

Ökonometrische Verfahren und neuronale Netze am Beispiel der Zinsprognose

Diplomarbeit
von
Haiming Sa

Mai 1998

Betreuer: Prof. Dr. W. Menzel
Dipl.-Math. S. Gutjahr

Institut für Logik, Komplexität und Deduktionssysteme
Fakultät für Informatik
Universität Karlsruhe

31927



D F Sa

Inv.-Nr. 2000/265

Erklärung

Hiermit erkläre ich, die vorliegende Diplomarbeit persönlich und ohne unzulässige Hilfsmittel angefertigt zu haben. Alle Quellen sind im Literaturverzeichnis aufgeführt.

Haiming Sa

Karlsruhe, Mai 1998

Danksagung

Professor Dr. Wolfram Menzel gebührt der erste Dank, da ohne ihn diese Arbeit nicht zustande gekommen wäre.

Für die Betreuung während der Arbeit danke ich besonders Steffen Gutjahr. Mein Dank für wertvolle Diskussionen und Hinweise geht auch an Thomas Ragg und Olaf Liedtke.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Gliederung	3
2	Ökonometrische Verfahren zur Zeitreihenprognose	4
2.1	Zeitreihenmodelle	4
2.1.1	Weißes Rauschen	5
2.1.2	Moving-Average-Prozesse	5
2.1.3	Autoregressive Prozesse	5
2.1.4	Autoregressive Moving-Average-Modelle	6
2.1.5	Integrierte Prozesse	6
2.2	Kointegration und Fehlerkorrekturmodelle	7
2.2.1	Kointegrierte Zeitreihen	8
2.2.2	Fehlerkorrekturmodelle	8
3	Neuronale Netze zur Zeitreihenprognose	11
3.1	Neuronale Netze als Prognoseverfahren	11
3.2	Netzwerktopologie	12
3.3	Lernverfahren	13
3.3.1	Gradientenabstiegsverfahren	14
3.3.2	Resilient Propagation	15
3.4	Gütemaße zur Beurteilung von Prognosen	15
3.5	Netzwerkmodellierung	16
4	Zinsprognose mit Fehlerkorrekturmodell	18
4.1	Schätz- und Testmethode	18
4.1.1	Kleinste-Quadrate-Schätzer	18
4.1.2	Kritischer t-Wert	19

4.1.3	ADF-Test	20
4.2	Empirische Untersuchung der Zinsentwicklung der DM	22
4.2.1	Daten	22
4.2.2	Kointegration	26
4.2.3	Fehlerkorrekturmodell	30
5	Zinsprognose mit Neuronalen Netzen	37
5.1	Vorgehensweise	37
5.1.1	Voranalyse der Daten	37
5.1.2	Netzwerkaufbau	41
5.1.3	Einlernen des Netzwerks	47
5.1.4	Netzwerkbeurteilung	48
5.2	Empirische Untersuchung der Zinsentwicklung in DM	48
5.2.1	Verbindungsart der Perzeptronen	49
5.2.2	Regularisierung	49
5.2.3	Weitere Ein- und Ausgabemöglichkeiten	55
6	Zusammenarbeit von NN und ECM	61
6.1	Entscheidungskomitee	61
6.1.1	Bildung des Entscheidungskomitees	61
6.1.2	Empirische Untersuchungen	62
6.2	Datenvorverarbeitung mit dem Fehlerkorrekturmodell	64
6.2.1	Zweistufiges Prognosemodell	64
6.2.2	Empirische Untersuchungen	65
7	Bewertung der Prognoseverfahren	69
7.1	Der Effekt der Abweichungsterm	69
7.2	Ein Vergleich konkurrierender Verfahren	71
8	Zusammenfassung und Ausblick	73
8.1	Zusammenfassung	73
8.2	Ausblick	74
	Literaturverzeichnis	74

Kapitel 1

Einleitung

1.1 Motivation

In der angewandten Ökonomie sind zahlreiche Verfahren entwickelt worden, die sich mit ökonometrischen Zeitreihen beschäftigen. Bis in die jüngste Zeit hat sich der vorwiegende Teil der ökonometrischen Theorie unter der Annahme entwickelt, daß die zugrundeliegenden Zeitreihen stationär sind. Demgegenüber ist die Mehrzahl der Zeitreihen, die in der angewandten Ökonomie betrachtet werden, offensichtlich nicht stationärer Natur (vgl. [1]). Das Problem dabei ist, daß die normalen statistischen Konvergenzeigenschaften dann nicht mehr gelten, wenn eine Zeitreihe nichtstationär ist.

Wird ein Verfahren, das die Stationarität der zugrundeliegenden Zeitreihen verlangt, für ökonometrische Zeitreihen angewandt, die in der Tat nichtstationär sind, so kann dies dazu führen, daß voneinander unabhängige Variablen aufgrund ihres zufälligen Gleichverhaltens als zusammenwirkend festgestellt werden. Um solche Verwirrungen zu vermeiden, wurde basierend auf den Ideen von Box und Jenkins [2] vorgeschlagen, in dem nichtstationären Fall die Zeitreihen erst zu differenzieren, um sie stationär zu machen. Aber eine Nebenwirkung dieser Methode ist, daß durch das Differenzieren die Niveaugrößen der Zeitreihen verschwinden. Auf diese Weise gehen die Informationen über langfristige Abhängigkeiten, die in den Niveaugrößen der Zeitreihen vorliegen, verloren. In [3] wurde eine Methode zur Lösung dieser Problematik vorgestellt, wobei der Ansatz der Kointegration und die Verwendung von Fehlerkorrekturmodellen gefordert wurden.

Das Konzept der Kointegration ermöglicht, Variablen auf langfristige Gleichgewichtbeziehungen zu untersuchen. Ist eine Anzahl von Zeitreihen, die einzeln betrachtet nicht-

stationär sind, kointegriert, dann existiert zwischen diesen mindestens eine Linearkombination, die stationär ist. Diese Linearkombination kann dann als langfristige Beziehung angesehen werden. Engle und Granger haben in [3] bewiesen, wenn man eine solche langfristige Beziehung in gewisser Form in ein Modell mit differenzierten Variablen einführt, bildet man ein Fehlerkorrekturmodell. Und durch die langfristige Beziehung kann der Informationsverlust, der durch das Differenzieren entsteht, ausgeglichen werden.

Im Rahmen dieser Arbeit wird das Konzept der Kointegration und des Fehlerkorrekturmodells theoretisch diskutiert und anschließend zur Zeitreihenprognose eingesetzt. Das Ziel dabei ist, zu untersuchen, ob sich die nichtstationären Zeitreihen überzeugend von dem Konzept der Kointegration und des Fehlerkorrekturmodells analysieren und prognostizieren lassen. Für die empirische Untersuchung folgen wir dem Konzept von Granger. Es ist auch zu prüfen, ob sich die Theorie von Granger empirisch nachweisen läßt.

Für die Prognose der nichtstationären Zeitreihen versuchen wir die neuronale Netze (NN) einzusetzen. Es ist bewiesen, daß die neuronale Netze bei geeigneter Gestaltung jede beliebige Funktion approximieren können. Man hat daher die Hoffnung, daß die neuronalen Netze die Beziehung der zusammenwirkenden Faktoren und damit die Entwicklung der Zeitreihe abbilden können. Durch die empirische Untersuchung sind zwei Fragen zu beantworten:

1. Sind die neuronale Netze gegenüber den ökonometrischen Modellen zur Prognose der Finanzzeitreihe konkurrenzfähig?
2. Wird eine Zusammenarbeit von neuronalen Netzen und ökonometrischen Modellen erfolgreich sein?

In der Anwendung neuronaler Netze wird die Untersuchungen hauptsächlich in zwei Punkten eingehen

- Vorverarbeitung der Netzeingabe
- Regularisierung.

Man erwartet, daß die in der Variablen versteckte Informationen durch geeignete Vorverarbeitung besser präsentiert und damit auch vom neuronalen Netz besser bearbeitet werden können. Die Regularisierung sorgt dafür, daß die Netze nicht allzu stark an das Trainingsmuster angepaßt werden. Durch die Regularisierung können die Netze das in der eigentlichen Informationen enthaltene Rauschen ausfiltern und damit ihre Generalisierungsfähigkeit erhöhen.

1.2 Gliederung

Die Arbeit ist in sieben Kapiteln gegliedert. Die ersten zwei Kapitel beschäftigen sich mit der Theorie, wohingegen die nächsten vier Kapitel die empirischen Untersuchungen umfassen. Das letzte Kapitel stellt eine Zusammenfassung dar.

In **Kapitel 2** wird eine Einführung in Zeitreihenmodelle gegeben, und es wird speziell auf die Stationarität der Prozesse eingegangen. Ferner werden das Konzept kointegrierter Zeitreihen sowie das Fehlerkorrekturmodell vorgestellt.

Kapitel 3 befaßt sich mit der Grundlage der neuronalen Netze hinsichtlich der Anwendung der Zeitreihenprognose. Außerdem wird ein Schema für den Modellaufbau bei neuronalen Netzen vorgestellt.

In **Kapitel 4** werden Testverfahren für die Untersuchung des Integrationsgrades von Zeitreihen und Schätzverfahren vorgestellt. Anhand des Konzepts von Granger wird eine empirische Untersuchung der Zinsentwicklung der D-Mark durchgeführt. Dabei wird die Kointegration der einbezogenen Zeitreihen getestet und anschließend wird die Zinsprognose mit dem Fehlerkorrekturmodell durchgeführt.

In **Kapitel 5** werden die neuronale Netze bei der Zinsprognose angewandt. Die Vorgehensweise des Modellaufbaus und die dabei verwendete Methode werden zuerst vorgestellt. In den empirischen Untersuchungen werden zahlreiche Tests durchgeführt.

In **Kapitel 6** wird die Zusammenarbeit der neuronalen Netze und des Fehlerkorrekturmodells untersucht. Dabei wird ein Entscheidungskomitee sowie ein zweistufiges Prognosemodell gebildet.

In **Kapitel 7** wird ein umfassender Vergleich zwischen den Ergebnissen verschiedener Prognosemethoden durchgeführt. Aus den Ergebnissen dieses Vergleichs werden das Konzept von Granger in der praktischen Anwendung und die Eignung der neuronalen Netze zur Zeitreihenprognose beurteilt.

Kapitel 8 ist eine Zusammenfassung der Arbeit. Dort werden die in diesem Kapitel aufgeworfenen Frage aufgrund der aus den empirischen Untersuchungen resultierten Ergebnisse beantwortet.

Kapitel 2

Ökonometrische Verfahren zur Zeitreihenprognose

2.1 Zeitreihenmodelle

Das Ziel der Zeitreihenanalyse ist die Identifikation der stochastischen Eigenschaften eines Prozesses. Wenn dies gelingt, kann anhand von geschätzten Parametern die Aufstellung der Prognose der zukünftigen Entwicklung versucht werden. Dabei setzen viele in der klassischen Ökonometrie entwickelten Zeitreihenmodelle voraus, daß es sich bei der Untersuchungszeitreihe um einen stationären Prozeß handelt.

Definition 2.1: Ein Prozeß heißt *strikt stationär*, wenn alle existierende Momente des Prozesses zeitinvariant sind. Ein Prozeß heißt *schwach stationär* (später nur *stationär* genannt) wenn gilt [4]:

1. $E[x_t] = \mu < \infty$,
2. $Var[x_t] = \sigma^2 < \infty$
3. $Cov[x_t, x_{t+h}] = \sigma_h < \infty$, für alle $t, h \in \mathbb{N}$.

Für jeden Zeitpunkt t und $t + h$ sind die Parameter μ , σ und σ_h konstant und endlich. Bei einem stationären Prozeß sind der Mittelwert und die Varianz unabhängig von der Zeit, und die Kovarianz zwischen x_t und x_{t+h} hängt nur von der Zeitverzögerung h ab, aber nicht von der Zeit t .

2.1.1 Weißes Rauschen

Definition 2.2: Ein Prozeß ε_t , $t \in \mathbf{N}$, heißt *weißes Rauschen*, falls für alle $t, h \in \mathbf{N}$ gilt [6]:

1. $E[\varepsilon_t] = 0$
2. $Var[\varepsilon_t] = \sigma^2 < \infty$
3. $Cov[\varepsilon_t, \varepsilon_{t+h}] = 0$

Der oben definierte Prozeß stellt das einfachste Beispiel eines stationären stochastischen Prozesses dar, der zeitlich unkorreliert ist.

2.1.2 Moving-Average-Prozesse

Ein Prozeß x_t , der aus der Linearkombination einer Folge von weißem Rauschen entsteht, heißt *Moving-Average-Prozeß der Ordnung q* , geschrieben als MA(q):

$$x_t = \sum_{i=0}^q \theta_i \varepsilon_{t-i} = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t \in \mathbf{N}, \quad \theta_i \in \mathbf{R}.$$

Der Koeffizient θ_0 kann ohne Einschränkung der Allgemeinheit auf Eins normiert werden. Ein Moving-Average-Prozeß ist immer stationär.

2.1.3 Autoregressive Prozesse

In einem *autoregressiven Modell der Ordnung p* mit $p \in \mathbf{N}$ (AR(p)) wird der Wert des stochastischen Prozesses x_t mit $t \in \mathbf{N}$ durch eine gewichtete Summe der p verzögerten Werte des Prozesses und dem Störterm ε_t dargestellt:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t, \quad t \in \mathbf{N}, \quad \phi_p \in \mathbf{R}.$$

Jeder autoregressive Prozeß ist invertierbar, d.h. daß der Prozeß ε_t durch den Prozeß x_t dargestellt werden kann.

Ein Lagoperator L ist definiert durch:

$$L^i x_t = x_{t-i}$$

Mit Hilfe des Lagoperators L kann der AR(p)-Prozeß auch geschrieben werden als:

$$\begin{aligned}(\phi_0 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)x_t &= \varepsilon_t \\ \Phi(L)x_t &= \varepsilon_t, \quad t \in \mathbf{N}, \quad \phi_p \in \mathbf{R},\end{aligned}$$

wobei $\phi_0 = 1$ ist. Damit ein AR(p)-Prozeß stationär ist, müssen die Lösungen der Gleichung $\Phi(z) = 0$ alle außerhalb des Einheitskreises liegen, also $|z| > 1$, wobei $\Phi(z)$ das charakteristische Polynom $1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$ ist [6].

2.1.4 Autoregressive Moving-Average-Modelle

Eine Kombination von AR- und MA-Prozessen resultiert in folgender Form:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

und wird als *autoregressives Moving-Average-Modell der Ordnung p, q* , also ARMA(p, q) bezeichnet. Mit Hilfe des Lagoperators L kann der ARMA(p, q)-Prozeß auch geschrieben werden als:

$$\begin{aligned}(1 - \phi_1 L - \dots - \phi_p L^p)x_t &= (1 + \theta_1 L + \dots + \theta_q L^q)\varepsilon_t \\ \Phi(L)x_t &= \Theta(L)\varepsilon_t.\end{aligned}$$

Der ARMA-Prozeß ist stationär, wenn die Lösungen z der charakteristischen Gleichung $\Phi(z) = 0$ alle außerhalb des Einheitskreises liegen. Der Prozeß x_t ist invertierbar, wenn die Lösungen z der charakteristischen Gleichung $\Theta(z) = 0$ ebenfalls die Bedingung $|z| > 1$ erfüllen [6]. Jeder stationäre, invertierbare ARMA(p, q)-Prozeß hat eine stationäre AR(∞)-Darstellung, und jeder stationäre ARMA(p, q)-Prozeß läßt sich als MA(∞)-Prozeß darstellen mit den Parametern θ_i , welche die Bedingung $\sum_{i=0}^{\infty} \theta_i < \infty$ erfüllen [6].

2.1.5 Integrierte Prozesse

Hier werden Zeitreihen betrachtet, die nichtstationär sind, die aber durch einfache oder mehrfache Differenzenbildung zu stationären gewandelt werden können. Diese sogenannte Klasse der integrierten Zeitreihen deckt natürlich nur einen Teil möglicher Nicht-Stationarität ab, die aber kann das Verhalten der in der Ökonomie beobachteten Zeitreihen hinreichend genau beschreiben [6]. Im allgemeinen werden stochastische Prozesse,

die durch d -fache Differenzenbildung stationär gemacht werden können, als integrierte Zeitreihen vom Grad d bezeichnet:

Definition 2.3: Ein stochastischer Prozeß x_t , $t \in \mathbb{N}$, der nach d -facher Differenzenbildung eine stationäre invertierbare ARMA-Darstellung hat, wird als *integriert vom Grade d* , $d \in \mathbb{N}$ bezeichnet, mit der Schreibweise $x_t \sim I(d)$ [3].

Ist der ARMA-Prozeß x_t nichtstationär, aber solcher Art, daß genau d Wurzeln von $\Phi(L)$ auf dem Einheitskreis und die restlichen außerhalb liegen, dann kann der Prozeß geschrieben werden als [5]

$$\omega(L)(1-L)^d x_t = \Theta(L)\varepsilon_t, \quad (2.1)$$

wobei $\omega(L)$ ein stationärer autoregressiver Operator der Ordnung $p-d$ ist, und der Operator $(1-L)^d$ d Wurzeln der Größe Eins hat. Eine Darstellung dieser Art wird ein *autoregressiver integrierbarer Moving-Average-Prozeß der Ordnung $p-d, d, q$* (ARIMA($p-d, d, q$)) genannt, und x_t wird in diesem Fall als homogen nichtstationär der Ordnung d bezeichnet.

2.2 Kointegration und Fehlerkorrekturmodelle

Das schwerwiegende Problem, das entsteht, wenn versucht wird, Zeitreihen zu analysieren, die integriert von der Ordnung k sind, mit $k > 0$, d.h. die nichtstationär sind, resultiert aus der Tatsache, daß in diesem Fall die normalen statistischen Eigenschaften der ersten und zweiten Momente nicht mehr gelten [7].

Wenn die Voraussetzung der Stationarität mißachtet wird und weiterhin Standardverfahren für ökonomische Zeitreihen angewendet werden, so kann dies dazu führen, daß zweifelsfrei unabhängige Variablen wegen ihres zufälligen Gleichrends als systematisch korreliert betrachtet werden. Diese Beobachtung systematischen Gleichverhaltens unabhängiger Zeitreihen wird als *Scheinkorrelation* bezeichnet [8]. Um die Schwierigkeiten, die aus den Scheinkorrelationen resultieren, zu umgehen, wurde basierend auf den Ideen von Box und Jenkins [2] vorgeschlagen, die zu betrachtenden Zeitreihen erst zu differenzieren, um sie anschließend stationär zu machen. Eine Folge dieser Prozedur ist jedoch, daß durch die alleinige Analyse der Differenzen ökonomischer Zeitreihen alle Informationen über mögliche langfristige Abhängigkeiten, die im Niveau der Zeitreihen vorliegen, verloren gehen. Eine Möglichkeit, dieses Problem zu lösen, ist die Verwendung von Fehlerkorrekturmodellen,

die auf dem Konzept der kointegrierten Zeitreihen basieren.

2.2.1 Kointegrierte Zeitreihen

Häufig ist es von Interesse festzustellen, ob ein System ein langfristig gültiges Gleichgewicht hat, oder ob zwischen nichtstationären Zeitreihen bestimmte Abhängigkeiten bestehen. Ein empirischer Test auf derartige Beziehungen, d.h. auf einen gemeinsamen stochastischen Trend in den Daten, wird im Rahmen des theoretischen Konzeptes der Kointegration ermöglicht. Das Vorgehen wurde erstmals von Granger und Weiss [10] vorgeschlagen und wird im folgenden vorgestellt.

Sind zwei Zeitreihen x_t und y_t integriert von Grad d ($x_t \sim I(d), y_t \sim I(d)$), dann gilt im allgemeinen, daß auch die Linearkombination $z_t = x_t - \beta y_t$ integriert von Grad d ist ($z_t \sim I(d)$). Wenn ferner sich die zwei Zeitreihen langfristig parallel entwickeln und die Differenz der beiden stochastisch gesehen konstant bleibt, dann ist es möglich, daß eine langfristige Gleichgewichtsbeziehung zwischen den beiden Zeitreihen besteht, so daß $z_t \sim I(d - b)$ mit $b > 0$ gilt. In diesem Fall sind die beiden Zeitreihen kointegriert. Kointegration läßt sich formal folgendermaßen definieren, wobei zu beachten ist, daß bei der Betrachtung eines multivariaten Systems, x_t und z_t Vektoren von Zeitreihen darstellen und nicht mehr nur einzelne Zeitreihen:

Definition 2.4: Die Komponenten eines Vektors x_t sind *kointegriert von Grade d und b* , bezeichnet durch $x_t \sim CI(d, b)$, wenn alle Komponenten von x_t integriert von Grade d sind, also $x_t \sim I(d)$, und ein Vektor β , mit $\beta \neq 0$, existiert, so daß $z_t = \beta x_t \sim I(d - b)$, $b > 0$ [10].

Unsere Untersuchungen konzentrieren sich auf den Fall $d = 1$, $b = 1$, d.h. auf Zeitreihen und Linearkombinationen, die integriert vom Grad Eins bzw. Null sind. Wenn die Komponenten von $x_t \sim I(1)$ nicht kointegriert sind, so ist z_t nichtstationär und hat z_t einen von der Zeit t abhängigen Mittelwert. Liegt dagegen Kointegration vor, so ist z_t stationär, also $z_t \sim I(0)$.

2.2.2 Fehlerkorrekturmodelle

Eine sehr wichtige Folgerung, die sich aus der Kointegrationsanalyse ergibt, resultiert aus dem "Granger-Repräsentations-Theorem" [3]. Das Theorem besagt, daß wenn eine Anzahl von Variablen kointegriert ist, daß dann eine gültige Fehlerkorrektur-Darstellung

der Daten existiert. In einem multivariaten System mit einem 3-dimensionalen Zeitreihenvektor $(x_t, y_t, z_t)^T$, dessen Komponenten alle integriert vom Grad Eins sind, kann das Fehlerkorrekturmodell (Error correction modell (ECM)) in der Form [4]

$$(1 - L)x_t = \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1 - L)x_{t-i} + \sum_{i=1}^m \alpha_{2i}(1 - L)y_{t-i} + \sum_{i=1}^k \alpha_{3i}(1 - L)z_{t-i} + \alpha_4 err_{t-1} + \varepsilon_t \quad (2.2)$$

dargestellt werden, wobei ε_t ein stationärer Störterm und $err_t = x_t - (\beta_0 + \beta_1 y_t + \beta_2 z_t)$ ist. Der Vektor β wird dann als *Kointegrationsvektor* bezeichnet.

