras då högre ordningens element (förklaras ingående under rubriken "Högre ordningens system") används vid upprepad inkodning enligt ovanstående blir ännu bättre. Då HNeT Discovery Package används noteras att en stimulusvektor som utökas till ungefär 1000 högre ordningens element enkelt lär in 1000 slumpmässigt genererade stimulus-respons associationer. Efter ett litet antal (<4) upprepade inkodningar blir medelfelet i de återgivna responsen i regel mindre än  $\pm 1$  % av motsvarande önskade responser. Dessa empiriska resultat illustrerar HNeT-systemets förmåga att konvergera till en mycket bra svarsavgivning efter endast ett fåtal upprepade inkodningar.

## Vikten av symmetri

Symmetri i datamängdernas fördelningar är av största vikt i HNeT-processen. Förmågan att superponera ett flertal stimulus-respons associationer på en och samma uppsättning komplexa tal (korrelationsmatrisen) bygger på en summering av komplexa tal som är symmetriskt fördelade kring origo i det komplexa talplanet. Om kravet på symmetri är uppfyllt tenderar den totala förskjutningen p.g.a. den produkt som beräknas såväl vid inkodning som avkodning att öka på ett sätt som motsvarar en random walk. Allmänt ger en vektoriell summering av N st



steg av en längdenhet i en random walk en fördelning kring startpunkten för "vandringen" med ett väntevärde som går mot  $\sqrt{N}$  längdenheter. Figur 5 visar en idealt symmetrisk fördelning.



Figur 5 - Symmetriskt fördelade komplexa tal

En sådan symmetri är en garanti för att ett stimulusmönster som är helt skilt från de tidigare inkodade ger en avgiven respons motsvarande en vektoriell summering av steg i en random walk i det komplexa planet. Medelbeloppet för dessa random walks ger medelkonfidensen för icke-igenkända responser. På samma sätt ger tidigare inkodade stimulusmönster en respons vid avkodning med ett absolutbelopp eller konfidensvärde som ligger nära ett. Betrakta som exempel en HNeT-cell med  $M=1000$  synaptiska indata (vilket är lite för ett biologiskt neuron) och med  $N=250$  inkodade mönster. Då cellen exponeras för redan inkodade stimulusvektorer genereras responser med absolutbelopp i närheten av 1. Icke inkodade stimuli genererar däremot vid avkodning

responser med ett absolutbelopp av i medeltal  $\sqrt{\frac{N}{M}}$ . Insättningen av siffror enligt ovanstående ger ett konfidensvärde på 0.5. Detta värde används för att särskilja responser genererade av igenkända stimuli från responser genererade av icke-igenkända stimuli.

Den nödvändiga symmetrin erhålles på två olika sätt:

- 1) Användning av sigmoida transformer för att omvandla normalfördelningar eller approximativa normalfördelningar härstammande från de reellvärda obehandlade stimulus- och responsuppsättningarna till den komplexvärda form som används i HNeT-metoden.
- 2) Utökning av indata till högre ordningens element enligt metod som beskrivs nedan (se även "Högre ordningens system").

Bildandet av s.k. högre ordningens element definieras nedan:

$$\prod_{k=1}^K \lambda_k e^{i\theta_k} \quad \text{där } K \text{ definierar ordningen hos elementen} \quad (41)$$

Antalet olika kombinationer av K:te ordningen som kan bildas av en uppsättning data med M värden ges av:

$$\frac{(M + K - 1)!}{(M - 1)! K!} \quad (42)$$

På detta sätt kan uppsättningar av indata av liten storlek utökas till väldigt stora uppsättningar, vilket ökar antalet stimulus-respons associationer som kan inkodas med bibehållen noggrannhet vid responsavgivning. En annan effekt av utökningen är att fördelningen för den resulterande uppsättningen går asymptotiskt mot ett nästan optimalt symmetriskt tillstånd oberoende av den ursprungliga fördelningen. Symmetri kan uppnås för i stort sett alla fördelningar, med undantag för de exceptionella, patologiska, fallen (t.ex. ett singulärt tillstånd där alla informationselement är orienterade längs den positiva reella axeln.).

