

Inhaltsverzeichnis

2.1 Einleitung . . . . . 27

2.2 Lineare Regression . . . . . 31

    2.2.1 Einfache Regression . . . . . 31

    2.2.2 Mehrfache Regression . . . . . 34

2.3 Exponentielles Glätten . . . . . 35

    2.3.1 Einfaches exponentielles Glätten . . . . . 35

    2.3.2 Das Holt-Winters-Verfahren . . . . . 39

2.4 Abschließende Bemerkungen . . . . . 46

Literatur . . . . . 47

2.1 Einleitung

Der Aufwand der Einarbeitung in statistische Softwarepakete stellt eine nicht zu unterschätzende Hürde vor der praktischen Anwendung von quantitativen Prognoseverfahren dar, mit denen zum Beispiel begründete Vorhersagen für zukünftige Auftragseingänge erstellt werden können. In diesem Beitrag wird die Anwendung der Open Source Software gretl (für „Gnu Regression, Econometrics and Time-series Library“) auf einige grundlegende deskriptive Prognoseverfahren dargestellt: die lineare Regression, das exponentielle Glätten und das Holt-Winters-Verfahren der exponentiellen Glättung von Zeitreihen mit

Prof. Dr. Thomas Christiaans ✉  
FOM Hochschule für Oekonomie & Management gGmbH, Birlenbacher Straße 18,  
57078 Siegen, Deutschland  
e-mail: thomas.christiaans@fom.de

Trend- und Saisonkomponente. Unter anderem zeichnen die folgenden Eigenschaften gretl aus:

1. Die Software ist kostenlos unter der GNU General Public License verfügbar.
2. Sie ist einfach zu installieren und dank einer menügesteuerten Benutzeroberfläche auch einfach zu bedienen, wodurch die Einstiegshürde niedriger gelegt wird. Fortgeschrittene Anwender können jedoch auch eine Konsole oder Skripte verwenden.
3. Die Rechengenauigkeit ist sehr hoch und dank des Open-Source-Konzepts ist die Software, die im Bereich der Ökonometrie und Zeitreihenanalyse das „Aushängeschild“ unter den Open-Source-Programmen darstellt, keine Black Box (vgl. Yalta und Yalta 2010).
4. Allerdings ist ihr Anwendungsbereich durch die Fokussierung auf zeitreihenanalytische und ökonometrische Methoden weniger umfassend als derjenige des ebenfalls freien statistischen Programmpaketes R, dessen Anwendung auf Prognosetechniken im Beitrag „Szenarioanalyse als Prognoseinstrument mit einem Beispiel zur Kundenbindung“ in diesem Band dargestellt wird.

Sie erhalten die aktuelle Version von gretl für verschiedene Betriebssysteme im Internet kostenlos auf [gretl.sourceforge.net](http://gretl.sourceforge.net). Für dieses Kapitel ist die Version 1.9.12 für Windows verwendet worden. Ein Handbuch von den gretl-Entwicklern liegt der Software als pdf-Datei bei (Cottrell und Lucchetti 2012).

Die Prognoserechnung selbst kann man grob einteilen in zeitreihenanalytische Verfahren, bei denen aus dem vergangenen zeitlichen Verlauf einer Variablen auf ihren zukünftigen Verlauf geschlossen wird, und ökonometrische Verfahren, bei denen der Wert einer Variablen in Abhängigkeit von den Werten einer oder mehrerer anderer Variablen prognostiziert wird. Im Fall einer Punktprognose ist das Ergebnis der Prognose eine Zahl, im Fall einer Intervallprognose ist es ein sogenanntes „Konfidenzintervall“. Ein Prognoseverfahren heißt „deskriptiv“, wenn es ohne explizite Betrachtung stochastischer Zusammenhänge durchgeführt wird. Das bedeutet, dass die Methoden der induktiven Statistik, mit denen zum Beispiel Wahrscheinlichkeiten dafür angegeben werden können, dass ein Prognoseintervall das wahre Ergebnis enthält, nicht verwendet werden, sodass nur Punktprognosen möglich sind. Die theoretische Untermauerung der Prognoseergebnisse ist insofern eingeschränkt. Auf der Habenseite steht jedoch die relativ einfache Anwendung der deskriptiven Methoden. Die Methoden der induktiven Statistik werden daher in diesem Beitrag bewusst vernachlässigt.

Eine Zeitreihe ist eine zeitlich geordnete Folge von Beobachtungen eines statistischen Merkmals, zum Beispiel der Auftragseingänge eines Unternehmens. Für jeden Zeitpunkt oder Zeitraum  $t$  aus einer Menge von Beobachtungszeitpunkten oder -räumen liegt dabei eine Beobachtung  $y_t$  des Merkmals vor. Dabei handelt es sich grundsätzlich um ein kardinalskaliertes Merkmal, bei dem die Abstände zwischen den einzelnen Zeitreihenwerten interpretiert und Rechenoperationen wie die Bildung von Durchschnittswerten sinnvoll angewendet werden können. Handelt es sich bei dem Merkmal um eine zeitpunktbezogene Bestandsgröße (zum Beispiel Lagerbestände), so sind die Werte  $t$  als Zeitpunkte,

**Tab. 2.1** Monatliche Auftragseingänge im Wohnungsbau in Deutschland (Wertindizes)

Jahr	Jan.	Feb.	Mär.	Apr.	Mai.	Jun.	Jul.	Aug.	Sep.	Okt.	Nov.	Dez.
2000	151,9	159,7	234,0	193,5	195,3	205,4	180,8	173,2	177,8	160,2	138,7	136,7
2001	128,0	131,1	181,1	157,2	169,3	190,5	141,8	143,4	153,7	135,7	108,7	106,0
2002	103,1	109,6	146,2	144,0	136,9	148,2	123,1	124,4	135,2	118,6	120,3	123,1
2003	87,8	96,1	136,6	122,9	117,3	131,8	126,3	117,0	135,0	124,2	108,4	117,2
2004	72,3	102,1	126,1	110,7	116,0	129,9	111,3	106,5	125,9	103,1	90,6	98,5
2005	69,3	79,4	112,8	100,0	108,9	119,1	102,6	101,7	112,7	98,8	90,3	104,2
2006	77,1	82,2	112,1	109,7	114,2	124,5	119,1	106,2	116,3	101,3	89,4	98,8
2007	76,5	80,0	106,2	107,8	107,0	116,7	105,3	103,6	108,3	100,0	80,4	90,7
2008	73,3	81,3	103,5	94,5	96,7	111,4	104,8	92,5	114,1	94,4	87,9	79,5
2009	57,3	69,8	96,0	105,4	95,7	108,2	108,7	96,3	110,5	103,7	89,8	89,1
2010	55,1	76,0	116,2	113,4	107,9	111,4	113,6	112,7	130,9	106,5	96,6	101,3
2011	77,0	97,9	144,1	134,6	152,3	127,7	132,2	131,3	128,7	124,3	123,7	127,7
2012	103,8	117,6	153,4	145,2	134,8	167,8	142,9	143,3	140,4	142,0	127,2	121,6
2013	106,9	111,6	152,5	152,9	146,1	187,1	165,8					

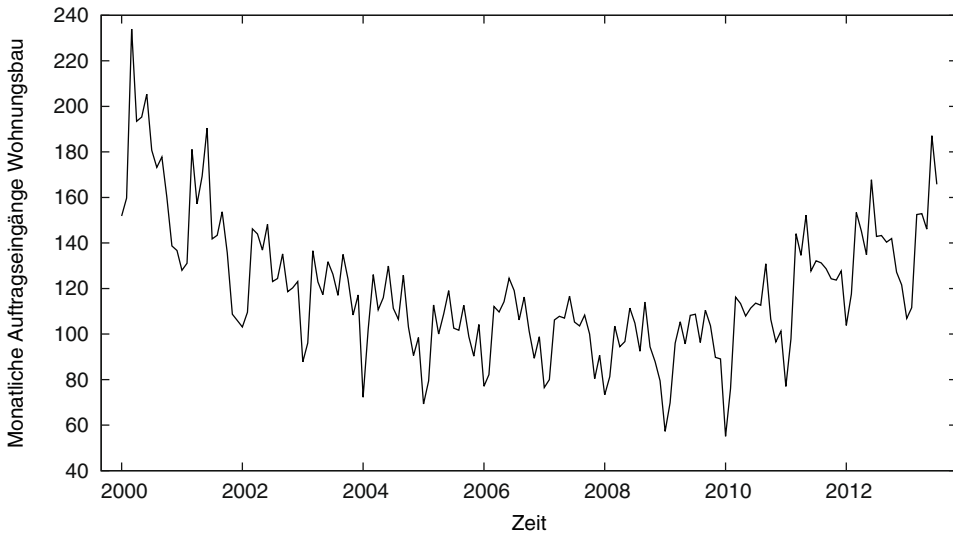
Quelle: Statistisches Bundesamt, Bauhauptgewerbe Lange Reihen Auftragseingang, Wiesbaden, Juli 2013.

handelt es sich um eine zeitraumbezogene Stromgröße (zum Beispiel Auftragseingänge), so sind sie als Zeiträume beziehungsweise Perioden zu interpretieren. Im Folgenden handelt es sich bei allen Beispielen um Stromgrößen.