Wenn Variablen kointegriert sind, dann können aufgrund des Granger-Repräsentations-Theorems durch die Verwendung von Fehlerkorrekturmodellen Informationen im Niveau der Variablen beibehalten werden. Dadurch werden die beiden Ansätze, das Schätzen im Niveau sowie in den ersten Differenzen, in Einklang gebracht, und somit das Problem der Scheinkorrelationen wird umgangen. Das Fehlerkorrekturmodellen kombiniert langfristige Gleichgewichtsbeziehungen mit den kurzfristigen Entwicklungen der Zeitreihen. Innerhalb eines solchen Modells wird ein Teil der Abweichung vom langfristigen Gleichgewicht in einer Periode in der darauffolgenden korrigiert. In die obere Gleichung (2.2) geht der Fehlerkorrektur-Term (err_{t-1} , auch Abweichungsterm genannt) mit einer Verzögerung von einer Zeiteinheit ein.

Zum Aufbau von Fehlerkorrekturmodellen schlagen Engle und Granger [3] ein einfaches Zwei-Schritt-Verfahren vor. Nach dem Test auf die Kointegration wird in einem ersten Schritt mit der Methode des Kleinsten-Quadratischen-Fehlers der Kointegrationsvektor β geschätzt. Im zweiten Schritt können die anderen Parameter des Fehlerkorrekturmodells ermittelt werden, indem die Schätzwerte der Kointegrationsgleichung in ein allgemeines Fehlerkorrekturmodell einbezogen werden. Dies wird dadurch erreicht, daß das Residuum der Kointegrationsgleichung einfach als weiterer Parameter ins Fehlerkorrekturmodell eingeht.

Für den Fall dreier Variablen x_t , y_t und z_t , die $CI(1, 1)$ sind, wird im ersten Schritt der Kointegrationsvektor β mit

$$x_t = \beta_0 + \beta_1 y_t + \beta_2 z_t \quad (2.3)$$

geschätzt. Der Fehler der Kointegrationsgleichung err_t ergibt sich dann durch

$$err_t = x_t - (\hat{\beta}_0 + \hat{\beta}_1 y_t + \hat{\beta}_2 z_t). \quad (2.4)$$

Um die Kointegration der Variablen in (2.3) zu gewährleisten, muß err_t stationär sein. Somit geht er in das Fehlerkorrekturmodell ein, mit

$$\begin{aligned} (1-L)x_t &= \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1-L)x_{t-i} + \sum_{i=1}^m \alpha_{2i}(1-L)y_{t-i} \\ &\quad + \sum_{i=1}^k \alpha_{3i}(1-L)z_{t-i} + \alpha_4 err_{t-1} + \varepsilon_t, \end{aligned} \quad (2.5)$$

wobei ε_t weißes Rauschen darstellt.

Kapitel 3

Neuronale Netze zur Zeitreihenprognose

3.1 Neuronale Netze als Prognoseverfahren

Konventionelle ökonomische Prognoseverfahren basieren auf einem vorab wohldefinierten Modell. Sie unterstellen ein vorformuliertes Gleichungssystem, dessen Größen und Größenbeziehungen determiniert sind. Man geht davon aus, daß sich die Strukturen der Wirklichkeit in einem vorliegenden Prognosemodell darstellen lassen. Modellvorgabe im ökonomischen Sinn bedeutet, daß a priori sowohl eine Festlegung auf bestimmte Variablen und deren Vorverarbeitung, als auch eine mathematische Beschreibung ihrer Abhängigkeiten erfolgen muß [14]. Es werden folglich konkrete Funktionstypen unterstellt.

Für ein neuronales Netz muß die genaue parametrisierte Beschreibung der zugrundeliegenden Funktion a priori nicht bekannt sein. Das Netzwerk versucht durch Lernen die den Daten zugrundeliegende Funktion abzubilden. Die hierbei stattfindende Verknüpfung von Linearkombinationen der Aktivierungsfunktionen eignet sich zur Approximation beliebiger komplexer Funktionen. Neuronale Netze besitzen somit die Fähigkeit, relevante Strukturzusammenhänge in den Daten im Rahmen seines Lernmodus zu erkennen und selbständig zu gewichten.

Eine gänzliche Modellfreiheit ist jedoch auch bei neuronalen Netzen nicht gegeben. Eine bestimmte Modellvorstellung ist notwendig. Der Datenwahl kommt dabei eine große Bedeutung zu, da das Lernen neuronales Netzes schließlich auf den ausgewählten Da-

ten basiert. Modellvorstellung im neuronalen Sinn bedeutet folglich im wesentlichen die Festlegung auf eine möglichst repräsentative Teilmenge an Variablen und deren möglichst effiziente Aufbereitung [14]. Im Vergleich zur ökonomischen Modellvorgabe handelt es sich hierbei um eine reduzierte Problemstellung, da auf eine konkrete Festlegung der parametrisierten Beschreibung der Variablen verzichtet werden kann.

3.2 Netzwerktopologie

In [11] wurde bewiesen, daß ein neuronales Netz mit einer einzigen inneren Schicht bereits in der Lage ist, eine beliebige stetige Funktion zu approximieren. Aus diesem Grund wird hier ausschließlich die Drei-Schicht-Topologie als Standardtopologie verwendet. In Abbildung 3.1 ist ein vorwärtsgerichtetes Netzwerk mit Drei-Schicht-Topologie dargestellt¹.

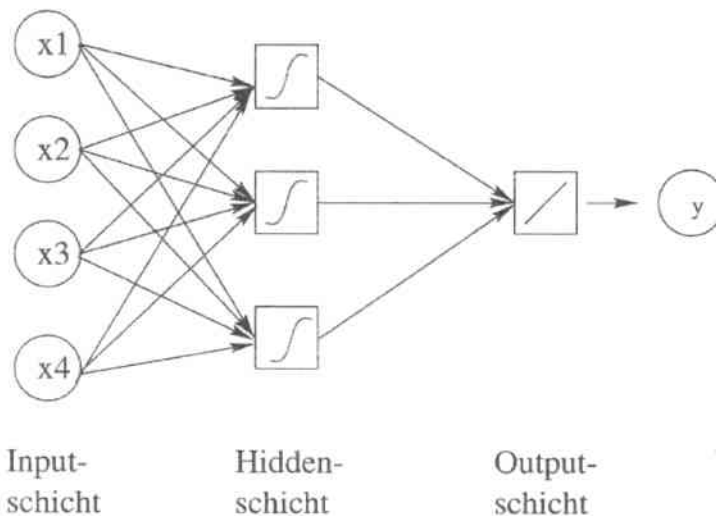


Abbildung 3.1 Drei-Schicht-Topologie für ein neuronales Netz

In der Inputschicht finden sich eine ganze Reihe von Neuronen, deren einzige Aufgabe es ist, Informationen aus der Zeitreihen aufzunehmen. In ihnen findet keine Überlagerung verschiedener Variablen statt. Die folgende Hiddenschicht verarbeitet in jeder Zelle die Information. Die Outputschicht verarbeitet die Information weiter und gibt dann den Output des Netzes aus.

¹vgl. [15]

Eine einfache Erweiterung des Standardmodells ist die Einbeziehung linearer Konnektoren als unmittelbare Verbindungen zwischen Input und Output. In Abbildung 3.2 ist eine solche Erweiterung abgebildet².

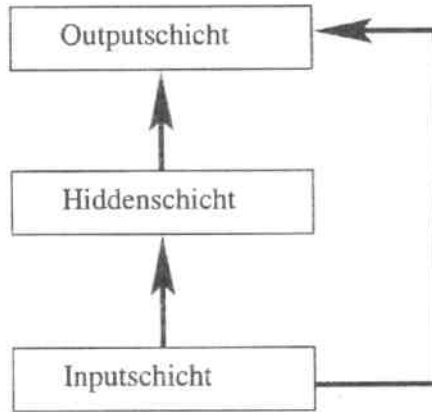


Abbildung 3.2 Drei-Schicht-Topologie mit linearer Erweiterung

Der lineare Teil des Modellbaus kann durch die Parameter in der Direktverbindung zwischen Input und Output eventuell besser repräsentiert werden, während die nicht-linearen Teile durch die Zwischenschicht des Netzwerkes modelliert werden [16]. Empirische Untersuchungen sprachen dafür, daß die Zeitreihen meistens als Zusammenspiel vieler Faktoren dargestellt und ihnen hauptsächlich nichtlineare Beziehung zugesprochen werden [16]. Man sollte aber nicht deswegen glauben, daß man ganz auf die linearen Anteile ökonomischer Zusammenhängen verzichten könnte.

3.3 Lernverfahren

Ein neuronales Netz ist ein adaptives System. Beim Vorgang des sogenannten *Einlernens* werden die Gewichte eines neuronalen Netzes in einem iterativen Prozeß angepaßt. Durch Modifikation der Gewichte w_i wird der Zusammenhang zwischen Eingabe und Ausgabe des Netzes so verändert, daß dessen Ausgabe sich dem Zielwert annähert.

Der Mustermenge D ist die Zusammensetzung der Eingabemustermenge X und der Ausgabemustermenge Y :

$$D = \{(X_i, Y_i) | i = 1 \dots N\}, \quad \text{mit } X_i \in \mathbf{R}^I \quad \text{und} \quad Y_i \in \mathbf{R},$$

²vgl. [16]

mit I Eingabeneuronen und N Mustern.

Die Güte des Netzes wird dabei bei gegebenen Mustern D durch ein Fehlermaß $E_D(D, w)$ ³ des Netzwerks in Abhängigkeit der Gewichte ausgedrückt. Ziel jeder Lernregel ist es, diesen Fehler E_D zu minimieren.

Es gibt verschiedene Fehlermetriken, bezüglich derer der Fehler minimiert werden kann. Hier wird der quadratische Fehler verwendet:

$$E_D(w) = \frac{1}{2}(y_i - f(x_i, w))^2. \quad (3.1)$$

wobei f die von dem Netz abgebildete Funktion ist.

Der über alle Mustern summierte Fehler ist dann:

$$E_D(D, w) = \frac{1}{2} \sum_{i=1}^N (y_i - f(x_i, w))^2, \quad (3.2)$$

Die Lernverfahren, die den Fehler einzelnes Musters nach (3.1) verwenden, heißen *Musterlernen*, während die anderen, die den summierten Fehler nach (3.2) verwenden, heißen *Epochenlernen*. Zur Gewichtsänderung werden bei Epochenlernen im Gegensatz zum Musterlernen, bei dem die Gewichte für jedes einzelne Muster geändert werden, alle Muster berücksichtigt. In späteren Untersuchungen wird für die Gewichtsänderung ein nicht triviales Verfahren, *Resilient Propagation* (RPROP) [18] verwendet. RPROP ist ein erweitertes Gradientenabstiegsverfahren für Epochenlernen.

3.3.1 Gradientenabstiegsverfahren

Bei einem Gradientenabstiegsverfahren wird versucht, ein lokales Minimum von E_D durch die Gewichtsänderungen zu finden. Ausgehend von $w(0)$ wird ein iterativer Prozeß durchgeführt, gemäß

$$w(t+1) = w(t) - \eta \frac{\partial E_D}{\partial w}(w(t)).$$

$\eta > 0$ ist eine Konstante, die sogenannte *Schrittweite*.

Es wird also in Richtung des steilsten Abstiegs im Fehleraum jeweils der nächste Punkt gefunden. Der Abstand zwischen $w(t)$ und $w(t+1)$ hängt von der Schrittweite η und der partiellen Ableitung $\frac{\partial E_D}{\partial w} w(t)$ ab. Ist E_D konvex, so ist ein lokales Minimum zugleich globales Minimum. Im Normalfall, muß es aber nicht eindeutig bestimmt sein [19].

³Der Index D deutet an, daß es sich hierbei um den Fehler bzgl. der Mustermenge D handelt.

3.3.2 Resilient Propagation

Resilient Propagation ist ein Verfahren, bei dem jedes einzelne Gewicht dynamisch und lokal angepasst wird. Die Idee dabei ist: für die Änderung der Gewichte soll nur noch das Vorzeichen der partiellen Ableitungen von E_D betrachtet werden, ihrer Betrag soll dabei keine Rolle mehr spielen. Bei jeder Gewichtsänderung wird sich jeder einzelne Gewichte w_{ij} nun gemäß dieses Vorzeichens um einen bestimmten Betrag verändert, wobei dieser Betrag [19]

- lokal für w_{ij} ist und vom Zeitpunkt abhängt
- ausgehend von seinem Wert zum vorausgehenden Zeitpunkt fortgeschrieben wird in Abhängigkeit davon, ob die Gradientenkomponente (i, j) zuletzt ihr Vorzeichen gewechselt hat oder nicht.

Mit $E'_{ij}(t) = \frac{\partial E}{\partial w_{ij}}(w_{ij}(t))$ werden die Gewichtsänderungen durchgeführt gemäß [19]

$$w_{ij}(t+1) = \begin{cases} w_{ij}(t) - \Delta_{ij}(t) & \text{falls } E'_{ij}(t) > 0 \\ w_{ij}(t) & \text{falls } E'_{ij}(t) = 0 \\ w_{ij}(t) + \Delta_{ij}(t) & \text{falls } E'_{ij}(t) < 0 \end{cases}$$

wobei

$$\Delta_{ij}(t) = \begin{cases} \eta^+ * \Delta_{ij}(t-1) & \text{falls } E'_{ij}(t-1) * E'_{ij}(t) > 0 \\ \Delta_{ij}(t-1) & \text{falls } E'_{ij}(t-1) * E'_{ij}(t) = 0 \\ \eta^- * \Delta_{ij}(t-1) & \text{falls } E'_{ij}(t-1) * E'_{ij}(t) < 0 \end{cases}$$

und η^+ , η^- Konstanten sind mit: $\eta^+ > 1$, $0 < \eta^- < 1$. Aufgrund experimenteller Untersuchungen lassen sich η^+ , η^- fixieren zu $\eta^+ = 1.2$, $\eta^- = 0.5$ [19].

3.4 Gütemaße zur Beurteilung von Prognosen

Um die Ergebnisse bei der Prognose beurteilen zu können, existiert in der Literatur eine Menge Beurteilungs- oder Gütemaßen. Im folgenden werden zwei wichtige Gütemaße vorgestellt.

Ein gängiges Maß ist der in der Fehlerfunktion für Gewichtsänderung verwendeten quadratische Fehler E zwischen dem vorhergesagten und tatsächlich eingetretenen Wert:

$$E = \frac{1}{2} \sum_{i=1}^N (y_i - f(x_i, w))^2$$

Ein rein qualitatives Maß stellt die sogenannte *Trefferquote* dar, die den Anteil richtig vorhergesagter Trends (Kurs steigt bzw. Kurs fällt) an der Gesamtanzahl erfolgter Prognosen angibt.

3.5 Netzwerkmodellierung

Der wesentliche Teil in der Anwendung der neuronalen Netze ist die Netzwerkmodellierung. Dazu gehören die Vorverarbeitung der Daten, die Wahl der Netztopologie sowie die Anwendung von Techniken wie Regularisierung durch Weight Decay. Das Ziel dabei ist, eine Netzwerkarchitektur zu bestimmen, die einerseits klein genug ist, um das in den Beobachtungen enthaltenen Rauschen auszufiltern und andererseits groß genug ist, um den funktionalen Anteil in den Beobachtungen abzubilden [15].

Der Prozeß der Netzwerkmodellierung ist grundsätzlich sehr ähnlich wie diejenige, die in der ökonomischen Modellbildung verwendet wird. Sie gliedert sich in fünf Schritte [15]:

1. Voranalyse der Daten
2. Spezifikation des Modells
3. Schätzung der Parameter
4. Beurteilung der Modells
5. Auswahl eines Modells aus allen alternativen Modellen

In Abbildung 3.3 sind diese fünf Schritte in einem Prozeßdiagramm abgebildet⁴.

In dem Modellbildungsprozeß beziehen sich die ersten vier Schritte auf den Aufbau eines einzelnen Modells, der letzte Schritt betrifft die Auswahl eines Modells aufgrund

⁴vgl. [15]

der Bewertungen der Netzwerke nach bestimmten Kriterien. Dieser Prozeß wird so lange wiederholt, bis eines der resultierenden Netzwerke für die Aufgabe nach den gegebenen Kriterien akzeptabel ist.

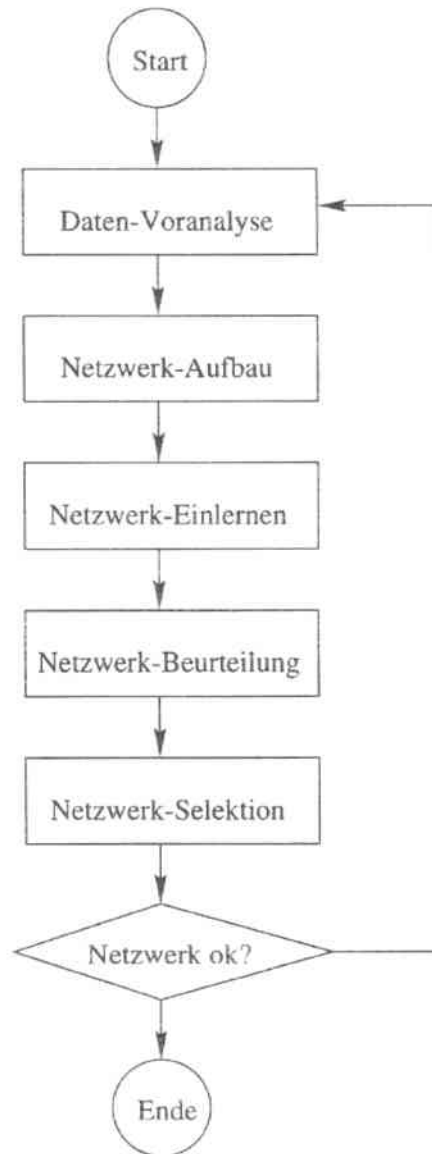


Abbildung 3.3 Ein Modellierungsprozeß für neuronale Netze

Kapitel 4

Zinsprognose mit Fehlerkorrekturmodell

Das in Kapitel 2 dargestellte Fehlerkorrekturmodell soll im folgenden zur Lösung des praktischen Problems getestet werden. Dabei wird das Fehlerkorrekturmodell zur Prognose der zehnjährigen Zinsenentwicklung in DM auf Tages-, Wochen sowie Monatsbasis angewandt. Bevor wir das Modell für ein konkretes Problem konstruieren, machen wir eine kurze Vorstellung über einige Methoden, die in dem Modellaufbau zum Einsatz kommen werden.

4.1 Schätz- und Testmethode

4.1.1 Kleinste-Quadrate-Schätzer

In der Form des kointegrierten Modells existiert (mindestens) eine Linearkombination zwischen den kointegrierten Variablen (siehe Definition 2.4). Diese Linearkombination stellt im Fall von $CI(1,1)$ definitionsgemäß einen stationären Prozeß dar. Der Kleinste-Quadrate-Schätzer dient dazu, diesen Kombinationsparameter zu schätzen.

Gegeben seien die Eingabematrix X und der Ausgabevektor Y eines Prozesses mit

$$X = \begin{pmatrix} x_1^1 & \cdots & x_1^m \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^m \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Gesucht ist der Vektor $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}$ derart, daß $(X\alpha - Y)^2$ minimal ist.

Multipliziert man $(X\alpha - Y)^2$ aus, hat man:

$$\begin{aligned} (X\alpha - Y)^2 &= (X\alpha - Y)^T(X\alpha - Y) \\ &= (X\alpha)^T(X\alpha) - (X\alpha)^TY - Y^T(X\alpha) + Y^TY \\ &= \alpha^T X^T X \alpha - 2\alpha^T X^T Y + Y^T Y. \end{aligned}$$

Um $(X\alpha - Y)^2$ zu minimieren, setzt man die Ableitung von $(X\alpha - Y)^2$ nach α gleich 0 und löst die Gleichung $\frac{\partial(X\alpha - Y)^2}{\partial\alpha} = 0$ nach α auf. Da $X^T X$ symmetrisch ist, gilt

$$\frac{\partial(\alpha^T X^T X \alpha)}{\partial\alpha} = 2X^T X \alpha.$$

Somit hat man:

$$\begin{aligned} \frac{\partial(X\alpha - Y)^2}{\partial\alpha} &\stackrel{!}{=} 0 \\ \Leftrightarrow 2X^T X \alpha - 2X^T Y &= 0 \\ \Leftrightarrow X^T X \alpha &= X^T Y \\ \Leftrightarrow \alpha &= (X^T X)^{-1} X^T Y. \end{aligned}$$

Der Kleinste-Quadrate-Schätzer für α ist somit $(X^T X)^{-1} X^T Y$.

4.1.2 Kritischer t-Wert

Nachdem der Kombinationsparameter α für die Linearkombination geschätzt werden, ist es wichtig zu wissen, ob jede Komponente α_i von α signifikant von 0 unterschiedlich ist. Es geht also darum zu überprüfen, ob die zu α_i gehörige erklärende Variable tatsächlich zur Erklärung von Y beiträgt. Dafür wird ein *t-Test* für die Hypothese

$$H_0 : \alpha_i = 0 \quad \text{gegen} \quad H_1 : \alpha_i \neq 0 \quad \text{mit} \quad i = 1, 2, \dots, m$$

durchgeführt. Bei einem t-Test wird die Gültigkeit der gestellten Hypothese geprüft. Unter $H_0 : \alpha_i = 0$, muß der Schätzwert $\hat{\alpha}_i$ in ein bestimmtes um 0 symmetrisches Intervall fallen, dessen Grenzen von der vorgegebenen Sicherheitswahrscheinlichkeit $(1-p)$ abhängen. Dieses Intervall stellt den sogenannten *Annahmereich* dar. p ist die Wahrscheinlichkeit, daß bei Gültigkeit von H_0 $\hat{\alpha}_i$ einen Wert außerhalb des Annahmereichs annimmt. Sie ist die sogenannte *Irrtumswahrscheinlichkeit*, mit der man H_0 verwirft, obwohl H_0 zutrifft.

