



Dr. rer. nat. Hans UHLIG, Jagersredder 27a, 22397 Hamburg,  
Tel. 040 / 605 51 50

Der folgende Artikel wurde in dieser Form der amerikanischen Finanzzeitschrift ‚Technical Analysis of Stocks and Commodities‘, Seattle, Washington, U.S.A. (auf Englisch) zur Publikation eingereicht und erschien dort in etwas gekürzter Form im Septemberheft 2001. Es wurden jedoch weder Textauszüge noch Bilder jenes Artikels hier verwendet.

## **Zusammenfassung**

---

In diesem Artikel wird die sogenannte ‚nearest neighbour‘ Vorhersagemethode als Prognoseinstrument für Finanzmärkte vorgestellt. Im deutschen Sprachraum vielleicht einigen bekannt unter der Formulierung: Vorhersage mit ‚nächsten Nachbarn‘. Als Beispiel dient der amerikanische Standard & Poor’s 500 Index, der für Portfolioverwalter in den U.S.A. auch allgemein als Vergleichsstandard (Benchmark) dient. Auf der Basis von Wochenschlusskursen von 1980 bis 2000, das sind etwa 1000 Werte, wurde für ein halbes Jahr jeweils die Indexänderung für eine Woche im voraus prognostiziert. Dabei kam es nur darauf an, ob die Änderung zu einem positiven oder zu einem negativen Ertrag führte.

Die fiktiven Erträge dieser Methode wurden mit der naiven Strategie ‚kaufen und behalten‘ verglichen. Es zeigte sich, dass die ‚nearest neighbour‘ Methode deutlich besser abschnitt, als die naive Strategie. Während der Index in der untersuchten Zeit deutlich verlor, konnte die hier vorgestellte Methode einen kleinen Gewinn verbuchen.

Der besondere Charme des Verfahrens besteht darin, dass die Vorhersagen vollständig transparent sind und dass der Anwender lediglich ein Tabellenkalkulationsprogramm benötigt, beispielsweise das bekannte Microsoft-Produkt Excel®, oder das völlig ebenbürtige, kostenlos erhältliche Kalkulationsprogramm aus dem OpenOffice® Paket.

---

## **Vorhersage mit nächsten Nachbarn ( k-nearest neighbour prediction)**

Die internationalen Aktienmärkte sind seit ihrem langen Höhenflug bis zum Frühjahr 2000 nicht mehr reine Einbahnstraßen, in denen ein Gewinnjahr auf das andere folgt. Vor dem Platzen der dot.com Blase war eine der besten Anlagestrategien "buy and hold" bzw. "kaufen und behalten". Dieses, im historischen Vergleich betrachtet, recht ungewöhnliche Marktverhalten hat zu einer weit verbreiteten Überschätzung der Methode „buy and hold“ und gleichzeitig zu einer Geringschätzung von zeitlich befristeten Marktengagements geführt. Inzwischen hat sich das Bild etwas geändert. Die Indizes erreichen nicht mehr immer neue Rekordhöhen, sondern sie brechen zum Teil dramatisch ein. In diesen Zeiten ist „kaufen und behalten“ nicht mehr die Investmentstrategie der Wahl, wie der vorurteilsfreie Leser sicherlich zugeben wird. Ein Blick auf andere Ansätze zur Geldanlage am Aktienmarkt erscheint daher ratsam.

Hier werde ich Ihnen zeigen, wie die Chaos-Theorie uns helfen kann, bessere Anlageentscheidungen zu treffen. Ich werde Sie mit einer Methode vertraut machen, die aus der Chaos-Forschung stammt und „nearest neighbour prediction“ bzw. „Vorhersage mit nächsten Nachbarn“ genannt wird. Im weiteren Verlauf werde ich diese Methode kurz NNP nennen.

Die Bezeichnung ‚Vorhersage mit nächsten Nachbarn‘ ist vielleicht etwas irreführend. Damit ist allgemein die Vorhersage der Entwicklung eines dynamischen Systems – hier ist es ein Markt – gemeint, ausgehend von einer bestimmten Situation, bzw. einem bestimmten Zustand, der vorher definiert sein muss. Dieser Marktzustand, der durch ein bestimmtes Muster von jeweils mehreren vorhergehenden Marktbewegungen charakterisiert ist, wird verglichen mit allen Marktzuständen (Mustern), die in der Vergangenheit aufgetreten sind. Von allen diesen Marktmustern sucht man sich die ähnlichsten heraus. Wie viele der ähnlichsten Muster man betrachtet, legt die Variable ‚k‘, siehe Titel, fest. Dann sieht man sich an, wie sich der Markt, ausgehend von diesen, besonders ähnlichen, Zuständen damals entwickelt hat und schließt daraus auf die künftige Entwicklung. Dazu wird über die Marktbewegungen, die den ähnlichsten Marktsituationen in der Vergangenheit folgten, gemittelt. Der Mittelwert dieser Bewegungen ist dann der Schätzwert für die erwartete Bewegung ausgehend von der aktuellen Marktsituation.

Wenn wir Begriffe aus der ‚technischen Analyse‘ verwenden, könnte man die NNP als eine Art von automatischer Chart-Interpretation bezeichnen. Hier wird auf der Grundlage von wöchentlichen Schlusskursen die NNP benutzt, um die Bewegungsrichtung des Aktienindex S&P 500 für eine Woche vorher zu sagen. Hätte man gemäß dieser Vorhersagen investiert, dann hätte man in den Jahren von Anfang 1996 bis Ende 2000 deutlich besser abgeschnitten, als mit der Methode „kaufen und behalten“, selbst in Zeiten, als „kaufen und behalten“ eine gute Strategie war. Diese Methode hat viele Vorzüge zu bieten: man benötigt nur ein Kalkulationsprogramm. Das Funktionsprinzip ist leicht zu verstehen. Die Entscheidungen sind immer durchschaubar und nachvollziehbar, was ein Vorteil ist gegenüber Neuronalen Netzen. Es gibt außerdem keine lange Trainingsphase. Zudem ist die Methode recht robust und wird von selbst immer besser, je mehr Daten mit der Zeit hinzukommen.