## Dynamiskt minne

I HNeT-processen definierar varje stimulus-response association en generaliseringsförmåga i den meningen att stimulusuppsättningar med en viss deterministisk likhet med inkodade stimulusmönster avbildas på önskad respons. Om en dynamisk dämpning appliceras på korrelationsmatrisen,  $[X]$ , observeras minneskaraktistik som innebär att konfidensvärdena hos tidigare inkodade associationer dämpas över tiden. Detta innebär också att det deterministiska felet eller bruset vid responsavgivning ökar p.g.a. dessa dämpade avbildningar. Avbildningar kan undertryckas eller dämpas kraftigt och "skrivas över" eller ersättas av nya avbildningar. Nedan definieras en form av minnesdämpning där en första ordningens dämpning appliceras på elementen i korrelationsmatrisen  $x_j \in [X]$ .

$$x_j = \frac{1}{c} \int_{t_0}^T \lambda_{j,i} \gamma_i e^{i(\phi_i - \theta_{j,i})} e^{-\frac{t-T}{\tau}} dt \quad (43)$$

där  $\tau$  är tidskonstanten för dämpningen  
 $T$  är aktuell tid

Med ovanstående definitioner gäller att långtidsminne eller permanent minne karakteriseras av liten dämpning (dvs stor tidskonstant  $\tau$ ) och korttidsminne eller närminne av en kraftigare dämpning. I HNeT implementeras dynamiskt minne genom att en övre gräns (ett tröskelvärde) för medelabsolutbeloppen i korrelationsmatrisen sättes. Detta tröskelvärde bestämmer när dämpning ska påbörjas. I HNeT är det förvalda värdet (defaultvärdet) på tröskeln 1.0, dvs långtidsminne erhålles. Detta värde kan emellertid modifieras av användaren varvid alla stadier mellan extremt närminne och permanent minne kan erhållas. Vid inkodning tills mättnad uppstår (dvs antalet inkodade slumpmässigt genererade mönster  $\gg$  antalet synaptiska indata eller storleken på cellen) gäller alltid följande för medelabsolutbeloppet i korrelationsmatrisen:

$$|x_j| \leq \sqrt{M} \quad (44)$$

där  $M$  är antalet synapser för stimuli (ev. med högre ordningens element) som tillförs cellen.

Minnesprofilen som erhålles vid ett lågt tröskelvärde ger allmänt lägre värde på absolutbeloppen (konfidensvärdena) hos avgivna respons för en större del av indata rummet, då dessa motsvarar icke-igenkända stimulus. Detta korttidsminne innebär att dämpningen startar vid relativt låga antal inkodade stimulus-respons associationer. Det uppvisar också mindre värden på felet i de avgivna responser som motsvarar nyligen inkodade stimuli.

Då permanent minne nås stabiliseras korrelationselementens absolutbelopp vid ett värde som ligger lite lägre än den teoretiska övre gräns som ges av (44), oberoende av antalet inkodade associationer. Med andra ord så möjliggör HNeT-processen en inkapsling av ett obegränsat antal godtyckliga stimulus-respons associationer i en korrelationsmatris. Felet i avgivna responser ökar i takt med antalet associationer och komplexiteten hos dessa. På motsvarande sätt blir konfidensvärdena vid långtidsminne för icke-igenkända stimuli större, vilket innebär en minskad möjlighet till särskiljning mellan "säkra" och "osäkra" svar.

## Högre ordningens system

Konventionella ANN-modeller har en viss begränsning vad gäller antalet associationer som med föreskriven noggrannhet kan inkodas i det neurala nätverket. I metoden som används i HNeT övervinns till viss del dessa begränsningar genom bildande av högre ordningens element eller produkter ur den ursprungliga uppsättningen indata  $[S]$ . Felet vid responsavgivning enligt (18) gäller även för det fall att stimulusuppsättningen har utökats till högre ordningens element under förutsättning att dessa högre ordningens element bildar åtskilda (unika) uppsättningar. Med unik uppsättning avses alla möjliga olika (oavsett kommutativ ordning) kombinationer av obehandlad indata.