Die Beobachtungszeiträume werden in der theoretischen Darstellung einfach durchnummeriert, zum Beispiel  $t = 1, \dots, T$ , wobei dann  $T$  der letzte Zeitraum der Reihe ist und vorausgesetzt wird, dass alle Perioden gleich lang sind. In der Praxis steht jeder dieser Werte für ein reales Datum, zum Beispiel  $t=1$  für Januar 2000,  $t=2$  für Februar 2000,  $t=13$  für Januar 2001 usw. Die Tab. 2.1 enthält reale Monatsdaten der Auftragseingänge im Wohnungsbau als Wertindizes<sup>1</sup> (Basisjahr 2005) für den Zeitraum von Januar 2000 bis Juli 2013. Für dieses Beispiel würde der letzte Zeitraum  $t = T = 163$  für Juli 2013 stehen, da insgesamt 163 Zeitreihenwerte vorliegen.

Einen wichtigen ersten Eindruck von einer Zeitreihe erhält man durch ihre grafische Darstellung beziehungsweise das sogenannte Zeitreihenpolygon, das durch Verbinden der einzelnen Zeitreihenpunkte mit Geradenstücken entsteht und das für die Daten der Tab. 2.1 in der Abb. 2.1 dargestellt wird. Man erkennt hier gut, dass diese Zeitreihe offenbar saisonale Schwankungen enthält, also etwa systematisch höhere Werte in den Sommer-

<sup>1</sup> Aus Platzgründen können Indizes hier nicht weiter erläutert werden, vgl. dazu ein beliebiges Lehrbuch der Statistik, zum Beispiel Lübke und Vogt (2015). Für die folgende Darstellung der Prognoseverfahren können Sie sich einfach vorstellen, dass der Wert der Auftragseingänge im April 2005 auf 100 Prozent normiert wurde. Im Juli 2013 ist der Wert dann gegenüber April 2005 um 65,8 Prozent gestiegen.



**Abb. 2.1** Zeitreihenpolygon der monatlichen Auftragseingänge im Wohnungsbau (Wertindizes)

monaten verglichen mit den Wintermonaten, und außerdem bis etwa 2009 einen fallenden und ab 2009 einen ansteigenden Trend aufweist. Dieses Erkenntnis ist wichtig bei der Auswahl eines passenden Prognoseverfahrens, das sowohl die Saisonschwankungen als auch den Trend berücksichtigen sollte.

Bei zeitreihenanalytischen Verfahren gibt es also nur die beobachtete Variable  $y_t$  als abhängige und die Zeit  $t$  als unabhängige Variable. Aus dem zeitlichen Verlauf heraus versucht man, auf Gesetzmäßigkeiten wie Trends (also einen langfristigen An- oder Abstieg der Zeitreihenwerte), konjunkturelle (also mittelfristige Schwankungen) oder kurzfristige saisonale Schwankungen zu schließen und diese Ergebnisse dann zur Prognose zu verwenden. Häufig stellt man sich die beobachteten Zeitreihenwerte als additive oder multiplikative Überlagerung einer Trend-, einer Konjunktur- und einer Saisonkomponente sowie eines unsystematischen Zufallseinflusses vor. Im sogenannten klassischen Komponentenmodell wird versucht, die Zeitreihe explizit in diese Komponenten zu zerlegen (vgl. dazu den Beitrag „Szenarioanalyse als Prognoseinstrument mit einem Beispiel zur Kundenbindung“ in diesem Band). Im vorliegenden Beitrag wird stattdessen im Abschn. 2.3 das exponentielle Glätten als Prognoseverfahren thematisiert, bei dem nicht die Zerlegung in Komponenten, sondern die Prognose im Mittelpunkt des Interesses steht, obwohl auch hier Trend- und Saisonkomponenten eine Rolle spielen. Schließlich werden Zeitreihen heute auch häufig explizit als stochastische Prozesse modelliert, etwa durch die in diesem Buch nicht thematisierten ARIMA-Modelle (vgl. dazu zum Beispiel Schlittgen und Streitberg 2001).

Die ökonometrischen Verfahren basieren auf dem Grundgedanken, dass zahlreiche ökonomische Variablen von anderen ökonomischen Variablen abhängen. Dementspre-

chend wird in der Ökonometrie geschätzt, wie eine (oder mehrere) unabhängige (exogene) Variable  $x$  eine (oder mehrere) abhängige (endogene) Variable  $y$  beeinflusst. Ökonometrische Prognosen sind daher häufig sogenannte „bedingte Prognosen“, bei denen die Prognose für den  $y$ -Wert in Abhängigkeit von dem  $x$ -Wert angegeben wird, etwa „wenn der Wert der exogenen Variablen  $x$  im Prognosezeitraum um fünf Prozent steigt, dann wird der Wert der endogenen Variablen  $y$  um drei Prozent steigen“. Dagegen sind Prognosen auf der Grundlage von Zeitreihenanalysen tendenziell unbedingte Prognosen, etwa „der Wert der endogenen Variablen  $y$  wird im Prognosezeitraum um drei Prozent steigen“. Allerdings werden auch hier manchmal Bedingungen angegeben, zum Beispiel „die Prognose gilt, wenn es nicht zu einer Abschwächung der Konjunktur aufgrund unvorhersehbarer exogener Einflüsse kommt“. Werden die zukünftigen  $x$ -Werte in einer ökonometrischen Analyse mit zeitreihenanalytischen Methoden prognostiziert, so kann man auch aus einer ökonometrischen Prognose eine unbedingte Prognose machen.

Ökonometrische Verfahren eignen sich deshalb besonders dann gut zur Prognose, wenn die zukünftigen  $x$ -Werte gut vorausgesagt oder sogar selbst beeinflusst werden können. Im Beitrag „Multiple Regression als Konzept zur Absatzprognose“ in diesem Band werden die Grundlagen der ökonometrischen Verfahren am Beispiel der Schätzung von Preis-Absatz-Funktionen dargestellt, die eine Prognose des zukünftigen Absatzes in Abhängigkeit vom selbst gesetzten Preis und weiteren absatzpolitischen Maßnahmen erlauben. Im vorliegenden Beitrag wird im Abschn. 2.2 kurz erläutert, wie eine solche ökonometrische Schätzung mit gretl durchgeführt werden kann. Bei dieser Gelegenheit ist zu betonen, dass die Stärken von gretl gerade bei den ökonometrischen Verfahren und den ARIMA-Modellen liegen. Das im vorliegenden Beitrag vorrangig beschriebene exponentielle Glätten ist aus gretl Sicht eher ein Randthema, auf das sich gretl trotzdem gut anwenden lässt.

---

## 2.2 Lineare Regression

### 2.2.1 Einfache Regression

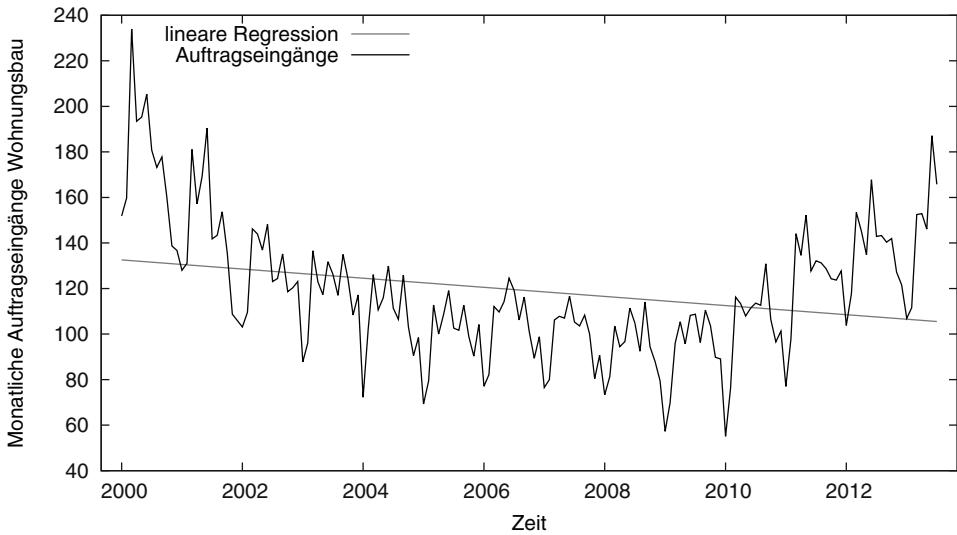
Die lineare Regression ist das grundlegende in der Ökonometrie verwendete Verfahren, um den Einfluss einer oder mehrerer unabhängiger Variablen auf eine oder mehrere abhängige Variablen statistisch zu ermitteln. Im Beitrag „Multiple Regression als Konzept zur Absatzprognose“ in diesem Band wird dieses Verfahren, einschließlich der mathematischen Herleitung der Formeln für die Schätzung der Regressionskoeffizienten, am Beispiel einer Preis-Absatz-Funktion erläutert. Daher wird hier lediglich kurz dargestellt, wie man ökonometrische Verfahren in gretl umsetzen kann.