## **Lokale Prediktoren im Vergleich zu globalen**

Die NNP ist eine lokale Vorhersage, im Gegensatz zu den globalen Vorhersagen, mit denen die meisten von Ihnen vertraut sein werden. Eine Methode, die über alle verfügbaren Daten mittelt, um Vorhersagen zu treffen, ist eine globale Methode. Autoregressive integrierte gleitende Durchschnitte (ARIMA) sind Beispiele für globale Methoden. Anders gehen einige ‚Technische Analysten‘ vor. Sie sehen sich den aktuellen Kursverlauf an und suchen darin nach bestimmten Mustern, die ähnlichen Situationen in der Vergangenheit entsprechen und die sie deuten können. Für die Interpretation und Vorhersage werden also nur einander ähnliche Kursverläufe herangezogen. Vorhersagen aufgrund von Chart-Mustern sind Beispiele für lokale Vorhersagen.

## **Gewöhnliche Chart-Interpretation und NNP – Gemeinsamkeiten und Unterschiede**

Das Lesen von Charts hat zumindest zwei kritische Aspekte: erstens ist es eine visuelle Methode und somit liegt die Einstufung von Ähnlichkeiten im Ermessen des Betrachters, zweitens gibt es auch bisweilen Kursverläufe die keinem bestimmten Muster entsprechen. Dazu wäre dann eine klare Aussage nicht möglich. Dagegen ist der NNP-Ansatz eine numerisch quantitative Methode, bei der Muster systematisch definiert und analysiert werden. Zunächst müssen wir definieren, wie ein Muster aufgebaut sein soll, d.h. wie viele Daten ein Muster bilden und welche Daten es sein sollen, dann haben wir festzulegen, wie die Ähnlichkeit zwischen zwei beliebigen Mustern quantitativ gemessen werden soll. Wenn wir dies erledigt haben, können wir jedes andere Muster mit einem Referenzmuster vergleichen und wir können alle Muster gemäß ihrer Ähnlichkeit zum Referenzmuster sortieren. Schließlich sehen wir uns die ähnlichsten Muster in der Vergangenheit an und stellen fest, wie sich der Markt verhalten hat, nachdem diese Muster aufgetreten sind. Wir erwarten, dass die Marktentwicklung nach dem Auftreten des Referenzmusters so ähnlich sein wird, wie nach dem Erscheinen der ähnlichsten Muster in der Vergangenheit. Diese ähnlichen Muster sind die sogenannten „,k' nächsten Nachbarn“. Die Variable ‚k‘ bezeichnet die Anzahl der nächsten Nachbarn, die berücksichtigt werden. In der Wahl von ‚k‘ sind wir frei. Die Vorhersage aufgrund der nächsten Nachbarn erfolgt so, dass man den Mittelwert der Entwicklungen der nächsten Nachbarn bildet und dieser ist der Erwartungswert, bzw. der Vorhersagewert. Nun war schon oft von Mustern die Rede, aber was ist eigentlich ein Muster? Welche Daten und wie viele davon bilden ein brauchbares Muster? Hier kommt die nichtlineare Statistik und die Chaos-Theorie ins Spiel.

### **Was die Chaos-Theorie uns lehren kann**

Die Chaos-Theorie besagt, dass man die vollständige Dynamik, eines chaotischen Systems, das aus mehreren Komponenten besteht, rekonstruieren kann und dass man dazu nur die Zeitreihe einer einzigen Komponente dieses Systems benötigt. Wenn das betrachtete System ein Markt ist, würde man die Zeitreihe der Marktpreise benutzen, um daraus die Marktdynamik zu rekonstruieren. Das bedeutet, wir würden nur die Marktpreise benötigen, um daraus unsere Muster zu bilden, die wir für die nächsten Nachbarn brauchen. Es bleibt die Frage: Ist ein Markt ein chaotisches System? Die Antwort ist: möglicherweise. Da Märkte wie Zufallssysteme aussehen können, werden sie entweder wirklich zufallsgesteuert sein, siehe ‚random walk‘-Hypothese, oder sie sind nichtlinear (chaotisch). Natürlich wird niemand bestreiten, dass die Märkte zufälligen Einflüssen unterliegen, aber die innere Dynamik ist höchstwahrscheinlich nicht zufällig, wie man mit etablierten Methoden der nichtlinearen Statistik überzeugend nachweisen konnte. Auch neuere Methoden, die von Chaos-Forschern entwickelt wurden, konnten dies bestätigen.

### **Bewahren der Information, die in den Daten enthalten ist**

Zwei Tests zum Nachweis nichtlinearer Beziehungen zwischen Daten sind besonders nützlich: der Chi<sup>2</sup>-Test auf Unabhängigkeit und der Test auf bedingte Entropie, bzw. auf Transinformation. Beide sind Statistikern wohl bekannt, aber Sie werden sie vielleicht nicht in Ihrem Kalkulationsprogramm finden. Mit dem Chi<sup>2</sup>-Test auf Unabhängigkeit können wir herausfinden, ob der Markt ein ‚Gedächtnis‘ hat und wie signifikant die Abhängigkeit von früheren Ereignissen ist. Durch das Messen der bedingten Entropie oder der Transinformation können wir bestimmen, wie viel Information im Durchschnitt über die zukünftige Entwicklung in den Daten der Gegenwart und der Vergangenheit enthalten ist und wie schnell diese Information mit zunehmendem Zeitabstand abnimmt. Wenn wir dies wissen, so kann uns das helfen, zu entscheiden, welche und wie viele Daten wir benötigen um daraus ein Abbild des Marktzustandes (Muster) zu bilden, das genügend Information enthält. Wenn Sie keinen Zugang zu Programmen besitzen, die solche Art von nichtlinearer Statistik berechnen können – und das wird die Regel sein - dann müssen Sie die geeigneten Daten durch Versuch und Irrtum herausfinden. Wenn Sie Glück haben und diese Programme nutzen können, sollten Sie bedenken, dass dies globale Methoden sind und nicht angepasst für lokale Prediktoren. Sie können nur grobe Schätzungen für brauchbare Daten abgeben.