Ett exempel är den Hermitska matrisen som ger en uppsättning av andra ordningens element. Denna kan beskrivas genom att representera en uppsättning stimuli som en radvektor och bilda:

$$[\mathbf{H}] = [\overline{S}]^T \cdot [S] \quad (45)$$

Diagonalen i denna matris innebär en redundans. Simuleringar visar att samma antal associationer kan inkodas med given svarsnoggrannhet om diagonalen utesluts. Antalet unika element som erhålles genom denna operation är:

$$N^2 - N \quad (46)$$

Metoden med Hermitska matriser kan generaliseras ytterligare genom att tillåta utökning till element av högre ordning ( $>2$ ). Ett indatafält utökat till 4:de ordningens element genereras t.ex. av följande produkt:

$$s_j = \prod_{k=1}^4 \lambda_{f(j,k)} e^{i\theta_{f(j,k)}} \quad (47)$$

där  $f(j, k)$  är någon godtycklig funktion som väljer ut element som funktion av k och j (index för aktuell synaps).

Denna operation ger både en rotation av argumentet och ett värde på konfidensen för det genererade elementet. En del av de individuella elementen  $\lambda e^{i\theta}$  kan konjugeras för att öka symmetrin, vilket ger bättre generaliseringsförmåga vid responsavgivning. Valet av element (vilket sköts av  $f(j, k)$ ) kan ske på ett "pseudo"-slumpmässigt sätt vilket ger en jämn påverkan från olika indata. Varje högre ordningens element motsvarar en "AND-operation" på elementen  $\lambda_{f(j,k)} e^{i\theta_{f(j,k)}}$  vilken ger konfidensvärdet för det utökade elementet. Detta koncept kan illustreras med följande exempel där som tidigare konfidensvärdet för element j representeras av absolutbeloppet av det komplexa elementet. Data med låg konfidens drar ned konfidensvärdet för motsvarande högre ordningens element. För fjärde ordningens element enligt (47) minskar resultatets konfidens för varje ingående låga värde på  $\lambda_1$  till  $\lambda_4$  enligt:

$$|s_j| = \lambda_1 \lambda_2 \lambda_3 \lambda_4 \quad (48)$$

Dessa högre ordningens element infogas i inkodningen enligt följande:

Enkel inkodning:

$$x_j = \sum_{l=1}^N \gamma_l e^{i\theta_l} \prod_{k=1}^K \lambda_{f(j,k)} e^{-i\theta_{f(j,k)}} \quad (49)$$

Avkodning:

$$r = \frac{1}{c} \sum_{k=1}^M x_j \prod_{k=1}^K \lambda_{f(j,k)}^* e^{i\theta_{f(j,k)}} \quad (50)$$

Här definierar K ordningen för de utökade elementen. Felet vid respons avgivning följer (18) även för element av högre ordning än ett under förutsättning att den utökade uppsättningen endast innehåller unika element enligt tidigare definition. Kravet på unika element sätter en övre gräns för antalet utökade element som kan erhållas för en given stimulusuppsättning och en given ordning. Denna gräns innebär i sin tur en begränsning av antalet associationer som kan inkodas med given noggrannhet.

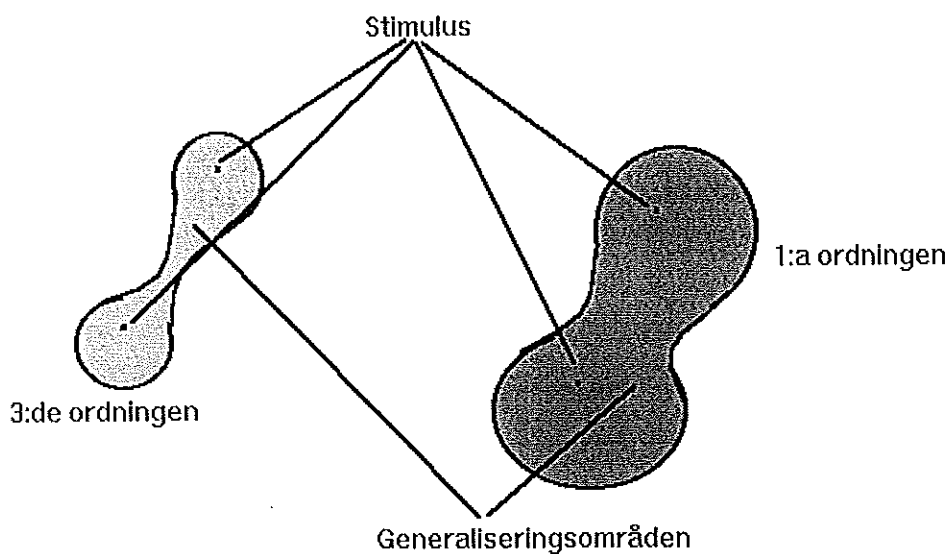
Tabell 7 visar antalet element som kan erhållas för uppsättningar av indata med 1 till 10 element. Konsekvenserna av resultatet enligt denna tabell är att ett stort antal stimulus-respons avbildningar kan

