Zeitreihenanalyse

Von der Beobachtung zum Modell

Ein Leitfaden mit Beispielrechnungen

Berufsspezialist (IHK DQR5)

LFB KI BW • iludis.de

Fahrplan: Vom Beobachten zum Modellieren

Dieses Skript führt schrittweise durch die Grundlagen der Zeitreihenanalyse. Der rote Faden folgt einer einfachen Logik:

Phase	Block	Inhalt
👁 Beobachten	1–2	Was sind Zeitreihen? Dekomposition in Komponenten
✅ Prüfen	3–4	Stationarität als Voraussetzung. ACF als Werkzeug
🔧 Herstellen	5	Differenzierung: Stationarität erzeugen
📊 Modellieren	6–7	AR: Vergangenheit nutzen. MA: Schocks modellieren
🔍 Auswählen	8	PACF + Entscheidungstabelle für die Modellwahl
🧩 Zusammenführen	9	ARIMA(p,d,q) als Baukasten
Grenzen des Modells	10	Was kann ARIMA, was nicht?
SARIMA als Erweiterung	11	Saisonalität modellieren
Der ADF-Test	12	

Eine durchgehende Mini-Zeitreihe dient als Anker für alle Beispielrechnungen. So entsteht ein zusammenhängendes Bild statt isolierter Übungen.

1) Was sind Zeitreihen?

Eine Zeitreihe ist eine geordnete Folge von Datenpunkten, die in regelmäßigen zeitlichen Abständen erhoben werden. Der entscheidende Unterschied zu „normalen" Datensätzen: Die Reihenfolge der Daten trägt Information. Vertauscht man die Zeilen eines normalen Datensatzes, ändert sich nichts. Vertauscht man die Punkte einer Zeitreihe, geht die gesamte Struktur verloren.

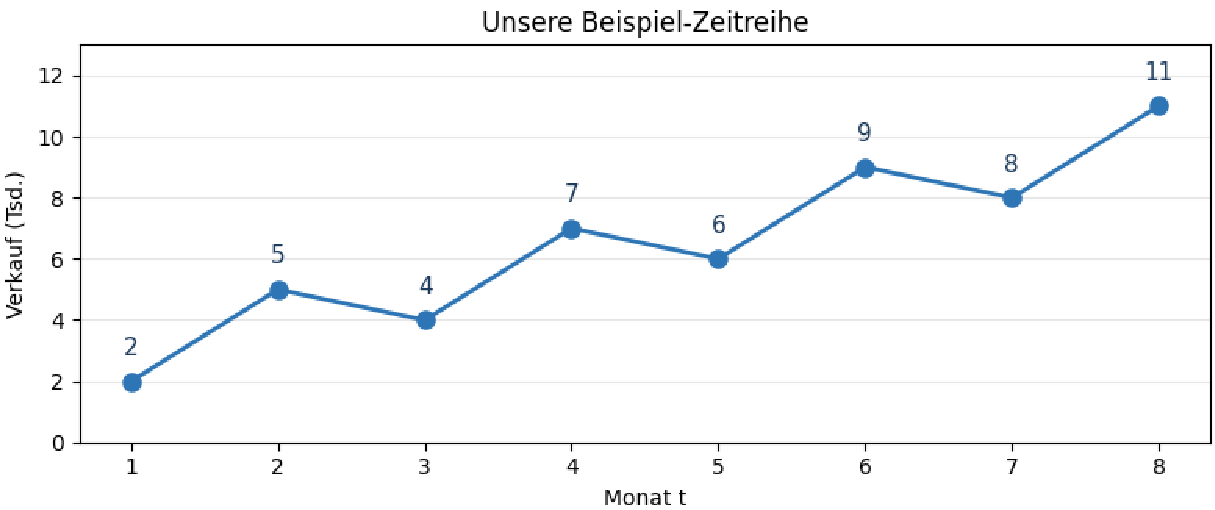
Alltagsbeispiele

Zeitreihen begegnen uns überall: tägliche Temperaturmessungen, monatliche Umsatzzahlen, stündliche Aktienkurse, jährliche CO_2-Konzentrationen, sekundündliche Sensordaten in der Produktion.