## Definieren eines Musters

In vielen Vorversuchen, die ich auf diesem Gebiet durchgeführt habe, hat sich gezeigt, dass es günstig ist, mit den natürlichen Logarithmen des Index zu arbeiten und dass vier Daten pro Muster genügen, um den Marktzustand aufgrund von Wochenkursen zu jeder Zeit zu charakterisieren. Für den S&P 500 wurden wöchentliche Differenzen der Logarithmen von Wochenschlusskursen für die Muster verwendet und zwar die von 1, 2, 4 und 5 Wochen früher als die Vorhersageweche.

## Messen der Ähnlichkeit

Die Ähnlichkeit wird gemessen durch Berechnung des absoluten Abstandes, bzw. des Betrages des Abstandes. Diese Art der Abstandsmessung wird auch als Manhattan Metrik oder City Block Norm bezeichnet. Der Abstand aller anderen Muster im Vergleich zum Referenzmuster (siehe die Formel in Feld H16 des Bildschirmabdrucks) wird berechnet. Die Bezeichnung , abs(.) , bewirkt, dass jede einzelne Differenz ein positives Vorzeichen bekommt. Dadurch vermeidet man, dass sich Differenzen mit verschiedenem Vorzeichen gegenseitig aufheben können. Das , \$ , in der Formel bezeichnet einen absoluten Zellbezug. Wenn man es weglassen würde, so würde nur ein Muster mit dem Referenzmuster verglichen werden und alle anderen Muster nur miteinander. Was ist mit anderen Abstandsmaßen, werden Sie vielleicht fragen. Es gibt sie und sie werden an anderer Stelle noch diskutiert werden.

date	S&P500	ln(S&P500)	week 5	week 4	week 2	week 1	"distance"	week 0	week 0
20001215	1312,15	7,1794223	0,0013	-0,019	0,0407	-0,043			
20001208	1369,89	7,2224857	-0,043	0,0013	-0,02	0,0407	0	-0,04306	0,018238
20001201	1315,23	7,1817668	0,0336	-0,043	-0,019	-0,02	0	0,040719	0,001243
20001124	1341,77	7,2017449	-0,012	0,0336	0,0013	-0,019	0	-0,01998	0,006843
20001117	1367,72	7,2209004	0,0164	-0,012	-0,043	0,0013	0	-0,01916	0,043938
20001110	1365,98	7,2196274	-0,025	0,0164	0,0336	-0,043	0	0,001273	0,027039
20001103	1426,69	7,2631124	-0,019	-0,025	-0,012	0,0336	0	-0,04348	-0,00744
19860201	211	5,3518581	-0,009	-0,024	-0,01	0,024	0,023222315	0,014118	
19970906	929,05	6,8341626	-0,014	-0,036	-0,026	0,0324	0,030689035	-0,00555	
19880130	257	5,5490761	-0,02	-0,016	-0,024	0,0437	0,031159919	-0,02762	
19941015	469	6,1506028	0,0064	-0,026	-0,015	0,0303	0,032560111	-0,01072	
19941217	458	6,1268692	-0,002	-0,02	-0,016	0,0266	0,032586983	0	
19990621	1342,84	7,2025421	-0,006	-0,022	-0,026	0,0373	0,034395639	-0,02071	
19840224	163	5,0937502	-0,012	-0,024	0	0,0186	0,036359412	0	
19971122	963,09	6,8701469	-0,003	-0,029	0,0009	0,0367	0,027325137	-0,00802	

**Bild 1 – Kalkulationsblatt mit NNP-Methode**

Siehe Beschreibung in den Texten im vorigen und nachfolgenden Absätzen

## Sortieren der Muster nach ihrem Abstand

Nachdem man die Abstände aller Datensätze (Muster) in der Datenbank relativ zum Referenzdatensatz bestimmt hat, werden die Datensätze nach ihren Abständen zum

Referenzmuster sortiert. Die Datensätze, die nach der Sortierung am dichtesten beim Referenzmuster stehen, sind die nächsten Nachbarn. Wir wissen, wie sich die nächsten Nachbarn entwickelt haben, das heißt, welche Marktbewegung dem jeweiligen Muster folgte. Alles was wir tun müssen, ist, den Mittelwert dieser jeweils folgenden Marktbewegung von allen nächsten Nachbarn zu bilden und wir haben eine Schätzung, wie die Marktbewegung aussehen wird, die dem Referenzmuster folgt. Grundsätzlich ist man frei in der Wahl, wie viele nächste Nachbarn man berücksichtigen sollte. Ich habe herausgefunden, dass vier nächste Nachbarn ein guter Ausgangswert sind. Darum ist die Variable  $k=4$  ( bei  $k$  nächste Nachbarn). Wenn man sich auf einen Wert festgelegt hat, sollte man dabei bleiben.

## **NNP beim S&P500**

Der erste Bildschirmabdruck zeigt eine Beispielvorhersage für den S&P500. Als Datenbank benutzte ich die Wochenschlusskurse der Zeitspanne von Jan 1980 bis Dezember 2000, die ungefähr 1100 Daten umfassen. Die Muster, die verwendet wurden, sind in den Spalten D bis G gezeigt. Spalte I zeigt, wie sich der Markt bewegt hat in der Woche, die nach dem Auftreten des Musters folgte. Diese Spalte ist nicht Teil des Musters und sie wird daher auch nicht benutzt, um den Abstand zum Referenzmuster zu bestimmen. Spalte I wird aber benötigt, um eine Schätzung (Vorhersage) für die Entwicklung des Marktes abzugeben, die dem Auftreten des Referenzmusters folgt. In der markierten Zelle (H16) steht die Formel für die Abstandsberechnung. Diese Formel wurde in die Spalte H von allen Datensätzen vor dem 3. November 2000 kopiert. Und dann wurden diese Datensätze (Muster) gemäß ihrem Abstand aufsteigend sortiert. Der ähnlichste Datensatz steht in Reihe 17 und die nächsten drei weiteren nächsten Nachbarn in den Reihen 18 bis 20. Spalte A zeigt uns, wann diese ähnlichsten Muster in der Vergangenheit aufgetreten waren: 1986, 1997, 1988 und 1994, also zeitlich weit zurück. Das ähnlichste Muster ist in Zeile 17 und in Zelle I17 sehen wir, wie sich der Markt nach dem Auftreten des Musters entwickelt hat. Damals stieg der Kurs des Index. Aber bei den drei weiteren nächsten Nachbarn verlief die Entwicklung anders. Dadurch ist der Mittelwert aus den Zellen (L17:L20) eine Abwärtsbewegung von  $-0.0074$  wie es in Zelle J16 errechnet wurde. Dieses Beispiel zeigt den Vorteil, den es bietet, wenn man über mehrere nächste Nachbarn mittelt, um die zu erwartende Entwicklung abzuschätzen. Siehe dazu auch die Diskussion zur Robustheit, weiter unten.

