

---

# Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren

## 2

Michael Schröder

---

### 2.1 Allgemeine Überlegungen

#### 2.1.1 Anforderungen an Verfahren für kurzfristige Prognoserechnungssysteme

**Genauigkeit der Vorhersage.** Bei der Forderung nach einer möglichst hohen Genauigkeit der Vorhersage ist zu berücksichtigen, dass zusätzliche Genauigkeit in der Regel mit erhöhtem Aufwand verbunden ist. Man hat daher stets diesen erhöhten Aufwand mit den Erträgen aus der besseren Prognose zu vergleichen. Letztere können in Kostensenkungen bestehen. (Beispiel: Bei genauerer Prognose im Rahmen eines Lagerdispositionssystems können die Sicherheitsbestände niedriger gehalten werden.) Der Ertrag kann jedoch auch eine Risikominderung sein, etwa weil durch eine bessere Vorhersage im Rahmen eines Planungsmodells die Gefahr einer Fehlentscheidung geringer ist.

**Reagibilität und Stabilität der Verfahren.** Eine vollkommene Übereinstimmung zwischen Prognosewert und später tatsächlich beobachtetem Wert kann nur zufälliger Natur sein, da die hier behandelten Vorhersageverfahren lediglich Schätzwerte für die zu prognostizierende Größe liefern. In der Regel wird also der Beobachtungswert vom Prognosewert abweichen. Es stellt sich jeweils die Frage, ob diese Abweichung zufälliger Natur oder ob sie das erste Anzeichen dafür ist, dass sich die Zeitreihe in ihrem Verhalten grundlegend geändert hat (ob z. B. ein bisher steigender Trend in einen fallenden Trend übergeht). Im ersten Fall soll die Methode in der Lage sein, die Abweichung dadurch „herauszuglätten“, dass sie durch die Verwendung von vielen Vergangenheitsdaten bei der Durchschnittsbildung den Einfluss von zufälligen Abweichungen klein hält. Im zweiten Fall, wenn sich also

tatsächlich eine neue Entwicklung der Zeitreihe anbahnt, sollte das Verfahren die älteren Daten möglichst gering gewichten und die Prognose aus den jüngsten Daten ableiten, damit die Anpassung an die neue Entwicklung möglichst rasch erfolgt.

Es gilt also, einen befriedigenden Kompromiss zwischen Stabilität bei Zufallsabweichungen und Reagibilität auf Änderungen im Verhalten der Zeitreihe zu finden.

**Eingriffsmöglichkeiten in das Verfahren.** Es ist bei der Gestaltung eines Prognosesystems zu prüfen, ob eine Verbesserung der Vorhersage dadurch erreicht werden kann, dass man den menschlichen Experten einbezieht. Das kann aus folgenden Gründen vorteilhaft sein: Der Mensch hat nicht die Möglichkeit, alle Vergangenheitsinformationen zu verarbeiten. Er mag aber eine Reihe von – für das Prognosesystem „externen“ – Informationen besitzen, deren Berücksichtigung zu einer Verbesserung der Vorhersage führen kann.

Darüber hinaus verfügt er über ausgeprägte Fähigkeiten, Regelmäßigkeiten oder spezifische Abweichungen zu erkennen (Mustererkennung, Pattern Recognition). Es kann die Prognose verbessert werden, wenn dem Menschen die Möglichkeit gegeben wird, Besonderheiten in der Zeitreihenentwicklung zu interpretieren und seine Informationen dem Prognosesystem zur Verfügung zu stellen.

Prognoserechnungssysteme auf der Basis von Zeitreihen verzichten darauf, Kausalbeziehungen zu konstruieren, und gehen sozusagen mathematisch-statistisch „vorurteilsfrei“ an das Problem heran. Diese Beschränkung kann aufgehoben werden, wenn der Mensch als Informations- und Intelligenzträger eingeschaltet wird.

**Rechenzeit und Speicherplatzbedarf.** Der für die Anwendung eines Prognoserechnungssystems erforderliche Aufwand an Rechenzeit und Speicherkapazität wird durch die Anzahl der Elemente, für die man eine Vorhersage zu treffen hat, durch die Häufigkeit der Vorhersagen und die Methodenauswahl bestimmt. Kurzfristige Vorhersagen werden bei betriebswirtschaftlichen Anwendungen wie beispielsweise Lagerhaltungssystemen in kurzen Zeitabständen und für eine große Anzahl von Elementen durchgeführt. Unter diesen Voraussetzungen müssen selbst bei den heutigen Preis-Leistungs-Verhältnissen von Rechnern und Datenspeichern entsprechende Anforderungen an die Verfahren gestellt werden. Beispielsweise werden sie so konzipiert, dass sie entweder nicht alle vorhandenen Vergangenheitsinformationen benötigen oder dass sie in der Lage sind, diese in verdichteter Form zu erstellen bzw. zu verarbeiten.

## 2.1.2 Datenauswahl und -analyse

Bei der Konzeption eines Zeitreihen-Prognosemodells wird man sich zu Beginn der Überlegungen mit der Frage beschäftigen, welche der verfügbaren Daten zur Bildung der Zeitreihen verwendet werden sollen. Diese Entscheidung wird nicht nur von der Aktualität, Qualität und Form der verfügbaren Daten beeinflusst, sondern

auch von den verschiedenen Geschäftszusammenhängen im Umfeld der Prognoseaufgabe (vgl. dazu [1, S. 219 ff.]).

### **Datenquellen**

Die Problematik der Entscheidung, an welcher Stelle des Datenflusses die Daten zweckmäßig zu erfassen sind, soll am Beispiel der Umsatzprognose dargestellt werden:

Wählt man die Zahlungseingänge als Basis, so stützt man sich auf die effektiv eingetroffenen Gegenleistungen, erhält die Daten jedoch zu einem relativ späten Zeitpunkt. Entscheidet man sich für die Rechnungsausgänge, so hat man die Daten um das Zahlungsziel früher, jedoch sind unter Umständen Korrekturen für Gutschriften und Retouren erforderlich. Nimmt man die eintreffenden Kundenaufträge (Bestellungen), so stehen die Daten noch früher bereit, jedoch kann sich das Problem ergeben, dass ein Teil der Kundenaufträge storniert wird oder aus anderen Gründen nicht zur Auslieferung kommt.

Dieses Problem kann noch weiter verfolgt werden: Angenommen, man ist Produzent eines Artikels, der zunächst an einen Großhändler, von dort an einen Einzelhändler und von dort an den Endverbraucher geliefert wird. Wo setzt man für welche Prognosezwecke an? Für den eigenen unmittelbaren Absatz sind nur die Lieferungen an den Großhändler interessant. Will man die Prognosedaten für die Produktionsplanung verwenden, so wünscht man möglicherweise Kenntnisse über die Verkäufe des Großhändlers an die Einzelhändler oder gar der Einzelhändler an die Kunden. Letztere eilen den Bestellungen des Großhändlers mit einer gewissen Phasenverschiebung voraus; jedoch wird die Prognose schwieriger, da man nicht weiß, welche Lagerhaltungspolitik die Groß- und Einzelhändler betreiben. Um Kenntnis von den Verkäufen des Einzelhandels zu erlangen, gibt es verschiedene Hilfsmittel, z. B. Haushaltspanels, die Rücksendung von Garantiekärtchen oder Verlosungen, an denen der Käufer teilnimmt, wenn er Lose einsendet, die dem verkauften Artikel beigelegt waren. (In solchen Fällen stellt sich die Frage, ob diese Maßnahmen nicht mehr kosten, als die Verbesserung der Prognose wert ist.)

Dieses Beispiel zeigt: Es ist sinnvoll und auch notwendig, dass sich der Konstrukteur eines Prognosesystems in einem ersten Schritt mit den Geschäftsfällen befasst, die sich im Umfeld der Prognoseaufgabe abspielen. Ähnliche Vorschläge und Beispiele finden sich bei Armstrong ([1, S. 219 ff.]).

### **Prognoseobjekt und Verdichtungsebene**

Das „quasi-natürliche“ Objekt der Prognose ist die Zeitreihe der Produkte bzw. Artikel, die nachgefragt oder verkauft wurden. Man gewinnt sie aus den Modulen Auftragseingangserfassung oder Fakturierung (einschließlich der Erfassung an der Ladenkasse bzw. „Point of Sale“ im Einzelhandel) der betrieblichen Informationsverarbeitung.

Oft weisen diese Zeitreihen aber zu gravierende Unregelmäßigkeiten auf, die zufallsbedingt sein mögen. In solchen Konstellationen empfiehlt es sich, stärker zu aggregieren, z. B. die Vorhersage auf der Ebene der Produktgruppen durchzuführen und später die Ergebnisse mithilfe von Schlüsseln auf die Einzelerzeugnisse zu verteilen.

Bei starker Variantenfertigung entfallen auf die einzelnen Ausprägungen zu wenige Daten, sodass die statistische Basis nicht ausreicht. Man hört Aussagen, wie: „Wir stellen zwar pro Jahr einige tausend Produkte her, aber keine zwei davon sind zu hundert Prozent gleich.“ Eine Möglichkeit liegt darin, die Prognose auf der zweithöchsten Stücklistenstufe anzusiedeln, auf der die Varianten (z. B. „Extras“) nicht differenziert sind. Eine andere Alternative bedingt, nicht Enderzeugnisse, sondern Merkmale zum Gegenstand der Vorhersage zu machen („Merkmalsbasierte Vorhersage“, „Characteristic-based Forecasting“ [5]). So wird z. B. in der Automobilindustrie die Zahl der Schiebedächer oder die Menge an blauem Lack prognostiziert (siehe auch die Ausführungen von Mertens, Zeller und Große-Wilde in Kapitel 27 dieses Buches).

## **Datenanalyse**

In einem zweiten Schritt setzt man sich dann im Detail mit den Daten auseinander und stellt für eine größere Zahl von Perioden die Daten graphisch in Diagrammen dar, um allmählich ein Gefühl für deren Schwankungen und Periodizitäten zu erhalten. Die in der Praxis des Wirtschaftslebens verfügbaren Zeitreihen können mit Messfehlern behaftet oder unvollständig sein. Beides kann zu gravierenden Prognosefehlern führen und erfordert eine Bereinigung der Daten vor der Berechnung von Vorhersagewerten. Spezielle Anforderungen stellen auch Zeitreihen, die wiederholt Nullwerte enthalten und ein sporadisches Nachfrageverhalten repräsentieren. Armstrong ([1, S. 222 ff.]) schlägt hier vor, die Messwerte zu verdichten (z. B. zeitlich oder räumlich) oder die Nullwerte durch Mittelwerte aus früheren und späteren Methoden zu ersetzen. (In Kapitel 4 dieses Buches werden für diesen Nachfragetyp geeignete Verfahren beschrieben.)

Diese Voruntersuchung des Datenmaterials darf aber nicht dazu verleiten, bereits hier verallgemeinernde Aussagen über die Zeitreihen abzugeben, da die Gefahr groß ist, dass atypische Datenkonstellationen als typisch angesehen und daraus falsche Schlüsse für die Verfahrensauswahl gezogen werden. Diese Untersuchung hat vielmehr den Zweck, die Auswahl von Verfahren für eine mathematisch-statistische Datenanalyse zu erleichtern. Im Speziellen geht es darum, zufällige Schwankungen von systematischen zu unterscheiden und die Bildungsgesetze der systematischen Schwankungen festzustellen.

### 2.1.3 Prognoseintervall und Vorhersagezeitraum

Wir haben oben bereits festgehalten, dass wir uns mit Vorhersageverfahren befassen, die auf der Extrapolation von Zeitserien beruhen. Diese Zeitreihen bestehen aus Beobachtungswerten  $x_t$  ( $t = 0, 1, 2, \dots, T$ ), die jeweils in gleichen, diskreten Zeitabständen ermittelt wurden. Wenn nach jeder neu eintreffenden Beobachtung die Vorhersage wiederholt werden soll, ist das Prognoseintervall gleich dem Zeitabstand zwischen zwei Beobachtungen. Unter dem Vorhersagezeitraum versteht man die Länge der Periode, auf die sich ein Vorhersagewert bezieht. Diese Periode setzt sich aus einem oder mehreren Prognoseintervallen zusammen.