Unsere Beispiel-Zeitreihe

Für den gesamten Kurs verwenden wir eine kleine Zeitreihe mit 8 Werten — fiktive monatliche Verkaufszahlen (in Tausend Stück):

Monat t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Wert x_t	2	5	4	7	6	9	8	11

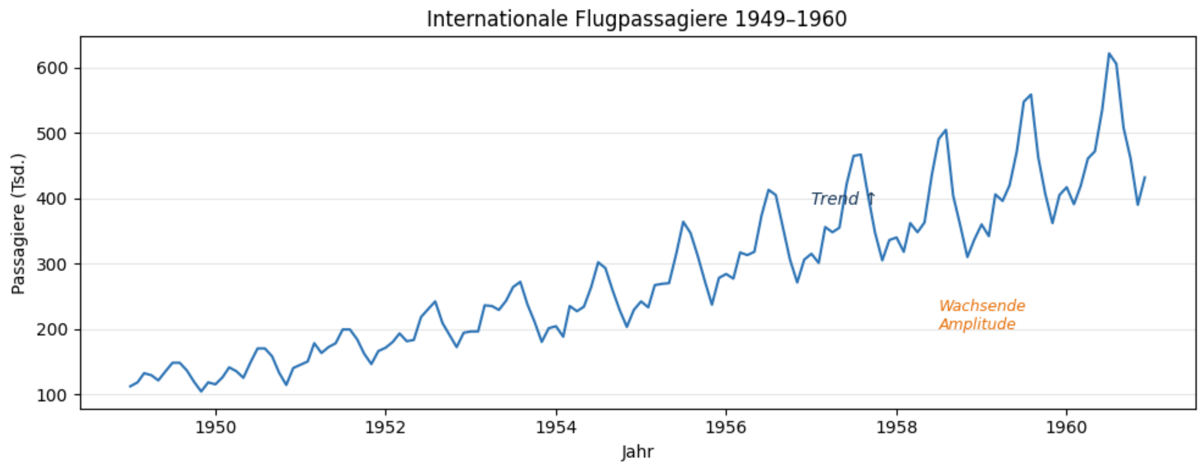


Was fällt auf? Die Werte steigen grob an (Trend), aber es gibt ein Auf und Ab (Schwankung).

Das werden wir in Block 2 genauer zerlegen.

Airpassengers-Zeitreihe

Hier die berühmte ‘Air-Passengers’-Zeitreihe, die eine ähnliche Bedeutung besitzt wie der Iris-Datensatz:



> 📌 **Merksatz:** Eine Zeitreihe ist eine zeitlich geordnete Folge von Messwerten. Die Reihenfolge ist entscheidend — sie trägt die Information.

2) Anatomie einer Zeitreihe — Dekomposition

Jede Zeitreihe lässt sich gedanklich in Bausteine zerlegen. Das klassische Komponentenmodell kennt drei:

Die drei Komponenten

Trend (T)

Die langfristige Entwicklungsrichtung. Steigt der Umsatz über Jahre? Sinkt die Arbeitslosenquote? Der Trend beschreibt die „Grundrichtung“, wenn man alle kurzfristigen Schwankungen glätten würde.

Saisonalität (S)

Regelmäßig wiederkehrende Muster mit fester Periode. Eisverkauf steigt jeden Sommer, Heizkosten jeden Winter. Die Periode ist bekannt und konstant (z. B. 12 Monate, 4 Quartale, 7 Wochentage).

Residuum / Rauschen (ε)

Alles, was nach Abzug von Trend und Saisonalität übrig bleibt. Das Residuum enthält zufällige Schwankungen (Rauschen), aber auch unerklärte Effekte. Ein gutes Modell reduziert das Residuum auf reines Rauschen ohne erkennbare Muster.