Mit dem zweiseitigen t-Test wird für jedes $\hat{\alpha}_i$ ein t-Wert t_i durch

$$t_i = \frac{\hat{\alpha}_i}{\sqrt{\widehat{Var}[\hat{\alpha}_i]}}$$

ermittelt. Für die Sicherheitswahrscheinlichkeit gilt

$$1 - p = P_{H_0}(t_i \in [-t_{krit}, t_{krit}]),$$

wobei $[-t_{krit}, t_{krit}]$ der Annahmereich für H_0 und t_{krit} der *kritische t-Wert* ist, bei dem die Hypothese H_0 gerade abgelehnt werden würde. Ist der ermittelte t-Wert t_i für α_i innerhalb des Annahmereichs $[-t_{krit}, t_{krit}]$, ist H_0 mit Irrtumswahrscheinlichkeit von p nicht zu verwerfen; gilt $t_i \notin [-t_{krit}, t_{krit}]$, dann ist H_0 mit Irrtumswahrscheinlichkeit von p abzulehnen, und somit ist α_i mit einer maximale Irrtumswahrscheinlichkeit von p von 0 unterschiedlich.

4.1.3 ADF-Test

Ein wichtiger Schritt bei der Überprüfung der Kointegrationsbedingung ist den Integrationsgrad (siehe Definition 2.3) jeder Variablen zu testen. Dazu wird der ADF-Test (Augmented Dickey-Fuller Unit Root Test) verwendet.

Für diesen Test wird von einem AR(1)-Prozeß ausgegangen:

$$x_t = \phi x_{t-1} + \varepsilon_t, \quad t \in \mathbf{N} \quad (4.1)$$

mit ε_t ein weißes Rauschen. Ist $\phi > 1$, divergiert der Prozeß x_t ; gilt $\phi = 1$, so ist er ein nichtstationärer Prozeß. Die Bedingung für einen stationären Prozeß ist $|\phi| < 1$.

Der ADF-Test ist ein t-Test für den Koeffizient ϕ aus der Regressionsgleichung (4.1). Zu testen ist, ob in (4.1) $\phi = 1$ ist. Dabei wird überprüft, ob das x_t einem nichtstationären Prozeß $x_t = x_{t-1} + \varepsilon_t$ folgt, oder ob ein stationärer AR(1)-Prozeß vorliegt.

Um die Teststatistik abzuleiten, wird zunächst die Gleichung (4.1) als Testregression betrachtet. Zieht man von beiden Seiten der Gleichung x_{t-1} ab, so erhält man:

$$(1 - L)x_t = \bar{\phi}x_{t-1} + \varepsilon_t, \quad (4.2)$$

mit $\bar{\phi} = \phi - 1$ und ε_t ein weißes Rauschen. Die Hypothese lautet jetzt:

$$H_0 : \bar{\phi} = 0 \quad \leftrightarrow \quad H_1 : \bar{\phi} \neq 0. \quad (4.3)$$

H_1 wird zwar als "ungleich Null" formuliert, aber um die Stationarität des Prozesses zu gewährleisten, muß der Koeffizient im Bereich $(-1, 0)$ liegen, so daß der t-Wert auch unter 0 liegen soll. Es wird als Teststatistik die t-Wert für $\bar{\phi}$ betrachtet:

$$\hat{t}_{\bar{\phi}} = \frac{\hat{\bar{\phi}}}{\sqrt{\text{Var}[\hat{\bar{\phi}}]}}$$

Die Verteilung dieser Teststatistik ist unter H_0 keine Standardnormalverteilung und wurde von Dickey [20] tabelliert.

Um die restriktiven Annahmen für den Störprozeß ε_t in (4.2) zu verallgemeinern, ist eine Erweiterung um einen ARMA(p,q)-Prozeß nötig:

$$\begin{aligned} (1 - \phi_1 L - \dots - \phi_p L^p)\varepsilon_t &= (1 + \theta_1 L + \dots + \theta_q L^q)u_t \\ \Phi(L)\varepsilon_t &= \Theta(L)u_t. \end{aligned}$$

Ist der Moving-Average-Teil invertierbar, so hat u_t bekanntermaßen eine $\text{AR}(\infty)$ -Repräsentation [7]:

$$\begin{aligned} u_t &= \Theta(L)^{-1}\Phi(L)\varepsilon_t \\ &= \varepsilon_t - \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i} \end{aligned}$$

mit $\alpha_i + \theta_1 \alpha_{i-1} + \theta_2 \alpha_{i-2} + \dots + \theta_q \alpha_{i-q} = 0, i > q$.

Die Gleichung (4.2) läßt sich nun umschreiben [7]:

$$(1 - L)x_t = \bar{\phi}x_{t-1} + \sum_{i=1}^{\infty} \alpha_i (1 - L)x_{t-i} + u_t.$$

Die Hypothesen dieser Testregression bleiben identisch (4.3). Dieser Test wird als ADF-Test bezeichnet.

4.2 Empirische Untersuchung der Zinsentwicklung der DM

In diesem Abschnitt wird die Eignung des Fehlerkorrekturmodells zur Vorhersage der ökonomischen Zeitreihe empirisch untersucht. Dabei werden das Fehlerkorrekturmodell und das Modell ohne Information über langfristige Gleichgewichtsbeziehungen der Zeitreihen zur Prognose der zehnjährigen Zinsentwicklung in DM angesetzt. Anschließend werden die beiden Modelle gemäß ihrer Prognoseleistungen verglichen.

4.2.1 Daten

Es wurden zunächst die ökonomische Faktoren ausgesucht, die die Zinsentwicklung bzw. die Ursache der Zinsentwicklung widerspiegeln. Bei der Entwicklung für das ökonometrische Modell wurde angenommen, daß für die Entwicklung des langfristigen Zinsen in Deutschland, der kurzfristige Zins in Deutschland und der langfristige Zins in der USA entscheidend sind. Unter dieser Annahme wurden letztlich der zehnjährige Zins in D-Mark (D10Y), der zehnjährige Zins in US-Dollar (U10Y) sowie die Frankfurt-Interbanken-Dreimonaten-Zins (F3M) in die Analyse einbezogen. Die verwendete Datenmenge umfaßt die täglichen Werte von 1. März 1991 bis zum 1. April 1997. In Abbildung 4.1 wurden die drei Zeitreihen in diesem Zeitraum dargestellt. Zum Modellaufbau wurden allerdings nur Daten von 1. März 1991 bis zum 28. Februar 1994 verwendet. Und die Daten im übrigen Zeitraum wurden zum Modelltest verwendet.

Differenzreihe

Was bei der Prognose tatsächlich interessiert, sind die Veränderungen der Zeitreihe. Für die drei Zeitreihen D10Y, U10Y und F3M werden Differenzreihen $D10Y_nDIFF^1$, $U10Y_nDIFF$ und $F3M_nDIFF$ gebildet:

$$X_nDIFF_t = (1 - L^n)X_t = X_t - X_{t-n}$$

In Abbildung 4.2 - 4.4 werden die Tagesdifferenzreihen, Wochendifferenzreihen (5 Tage Differenz) und Monatsdifferenzreihen (20 Tage Differenz) dargestellt.

¹Mit X_nDIFF wird eine n-Tage-Differenzreihe von X bezeichnet.

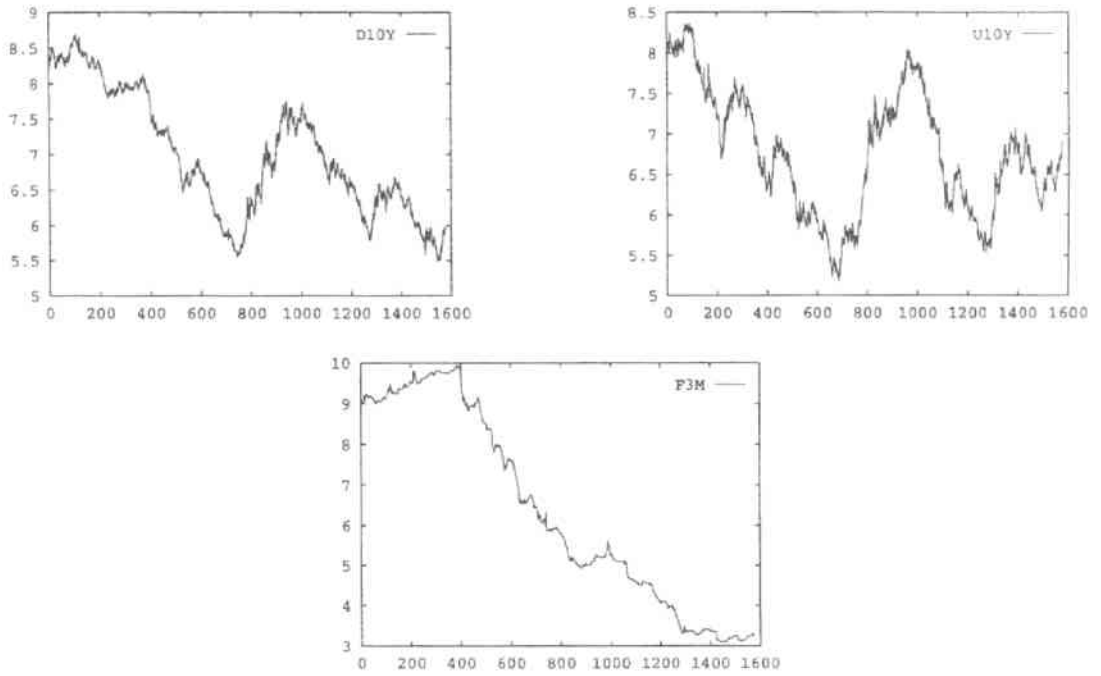


Abbildung 4.1 Tageswert der verwendeten Zeitreihen

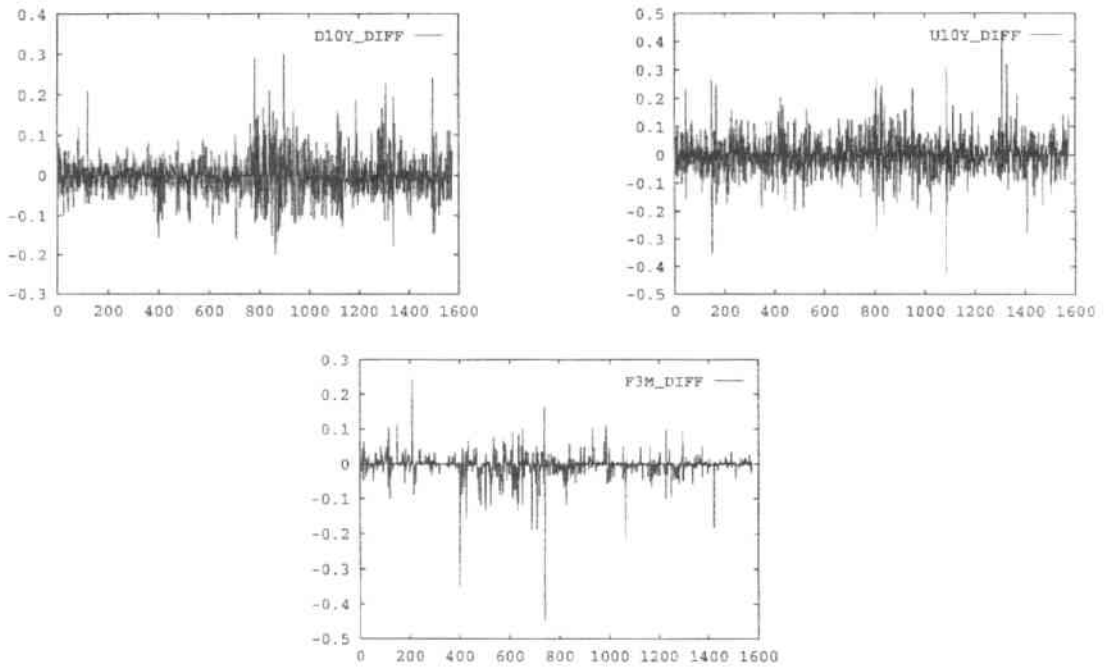


Abbildung 4.2 Tagesdifferenz der verwendeten Zeitreihen

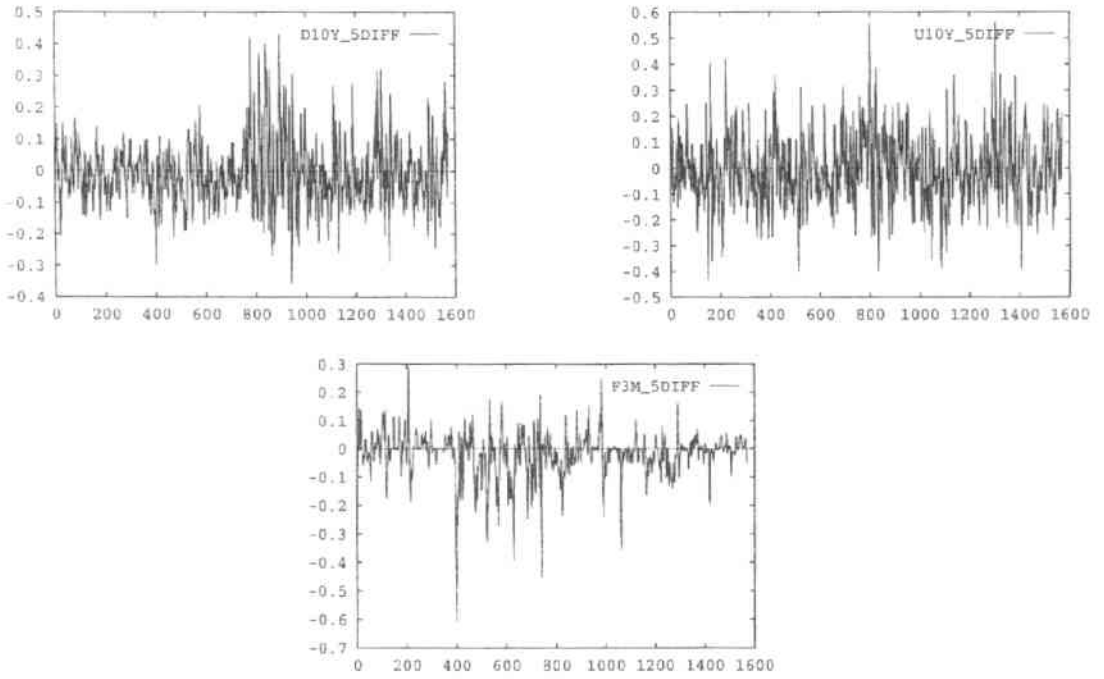


Abbildung 4.3 Wochendifferenz der verwendeten Zeitreihen

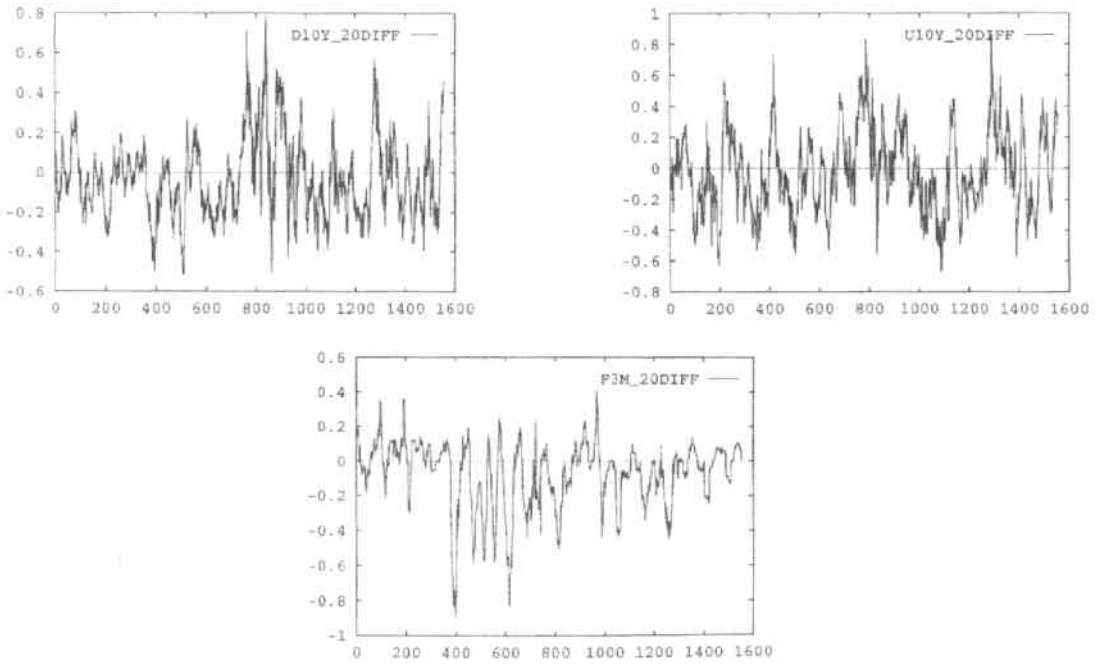


Abbildung 4.4 Monatsdifferenz der verwendeten Zeitreihen

Korrelation

Tabelle 4.1 stellt die Korrelationskoeffizienten der Zeitreihen D10Y und U10Y bzw. D10Y und F3M dar².

	Lags	U10Y	F3M
D10Y	0	0,788	0,758
	-1	0,785	0,758
	-2	0,781	0,759
	-3	0,778	0,759
	-4	0,775	0,760
	-5	0,772	0,761
	-20	0,722	0,767

*Tabelle 4.1 Korrelationskoeffizient von D10Y und U10Y
bzw. von D10Y und F3M*

Bis zur 20-Tage-Verschiebung zeigt es hohe Korrelationen zwischen D10Y und den anderen beiden Zeitreihen. Die Entwicklung von D10Y scheint mit der Entwicklung von U10Y und von F3M stark abhängig zu sein.

Tabelle 4.2 stellt die Korrelationskoeffizienten der Differenzreihen D10Y_DIFF und U10Y_DIFF bzw. D10Y_DIFF und F3M_DIFF dar. Die Korrelationen der Differenzreihen sind zwar nicht ignorierbare (zwischen D10Y_DIFF und U10Y_DIFF bzw. D10Y_DIFF und F3M_DIFF ohne Lags jeweils 0,226 und 0,116), aber sie sind wesentlich kleiner als die Korrelationen der ursprünglichen Zeitreihen. Faßt man die Ergebnisse aus den beiden Tabellen zusammen, stellt man fest, daß D10Y zwar einem ähnlichen Trend wie U10Y bzw. F3M folgt, aber der lineare Zusammenhang zwischen ihrer täglichen Veränderung ist vergleichsmäßig schwach.

²Mit X-n wird eine aus einer Verschiebung von Zeitreihe X für n Tage nach hinten ergebene Zeitreihe bezeichnet.

	Lags	U10Y_DIFF	F3M_DIFF
D10Y_DIFF	0	0,226	0,116
	-1	-0,014	0,143
	-2	-0,004	0,030
	-3	-0,031	0,0
	-4	-0,004	0,024
	-5	-0,016	0,038
	-20	-0,004	0,017

*Tabelle 4.2 Korrelationskoeffizient von D10Y_DIFF und U10Y_DIFF
bzw. D10Y_DIFF und F3M_DIFF*

Tabelle 4.3 stellt die Korrelationskoeffizienten der Mehrtagdifferenzreihen dar. Auch hier zeigt sich, mit dem Vergleich zu den ursprünglichen Zeitreihen ein deutlich niedriger Korrelationsgrad. Das heißt, obwohl D10Y langfristig dem ähnlichen Trend mit U10Y und F3M folgt, hängen die kurzfristige entwicklungen der Zeitreihen linear nicht so stark voneinander ab. Um dennoch ein Modell für eine kurzfristige Prognose mit diesen drei Zeitreihen aufbauen zu können, hat die Diskussion in Kapitel 2 die Anwendung von Fehlerkorrekturmodell resultiert.

	U10Y_nDIFF*	F3M_nDIFF
D10Y_2DIFF	0,337	0,170
D10Y_3DIFF	0,363	0,191
D10Y_5DIFF	0,394	0,208
D10Y_20DIFF	0,476	0,238

* : Der n in U10Y_nDIFF und F3M_nDIFF nimmt den Wert an, der dem Wert von n in D10Y_nDIFF entspricht.

Tabelle 4.3 Korrelationskoeffizient von Mehrtagdifferenzreihe

4.2.2 Kointegration

Im Kapitel 2 ist ein Zwei-Schritt-Verfahren zum Schätzen von Fehlerkorrekturmodellen von Engle und Granger vorgestellt worden. Im ersten Schritt wird auf die Kointegration der Zeitreihe getestet. Wenn der Test bestanden ist, wird der Kointegrationsvektor

geschätzt. Nach (2.3) ergibt sich die folgende Gleichgewichtshypothese:

$$D10Y_t = \beta_0 + \beta_1 U10Y_t + \beta_2 F3M_t \tag{4.4}$$

Es muß geprüft werden, ob (4.4) eine gültige Kointegrationsbeziehung darstellt und somit die Voraussetzung zur Spezifikation eines Fehlerkorrekturmodells erfüllt ist. Zunächst ist es dazu erforderlich, den Intergrationsgrad der einzelnen Zeitreihe zu bestimmen. Der ADF-Test wurde auf D10Y, U10Y und F3M durchgeführt. Dabei wurde getestet, ob die drei Zeitreihen nichtstationäre Prozesse darstellen. Somit lautete die Hypothese:

$$H_0 : D10Y, U10Y \text{ und } F3M \text{ stellen jeweils einen nichtstationären Prozeß dar.}$$

Die Testergebnisse befinden sich in Tabelle 4.4.