Im Folgenden sind einige Faktoren angeführt, die bei der Wahl der Größe von Prognoseintervall und Vorhersagezeitraum zu berücksichtigen sind. (In der Regel wird der Vorhersagezeitraum ein Vielfaches des Prognoseintervalls sein. Brown gibt als Faustregel an, dass das Prognoseintervall etwa ein Viertel bis ein Zehntel des Vorhersagezeitraums sein soll (vgl. [4, S. 43]).)

Für die Größe des Prognoseintervalls gilt:

1. Eine untere Grenze ist dadurch gegeben, dass die Laufzeitpunkte der Prognoseprogramme nicht vor denen jener Programme liegen können, welche die Beobachtungswerte für das Prognoseprogramm bereitstellen (z. B. liefern Lagerbestandsführungsprogramme Informationen über die Lagerabgänge, die dann wieder die Grundlage für die Bedarfsprognose sind).
2. Kurze Prognoseintervalle bzw. häufige Prognosen erfordern einen hohen Datenverarbeitungsaufwand und können dazu führen, dass das Modell stark auf Zufallsschwankungen reagiert.
3. Bei großen Prognoseintervallen hingegen erhöht sich die Gefahr, dass Veränderungen im Verhalten der Zeitreihe nicht rechtzeitig erkannt werden. Dem steht allerdings der Vorteil gegenüber, dass bei einer geringeren Anzahl von Prognosen weniger Planrevisionen durchzuführen sind.

Für die Größe des Vorhersagezeitraums gilt:

1. Eine untere Grenze ist durch die Zeitspanne vorgegeben, die zwischen dem Veranlassen einer Aktion aufgrund der Prognose und dem Eintreten des durch sie beabsichtigten Effekts verstreicht. Beispielsweise muss der Vorhersagezeitraum in einem Lagerhaltungssystem zumindest so groß wie die Zeitspanne sein, die zwischen einer Neubestellung und dem Eintreffen des Gutes in dem Lager vergeht.
2. Nach oben wird die Länge des Vorhersagezeitraums vor allem durch die geforderte Genauigkeit und Zuverlässigkeit des Verfahrens bestimmt, da sich mit zunehmendem Vorhersagezeitraum die Gefahr von Fehlprognosen – z. B. durch eine Trendänderung – erhöht.

## 2.2 Modelle zur Abbildung von Zeitreihen

Nachdem wir das Verhalten einer Zeitreihe analysiert haben, stehen wir vor der Aufgabe, ihre systematischen Veränderungen über die Zeit durch ein Modell darzustellen.

Gedanklich wollen wir so vorgehen, dass wir die beobachteten Werte als das Ergebnis eines „Prozesses“ plus einer nicht vorhersagbaren Zufallsabweichung auffassen.

Der Begriff „Prozess“ kann hier durchaus in einem physikalischen Sinne verstanden werden; man denke etwa an einen radioaktiven Zerfall oder an die Schwingungen eines Pendels. Zu bestimmten diskreten Zeitabständen führen wir nun eine Messung durch und halten die Werte  $x_t$  fest. Bei vielen physikalischen Prozessen ist die zugrunde liegende Gesetzmäßigkeit genau bekannt. Die Daten, mit denen man zu tun hat, sind ebenfalls das Ergebnis eines Prozesses, der sich aber niemals genau erfassen lässt, da er durch zu viele Faktoren und Interaktionen determiniert ist. Wir sind daher gezwungen, diesen uns unbekannten Prozess in einem deskriptiven Modell darzustellen. Bezüglich der Zufallsabweichungen, die ebenfalls in die Beobachtungswerte eingehen, wird angenommen, dass

1. der Erwartungswert null,
2. die Varianz konstant und
3. die Verteilung annähernd normal ist.

Die unbekannten Prozesse werden als deterministische Funktionen der Zeit betrachtet. Die Modelle haben die Aufgabe, diese Prozesse möglichst genau zu beschreiben. Für die Beobachtungswerte gilt:

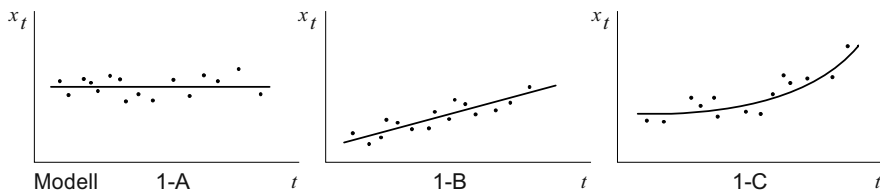
$$x_t = P_t + e_t \quad (P_t = \text{Prozess} , e_t = \text{Zufallsabweichung})$$

Im Folgenden werden nun eine Übersicht über die grundsätzlichen Möglichkeiten zur Modellbildung bei der kurzfristigen Zeitreihenprognose gegeben und die wichtigsten Modelle mathematisch dargestellt.

### 2.2.1 Grafische Übersicht über die Möglichkeiten zur Modellbildung

Die Modelle der Gruppe 1 (Abb. 2.1) haben gemeinsam, dass sie keine periodische Veränderung über der Zeit aufweisen.

In Modell A wird angenommen, dass der zugrunde liegende Prozess über die Zeit konstant ist, in Modell B, dass der Prozess eine linear, im Modell C, dass er

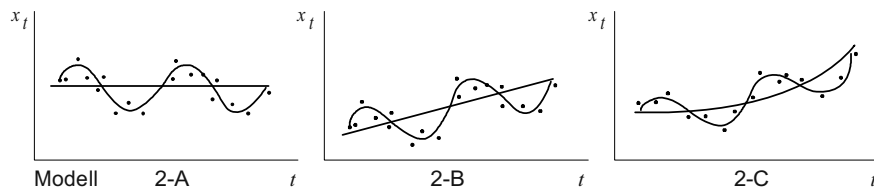


**Abb. 2.1** Modelle für Zeitreihen ohne Periodizität ([15, S. 312])

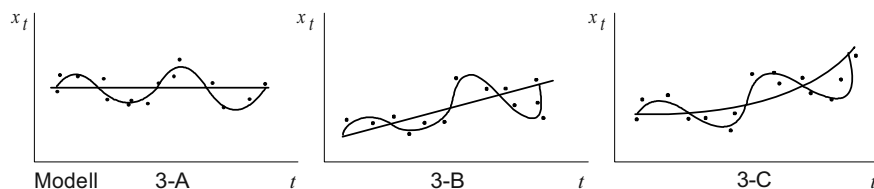
eine nichtlinear wachsende Funktion der Zeit ist. Dabei stellen die ausgezogenen Linien den Prozess dar, der durch das Modell beschrieben werden soll, während die Abstände der Beobachtungswerte von diesen Linien das Ergebnis der nicht vorher-sagbaren Zufallsabweichungen sind.

Die Darstellungen in den Abb. 2.2 und 2.3 sind zwei weitere Modellgruppen, die dadurch entstanden, dass den Modellen der Gruppe 1 zyklische Schwankungen überlagert wurden.

In der Modellgruppe 2 sind diese Schwankungen additiv überlagert, das heißt, dass die Größe ihrer Amplituden von der Höhe des Grundwertes unabhängig ist. In der Modellgruppe 3 hingegen sind die Schwankungen multiplikativ überlagert, sodass die Amplituden zeitabhängig sind.



**Abb. 2.2** Modelle für Zeitreihen, bei denen periodische Schwankungen den Grundwert additiv überlagern ([15, S. 313])



**Abb. 2.3** Modell für Zeitreihen, bei denen periodische Schwankungen den Grundwert multiplikativ überlagern ([15, S. 314])

## 2.2.2 Mathematische Beschreibung der wichtigsten Modelle

Im Folgenden werden die statistischen Modelle behandelt, die zur Beschreibung der datenerzeugenden Prozesse Verwendung finden (vgl. hierzu [4, S. 57 ff.]).

### Konstantes Modell

Betrachtet man das Modell 1-A (Abb. 2.1), so geht daraus die Annahme hervor, dass der zugrunde liegende Prozess über die Zeit annähernd konstant ist. Wir verwenden daher zur Beschreibung des Verlaufs ein Modell von der Form  $P_t = a$ , wobei  $a$  der „wahre“ Wert der Konstanten ist, den wir nicht kennen. (Die Gleichung der Geraden, die in Abb. 2.1 durch die Punkte gelegt wurde, ist uns nicht bekannt.)

Beziehen wir die Zufallsabweichungen ( $e_t$ ) in unsere Überlegungen mit ein, so können wir sagen, dass die Beobachtungswerte  $x_t$  Stichproben aus einer bestimmten Verteilung mit dem Mittelwert  $a$  sind. Da dieser Mittelwert unbekannt ist, ist man gezwungen, aus den vorliegenden Vergangenheitsdaten einen Schätzwert zu errechnen.

Die Indizierung in diesem Beitrag erfolgt in Anlehnung an Brown [4]. Für die Zeit als unabhängige Variable wird ( $t$ ) verwendet ( $t = 0, 1, 2, \dots, T$ ), wobei die Gegenwart den Index  $T$  erhält. Beispielsweise ist  $x_T$  der Beobachtungswert, der in der letzten Beobachtungsperiode eingetroffen ist.

Ein Schätzwert, der nach dem Eintreffen von  $x_T$  aus allen oder einer bestimmten Menge  $N$  Vergangenheitsdaten errechnet wurde, soll durch das Symbol  $\hat{\phantom{x}}$  und durch den Index  $T$  gekennzeichnet werden. Beispiel:  $\hat{a}_T$  ist ein Schätzwert für den Wert  $a$  in einem konstanten Modell (siehe oben), der nach dem Eintreffen von  $x_T$  berechnet wurde.

Eine Vorhersage, die in der Periode  $T$  auf Basis der Schätzwerte  $\hat{a}_T, \hat{b}_T, \dots$  usw. abgegeben wird, erhält ebenfalls das Symbol  $\hat{\phantom{x}}$  und den Index  $T$ . Zur Kennzeichnung der Periode, für die die Vorhersage gilt, wird  $(T + i)$  verwendet. Führen wir nun eine Prognose für die Periode  $(T + i)$  durch, so bedeutet das, dass wir den Mittelwert der Verteilung in der Periode  $(T + i)$  vorhersagen.

Die Vorhersagegleichung lautet:  $\hat{x}_{T+i} = \hat{a}_T$ . Für ihre Gültigkeit müssen folgende Annahmen zutreffen:

1. Der den Daten zugrunde liegende Prozess kann durch eine Konstante dargestellt werden.
2. Diese Konstante (der Koeffizient des Modells) kann man durch eine Mittelwertbildung über Vergangenheitsdaten abschätzen.
3. Der Mittelwert der Verteilung kann durch den Wert  $\hat{a}_T$  mit hinreichender Genauigkeit dargestellt werden.



## Lineares Modell

Gehen wir von dem zweiten Modell in Abb. 2.1 aus, so sieht man, dass die beobachteten Werte einen Trend aufweisen, der über der Zeit konstant ist (gleiche Zuwachsraten). In diesem Fall nehmen wir an, dass sich der zugrunde liegende Prozess durch ein Modell von der Form  $P_t = a + bt$  abbilden lässt. Dabei stellt  $a$  den Durchschnittswert zu dem Zeitpunkt dar, in dem  $t$  als null angenommen wird (relative Zeit). Der Trendwert  $b$  gibt die Änderungsrate des ordinalen Wertes innerhalb einer Periode an. Da wiederum die Werte von  $a$  und  $b$  unbekannt sind, ist eine Schätzung aus den Daten der Vergangenheit notwendig. Diese Schätzung soll nach Ablauf einer bestimmten Periode durch die Werte  $\hat{a}_T$  und  $\hat{b}_T$  dargestellt werden.

Beachten wir, dass unsere Beobachtungswerte durch die Punktwolke in Abb. 2.1 gegeben sind. Die Schätzwerte unterliegen zeitlichen Schwankungen. Die Vorhersagegleichung ist bei einem linearen Modell durch  $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$  gegeben. Dabei ist zu berücksichtigen, dass die Vorhersage nur von der *relativen* Zeit abhängt und daher gilt, dass  $\hat{a}_T = \hat{a}_{T-1} + \hat{b}_{T-1}$ .

## Modelle höherer Ordnung

Liegt der Fall vor, dass auch der Trendanstieg über die Zeit (bzw. innerhalb unseres Beobachtungszeitraums) nicht mehr konstant ist (vgl. Modell 3 in Abb. 2.1), so muss dieser Sachverhalt durch eine Erweiterung des Modells wiedergegeben werden, beispielsweise durch  $P_t = a + bt + ct^2$ . Es ist erkennbar, dass wir im Falle eines quadratischen Modells drei Koeffizienten abschätzen müssen. (Allgemein gilt, dass bei Verwendung eines Modells  $n$ -ter Ordnung  $(n+1)$  Koeffizienten zu schätzen sind.)