## **Ergebnisse**

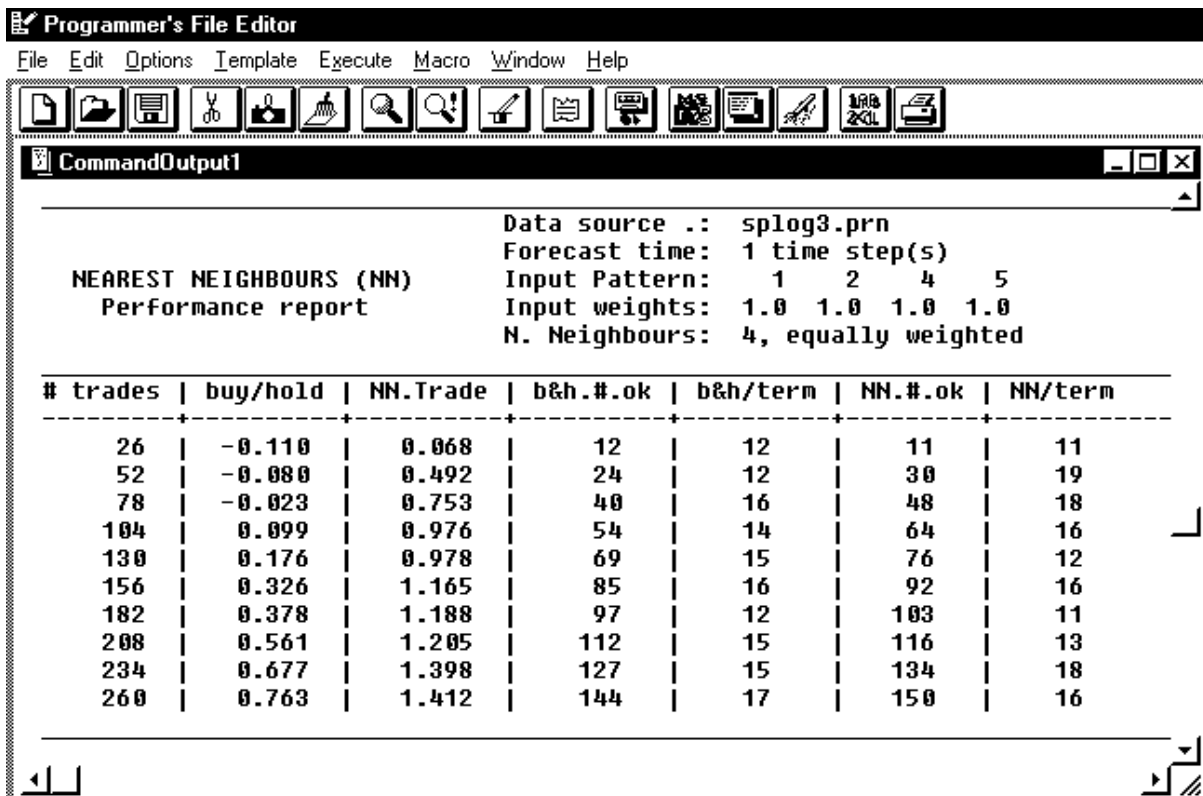
Die Methode, die hier vorgestellt wird, ist die einfachste Form der NNP. Aber, wenn sie richtig durchgeführt wird, funktioniert sie ohne weitere Verbesserungen und kann die Methode „kaufen und halten“ übertreffen. Ich habe die NNP für die letzten 26 Wochen des Jahres 2000 mit Hilfe eines Kalkulationsprogramms durchgeführt. Die Daten sind im Block L1 : N8 des abgebildeten Ausschnitts eines Kalkulationsblattes zusammengefasst. In den Feldern L5:N5 ist die Bilanz der letzten 26 Wochen, also eines Halbjahres zu sehen. Spalte M enthält die Ergebnisse der Methode „kaufen und halten“. Ein Verlust von knapp 11% wäre erzielt worden. Die Spalte N zeigt die mit dem NNP erreichbaren Ergebnisse. Für das letzte Halbjahr wäre ein kleiner Ertrag von 6% erzielbar gewesen. In den nachfolgenden Feldern der Spalten sind Einzelergebnisse für die verschiedenen Wochen angegeben. So kann man jede einzelne Entscheidung nachlesen. In den Feldern M8 und N8 sieht man, dass die NNP zwar nur elf Mal die richtige Marktrichtung vorhergesagt hat, und dennoch im Mittel einen leichten Gewinn erzielt hätte, während die Methode „kaufen und behalten“ zwölf Mal richtig gewesen wäre, aber dennoch zu einem geringen Verlust geführt hätte.