In der zweiten Spalte der Tabelle 4.4 sind die von dem ADF-Test ermittelte t-Werte für jede Zeitreihe gezeigt. In den letzten drei Spalten sind die kritische Werte für drei Irrtumswahrscheinlichkeiten 1%, 5% sowie 10% gezeigt. Da die ermittelten t-Werte generell unter 0 liegen, wird auch nur die untere Grenze der jeweiligen Annahmebereiche gezeigt. Wenn der t-Wert einer Zeitreihe vom ADF-Test kleiner als die untere Grenze eines Annahmebereich ist, dann ist es mit der zugehörigen Irrtumswahrscheinlichkeit abzulehnen, daß diese Zeitreihe einen nichtstationären Prozeß darstellt, also ist sie dann ein stationärer Prozeß.

	t-Wert von ADF	kritischer t-Wert		
		1%	5%	10%
D10Y	-1,731	-3,969	-3,415	-3,130
U10Y	-1,857	-3,969	-3,415	-3,130
F3M	-1,590	-3,969	-3,415	-3,130

Tabelle 4.4 Ergebnisse der ADF-Tests auf betreffende Zeitreihen

Laut den Testergebnissen aus Tabelle 4.4 stellt man fest, daß die t-Werte allen drei Zeitreihen im Annahmebereich der Hypothese mit einer Irrtumswahrscheinlichkeit von 10% liegen. Das bedeutet: die Hypothese, daß D10Y, U10Y und F3M jeweils einen nichtstationären Prozeß darstellen, kann bei einer Irrtumswahrscheinlichkeit von 10% nicht

verworfen werden³. Somit sind die Zeitreihen nichtstationär. Um Stationarität zu erzielen, ist deshalb eine Differenzenbildung notwendig. Führt man den gleichen ADF-Test auf Differenzreihen durch (Ergebnisse in Tabelle 4.5), zeigt sich, daß die t-Werte aller Differenzreihen außerhalb des Annahmebereichs mit einer Irrtumswahrscheinlichkeit von 1% liegen. Die Hypothese, daß die betreffende Differenzreihen nichtstationäre Prozesse darstellen, kann bei einer Irrtumswahrscheinlichkeit von maximal 1% abgelehnt werden. Die Differenzreihen sind damit stationär. Und nach Definition 2.3 sind die ursprüngliche Zeitreihen D10Y, U10Y und F3M $I(1)$. Die grundlegende Bedingung, daß nur Zeitreihen des gleichen Integrationsgrades kointegriert sein können, sind damit erfüllt.

	t-Wert von ADF	kritischer t-Wert		
		1%	5%	10%
D10Y_DIFF	-41,92	-2,57	-1,94	-1,62
U10Y_DIFF	-40,93	-2,57	-1,94	-1,62
F3M_DIFF	-34,63	-2,57	-1,94	-1,62

Tabelle 4.5 Ergebnisse der ADF-Tests auf Differenzreihen

In einem nächsten Schritt kann der Kointegrationsvektor für (4.4) geschätzt werden. Das Kleinste-Quadrate-Verfahren brachte die nachstehenden Ergebnisse. Die t-Werte der Koeffizienten sind unter den zugehörigen Koeffizienten im Klammer geschrieben.

$$D10Y_t = 1.21 + 0,67U10Y_t + 0,20F3M_t + err_t, \quad (4.5)$$

(18.2) (64.1) (58.9)

Der kritische t-Wert für die Ablehnung der Hypothese, daß die geschätzten Koeffizienten nicht signifikant von 0 unterschiedlich sind, beträgt mit einer Irrtumswahrscheinlichkeit von maximal 0.05% 3,291. D.h. der Annahmebereich dieser Hypothese ist $[-3,291; 3,291]$. Die berechnete t-Werte der Koeffizienten sind weit außer dem Annahmebereich, also die geschätzten Koeffizienten sind signifikant von 0 unterschiedlich. Damit ist es bewiesen, daß die geschätzten Koeffizienten nicht durch einen Zufall entstanden sind, sondern ein stabiles Gleichgewicht nach der Gleichung (4.4) bilden. err_t in der Gleichung (4.5) ist die Abweichung der langfristigen Beziehung zwischen D10Y, U10Y und F3M. Der Verlauf von err_t ist in Abbildung 4.5 graphisch dargestellt.

³Wenn man diese Hypothese doch noch verwirft, macht man mit einer Wahrscheinlichkeit bis zu 10% Fehler. Das ist in der Statistik nicht akzeptierbar.

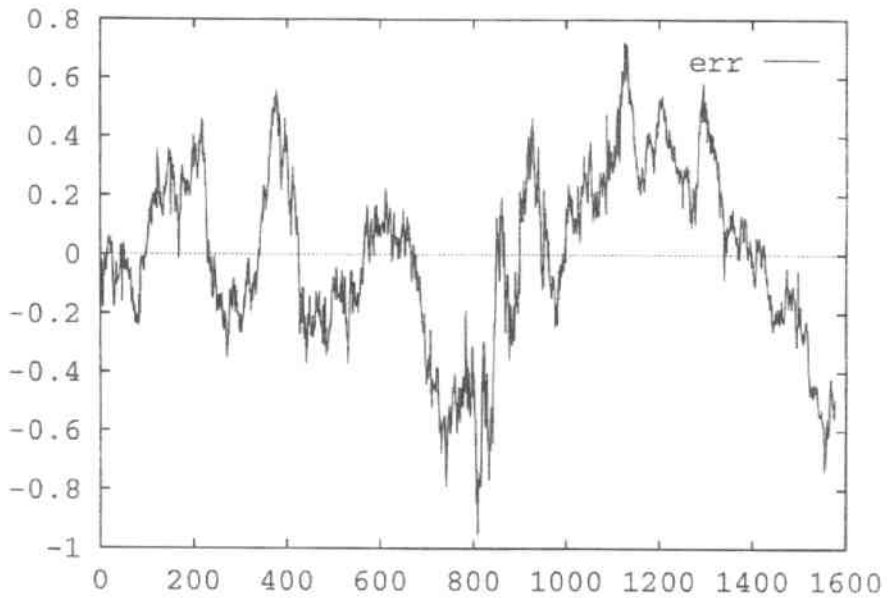


Abbildung 4.5 Abweichung der langfristigen Beziehung von $D10Y$, $U10Y$ und $F3M$

Nun kann geprüft werden, ob sich für (4.4) eine Kointegrationsbeziehung nachweisen läßt. Eine stabile Kointegrationsbeziehung zwischen den Modellvariablen in (4.4) ist nur dann empirisch haltbar, wenn err_t als stationärer Prozeß aufgefaßt werden kann. Es muß also getestet werden, ob err_t integriert vom Grad 0 ist. Ergebnisse aus Tabelle 4.6 können mit einer Irrtumswahrscheinlichkeit von maximal 1% die Hypothese der Nicht-Stationarität verwerfen. Damit ist (4.5) als konsistente Schätzung der langfristigen Beziehung zwischen den Modellvariablen aufzufassen.

	t-Wert von ADF	kritischer t-Wert		
		1%	5%	10%
err_t	-3.57	-2.57	-1.94	-1.62

Tabelle 4.6 Ergebnisse des ADF-Tests auf err_t

4.2.3 Fehlerkorrekturmodell

Modellaufbau

Aufgrund der Aussage des Granger-Repräsentations-Theorems, daß kointegrierte Zeitreihen immer eine Fehlerkorrekturdarstellung besitzen, wird im zweiten Schritt des von Engle und Granger vorgestellten Verfahrens nach (2.5) folgendes Fehlerkorrekturmodell konstruiert⁴:

$$\begin{aligned} (1 - L^b)D10Y_t &= \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1 - L^b)D10Y_{t-bi} + \sum_{i=1}^m \alpha_{2i}(1 - L^b)U10Y_{t-bi} \\ &+ \sum_{i=1}^k \alpha_{3i}(1 - L^b)F3M_{t-bi} + \alpha_{41}err_{t-b} \end{aligned} \quad (4.6)$$

Entsprechend der Zielsetzung der Studie, Prognosen auf Sicht von 1 Tag, 1 Woche (5 Tage) bzw. 1 Monat (4 Wochen) zu erstellen, erfolgt die Modellspezifikation auf Basis von 1, 5 bzw. 20 Tag(en):

$$\begin{aligned} (1 - L)D10Y_t &= \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1 - L)D10Y_{t-i} + \sum_{i=1}^m \alpha_{2i}(1 - L)U10Y_{t-i} \\ &+ \sum_{i=1}^k \alpha_{3i}(1 - L)F3M_{t-i} + \alpha_{41}err_{t-1} \end{aligned} \quad (4.7)$$

$$\begin{aligned} (1 - L^5)D10Y_t &= \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1 - L^5)D10Y_{t-5i} + \sum_{i=1}^m \alpha_{2i}(1 - L^5)U10Y_{t-5i} \\ &+ \sum_{i=1}^k \alpha_{3i}(1 - L^5)F3M_{t-5i} + \alpha_{41}err_{t-5} \end{aligned} \quad (4.8)$$

$$\begin{aligned} (1 - L^{20})D10Y_t &= \alpha_{00} + \sum_{i=1}^n \alpha_{1i}(1 - L^{20})D10Y_{t-20i} + \sum_{i=1}^m \alpha_{2i}(1 - L^{20})U10Y_{t-20i} \\ &+ \sum_{i=1}^k \alpha_{3i}(1 - L^{20})F3M_{t-20i} + \alpha_{41}err_{t-20} \end{aligned} \quad (4.9)$$

⁴Um das Problem zu vereinfachen, wird der Störterm ε_t in (2.5) in den empirischen Untersuchungen nicht betrachtet.

Parameterschätzung und Prognose

Ausgehend von einem in Bezug auf die Lagstruktur relativ breiten Ansatz wurde anhand einer Ausschöpfungsstrategie eine systematische Modellvereinfachung vorgenommen. Bei dieser Ausschöpfungsstrategie wird die Anzahl der Regressoren in der Modellstellung sukzessiv reduziert. Man schätzt für das bestehende Modell die Parameter, berechnet die entsprechende t-Werte und streicht dann die Regressoren mit nicht signifikanten Koeffizienten. Dieser Prozeß wird solange wiederholt bis das Endmodell mit nur signifikanten Regressoren erreicht wird. An dieser Stelle ist anzumerken, daß bewußt ausschließlich verzögerte Regressoren in das Modell aufgenommen wurden. Damit liegen zum Zeitpunkt t die benötigten Inputs zur Prognose der Zinsveränderung in $t+1$ (1 steht für einen Zeiteinheit, also in unserem Fall jeweils 1 Tag, 1 Woche oder 1 Monat) bereits vor.

Mit bis zu 5 Zeiteinheiten verzögerten Variablen bilden wir das Ausgangsmodell. Für die Tagesbasis sieht es so aus:

$$(1-L)D10Y_t = \alpha_{00} + \sum_{i=1}^5 \alpha_{1i}(1-L)D10Y_{t-i} + \sum_{i=1}^5 \alpha_{2i}(1-L)U10Y_{t-i} + \sum_{i=1}^5 \alpha_{3i}(1-L)F3M_{t-i} + \alpha_{41}err_{t-1} \quad (4.10)$$

Wir schätzen die Koeffizienten a_{ij} in der Gleichung (4.10) mit dem Kleinste-Quadrate-Schätzer. Tabelle 4.7 zeigt die geschätzten Koeffizienten und die entsprechenden t-Werte.

Der kritische t-Wert für die Ablehnung der Hypothese, daß die geschätzten Koeffizienten nicht signifikant von 0 unterschiedlich sind, beträgt für eine Irrtumswahrscheinlichkeit von maximal 15% 1,04. Also sind die Regressoren, dessen geschätzten Koeffizient einen t-Wert kleiner als 1,04 hat, definitiv nicht signifikant. Durch Eliminierung dieser Regressoren bekommen wir eine neue Gleichung:

$$(1-L)D10Y_t = \alpha_{00} + \alpha_{11}(1-L)D10Y_{t-1} + \alpha_{14}(1-L)D10Y_{t-4} + \alpha_{21}(1-L)U10Y_{t-1} + \alpha_{24}(1-L)U10Y_{t-4} + \alpha_{31}(1-L)F3M_{t-1} + \alpha_{34}(1-L)F3M_{t-4} + \alpha_{35}(1-L)F3M_{t-5} + \alpha_{41}err_{t-1} \quad (4.11)$$

Die Koeffizienten a_{ij} in der Gleichung (4.11) werden nochmal geschätzt. Tabelle 4.8 zeigt die geschätzten Koeffizienten und die entsprechenden t-Werte.

Regressor	Koeffizient	geschätzter Koeffizient	t-Wert
	α_{00}	-0.003	-2,17
$(1-L)D10Y_{t-1}$	α_{11}	-0,079	-2,17
$(1-L)D10Y_{t-2}$	α_{12}	-0,032	-0,86
$(1-L)D10Y_{t-3}$	α_{13}	0,014	0,38
$(1-L)D10Y_{t-4}$	α_{14}	0,050	1,32
$(1-L)D10Y_{t-5}$	α_{15}	-0,001	-0,02
$(1-L)U10Y_{t-1}$	α_{21}	0,104	4,81
$(1-L)U10Y_{t-2}$	α_{22}	0,014	0,65
$(1-L)U10Y_{t-3}$	α_{23}	0,015	0,70
$(1-L)U10Y_{t-4}$	α_{24}	0,033	1,49
$(1-L)U10Y_{t-5}$	α_{25}	0,021	0,97
$(1-L)F3M_{t-1}$	α_{31}	-0,037	-1,07
$(1-L)F3M_{t-2}$	α_{32}	-0,021	-0,61
$(1-L)F3M_{t-3}$	α_{33}	-0,017	-0,51
$(1-L)F3M_{t-4}$	α_{34}	0,091	2,64
$(1-L)F3M_{t-5}$	α_{35}	0,057	1,70
err_{t-1}	α_{41}	-0,016	-3,16

Tabelle 4.7 Schätzergebnis des Ausgangmodells

Regressor	Koeffizient	geschätzter Koeffizient	t-Wert
	α_{00}	-0.003	-2,24
$(1-L)D10Y_{t-1}$	α_{11}	-0,068	-1,93
$(1-L)D10Y_{t-4}$	α_{14}	0,050	1,40
$(1-L)U10Y_{t-1}$	α_{21}	0,102	4,77
$(1-L)U10Y_{t-4}$	α_{24}	0,032	1,50
$(1-L)F3M_{t-1}$	α_{31}	-0,045	-1,35
$(1-L)F3M_{t-4}$	α_{34}	0,086	2,59
$(1-L)F3M_{t-5}$	α_{35}	0,053	1,59
err_{t-1}	α_{41}	-0,018	-3,59

Tabelle 4.8 Schätzergebnis des Zwischenmodells

Der kritische t-Wert für eine Irrtumswahrscheinlichkeit von maximal 5% beträgt 1,65. Eliminieren wir die Regressoren, deren geschätzter Koeffizient einen t-Wert kleiner als 1,65 hat, haben wir wieder eine neue Gleichung:

$$(1 - L)D10Y_t = \alpha_{00} + \alpha_{11}(1 - L)D10Y_{t-1} + \alpha_{21}(1 - L)U10Y_{t-1} \\ + \alpha_{34}(1 - L)F3M_{t-4} + \alpha_{41}err_{t-1}$$

Dieser Prozeß wird so weit durchgeführt, bis ein kleineres parametrisiertes Endmodell erreicht wird, das nur signifikante Regressoren enthält. Dabei ist zu beachten, daß man mit dem Streichen der Regressoren aufhören sollte, bevor einer der grundlegenden Faktoren ganz aus dem Modell verschwindet. D.h., in unserem Fall müssen $D10Y_{t-i}$, $U10Y_{t-i}$ und $F3M_{t-i}$ jeweils mindestens einmal im Endmodell auftauchen. Wenn diese Bedingung nicht erfüllt werden kann, ohne die nicht signifikanten Regressoren im Modell stehen zu lassen, muß man dann die grundlegenden Faktoren neu wählen.

Anhand der o.g. Vorgehensweise wurde schließlich mit

$$(1 - L)D10Y_t = -0,003 - 0,069(1 - L)D10Y_{t-1} + 0,103(1 - L)U10Y_{t-1} \\ + 0,093(1 - L)F3M_{t-4} - 0,019err_{t-1} \quad (4.12)$$

$$(1 - L^5)D10Y_t = -0,010 + 0,073(1 - L^5)D10Y_{t-10} + 0,125(1 - L^5)D10Y_{t-20} \\ + 0,096(1 - L^5)U10Y_{t-5} + 0,135(1 - L^5)F3M_{t-25} \\ - 0,079err_{t-5} \quad (4.13)$$

$$(1 - L^{20})D10Y_t = -0,051 + 0,227(1 - L^{20})D10Y_{t-20} + 0,213(1 - L^{20})D10Y_{t-80} \\ - 0,114(1 - L^{20})U10Y_{t-60} + 0,101(1 - L^{20})F3M_{t-40} \\ - 0,357err_{t-20} \quad (4.14)$$

jeweils ein wesentlich kleineres parametrisiertes Fehlerkorrekturmodell geschätzt. Die t-Werte der geschätzten Koeffizienten sind in Tabelle 4.9 zusammengestellt. Die kritischen t-Werte für die Irrtumswahrscheinlichkeiten von maximal 2,5% und 0,05% betragen 1,96 bzw. 3,29. Damit sind die Regressoren in den Endmodellen für die Tages- und Wochenbasis mit einer Wahrscheinlichkeit von mindestens 97,5% signifikant. Die Regressoren in

den Endmodellen für die Monatsbasis sind mit einer Wahrscheinlichkeit von mindestens 99,95% signifikant. Dabei ist besonders zu beachten, daß der Abweichungsterm err_t im Endmodell für alle drei Basen für eine Wahrscheinlichkeit von mindestens 99,95% signifikant ist. Abweichungen der langfristigen Gleichgewichtsbeziehung in t haben demnach einen gewichtigen Einfluß hinsichtlich der Zinsveränderung von t auf $t+1$.

Periode	Regressor	Koeffizient	geschätzter Koeffizient	t-Wert
Tag		α_{00}	-0,003	-2,55
	$(1 - L)D10Y_{t-1}$	α_{11}	-0,069	-1,96
	$(1 - L)U10Y_{t-1}$	α_{21}	0,103	4,80
	$(1 - L)F3M_{t-4}$	α_{34}	0,093	2,86
	err_{t-1}	α_{41}	-0,019	-3,83
Woche		α_{00}	-0,010	-3,70
	$(1 - L)D10Y_{t-10}$	α_{12}	0,073	2,12
	$(1 - L)D10Y_{t-20}$	α_{14}	0,125	3,65
	$(1 - L)U10Y_{t-5}$	α_{21}	0,096	4,49
	$(1 - L)F3M_{t-25}$	α_{35}	0,135	5,16
	err_{t-5}	α_{41}	-0,079	-7,53
Monat		α_{00}	-0,051	-7,04
	$(1 - L)D10Y_{t-20}$	α_{11}	0,227	6,39
	$(1 - L)D10Y_{t-80}$	α_{14}	0,213	5,47
	$(1 - L)U10Y_{t-60}$	α_{23}	-0,114	-4,87
	$(1 - L)F3M_{t-40}$	α_{32}	0,101	4,38
	err_{t-20}	α_{41}	-0,357	-14,81

Tabelle 4.9 Schätzergebnisse der Fehlerkorrekturmodellen

Wenn man den Abweichungsterm ($\alpha_{41}err_{t-1}$) in der Gleichung (4.6) wegläßt, hat man ein lineares Modell ohne Information über langfristige Gleichgewichtsbeziehungen der Zeitreihen. Da wir solches Modell mit dem Fehlerkorrekturmodell vergleichen wollen, müssen wir auch ihn zur Prognose einsetzen. Mit dem oben beschriebenen Ausschöpfungsprozeß wurde dieses Modell für alle drei Basen geschätzt. Die nachstehende Gleichungen stellen die jeweilige Endmodelle dar. Die t-Werte der geschätzten Koeffizienten sind in Tabelle 4.10 zusammengestellt:

$$(1 - L)D10Y_t = -0,002 - 0,067(1 - L)D10Y_{t-1} + 0,118(1 - L)U10Y_{t-1} \\ + 0,044(1 - L)U10Y_{t-4} + 0,091(1 - L)F3M_{t-4} \quad (4.15)$$

$$(1 - L^5)D10Y_t = -0,006 + 0,150(1 - L^5)D10Y_{t-20} + 0,148(1 - L^5)U10Y_{t-5} \\ + 0,095(1 - L^5)U10Y_{t-10} + 0,051(1 - L^5)U10Y_{t-15} \\ + 0,109(1 - L^5)F3M_{t-25} \quad (4.16)$$

$$(1 - L^{20})D10Y_t = -0,094 - 0,311(1 - L^{20})D10Y_{t-60} - 0,189(1 - L^{20})D10Y_{t-100} \\ + 0,071(1 - L^{20})U10Y_{t-100} + 0,070(1 - L^{20})F3M_{t-80} \\ + 0,199(1 - L^{20})F3M_{t-100} \quad (4.17)$$

Periode	Regressor	Koeffizient	geschätzter Koeffizient	t-Wert
Tag		α_{00}	-0,002	-1,826
	$(1 - L)D10Y_{t-1}$	α_{11}	-0,067	-1,911
	$(1 - L)U10Y_{t-1}$	α_{21}	0,118	5,558
	$(1 - L)U10Y_{t-4}$	α_{24}	0,044	2,065
	$(1 - L)F3M_{t-4}$	α_{34}	0,091	2,77
Woche		α_{00}	-0,006	-2,04
	$(1 - L)D10Y_{t-20}$	α_{14}	0,150	4,25
	$(1 - L)U10Y_{t-5}$	α_{21}	0,148	7,13
	$(1 - L)U10Y_{t-10}$	α_{22}	0,095	4,56
	$(1 - L)U10Y_{t-15}$	α_{23}	0,051	2,43
	$(1 - L)F3M_{t-25}$	α_{35}	0,109	4,08
Monat		α_{00}	-0,094	-13,66
	$(1 - L)D10Y_{t-60}$	α_{13}	-0,311	-8,29
	$(1 - L)D10Y_{t-100}$	α_{15}	-0,189	-4,18
	$(1 - L)U10Y_{t-100}$	α_{25}	0,071	2,63
	$(1 - L)F3M_{t-80}$	α_{34}	0,070	2,92
	$(1 - L)F3M_{t-100}$	α_{35}	0,199	7,45

Tabelle 4.10 Schätzergebnisse der linearen Modellen nach Box und Jenkins

Zur Ableitung der Prognosewerte wurde Gleichung (4.12) - (4.17) zunächst über den Zeitraum 01.03.1991 bis 28.02.1994 geschätzt und daraufhin die Zinsveränderung von jeweils 1 Tag, 1 Woche bzw. 1 Monat für den Zeitraum 01.03.94 bis 31.03.97 vorhergesagt. Die Prognoseergebnisse der beiden Methoden sind in Tabelle 4.10 zusammengefasst.