Bei der einfachen linearen Regression wird nur eine unabhängige Variable (Regressor, exogene Variable) verwendet, mit der eine abhängige Variable (Regressand, endogene Variable) erklärt werden soll. Als Beispiel dienen die Daten der Tab. 2.1. Gesucht ist eine Regressionsgerade  $y = a + bt$ , die Schätzwerte  $\hat{y}_t$  für die echten Auftragseingänge  $y_t$  in

Abhängigkeit von der Zeit  $t$  als unabhängiger Variable liefert, wozu die Regressionskoeffizienten  $a$  und  $b$  geschätzt werden müssen. Die geschätzten Regressionskoeffizienten werden mit  $\hat{a}$  und  $\hat{b}$  bezeichnet. Angesichts des Umfangs dieses Datensatzes ist unmittelbar klar, dass eine Berechnung per Hand wenig Sinn macht, sodass der Einsatz von Software unverzichtbar ist:

1. Starten Sie gretl und geben Sie die Daten ein, was entweder per Hand oder per Import möglich ist. Zur Eingabe per Hand wählen Sie im Menü Datei | Neuer Datensatz und geben 163 bei Anzahl Beobachtungen ein. Wählen Sie anschließend Zeitreihe und Monatlich. Als Startbeobachtung geben Sie 2000:01 ein. Setzen Sie einen Haken bei Mit Dateneingabe beginnen und wählen Sie Anwenden. Als Namen für die erste Variable wählen Sie Auftragseingänge. Sie können nun die Daten aus der Tab. 2.1 eingeben und anschließend durch Auswahl des grünen Hakens bestätigen.
2. Bei 163 Daten ist es allerdings sinnvoller, diese Daten zu importieren. Öffnen Sie dazu die Excel-Datei<sup>2</sup> und kopieren Sie diejenigen Daten, die Sie verwenden möchten, in eine neue Tabelle (zum Beispiel Tabelle 1), die nur diese Daten in Spalte A enthält. Im Feld A:1 sollten Sie den Namen der Variablen (zum Beispiel Auftragseingänge) eingeben. Speichern Sie die Excel-Datei. In gretl wählen Sie nun im Menü Datei | Öffne Daten | Import | Excel und öffnen die soeben gespeicherte Excel-Datei. Bei Tabellenblatt zum Import wählen Sie Tabelle 1 und bestätigen mit OK und anschließend wählen Sie Ja und Zeitreihe sowie Monatlich. Als Startwert geben Sie 2000:01 ein und bestätigen schließlich mit Anwenden.
3. Im gretl-Fenster sehen Sie nun in der Spalte Variablenname die Einträge const und Auftragseingänge. Um die Regression durchführen zu können, wird noch die Variable Zeit (time) benötigt. Sie erzeugen sie mit Hinzufügen | Zeittrend.
4. Wählen Sie nun im Menü Modell | Kleinste Quadrate (OLS). Durch Markieren der Variablen und Anklicken der jeweiligen Pfeile können Sie Auftragseingänge als abhängige sowie const und time als unabhängige Variablen auswählen. Bestätigen Sie mit OK.
5. Als Ergebnis erhalten Sie ein neues Fenster mit einer vollständigen Auswertung der linearen Regression. Die geschätzte Regressionsgerade lautet  $y = 132,733 - 0,167t$ . Die Software gibt auch einige statistische Testgrößen aus, die in der rein deskriptiven Betrachtung nicht interpretiert werden können (einige werden im Beitrag „Multiple Regression als Konzept zur Absatzprognose“ in diesem Band erläutert). Der angegebene Wert von R-Quadrat in Höhe von 0,072 ist jedoch auch ohne Wahrscheinlichkeitstheoretische Betrachtungen aussagekräftig. Er besagt, dass der lineare Zusammenhang zwischen der Zeit  $t$  und den Auftragseingängen nur sehr schwach ist

<sup>2</sup> Sie finden diese Daten (gegebenenfalls in einer aktualisierten Fassung) im Excel-Format auf <https://www.destatis.de/DE/Publikationen/Thematisch/Bauen/BaugewerbeKonjunktur/Lange-ReihenAuftragseingangBaugewerbe.html>.



**Abb. 2.2** Zeitreihenpolygon und lineare Regression zur Tab. 2.1

beziehungsweise gar nicht besteht. Mit anderen Worten: Die lineare Regression der Auftragseingänge auf die Zeit ist nicht sinnvoll.

6. Wenn Sie im Menü des Ergebnisfensters Graphen | geschätzte vs. tatsächliche | gegen Zeit wählen, erhalten Sie das in der Abb. 2.2 dargestellte Zeitreihenpolygon mit der geschätzten Regressionsgeraden. Wenn Sie in gretl mit der rechten Maustaste auf die Abbildung klicken, öffnet sich ein selbsterklärendes Menü, mit dem Sie die Abbildung zum Beispiel noch überarbeiten und in verschiedenen Formaten abspeichern können.
7. Obwohl auch die grafische Darstellung noch einmal verdeutlicht, dass die hier durchgeführte lineare Regression nicht sinnvoll ist, weil die Auftragseingänge schlicht auch nicht näherungsweise linear von der Zeit abhängen, soll noch kurz erläutert werden, wie eine Prognose mit gretl durchgeführt werden kann. Manuell können Sie einfach die geschätzte Regressionsgerade  $y = 132,733 - 0,167t$  verwenden und zum Beispiel für August 2013 den Wert  $t=164$  oder Dezember 2013 den Wert  $t=168$  eingeben, um die Prognose für die entsprechenden Auftragseingänge zu erhalten. Mit gretl muss zunächst ein sogenannter „Out-of-Sample-Wert“ der Variablen hinzugefügt werden. Wählen Sie im Menü des gretl-Hauptfensters Daten | Beobachtungen hinzufügen und geben Sie zum Beispiel 5 ein. Die Zeittrendvariable und die Auftragseingänge laufen nun bis Dezember 2013, wobei die Auftragseingänge ab August 2013 noch frei sind und nun prognostiziert werden können. Wechseln Sie zum Ergebnisfenster und wählen im Menü Analyse | Prognosen und bestätigen Sie mit OK. Als Ergebnis erhalten Sie unbedingte Prognosewerte der Auftragseingänge von August bis Dezember 2013 sowie eine grafische Darstellung der Prognosegeraden.

Die lineare Regression ist in diesem Beispiel deshalb nicht sinnvoll, weil die Zeitreihe offenbar saisonale Schwankungen und für einen ersten Zeitraum einen absteigenden Trend aufweist, der sich dann später in einen aufsteigenden Trend umkehrt, was anhand der Abb. 2.1 unmittelbar zu erkennen ist. Ohne Berücksichtigung der saisonalen Schwankungen ist daher auch keine sinnvolle Prognose möglich (vgl. dazu den Abschn. 2.3.2). Interessiert man sich lediglich für den Trend, so kann die lineare Regression dann sinnvoll sein, wenn die Daten in mehrere Teilbereiche aufgeteilt werden. Da der sogenannte Strukturbruch, bei dem sich der Trend ändert, etwa im Jahr 2009 liegt, kann man zum Beispiel zwei getrennte lineare Regressionen für den ersten Zeitraum mit absteigendem und den zweiten Zeitraum mit ansteigendem Trend durchführen.

### 2.2.2 Mehrfache Regression

Bei der linearen Mehrfachregression werden mindestens zwei unabhängige Variablen betrachtet. Im Beitrag „Multiple Regression als Konzept zur Absatzprognose“ in diesem Band ist das Modell  $y = a + bp + cx$  betrachtet worden, wobei  $y$  für die Absatzmenge,  $p$  für den Preis und  $x$  für die Ausgaben für die Verkaufsförderung steht. Hier wird nur kurz gezeigt, wie dieses Modell mit den hypothetischen Daten der Tab. 1.1 aus dem Kap. 1 mit gretl geschätzt werden kann. Das Vorgehen unterscheidet sich nicht prinzipiell von demjenigen im Fall der linearen Einfachregression, weshalb die folgende Beschreibung kurz gefasst wird:

1. Starten Sie gretl und geben Sie die Daten aus der Tab. 1.1 für Absatzmenge, Preis und Ausgaben für die Verkaufsförderung ein. Anders als im vorherigen Beispiel wählen Sie nun Querschnittsdaten statt Zeitreihe. Nach Eingabe der Daten für die erste Variable Absatz wählen Sie im Menü jeweils Hinzufügen | Definiere neue Variable, um die weiteren Variablen Preis und Ausgaben einzugeben (alternativ können Sie auch das Pluszeichen bei Eingabe der ersten Variablen anklicken).
2. Wählen Sie nun im Menü Modell | Kleinste Quadrate (OLS). Durch Markieren der Variablen und Anklicken der jeweiligen Pfeile können Sie Absatzmenge als abhängige sowie const, Preis und Ausgaben als unabhängige Variablen auswählen. Bestätigen Sie mit OK.
3. Als Ergebnis erhalten Sie wiederum eine vollständige Auswertung der Mehrfachregression mit denselben Ergebnissen, die auch bereits im Beitrag „Multiple Regression als Konzept zur Absatzprognose“ angegeben worden sind (neben weiteren Testgrößen, die gretl automatisch berechnet). Die Abb. 2.3 zeigt die Ausgabe von gretl. Zur Interpretation wird auf das Kap. 1 verwiesen, in dem die wichtigsten der angegebenen Ergebnisse erläutert werden, sowie auf ein Lehrbuch der Ökonometrie, zum Beispiel Hackl (2013).