inkapslas i en korrelationsmatris (dvs en neuroncell) vid små uppsättningar av stimulus. Betrakta t.ex. ett litet stimulusfält med 20 element som utökas till 8:de ordningens element. I detta fall kan ungefär  $6 \cdot 10^6$  unika element bildas, vilket innebär att motsvarande antal unika avbildningar kan inkodas.

En annan observation är att området kring varje stimulus inom vilket generalisering kan ske krymper då ordningen på de utökade elementen ökar (se figur 6). Detta innebär således en möjlighet att kontrollera generaliseringsförmågan i neuroncellen. Utökning av stimulusfälten till högre ordningens element motsvarar i någon mening användningen av gömda lager i konventionella backpropagationnätverk.

	Ordning				
Stimulus storlek	1	3	5	7	10
1	1	1	1	1	1
5	5	35	126	330	1001
10	10	220	2002	11440	92378
15	15	680	11628	116280	1961256
20	20	1540	42504	657800	20030010

Tabell 7 - Antalet utökade element som funktion av ordningen och stimulusfältets storlek.



Figur 6 - Generaliseringsområdenas beroende av ordningen av utökade element.

## Bilaga 2

### Undersökning av HNeT Discovery Package

HNeT Discovery Package är en implementation av den holografiska metod för neurala nätverk som beskrivits i detta arbete. Programmet tillverkas av AND Artificial Neural Devices Corporation. Nedanstående resonemang bygger i stor utsträckning på den till programmet medföljande manualen [4]. Av praktiska skäl kan inte alla begrepp och facktermer förklaras här. Vid oklarheter hänvisas till manualen [4] till HNeT Discovery Package version 3.1 for Windows.

Ena delen av detta examensarbete går ut på att finna optimala inställningar för ett antal användardefinierade parametrar i programmet. Speciellt studeras inställningarna för den typ av data som kommer i fråga vid prediktion av OMX-index. Härvid användes försöksplanering enligt Taguchi-metoden. Avsikten med denna metod är att med ett mindre antal försök där flera parametrar ändras enligt ett visst mönster erhålla maximal kunskap om processen. Alternativet att variera endast en variabel per försök kräver väsentligt fler försök och kan även ge direkt missvisande resultat om parametrarna har ett inbördes beroende.

#### HNeTs parametrar

Nedan förklaras en del av de parametrar som ingår i HNeT Discovery Package.

#### Stimulus/response conversion

Denna parameter avgör vilken metod som används för att omvandla den obehandlade indatan (reella tal) till det komplexa format som används internt i HNeT. Det centrala är att fördelningen av de komplexa talen är symmetrisk i det komplexa talplanet. Alternativen är *Linear Conversion* och *Sigmoid Conversion*. För responsdata finns även alternativet *Real Conversion*. De olika metoderna beskrivs nedan.

*Linear Conversion* omvandlar data enligt följande formel:

$$\theta = \pi \left( 1 + c \left( \frac{s - \mu}{\sigma} \right) \right)$$

- där  $\theta$  är argumentet för det bildade komplexa talet
- $c$  är en användardefinierad koefficient
- $s$  är stimulus- eller responselementet.
- $\mu$  är väntevärdet för fördelningen av stimulus- eller responsdata
- $\sigma$  är standardavvikelsen för fördelningen av stimulus- eller responsdata

*Sigmoid Conversion:*

$$\theta = 2\pi \left(1 + e^{\frac{-c(s-\mu)}{\sigma}}\right)^{-1}$$

*Real Response Conversion:*

$$M = \frac{r - \mu}{\sigma}$$

där  $M$  är absolutbeloppet för det bildade komplexa talet. *Real Response Conversion* kan användas då svaret är binärt (typ ”upp”/”ned”, ”ja”/”nej”). Argumentet sätts antingen till 0 eller  $\pi$  beroende på om responsen är mindre eller större än  $\mu$ . Vid användning av denna omvandlingsmetod fås vid avkodning inte både svar och konfidensvärde. I stället motsvaras konfidensen, eller ”säkerheten”, av närheten till ett önskat svar. Om de riktiga svaren t.ex. definieras som  $\pm 1$  så är avgivna svar med högre absolutbelopp (dvs svar närmare +1 eller -1) ”säkrare” än svar än svar kring nollstrecket.