Periode	ECM (mit err_t)			Modelle ohne err_t		
	Quadratischfehler		Treffer in %	Quadratischfehler		Treffer in %
	Test	Train		Test	Train	
Tag	1,209	0,450	64,610	1,214	0,456	62,972
Woche	6,108	1,847	54,177	6,371	1,946	51,266
Monat	21,742	6,223	67,799	32,191	6,682	55,220

Tabelle 4.11 Zinsprognose mit linearen ökonomischen Modellen

In allen drei Fällen sind die Prognoseleistungen des Fehlerkorrekturmodells (mit dem Abweichungsterm err_t) gegen das Modell ohne Information über langfristige Gleichgewichtsbeziehung der Zeitreihen (ohne err_t) überlegen. Damit ist das Granger-Repräsentations-Theorem bestätigt worden. Wenn die betreffenden Zeitreihen kointegriert sind, dann kann der Abweichungsterm die in den Niveauvariablen enthaltenen langfristigen Informationen übertragen und damit die Prognoseleistung erhöhen. Die Prognosen für Tages- und Monatsbasis haben den Trend der Zinsentwicklung über 64% erkannt. Daß die Prognosen für Wochenbasis deutlich weniger erkannt haben, mag an den in den Wochen enthaltenen Feiertagen liegen. Da wir permanent 5 Tage für eine Woche rechneten, wurde bei einer Woche der tatsächliche Wochenzyklus zerstört, d.h. nicht mehr vom Montag auf den nächsten Montag sondern vom Montag auf den nächsten Dienstag prognostiziert, wenn in dieser Woche ein Feiertag war. Auf Tagesbasis hat man dieses Problem offensichtlich nicht. Und auf Monatsbasis sind die Feiertage wahrscheinlich aufgrund des längeren Zyklus weniger wichtig, da je länger der Zyklus ist, desto weniger wirkt einzelner Tag auf den gesamten Zyklus.

An dieser Stelle ist anzumerken, daß sich Fehlerkorrekturmodelle als ein brauchbares Instrument zur Prognose von Zinsentwicklungen erwiesen haben. Es konnte gezeigt werden, daß eine langfristig stabile Beziehung zwischen D10Y, U10Y und F3M besteht. Abweichungen der langfristigen Gleichgewichtsbeziehung sind damit für kurzfristige Prognosen von hohem Informationswert.

Kapitel 5

Zinsprognose mit Neuronalen Netzen

In Kapitel 3 wurden Neuronale Netze als Prognoseverfahren für Zeitreihen vorgestellt. Im Weiteren soll die Eignung Neuronaler Netze zur Prognose der Zinsentwicklung auf der Tages-, Wochen- bzw. Monatsbasis untersucht werden. Das Ziel dabei ist herauszufinden, ob neuronale Netze in diesem Fall der klassischen ökonomischen Verfahren (hier speziell Fehlerkorrekturmodell) gegenüber eine wettbewerbsfähige Methode darstellen. Dies bezüglich werden die gleichen Daten verwendet, die zuvor auch im Fehlerkorrekturmodell verwendet wurden.

5.1 Vorgehensweise

Zunächst wird die Vorgehensweise dargestellt, nach der die Anwendung der neuronalen Netzwerke zur Zinsprognose durchgeführt ist. Dabei wird auf die einzelnen Punkte des in Abschnitt 3.3 vorgestellten Modellierungsprozesses eingegangen.

5.1.1 Voranalyse der Daten

Zu der Voranalyse der Daten gehören vor allem statistische Untersuchungen, die in Kapitel 4 während des Aufbaus des Fehlerkorrekturmodells schon durchgeführt wurden. Ferner kommt bei den neuronalen Netzen der Variablenidentifikation eine große Bedeutung zu.

Bei der Anwendung von neuronalen Netzen hat man die Hoffnung, daß die zwischen Eingabe- und Ausgabevariablen vorhandene Zusammenhänge abgebildet werden können. Man vermutet, daß die Bewegungen im Finanzmarkt hauptsächlich von dem Zusammenwirken einiger Faktoren verursacht sind. Das heißt aber nicht, daß man alle verfügbaren

und in irgendeinem vermuteten Zusammenhang stehenden Variablen als Eingaberepräsentation benutzen soll. Die hohe Erwartung, daß das neuronale Netz die Fähigkeit besitzt, die wichtigen und richtigen Variablen zu entdecken, hat sich in vielen Fällen als falsch erwiesen [21].

Ein Problem bei der ökonometrischen Anwendung ist die Unklarheit, welche Faktoren im Finanzmarkt die entscheidende Rolle spielen. Somit sind unter der möglichen Eingangsgröße der neuronalen Netze redundante bzw. miteinander korrelierte Variablen vorhanden. Die redundante Information läßt das neuronale Netz verwirren und reduziert seine Generalisierungsfähigkeit. Und bei starker Korrelation sind die involvierten Merkmale nicht mehr klar identifizierbar. Dies führt zu einem schlechten Konvergenz des Lernalgorithmus [16]. Bei der Modellspezifikation muß man die Netzeingabe so bilden, daß sie möglichst informationsenthältig und repräsentativ sein.

Ein anderes Problem ist, daß die optimale Kodierung der Eingangsvariablen nicht ohne weiteres zu erhalten sind. Selbst wenn alle wichtigen Variablen festgestellt werden können, ist ihre natürliche Form nicht unbedingt geeignet für das neuronale Netz. Je besser die in den Eingabevariablen versteckte Information präsentiert ist, desto schneller kann das neuronale Netz lernen. Um die optimale Darstellung der verborgenen Information zu erhalten, werden zahlreiche Transformationen auf die Eingangsgröße angewandt. In [26] wurde gezeigt, daß bei gleichen Variablen unterschiedliche Kodierungen zu unterschiedlichen Leistungen der neuronalen Netze führten.

In Abschnitt 4.2.1 wurde beim Aufbau der ökonometrischen Modellen zur Prognose der Entwicklung von D10Y letztlich nur U10Y und F3M einbezogen. Da eines unserer Ziele ist, die Anwendung des Fehlerkorrekturmodells und des neuronalen Netzes zur Prognose der Zinsentwicklung zu vergleichen, wurden bei der Anwendung neuronaler Netzes auch nur die drei Zeitreihen betrachtet. Tabelle 5.1 zeigt die Korrelationen der drei Zeitreihen.

	D10Y	U10Y	F3M
D10Y	1	0.788	0.758
U10Y	0.788	1	0.357
F3M	0.758	0.357	1

Tabelle 5.1 Korrelationsmatrix

Die Korrelation zwischen U10Y und F3M ist den anderen gegenüber deutlich kleiner. Damit ist es nicht zu verwerfen, daß U10Y und F3M unterschiedliche Informationen über D10Y enthalten. Die Untersuchungen in Kapitel 4 zeigen, daß D10Y langfristig einem ähnlichen Trend wie U10Y oder F3M folgt, und daß zwischen den drei Zeitreihen eine Kointegrationsbeziehung besteht. Somit ist festzuhalten, daß U10Y und F3M als grundlegende Merkmale ins Netz geführt werden.

Das neuronale Netz hat gegenüber der herkömmlichen ökonomischen Methode einen wichtigen Vorteil — die Flexibilität. Dazu gehört, daß neuronale Netze neben der zugrundeliegenden Zeitreihe auch viele derivative Größen als Eingabe nehmen kann. Durch geeignete Zusammensetzung statistischer Merkmale kann das Verhalten der Zeitreihe besser beschrieben werden. Aus den drei ursprünglichen Zeitreihen kann man eine Vielzahl von Netzeingaben bilden. Es gibt mehrere Möglichkeiten vielfältige Netzeingaben aus den gleichen realen Zeitreihen zu bilden. In dieser Studie werden davon drei wichtige Möglichkeiten untersucht:

1. unterschiedliche Kodierung
2. unterschiedliche Skalierung
3. technische Indikatoren

In Tabelle 5.2 werden einige übliche Transformationen der Zeitreihe in der oben genannten Kategorie aufgelistet.

Durch die verschiedenen Kodierungen ist es möglich, eine Zeitreihe aus mehreren Perspektiven zu sehen. Aus direkter Beobachtung bekommt man die Werte, die die Zeitreihe annimmt. Aber in der Tat ist die Veränderungen der Zeitreihen in Bezug auf ihre Werte viel repräsentativer. Schließlich ist der Ablauf der Zeitreihe durch ihre Veränderungen beschrieben. Und um die Veränderungen der Zeitreihe richtig zu verstehen, ist es wiederum wichtig, das Verhältnis der Veränderungen gegenüber dem Niveau der Zeitreihe zu wissen.

	Bezeichnung	Bedeutung
Kodierung	absdiff	$x_t - x_{t-1}$
	bindiff	$\frac{x_t - x_{t-1}}{ x_t - x_{t-1} }$
	reldiff	$\frac{x_t - x_{t-1}}{x_{t-1}}$
	logdiff	$\ln x_t - \ln x_{t-1}$
Skalierung	scale -n m	lineare Skalierung zwischen -n und m
TI	ma	$\frac{1}{k} \sum_{i=0}^k x_{t-i}$
	schiefe	$\frac{E[x^3]}{\sigma^3}$
	vola	σ^2
	kurtosis	$\frac{E[x^4]}{\sigma^4} - 3$
	rsi	$\frac{\sum_{x_t > x_{t-1}} (x_t - x_{t-1})}{\sum x_t - x_{t-1} }$
	n.diff	$\sum_{k=0}^n \binom{n}{k} x_{t-k} (-1)^k$
	diff(x_t, y_t)	$x_t - y_t$

Tabelle 5.2 Technische Transformationen

Die Skalierung dient dazu, die unterschiedlichen Wertbereiche der zugrundeliegenden Zeitreihen in einen gemeinsamen absoluten Wertbereich zu bringen. Wichtig dabei ist, daß das Verhältnis zwischen den einzelnen Punkten innerhalb einer Zeitreihe nicht geändert werden darf. Das heißt, die positiven Werte dürfen nicht durch die Skalierung negativ werden und auch umgekehrt. Also muß der Nullpunkt dabei fixiert werden. Somit ist garantiert, daß die Veränderungsrichtung der Zeitreihe nicht durch Skalierung künstlich geändert wird. Das Anpassungsverhalten des Netzes wird dadurch nivelliert und Anpassungssprünge, die ihre Ursache in unterschiedlich hohen Absolutwerten der Zeitreihen haben, werden verhindert.

Die technischen Indikatoren sind entweder bekannte Maße in der Statistik zur Beschreibung einer Datenmenge oder wichtige Indikatoren in der Charttechnik zur Entscheidung des Kursverlaufs. Durch solche technische Indikatoren können die Eigenschaften der Zeitreihe näher geklärt werden.

5.1.2 Netzwerkaufbau

Der Netzwerkaufbau ist ein umfangreicher und deswegen auch ein schwieriger Teil des Modellierungsprozesses. Darunter versteht man sowohl die Wahl der Netztopologie als auch die Auswahl eines geeigneten Regularisierungsverfahrens.

Netztopologie

Zur Netztopologie wird, wie in Kapitel 3 schon erwähnt, die vorwärtsgerichtete Dreischicht-Topologie gewählt. Dabei ist die Verbindungsart der Schichten noch unbestimmt. Abbildung 5.1 zeigt die zwei vorgestellten Verbindungsarten¹.

Die zwei Verbindungsarten unterscheiden sich um die direkte Verbindungen von Inputschicht nach Outputschicht. Ein Netz ohne diese Verbindungen eignet eventuell mehr dazu, eine Beziehung, die überwiegend aus nicht-linearer Funktion besteht, zu approximieren. Die direkte Verbindungen von Inputschicht nach Outputschicht ist zur Abbildung des lineare Teils einer Beziehung wahrscheinlich hilfreich. Es wird im folgenden beide Verbindungsarten untersucht.

¹vgl. [16]

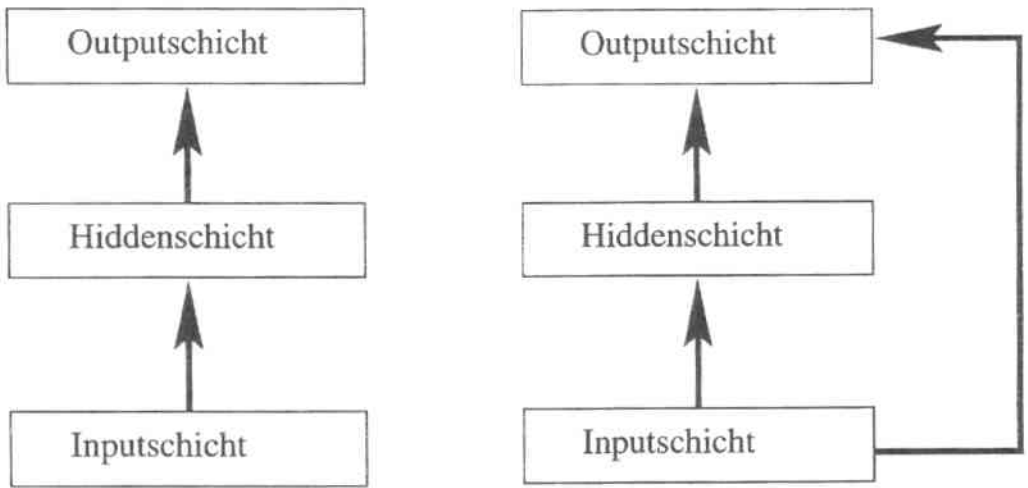


Abbildung 5.1 Zwei Verbindungsarten der Schichten

Letztlich bleibt noch die Frage zu klären, wie groß die Hiddenschicht sein soll? Wie auch viele andere Frage bei neuronalen Netze ist dies problemspezifisch. Es gibt eine Vielzahl an Literatur zur Wahl der entsprechenden Netzwerkgröße. Das Problem dabei ist, daß viele der dort vorgestellten Verfahren nicht frei von Kritik sind und zum Teil nur als Heuristiken angewendet werden können [15]. In unserer Untersuchung werden für jedes Netz mehrere Hiddenschichten mit verschiedenen Größen (3 bis 10 Hiddenunits) getestet. Es folgt die Wahl einer Hiddenschichtgröße aufgrund der Güte auf Testdaten.

Regularisierung

Bei der Anwendung neuronaler Netze ist es nicht unbedingt von großem Interesse, das Netz exakt an die Trainingsdaten anzupassen, sondern vielmehr wünscht man ein neuronales Netz, das die Generalisierungsfähigkeit auf unbekanntem Daten aufweist. Wenn ein Netz zu stark auf der Trainingsmenge trainiert wird, so ist es sehr stark an die Trainingsmenge angepasst, verliert aber die Generalisierungsfähigkeit. Besonders wenn die Daten stark verrauscht sind, ist es nicht so sinnvoll die einzelnen Trainingsdaten so exakt einzulernen, da das Netz dann an die durch das Rauschen verzerrten Daten angepasst wird. Eine sehr starke Anpassung des Netzes an die verzerrten Trainingsdaten führt zu einer großen Verwirrung, wenn das Netz später mit unbekanntem Daten arbeiten soll. Das ist der sogenannte *Overfitting*-Effekt.

Um den Overfitting-Effekt zu umgehen, ist eine Regularisierung notwendig. Hierbei wird zum Fehler E_D zusätzlich die Komplexität des Netzwerkes bestraft. Die zu minimierende Zielfunktion besteht daher nicht mehr einfach aus dem zuvor gewählten Fehlermaß aus den Trainingsdaten, sondern auch zusätzlich aus einer Funktion bezüglich der Komplexität des Netzes. Die Komplexität ist dabei abhängig von der Größe und Anzahl der verwendeten Gewichte. Die zu minimierende Zielfunktion ergibt sich als:

$$E(D, w) = \beta E_D(D, w) + \alpha E_W(w) \quad (5.1)$$

$E_D(D, w)$ ist die Summe der Fehler aus den gesamten Daten D , $E_W(w)$ der entsprechende Bestrafungsterm in Abhängigkeit der Gewichte w . Die Parameter α und β gewichten zwischen der Straffunktion und dem Fehler aus den Daten. Ein einfacher und weit verbreiteter Strafterm ist *Weight-Decay*, der über alle W quadrierten Gewichte summiert:

$$E_W(w) = \frac{1}{2} \sum_{i=1}^W w_i^2.$$

Der Weight-Decay Term bestraft dabei die Gewichte um so stärker, je weiter sie von Null abweichen.

Die Schwierigkeit, bei der Anwendung von Regularisierungstechniken wie Weight-Decay ist, daß für die Gewichtsparameter α und β entsprechende Werte gefunden werden müssen. In Rahmen dieser Arbeit wird eine in [25] vorgestellte Methode untersucht, mit der man durch die Anwendung der Bayes'schen Inferenz analytisch zu Lösungen für α und β kommt.

Bayes'sches Lernen

Beim Bayes'schen Lernen wird das Vertrauen in die Größe von verschiedenen Gewichten des Netzes durch eine Wahrscheinlichkeitsverteilung über den Gewichten zum Ausdruck gebracht. Das heißt, daß man voraussetzt, daß die Gewichte eine gewisse Verteilung $p(w)$ annimmt. Nach der Regel von Bayes hat man eine a-posteriori Verteilung:

$$p(w|D) = \frac{p(D|w) \cdot p(w)}{p(D)}.$$

Dabei gibt $p(D|w)$ die Wahrscheinlichkeit an, daß das Modell mit Gewichtsvektor w bei Eingabe x die Ausgabe y erzeugt.

Um die Verteilung $p(w|D)$ auszuwerten, muß man $p(w)$ und $p(D|w)$ voraussetzen. Für

$p(w)$ wählt man im einfachsten Fall eine Normalverteilung um 0 mit Varianz $\frac{1}{\alpha}$. Für die einzelnen Gewichte w_i gilt:

$$p(w_i) = \left(\frac{\alpha}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\alpha}{2}w_i^2}$$

Die Verteilung $p(w)$ ergibt sich dann als Produkt der Verteilungen für die einzelnen Gewichte w_i :

$$\begin{aligned} p(w) &= \prod_{i=1}^W p(w_i) \\ &= \left(\frac{\alpha}{2\pi}\right)^{\frac{W}{2}} e^{-\frac{\alpha}{2} \sum_{i=1}^W w_i^2} \\ &= C_\alpha e^{-\alpha E_W}, \end{aligned}$$

wobei $C_\alpha = \left(\frac{\alpha}{2\pi}\right)^{\frac{W}{2}}$ ist.

Über die Verteilung der Zielwerte machen wir die einfache Annahme, daß sie mit normalverteilten Rauschen mit Mittelwert 0 und Varianz $\frac{1}{\beta}$ behaftet sind. Die Wahrscheinlichkeit des tatsächlichen Zielwerts y bei Eingabe x zu schätzen ist gegeben durch [25]:

$$p(y|x, w) = \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\beta}{2}(f(x, w) - y)^2}.$$

Damit erhält man, falls die Datenpunkte unabhängig sind:

$$\begin{aligned} p(D|w) &= \prod_{i=1}^N p(y_i|x_i, w) \\ &= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\beta}{2} \sum_{i=1}^N (f(x_i, w) - y_i)^2} \\ &= C_\beta e^{-\beta E_D}, \end{aligned}$$

wobei $C_\beta = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}$ ist.

Mit der Regel von Bayes ergibt sich dann für die a-posteriori Verteilung der Gewichte

$$p(w|D) = \frac{1}{p(D)} C_\beta C_\alpha e^{-\beta E_D - \alpha E_W}.$$

Die Funktion $p(w|D)$ nimmt an der Stelle ihren maximalen Wert, wo w nach Beobachtung von D am wahrscheinlichsten ist. Somit maximiert der optimale Gewichtsvektor den Wert dieser Funktion. Da $\frac{1}{p(D)}$, C_α und C_β von den Gewichten nicht abhängen, findet man diesen optimalen Gewichtsvektor, indem man den negativen Exponenten

$$\beta E_D + \alpha E_W$$

minimiert. Die zu minimierende Funktion entspricht der früheren Fehlerfunktion (5.1) mit dem quadratischen Fehlerterm und dem Weight-Decay Strafterm.

Um geeignete Werte für α und β zu finden, wertet man wieder die a-posteriori Verteilung von α und β aus:

$$p(\alpha, \beta|D) = \frac{p(D|\alpha, \beta)p(\alpha, \beta)}{p(D)}.$$

Die a-priori Verteilung für α und β wählt man so, daß sie keinen Wert bevorzugt. Dann findet man die optimalen Werte für α und β , indem man $p(D|\alpha, \beta)$ maximiert. $p(D|\alpha, \beta)$ berechnet man durch:

$$\begin{aligned} p(D|\alpha, \beta) &= \int p(D|w, \alpha, \beta)p(w|\alpha, \beta)dw \\ &= \int p(D|w, \beta)p(w|\alpha)dw \\ &= C_\alpha C_\beta \int e^{-\beta E_D - \alpha E_W} dw \end{aligned}$$

Da E_D und E_W nicht-lineare Funktionen von w sind, ist $e^{-\beta E_D - \alpha E_W}$ nicht integrierbar. Um $p(\alpha, \beta|D)$ noch zu berechnen, bildet man mit $E(w) := \beta E_D + \alpha E_W$ die Taylor-Entwicklung um das Minimum w_{min} [25]:

$$E(w) = E(w_{min}) + \frac{1}{2}(w - w_{min})^T \mathbf{H}(w - w_{min})$$

wobei \mathbf{H} die Hessematrix der gesamten Fehlerfunktion $E = \beta E_D + \alpha E_W$ an der Stelle w_{min} ist².

²Es gilt, $\mathbf{H} = \beta \mathbf{H}_D + \alpha \mathbf{I}$, und \mathbf{H}_D ist die Hessematrix bezüglich E_D an der Stelle w_{min} .

Man erhält dann für das Integral [11]

$$\int e^{-E(w)} dw = e^{-E(w_{min})} (2\pi)^{N/2} |\mathbf{H}|^{-1/2}.$$

Man maximiert nun den Logarithmus von $p(D|\alpha, \beta)$:

$$\begin{aligned} \ln p(D|\alpha, \beta) &= \frac{W}{2} \ln \alpha + \frac{N}{2} \ln \beta - \alpha E_W(w_{min}) - \beta E_D(w_{min}) \\ &\quad - \frac{1}{2} \ln |\mathbf{H}| + \text{const.} \end{aligned}$$

Um die Maxima von $\ln p(D|\alpha, \beta)$ zu finden, rechnet man die partielle Ableitungen und setzt sie gleich 0.