Haben wir z. B. das oben angegebene quadratische Modell zur Darstellung unseres Prozesses ausgewählt, so ist unsere Vorhersage gegeben durch die Gleichung  $\hat{x}_{T+1} = \hat{a}_T + \hat{b}_T i + \hat{c}_T i^2$ .

Ein Verfahren zur Abschätzung der Koeffizienten von Modellen höherer Ordnung wird z. B. von Box und Jenkins [3] beschrieben.

## Modelle für Zeitreihen mit periodischen Schwankungen

Viele Zeitreihen in der Praxis weisen periodische Schwankungen auf (Abb. 2.2 und 2.3). Werden diese im Modell berücksichtigt, so vergrößert sich der erforderliche mathematisch-statistische Aufwand erheblich.

Lassen sich die Schwankungen durch die Verwendung von Winkelfunktionen darstellen, so gelangt man beispielsweise zu einem Modell von der Form  $P_t = a \cos(2\pi t/12)$ .

Diese Verfahren werden hier nicht näher erörtert, sie stehen im Mittelpunkt des in Kapitel 3 dieses Buches enthaltenen Beitrags.

## 2.3 Methoden zur Abschätzung der Modellkoeffizienten

### 2.3.1 Übersicht

Nach der Auswahl eines Modells, das in der Lage ist, das Verhalten einer Zeitserie zu beschreiben, benötigt man Verfahren, um die unbekannten Modellkoeffizienten abzuschätzen.

Eine erste Einteilung dieser Verfahren richtet sich danach, auf welche Modelle sie anwendbar sind. Zur Abschätzung des Koeffizienten im konstanten Modell kann entweder

- ein Verfahren der gleitenden Durchschnitte oder
- exponentielles Glätten erster Ordnung

verwendet werden (vgl. dazu den folgenden Abschnitt 2.3.2).

Sind zwei Koeffizienten in einem linearen Modell zu bestimmen, so stehen folgende Verfahren zur Verfügung (vgl. dazu Abschnitt 2.3.3):

1. Exponentielles Glätten mit Trendkorrektur
2. Exponentielles Glätten zweiter Ordnung nach Brown
3. Exponentielles Glätten zweiter Ordnung nach Holt
4. Exponentielles Glätten zweiter Ordnung mit gedämpftem Trend
5. Exponentielles Glätten mit Fehlerdifferenz nach Holt

Als Einteilungskriterium haben wir hier die Ordnung des Polynoms verwendet, mit dessen Hilfe wir den Prozess abbilden wollen. Eine andere Einteilungsmöglichkeit besteht darin, dass man unterscheidet, ob die Vergangenheitsdaten gleich gewichtet bei der Berechnung des Mittelwerts verwendet werden (Regression und einfache gleitende Durchschnitte) oder ob sie mit verschiedenen Gewichten zum Mittelwert beitragen (gewogene gleitende Durchschnitte, alle Verfahren des exponentiellen Glättens). Im letzten Fall berücksichtigt man, dass jüngere Daten meist stärker für die zukünftige Entwicklung relevant sind als dies für ältere Daten gilt.

Bei den Verfahren der exponentiellen Glättung kann man schließlich noch eine Unterscheidung nach der Zahl der verwendeten Glättungsfaktoren treffen. Den Schwerpunkt unserer Darstellung werden wir auf Methoden legen, die mit *einem* Glättungsfaktor auskommen, weil diese Verfahren vergleichsweise einfach sind und daher auch am häufigsten benutzt werden. Daneben sind aber auch Algorithmen bekannt, die zwei, drei oder noch mehr Glättungsfaktoren verwenden.

Die Tab. 2.1 gibt einen Überblick über die Verfahren nach den oben beschriebenen Einteilungskriterien.

Hyndman geht bei der Klassifikation von der Trendkomponente aus und kommt dann zu fünf (kein Trend, additiv, multiplikativ, mit und ohne Dämpfung) und un-

**Tab. 2.1** Übersicht über die Verfahren zur Abschätzung der Modellkoeffizienten

Verfahren	Anwendung auf			Gewichtung		Zahl der Glättungs- faktoren
	Konstanter Prozess	Linearer Prozess	Polynome höherer Ordnung	Gleich	Exponen- tiell	
Gleitende Durchschnitte erster Ordnung	X			X		0
Exponentielles Glätten erster Ordnung	X				X	1
Exponentielles Glätten mit Trendkorrektur		X			X	1
Exponentielles Glätten zweiter Ordnung (Brown)		X			X	1
Exponentielles Glätten zweiter Ordnung (Holt)		X			X	2
Exponentielles Glätten mit gedämpftem Trend		X	(X)		X	2
Exponentielles Glätten mit Fehlerdifferenz (Holt)		X	(X)		X	3

ter Berücksichtigung von Saisonschwankungen zu 15 verschiedenen Methoden des exponentiellen Glättens ([11, S. 11 ff.]).

### 2.3.2 Methoden für das konstante Modell

In diesem Punkt wollen wir uns mit verschiedenen Methoden der Mittelwertbildung befassen, die es ermöglichen, sukzessive eine möglichst gute, mit jeder neuen Beobachtung wiederholte Schätzung der Modellkoeffizienten durchzuführen.

Es wird angenommen, dass der den Daten zugrunde liegende Prozess konstant ist (vgl. Abschnitt „Gleitende Durchschnitte erster Ordnung“), also  $P_t = a$ , die beobachteten Werte durch  $x_t = P_t + e_t$  dargestellt werden können und  $e_t$  Stichproben einer bestimmten Verteilung mit dem Mittelwert 0 und der Varianz  $\sigma^2$  sind.

Da der „wahre“ Wert  $a$  nicht bekannt ist, muss versucht werden, aus den vorhandenen Daten  $x_t$  einen möglichst guten Schätzwert  $\hat{a}_T$  zu berechnen, der dann als Prognosewert  $\hat{x}_{T+i} = \hat{a}_T$  verwendet werden kann.

### Gleitende Durchschnitte erster Ordnung

Bei gleitenden Durchschnitten wird aus  $N$  Werten  $x_t$  ein Mittelwert dadurch gebildet, dass man jeweils die jüngste Beobachtung statt der ältesten in die Berechnung einbezieht. Dieser gleitende Durchschnitt kann für beliebige Perioden  $t$  mit der Formel

$$M_t = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-N+1}}{N} \quad (2.1)$$

berechnet werden.

Streng genommen gilt der in  $T$  berechnete Mittelwert nur für die Perioden  $T - (N - 1)/2$ . Solange aber der Prozess konstant ist, machen wir keinen Fehler, wenn wir diesen Wert als Prognosewert für die Periode  $(T + i)$  verwenden. (Weist die Zeitreihe einen Trend auf, so ist diese Vorgehensweise nicht mehr zulässig.)

Hat man nach Eintreffen von  $x_T$  einen neuen Mittelwert  $M_T$  aus den Daten  $x_t$  ( $t = T, \dots, T - N + 1$ ) berechnet, so wird  $M_T$  für den Schätzwert  $\hat{a}_T$  benutzt. Eine Vorhersage für die Periode  $(T + i)$  ist durch  $\hat{x}_{T+i} = \hat{a}_T$  gegeben.

Rechentechnisch ist es sehr umständlich, wenn wir in jeder Periode summieren und durch  $N$  dividieren müssen. Es ist einfacher, die Berechnung nach der Formel

$$M_t = M_{t-1} + (x_t - x_{t-N})/N \quad (2.2)$$

durchzuführen. Man berechnet also den neuen Mittelwert dadurch, dass zu dem der Vorperiode der  $N$ -te Teil der Differenz aus dem neuen Beobachtungswert  $x_T$  und dem Wert  $x_{t-N}$  addiert wird.

Setzen wir  $(x_t - x_{t-N}) = d_t$  und  $1/N = k$ , so erhalten wir:

$$M_t = M_{t-1} + k \cdot d_t \quad (2.3)$$

Zur Berechnung benötigen wir den Mittelwert der Vorperiode und die Beobachtungen  $x_t$  bis  $x_{t-N}$ .

Diese Methode erfüllt das Optimalitätskriterium einer Minimierung der Abstandsquadrate zwischen Modell- und Beobachtungswerten (zum Beweis vgl. Abschnitt „Exponentielles Glätten mit Trendkorrektur“), ist aber an folgende Annahmen (siehe dazu Abschnitt „Konstantes Modell“) gebunden:

1. Der Prozess ist durch eine Konstante darstellbar.
2. Diese kann durch einen Mittelwert über Vergangenheitsdaten abgeschätzt werden.
3. Der Mittelwert der Verteilung im Vorhersagezeitraum ( $i$  Perioden) wird durch den Schätzwert  $\hat{a}_T$  mit hinreichender Genauigkeit repräsentiert.

Nachteile dieser Methode:

1. Die obigen Annahmen sind für viele Zeitserien in der Praxis nicht zutreffend.
2. Der Aufwand für die Speicherung und Verarbeitung der Vergangenheitsdaten nimmt mit steigender Zahl von Prognoseobjekten und größeren  $N$  stark zu.
3. Tritt der Fall ein, dass sich  $a$  (beispielsweise durch Änderung der Marktsituation) von  $a_1$  auf  $a_2$  ändert, dann dauert es genau  $N$  Perioden, bis die neuen Schätzwerte für eine Vorhersage brauchbar sind. Das heißt aber, dass wir streng genommen diese  $N$  Perioden lang keine Vorhersage machen können.
4. Wird aus diesem Grund  $N$  verkleinert, so besteht die Gefahr, dass Zufallsschwankungen zu stark in die Vorhersagen eingehen.
5. Durch den konstanten Faktor  $1/N = k$  wird allen Daten das gleiche Gewicht zugeordnet. Hingegen ist die Annahme plausibel, dass die Daten der jüngsten Vergangenheit die Entwicklung besser repräsentieren als die älteren Daten und daher eine unterschiedliche Gewichtung zur Verbesserung der Prognose beitragen kann.

Wir wollen diese Nachteile in Form von Anforderungen an ein besser geeignetes Verfahren zusammenfassen:

1. Die Vergangenheitsentwicklung muss durch eine möglichst geringe Anzahl von Informationen dargestellt werden können.
2. Es muss die Möglichkeit vorhanden sein, mit geringem Aufwand die Reaktionszeit (Rate of Response) des Systems zu ändern.
3. Auch bei trendähnlichen Tendenzen soll das Verfahren in der Lage sein, sich diesen anzupassen.
4. Die Vergangenheitsdaten sollen nicht gleich, sondern proportional zu ihrer Relevanz für die zukünftige Entwicklung gewichtet werden.

### **Gewogene gleitende Durchschnitte**

Ein erster Versuch, die Nachteile des Verfahrens der ungewogenen gleitenden Durchschnitte zu verringern, besteht darin, die  $N$  Werte  $x_t$ , die jeweils zur Mittelwertbildung herangezogen werden, mit verschiedenen Koeffizienten zu gewichten. Die Auswahl der Koeffizienten wird in der Weise getroffen, dass die Daten proportional zu ihrer Relevanz für die zukünftige Entwicklung in den Mittelwert eingehen.

Damit aber durch die Gewichtung keine systematischen Komponenten eingeführt werden, muss für die Gewichtungsfaktoren  $k$  gelten:

$$\sum_{n=1}^N k_n = 1$$

Wenn die Zeitreihe also stellenweise Trendänderungen oder andere Schwankungen aufweist, wird dieses Verfahren bessere Ergebnisse als ungewogene gleitende Durchschnitte liefern. Diese Tatsache ist aus Abb. 2.4 deutlich zu ersehen. Man kann die Daten auch von unerwünschten Saisoneinflüssen bereinigen, wenn man die Durchschnittsbildung auf vollständige Saisonzyklen bezieht (z. B. mit vier, acht oder zwölf Monatsperioden) [2].

Der Nachteil der geringen Flexibilität in Bezug auf die Reaktionszeit des Verfahrens der ungewogenen gleitenden Durchschnitte bleibt aber auch bei dieser Methode bestehen. Dazu kommt noch, dass die Bestimmung der Koeffizienten  $k_n$  aufwändig ist.