Modell 1: KQ, benutze die Beobachtungen 1–15				
Abhängige Variable: Absatzmenge				
	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	1100,12	60,0601	18,3170	0,0000
Preis	−172,113	15,4343	−11,1514	0,0000
AusgabenVerkaufsfoerderung	0,338083	0,0759405	4,4520	0,0008
Mittel d. abh. Var.	669,4000	Stdabw. d. abh. Var.	58,08836	
Summe d. quad. Res.	3608,187	Stdfehler d. Regress.	17,34019	
$R^2$	0,923619	Korrigiertes $R^2$	0,910889	
$F(2, 12)$	72,55403	P-Wert( $F$ )	1,99e−07	
Log-Likelihood	−62,40591	Akaike-Kriterium	130,8118	
Schwarz-Kriterium	132,9360	Hannan–Quinn	130,7892	

**Abb. 2.3** Ausgabe der Ergebnisse der linearen Regression zur Tab. 1.1 mit gretl

4. Die geschätzte Regressionsgerade lautet  $y = 1100,12 - 172,113p + 0,338x$ . Sie können damit direkt Absatzprognosen erstellen, indem Sie numerische Werte für  $p$  und  $x$  einsetzen. Zum Beispiel erhält man für  $p=4$  und  $x=500$  die bedingte Absatzprognose  $\hat{y} = 580,71$ . Mit gretl muss zunächst wieder ein Out-of-Sample-Wert hinzugefügt werden (Hauptfenster Daten | Beobachtungen hinzufügen). Anschließend bearbeiten Sie die beiden Regressoren Preis und Ausgaben und geben die Werte 4 und 500 für die jeweils freien Beobachtungen ein. Im Ergebnisfenster wählen Sie im Menü Analyse | Prognosen und bestätigen mit OK.

## 2.3 Exponentielles Glätten

### 2.3.1 Einfaches exponentielles Glätten

Das einfache exponentielle Glätten (auch „exponentielles Glätten erster Ordnung“ genannt) ist ein grundlegendes Verfahren zur Erstellung kurzfristiger Prognosen bei Zeitreihen, das ursprünglich 1956 von Brown vorgeschlagen wurde und in Brown (1959) dargestellt wird.<sup>3</sup> Wenn  $y_t$  der Zeitreihenwert für den Auftragseingang in der Periode  $t$  ist, dann wird der geglättete Wert  $n_t$  durch

$$n_t = \alpha y_t + (1 - \alpha) n_{t-1} \tag{2.1}$$

berechnet, wobei  $0 < \alpha < 1$  der sogenannte Glättungsparameter ist, dessen numerischer Wert frei festgelegt werden kann.<sup>4</sup> Für den ersten Glättungswert  $n_1$  muss ein Startwert

<sup>3</sup> Die Bezeichnung rührt daher, dass man zeigen kann, dass der Einfluss der vergangenen Zeitreihenwerte auf die aktuelle Prognose mit der zeitlichen Entfernung exponentiell abnimmt.

<sup>4</sup> Im Beitrag „Szenarioanalyse als Prognoseinstrument mit einem Beispiel zur Kundenbindung“ in diesem Band wird gezeigt, wie man diesen und andere Glättungsparameter auch optimal berechnen kann.

angegeben werden, da  $n_0$  unbekannt ist. In der Regel wird  $n_0 = y_1$  und damit automatisch auch  $n_1 = y_1$  gewählt (zu erkennen, indem  $n_0 = y_1$  in Gl. 2.1 eingesetzt wird). In der Praxis reicht es dementsprechend aus, als Startwert gleich  $n_1 = y_1$  festzulegen. Die geschätzten Werte der Zeitreihenvariablen werden im Folgenden durch ein Dach gekennzeichnet. Als „Ein-Schritt-Prognose“ bezeichnet man ausgehend von einem Zeitpunkt  $t$  die Prognose für die Folgeperiode  $t + 1$ . Beim einfachen exponentiellen Glätten ist diese Ein-Schritt-Prognose durch

$$\hat{y}_{t+1} = n_t \quad (2.2)$$

gegeben. Der Prognosewert  $\hat{y}_{t+1}$  ist also gemäß Gl. 2.1 ein mit dem Glättungsparameter gewichtetes arithmetisches Mittel aus dem vorangehenden, tatsächlichen Zeitreihenwert  $y_t$  und seinem Prognosewert  $\hat{y}_t = n_{t-1}$  aus der Vorperiode. Je kleiner  $\alpha$  ist, desto weniger stark beeinflussen die aktuellen Werte die Prognose und desto stärker ist daher der Glättungseffekt. Bei einem hohen Wert von  $\alpha$  haben die aktuellen Werte dagegen ein hohes Gewicht und der Glättungseffekt ist gering. Soll am Ende der letzten Periode  $T$  des sogenannten „Stützbereichs“  $t = 1, \dots, T$  mit beobachteten Daten eine Mehr-Schritt-Prognose für  $h$  Perioden in der Zukunft erfolgen ( $h$  heißt dann „Prognosehorizont“), so wird  $n_T$  verwendet:

$$\hat{y}_{T+h} = n_T. \quad (2.3)$$

Der Glättungswert  $n_t$  wird auch „Niveauwert“ genannt. Diese Bezeichnung ergibt sich aus der Überlegung, dass das einfache exponentielle Glätten dann ein sinnvolles Verfahren ist, wenn die Zeitreihe keinen systematischen Trend und keine systematischen Schwankungen aufweist, also als Summe eines Niveaus  $n_t$  und einer Zufallsschwankung  $\varepsilon_t$  darstellbar ist:  $y_t = n_t + \varepsilon_t$ . Durch die Glättung soll der Zufallseinfluss  $\varepsilon_t$  herausgerechnet werden. Eine wichtige Eigenschaft des exponentiellen Glättens ist, dass der Einfluss der zurückliegenden Zeitreihenwerte auf die aktuellen Prognosen umso kleiner wird, je weiter die Zeitreihenwerte in der Vergangenheit liegen. Wenn sich das Niveau also (hoffentlich unsystematisch) geändert hat, gehen die alten Werte mit noch unterschiedlichem Niveau schwächer in die aktuelle Prognose ein, als die neueren Werte, die das aktuelle Niveau repräsentieren.

Die Tab. 2.2 enthält ein einfaches, hypothetisches Zahlenbeispiel, das sich auch zur Berechnung per Hand eignet. Die erste Zeile enthält die Periode, die zweite Zeile die beobachteten Zeitreihenwerte der hypothetischen Auftragseingänge  $y_t$  bis zur Periode 6. Für die Periode 7 soll eine Prognose erstellt werden. Als Glättungsparameter wird  $\alpha = 0,5$  verwendet.

**Tab. 2.2** Hypothetische Auftragseingänge

Periode $t$	1	2	3	4	5	6	7
Auftragseingänge $y_t$	5	3	6	4	6	5	–
geglättete Werte $n_t$	5	4	5	4,5	5,25	5,125	–
Ein-Schritt-Prognose $\hat{y}_t = n_{t-1}$	–	5	4	5	4,5	5,25	5,125

Die Anwendung der Gl. 2.1 für  $\alpha = 0,5$  und  $n_1 = y_1 = 5$  liefert:

$$n_2 = 0,5 \cdot y_2 + 0,5 \cdot n_1 = 0,5 \cdot 3 + 0,5 \cdot 5 = 4,$$

$$n_3 = 0,5 \cdot y_3 + 0,5 \cdot n_2 = 0,5 \cdot 6 + 0,5 \cdot 4 = 5,$$

$$n_4 = 0,5 \cdot 4 + 0,5 \cdot 5 = 4,5,$$

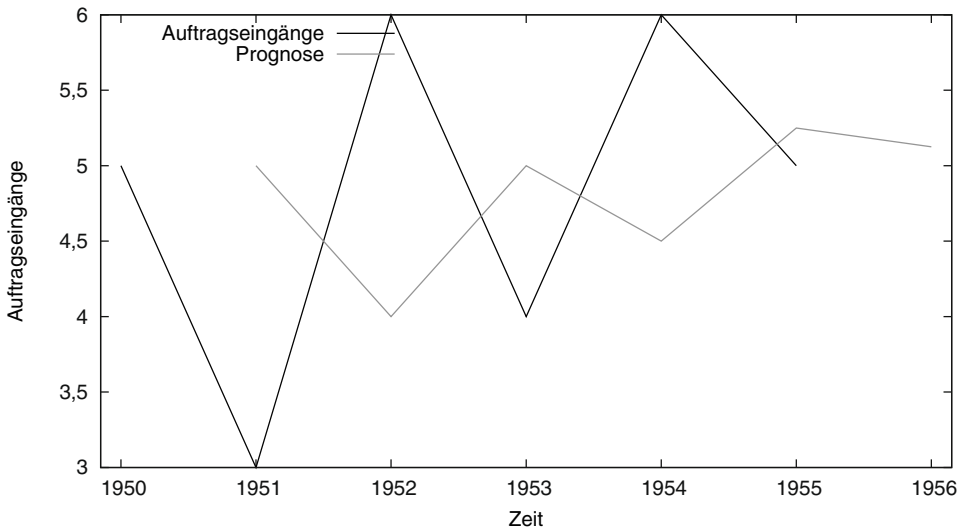
$$n_5 = 0,5 \cdot 6 + 0,5 \cdot 4,5 = 5,25,$$

$$n_6 = 0,5 \cdot 5 + 0,5 \cdot 5,25 = 5,125.$$

Die geglätteten Werte werden in der dritten Zeile der Tabelle angegeben. Gemäß der Gl. 2.2 ergeben sich die in der vierten Zeile angegebenen Ein-Schritt-Prognosen für alle Zeiträume  $t = 2, \dots, 7$ , indem die geschätzten Niveauewerte um jeweils eine Periode in die Zukunft verschoben werden. Insbesondere erhält man damit in der Periode 6 die Prognose der Auftragseingänge der Periode 7:  $\hat{y}_7 = n_6 = 5,125$ , die gegebenenfalls zu runden ist.