Seien λ_i die Eigenwerte von $\beta \mathbf{H}_D$, dann ergibt partielles Differenzieren nach α und Nullsetzen

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{\partial \ln p(D|\alpha, \beta)}{\partial \alpha} \\ &= \frac{W}{2\alpha} - E_W(w_{min}) - \alpha \frac{\partial}{\partial \alpha} E_W(w_{min}) - \frac{\partial}{\partial \alpha} \frac{1}{2} \ln \left(\prod_i^N (\lambda_i + \alpha) \right) \\ &= \frac{W}{2\alpha} - E_W(w_{min}) - \alpha \frac{\partial}{\partial \alpha} E_W(w_{min}) - \frac{1}{2} \sum_i^N \frac{1}{\lambda_i + \alpha}. \end{aligned}$$

Da λ_i die Eigenwerte von $\beta \mathbf{H}_D$ sind, gilt $\frac{d\lambda_i}{d\beta} = \frac{\lambda_i}{\beta}$. Damit ergibt partielles Differenzieren nach β und Nullsetzen

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{\partial \ln p(D|\alpha, \beta)}{\partial \beta} \\ &= \frac{N}{2\beta} - E_D(w_{min}) - \beta \frac{\partial}{\partial \beta} E_D(w_{min}) - \frac{\partial}{\partial \beta} \frac{1}{2} \ln \left(\prod_i^N (\lambda_i + \alpha) \right) \\ &= \frac{N}{2\beta} - E_D(w_{min}) - \beta \frac{\partial}{\partial \beta} E_D(w_{min}) - \frac{1}{2\beta} \sum_i^N \frac{\lambda_i}{\lambda_i + \alpha}. \end{aligned}$$

An dieser Stelle vernachlässigt man die Abhängigkeit des Minimums w_{min} von α und β , und reduziert diesen Fehler später durch eine iterative Approximation. So hat man

$$0 = \frac{W}{2\alpha} - E_W(w_{min}) - \frac{1}{2} \sum_i^N \frac{1}{\lambda_i + \alpha}$$

$$0 = \frac{N}{2\beta} - E_D(w_{min}) - \frac{1}{2\beta} \sum_i^N \frac{\lambda_i}{\lambda_i + \alpha}$$

Zusammenfassend ergibt sich für α und β :

$$\left\{ \begin{array}{l} \alpha = \frac{1}{2E_W(w_{min})} (W - \sum_i^N \frac{\alpha}{\lambda_i + \alpha}) \\ \beta = \frac{1}{2E_D(w_{min})} (N - \sum_i^N \frac{\lambda_i}{\lambda_i + \alpha}) \end{array} \right. \quad (5.2)$$

α und β werden am Anfang mit kleinen Werten initialisiert. Man trainiert das Netz und findet $w_{min}(\alpha, \beta)$, dann bestimmt man das neue α und β gemäß (5.2). Die beiden Schritte werden abwechselnd wiederholt bis zur Konvergenz von α und β . Durch diesen iterativen Algorithmus kann man die optimalen Werte sowohl für α und β als auch für w bestimmen.

5.1.3 Einlernen des Netzwerks

Die Gewichtsschätzung wird mit dem Verfahren Resilient-Propagation durchgeführt. Weiterhin wird der Bestrafungsterm wie Weight-Decay eingesetzt um den Overfitting zu vermeiden. Dabei wird auch das Bayes'sche Verfahren eingesetzt um die Gewichtsparameter zu berechnen. Die gesamte Daten wird in eine Trainingsmenge und eine Testmenge eingeteilt. Um später die Ergebnisse von neuronalen Netzen mit den Ergebnissen von Fehlerkorrekturmodellen vergleichen zu können, werden für neuronale Netze die gleichen Trainings- und Testzeitraum wie für die Fehlerkorrekturmodelle verwendet. Das Training wird solange fortgeführt, bis die Abbruchbedingung erfüllt ist.

Die Gewichte der Netze werden mit zufälligen Werten aus dem Intervall $[-1; 1]$ initialisiert. Für das Lernverfahren Resilient-Propagation wird der maximale Gewichtsänderungsschritt auf 0,01 festgesetzt.

5.1.4 Netzwerkbeurteilung

Zur Netzwerkbeurteilung werden die Gütemaße, die in Kapitel 3 vorgestellt worden sind, verwendet. Dabei ist noch zu beachten, daß die Ermittlung der Gütemaßstäbe nur anhand der Testdaten erfolgt. Die Gütemaßstäbe, die anhand der Traingdaten ermittelt werden, haben zur Beurteilung der Generalisierungsfähigkeit der Netze keine Bedeutung, da die Netzgewichte im Zuge der Gewichtsschätzung genau für diese Daten optimiert wurden. Sie können daher selten eine valide Abschätzung der tatsächlichen Modellgüter auf unbekanntem Daten darstellen. Die Testdaten stammen aus dem auf den Trainingdaten folgenden Zeitraum. Sie bleiben während der Parameterschätzung dem Netz unbekannt. Für die Testmenge ist die Generalisierungsfähigkeit der Netze entscheidend. Daher dienen die Testdaten der eigentlichen Beurteilung der Prognosegüte eines Netzes.

5.2 Empirische Untersuchung der Zinsentwicklung in DM

In diesem Abschnitt werden neuronale Netze zur Prognose der Zinsentwicklung in D-Mark auf Tages-, Wochen- bzw. Monatsbasis aufgebaut, indem die oben vorgestellten unterschiedlichen Verfahren eingesetzt werden. Um mit dem Fehlerkorrekturmodell vergleichen zu können, wird zuerst der Eingabevektor vom Fehlerkorrekturmodell übernommen und für das neuronale Netz verwendet. Später werden auch andere technische Indikatoren untersucht. Anhand der Gütemaßstäbe wird das beste Modell ausgewählt und daraufhin weitere Verbesserungsmaßnahme durchgeführt.

Bei der Netzwerkmodellbildung wird in der Praxis die in Abbildung 3.3 gezeigt Reihenfolge nicht immer streng eingehalten. So wird zum Beispiel die Verbindungsart der Netze (gehört zum zweiten Schritt) zuerst entschieden, bevor mögliche Eingabevektoren (gehört zum ersten Schritt) getestet werden. Jedes aufgebauten Netz wurde mit mehrere Gewichtsinitialisierungen trainiert. Untersuchungen zeigten, daß die Netze bei unterschiedlichen Gewichtsinitialisierungen sehr stabile Ergebnisse liefern. Aus diesem Grund, wurde bei der Ergebniszusammenfassung immer das Prognoseergebnis von einer beliebiger Gewichtsinitialisierung verwendet.

5.2.1 Verbindungsart der Perzeptronen

In Abbildung 5.1 sind zwei unterschiedliche Verbindungsarten dargestellt. Die zwei Verbindungsarten unterscheiden sich in den direkten Verbindungen, die sogenannten *shortcuts*, von Inputschicht nach Outputschicht. Man vermutet, daß die direkten Verbindungen zwischen der Input- und Outputschicht den linearen Zusammenhang besser approximieren können. Das gute Ergebnis des Prognosetests mit dem linearen Fehlerkorrekturmodell deutet darauf hin, daß der lineare Teil in der gesamten Beziehung zwischen D10Y, U10Y und F3M doch eine bedeutende Rolle spielt. Es muß festgestellt werden, welche Verbindungsart für die Zinsprognose besser geeignet ist. Neuronale Netze mit den zwei Verbindungsarten sind aufgebaut und getestet worden. Tabelle 5.3 stellt das Testergebnis dar.

Periode	mit <i>shortcuts</i>				ohne <i>shortcuts</i>			
	Hid.	Quad. Fehler		Treffer in %	Hid.	Quad. Fehler		Treffer in %
		Test	Train			Test	Train	
Tag	3	1,219	0,450	64,358	3	1,215	0,449	64,736
Woche	5	6,277	1,712	57,722	5	6,522	1,720	55,949
Monat	4	25,829	4,977	69,308	8	30,268	4,438	68,428

Tabelle 5.3 Zinsprognose der Netzen mit zwei verschiedene Verbindungsart

Das Testergebnis zeigt, daß in den Wochen- und Monatsprognosen die Netze mit Shortcuts im Testzeitraum kleine quadratische Fehler und höhere Trefferquote haben und in der Tagesprognose ist es gerade das Gegenteil. Das mag man so verstehen, daß der lineare Zusammenhang zwischen der Ein- und Ausgabe in der langfristigen Beziehung eine größere Rolle als in der kurzfristigen Beziehung spielt. In der weiteren Untersuchung werden dementsprechend für Wochen- und Monatsprognosen Netze mit Shortcuts und für Tagesprognosen Netze ohne Shortcuts verwendet.

5.2.2 Regularisierung

Vergleicht man Tabelle 4.11 mit Tabelle 5.3, stellt man fest, daß das neuronale Netz in der Testphase größere quadratische Fehler als das Fehlerkorrekturmodell erzeugt, obwohl es in der Trainingsphase kleinere quadratische Fehler erzeugt. Dies bedeutet, daß das neuronale Netz zu stark an den Trainingsmustern angepasst ist und dadurch seine

Generallisierungsfähigkeit verliert. Also sind die neuronalen Netze hier übertrainiert. Dies ist in Tabelle 5.4 deutlich gezeigt.

Periode	Neuronales Netz		Fehlerkorrekturmodell	
	Test	Train	Test	Train
Tag	1,215	0,449	1,209	0,450
Woche	6,277	1,712	6,108	1,847
Monat	25,829	4,977	21,742	6,223

Tabelle 5.4 Vergleich der quadratischen Fehlern zwischen NN und ECM

Weight-Decay

Für ein übertrainiertes Netz ist eine Regularisierung notwendig. Wir setzen zunächst die Weight-Decay Technik ein. Gemäß (5.1) wird ein zusätzlicher Strafterm, der durch einen Parameter α gewichtet ist, in der vorherigen Fehlerfunktion eingeführt. Der alte Fehlerterm ist ebenfalls durch einen Parameter β gewichtet. In der praktischen Umsetzung werden α und β in einem Parameter γ umgewandelt

$$\gamma = \frac{\alpha}{\beta}$$

In der ersten Untersuchung stellen wir den Gewichtsparameter γ bei der Netzspezifikation per Hand ein. Für $\gamma = \{0,5; 1; 2\}$ werden Prognosetests durchgeführt. Das Ergebnis des jeweils besten Netzes ist in Tabelle 5.5 zusammengefaßt.

Periode	Hid.	γ	Quad. Fehler		Treffer in %
			Test	Train	
Tag	5	0,5	1,222	0,454	64,232
Woche	7	1	5,883	1,877	57,595
Monat	6	0,5	20,730	5,945	69,308

Tabelle 5.5 Zinsprognose von neuronalen Netzen mit Weight-Decay

Der Weight-Decay-Term soll die zu starke Anpassung des Netzes an der Trainingsdaten verhindern und dadurch die Fehler auf Testdaten reduzieren. In Tabelle 5.6 ist ein Vergleich

zwischen neuronalen Netzen mit Weight-Decay, ohne Weight-Decay und dem Fehlerkorrekturmodell nach ihren quadratischen Fehler aus Training- und Testdaten dargestellt.

Periode	Neuronales Netz					Fehlerkorrekturmodell	
	mit <i>Weight-Decay</i>			ohne <i>Weight-Decay</i>			
	γ	Test	Train	Test	Train	Test	Train
Tag	0,5	1,222	0,454	1,215	0,449	1,209	0,450
Woche	1	5,883	1,877	6,277	1,712	6,108	1,847
Monat	0,5	20,730	5,945	25,829	4,977	21,742	6,223

Tabelle 5.6 Quadratische Fehlern von NN mit bzw. ohne *Weight-Decay* und ECM

Bei der Monatsprognose hat die Einführung des Weight-Decay-Terms den quadratischen Fehler auf Trainingsdaten leicht erhöht, dieser ist aber immer noch kleiner als der Fehler vom Fehlerkorrekturmodell. Dafür ist der Fehler auf Testdaten durch Verwendung von Weight-Decay-Term enorm reduziert. Das Netz mit Weight-Decay-Term erzeugte den kleinsten Fehler auf Testdaten. Somit ist es eine gelungene Regularisierung. Bei der Wochenprognose erzeugt das neuronale Netz mit Weight-Decay-Term auch den kleinsten Fehler auf Testdaten, aber seiner Fehler auf Trainingsdaten ist am größten. Trotzdem ist das Ziel der Regularisierung erreicht. Bei der Tagesprognose erzeugt das neuronale Netz mit Weight-Decay-Term die größten Fehler auf beiden Datenbereichen. Das weist auf eine Überregularisierung hin.

Eine Überregularisierung entsteht, wenn der Weight-Decay-Term zu stark gewichtet ist. In diesem Fall nimmt der tatsächliche Fehlerterm βE_D nur einen kleinen Anteil des gesamten Fehlers E ein, und dagegen spielt der Weight-Decay-Faktor αE_W in der Fehlerfunktion eine dominierende Rolle. So wird das Netz durch das Training wenig den Mustern angepasst, viel mehr wird das Weight-Decay-Term αE_W minimiert. Da die Fehlerfunktion überwiegend aus dem Weight-Decay besteht, ist sie dann auch minimiert, wenn der Weight-Decay minimal ist, obwohl der eigentliche Fehler von den Mustern noch nicht minimiert ist. Und dadurch erzeugt das Netz auf beiden Datenbereichen die größten Fehler. Um dieses Problem zu lösen muß man die geeigneten Werte für die Gewichtsparameter α und β bestimmen.

Bayes'sches Lernen

In Abschnitt 5.1.2 wurde das Bayes'sche Lernen vorgestellt, wobei die Gewichtsparameter in Gleichung (5.1) automatisch eingestellt werden können. Unter der Voraussetzung, daß die Gewichte und die Zielwerte des Netzes eine gewisse Verteilung annehmen, werden bei dem Bayes'schen Lernen die Gewichtsparameter α und β , ausgehend von einem kleinen Wert, während des Trainings iterativ optimiert. Im weiteren wird das Bayes'sche Lernen in der Zinsprognose eingesetzt, um eine anpassende Regularisierung durchzuführen. Die Testergebnisse sind in Tabelle 5.7 gezeigt.

Periode	Hid.	γ	Quad. Fehler		Treffer in %
			Test	Train	
Tag	4	0,008	1,209	0,450	64,862
Woche	5	0,02	6,055	1,825	56,203
Monat	10	0,26	20,282	5,695	69,434

Tabelle 5.7 Zinsprognose von neuronalen Netzen mit Bayes'sches Lernen

Die Werte von Parameter γ , die in Tabelle 5.7 gezeigt sind, sind die optimalen Werte von γ , die durch Bayes'sches Lernen berechnet wurden. In Tabelle 5.8 stellen wir die quadratischen Fehler auf Trainings- und Testdaten von Bayes'schem Lernen, nicht regularisiertem Netz und dem Fehlerkorrekturmodell zusammen, um die verschiedenen Methoden zu vergleichen.

Periode	NN					ECM	
	mit Bayes'sches Lernen			ohne Regularisierung		Test	Train
	γ	Test	Train	Test	Train		
Tag	0,008	1,209	0,450	1,215	0,449	1,209	0,450
Woche	0,02	6,055	1,825	6,277	1,712	6,108	1,847
Monat	0,26	20,282	5,695	25,829	4,977	21,742	6,223

Tabelle 5.8 Quadratische Fehlern von NN und ECM

Für die Tagesprognose ist der Fehler von Bayes'sches Lernen auf Testdaten kleiner und auf Trainingsdaten größer als die vom Netz ohne Regularisierung. Im Vergleich zum Fehler-

korrekturmodell hat Bayes'sches Lernen auf beiden Datenbereichen gleiche Fehler erzeugt. Dabei ist der Parameter γ sehr klein gestellt. Also der Strafterm, der die Komplexität des Netzes bestrafen soll, ist gegenüber den Fehler des Netzes wenig gewichtet. Das heißt, daß das Netz wenig Regularisierung braucht. Daher ist das Ziel der Regularisierung zwar erreicht, aber die Wirkung ist sehr gering. Für Wochen- und Monatsprognosen sind die Regularisierungen wirksamer. In beiden Fällen sind die Fehler vom Bayes'schen Lernen auf Testdaten deutlich kleiner und auf Trainingsdaten etwas größer als die Fehler von Netzen ohne Regularisierung. Im Vergleich zum Fehlerkorrekturmodell sind die Fehler vom Bayes'schen Lernen auf beiden Datenbereichen kleiner. Dies weist auf die anpassenden Regularisierungen hin.

An dieser Stelle wollen wir die Wirkung der Anwendung von Bayes'sches Lernen veranschaulichen. Als Beispiel nehmen wir das Netz für die Monatsprognose. Zuerst wird in Abbildung 5.1 gezeigt, wie sich der Gewichtsparameter γ im Lauf des Trainings verändert.

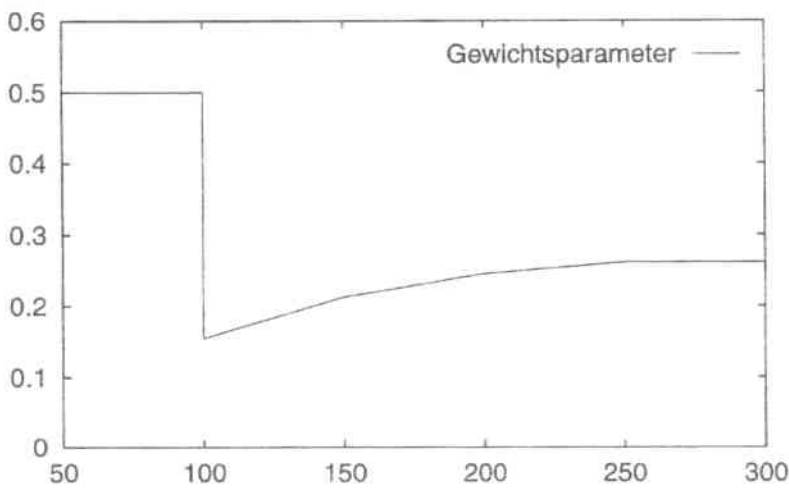


Abbildung 5.1 Veränderung vom Gewichtsparameter γ im Lauf des Trainings

Der Trainingsprozess ist so eingestellt, daß γ erst nach 100 Epochen zum ersten Mal berechnet wird und danach für jede weiteren 50 Epochen neu berechnet wird. Jeder neu berechnete Wert wird sofort in der Fehlerfunktion eingesetzt, um den neuen Fehler zu berechnen. γ nähert sich dabei schrittweise an seinen optimalen Wert. Dieser approximierte optimale Wert wird zum Schluß in die entgeltliche Fehlerfunktion eingesetzt.

Abbildung 5.2 zeigt den Verlauf des quadratischen Fehlers auf Trainings- und Testdaten nach Epochen. Abbildung 5.2 (a) ist von einem nicht regulierten Netz und (b) ist von einem Netz mit Bayes'schem Lernen.

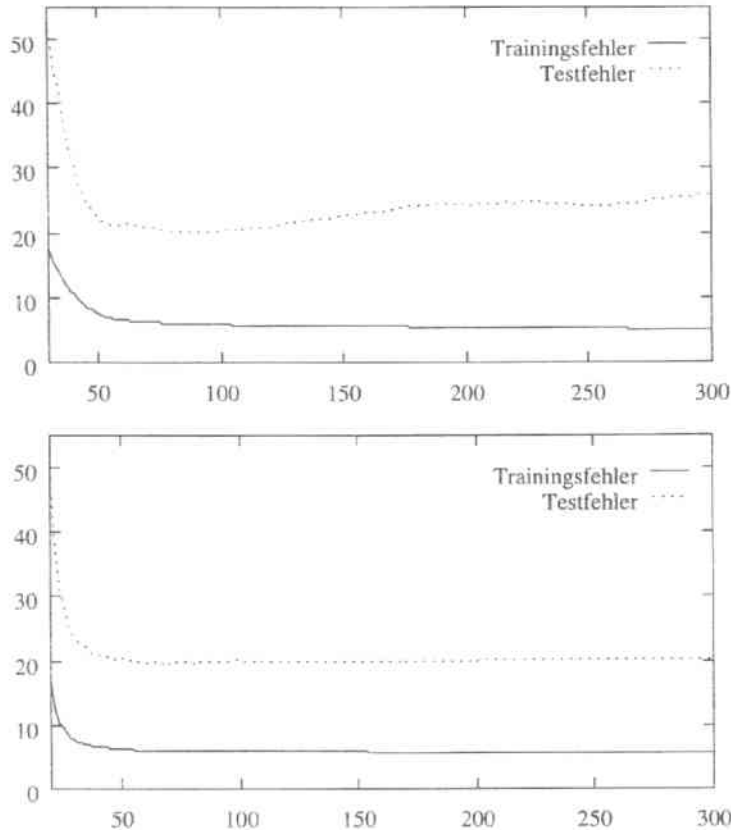


Abbildung 5.2 (a) Fehler vom Netz ohne Regularisierung, (b) Fehler vom Netz mit Bayes'schem Lernen

In Abbildung 5.2 (a) ist der Overfitting-Effekt deutlich zu sehen. Der Fehler auf Testdaten vom nicht regulierten Netz nimmt nach ungefähr 100 Epochen wieder zu, obwohl der Fehler auf Trainingsdaten noch ständig abnimmt. Dieser Effekt ist in Abbildung 5.2 (b) kaum zu sehen. Der Fehler auf Testdaten ist bei ungefähr 80 Epochen minimiert, und dies wird auch im Rest des Trainings beibehalten. Das ist dadurch erreicht, daß der zusätzliche Strafterm E_W die zu starke Anpassung des Netzes an den Trainingsdaten verhindert. Damit kann das Netz die optimale Approximation, die den ganzen Datenbereich anpasst, abbilden. Das Bayes'sche Lernen dient dazu, ein geeignetes Verhältnis zwischen E_D und E_W zu bestimmen.

5.2.3 Weitere Ein- und Ausgabemöglichkeiten

Im weiteren werden andere Ein- und Ausgabemöglichkeiten untersucht. Da unterschiedliche Kodierungen unterschiedliche Zielwerte zur Folge haben, ist es nicht mehr möglich die vom Netz erzeugten quadratischen Fehler zu verwenden, um die unterschiedlichen Kodierungen zu vergleichen. Außerdem ist die Trefferquote diejenige, für die man sich schließlich bei einer Zinsprognose interessiert, deswegen werden wir uns später nur noch an der Trefferquote orientieren.