	H	I	J	K	L	M	N	O	P	Q
1	Trade statistics									
2										
3	Method: buy & hold NNP									
4	-----									
5	Differences	Moves to be predicted			Result	-0,10982	0,060191			
6	between	actual	expected		Std.Dev.	0,0233328	0,023609			
7	patterns	-----	-----		# trades	26	26		b&h.ok	NN.ok
8	"distance"	week 0	week 0		correct	12	11		12	11
9	-----									
10	0									
11	0	-0,04306	0,018238		1	-0,043063	-0,04306		0	0
12	0	0,040719	0,001243		2	0,0407189	0,040719		1	1
13	0	-0,01998	0,006843		3	-0,019978	-0,01998		1	1
14	0	-0,01916	0,043938		4	-0,019155	-0,01916		1	1
15	0	0,001273	0,027039		5	0,001273	0,001273		2	2
16	0	-0,04348	-0,00744		6	-0,043485	0,043485		2	3
17	0	0,033578	0,001646		7	0,033578	0,033578		3	4
18	0	-0,0125	0,010138		8	-0,012498	-0,0125		3	4
19	0	0,016427	0,002451		9	0,0164271	0,016427		4	5
20	0	-0,02502	-0,00258		10	-0,025023	0,025023		4	6
21	0	-0,01934	0,014253		11	-0,019343	-0,01934		4	6
22	0	-0,00846	0,000433		12	-0,008464	-0,00846		4	6
23	0	-0,01173	0,024772		13	-0,011728	-0,01173		4	6
24	0	-0,01938	-0,00612		14	-0,019384	0,019384		4	7

**Bild 2 – Kalkulationsblatt mit Tabelle der NNP Handelsstatistik**

Die ‚Handelsstatistik‘ ist in den Spalten L ..N, Reihen 1 ..8 zu sehen. Hier sehen Sie einen kurzen Leistungsvergleich zwischen NNP und ‚kaufen und halten‘, einschließlich Standardabweichungen der Erträge, die beinahe identisch sind. Sie unterscheiden sich erst in der vierten Dezimalstelle.

Ein Test mit so wenig Daten kann kaum als repräsentativ angesehen werden, aber andererseits wäre es mit großem Zeitaufwand verbunden und sicher auch fehlerträchtig, wollte man eine große Zahl von Tests mit einem Kalkulationsprogramm durchführen. Daher habe ich die Brauchbarkeit der Methode nicht nur für 26 Wochen, sondern über die letzten zwölf Jahre untersucht. Dazu habe ich ein Programm verwendet, das ich selbst geschrieben habe. Es ist in der Sprache ‚C‘ geschrieben und kann auch auf älteren Computern in wenigen Sekunden riesige Datenmengen verarbeiten. Es ist ein reines Analyseprogramm und nicht besonders anwenderfreundlich. Einige Bildschirmausdrucke von Berichten für die letzten fünf Jahre (1996-2001) für verschiedene Parameter (z.B. Anzahl nächster Nachbarn, Anzahl und Struktur der Daten für die Muster, Gewichtung der Eingaben und der einzelnen nächsten Nachbarn) sind auf den folgenden Seiten zu sehen. Das Problem bei Langzeitstudien ist, dass die Datenbasis für die Suche nach nächsten Nachbarn immer kleiner wird, je weiter man in die Vergangenheit blickt, weil nur solche Daten, die älter sind als das Referenzmuster für die Suche nach ähnlichen Marktsituationen verwendet werden.



**Bild 3 – Ausgabe des NNP Programms**

„Ertragsstatistik“ für die NNP Methode mit den gleichen Parametern, wie jene, die im Kalkulationsblatt verwendet wurden, aber hier für die letzten Jahre 1996 bis 2000 einschließlich. Erläuterung der Spalten von links nach rechts: Number of Trades, d.h. die Anzahl wöchentlicher Entscheidungen, nächste Spalte: Ergebnis von ‚kaufen und halten‘, nächste Spalte: Ergebnis der NNP Vorhersagen, nächste Spalte: wie oft ‚kaufen und halten‘ richtig war, bezogen auf die Anzahl der Entscheidungen (Spalte ganz links), nächste Spalte: wie oft ‚kaufen und halten‘ richtig war im jeweils betrachteten Halbjahreszeitraum. Die nächsten beiden Spalten geben die entsprechenden Daten für die NNP an. Die Reihen von oben bis unten geben die summierten Ergebnisse für die Anzahl der Entscheidungen an. Ausgenommen davon sind die Spalten mit den Überschriften ‚b&h/term‘ und ‚NN/term‘

Wenn Sie die Statistik-Tabelle in Bild 2 und die erste Reihe in dieser Tabelle vergleichen, dann sehen Sie eine kleine Abweichung (weniger als 1%) zwischen den beiden Ausgaben. Dies kommt daher, dass im Kalkulationsblatt die Logarithmen berechnet und mit der vollen Zahl von 18 Dezimalstellen auch verwendet wurden. Dagegen hatte die Datei mit den Datensätzen für das NNP Programm nur Logarithmen  $\ln(\text{Index})$  mit maximal 6 Dezimalstellen für jede Zahl. Der Rundungsfehler ist unvermeidlich aber meistens sehr gering, aber bisweilen kann dies zu veränderten nächsten Nachbarn führen und somit größere Unterschiede bewirken. Diese werden sich aber in der Regel gegenseitig aufheben.

### Robustheit der Methode

Die Methode scheint recht robust zu sein, das bedeutet rauschtolerant und unempfindlich gegenüber geringfügigen Abweichungen. Ein Muster besteht aus vier Daten, so dass ungewöhnliche Einzelwerte das Muster nicht allzu sehr verändern können. Auch das Verwenden der absoluten Differenz als Distanzmaß hilft dabei, die Methode robust zu halten (siehe dazu die Diskussion weiter unten). Die Entscheidung über die künftige Markttrichtung wird getroffen, indem man den Mittelwert der Entwicklungen von den vier ähnlichsten Situationen in der Vergangenheit bildet. Auch das verringert den Einfluss möglicher Ausnahmewerte.

### Modifikationen bei der Bestimmung der nächsten Nachbarn

Man kennt mehrere Abwandlungen der Methode und einige davon können wirklich Verbesserungen bewirken.