### **Exponentiell gewogene Durchschnitte (exponentielles Glätten erster Ordnung)**

In diesem Abschnitt wird ausführlich das Verfahren des exponentiellen Glättens (Exponential Smoothing) beschrieben. Ausführlich deshalb, weil exponentielles Glätten weniger nur ein Verfahren als vielmehr ein Prinzip darstellt, welches für viele spezielle Vorhersagemethoden von grundlegender Bedeutung ist. Dazu kommt noch, dass sich alle Autoren über das Grundprinzip des exponentiellen Glättens einig sind, jedoch über die verschiedenen Erweiterungen, Verallgemeinerungen und speziellen Verfahren in der Literatur sehr verschiedene Ansichten vertreten werden. Daher besitzt das Prinzip des exponentiellen Glättens eine gewisse Allgemeingültigkeit, die – will man die darauf aufbauenden höheren Verfahren verstehen – eine eingehende Erörterung rechtfertigt (vgl. dazu auch [11, S. 13 ff.]).

### **Übergang von gleitenden Durchschnitten zum exponentiellen Glätten**

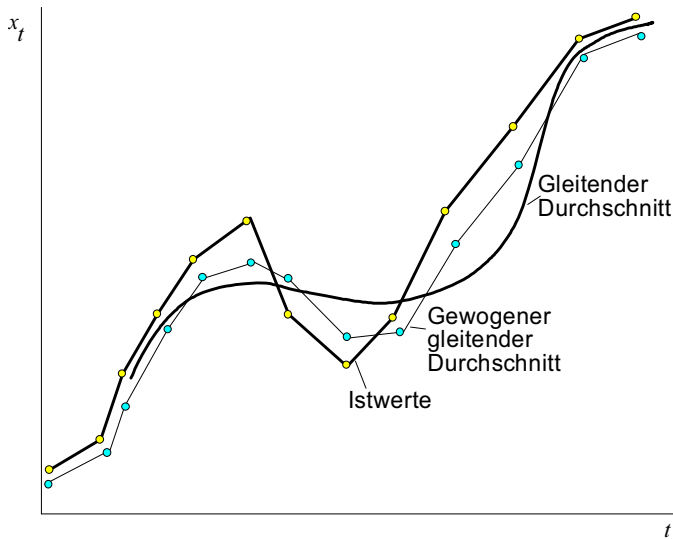
Ein Nachteil der Methode der gleitenden Durchschnitte ist die Notwendigkeit, dass zur Berechnung alle  $N$  Werte  $x_t$  gespeichert sein müssen. Da den Überlegungen ein konstanter Prozess zugrunde gelegt wurde, ist die Prognose für die Periode  $(T + i)$  durch  $\hat{x}_{T+i} = \hat{a}_T = M_T$  gegeben:

Wenn der Mittelwert  $M_T = \hat{a}_T$  als Vorhersagewert für die zukünftige Beobachtung  $\hat{x}_{T+i}$  verwendet wird, dann kann er auch als Schätzwert für den Beobachtungswert  $x_{T-N}$  eingesetzt werden. Wird dieser Vorgang laufend durchgeführt und jeweils anstatt des ältesten Wertes  $x_{T-N}$  der verfügbare Wert  $M_{T-1}$  eingesetzt, so erhält man für eine näherungsweise Berechnung des Mittelwerts

$$\hat{M}_t = \hat{M}_{t-1} + (x_t - \hat{M}_{t-1})1/N . \quad (2.4)$$

Setzt man nun  $\hat{M}_t = S_t(x)$  und  $1/N = A$ , so erhält man

$$S_t(x) = S_{t-1}(x) + A(x_t - S_{t-1}(x)) . \quad (2.5)$$



**Abb. 2.4** Vorhersagen nach den Verfahren der gleitenden Durchschnitte und der gewogenen gleitenden Durchschnitte ([11, S. 239])

(Die Bezeichnung  $S_t(x)$  wird im Folgenden für alle exponentiell gewogenen (geglätteten) Mittelwerte verwendet.)

Damit haben wir durch einen Analogieschluss zur Gleichung (2.2) die Grundformel des exponentiellen Glättens hergeleitet. Durch Umrechnungen ergibt sich:

$$S_t(x) = Ax_t + (1 - A)S_{t-1}(x) \quad (2.6)$$

Dieses  $S_t(x)$  ist ein exponentiell geglätteter Mittelwert, der zur Abschätzung von  $a_t$  verwendet wird.  $A$  wird als Glättungsfaktor bezeichnet, der nur ähnlich, aber keineswegs gleich dem Wert  $1/N$  ist, wie sich aus einer mathematischen Betrachtung im nächsten Abschnitt ergeben wird.

### Prinzip des exponentiellen Glättens

Wir wollen die Konsequenz unserer Vorgehensweise einer mathematischen Betrachtung unterziehen. Dazu gehen wir von der Formel (2.6)

$$S_t(x) = Ax_t + (1 - A)S_{t-1}(x)$$

aus und setzen

$$S_{t-1}(x) = Ax_{t-1} + (1 - A)S_{t-2}(x)$$

in Gleichung (2.6) ein. Wir erhalten

$$S_t(x) = Ax_t + A(1 - A)x_{t-1} + (1 - A)^2 S_{t-2}(x) .$$

Berechnet man den Wert für  $S_{t-2}(x)$  und setzt ihn wieder in die letzte Gleichung ein, so ergibt sich

$$S_t(x) = Ax_t + A(1 - A)x_{t-1} + A(1 - A)^2 x_{t-2} + (1 - A)^3 S_{t-3}(x)$$

und durch weiteres Einsetzen

$$S_t(x) = Ax_t + A(1 - A)x_{t-1} + \cdots + A(1 - A)^i x_{t-i}$$

bzw.

$$S_t(x) = A \sum_{i=0}^{\infty} (1 - A)^i x_{t-i} . \quad (2.7)$$

Man erkennt daraus, dass jeder Wert  $x_{t-i}$  ( $i = 0, 1, 2, \dots$ ) mit dem Koeffizienten  $A(1 - A)^i$  gewichtet wird.

Die Summe dieser Koeffizienten ist 1, wie leicht gezeigt werden kann: Setzt man  $A = p$  und  $(1 - A) = q$ , so ergibt der Ausdruck  $\sum_{i=0}^{\infty} A(1 - A)^i$  eine geometrische Reihe von der Form  $pq^0 + pq^1 + pq^2 + \cdots + pq^i + \cdots$  mit der Summe  $p/(1 - q)$ . Dieser Wert ist 1, wenn für  $p$  wieder  $A$  und für  $q(1 - A)$  eingesetzt wird. Dadurch ist sichergestellt, dass durch die Gewichtung kein systematischer Fehler in die Berechnungen eingeht.

Wir wollen wiederholen, welche Folgen die Substitution von  $x_{t-N}$  durch den Wert  $M_{t-1}$  (vgl. dazu Gleichungen (2.2) und (2.4)) hat. Wie Abb. 2.5 zeigt, wurden beim Verfahren der gleitenden Durchschnitte jeweils  $N$  Daten  $x_t$  mit dem gleichen Gewicht zur Berechnung des Durchschnitts benutzt.

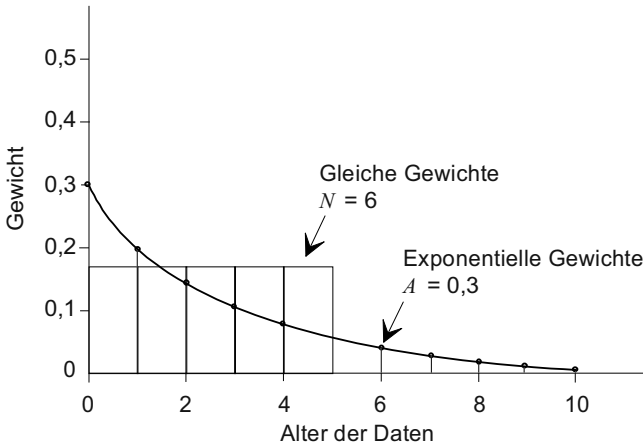
Durch die Substitution  $x_{t-N} = M_{t-1}$  sind folgende Änderungen eingetreten:

1. Es werden *alle* Daten der Vergangenheit zur Berechnung des Durchschnitts herangezogen.
2. Die Gewichtung der Daten erfolgt exponentiell fallend mit  $A(1 - A)^i$ .

Praktisch allerdings werden die älteren Daten – je nach Größe von  $A$  – ihren Einfluss auf den Durchschnitt nach wenigen Perioden verlieren.

Diese Aussagen gelten aber nur für den Fall, dass  $0 < A < 1$  ist. Wird  $A$  gleich null gewählt, so folgt aus (2.6), dass der neue Glättungswert jeweils gleich dem vorhergegangenen ist und daher überhaupt nicht auf Veränderung des numerischen Werts von  $x_t$  reagiert. Ist  $A = 1$ , so ist der neue Glättungs- jeweils gleich dem neuen Beobachtungswert.





**Abb. 2.5** Gewichtung der Vergangenheitsdaten beim Verfahren der gleitenden Durchschnitte und beim exponentiellen Glätten ([4, S. 102])

### Bestimmung des Glättungsfaktors

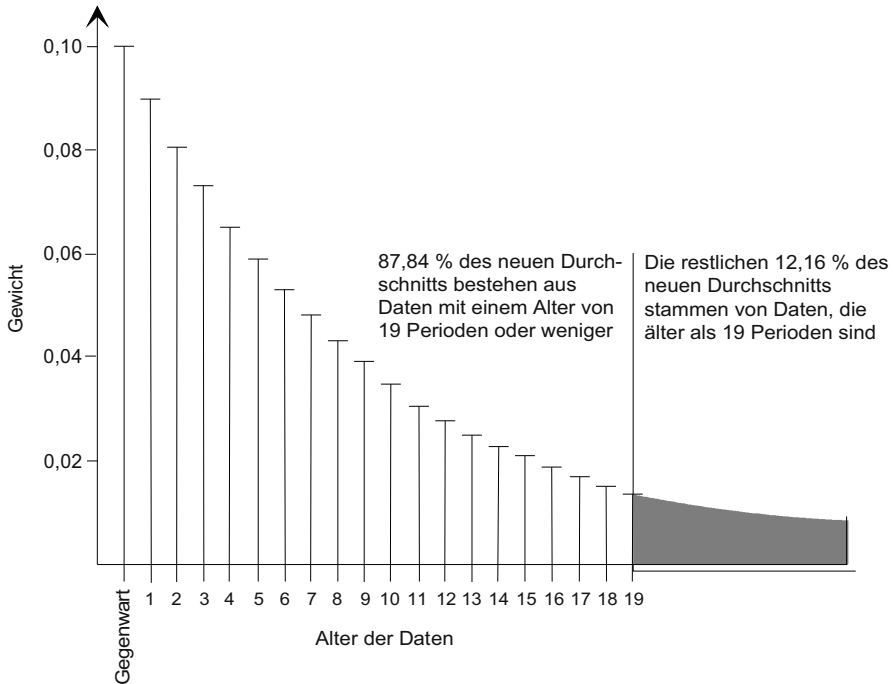
Aus den bisherigen Ausführungen geht hervor, dass die Größe des Wertes  $A$  entscheidend für die Reagibilität (bzw. Stabilität) des Verfahrens in Bezug auf Zufallsschwankungen ist. Durch die Wahl von  $A$  wird bestimmt, mit welchem Gewicht die Vergangenheitsdaten in den Glättungswert eingehen.

Bei dem Verfahren der gleitenden Durchschnitte legt man die Reagibilität des Systems direkt durch die Anzahl der Beobachtungen ( $N$ ) fest, die zur Berechnung des Mittelwerts herangezogen wurden, ohne eine unterschiedliche Gewichtung vorzunehmen. Beim exponentiellen Glätten wird indirekt über den Glättungsfaktor die Zahl der Werte bestimmt, die in dem Mittelwert hauptsächlich berücksichtigt werden sollen. Zur Verdeutlichung dienen die Abb. 2.6 und 2.7.

Wenn die Größe  $A$  für die Brauchbarkeit des Verfahrens entscheidend ist, dann muss nach Methoden gesucht werden, mit denen man ein geeignetes  $A$  bestimmen kann. Reif ([16, S. 9]) gibt als Erfahrungsregel an, dass  $A$  zwischen 0,1 und 0,3 zu wählen ist, Hyndman weist darauf hin, dass mit Zustandsraummodellen bessere Methoden zur Bestimmung des Glättungsparameter  $A$  und des Startwertes für die Vorhersage verfügbar sind ([11, S.14]).