Kodierung

In Abschnitt 5.1.1 sind vier Kodierungsmöglichkeiten der Differenzreihe angesprochen. Alle vier Kodierungen wurden getestet. Die Testergebnisse sind in Tabelle 5.9 zusammengestellt.

Kodierung	Periode	Hid.	γ	Treffer in %
bindiff	Tag	9	4,83	60,076
	Woche	9	4,45	58,987
	Monat	9	3,14	68,931
absdiff	Tag	4	0,01	64,862
	Woche	7	1	57,595
	Monat	10	0,26	69,434
reldiff	Tag	7	0,0001	63,602
	Woche	3	1	61,646
	Monat	3	0,002	68,428
logdiff	Tag	3	0,0001	63,350
	Woche	9	1	61,013
	Monat	4	0	70,063

Die besten Trefferquoten jeder Periode sind fett gedruckt.

Tabelle 5.9 Zinsprognose mit unterschiedlichen Kodierungen

Die Testergebnisse zeigen, daß die Kodierung die Prognoseleistung deutlich beeinflusst. Wenn die Kodierung sich ändert, ändert sich sowohl die optimale Topologie als auch die optimale Regularisierung der Netze. Mit anderen Worten, man findet kein optimales Netz

für das gleiche Problem mit der gleichen Eingangsinformation, wenn die Kodierung nicht optimal ist. Auch ist zu beachten, daß die optimale Kodierung für eine Prognosebasis nicht unbedingt optimal für die anderen Prognosebasen ist. In Tabelle 5.10 ist für jede Prognosebasis die beste Prognoseergebnisse zusammengefaßt.

Periode	Kodierung	Hid.	γ	Treffer in %
Tag	absdiff	4	0,01	64,862
Woche	reldiff	3	1	61,646
Monat	logdiff	4	0	70,063

Tabelle 5.10 Zinsprognose mit optimaler Kodierung

Tabelle 5.10 zeigt, daß **absdiff** für Tagesprognosen die beste Kodierung ist, während **reldiff** und **logdiff** jeweils für Wochenprognosen und Monatsprognosen die besten Wahlen sind. In der Zukunft wird für die verschiedenen Prognosebasis, nur die entsprechende optimale Kodierung weiterverwendet.

Skalierung

In folgender Untersuchung werden die Eingabevariablen zwischen -1 und 1 linear skaliert. Um die Vorzeichen der Werte zu behalten, werden praktisch den 0 fixiert und die positiven Werte zu positiven Werte bzw. die negativen Werte zu negativen Werte skaliert. Tabelle 5.11 zeigt das Testergebnis von den Prognosen mit skalierten Daten.

Periode	Kodierung	Hid.	γ	Treffer in %
Tag	absdiff	5	0,14	64,610
Woche	reldiff	7	1	60,760
Monat	logdiff	6	0,53	70,315

Tabelle 5.11 Zinsprognose mit Variablen skaliert zwischen -1 und 1

Vergleich Tabelle 5.11 mit 5.10, stellt man fest, daß durch die Skalierung die Trefferquoten für Monatsprognosen leicht erhöht aber für Tagesprognosen und Wochenprognosen ein wenig zurückgegangen sind. Also ist eine Skalierung der Eingangsvariablen nicht unbedingt

notwendig. Selbst wenn die Skalierung für eine Prognosebasis eine positive Wirkung hat, kann sie für die anderen Prognosebasen negative Wirkungen haben.

Technische Indikatoren

Bislang haben wir bis auf die Kodierungsart und die Skalierung die gleichen Eingaben für die neuronalen Netze wie für das Fehlerkorrekturmodell verwendet. In diesem Abschnitt versuchen wir die Anwendung technischer Indikatoren zu untersuchen. Zunächst testen wir Netze, dessen Eingaben nur aus Differenzreihen von D10Y und ihrer technischen Indikatoren bestehen. Abbildung 5.3 zeigt das Eingabe-Schema.

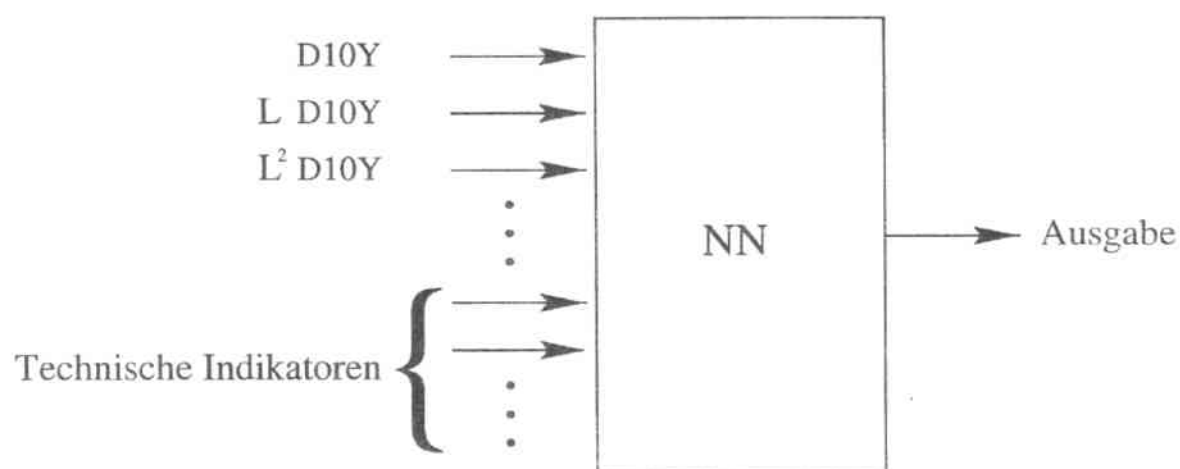


Abbildung 5.3 Neuronales Netz mit technischen Indikatoren als Eingabe

In Tabelle 5.2 sind die zur Verfügung stehenden technischen Indikatoren aufgelistet. Die ökonomischen Zeitreihen sind naturbedingt meistens stark verrauscht. Wenn eine neue Variable in den Eingangsvektor eingefügt wird, tritt mit der ergänzenden Information auch eine Menge Rauschen bei. Es ist deswegen bei der Anwendung technischer Indikatoren zu beachten, daß die gesamte Information, die die Eingabe trägt, nicht durch hinzukommende Variable von mitwachsendem Rauschen überschattet wird.

Da die Anzahl der möglichen Kombinationen und Parametrisierungen technischer Indikatoren astronomisch ist, ist die komplette Durchsuchung nicht möglich. Mit Hilfe der *Mutual-Information-Technik*³, versuchen wir aus einer großen Auswahl die optimale Zusammensetzung technischer Indikatoren zu finden.

³Genaue Diskussion über Anwendung von Mutual Information in [26]

Die Prognosen auf unterschiedlichen Basen werden getrennt behandelt, da die optimale Zusammensetzung technischer Indikatoren dabei auch unterschiedlich ist. In Tabelle 5.12 sind die für jede Basis mit Mutual Information ausgewählte Zusammensetzung technischer Indikatoren gezeigt.

Periode	TI
Tag	ma+diff
Woche	ma+vola
Monat	ma+diff+kurtosis

Tabelle 5.12 Ausgewählte technische Indikatoren

Mit den ausgewählten Zusammensetzungen technischer Indikatoren und der Zeitreihe $D10Y$ werden die Netzeingaben gebildet. Das Ergebnis des Prognosetests mit solchen Netzeingaben ist in Tabelle 5.13 dargestellt.

Periode	TI	Hid.	γ	Treffer in %
Tag	ma+diff	4	0,01	58,186
Woche	ma+vola	8	0,003	59,241
Monat	ma+diff+kurtosis	5	7,67	59,748

Tabelle 5.13 Prognose mit technischen Indikatoren

Die Trefferquote für die Tages-, Wochen bzw. Monatsprognose, mit nur technischen Indikatoren, sind jeweils um 6,424, 1,519 bzw. 10,567 Punkte niedriger, als die Trefferquote von vorher (siehe Tabelle 5.10 bzw. 5.11). Das weist daraufhin, daß die Faktoren $U10Y$, $F3M$ und der Abweichungsterm nicht durch die technischen Indikatoren von $D10Y$ zu ersetzen sind. Sie enthalten Informationen, die nicht durch technische Transformationen von $D10Y$ erworben werden können.

Als nächstes testen wir Netze, die technische Indikatoren als zusätzliche Eingabe neben $D10Y$, $U10Y$, $F3M$ und dem Abweichungsterm haben. In Abbildung 5.4 ist die Bildung

des Eingabevektors veranschaulicht.

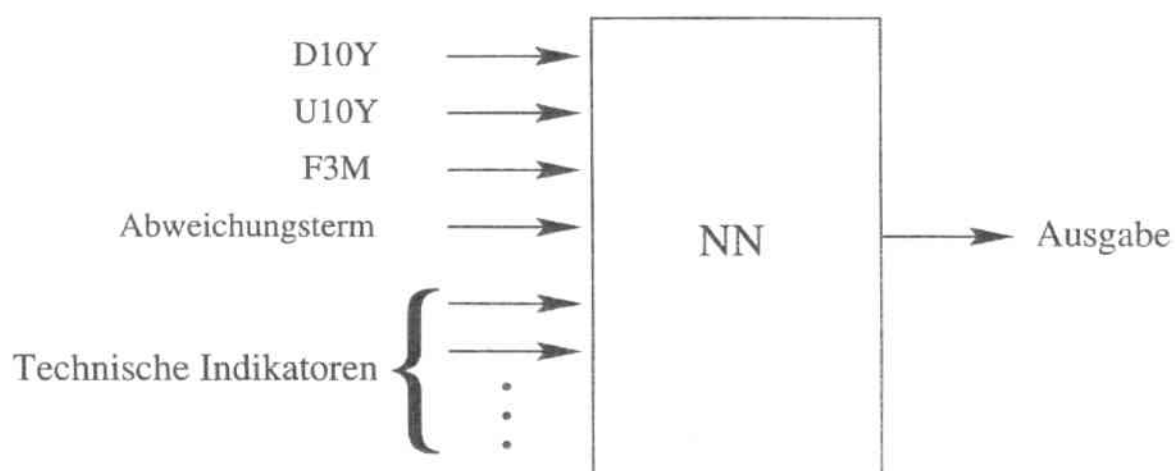


Abbildung 5.4 Neuronales Netz mit technischen Indikatoren als zusätzliche Eingabe

In Tabelle 5.14 sind die mit Mutual-Information ausgewählte technische Indikatoren dargestellt.

Periode	TI
Tag	rsi
	diff
	rsi+diff
Woche	ma
	kurtosis
	ma+kurtosis
Monat	ma
	kurtosis
	ma+kurtosis

Tabelle 5.14 Ausgewählte technische Indikatoren

Die ausgewählte Kombination von der technischen Indikatoren werden dann in die Eingabevektoren eingefügt. Tabelle 5.15 zeigt das Ergebnis des Prognosetests mit den neuen

Eingabevektoren.

Periode	TI	Hid.	γ	Treffer in %
Tag	rsi	4	0,01	64,987
	diff	4	0,01	64,232
	rsi+diff	7	0,004	64,862
Woche	ma	3	1	60,633
	kurtosis	3	1	60,253
	ma+kurtosis	6	1	60,506
Monat	kurtosis	5	0,32	67,547
	ma	4	0,42	66,793
	ma+kurtosis	7	0,36	69,182

Tabelle 5.15 Prognose mit technische Indikatoren

Die Anwendung technischer Indikatoren als zusätzliche Eingangsgröße, hat hier ebenfalls wenig Verbesserung gebracht. Bei den Tagesprognosen und den Wochenprognosen haben sich die Trefferquoten zwar verbessert, aber die Verbesserung ist sehr gering. Bei den Monatsprognosen fällt die Trefferquote zurück, wenn eine weitere Variable in den Eingangsvektor eingefügt wird. Das heißt, daß die hinzukommenden Variablen in diesem Fall überwiegend Rauschen mitbringen.

Am Ende dieses Kapitls fassen wir die bisherigen besten Netze mit ihren Prognoseergebnissen in Tabelle 5.16 zusammen.

Periode	Kodierung	Skalierung	TI	Hid.	γ	Treffer in %
Tag	absdiff	-	rsi	4	0,01	64,987
Woche	reldiff	-	-	3	1	61,646
Monat	logdiff	$[-1, 1]$	-	6	0,53	70,315

Tabelle 5.16 Einstellung der neuronalen Netze mit den bisher besten Trefferquoten

Kapitel 6

Zusammenarbeit von NN und ECM

In den letzten zwei Kapiteln haben wir das neuronale Netz und das Fehlerkorrekturmodell zur Prognose der Zinsentwicklung getestet. Die Ergebnisse zeigen, daß beide Methoden für alle drei Perioden akzeptierbare Prognoseleistungen liefern können. Da die beiden Modelle unterschiedlichen Mechanismen folgen, kommt man auf die Idee, daß es eventuell zu einer Verbesserung führt, wenn man die zwei Methoden zusammenarbeiten läßt. In diesem Kapitel wird der kombinierte Einsatz der beiden Methoden untersucht.

6.1 Entscheidungskomitee

6.1.1 Bildung des Entscheidungskomitees

Eine einfache Möglichkeit der Zusammenarbeit vom neuronalen Netz und dem Fehlerkorrekturmodell ist, daß man die beiden Methoden parallel laufen läßt und ihre Prognosen miteinander vergleicht. Man bildet dadurch ein Entscheidungskomitee. Das Entscheidungskomitee macht die entgeltliche Prognoseentscheidung, und dies wird nur dann für einen bestimmten Tag gemacht, wenn beide Methoden für diesen Tag die gleiche Richtung der Zinsentwicklung vorhersagen. Die Prognoseentscheidung des Entscheidungskomitees entspricht der Vorhersage der beiden Methoden. Wenn beide Methoden für einen Tag jeweils eine unterschiedliche Richtung der Zinsentwicklung vorhersagen, dann wird das Entscheidungskomitee für diesen Tag keine Prognoseentscheidung machen. Die Komiteebildung ist in Abbildung 6.1 dargestellt.

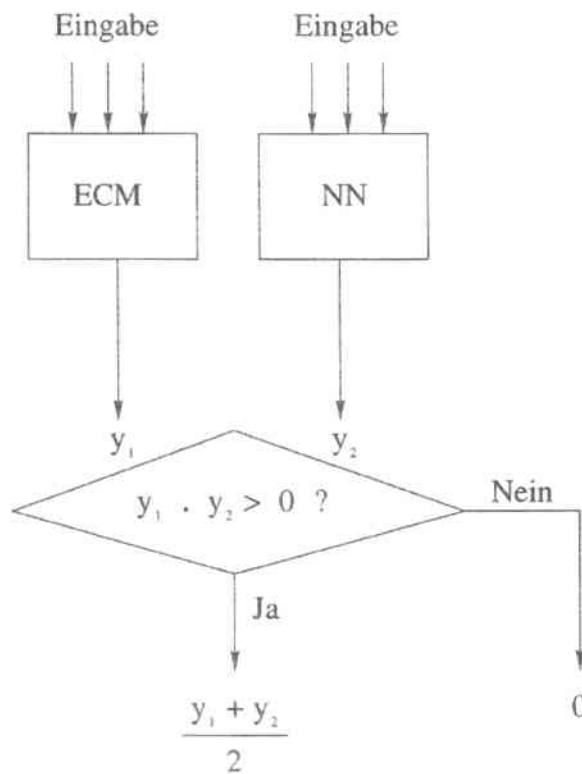


Abbildung 6.1 Entscheidungskomitee

Die Idee der Komiteebildung besteht darin, daß man durch die gleichzeitige Betrachtung beider Methoden versucht, eine sichere Entscheidung zu treffen. Das Fehlerkorrekturmodell und das neuronale Netz treffen ihre Entscheidungen nach ihren eigenen Mechanismen. So kann man sich vorstellen, wenn beide Methoden für einen Tag den gleichen Trend der Zinsentwicklung vorhersagen, dann liegt die Entscheidung für diesen Trend möglicherweise näher an der Wahrheit, als wenn beide Methoden für diesen Tag ganz verkehrte Trends vorhersagen. In den zweifelhaften Fälle, d.h., wenn die beiden Methoden die verkehrte Trends prognostizieren, schweigt das Entscheidungskomitee, also macht er dann keine Prognoseentscheidung. Das reduziert zwar die Anzahl der gemachten Prognosen, aber man hofft, daß man dadurch eine höhere Trefferquote erzielen kann.

6.1.2 Empirische Untersuchungen

Zunächst untersuchen wir, wie stark sich die Prognosen der beiden Methoden unterscheiden. Für das Fehlerkorrekturmodell übernehmen wir die Prognoseergebnisse in Tabelle 4.11. Für die neuronalen Netze übernehmen wir die Prognoseergebnisse von den Netzen.

dessen Gestaltung in Tabelle 5.16 beschrieben sind, da sie die beste Trefferquote für die jeweiligen Prognosebasis haben. In Tabelle 6.1 sind die Korrelationen zwischen den Prognosen der beiden verschiedenen Methoden dargestellt.

Periode	Korrelation
Tag	0,94
Woche	0,74
Monat	0,94

Tabelle 6.1 Korrelationen zwischen den Prognosen von ECM und NN

Die Korrelationskoeffizienten in Tabelle 6.1 zeigen, daß für die Tages- bzw. Monatbasis die Prognosen von neuronalen Netze und dem Fehlerkorrekturmodell hoch korreliert sind, d.h. die beiden Methoden haben sehr oft die gleichen Trends vorhergesagt. Für die Wochenbasis sind die Prognosen beider Methoden nicht sehr hoch korreliert, also unterscheiden sich die Prognosen der beiden Methoden in diesem Fall etwas mehr.

Wir bilden für jede Prognosebasis ein Entscheidungskomitee, und führen die Prognosetests durch. Die Prognoseergebnisse sind in Tabelle 6.2 zusammengestellt.

Periode	Treffer in %
Tag	65,545
Woche	59,835
Monat	70,516

Tabelle 6.2 Zinsprognose mit dem Entscheidungskomitee

Für die Tages- bzw. Monatbasis haben die Entscheidungskomitees die Trends der Zinsentwicklungen um 0,558% bzw. 0,201% (vgl. Tabelle 5.16) mehr als die neuronalen Netze und um 0,935% bzw. 2,717% (vgl. Tabelle 4.11) mehr als das Fehlerkorrekturmodell richtig prognostiziert. In diesen beiden Fällen haben die Entscheidungskomitees, aufgrund der uneinheitlichen Vorhersage der beiden Methoden, durch das Schweigen überwiegende Fehlentscheidung vermieden. Für die Wochenbasis hat das Entscheidungskomitee die Trends

der Zinsentwicklungen um 1,811% weniger als das neuronale Netz und um 5,657% mehr als das Fehlerkorrekturmodell richtig prognostiziert. Da das Fehlerkorrekturmodell für die Wochenbasis viel schlechter als das neuronale Netz funktioniert hat, hat er in vielen Fällen durch seine falschen Prognosen das Entscheidungskomitee dazu gebracht, keine Prognoseentscheidung zu machen, obwohl das neuronale Netz dort richtig prognostiziert hat. Damit stellt man fest, daß man bei der Bildung des Entscheidungskomitees besonders vorsichtig sein muß, wenn die einzelne Methode sehr unterschiedliche Prognoseleistungen haben. Denn durch die Komiteebildung wird die Prognoseleistung des gesamten Entscheidungskomitees in diesem Fall nicht unbedingt erhöht, sondern die Leistung der starken Methode könnte durch die Leistung der schwachen Methode ausgeglichen werden. Bei denjenigen Methoden, die ähnliche Prognoseleistungen haben, kann man durch die Bildung des Entscheidungskomitees eine durchaus höhere Prognosetrefferquote erzielen.

6.2 Datenvorverarbeitung mit dem Fehlerkorrekturmodell

Das neuronale Netz hat gegenüber den herkömmlichen ökonomischen Verfahren größere Flexibilität. Es kann aus Informationen, die die vielen anderen Methoden schwierig verarbeiten können, lernen. Diese Fähigkeit ermöglicht dem neuronalen Netz, die Ausgabe anderer Methoden zu übernehmen, und dadurch die Arbeit der anderen Methoden auszunutzen. Jedes Prognoseverfahren ist im Prinzip ein Mechanismus zur Informationsverarbeitung. Sie analysieren die zur Verfügung stehenden Information auf ihrer eigen Art und Weise und treffen dann ihre Entscheidungen. Wenn das neuronale Netz die Entscheidung eines anderen Verfahrens als Eingabe nimmt, betrachtet das Netz während seiner Arbeit die Meinung dieses Verfahrens mit.

6.2.1 Zweistufiges Prognosemodell

Das Fehlerkorrekturmodell wird nun als Vorverarbeitungsmechanismus eingesetzt. Zuerst wird die Prognose mit Fehlerkorrekturmodell gemacht, danach wird das Prognoseergebnis wieder als Eingangsvariable ins neuronale Netz geführt. Dadurch entsteht ein zweistufiges Prognosemodell. Die Netzeingabe entsteht dann aus der Ausgabe des Fehlerkorrekturmodells und den transformierten Variablen. Unter den transformierten Variablen versteht man die unterschiedlich kodierten bzw. skalierten Zeitreihen sowie die technischen Indikatoren. Die Struktur dieses zweistufige Prognosemodells ist in Abbildung 6.2 gezeigt.