Programmer's File Editor

File Edit Options Template Execute Macro Window Help

CommandOutput1

```

Data source .: splog3.prn
Forecast time: 1 time step(s)
Input Pattern: 1 2 4 5
Input weights:
N. Neighbours: 4, equally weighted

```

NEAREST NEIGHBOURS (NN)  
Performance report

# trades	buy/hold	NN.Trade	b&h.#.ok	b&h/term	NN.#.ok	NN/term
26	-0.110	0.055	12	12	10	10
52	-0.080	0.382	24	12	28	18
78	-0.023	0.610	40	16	44	16
104	0.099	0.880	54	14	62	18
130	0.176	0.882	69	15	74	12
156	0.326	1.120	85	16	89	15
182	0.378	1.138	97	12	99	10
208	0.561	1.198	112	15	115	16
234	0.677	1.399	127	15	133	18
260	0.763	1.518	144	17	152	19

**Bild 4 – Ausgabe des NNP Programms**

Dieser Ausdruck unterscheidet sich von dem in Bild 3 nur dadurch, dass die Eingaben nicht alle gleich, sondern individuell gewichtet waren. (Die Angaben der Werte wurden nachträglich gelöscht, weil ich nicht alles verraten möchte).

Eine mögliche Verfeinerung wäre es, die verschiedenen Eingabewerte für das Muster unterschiedlich zu gewichten, siehe Bild 4. Vielen wird es logisch erscheinen, dass die jüngste Marktbewegung wichtiger sein könnte für die kommende Bewegung, als die Bewegung vor, sagen wir einmal, drei Wochen. Aber meistens ist die jüngste Bewegung nicht so wichtig, wie man meint und die früheren Bewegungen haben mehr Einfluss als viele Anhänger der ‚random walk‘-Hypothese – glauben. Dennoch können individuell angepasste Gewichte die Vorhersagen verbessern.

Eine mögliche Modifikation betrifft die Gewichtung der nächsten Nachbarn. Bisher haben wir einen einfachen Mittelwert über die vier nächsten Nachbarn gebildet. Das heißt, der nächste Nachbar bekommt genauso viel Gewicht bei der Berechnung der wahrscheinlichen künftigen Entwicklung, wie der Nachbar an vierter Stelle. Die Methode kann verbessert werden, indem man jeden Nachbarn proportional zu seiner Ähnlichkeit mit dem Referenzmuster, bzw. umgekehrt proportional zu seinem Abstand davon gewichtet. Das heißt, je ähnlicher die Marktsituation in der Vergangenheit, desto mehr Gewicht bekommt sie. Ein Beispiel dafür ist in Bild 5 gezeigt. Die Gewichte müssen natürlich normalisiert werden (jedes einzelne so berechnete Gewicht muss durch das Gesamtgewicht geteilt werden).



Programmer's File Editor  
File Edit Options Template Execute Macro Window Help

CommandOutput1

```

Data source .. splog3.prn
Forecast time: 1 time step(s)
NEAREST NEIGHBOURS (NN) Performance report
Input Pattern: 1 2 4 5
Input weights: 1.0 1.0 1.0 1.0
N. Neighbours: 4, distance weighted

```

# trades	buy/hold	NN.Trade	b&h.#.ok	b&h/term	NN.#.ok	NN/term
26	-0.110	-0.001	12	12	13	13
52	-0.080	0.243	24	12	30	17
78	-0.023	0.485	40	16	48	18
104	0.099	0.790	54	14	65	17
130	0.176	0.793	69	15	77	12
156	0.326	0.980	85	16	93	16
182	0.378	1.009	97	12	104	11
208	0.561	0.991	112	15	115	11
234	0.677	1.232	127	15	135	20
260	0.763	1.182	144	17	149	14

**Bild 5 – Ausgabe des NNP Programms**

Diese Ausgabe zeigt, wie sich die Gewichtung der nächsten Nachbarn nach ihren Abständen bzw. ihrer Ähnlichkeit auswirkt. Sonst wurden die gleichen Parameter, wie bei Bild 3 verwendet.

Die beiden eben angesprochenen Modifikationen lassen sich auch miteinander kombinieren, wie es im folgenden Beispiel, in Abbildung 6 durchgerechnet wurde. Nicht immer führen solche Modifikationen zu verbesserten Ergebnissen.

Programmer's File Editor  
File Edit Options Template Execute Macro Window Help

CommandOutput1

```

Data source .. splog3.prn
Forecast time: 1 time step(s)
NEAREST NEIGHBOURS (NN) Performance report
Input Pattern: 1 2 4 5
Input weights:
N. Neighbours: 4, distance weighted

```

# trades	buy/hold	NN.Trade	b&h.#.ok	b&h/term	NN.#.ok	NN/term
26	-0.110	0.084	12	12	13	13
52	-0.080	0.231	24	12	29	16
78	-0.023	0.481	40	16	46	17
104	0.099	0.833	54	14	65	19
130	0.176	0.835	69	15	77	12
156	0.326	1.073	85	16	92	15
182	0.378	1.098	97	12	102	10
208	0.561	1.208	112	15	120	18
234	0.677	1.458	127	15	140	20
260	0.763	1.563	144	17	158	18

## **Bild 6 – Ausgabe des NNP Programms**

Ein Beispiel für abstandgewichtete nächste Nachbarn und gewichtete Eingaben.

Auch die Beschränkung auf die jeweils vier nächsten Nachbarn ist nicht verbindlich. Man könnte genau so gut drei oder fünf nächste Nachbarn verwenden. Ein Vorteil bei drei nächsten Nachbarn wäre, dass diese dichter am Referenzwert wären, und somit mehr Ähnlichkeit mit diesem gewährleistet wäre. Andererseits trägt das Einschließen eines weiteren Nachbarn zur Robustheit des Systems bei, die ebenfalls eine gewünschte Eigenschaft ist, siehe unten.

Es gibt noch Spielraum für weitere Verbesserungen an der NNP, wie sie hier beschrieben wurde. Bisher haben wir nur einen kleinen Teil der verfügbaren Daten berücksichtigt. Tageskurse könnten verwendet werden und längere Zeitreihen könnten ebenfalls zu einer weiteren Verbesserung beitragen. Sie werden vielleicht an ein anderes Distanzmaß denken, wie z.B. die Euklidische Metrik (Wurzel aus der Summe der quadrierten Einzeldifferenzen). Ich habe sie beim S&P500 ausprobiert, aber sie war nicht so zufriedenstellend, wie die Manhattan Metrik. Erstens war das Verhältnis richtige : falsche Entscheidungen schlechter. Und wegen der Berechnungen zweiter Ordnung (Quadrierung) bei der Euklidischen Metrik, waren die Ergebnisse weniger stabil. Das führte zu größeren Abweichungen in aufeinander folgenden Halbjahreszeiträumen.