Eine andere Möglichkeit besteht darin,  $A$  so zu wählen, dass etwa so viele Werte  $x_t$  wie bei dem Verfahren der gleitenden Durchschnitte zur Berechnung herangezogen werden. Hat man dort beispielsweise mit  $N = 50$  befriedigende Ergebnisse erzielt, so kann man ein  $A$  berechnen, bei dessen Verwendung ebenfalls annähernd die letzten 50 Werte in die Rechnung eingehen.

Zu diesem Zweck wollen wir ein durchschnittliches Alter  $\bar{T}$  unserer Daten berechnen. Im Falle der gleitenden Durchschnitte ist dieses Alter durch  $\bar{T} = (0 + 1 +$



**Abb. 2.6** Gewichtung der Vergangenheitsdaten mit  $A = 0,1$  ([16, S. 10])

$2 + \dots + N - 1)/N$  zu ermitteln. (Die jüngste Beobachtung ist null Perioden alt, die vorhergegangene eine Periode usw.)

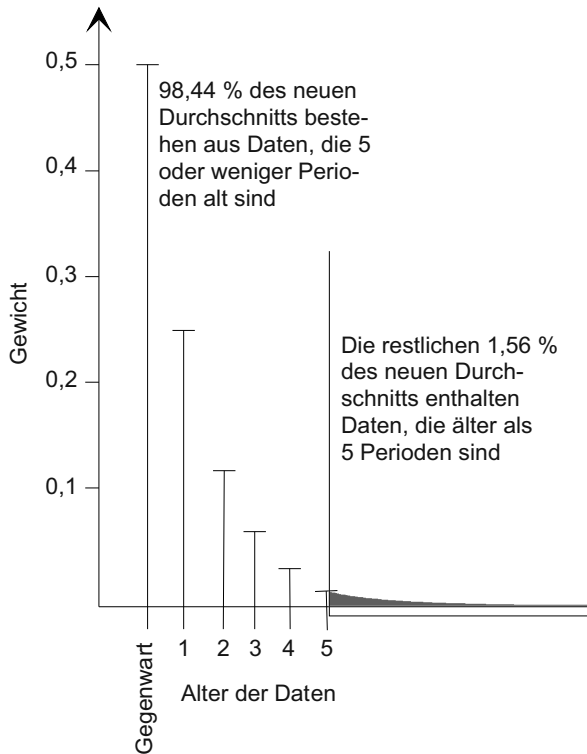
Die Summe über  $0, 1, 2, \dots, N - 1$  ist gleich  $(N - 1)N/2$ , sodass sich das mittlere Alter der Beobachtungen beim Verfahren der gleitenden Durchschnitte zu  $\bar{T} = (N - 1)/2$  ergibt.

Das durchschnittliche Alter wurde bestimmt durch die Summe der Produkte aus

1. dem Alter der Daten und
2. dem Gewicht, das ihnen das Verfahren verleiht.

Beim Verfahren der gleitenden Durchschnitte wird allen Daten das gleiche Gewicht  $1/N$  zugeordnet. Hingegen müssen wir beim exponentiellen Glätten das Alter der Daten mit dem entsprechenden Gewicht  $A(1 - A)^i$  multiplizieren, um das mittlere Alter der Daten zu erhalten. Setzen wir  $(1 - A) = B$ , so ist

$$\bar{T} = 0A + 1AB + 2AB^2 + \dots = A \sum_{i=0}^{\infty} iB^i.$$



**Abb. 2.7** Gewichtung der Vergangenheitsdaten mit  $A = 0,5$  ([16, S. 10])

Da für  $0 \leq B < 1$  gilt, dass

$$\sum_{k=0}^{\infty} k B^k = B/(1-B)^2$$

erhält man  $\bar{T} = AB/(1-B)^2$  bzw. da  $B = 1-A$ ,  $\bar{T} = (1-A)/A$ . Wenn wir nun den Glättungsfaktor so wählen wollen, dass er den Daten dasselbe Durchschnittsalter wie beim Verfahren der gleitenden Durchschnitte gibt, so erhalten wir  $(N-1)/2 = (1-A)/A$  und durch Auflösen nach  $A$

$$A = 2/(N+1) .$$

Wenn wir beispielsweise wünschen, dass sich jeweils die Gewichte der letzten fünf Daten zu annähernd 100 Prozent ergänzen (annähernd deshalb, weil beim exponentiellen Glätten immer noch ein Restprozentsatz älterer Daten zur Mittelwertbildung beiträgt und wir unsere Berechnung auf gleiches *Durchschnittsalter* gestützt haben), dann müssen wir ein  $A$  von 0,33 wählen. In Tab. 2.2 sind für einige

**Tab. 2.2** Entsprechende Werte für  $N$  und  $A$ , die gleiches Durchschnittsalter ergeben ([4, S. 108])

$N$ bei gleitendem Durchschnitt	Äquivalenter Wert für $A$
3	0,500
4	0,400
5	0,333
6	0,286
9	0,200
12	0,154
19	0,100
39	0,050
199	0,010

Werte von  $N$  entsprechende  $A$ , die beim exponentiellen Glätten Durchschnitte mit demselben Durchschnittsalter ergeben, zusammengestellt.

Ist nun die Annahme gerechtfertigt, dass mit einem bestimmten  $N$  bei der Methode der gleitenden Durchschnitte befriedigende Ergebnisse erzielt werden können, so kann man mithilfe dieser Tabelle ein äquivalentes  $A$  auswählen.

Bei der Bestimmung des Faktors  $A$  ist noch zu beachten:

1. Ergeben empirische Untersuchungen, dass  $A$  größer als 0,3 zu wählen sei, so wäre es zweckmäßig, zu untersuchen, ob ein konstantes Modell überhaupt noch zugrunde gelegt werden kann oder ob die Daten nicht doch einen Trend oder ein Saisonverhalten aufweisen.
2. Wenn man bei Anwendung der gleitenden Durchschnitte ein  $N = 12$  gewählt hat, um Saisonschwankungen auszugleichen, dann ist ein äquivalentes  $A$  von 0,154 beim exponentiellen Glätten nicht geeignet, den gleichen Effekt zu erzielen.

### Reaktion auf plötzliche Veränderungen

Weitere Einsichten in die Bedeutung des Glättungsfaktors  $A$  kann man dadurch gewinnen, dass man die Zeitserie bewusst verändert und beobachtet, welche Auswirkungen diese Änderung auf die folgenden Perioden hat. Dabei sollte man von „reinen“ Daten ausgehen, also von Daten, die frei von Zufallsabweichungen sind (vgl. dazu [12, S. 241 f.]).

### Reaktion auf einen Impuls

Wir nehmen nun an, dass die Werte  $x_t$  bis zu einer Periode  $t$  konstant sind ( $\dots x_{t-2} = x_{t-1} = x_t$ ) und in der Periode  $(t + 1)$  ein Impuls in der Form auftritt, dass  $x_{t+1} = (1 + x_t)$  und  $x_{t+2}$  wieder gleich  $x_t$  sind.

Das heißt aber auch, dass der Glättungswert  $S_t(x)$  bis zum Zeitpunkt  $t$  gleich  $x_t$  war, da die Daten keine Zufallsabweichung enthalten:

$$S_{t+1}(x) = Ax_{t+1} + (1 - A)S_t(x)$$

Da nun  $S_t(x) = x_t$  und  $x_{t+1} = 1 + x_t$ , erhält man

$$\begin{aligned} S_{t+1}(x) &= A(1 + x_t) + (1 - A)x_t = x_t + A = x_t + A(1 - A)^0 \\ S_{t+2}(x) &= Ax_{t+2} + (1 - A)S_{t+1}(x) . \end{aligned} \quad (2.8)$$

Da  $S_{t+1}(x) = (x_t + A)$  und  $x_{t+2} = x_t$  gilt  $S_{t+2}(x) = Ax_t + (1 - A)(x_t + A) = Ax_t + x_t + A - x_tA - A^2 = x_t + A(1 - A)^1$  und für die  $i$ -te Periode nach der Störung  $S_{t+i}(x) = x_t + A(1 - A)^{i-1}$ .

Damit erhalten wir die Störungsfunktion zu

$$e_i = A(1 - A)^{i-1} . \quad (2.9)$$

Diese Funktion gibt an, welchen Fehler der Glättungswert  $i$  Perioden nach Eintreten der Störung von einer Einheit noch aufweist. Die Größe des Fehlers ( $e$ ) ist also von der Wahl des  $A$  und von der Anzahl der Perioden abhängig, die seit Eintreten der Störung vergangen sind. Je größer  $A$  gewählt wird, desto stärker reagiert der Glättungswert auf den Impuls, benötigt dafür aber weniger Perioden, um wieder auf das alte Niveau zurückzufallen (vgl. Abb. 2.8).

### Reaktion auf eine Niveauänderung

Jetzt soll untersucht werden, welche Auswirkungen sich ergeben, wenn sich ab einer Periode  $t$  das Niveau des Prozesses so ändert, dass

$$x_t = x_{t-1} = x_{t-2} \text{ und } x_{t+1} = (1 + x_t) = x_{t+2} \dots$$

Für  $S_{t+1}(x)$  gilt ebenso wie im Abschnitt „Reaktion auf einen Impuls“ (vgl. (2.8)), dass

$$S_{t+1}(x) = x_t + A .$$

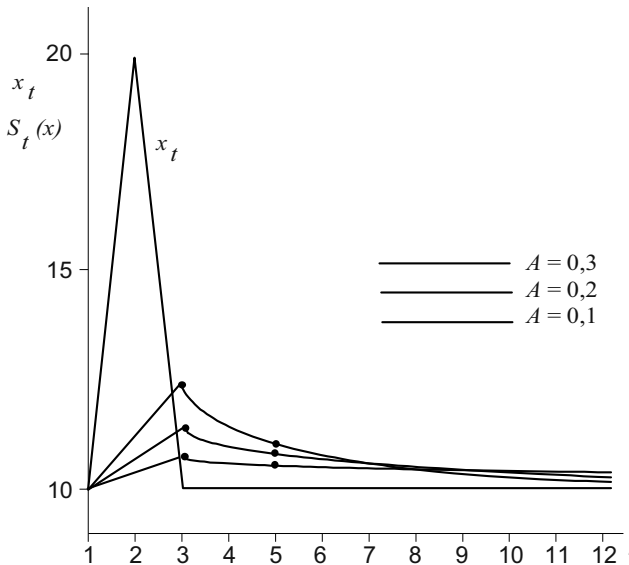
Wir erweitern zu

$$S_{t+1}(x) = (1 + x_t) - (1 - A) \quad (2.10)$$

$$S_{t+2}(x) = Ax_{t+2} + (1 - A)S_{t+1}(x) . \quad (2.11)$$

Setzt man (2.10) in (2.11) und  $x_{t+2} = 1 + x_t$  ein, so erhält man

$$\begin{aligned} S_{t+2}(x) &= A(1 + x_t) + (1 - A)[(1 + x_t) - (1 - A)] \\ &= A(1 + x_t) + (1 - A)(1 + x_t) - (1 - A)^2 \\ &= (1 + x_t) - (1 - A)^2 . \end{aligned}$$



**Abb. 2.8** Reaktion des Verfahrens der exponentiellen Glättung erster Ordnung auf einen Impuls ([12, S. 242])

Werden diese formalen Überlegungen weiter geführt, so erhält man für die  $i$ -te Periode  $S_{t+i}(x) = (1 + x_t) - (1 - A)^i$ . Die Störungsfunktion, die uns den in der Periode  $(t + i)$  noch vorhandenen Fehler angibt, ist dann

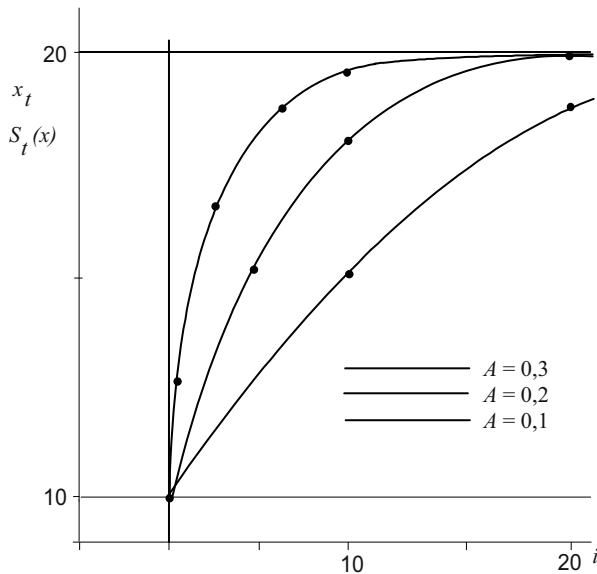
$$e_i = -(1 - A)^i. \quad (2.12)$$

Das bedeutet, dass sich im Falle einer Niveauänderung der Glättungswert umso schneller an das neue Niveau anpasst, je größer  $A$  gewählt wird (vgl. Abb. 2.9).