Die Berechnung per Hand ist nur bei solch einfachen Beispielen sinnvoll. Für umfangreichere Zeitreihen in der Praxis ist der Einsatz von Software unverzichtbar. Für das vorangehende Beispiel wird daher jetzt noch zur Illustration das Vorgehen beim Einsatz von gretl beschrieben:

1. Starten Sie gretl, wählen Sie im Menü Datei | Neuer Datensatz und geben Sie bei Anzahl Beobachtungen: 7 ein. Wählen Sie anschließend Zeitreihe und Jährlich. Das Jahr der Startbeobachtung ist für dieses hypothetische Beispiel unerheblich und kann auf dem Vorgabewert 1950 bleiben. Haken Sie Mit Dateneingabe beginnen ab und wählen Sie Anwenden. Als Namen für die erste Variable wählen Sie Auftragseingaenge. Sie können nun die Daten 5, 3 usw. aus der Tab. 2.2 eingeben und das letzte Feld freilassen. Bestätigen Sie durch Setzen des grünen Hakens und ignorieren Sie die Warnung.
2. Im gretl-Fenster sehen Sie nun in der Spalte Variablenname den Eintrag Auftragseingaenge (die Variable const können Sie ignorieren). Markieren Sie diese Zeile mit der Maus und wählen Sie dann aus dem Menü Variable | Filter | Exponentzieller gleitender Durchschnitt. Bei Gewicht auf derzeitiger Beobachtung tragen Sie 0,5 ein. Als ersten EMA-Wert wählen Sie das Mittel der ersten n Beobachtungen und tragen 1 ein. Setzen Sie jeweils einen Haken bei Plotte originale und geglättete Reihe und



**Abb. 2.4** Zeitreihenpolygon und Ein-Schritt-Prognosen zur Tab. 2.2 für  $\alpha = 0,5$

bei Geglättete Reihe speichern als. Klicken Sie OK. Als Ergebnis erhalten Sie eine Abbildung der Zeitreihenwerte  $y_t$  und der geglätteten Werte  $n_t$ , die Sie wie bereits im Abschn. 2.2.1 beschrieben noch bearbeiten und abspeichern können.

3. Im gretl-Fenster erscheint jetzt zusätzlich die abgespeicherte Reihe `ema_Auftragseingaenge` mit den geglätteten Werten  $n_t$  (der letzte Wert ist die Prognose für die Periode 7). Um eine Abbildung zu erstellen, die nicht die Originalzeitreihe und die geglätteten Werte (Zeilen 2 und 3 in der Tab. 2.2), sondern die Originalwerte und die Ein-Schritt-Prognosen gegenüberstellt (Zeilen 2 und 4 in der Tab. 2.2), markieren Sie die Zeile `ema_Auftragseingaenge` und wählen im Menü Hinzufügen | Lags gewählter Variablen. Bei Anzahl zu erzeugender Lags geben Sie 1 ein und bestätigen mit OK.
4. Im gretl-Hauptfenster erscheint nun unter `ema_Auftragseingaenge` eine weitere Variable `ema_Auftragseingaenge_1` (klicken Sie gegebenenfalls das Pluszeichen an). Markieren Sie `Auftragseingaenge` und `ema_Auftragseingaenge_1` (drücken Sie dazu die Taste `Strg`). Klicken Sie die rechte Maustaste und wählen Sie im sich öffnenden Menü `Zeitreihengraph`. Bestätigen Sie in einem Graphen mit OK. Als Ergebnis erhalten Sie die Abb. 2.4.

2.3.2 Das Holt-Winters-Verfahren

2.3.2.1 Zeitreihen mit Trend ohne Saison

Das einfache exponentielle Glätten ist für solche Zeitreihen sinnvoll, die weder einen Trend noch saisonale Schwankungen aufweisen. Enthält eine Zeitreihe dagegen zum Beispiel einen ansteigenden Trend, so wird das Niveau systematisch unterschätzt. Die Tab. 2.3 enthält die Jahresdurchschnittswerte der Wertindizes der in der Tab. 2.1 angegebenen Monatswerte. Hier ist von 2000 bis 2009 ein negativer, von 2009 bis 2012 ein positiver Trend zu erkennen.

In der Abb. 2.5 ist das Zeitreihenpolygon zusammen mit den für  $\alpha = 0,3$  erstellten Prognosen zu sehen. Die Abbildung zeigt das typische Verhalten der einfachen exponentiellen Glättung bei Zeitreihen mit Trend. Bei fallendem Trend liegen die Prognosen systematisch oberhalb der echten Werte, bei steigendem Trend systematisch darunter. Bei Zeitreihen mit systematischem Trend muss daher das einfache exponentielle Glätten durch ein Verfahren ersetzt werden, das diesen Trend berücksichtigt. Ähnlich verhält es sich bei Zeitreihen mit systematischen, saisonalen Schwankungen.

Tab. 2.3 Jährliche Auftragseingänge im Wohnungsbau Deutschland (Wertindizes)

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Index	175,6	145,5	127,7	118,4	107,8	100,0	104,2	98,5	94,5	94,2	103,5	125,1	136,7

Quelle: Statistisches Bundesamt, Bauhauptgewerbe Lange Reihen Auftragseingang, Wiesbaden, Juli 2013.

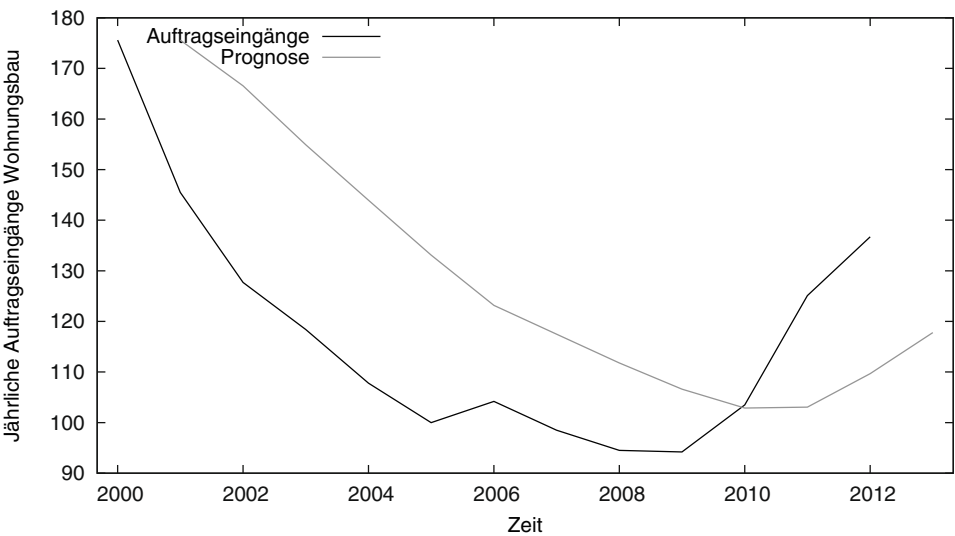


Abb. 2.5 Prognose aufgrund einfacher exponentieller Glättung ( $\alpha = 0,3$ ) zur Tab. 2.3

Geht man etwa davon aus, dass sich die Zeitreihe aus einem linear ansteigenden Trend  $n = a + bt$  und einem Zufallseinfluss  $\varepsilon_t$  zusammensetzt, so ist es sinnvoll, das geschätzte Niveau der Vorperiode, im Vergleich zur Gl. 2.1, um einmal die geschätzte Steigung der Vorperiode  $b_{t-1}$  anzupassen, denn das Niveau nimmt ja pro Periode um einmal die Steigung zu:

$$n_t = \alpha y_t + (1 - \alpha) (n_{t-1} + b_{t-1}). \quad (2.4)$$

Auch die Steigung selbst wird nicht als konstant angesehen und muss daher fortgeschrieben werden. Da die aktuelle Steigung unter Vernachlässigung des Zufallseinflusses gerade gleich der aktuellen Änderung des Niveaus, also gleich  $n_t - n_{t-1}$  ist, ergibt sich die geglättete neue Steigung mit dem zusätzlich zu wählenden Glättungsparameter  $\beta$  als gewichteter Durchschnitt des aktuellen und des vorangehenden Steigungswertes:

$$b_t = \beta (n_t - n_{t-1}) + (1 - \beta) b_{t-1}. \quad (2.5)$$

Als Startwerte werden nun jeweils Werte für das Niveau und für die Steigung benötigt:

$$n_1 = y_1 \text{ und } b_1 = y_2 - y_1 \text{ [mit Gln. 2.4 und 2.5 folgt daraus } n_2 = y_2 \text{ und } b_2 = b_1]. \quad (2.6)$$

Die Gln. 2.4 bis 2.6 stellen das Verfahren von Holt (1959) zur exponentiellen Glättung von Zeitreihen mit Trend dar. Innerhalb des Stützbereichs lauten die Ein-Schritt-Prognosen und damit die geschätzten Zeitreihenwerte:

$$\hat{y}_t = n_{t-1} + b_{t-1}. \quad (2.7)$$

Als Prognose am Ende der Periode  $T$  für den Zeitreihenwert der Periode  $T + h$  ( $h = 1, 2, \dots$ ) dient der geglättete Wert aus  $T$ , korrigiert um  $h$ -mal die in  $T$  geschätzte Steigung:

$$\hat{y}_{T+h} = n_T + h \cdot b_T. \quad (2.8)$$

Wenn möglich, wird nur die Ein-Schritt-Prognose ( $h = 1$ ) verwendet. Die Prognose für die Folgeperioden kann dann jeweils durch die aktualisierte Glättung erfolgen, nachdem die neuen Zeitreihenwerte bekannt sind.