Der hier vorgestellte NNP ist ein lokaler Prediktor nullter Ordnung, d.h. Der Zielwert ist nicht durch eine Funktion mit den Eingabewerten verknüpft. Es wird einfach der jeweilige Folgewert für jeden der nächsten Nachbarn als wahrer Zielwert angenommen und darüber gemittelt. So wie hier wurde auch von Farmer und Sidorowich verfahren. Es wäre auch möglich, die Eingabewerte und den Zielwert über eine Funktion miteinander zu verknüpfen. Beispielsweise könnte man bei allen 'k' nächsten Nachbarn die Eingabewerte mit Linearfaktoren so gewichten, dass die Funktionswerte möglichst geringe Abstände von den Zielwerten (den Folgewerten) erhalten. Diese, durch lineare Optimierung gefundenen Faktoren würde man dann auf das Referenzmuster anwenden und so den wahrscheinlichen Zielwert des Referenzmusters bestimmen. Auf diese Art wurden die nächsten Nachbarn als Prediktor erster Ordnung (d.h. In einer linearen Funktion) von Sugihara und May mit den Zielwerten verknüpft. Von Martin Casdagli, der viel mit nächsten Nachbarn experimentiert hat, wurden Prediktoren erster, zweiter und höherer Ordnung systematisch getestet. Er fand heraus, dass Prediktoren erster und zweiter Ordnung brauchbar waren, Funktionen höherer Ordnung dagegen keine Verbesserung brachten. Bei verrauschten Werten, mit denen man es hier zu tun hat, sind Prediktoren zweiter Ordnung bereits weniger geeignet, wie meine eigenen Versuche gezeigt haben. Zwar kann man durch die zweite Ordnung bessere Anpassung an Daten bekommen, doch durch die quadratische Funktion besteht bei den Werten zwangsläufig eine Tendenz zur Instabilität. Das bewirkt eine unerwünschte Einschränkung der Robustheit. Wägt man den möglichen Vorteil der besseren Angleichbarkeit an die Daten gegen die Nachteile der geringeren Robustheit und nicht zu vergessen, die aufwendigere Berechnung gegeneinander ab, so ist ein Prediktor erster Ordnung einem solchen zweiter Ordnung vorzuziehen.

Eine weitere mögliche Modifikation kann darin bestehen, dass man nicht Rohdaten verwendet, wie hier geschehen, sondern zunächst die Daten glättet, wie ich es beispielsweise für den ‚robusten Trend‘ durchgeführt habe, siehe dazu Versicherungswirtschaft Nov. 2001. Die Glättung führt dazu, dass aufeinander folgende Daten gemeinsame Informationen enthalten, so dass für geglättete Daten ganz andere Muster verwendet werden müssen. Dafür kann man jedoch mit weniger Daten für ein einzelnes Muster auskommen. Beim S&P500 konnte ich mit drei geglätteten Werten pro Muster auch die besten Modifikationen mit ungeglätteten Werten übertreffen, ohne Abbildung.

## **Anwendbarkeit des NNP auf verschiedene Märkte**

Sie werden sich fragen, ob die Methode auch auf andere Aktienmärkte, auf einzelne Aktien oder auf andere Finanzmärkte ganz allgemein angewendet werden kann. Nun, um ehrlich zu sein, gehört der S&P500, wegen seiner Marktbreite, zu den einfacheren Aufgaben. Eine gute Musterdefinition kann bei anderen Märkten anders aussehen, als in diesem Beispiel und das

gleiche kann für die anderen Parameter gelten. Die NNP kann auch zum Schätzen der Entwicklung anderer internationaler Aktienindizes (Nikkei 225, FTSE 100) erfolgreich eingesetzt werden. Sogar die Entwicklung des Ölpreises oder des Wechselkurses von Yen/US-\$ für die Jahre 1996 bis 2000 einschließlich, ließen sich mit der NNP recht gut vorhersagen. Dies sind nur zwei Beispiele aus den Rohstoff- und Währungsmärkten, aber die Anwendung der NNP ist keineswegs darauf beschränkt und sie kann auch andere Strategien als „kaufen und halten“ übertreffen.