## Initial statistics

Denna parameter bestämmer antalet högre ordningens element som ska bildas. Enligt tidigare diskussion ger ett högre antal högre ordningens element möjlighet att inkoda ett större antal distinkta stimulus-respons associationer. Generaliseringsförmågan minskar emellertid i motsvarande grad.

## Optimization

Denna parameter gör att användaren automatiskt kan rensa bort inparametrar (inklusive de som är av högre ordning) för att därigenom effektivisera efterföljande inkodning. När *Optimization* aktiveras plockas de inparametrar bort som har gett upphov till absolutbelopp i korrelationsmatrisen som är mindre än en viss användardefinierad gräns.

## Learning Rate

Denna parameter bestämmer i vilken utsträckning ny information ska inkodas i korrelationsmatrisen. Praktiskt innebär det att endast en viss del av differensen  $R_{diff}$  i (28) inkodas. Ett lägre värde på *Learning Rate* minskar känsligheten för brus eller enstaka felaktiga värden i träningsdatan. Om det finns ett starkt

och tydligt samband mellan stimulus och respons kan ett högt värde på *Learning Rate* användas, varvid antalet inkodningar kan minskas. I annat fall bör ett lägre värde användas.

## Memory Profile

Här definieras dämpningen som minnet (dvs korrelationsmatrisen) ska utsättas för. Ett lägre värde på denna parameter innebär att stimulus-respons associationer som varit inkodade sedan länge får mindre inflytande över avgivet respons än de som inkodats nyligen. Vitsen med detta är att bättre resultat kan uppnås då man arbetar med en över tiden föränderlig process.

## Thermalize

Med denna parameter kan generaliseringsförmågan ökas genom att vitt brus tillförs träningsdatan. Ju högre värde på *Thermalize* desto mer tillfört brus.

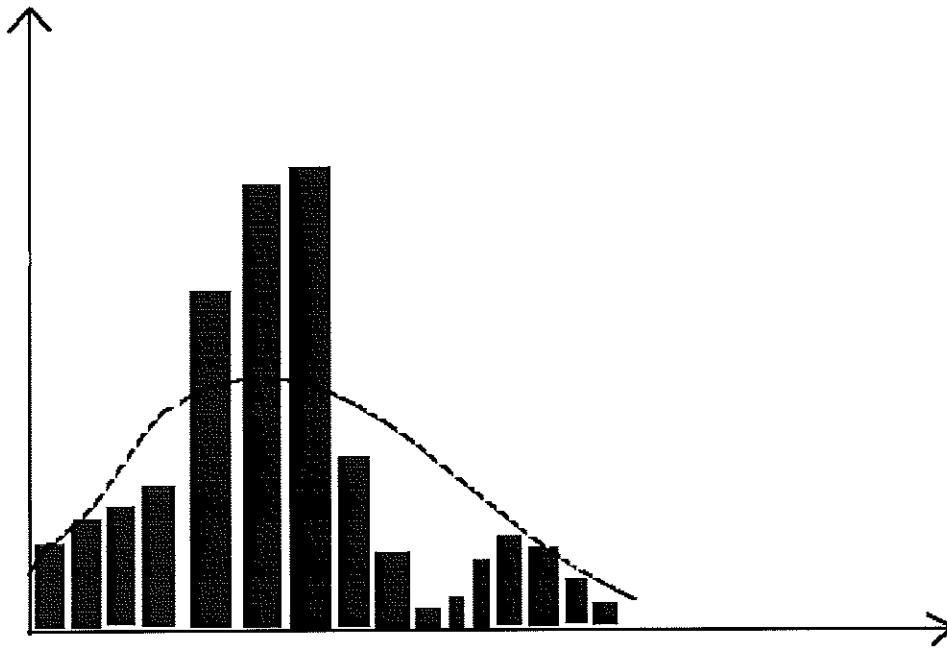
## Resultat

Nedan listas de resultat som erhållits efter gjorda undersökningar av HNeTs parametrar. I många fall stämmer mina resultat med de som återfinns i manualen [4] samt i [18]. En del nya saker och en del skillnader kan dock noteras.