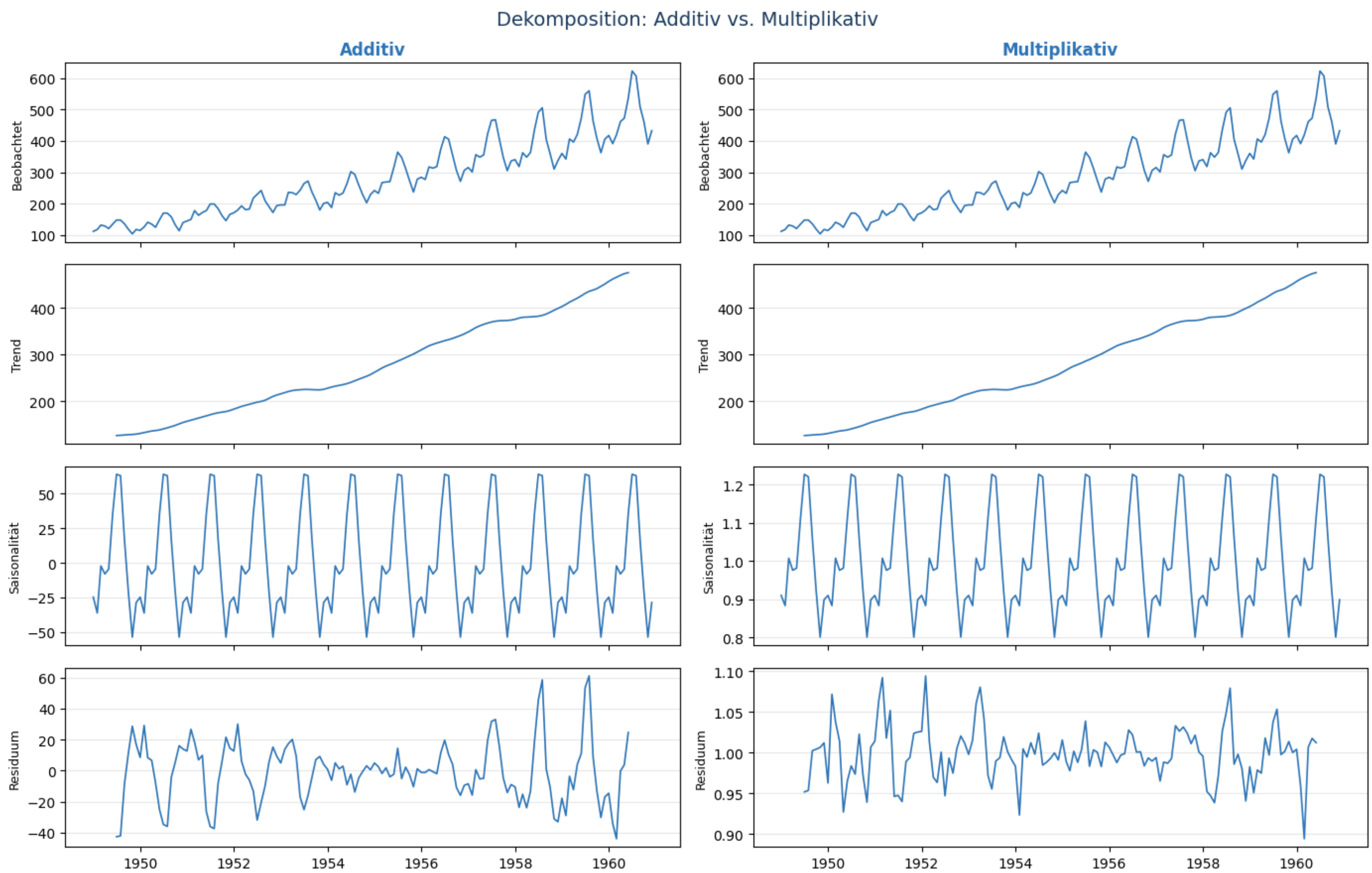
Additiv oder Multiplikativ?

Die entscheidende Frage ist: Wie hängen die Komponenten zusammen?

Modell	Formel	Wann passend?
Additiv	$x_t = T + S + \varepsilon$	Die saisonalen Schwankungen bleiben über die Zeit gleich groß (konstante Amplitude).
Multiplikativ	$x_t = T \times S \times \varepsilon$	Die Schwankungen wachsen mit dem Niveau (z. B. Flugpassagierdaten: höherer Trend → größere saisonale Ausschläge).