Das Holt-Winters-Verfahren ist in gretl nicht direkt implementiert. Allerdings existiert ein von Ignacio Diaz-Empananza programmiertes Funktionspaket, mit dem gretl um eine spezielle Version des Holt-Winters-Verfahrens erweitert werden kann. Die verfügbaren Funktionspakete finden Sie, indem Sie im Menü Datei | Funktionsdateien | Auf Server wählen (eine Internetverbindung vorausgesetzt). Wenn Sie das Paket HoltWinters markieren, können Sie mit der rechten Maustaste Installation auswählen. Anschließend können Sie das Paket mit Datei | Funktionsdateien | auf lokaler Maschine durch Doppelklick auf HoltWinters laden.<sup>5</sup>

<sup>5</sup> Für dieses Kapitel ist die Version 1.4 der HoltWinters-Funktionsdatei verwendet worden.

**Tab. 2.4** Arbeitstabelle Holt-Winters-Verfahren ohne Saison ( $\alpha = 0,3; \beta = 0,1$ )

Lineare Regression: $y = 76,21 + 11,53t$ , Startwerte: $n_1 = 76,21 + 11,53 = 87,74, b_1 = 11,53$				
Zeit	Daten $y_t$	Prognose $n_{t-1} + b_{t-1}$	$n_t$ gem. Gl. 2.4	$b_t$ gem. Gl. 2.5
t=2 (2009)	94,2	99,2700	97,7490	11,3779
t=3 (2010)	103,5	109,1269	107,4388	11,2091
t=4 (2011)	125,1	118,6479	120,5835	11,4027
t=5 (2012)	136,7	131,9862	133,4003	11,5441
Prognose $\hat{y}_{5+h} = n_5 + h \cdot b_5$				
t=6 (2013)	–	144,9444	–	–
t=7 (2014)	–	156,4885	–	–

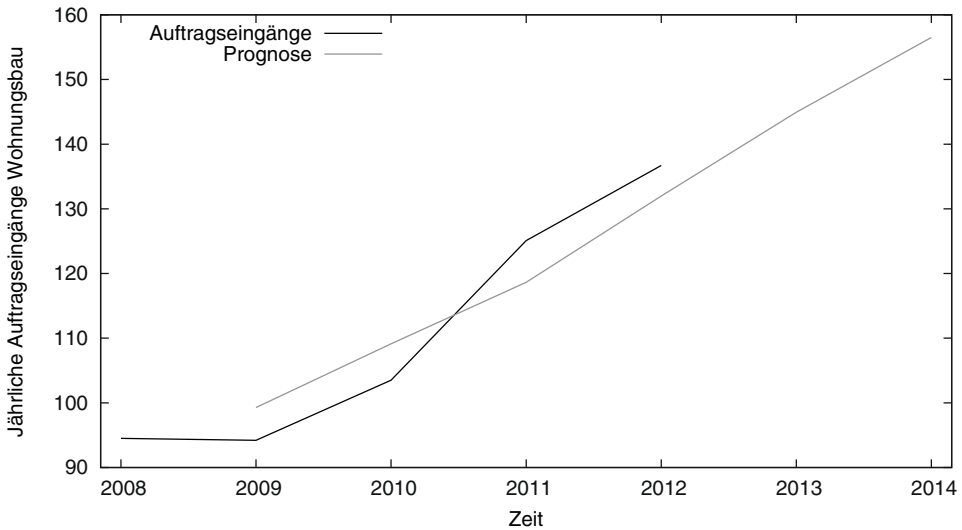
Wenn Sie jährliche Daten verwenden (also Daten ohne Saison), wird automatisch nur das Verfahren von Holt für Zeitreihen mit Trend und ohne Saison verwendet, das heißt, die Gln. 2.4, 2.5 und 2.8 werden zur Glättung und Prognose genutzt. Die Startwerte werden allerdings nicht gemäß Gl. 2.6, sondern mittels einer linearen Regression der Zeitreihenwerte auf die Zeit bestimmt, so wie sie im Abschn. 2.2.1. durchgeführt worden ist. Ist das Ergebnis der linearen Regression die Gerade  $y = \hat{a} + \hat{b}t$ , so werden folgende Startwerte verwendet:

$$n_1 = \hat{a} + \hat{b}, \quad b_1 = \hat{b}.$$

(2.9)

Als Beispiel werden nun die jährlichen Auftragseingänge der Tab. 2.3 betrachtet, wobei lediglich die Werte ab 2008 verwendet werden, um noch eine sinnvolle Berechnung per Hand zu ermöglichen. So wie im Kap. 1 oder im Abschn. 2.2.1 dargestellt, kann eine lineare Regression der Auftragseingänge auf die Zeit mit dem Ergebnis  $y = 76,21 + 11,53t$  erstellt werden, wobei  $t = 1$  für 2008 steht. Als Glättungsparameter werden  $\alpha = 0,3$  und  $\beta = 0,1$  verwendet. Die weitere Berechnung erfolgt dann wie anhand der Tab. 2.4 dargestellt.

- In gretl können Sie diese Berechnungen mithilfe einer Funktionsdatei durchführen:
1. Starten Sie gretl und geben Sie die Daten wie zuvor beschrieben per Hand ein (Menü Datei | Neuer Datensatz, 7 bei Anzahl Beobachtungen, Zeitreihe und Jährlich, Startbeobachtung 2008, Name für erste Variable Auftragseingaenge).
  2. Markieren Sie im gretl-Hauptfenster die Zeile Auftragseingaenge und wählen Sie im Menü Datei | Funktionsdateien | Auf lokaler Maschine (alternativ können Sie in der Fußzeile von gretl auf das Symbol fx klicken). Doppelklicken Sie auf die Zeile HoltWinters (Installation der Funktionsdatei wie zuvor beschrieben vorausgesetzt).
  3. Als dependent variable (series) wählen Sie Auftragseingaenge. Die smoothness parameter lassen Sie auf den Vorgabewerten. Bei list tragen Sie einen Namen ein, zum Beispiel Prognose und bestätigen mit OK.



**Abb. 2.6** Zeitreihenpolygone und Prognosen nach Holt zur Tab. 2.3 ab 2008

4. Im gretl-Hauptfenster erscheint nun eine neue Variable `Auftragseingaenge_H`, die die eben auch per Hand in der Tab. 2.4 berechneten Prognosewerte enthält.
5. Um eine grafische Darstellung zu erhalten, markieren Sie `Auftragseingaenge` und `Auftragseingaenge_H`. Mit der rechten Maustaste können Sie dann `Zeitreihengraph` auswählen und erhalten die Abb. 2.6.

### 2.3.2.2 Zeitreihen mit Trend und Saison

Bei monatlichen oder Quartalsdaten ist in ökonomischen Anwendungen in der Regel davon auszugehen, dass Saisonschwankungen vorliegen, dass also zum Beispiel ein Komponentenmodell der Form  $y_t = n_t + s_t + \varepsilon_t$  sinnvoll ist, bei dem sich die Zeitreihenwerte aus einem Niveau  $n_t$  gegebenenfalls mit Trend, einer Saisonkomponente  $s_t$  und einem Zufallseinfluss  $\varepsilon_t$  zusammensetzen. Das soeben beschriebene Verfahren für Zeitreihen mit Trend muss dann noch um eine Schätzung der Saisonkomponente  $s_t$  ergänzt werden (Winters 1960). Dazu wird in der Gl. 2.4 der aktuelle Wert  $y_t$  um die Saisonkomponente korrigiert, während die Schätzung der Steigung gemäß Gl. 2.5 gleich bleibt. Die geglättete Saisonkomponente selbst wird als gewichteter Durchschnitt des aktuellen und des vorangegangenen Saisonwertes (also demjenigen des gleichen Monats oder Quartals aus dem Vorjahr) geschätzt, wobei  $\gamma$  als zusätzlicher Glättungsparameter verwendet wird, wegen  $s_t = y_t - n_t - \varepsilon_t$  also durch:

$$s_t = \gamma (y_t - n_t) + (1 - \gamma) s_{t-p}.$$

Dabei ist  $p$  die Anzahl der Saisonperioden pro Jahr, zum Beispiel  $p=12$  bei Monatsdaten und  $p=4$  bei Quartalsdaten. Zusammengefasst lautet damit das Holt-Winters-



## Verfahren

$$n_t = \alpha (y_t - s_{t-p}) + (1 - \alpha) (n_{t-1} + b_{t-1}), \quad (2.10)$$

$$b_t = \beta (n_t - n_{t-1}) + (1 - \beta) b_{t-1}, \quad (2.11)$$

$$s_t = \gamma (y_t - n_t) + (1 - \gamma) s_{t-p}. \quad (2.12)$$

Zur Wahl der Startwerte existieren unterschiedliche Vorschläge. Eine einfache Wahl ist

$$n_p = \frac{1}{p} \sum_{t=1}^p y_t, \quad b_p = 0, \quad s_t = y_t - n_p, \quad t = 1, \dots, p. \quad (2.13)$$

Mit der ersten Gleichung wird der Startwert für das Niveau am Ende des ersten Jahres durch das arithmetische Mittel der Zeitreihenwerte im ersten Jahr gewählt. Die zweite Gleichung unterstellt, dass am Ende des ersten Jahres noch kein Trend vorliegt. Mit der dritten Gleichung werden die Saisonkomponenten als Abweichung der Zeitreihenwerte vom geschätzten Niveauwert am Ende des ersten Jahres geschätzt. Um die Startwerte gemäß Gl. 2.13 zu wählen, müssen mindestens  $p$  Beobachtungen vorliegen. Die Rekursion gemäß den Gln. 2.10 bis 2.12 beginnt dann bei  $t = p + 1$ .