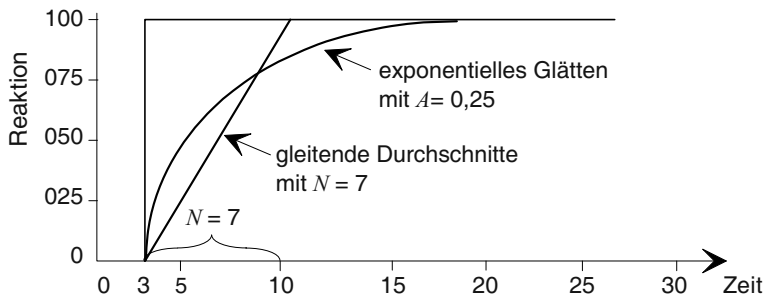
Die Abb. 2.8 und 2.9 zeigen, welche Auswirkungen plötzliche Veränderungen bei verschiedenen Werten von  $A$  haben, und das Optimierungsproblem bei der Wahl des Faktors  $A$ : Ein Modell mit großem  $A$  passt sich schneller den Niveauänderungen an, reagiert aber stark auf Zufallsabweichungen. Ein kleines  $A$  liefert hingegen stabile Vorhersagewerte, benötigt jedoch eine vergleichsweise lange Zeit, um Niveauänderungen zu verkraften.

Der durch die Störung verursachte Fehler wird theoretisch erst nach unendlich vielen Perioden null und der Glättungswert nähert sich nur asymptotisch dem Ausgangszustand bzw. dem neuen Niveau. Für  $A = 0,3$  wird dieser Zustand nach zehn Perioden annähernd erreicht.

Abbildung 2.10 zeigt einen Vergleich zwischen dem Verfahren der gleitenden Durchschnitte und dem exponentiellen Glätten. Daraus geht hervor, dass sich für  $N = 7$  der gleitende Mittelwert genau sieben Perioden nach der Niveauerhöhung an



**Abb. 2.9** Reaktion des Verfahrens des exponentiellen Glättens erster Ordnung auf eine Niveauänderung ([12, S. 243])



**Abb. 2.10** Vergleich zwischen dem Verfahren der gleitenden Durchschnitte und dem exponentiellen Glätten ([4, S. 114])

das neue Niveau angepasst hat, während sich die Kurve der exponentiell geglätteten Werte nur asymptotisch der neuen Niveau-Geraden nähert.

### Bedeutung und Vorteile des exponentiellen Glättens

Man spricht von einem Verfahren des exponentiellen Glättens erster Ordnung als einer Methode zur Abschätzung des Koeffizienten in einem konstanten Modell. Aus dieser Blickrichtung weist das exponentielle Glätten insbesondere im Vergleich mit dem Verfahren der gleitenden Durchschnitte folgende Vorteile auf:

1. Zur Berechnung unseres Glättungswertes werden nur der Glättungswert der Vorperiode, ein Wert für  $A$  und der neue Beobachtungswert  $x_t$  benötigt.
2. Es kann gezeigt werden, dass das Verfahren der exponentiellen Glättung die Abweichungssquadrate zwischen den neuen Beobachtungs- und den Prognosewerten minimiert ([4, S. 102]). Das Verfahren ist in Bezug auf dieses Kriterium optimal.
3. Wir sind in der Lage, durch die Änderung eines einzigen Parameters (des Glättungsparameters  $A$ ) die Zahl und das Gewicht der Vergangenheitsdaten zu bestimmen, die zur Bildung des Glättungswertes herangezogen werden, ohne in den Rechenformalismus eingreifen zu müssen. Es wird durch einen einzigen Parameter die Stabilität bzw. Reagibilität des Verfahrens determiniert.

Der zuletzt genannte Vorteil ist mit ein Grund dafür, im exponentiellen Glätten ein Prinzip zu sehen, auf dem eine Vielzahl anderer Verfahren basiert.

Wird der Glättungsparameter  $A$  in Abhängigkeit vom Vorhersagefehler geändert, so gelangt man zu adaptiven Verfahren, die sich automatisch an Veränderungen der Zeitreihencharakteristik anpassen können.

Eine andere Möglichkeit besteht darin, dass man das exponentielle Glätten erster Ordnung auf geglättete Daten anwendet. Ergebnis ist das exponentielle Glätten zweiter Ordnung, mit dem wir uns im Abschnitt „Exponentielles Glätten zweiter Ordnung (nach Brown)“ befassen werden.

Wenn die mit exponentiellem Glätten ermittelten Schätzwerte große Unsicherheiten aufweisen, kann es zielführend sein, die extrapolierten Werte mit subjektiven Schätzungen zu kombinieren ([1, S. 227]).

Man sieht bereits aus diesen Andeutungen, dass es eine Vielzahl von Verfahren gibt, die alle – trotz ihrer unterschiedlichen Ausprägungen – auf dem Prinzip des exponentiellen Glättens beruhen.

### **2.3.3 Methoden zur Abschätzung der beiden Koeffizienten im linearen Modell mit Trend**

Die bisher behandelten Methoden setzten voraus, dass der den Daten zugrunde liegende Prozess über der Zeit konstant verläuft. Es ist aber sowohl für die operativen Systeme (z. B. Parametrierung von Lagerhaltungssystemen) als auch für die Unternehmensstrategie sehr wichtig, zu erkennen, wenn sich im Unternehmen oder im Markt ein Trend entwickelt. Übersieht man das bzw. reagiert die Unternehmensleitung zu spät, so können die Folgen verheerend sein, wie die folgenden Beispiele zeigen:

- Der Kodak-Konzern übersah den Trend zur digitalen Fotografie und gab die Marktführerschaft an Canon ab.



- Nokia galt lange als führend bei der Ergonomie von Mobiltelefonen, erkannte aber zu spät den Trend zu berührungssensitiven Oberflächen.
- Sony und Samsung reagierten zu spät auf den Trend zu kleinen Laptops („Net-books“) und verlor deutliche Marktanteile an Asus und Acer.

Wir nehmen nun an, dass der Prozess eine linear zunehmende Funktion der Zeit ist (Modell 1-B, Abb. 2.1)  $P_t = a + b t$ .

Es sind jetzt Verfahren erforderlich, die es gestatten, die Modellkoeffizienten  $a$  und  $b$  abzuschätzen. Für die Daten gilt  $x_t = a + b t + e_t$ .

Dabei sind  $e_t$  Zufallsabweichungen, von denen wir wieder annehmen, dass sie einen Erwartungswert von null haben.

### Exponentielles Glätten mit Trendkorrektur

Wird exponentielles Glätten (erster Ordnung) auf einen linearen Prozess angewendet, so bezieht sich der Glättungswert  $S_t(x)$  auf eine frühere Periode, die unter Berücksichtigung der *unterschiedlichen* (exponentiellen) Gewichtung dem durchschnittlichen Alter der Daten entspricht. Das mittlere Alter der Daten beim exponentiellen Glätten wurde im Abschnitt „Bestimmung des Glättungsfaktors“ mit  $(1 - A)/A$  bestimmt. Ein Glättungswert, welcher in der Periode  $T$  aus *allen* Daten  $x_t$  berechnet wird, bezieht sich daher auf die Periode  $(T - (1 - A)/A)$ . Soll eine Vorhersage durchgeführt werden, so muss vorher der Glättungswert  $S_t(x)$  um das  $(1 - A)/A$ -fache des Trendanstiegs pro Periode  $\hat{b}_t$  korrigiert werden.

Bezeichnen wir nun den *korrigierten* Schätzwert für den Periodengrundwert mit  $\hat{a}_t$ , dann ist dieser mit

$$\hat{a}_t = S_t(x) + \frac{(1 - A)}{A} \hat{b}_t \quad (2.13)$$

zu berechnen.

Beim Verfahren des exponentiellen Glättens mit Trendkorrektur wird der Trendwert  $\hat{b}_t$  folgendermaßen *direkt* fortgeschrieben: Man berechnet einen *aktuellen* Trendwert  $b_t$  aus der Differenz der Glättungswerte  $S_t(x)$  und  $S_{t-1}(x)$

$$b_t = S_t(x) - S_{t-1}(x) . \quad (2.14)$$

Dieser wird mit exponentiellem Glätten (erster Ordnung) fortgeschrieben. (Zur Kennzeichnung des geglätteten Trendwerts wird  $S_t(b)$  verwendet.)

$$S_t(b) = S_{t-1}(b) + A(b_t - S_{t-1}(b)) . \quad (2.15)$$

Der Glättungswert  $S_t(b)$  wird zur Abschätzung von  $\hat{b}_t$  herangezogen

$$\hat{b}_t = S_t(b) . \quad (2.16)$$

Zur Berechnung benötigt man also die Formeln (2.13)–(2.16):

$$\begin{aligned} b_t &= S_t(x) - S_{t-1}(x) , && \text{(aktueller Trendwert)} \\ S_t(b) &= S_{t-1}(b) + A(b_t - S_{t-1}(b)) , && \text{(fortgeschriebener Trendwert)} \\ \hat{b}_t &= S_t(b) , \\ \hat{a}_t &= S_t(x) + \frac{(1-A)}{A} \hat{b}_t . && \text{(korrigierter Grundwert)} \end{aligned}$$

Die Vorhersage in einer Periode ( $T$ ) für eine Periode ( $T + i$ ) ist dann  $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$ .

### Exponentielles Glätten zweiter Ordnung (nach Brown)

Dieses Verfahren (Double Exponential Smoothing) ist sehr bekannt und verbreitet. Der Zusatz „nach Brown“ ist deswegen erforderlich, weil es auch noch andere Verfahren des exponentiellen Glättens für das lineare Modell gibt (vgl. Abschnitt „Zwei-Parameter-Modell nach Holt“ und folgende.).

Auch hier bestehen grundsätzlich zwei Möglichkeiten der Berechnung:

1. Verwendung von Glättungswerten erster und zweiter Ordnung
2. Direkte Fortschreibung der Modellkoeffizienten

Die Ergebnisse sind in jedem Fall gleich, der Unterschied ist nur rechentechnischer Natur (Beweis siehe Abschnitt „Vergleich zwischen exponentiellem Glätten mit Trendkorrektur und exponentiellem Glätten zweiter Ordnung (nach Brown)“).

### Verwendung von Glättungswerten erster und zweiter Ordnung

Der einfache Glättungswert  $S_t(x)$  (im Folgenden mit  $S_t^1(x)$  bezeichnet) errechnet sich aus Gleichung (2.6):

$$S_t^1(x) = Ax_t + (1 - A)S_{t-1}^1(x)$$

Wir fassen nun die errechneten Glättungswerte  $S_t^1(x)$  als Beobachtungswerte auf und schreiben sie mit exponentiellem Glätten erster Ordnung fort. Damit erhalten wir einen Glättungswert zweiter Ordnung, der einen Glättungswert über die Glättungswerte darstellt (im Folgenden mit  $S_t^2(x)$  bezeichnet).

$$S_t^2(x) = AS_t^1(x) + (1 - A)S_{t-1}^2(x) \quad (2.17)$$

oder

$$S_t^2(x) = S_{t-1}^2(x) + A(S_t^1(x) - S_{t-1}^2(x)) . \quad (2.18)$$

Nimmt man an, dass unsere Daten frei von Zufallsabweichungen sind, so liegen die Glättungswerte zweiter Ordnung auf einer Geraden parallel zu den Glättungswerten erster Ordnung bzw. zu den Beobachtungswerten. Der horizontale Abstand der Geraden zueinander ist konstant und beträgt  $(1 - A)/A$ , der vertikale Abstand ist daher durch  $b_t(1 - A)/A$  gegeben:

$$x_t - S_t^1(x) = S_t^1(x) - S_t^2(x) = b_t(1 - A)/A \quad (2.19)$$

Sind die Daten mit Zufallsabweichungen behaftet, so können mithilfe der Gleichung (2.19) Schätzwerte für den korrigierten Periodengrundwert und den Trendwert bestimmt werden:

$$\hat{a}_t = S_t^1(x) + S_t^1(x) - S_t^2(x) = 2S_t^1(x) - S_t^2(x) \quad (2.20)$$

Dabei ist  $\hat{a}_t$  ein Periodengrundwert, der bereits um den durch Trend verursachten „Time-Lag“ korrigiert wurde ( $\hat{a}_t = \hat{x}_t$ ).