### **Ratschläge zur Verwendung des beiliegenden Kalkulationsblattes**

Wer den NNP selbst ausprobieren möchte, der findet anbei ein Kalkulationsblatt mit den nötigen Formeln zur Berechnung. Auf dem Blatt sind die Daten des S&P500 eingetragen, damit man das Beispiel nachvollziehen kann. Man kann sich das Kalkulationsblatt auch kopieren und die Datensätze S&P500 gegen solche des DAX, des ATX, des SMI oder von Währungspaaren, wie US-\$/Euro oder US-\$-/Schweizer Franken austauschen. Mit den Mustern müsste man jedoch herumprobieren, um brauchbare Ergebnisse zu erzielen. Man sollte nicht erwarten, dass die optimale Zusammensetzung der Muster immer gleich ist. In die Spalten D bis G und I muss man Werte eintragen, nicht Formeln, andernfalls gibt es nach dem Sortieren ein heilloses Durcheinander. Geben sie zunächst die Formel zur Berechnung der wöchentlichen Differenzen der natürlichen Logarithmen der Indexwerte, bzw. Wechselkurse in die Spalte G, kopieren Sie die Formel von der ersten Datenzeile bis zur letzten und markieren Sie alle Formeln, gehen Sie dann in das Menü ‚Bearbeiten‘, wählen Sie ‚Inhalte einfügen‘ und wählen Sie daraus die Möglichkeit ‚Zahlen‘. Dann werden statt der Formeln die jeweiligen Zahlenwerte in den Zellen gespeichert. Diese können Sie dann versetzt, entsprechend der Musterdefinition, in die Spalten D bis F kopieren. Damit hat man die Muster und auch die Marktzustände definiert. Außerdem muss man die Werte noch zurück versetzt in die Spalte I einfügen. Denn in der Spalte I steht, wie sich der Markt in der jeweils folgenden Woche entwickelt hat. Ich empfehle, die erste Spalte mit dem Datum so zu formatieren, wie ich es im Beispiel vorgegeben habe, also Jahreszahl, Monat und Tag ohne Zwischenraum oder Trennzeichen. So entsteht aus dem Datum eine Zahl und die Reihenfolge der Zahlen entspricht der Datumsfolge. Nach einem solchen Format für den Datumsschlüssel kann jedes Kalkulationsprogramm problemlos Datensätze chronologisch sortieren. Die Formel zur Abstandsberechnung, die man zum Finden der ‚nächsten Nachbarn‘ braucht, muss man für jedes neue Referenzmuster etwas anpassen und in die Spalte H eintragen. Die Formel in Spalte H11 bestimmt, dass die Felder D,E,F,G aller folgenden Zeilen, mit den Feldern D,E,F und G der Zeile 11 verglichen werden sollen. Die Differenzen der Felder D,E,F und G mit den entsprechenden Referenzfelder in Zeile 11 werden aufsummiert und ergeben den Abstand des jeweiligen Marktzustands zum Marktzustand am 8. Dez. 2000. Wenn man die nächsten Nachbarn zu diesem Marktzustand finden möchte, dann muss man die Datensätze nach ihrem Abstand zum Datensatz in Zeile 11 sortieren. Ein vollständiger Datensatz umfasst die Spalten A bis I. Der Sortierbereich ist daher der Bereich A11 bis H1098, das sind die Spalten A bis H in den Zeilen 11 bis 1098. Zeile 11 enthält den Referenzdatensatz und Zeile 1098 enthält den ersten verfügbaren Datensatz. Als Sortierschlüssel wählt man die Spalte H und zwar den Bereich H11 bis H1098. Man sortiert aufsteigend, d.h. der Datensatz mit dem geringsten Abstand steht ganz oben. Natürlich ist das der Datensatz vom 8. Dezember 2000, denn es ist der Referenzdatensatz, dessen Abstand zu sich gleich null ist. Danach kommen die Nachbarn, geordnet nach Nähe zum Referenzdatensatz. In Feld J11 kann man die erwartete Marktbewegung ablesen. Sie entspricht dem Mittelwert aus den Feldern in Spalte I der ‚nächsten Nachbarn‘. Auf diese Weise kann man einen beliebigen Datensatz als Referenzdatensatz festlegen und die nächsten Nachbarn dazu herausfinden. Um die ursprüngliche Reihenfolge der Datensätze wieder herzustellen, muss man den Bereich A11 bis H1098 wieder sortieren, dieses Mal jedoch mit Spalte A11 bis A1098 als Sortierschlüssel. Anders als beim Sortieren nach Abstand muss man jedoch in absteigender Reihenfolge sortieren lassen, damit der Datensatz mit dem jüngsten Datum ganz oben steht.

## Eine Mahnung zur Vorsicht

Ein NNP ist keine Kristallkugel, mit der man in die Zukunft sehen kann. Auch wenn die Vorhersage korrekt durchgeführt wurde, lag sie in vielen Fällen falsch und so wird es auch weiterhin sein. Auf lange Sicht war das Verhältnis von richtigen zu falschen Entscheidungen zwischen 55 : 45 und 60 : 40 und der NNP schien mehrere große Bewegungen richtig vorhergesagt zu haben. Aber natürlich ist das keine Garantie für die Zukunft. Wenn man den NNP benutzt, wird man dadurch noch nicht zu einem erfolgreichen Wertpapierhändler. Ich kann nur die Botschaften betonen und wiederholen, die von erfahrenen Kennern der Aktienmärkte bei Interviews mit Finanzzeitschriften übermittelt wurden, nämlich, dass man unbedingt seine Hausaufgaben machen muss. Man muss ein Handelssystem haben und man muss auch ein Geldmanagement-System anwenden. Schließlich muss man auch noch die Selbstdisziplin eines Wertpapierhändlers aufbringen, die, wie uns die ‚behavioral finance‘ lehrt, eine recht seltene Gabe ist.

In diesem Abschnitt habe ich mich öfter auf die Chaos-Theorie bezogen: Sie ist der umgangssprachliche Begriff für die Theorie komplexer dynamischer Systeme. Neben der nichtlinearen Statistik, die als etabliert gelten kann, ist die zweite Grundlage dieser Theorie die Topologie, d. i. ein Zweig der Mathematik, der sich mit den Eigenschaften geometrischer Gebilde befasst, die auch bei Transformationen konstant bleiben. Von denen, die nur Laien auf dem Gebiet sind, wird diese Disziplin auch despektierlich als „Gummi-Mathematik“ bezeichnet. Diese beiden angesprochenen Theorien liefern die Werkzeuge und den wissenschaftlichen Hintergrund für die Verwendung des NNP. Weil man widerspruchsfreie Regeln für die Interpretation von Marktzuständen aufstellt und so die Subjektivität beim Beurteilen von Kursverläufen ausschließt, ist die Vorhersage mit nächsten Nachbarn nicht nur eine bloße ‚technische Analyse‘ sondern sie kann als wissenschaftliche Methode angesehen werden.

Auch wenn die hier vorgestellte Methode große Vorzüge besitzt, sollte man nicht gleich zu viel von ihr erwarten, wenn man sie auf Marktsituationen anwendet. Denn Märkte sind soziale Systeme und keine mathematischen Gleichungen oder physikalische Systeme, wie jene, die aus der Chaos-Theorie bekannt sind. Sie folgen nicht ewigen Gesetzen, sondern sie befolgen Regeln und gelegentlich, so scheint es, nicht einmal das. Ein Grund dafür ist, dass Einflüsse von außerhalb des Marktes vorübergehend die Dynamik stören können. Ein anderer Grund ist, dass sich Märkte an die Veränderungen in ihrer Umgebung anpassen und sich wie alle lebenden Systeme hin zu komplexeren Regeln des Zusammenlebens bewegen. Darum können Versuche, mathematische Modelle von Märkten aufzustellen, nur mit begrenztem Erfolg rechnen.