Innerhalb des Stützbereichs lauten die Ein-Schritt-Prognosen und damit die geschätzten Zeitreihenwerte:

$$\hat{y}_t = n_{t-1} + b_{t-1} + s_{t-p}. \quad (2.14)$$

Die Prognose am Ende der Periode  $T$  für den Zeitreihenwert der Periode  $T + h$  ( $h = 1, 2, \dots, p$ ) ist

$$\hat{y}_{T+h} = n_T + h \cdot b_T + s_{T-p+h}, \quad (2.15)$$

wobei zu berücksichtigen ist, dass für  $h > p$  keine Schätzungen der Saisonkomponente mehr vorliegen. Soll der Prognosehorizont vergrößert werden, so kann für  $h = p + 1, p + 2, \dots, 2p$  die Formel  $\hat{y}_{T+h} = n_T + h \cdot b_T + s_{T-2p+h}$  genutzt werden.

Tatsächlich werden zahlreiche Varianten des Holt-Winters-Verfahrens verwendet.<sup>6</sup> Die bisher beschriebene Methode wird auch als „Verfahren mit additiver Saisonkomponente“ bezeichnet, denn das zugrunde liegende Modell lautet  $y_t = n_t + s_t + \varepsilon_t$ . Daneben gibt es auch multiplikative Modelle der Form  $y_t = n_t s_t \varepsilon_t$ .<sup>7</sup> Dem Funktionspaket von gretl liegt ein Modell mit multiplikativer Saisonkomponente zugrunde, bei dem ohne Zufallseinfluss  $s_t = y_t / n_t$  ist. Auch die Startwerte werden aufwendiger bestimmt. Die verwendeten Gleichungen werden aus Platzgründen lediglich wie folgt angegeben:

$$n_t = \alpha (y_t / s_{t-p}) + (1 - \alpha) (n_{t-1} + b_{t-1}), \quad (2.16)$$

$$b_t = \beta (n_t - n_{t-1}) + (1 - \beta) b_{t-1}, \quad (2.17)$$

<sup>6</sup> Einen Überblick finden Sie zum Beispiel im Kap. 7 von Hyndman und Athanasopoulos (2013), einem frei zugänglichen Onlinelehrbuch der Prognoseverfahren.

<sup>7</sup> Multiplikative Modelle eignen sich vor allem dann, wenn die Schwankungen bei höherem Niveauwert eine tendenziell größere Amplitude aufweisen.

$$s_t = \gamma (y_t/n_t) + (1 - \gamma) s_{t-p}. \quad (2.18)$$

Für die Startwerte werden nun volle vier Jahre an Daten benötigt. Eine lineare Regression erfolgt für die ersten vier Jahre und die geschätzten Zeitreihenwerte  $\hat{a} + \hat{b}t$  für diese vier Jahre werden als Niveaus verwendet, um die Startwerte der Saisonkomponenten zu ermitteln:

$$s_t = \tilde{s}_t / \bar{s}, t = 1, \dots, p, \text{ wobei}$$

$$\tilde{s}_t = \frac{1}{4} \left( \frac{y_t}{\hat{a} + \hat{b}t} + \frac{y_{t+p}}{\hat{a} + \hat{b}(t+p)} + \frac{y_{t+2p}}{\hat{a} + \hat{b}(t+2p)} + \frac{y_{t+3p}}{\hat{a} + \hat{b}(t+3p)} \right) \text{ und} \quad (2.19)$$

$$\bar{s} = \frac{1}{p} \sum_{t=1}^p \tilde{s}_t.$$

Die Formeln für die Startwerte von Niveau und Steigung (pro Monat oder Quartal) lauten:

$$n_p = y_p, \quad b_p = \frac{1}{p^2} \sum_{t=1}^p (y_{t+p} - y_t). \quad (2.20)$$

Die Rekursion gemäß den Gln. 2.16 bis 2.18 erfolgt ab  $t = p + 1$ . Die Ein-Schritt-Prognosen innerhalb des Stützbereichs lauten:

$$\hat{y}_t = (n_{t-1} + b_{t-1}) s_{t-p}. \quad (2.21)$$

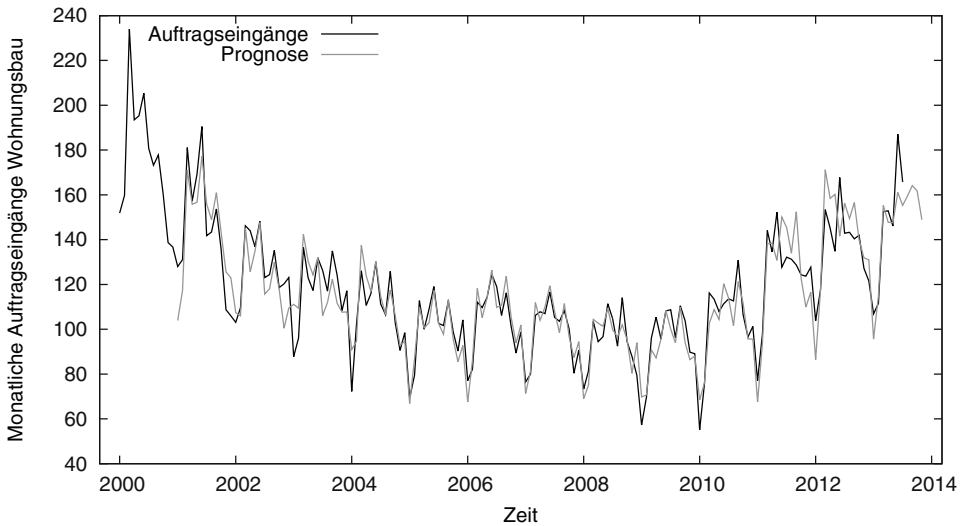
Die Prognose am Ende der Periode  $T$  für den Zeitreihenwert der Periode  $T + h$  ( $h = 1, 2, \dots, p$ ) ist

$$\hat{y}_{T+h} = (n_T + h \cdot b_T) s_{T-p+h}, \quad (2.22)$$

wobei wieder zu berücksichtigen ist, dass für  $h > p$  keine Schätzungen der Saisonkomponente mehr vorliegen. Für  $h = p + 1, p + 2, \dots, 2p$  verwendet man daher  $\hat{y}_{T+h} = (n_T + h \cdot b_T) s_{T-2p+h}$ .

Auf die aufwendige Berechnung eines Beispiels per Hand wird nun verzichtet. Zur Erstellung von Prognosen mit gretl werden die vollständigen Daten der Tab. 2.1 verwendet:

1. Starten Sie gretl und importieren Sie die Daten wie bereits im Abschn. 2.2.1 beschrieben. Um Prognosen für die Zukunft zu erstellen, müssen wiederum Out-of-Sample-Werte hinzugefügt werden (geben Sie 4 im Menü Daten | Beobachtungen hinzufügen ein).
2. Markieren Sie im gretl-Hauptfenster die Zeile Auftragseingaenge und wählen Sie im Menü Datei | Funktionsdateien | Auf lokaler Maschine. Doppelklicken Sie auf die Zeile HoltWinters.
3. Als dependent variable (series) wählen Sie Auftragseingaenge. Die smoothness parameter lassen Sie auf den Vorgabewerten. Bei list tragen Sie einen Namen ein, zum Beispiel Prognose. Bestätigen Sie mit OK.



**Abb. 2.7** Zeitreihenpolygon und Prognosen nach Holt/Winters zur Tab. 2.1

4. Im gretl-Hauptfenster erscheint nun neben der neuen Variablen `Auftragseingänge_H` auch die Variable `Auftragseingänge_W`, die die Prognosewerte des Holt-Winters-Verfahrens mit Saison enthält. Die letzten vier Werte (159,43, 164,10, 161,80 und 148,88) sind die Out-of-Sample-Prognosen für August bis November 2013.
5. Um eine grafische Darstellung zu erhalten, markieren Sie `Auftragseingänge` und `Auftragseingänge_W`. Mit der rechten Maustaste können Sie dann `Zeitreihengraph` auswählen und erhalten eine grafische Darstellung, die hier in der Abb. 2.7 dargestellt wird.