Da aber auch gilt, dass

$$S_t^1(x) - S_t^2(x) = b_t(1 - A)/A ,$$

erhält man einen Schätzwert für den Trendanstieg pro Periode durch

$$\hat{b}_t = A/(1 - A) (S_t^1(x) - S_t^2(x)) . \quad (2.21)$$

Soll nun in einer bestimmten Periode  $T$  eine Vorhersage für  $(T + i)$  Perioden durchgeführt werden, so lautet die Vorhersagegleichung  $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$ . Wir können also durch die Bildung von Glättungswerten erster und zweiter Ordnung Schätzwerte für die Koeffizienten  $a$  und  $b$  in einem linearen Modell bestimmen. Zu Beginn der Glättung sind die Glättungswerte der Vorperiode  $S_0^1(x)$  und  $S_0^2(x)$  erforderlich. Bezieht man die Formeln (2.20) und (2.21) auf die Periode  $t = 0$ , so erhält man

$$\hat{a}_0 = 2S_0^1(x) - S_0^2(x) \quad (2.22)$$

und

$$\hat{b}_0 = A/(1 - A) (S_0^1(x) - S_0^2(x)) . \quad (2.23)$$

Auflösen dieser beiden Gleichungen nach  $S_0^1(x)$  bzw.  $S_0^2(x)$  ergibt mit  $B = 1 - A$

$$S_0^1(x) = \hat{a}_0 - \hat{b}_0 B/A \quad (2.24)$$

$$S_0^2(x) = \hat{a}_0 - 2\hat{b}_0 B/A \quad (2.25)$$

Liegen repräsentative Vergangenheitsdaten vor, so kann man die Werte  $a_0$  und  $b_0$  z. B. mithilfe einer Regressionsrechnung bestimmen. Ist dies nicht der Fall, so muss versucht werden, durch den Vergleich mit ähnlichen Zeitreihen (z. B. von ähnlichen Produkten) möglichst gute Schätzungen für  $\hat{a}_0$  und  $\hat{b}_0$  zu erhalten. Sind die Schätzwerte sehr unsicher, so wird man gegebenenfalls zu Beginn mit einem größeren  $A$  rechnen, damit sich ein Fehler bei der Schätzung der Anfangswerte nicht lange auswirkt, und nach einigen Perioden das  $A$  wieder verkleinern.

Sind nun die Werte für  $\hat{a}_0$  und  $\hat{b}_0$  bestimmt, so können mit (2.24) und (2.25) die Glättungswerte für die Periode  $t = 0$  berechnet werden.

Die Glättungswerte für die Periode  $t = 1$  sind gemäß Gleichung (2.6) und (2.17):

$$\begin{aligned} S_1^1(x) &= Ax_1 + BS_0^1(x) \quad \text{und} \\ S_1^2(x) &= AS_1^1(x) + BS_0^2(x) \end{aligned}$$

Die gesuchten Modellkoeffizienten für  $t = 1$  findet man durch Einsetzen in die Gleichungen (2.20) und (2.21):

$$\begin{aligned} \hat{a}_1 &= 2S_1^1(x) - S_1^2(x) \\ \hat{b}_1 &= A/B (S_1^1(x) - S_1^2(x)) \end{aligned}$$

Mithilfe der Vorhersagegleichung  $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$  kann nun für eine beliebige Zahl von Perioden prognostiziert werden. Wird jeweils nur für eine Periode vorhergesagt ( $\hat{x}_{T+1}$ ), so kommt man durch Einsetzen von (2.20) und (2.21) in die Vorhersagegleichung unter Berücksichtigung der Beziehung  $B = 1 - A$  zu folgenden Gleichungen:

$$\begin{aligned} \hat{x}_{T+1} &= 2S_T^1(x) - S_T^2(x) + \frac{A}{1-A} (S_T^1(x) - S_T^2(x)) \\ \hat{x}_{T+1} &= S_T^1(x) + (S_T^1(x) - S_T^2(x)) + \frac{A}{1-A} (S_T^1(x) - S_T^2(x)) \\ \hat{x}_{T+1} &= S_T^1(x) + (S_T^1(x) - S_T^2(x)) \left(1 + \frac{A}{1-A}\right) \\ \hat{x}_{T+1} &= S_T^1(x) + (S_T^1(x) - S_T^2(x)) \frac{1}{B} \end{aligned} \tag{2.26}$$

### Direkte Fortschreibung der Modellkoeffizienten

Bei der oben (Abschnitt „Verwendung von Glättungswerten erster und zweiter Ordnung“) behandelten Methode ist die erforderliche Trendkorrektur durch die Differenz der Glättungswerte erster und zweiter Ordnung eingeführt worden. Um den korrigierten Periodengrundwert  $\hat{a}_t$  zu erhalten, wurde  $S_t^1(x)$  um den Betrag  $(S_t^1(x) - S_t^2(x))$  vermehrt. Abbildung 2.11 verdeutlicht diesen Sachverhalt.

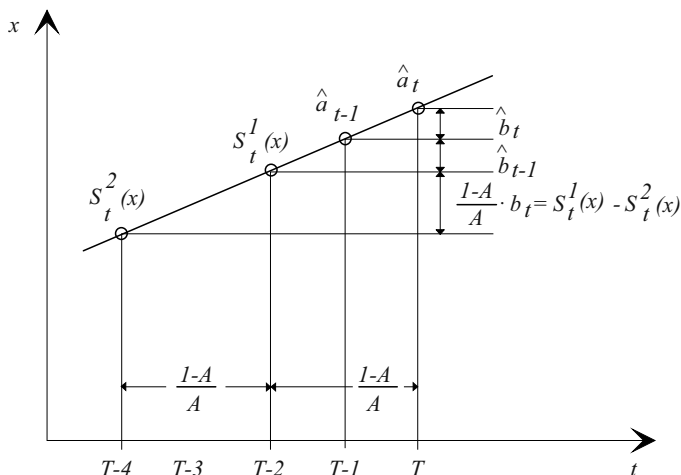


Abb. 2.11 Möglichkeiten der Trendkorrektur

Beim Verfahren des exponentiellen Glättens mit Trendkorrektur wurde der Glättungsfaktor erster Ordnung um den Betrag  $\frac{1-A}{A}b_t$  korrigiert, der Trendanstieg pro Periode  $b_t$  wurde getrennt bestimmt bzw. fortgeschrieben (vgl. Abschnitt 2.3.1). Aus Abb. 2.11 geht hervor, dass als dritte Möglichkeit zur Trendkorrektur die Beziehung

$$\hat{a}_t = \hat{a}_{t-1} + \hat{b}_{t-1} \quad (2.27)$$

verwendet werden kann.

Diese Methode hat den Vorteil, dass keine Glättungswerte zu berechnen sind und die Modellkoeffizienten direkt fortgeschrieben werden können. Zu diesem Zweck wird die Trendgerade gedanklich in einen Trendwert und in einen Grundwert zerlegt, die über der Zeit als annähernd konstant angesehen werden. Da jetzt zwei konstante Prozesse vorliegen, können die Koeffizienten mit exponentiellem Glätten erster Ordnung fortgeschrieben werden, ohne dass eine Trendkorrektur durchzuführen ist.

Die Umformung der Formeln des exponentiellen Glättens zweiter Ordnung zu Formeln für direkte Fortschreibung der Modellkoeffizienten soll am Beispiel des konstanten Modells gezeigt werden.

Als Schätzwert für  $\hat{a}_t$  wird der Glättungswert erster Ordnung verwendet:

$$\hat{a}_t = S_t^1(x) = Ax_t + (1-A)S_{t-1}^1(x) \quad (2.28)$$

Eine Vorhersage für die nächste Periode ist im konstanten Modell durch  $\hat{x}_{t+1} = \hat{a}_t = S_t^1(x)$  gegeben.

Es muss aber auch  $\hat{x}_t = \hat{a}_{t-1} = S_{t-1}^1(x)$  gelten. Setzt man nun in der Gleichung (2.28) für  $S_{t-1}^1(x)$  den Wert  $\hat{a}_{t-1}$  ein, so erhält man nach Umformen

$$\hat{a}_t = \hat{a}_{t-1} + A(x_t - \hat{a}_{t-1}) \quad (2.29)$$

als Formel für die direkte Fortschreibung des Koeffizienten im konstanten Modell.

Die Ableitung der Formel (2.29) aus der Formel (2.28) ist beim konstanten Modell trivial. Für das lineare Modell ist diese Vorgehensweise nicht so einfach. Wir geben hier nur die grundlegenden Rechenschritte und das Ergebnis wieder.

Man geht von den Formeln

$$\begin{aligned} \hat{a}_t &= 2S_t^1(x) - S_t^2(x) \text{ und} \\ \hat{b}_t &= A/B (S_t^1(x) - S_t^2(x)) \end{aligned}$$

aus (vgl. (2.20) und (2.21)), in die man für

$$\begin{aligned} S_t^1(x) &\text{ und } S_t^2(x) \\ S_t^1(x) &= Ax_t + (1-A)S_{t-1}^1(x) \text{ und} \\ S_t^2(x) &= AS_t^1(x) + (1-A)S_{t-1}^2(x) \end{aligned}$$

einsetzt (vgl. (2.6) und (2.17)).

In den daraus resultierenden Ausdrücken für  $\hat{a}_t$  und  $\hat{b}_t$  werden die Glättungswerte durch

$$\begin{aligned} S_t^1(x) &= \hat{a}_t - \hat{b}_t B/A \text{ und} \\ S_t^2(x) &= \hat{a}_t - 2\hat{b}_t B/A \end{aligned}$$

ersetzt (vgl. (2.24) und (2.25)). (Diese Formeln gewinnt man durch Auflösen der Gleichungen (2.20) und (2.21) nach  $S_t^1(x)$  bzw.  $S_t^2(x)$ .)

Nach verschiedenen Umformungen erhält man als Resultate:

$$\hat{a}_t = \hat{x}_t + (1 - B^2)(x_t - \hat{x}_t) \quad (2.30)$$

$$\hat{b}_t = \hat{b}_{t-1} + (1 - B)^2(x_t - \hat{x}_t) \quad (2.31)$$

In diesen Formeln ist  $\hat{x}_t$  der um den Trend korrigierte Periodengrundwert ( $\hat{x}_t = \hat{a}_{t-1} + \hat{b}_{t-1}$ ), der als „alter Glättungswert“ in die Formeln eingeht, hingegen stellt  $\hat{a}_t$  den fortgeschriebenen „neuen Glättungswert“ dar, welcher unter Berücksichtigung der Differenz zwischen neuem Beobachtungswert  $x_t$  und dem „alten Glättungswert“  $\hat{x}_t$  mit exponentiellem Glätten erster Ordnung bestimmt wird ([4, S. 140]).

Eine Vorhersage für die Periode  $(T + i)$  ist durch  $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$  gegeben.

Es ist aber zu beachten, dass es sich hier lediglich um eine andere Berechnungsmethode handelt, die dieselben Werte wie die im vorhergegangenen Abschnitt behandelte Methode ergibt. In dieser Form kann aber das Verfahren von Brown mit anderen besser verglichen werden.

### Vergleich zwischen exponentiellem Glätten mit Trendkorrektur und exponentiellem Glätten zweiter Ordnung (nach Brown)

Diese beiden Verfahren unterscheiden sich lediglich in der Organisation des Rechenvorgangs, führen aber zu den gleichen Ergebnissen. Der korrigierte Periodengrundwert ist beim exponentiellen Glätten mit Trendkorrektur (2.13):

$$\hat{a}_t = S_t(x) + B/A\hat{b}_t$$

Der Trendwert  $b_t$  wird direkt mit der Formel des exponentiellen Glättens erster Ordnung

$$\hat{b}_t = \hat{b}_{t-1} + A(b_t - \hat{b}_{t-1}) = S_t(b)$$

und  $\hat{b}_{t-1} = S_{t-1}(b)$  fortgeschrieben, wobei der jeweils neue „Beobachtungswert“  $b_t$  durch

$$b_t = S_t^1(x) - S_{t-1}^1(x)$$

gegeben ist (vgl. (2.14)). Da man den Periodengrundwert mithilfe der Glättungswerte bestimmt, werden beide zuvor beschriebenen Berechnungsformen (Abschnitte „Verwendung von Glättungswerten erster und zweiter Ordnung“ und „Direkte Fortschreibung der Modellkoeffizienten“) verwendet.