## Stimulus/response conversion

De flesta av mina stimulus har en fördelning som påminner om normalfördelningen, varför *Sigmoid Conversion* ger bäst symmetri. Symmetrin förbättras ytterligare genom användning av *Dynamic Sample Interval* med värdet 20. För att undvika värden vid reella axeln sätts *Sigmoid Coefficient* till 1.1 - 1.2. Genom att gå in under *Distribution* och studera fördelningen och sedan skära bort lämpliga partier i träningsfilen erhålls ytterligare förbättringar (se nedanstående skiss).





En del records innehållande värde ovanför den streckade tänkta normalfördelningen skärs bort, vilket ger bättre symmetri då Sigmoid Conversion används.

Min modell bygger på att riktiga svaret är  $+1/-1$  motsvarande "upp"/"ner", varför *Real Response Conversion* är bästa val.

## Initial Statistics

Default-värdet visar sig ligga något för lågt i många fall. Här gäller regeln att ju bättre förklaringsvärde indata har desto större värde kan användas. Det gäller även att beakta på vilket sätt testfilen väljes. Pondera att man bygger upp sin träningsfil med kurshistorik för de senaste 10 dagarna presenterade i varje record. Om man skapar en testfil genom att slumpmässigt välja ut ca 20 % av träningsfilen och inte låta nätverket träna på dessa records kommer man att märka att ett väldigt stort värde på *Initial Statistics* (ofta maxvärdet) ger bäst resultat. Detta beror på att nätverket har sett två records som är väldigt lika den record som nätverket ska testköras på, nämligen recorden före och efter den borttagna (dagen innan och dagen efter). På grund av detta gynnas inlärningsprocessen av högre *Initial Statistics*, då det mer blir fråga om memorisering än generalisationsförmåga. Ett bättre sätt att träna nätverket är att välja de sista 10-30 procenten som testfil och när resultatet är tillfredsställande så gör man om träningsförfarandet med hela träningsfilen.

## Learning Rate

Samma slutsats som i [18] dvs lågt värde (ca 10) användes. Särskilt i ett sådant här fall, dvs prediktion av OMX-index, så är det extra viktigt att använda ett mycket lågt värde beroende på den stora mängd brus som ingår.

## Övriga parametrar

*Thermalize* bör inte användas, vilket är naturligt med tanke på den stora mängd brus som redan ingår i indata. Även när det gäller *Memory Profile* är resultatet att default-värdet (som ger långtidsminne) är bästa alternativ. Min utgångspunkt när det gällde *Shuffle* var att användning av detsamma endast kunde leda till bättre eller minst lika bra resultat. Detta har ej kunnat styrkas av mina undersökningar, dvs resultatet kan bli sämre av *Shuffle*. Precis samma sak gäller *Sigmoid Postprocessing*. Har man bra symmetri från början blir resultatet ofta sämre av *Sigmoid Postprocessing*. Ofta fås ett "instabilt" nätverk som "kastar sig" mellan olika värden. I vissa modeller blir resultatet bättre om *Initial Statistics* väljs stort från början och *Optimization* sedan används för att komma ner till ett värde som ger god generaliseringsförmåga.

## Brister

Rent generellt är HNeT Discovery Package väldigt användarvänligt och "buggfritt". Några fel och brister finns dock:

- Fördelningen av svaren i komplexa talplanet vid *Real Response Conversion* stämmer inte med manualens beskrivning av hur *Real Conversion* fungerar. Enligt denna ska alla värden ligga på reella axeln vilket de inte gör om en kontroll genom att gå in på *distribution* görs.
- Möjlighet att välja olika omvandlingsmetoder för olika indata saknas. Detta är den allvarligaste brist som kunnat konstateras, eftersom det på grund av detta inte går att "skräddarsy" omvandlingen och på så sätt nå optimal symmetri. Denna brist lär emellertid vara åtgärdad i "proffsversionen" av programmet, HNeT Professional Development System.