### 2.3.2.3 Automatisierung der Prognose durch ein gretl-Skript

Um die hier dargestellten Möglichkeiten in der Praxis sinnvoll anwenden zu können, bietet es sich an, gretl-Skripte zu verwenden. Wenn Sie zum Beispiel jeden Monat einen neuen Zeitreihenwert erheben und dann jeweils vier Monatswerte im Voraus prognostizieren möchten, kann man diese wiederholte Prognose durch ein Skript erheblich abkürzen. Erstellen Sie dazu mit einem Editor eine reine Textdatei mit der Dateiendung `.inp` mit folgendem Inhalt:

```
open LW:/pfad/daten.xlsx
setobs 12 2000:01
addobs +4
include HoltWinters.gfn
list prognose=HoltWinters(Auftragseingänge, 0.3, 0.1, 0.7)
```

```
print prognose -byobs
gnuplot Auftragseingaenge Auftragseingaenge_W -with-lp -time-series
-output=LW:/pfad/plot.eps.
```

Mit der ersten Zeile öffnen Sie die Datendatei `daten.xlsx`, wobei hier unterstellt wird, dass Sie eine Excel-Datei verwenden (LW steht für das Laufwerk und `pfad` für den Pfad zu dem Verzeichnis, in dem Sie ihre Datendatei abgespeichert haben). Nutzen Sie lediglich die erste Tabelle der Excel-Datei und geben Sie im Feld A:1 den Namen der Variablen, zum Beispiel `Auftragseingaenge` ein. Die Zahlenwerte sollten direkt darunter alle in der ersten Spalte stehen. Mit der zweiten Zeile setzen Sie fest, dass die Daten monatlich vorliegen und dass die Startbeobachtung aus Januar 2000 ist. Durch die dritte Zeile fügen Sie vier noch leere Beobachtungen hinzu, für die die Prognosen erstellt werden sollen. Die vierte Zeile lädt die HoltWinters-Funktionsdatei, mit der fünften und sechsten Zeile wird die Prognose erstellt und numerisch ausgegeben. Für die Glättungsparameter (0.3, 0.1 und 0.7) können Sie natürlich auch andere Werte verwenden. Die siebte Zeile (die hier aus Platzgründen auf zwei Zeilen umgebrochen wird) schließlich erzeugt eine Abbildung und speichert sie in der Datei `plot.eps` im Encapsulated PostScript-Format. Wenn Sie anstelle von `.eps` die Dateierendung `.pdf` angeben, erhalten Sie eine Abbildung im pdf-Format.

Um eine Prognose durchzuführen, müssen Sie nun lediglich die Datei mit der Endung `.inp` doppelt anklicken (eventuell müssen Sie nach dem ersten Doppelklick noch `gretl` als Standardprogramm zum Öffnen dieses Dateityps auswählen). Damit starten Sie `gretl` und öffnen ein Skriptfenster, in dem Sie durch Klick auf das Zahnradsymbol das Skript ausführen können. Wenn Sie die Prognose im Folgemonat aktualisieren wollen, müssen Sie lediglich Ihre Datendatei `daten.xlsx` um den neu erhobenen Zeitreihenwert ergänzen und können dann die Skriptdatei für eine erneute Prognose starten.

---

## 2.4 Abschließende Bemerkungen

In diesem Beitrag sind einige grundlegende Verfahren der Prognoserechnung auf rein deskriptive Weise erörtert worden, das heißt ohne Betrachtung der zugrunde liegenden stochastischen Modelle. Der Schwerpunkt ist dabei auf die Anwendung der Open Source-Software `gretl` gelegt worden. Zu betonen ist, dass `gretl` gerade im Bereich der Ökonometrie und der hier nicht behandelten ARIMA-Modelle auch professionellen und spezialisierten Ansprüchen genügt. Die Bedeutung, die `gretl` mittlerweile in diesen Bereichen hat, erkennt man auch daran, dass zum Beispiel Hackl (2013) in der Neuauflage seines Lehrbuchs der Ökonometrie in Anhängen die praktische Anwendung der Methoden nicht mehr nur anhand des kommerziellen EViews, sondern auch mit `gretl` beschreibt. Zum Lehrbuch *Principles of Econometrics* von Hill et al. (2011) gibt es eine umfangreiche, frei erhältliche Anleitung zur Umsetzung mit `gretl` (Adkins 2013).

Die Berücksichtigung induktiver Methoden erlaubt neben den hier betrachteten Punktprognosen beispielsweise auch Intervallprognosen mit Wahrscheinlichkeitsaussagen (vgl. dazu den Beitrag „Szenarioanalyse als Prognoseinstrument mit einem Beispiel zur Kundenbindung“ in diesem Band) oder statistische Tests zur Verbesserung der Modellauswahl und -interpretation. Am prinzipiellen Vorgehen mit gretl ändert sich jedoch nichts.

Die in diesem Beitrag dargestellte exponentielle Glättung nach Holt/Winters ist bis in die Gegenwart hinein die Grundlage für weitere Forschung und die Entwicklung von Varianten dieser Methode zum Beispiel zur Anpassung an spezielle Situationen (etwa starke Ausreißer in den Daten). Eine kurze Übersicht finden Sie in Goodwin (2010), eine umfangreiche Darstellung in Gardner (2006).

In diesem Beitrag ist über die Auswahl der Glättungsparameter beim exponentiellen Glätten wenig gesagt worden. In der Praxis kann man zum Beispiel, ausgehend von den in gretl voreingestellten Parametern, durch Auswahl anderer numerischer Werte zwischen null und eins anhand der grafischen Darstellung von Zeitreihen und Prognosen austesten, ob sich die Prognosen verbessern. Ein systematischerer Weg ist es, diese Parameter so zu bestimmen, dass zum Beispiel die mittlere quadratische Abweichung der Ein-Schritt-Prognosen von den echten Werten für den Zeitbereich mit bereits bekannten Zeitreihenwerten minimiert wird. Im Kap. 10 wird gezeigt, wie eine derartige Optimierung mit dem Statistikpaket R durchgeführt werden kann.

Die relativ einfache Umsetzung von Prognoseverfahren mithilfe von Software sollte nicht dazu verleiten, sich allein auf statistische Prognosen zu verlassen. Mit einer Punktprognose werden Sie aller Wahrscheinlichkeit nach so gut wie nie richtig liegen, sondern bestenfalls in der Nähe des tatsächlich eintretenden Wertes. Ein Beispiel für eine eingehende Analyse der Fehler von Prognosen liefern Juhn und Loungani (2002), die unter anderem die Genauigkeit von Wachstumsprognosen des realen BIPs in zahlreichen industrialisierten Ländern in den 1990er-Jahren überprüft haben. Der durchschnittliche Fehler der Prognosen am Jahresanfang für das jeweils laufende Jahr betrug mehr als 1,1 Prozentpunkte und vergrößerte sich mit dem Prognosehorizont noch erheblich. Angesichts eines tatsächlichen durchschnittlichen Wachstums von 2,3 Prozent in diesem Zeitraum lagen die Prognosen also deutlich falsch. Eine kurze Darstellung einiger Gründe für Prognosefehler findet man bei Sandte (2004). Für die betriebliche Praxis folgt, dass Entscheidungen nicht allein aufgrund von statistischen Prognosen gefällt werden dürfen, sondern dass die Entscheidungsfindung von zahlreichen Gesichtspunkten abhängig gemacht werden sollte, unter denen die statistische Prognose lediglich ein (wenn auch wichtiges) Kriterium darstellt.

---

## Literatur

- Adkins, L. C. (2013). *Using gretl for Principles of Econometrics*, 4. Auflage, Version 1.041. <http://www.LearnEconometrics.com/gretl.html>. Zugegriffen: 16.01.2014
- Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. New York.: McGraw-Hill.

- Cottrell, A., & Lucchetti, R. (2012). *Gretl User's Guide, Teil des gretl-Programmpaketes, hier Version 1.9.12*. <http://gretl.sourceforge.net>
- Gardner Jr., E. S. (2006). Exponential Smoothing: The state of the Art - Part II. *International Journal of Forecasting*, 22(4), 637–666.
- Goodwin, P. (2010). The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong. *Foresight*, 19, 30–33.
- Hackl, P. (2013). *Einführung in die Ökonometrie* (2. Aufl.). München: Pearson.
- Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). *Principles of Econometrics* (4. Aufl.). New York.: Wiley.
- Holt, C. (1959). *Forecasting seasonals and trends by exponentially weighted moving averages*. ONR Research Memorandum, Bd. 52.
- Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: Principles and Practice*. <http://otexts.org/fpp/>. Zugegriffen: 16.01.2014
- Juhn, G., & Loungani, P. (2002). Further Cross-Country Evidence on the Accuracy of the Private Sector's Output Forecasts. *IMF Staff Papers*, 49(1), 49–64.
- Lübke, K., & Vogt, M. (2015). *Angewandte Wirtschaftsstatistik: Daten und Zufall*. Wiesbaden: Springer Gabler.
- Sandte, H. (2004). Grenzen von Prognosen - oder: Warum Prognostiker irren (dürfen). *Das Wirtschaftsstudium*, 33, 189–190.
- Schlittgen, R., & Streitberg, B. H. J. (2001). *Zeitreihenanalyse* (9. Aufl.). München: Oldenbourg.
- Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.
- Yalta, A. T., & Yalta, A. Y. (2010). Should Economists Use Open Source Software for Doing Research? *Computational Economics*, 35(4), 371–394.

Markt- und Absatzprognosen

Modelle - Methoden - Anwendung

Gansser, O.; Krol, B. (Hrsg.)

2015, XXIV, 370 S. 90 Abb., Hardcover

ISBN: 978-3-658-04491-6