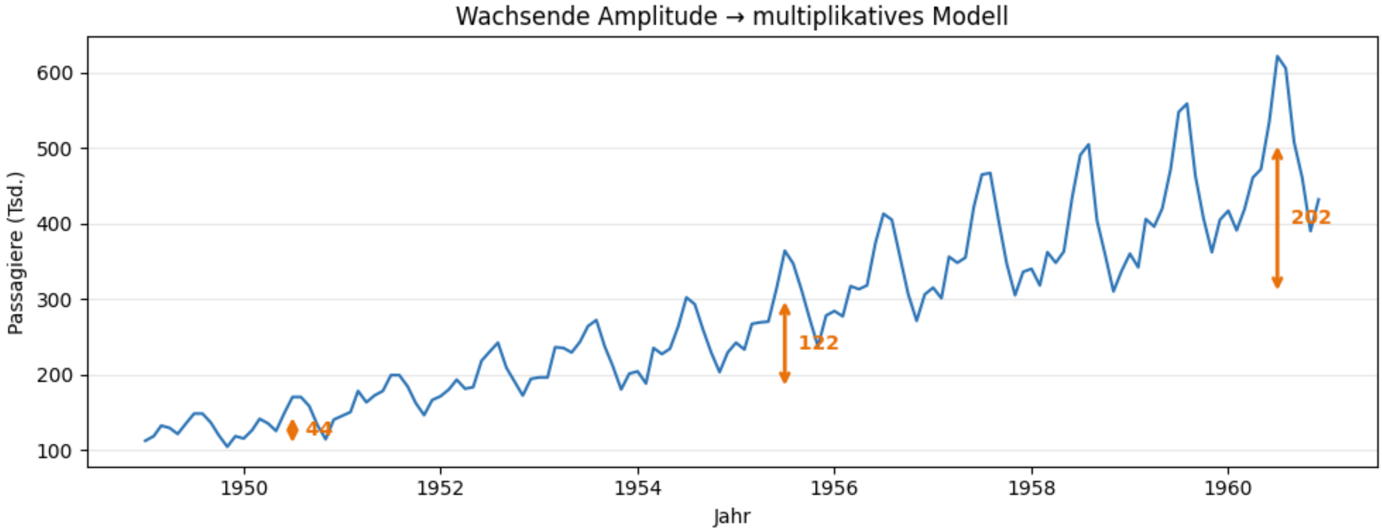
> ⚠ **Hinweis:** Praxis-Faustregel: Schauen Sie den Zeitreihenplot an. Wenn die Höhe der saisonalen „Zähne" über die Zeit gleich bleibt → additiv. Wenn die Zähne mit dem Niveau wachsen → multiplikativ. Im multiplikativen Fall hilft oft eine Log-Transformation, um das Problem auf den additiven Fall zurückzuführen.

> 📌 **Merksatz:** Eine Zeitreihe = Trend + Saisonalität + Residuum (additiv) oder Trend × Saisonalität × Residuum (multiplikativ). Die Dekomposition macht jede Komponente einzeln sichtbar und analysierbar.



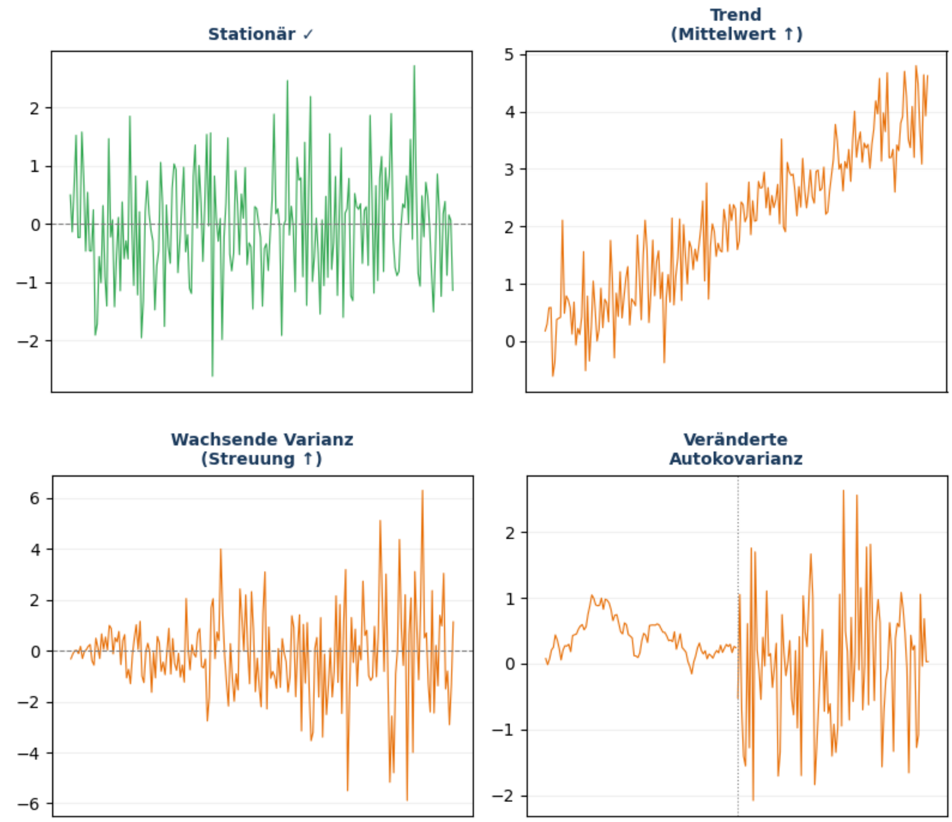
Amplitude-Check (warum multiplikativ?):