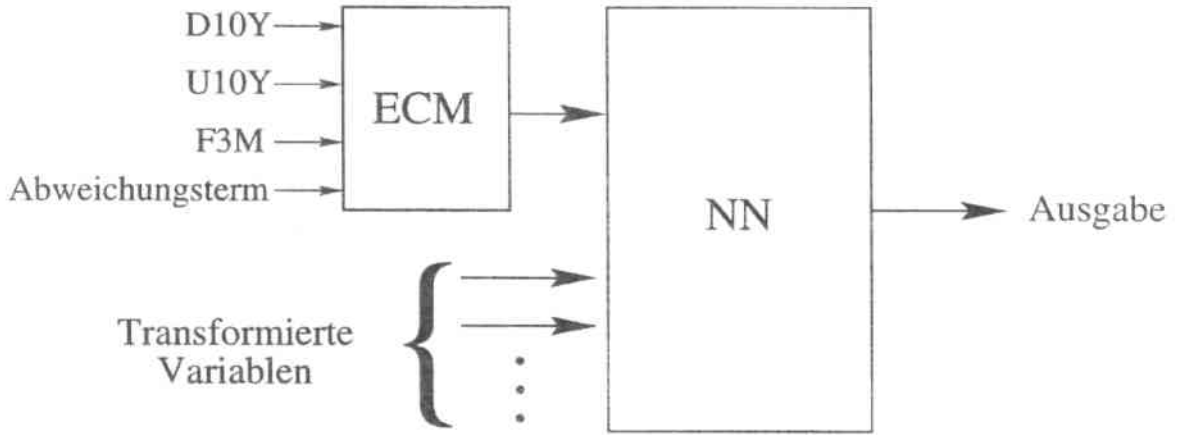


Abbildung 6.2 Zweistufiges Prognosemodells

Durch den Einsatz des Fehlerkorrekturmodells wird zwar keine zusätzliche Information für das neuronale Netz erworben, aber man hat dabei die Hoffnung, daß durch die Verarbeitung des Fehlerkorrekturmodells die bestehende Information besserre präsentiert ist und damit die Arbeit des neuronalen Netzes erleichtert wird.

6.2.2 Empirische Untersuchungen

Die Prognosetests werden in zwei Schritten durchgeführt. Im ersten Schritt werden die Zinsentwicklungen auf verschiedenen Basen mit dem Fehlerkorrekturmodell prognostiziert. Wir übernehmen hier die Prognoseergebnisse des Fehlerkorrekturmodells in Tabelle 4.11. Im zweiten Schritt werden mehrere Eingabevektoren für die neuronalen Netze gebildet. Jeder Eingabevektor besteht aus der Ausgabe des Fehlerkorrekturmodells und der unterschiedlichen Zusammensetzung der transformierten Zeitreihen. Mit jedem Eingabevektor wurden mehrere Prognosemodelle durch verschiedene Parameterisierungen aufgebaut. Alle Prognosemodelle wurden getestet. In Tabelle 6.3 werden für jede Basis die Eingabevektoren der besten drei Prognosemodelle gezeigt.

Periode	Eing.-Nr.	Eingabevektor
Tag	1	<i>absdiff</i> von D10Y, U10Y und F3M mit Lags 1 bis 3 Abweichungsterm mit Lags 1 bis 3 Ausgabe von ECM mit Lags 1
	2	<i>absdiff</i> von D10Y, U10Y und F3M mit Lags 1 bis 3 Abweichungsterm mit Lags 1 Ausgabe von ECM mit Lags 1 bis 3
	3	<i>absdiff</i> von D10Y, U10Y und F3M mit Lags 1 bis 3 Abweichungsterm mit Lags 1 bis 3 Ausgabe von ECM mit Lags 1 bis 3
Woche	1	<i>reldiff</i> von D10Y, U10Y und F3M mit Lags 1 bis 3 Abweichungsterm mit Lags 1 bis 3 <i>ma</i> von <i>reldiff</i> von D10Y Ausgabe von ECM mit Lags 1 bis 3
	2	<i>reldiff</i> von D10Y mit Lags 2 und 4 <i>reldiff</i> von U10Y Lags 1 <i>reldiff</i> von F3M mit Lags 5 Abweichungsterm mit Lags 1 <i>ma</i> von <i>reldiff</i> von D10Y Ausgabe von ECM mit Lags 1
	3	<i>reldiff</i> von D10Y, U10Y und F3M mit Lags 1 bis 3 Abweichungsterm mit Lags 1 Ausgabe von ECM mit Lags 1 bis 3
Monat	1	<i>scale(logdiff)</i> von D10Y, U10Y, F3M mit Lags 1 bis 3 skalierter Abweichungsterm mit Lags 1 skalierte Ausgabe von ECM mit Lags 1
	2	<i>scale(logdiff)</i> von D10Y, U10Y, F3M mit Lags 1 bis 3 skalierter Abweichungsterm mit Lags 1 skalierte <i>kurtosis</i> von <i>logdiff</i> von D10Y skalierte Ausgabe von ECM mit Lags 1
	3	<i>scale(logdiff)</i> von D10Y, U10Y, F3M mit Lags 1 bis 3 skalierter Abweichungsterm mit Lags 1 bis 3 skalierte Ausgabe von ECM mit Lags 1

Tabelle 6.3 Ausgewählte Netzeingabe für die zweistufigen Modelle

In Tabelle 6.4 sind dann die entsprechenden Prognoseergebnisse dargestellt.

Periode	Eing.-Nr.	Hid.	γ	Treffer in %
Tag	1	4	0,01	65,365
	2	8	0,01	65,365
	3	5	0,004	65,239
Woche	1	6	1	62,152
	2	7	0,5	60,127
	3	5	1	59,494
Monat	1	6	0,42	73,459
	2	10	0,40	71,950
	3	8	0,21	69,686

Tabelle 6.4 Zinsprognose mit den zweistufigen Modellen

Die Testergebnisse zeigen, daß allein durch die Einführung von Ausgabe des Fehlerkorrekturmodells, noch keine gute Netzeingabe garantiert werden kann. Auch hier spielt die Zusammensetzung anderer Variablen eine bedeutende Rolle. Für die unterschiedliche Kombination der Variablen variieren die Trefferquoten über mehrere Prozente hinaus. In Tabelle 6.5 sind die besten Prognoseergebnisse für die drei Prognosebasen zusammengestellt.

Periode	Hid.	γ	Treffer in %
Tag	4	0,01	65,365
Woche	6	1	62,152
Monat	6	0,42	73,459

Tabelle 6.5 Die besten Prognoseergebnisse der zweistufigen Modellen

Die Trefferquoten der zweistufigen Modellen haben sich gegenüber den einzelnen Methoden für alle Perioden erhöht. Bei der Monatsprognosen war die Verbesserung der Trefferquote besonders groß. Die beste Trefferquote der neuronalen Netze war 70,315% (siehe Tabelle 5.16). Sie ist jetzt um 3,144% auf 73,459% gestiegen. Bei der Tages- und Wochenprognose sind die Verbesserungen der Trefferquoten jeweils 0,378% und 0,343% (vgl.

Tabelle 5.16). Somit ist das zweistufige Modell mit einer geeigneten Bildung der Eingabe durchaus konkurrenzfähig.

Zum Schluß fassen wir die besten Prognoseergebnisse von der Zusammenarbeit der neuronalen Netzen mit dem Fehlerkorrekturmodell in Tabelle 6.6 zusammen.

Periode	Methode	Treffer in %
Tag	Entscheidungskomitee	65,545
Woche	Zweistufiges Modell	62,152
Monat	Zweistufiges Modell	73,459

Tabelle 6.6 Die besten Prognoseergebnisse von der Zusammenarbeit

Kapitel 7

Bewertung der Prognoseverfahren

In diesem Kapitel werden wir über die untersuchten Prognoseverfahren eine komplette Bewertung machen. Dabei möchten wir zunächst feststellen, ob die Einführung des Abweichungsterms ins linearen Modell eine positive Wirkung in der Zinsprognose hat, mit anderen Worten, ob sich das "Granger-Repräsentations-Theorem" empirisch nachweisen lässt. Weiter möchten wir die Konkurrenzfähigkeit des neuronalen Netzes gegenüber dem Fehlerkorrekturmodell zur Zinsprognose feststellen. Als letztes möchten wir die Zusammenarbeit der beiden Verfahren bewerten.

7.1 Der Effekt der Abweichungsterm

An dieser Stelle müssen wir an die Ursache der Anwendung, von der Abweichung der langfristige Beziehung erinnern. Untersuchungen in Kapitel 4 zeigen, daß angesichts der kurzfristige Zinsentwicklung Scheinkorrelationen zwischen D10Y und U10Y bzw. D10Y und F3M auftreten. Das heißt, während starke Korrelationen zwischen D10Y und U10Y bzw. D10Y und F3M existieren, sind ihre kurzfristigen Veränderungen wesentlich weniger korreliert. Und damit stellen die hohe Korrelationen der ursprüngliche Zeitreihen für die kurzfristige Prognose von D10Y eine Verwirrung dar. Um das Problem zu lösen, verwendet man nur die Differenzreihe im Prognosemodell. Dabei gehen aber die Informationen hinsichtlich potentieller langfristiger Beziehungen, die sich in den Niveauvariablen widerspiegeln, verloren. Durch die Einführung der Abweichung der langfristige Beziehung zwischen D10Y, U10Y und F3M hofft man, daß dieser Informationsverlust dann ausgeglichen werden kann.

In Kapitel 4 werden mit dem linearen Modell, ohne Anwendung des Abweichungsterms

und mit dem Fehlerkorrekturmodell, wobei der Abweichungsterm eingesetzt wird, Zinsprognosen durchgeführt. Die erzielte Ergebnisse sind in Tabelle 7.1 nochmal präsentiert (vgl. Tabelle 4.11).

Periode	ECM (mit err_t)		Modelle ohne err_t	
	Quad. Fehler	Treffer in %	Quad. Fehler	Treffer in %
Tag	1,209	64,610	1,214	62,972
Woche	6,108	54,177	6,371	51,266
Monat	21,742	67,799	32,191	55,220

Tabelle 7.1 Zinsprognose mit linearen ökonomischen Modellen

Um den Vergleich der verschiedenen Modelle zu erleichtern, stellen wir die Differenz der quadratischen Fehler und der Trefferquoten in Tabelle 7.2 dar. In Tabelle 7.2 bezeichnen QF_F und TQ_F bzw. QF_L und TQ_L jeweils den quadratischen Fehler und die Trefferquote vom Fehlerkorrekturmodell bzw. vom linearen Modell ohne Abweichungsterm. In der Spalte, die mit **Prozentuell** bezeichnet ist, stehen die prozentuellen Differenzen, also je nachdem $\frac{QF_F - QF_L}{QF_L}$ oder $\frac{TQ_F - TQ_L}{TQ_L}$ in Prozent.

Periode	$QF_F - QF_L$		$TQ_F - TQ_L$	
	Absolut	Prozentuell	Absolut	Prozentuell
Tag	-0,005	-0,41%	1,638	2,60%
Woche	-0,263	-4,13%	2,911	5,68%
Monat	-10,449	-32,61%	12,579	22,78%

Tabelle 7.2 Differenz der Zinsprognose mit verschiedenen linearen Modellen

Die Ergebnisse in Tabelle 7.2 zeigen, daß die Anwendung des Abweichungsterms für die Tagesprognosen eine geringe Auswirkung hat. Dabei wurde der quadratische Fehler lediglich nur um 0,41% reduziert. Die Trefferquote erhöhtet sich um 1,638 Punkte, das sind gerade 2,60%. Für die Wochen- und Monatsprognosen hat sich dabei deutlich mehr getan. Die quadratische Fehler haben sich jeweils um 4,13% bzw. 32,61% reduziert, und die

Trefferquoten sind um 2,911 bzw. 12,579 Punkten gestiegen. Die Verbesserungen der Trefferquoten sind jeweils 5,68% und 22,78%. Dieses Ergebnis kann man als eine Bestätigung der ökonomischen Hypothese annehmen, daß die in dem Abweichungsterm enthaltene Information die langfristige Beziehung widerspiegelt. Je länger die Prognosebasis ist, desto einflussreicher ist diese Information.

7.2 Ein Vergleich konkurrierender Verfahren

Anhand der Trefferquote der Prognosetests können wir das Fehlerkorrekturmodell mit den neuronalen Netze vergleichen. In Tabelle 7.5 sind die Prognoseergebnisse vom Fehlerkorrekturmodell und der neuronalen Netze zusammengefaßt. In Tabelle 7.5 sind auch die Differenzen der Trefferquoten gezeigt, wobei TQ_N und TQ_E jeweils die Trefferquote der neuronalen Netze bzw. des Fehlerkorrekturmodells bezeichnen.

Periode	NN	ECM	$TQ_N - TQ_E$	
			Absolut	Prozentuell
Tag	64,987	64,610	0,377	0,58%
Woche	61,646	54,178	7,468	13,78%
Monat	70,315	67,799	2,516	4,68%

Tabelle 7.5 Ein Vergleich des Fehlerkorrekturmodells mit den neuronalen Netzen

Bei der Tagesprognose trifft das neuronale Netz den Zinstrend 0,377% mehr als das Fehlerkorrekturmodell. Die Trefferquote ist dabei um 0,58% erhöht. Diese Prognoseverbesserung ist aber nicht sehr deutlich. Bei der Wochenprognose hat das neuronale Netz die größte Trefferquoteerhöhung. Diese beträgt 7,468 Punkten bzw. 13,78%. Damit ist das neuronale Netz dem Fehlerkorrekturmodell bezüglich der Wochenprognose einen großen Schritt voraus. Für die Monatsprognosen hat das neuronale Netz die Trefferquote um 2,516 Punkten bzw. 4,68% erhöht. Dies gilt als eine deutliche Verbesserung. Insgesamt ist das neuronale Netz gegen das Fehlerkorrekturmodell doch konkurrenzfähig.

Zum Schluß machen wir einen Vergleich zwischen dem Fehlerkorrekturmodell und der Zusammenarbeit neuronaler Netze mit dem Fehlerkorrekturmodell. In Tabelle 7.6 befinden sich die jeweiligen Trefferquoten und die Differenz der Trefferquoten.

Periode	NN+ECM	ECM	$TQ_{N+E} - TQ_E$	
			Absolut	Prozentuell
Tag	65,545	64,610	0,935	1,45%
Woche	62,152	54,178	7,974	14,72%
Monat	73,459	67,799	5,660	8,35%

Tabelle 7.6 Ein Vergleich zwischen dem ECM und der Zusammenarbeit der NN mit dem ECM

Bei der Tagesprognose hat sich die Trefferquote durch die Zusammenarbeit neuronaler Netz mit dem Fehlerkorrekturmodell um 0,935 Punkten bzw. 1,45% erhöht. Bei der Wochenprognose ist die Trefferquote durch die Zusammenarbeit um 7,974 Punkten bzw. 14,72% angestiegen. Bei der Monatsprognose liegt die Trefferquoteerhöhung bei 5,660 Punkten bzw. 8,35%. Insgesamt ist die Zusammenarbeit dem Fehlerkorrekturmodell definitiv überlegen.

Kapitel 8

Zusammenfassung und Ausblick

8.1 Zusammenfassung

Ziel dieser Arbeit war:

1. Die Theorie von Granger empirisch zu untersuchen
2. Die Fähigkeit, neuronaler Netze gegenüber dem Fehlerkorrekturmodell zur Zinsprognose, zu untersuchen
3. Die Zusammenarbeit neuronaler Netze mit dem Fehlerkorrekturmodell zu testen

Im zweiten und dritten Kapitel wurden die Ansätze des Fehlerkorrekturmodells und der neuronalen Netze theoretisch diskutiert. In den nachfolgenden vier Kapiteln wurden reichliche empirische Untersuchungen durchgeführt. Im siebten Kapitel wurden die Testergebnisse zusammengefaßt und bewertet.

Der Vergleich der Testergebnisse in Kapitel 7 zeigt, daß die Einführung der Abweichung langfristiger Gleichgewichtsbeziehung Trefferquotenerhöhungen zur Zinsprognose auf allen drei Basen ermöglicht hat. Das bedeutet, daß für die Beschreibung der Zeitreihenentwicklung die langfristige Beziehung der kointegrierten Zeitreihen durchaus eine Korrektur eines Modells, das um die Scheinkorrelation zu vermeiden nur auf differenzierten Reihen basiert ist, darstellt. Und je länger die Basisperiode ist, desto stärker ist die Korrektur. Die Theorie von Granger ist damit empirisch nachgewiesen worden.

Mit den neuronalen Netze haben wir auf alle drei Basen bessere Prognoseergebnissen

gegenüber dem Fehlerkorrekturmodell erzielt. Bezüglich der Tagesbasis ist die Trefferquotenerhöhung nicht sehr groß. Aber auf dem Wochen- und Monatsbasis hat das neuronale Netz überzeugende Prognoseverbesserungen geliefert. Damit ist festzustellen, daß die neuronalen Netze gegenüber dem Fehlerkorrekturmodell auf einer längerer Basisperiode durchaus konkurrenzfähig sind.

Die Zusammenarbeit neuronaler Netze mit dem Fehlerkorrekturmodell hat gegenüber dem Fehlerkorrekturmodell eine noch größere Trefferquotenerhöhung geliefert. Damit ist der kombinierte Einsatz eines neuronalen Netzes mit einem Fehlerkorrekturmodell der Anwendung jeder einzelnen Methode überlegen. Für eine weitere Entwicklung des Prognosesystems ist dies besonders zu empfehlen.

8.2 Ausblick

Es ergeben sich nun weitere Schwerpunkten, die ein Ziel weiterer Forschung im Bereich der Zeitreihenprognose mit neuronalen Netzen darstellen.

Zum einen ist es eine erweiterte Datenvoranalyse. Da die neuronalen Netze die Fähigkeit besitzen, komplizierte Funktionen zu approximieren, ist es sinnvoll, weitere möglicherweise zusammenwirkende ökonomische Faktoren in der Datenvoranalyse einzubeziehen. Es besteht die Möglichkeit, daß die neuronalen Netze aus solchen Variablen nützliche Informationen gewinnen.

Zum anderen liegt es in einer weiteren Untersuchung des kombinierten Einsatzes neuronaler Netze mit dem Fehlerkorrekturmodell. In unserer Untersuchung wurde der Fehlerkorrekturmodell als Vorverarbeitungsmechanismus für die neuronale Netze verwendet. Man kann aber auch umgekehrt die Ausgabe der neuronalen Netze in einem Fehlerkorrekturmodell betrachten. Bei der Bildung des Entscheidungskomitees kann man auch andere Kriterien untersuchen, z.B. die Prognoseergebnisse unterschiedlicher Methoden nach der Prognoseleistung der Methode unterschiedlich gewichtet zu betrachten. Dadurch schenkt man den besseren Methoden mehr Vertrauen, wohingegen die schwächere Methoden mehr oder weniger vernachlässigt werden.

Literaturverzeichnis

- [1] Nelson, C.R. & Plosser, C.: *Trends and Random Walk in Macroeconomic Time Series*, Journal of Monetary Economics, 10, 139-162, 1982
- [2] Box, G.E.P. & Jenkins, G.M.: *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970
- [3] Engle, R.F. & Granger, C.W.J.: *Co-Integration and Error Correction: Representation, Estimation, and Testing*, Econometrica: Vol. 55, Nr.2, S. 251-276, 1987
- [4] Kim, Jeong-Ryeol: *Analyse kointegrierter Modelle*, HAAG & HERCHEN Verlag, Frankfurt, 1994
- [5] Box, G.E.P. & Jenkins, G.M.: *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976
- [6] Sauer, Egbert: *Aktienindexprognosen mit Fehlerkorrekturmodellen und dem ökonomisch relevanten Zins*, Peter Lang, Frankfurt, 1996
- [7] Gerhards, Tilmann: *Theorie und Empirie flexibler Wechselkurse*, Physica, Heidelberg, 1994
- [8] Granger, C.W.J. & Newbold, P.: *Spurious Regressions in Econometrics*, Journal of Econometrics: 2, 111-120, 1974
- [9] Sims, C.A.: *New Methods in Business Cycle Research*, Federal Reserve Band of Minneapolis, 1977
- [10] Granger, C.W.J. & Weiss, A.A.: *Time Series Analysis of Error-Correction Models*, Academic Press: 255-278, New York, 1983
- [11] Bishop, M.: *Neural Networks for Pattern Recognition*, Oxford University Press, 1996
- [12] MadKay, D.J.C.: *Bayesian Interpolation*, Neural Computation, 4(3):415-417, 1992

- [13] MadKay, D.J.C: *A practical Bayesian framework for backpropagation networks*, Neural Computation, 4(3):448-472, 1992
- [14] Baun, S.: *Neuronalen Netze in der Aktienkursprognose*, in: Rehkugler H. und Zimmermann, H.G., *Neuronale Netze in der Ökonomie*, Franz Vahlen, München, 1994
- [15] Bartlmae, K.: *Multi-Task Lernen bei neuronalen Netzen zur Aktienkursprognose*, Diplomarbeit bei der Fakultät für Informatik, Universität Karlsruhe, Karlsruhe, 1997
- [16] Zimmermann, H.G.: *Neuronalen Netze als Entscheidungskalkül*, in: Rehkugler H. und Zimmermann, H.G., *Neuronale Netze in der Ökonomie*, Franz Vahlen, München, 1994
- [17] Poddig, T.: *Mittelfristige Zinsprognose mittels KNN und ökonometrischer Verfahren*, in: Rehkugler H. und Zimmermann, H.G., *Neuronale Netze in der Ökonomie*, Franz Vahlen, München, 1994
- [18] Riedmiller, M. & Braun, H.: *A direct adaptive method for faster backpropagation learning*, Proceedings of the IEEE International Conference on neural networks, 1993
- [19] Menzel W.: *Theorie der Neuronalen Netze*, Skriptum zur Vorlesung "Theorie der Neuronalen Netze", Karlsruhe, 1996
- [20] Dickey: *Estimation and Hypothesis Testing in Nonstationary Time Series*, Ph.D. dissertation, Iowa State University, 1976
- [21] Steurer, E.: *Wechselkursprognose: Fehlerkorrekturmodelle im Vergleich mit Neuronalen Netzen*, in: Bol, G., Schwenker, F., *Finanzmarktanalyse und -prognose mit innovativen quantitativen Verfahren*, Physika, Heidelberg, 1996
- [22] Gutjahr S.: *Improving the Determination of Hyperparameters in Bayesian Learning*, Proceedings of the Ninth Australian Conference on Neural Networks, Brisbane, 1998
- [23] Heil J.: *Einführung in die Ökonometrie*, Oldenbourg Verlag GmbH, München, 1989
- [24] Frohn J.: *Grundausbildung in Ökonometrie*, Walter de Gruyter, Berlin, 1995
- [25] Menzel W.: *Theorie der Neuronalen Netze*, Skriptum zur Vorlesung "Theorie der Neuronalen Netze", Karlsruhe, 1997
- [26] Sa, H.: *Eingabeoptimierung für neuronale Netze mit Mutual Information*, Studienarbeit bei der Fakultät für Informatik, Universität Karlsruhe, Karlsruhe, 1996