Die entsprechenden Formeln für exponentielles Glätten zweiter Ordnung sind:

$$\begin{aligned}\hat{a}_t &= \hat{x}_t + (1 - B^2)(x_t - \hat{x}_t) \\ \hat{b}_t &= \hat{b}_{t-1} + (1 - B)^2(x_t - \hat{x}_t) \\ (\hat{x} &= \hat{a}_{t-1} + \hat{b}_{t-1})\end{aligned}$$

Durch eine Reihe von Umformungen kann gezeigt werden, dass die Formeln (2.13) und (2.15) mit den Formeln (2.30) und (2.31) identisch sind ([17, S. 53 ff.]).

### Zwei-Parameter-Modell nach Holt

Bei allen bisher besprochenen Verfahren des exponentiellen Glättens wurde ein einziger Glättungsparameter ( $A$  bzw. für  $(1 - A) = B$ ) verwendet. Für das lineare Modell bedeutet das, dass sowohl Periodengrundwert als auch Trendwert mit dem gleichen Parameter geglättet wurden. Wir wollen nun Verfahren behandeln, die zwei (bzw. drei) verschiedene Glättungsparameter benutzen.

Holt war der erste Autor, der exponentielles Glätten für die kurzfristige Vorhersage verwendet hat ([10, S. 104]). Im Gegensatz zu dem Verfahren von Brown werden hier Periodengrundwert und Trendwert mit zwei verschiedenen Glättungsparametern fortgeschrieben. Die Grundformeln lauten:

$$\hat{a}_t = Cx_t + (1 - C)(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (2.32)$$

$$\hat{b}_t = D(\hat{a}_t - \hat{a}_{t-1}) + (1 - D)\hat{b}_{t-1} \quad (2.33)$$

$C$  und  $D$  sind die Glättungsparameter, für die gilt  $0 < C < 1$  und  $0 < D < 1$ . Auch hier ist die Vorhersagegleichung für eine Periode in die Zukunft  $\hat{x}_{t+1} = \hat{a}_t + \hat{b}_t$ , sodass auch gelten muss  $\hat{x}_t = \hat{a}_{t-1} + \hat{b}_{t-1}$ .

Um diese Formeln mit den Formeln des Ein-Parameter-Modells von Brown (2.30) vergleichen zu können, führen wir folgende Umformung durch: Wir setzen  $(\hat{a}_{t-1} + \hat{b}_{t-1}) = \hat{x}_t$  in (2.32) ein und formen um zu

$$\hat{a}_t = \hat{x}_t + C(x_t - \hat{x}_t) \quad (2.34)$$

bzw. mit  $\hat{x}_t = \hat{a}_{t-1} + \hat{b}_{t-1}$

$$\hat{a}_t - \hat{a}_{t-1} = \hat{b}_{t-1} + C(x_t - \hat{x}_t) \quad (2.35)$$

Der Ausdruck (2.35) wird in (2.33) eingesetzt und man erhält für  $\hat{b}_t$

$$\hat{b}_t = \hat{b}_{t-1} + E(x_t - \hat{x}_t); \quad (E = CD) \quad (2.36)$$

Wir stellen nun die Formeln des Holt'schen Zwei-Parameter-Modells (vgl. (2.34) und (2.36))

$$\begin{aligned} \hat{a}_t &= \hat{x}_t + C(x_t - \hat{x}_t) \\ \hat{b}_t &= \hat{b}_{t-1} + E(x_t - \hat{x}_t) \end{aligned}$$

den Formeln von Brown für das Ein-Parameter-Modell (vgl. (2.30) und (2.31))

$$\begin{aligned} \hat{a}_t &= \hat{x}_t + (1 - B^2)(x_t - \hat{x}_t) \\ \hat{b}_t &= \hat{b}_{t-1} + (1 - B)^2(x_t - \hat{x}_t) \end{aligned}$$

gegenüber.

Daraus ist nun ersichtlich, dass das Brown'sche Modell ein Sonderfall des Zwei-Parameter-Modells von Holt ist, da es die beiden Parameter  $C$  und  $D$  so beschränkt, dass

$$C = (1 - B^2) \quad \text{und} \quad E = CD = (1 - B)^2 \quad \text{bzw.} \quad D = \frac{(1 - B)^2}{C} \quad \text{ist.}$$



### Zwei-Parameter-Modell mit gedämpftem Trend

Werden die Trendkomponenten der Gleichungen (2.32) und (2.33) mit einem zusätzlichen Faktor  $\Phi$ ,  $0 \leq \Phi \leq 1$ , multipliziert, so erhält man einen abgeschwächten Trend [6]:

$$\hat{a}_t = Cx_t + (1 - C)(\hat{a}_{t-1} + \Phi\hat{b}_{t-1}) \quad (2.37)$$

$$\hat{b}_t = D(\hat{a}_t - \hat{a}_{t-1}) + (1 - D)\Phi\hat{b}_{t-1} \quad (2.38)$$

Die Ermittlung des Prognosewertes unterscheidet sich dann von der bisherigen ( $\hat{x}_{T+i} = \hat{a}_T + \hat{b}_T i$ ) insofern, als man im zweiten Summanden über Dämpfungsfaktoren, die für die betrachteten Zukunftsperioden unterschiedlich gewählt werden können, summieren muss:

$$\hat{x}_{T+i} = \hat{a}_T + \sum_{j=1}^i \Phi_j \hat{b}_T \quad (2.39)$$

Für  $\Phi = 1$  resultiert als Spezialfall das ursprüngliche Holt-Modell, für  $\Phi = 0$  das exponentielle Glättungsmodell erster Ordnung.

Das Verfahren mit gedämpftem Trend stellt eine beachtenswerte Weiterentwicklung dar, da sich auf diese Weise Prozesse, deren Werte in einer „Grauzone“ zwischen „mit und ohne Trend“ variieren, besser modellieren lassen. Ein Beispiel zum praktischen Einsatz (Prognose der Personenbeförderung bei der Schweizerischen Bundesbahn für die Jahre 1987 bis 1996) findet sich in ([19, S. 48]).

### Drei-Parameter-Modell mit Fehlerdifferenzausdruck

Zum Abschluss wollen wir noch ein Drei-Parameter-Modell zeigen: Wir erweitern das Holt'sche Modell durch die Einführung eines Fehlerdifferenzausdrucks: Der Vorhersagefehler  $e_t$  ist die Differenz zwischen dem Beobachtungswert der Periode  $t$  und dem Vorhersagewert für die gleiche Periode  $t$ , der in  $(t - 1)$  errechnet wurde, und er entspricht damit dem Ausdruck  $(x_t - \hat{x}_t)$ , der in den Gleichungen bisher erscheint ( $x_t$  = Beobachtungswert,  $\hat{x}_t = \hat{a}_{t-1} + \hat{b}_{t-1}$ ). Die Fehlerdifferenz ist  $(e_t - e_{t-1})$  und wird mit einem dritten Parameter geglättet. Man erhält folgende Gleichungen (vgl. dazu (2.32) und (2.33)):

$$\hat{a}_t = Fx_t + (1 - F)\hat{x}_t + H(e_t - e_{t-1}) \quad (2.40)$$

$$\hat{b}_t = G(\hat{a}_t - \hat{a}_{t-1}) + (1 - G)\hat{b}_{t-1} \quad (2.41)$$

$F, G, H$  sind Glättungsparameter.

## 2.4 Grenzen und Ausblick

Wir sind in diesem Beitrag von einfachen Verfahren der gleitenden Durchschnitte ausgegangen und haben daraus das Prinzip des exponentiellen Glättens abgeleitet. Dann haben wir Verfahren gezeigt und verglichen, die auf dem exponentiellen Glätten beruhen und zur Abschätzung der Koeffizienten in Trendmodellen verwendet werden können, bis hin zu Mehr-Parameter-Modellen.

Betrachtet man die zunehmende Komplexität der Verfahren, so stellt sich die Frage, ob der zusätzliche Aufwand für die mathematisch-statistischen Verfahren die Genauigkeit der Vorhersagen erhöhen kann. In einer Veröffentlichung aus dem Jahre 2001 hat Armstrong 32 Anwendungsberichte ausgewertet und darunter nur fünf gefunden, in denen komplexere Methoden zu einer Verbesserung der Vorhersagen geführt haben ([1, S. 227 ff.]). Ist damit eine Grenze für die Anwendbarkeit von exponentiellem Glätten für kurzfristige Zeitreihenprognosen erreicht worden? Hyndman verneint das und zeigt, dass sich durch die Verbindung von exponentiellem Glätten mit Zustandsraummodellen in einem „Statistical Framework“ ganz entscheidende Vorteile bei der Auswahl der Modelle, bei der Bestimmung der Initialwerte und der Glättungsparameter in einem automatisiertem Umfeld erzielen lassen ([11] und Kapitel 18 in diesem Sammelband).

---

## Literaturverzeichnis

- [1] Armstrong, J.S. (Hrsg.): Principles of Forecasting: A Handbook for Researchers and Practitioners, Boston (2001)
- [2] Armstrong, J.S.: Long-Range Forecasting: From Crystal Ball to Computer, 2. Aufl., New York (1985)
- [3] Box, G.E.P. und Jenkins, G.M.: Time Series Analysis, Forecasting and Control, San Francisco (1970)
- [4] Brown, R.G.: Smoothing, Forecasting and Prediction of Discrete Time Series, Englewood Cliffs (1963)
- [5] Dickersbach, J.T.: Characteristic Based Planning with mySAP SCM, Berlin, Heidelberg (2005)
- [6] Gardner Jr., E.S.: Exponential Smoothing: The State of the Art, J. Forecast. **4**, 1 ff. (1985)
- [7] Gardner Jr., E.S. und McKenzie, E.: Model Identification in Exponential Smoothing, J. Oper. Res. Soc. **39**, 863 ff. (1988)
- [8] Harrison, P.J.: Short Term Sales Forecasting, Appl. Stat. J. Roy. St. C. **14**, 102 ff. (1965)
- [9] Harrison, P.J.: Exponential Smoothing and Short-Term Sales Forecasting, Manage. Sci. **13**, 821 ff. (1967)
- [10] Holt, C.C., Modigliani, F., Muth, J.F. und Simon, H.A.: Planning Production, Inventories and Work Force, Englewood Cliffs (1960)
- [11] Hyndman, R.J., Koehler A.B., Ord, J.K. und Snyder, D.: Forecasting with Exponential Smoothing, Berlin, Heidelberg (2008)
- [12] Lewandowski, R.: Modelle und Methoden der ökonomischen Vorhersage, Elektronische Datenverarbeitung **11**, 235 ff. (1969)

- [13] Makridakis, S., Wheelwright, S.C. und Hyndman, R.J.: Forecasting, Methods and Applications, 3. Aufl., Indianapolis (1998)
- [14] Müller-Merbach, H.: Operations Research, 3. Aufl., München (1973)
- [15] Pegels, C.: Exponential Forecasting: Some New Variations, Manage. Sci. **15**, 311 ff. (1969)
- [16] Reif, K.: Bedarfsvorhersage mittels mathematisch-statistischer Verfahren, IBM Form 81518 (1966)
- [17] Schröder, M.: Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren, in: Mertens, P. (Hrsg.), Prognoserechnung, 2. Aufl., Würzburg, S. 53 ff. (1975)
- [18] Wiese, K.H.: Exponential Smoothing – eine Methode der statistischen Bedarfsvorhersage, IBM Form 78129 (1964)
- [19] Weber, K.: Prognosemethoden und -Software, Idstein (1991)
- [20] Winters, P.R.: Forecasting Sales by Exponentially Weighted Moving Averages, Manage. Sci. **6**, 324 ff. (1960)

Prognoserechnung

Mertens, P.; Rässler, S. (Hrsg.)

2012, XIX, 665 S. 156 Abb., Hardcover

ISBN: 978-3-7908-2796-5

A product of Physica-Verlag Heidelberg