man sieht direkt, dass die saisonale Schwankung von ~44 auf ~202 wächst



3) Stationarität: die zentrale Voraussetzung

Bevor wir Zeitreihen modellieren können, brauchen wir eine fundamentale Eigenschaft: **Stationarität**. ARIMA-Modelle setzen sie voraus. Ohne Stationarität ändern sich die statistischen Eigenschaften der Reihe über die Zeit — und ein Modell, das auf einem Zeitabschnitt trainiert wurde, versagt auf einem anderen.



Die grüne Kurve verläuft stationär: im Laufe der Zeit verändern sich weder

der Trend *(z.B. Mittelwert steigt)*,

noch ändert sich die Varianz *(z.B. stärker werdende Streuung)*,

noch ändert sich die Kovarianz *(also die Stärke der Abhängigkeit der Datenpunkte zu ihren Vorgängern)*.

Auffrischung: Varianz und Kovarianz

Die **Varianz** misst, wie stark Werte um ihren Mittelwert streuen. Große Varianz bedeutet starke Schwankungen, kleine Varianz bedeutet enge Bündelung um den Mittelwert.

Die **Kovarianz** misst, ob und wie zwei Variablen gemeinsam schwanken. Positive Kovarianz: Wenn A steigt, steigt tendenziell auch B. Negative Kovarianz: Wenn A steigt, fällt B. Kovarianz nahe null: Kein gemeinsamer Trend.

Neu: Autokovarianz — Kovarianz über Lags

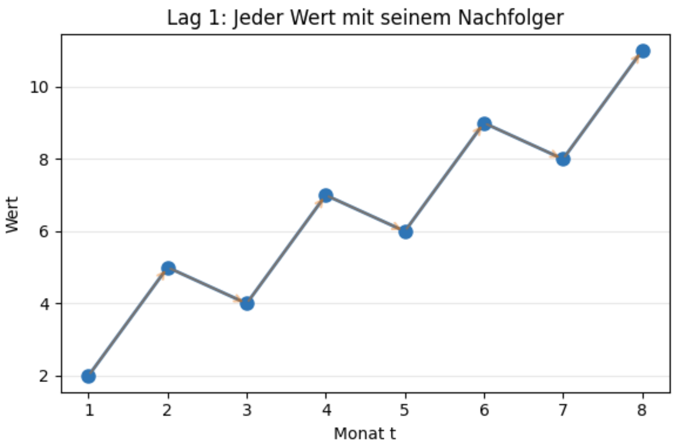
In der Zeitreihenanalyse kommt ein neuer Gedanke hinzu: Wir berechnen die Kovarianz einer Zeitreihe **mit sich selbst**, aber zeitversetzt. Das nennt man **Autokovarianz**.

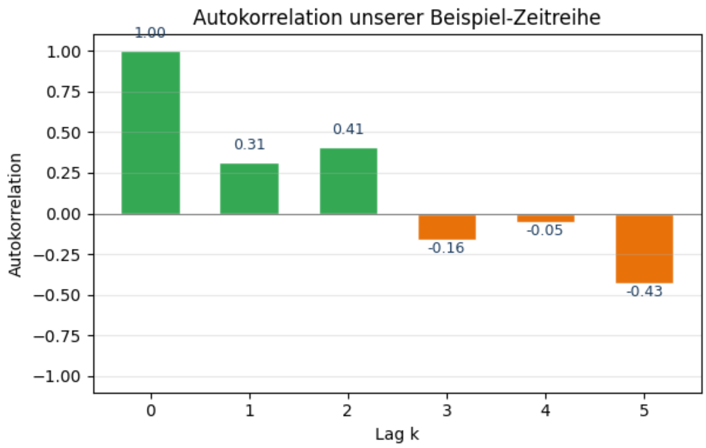
Der **Lag** (Verzögerung) gibt an, um wie viele Zeitschritte wir die Reihe gegen sich selbst verschieben:

	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8
Original x_t	2	5	4	7	6	9	8	11
Lag 1: x_{t-1}	—	2	5	4	7	6	9	8
Lag 2: x_{t-2}	—	—	2	5	4	7	6	9

Die Autokovarianz bei Lag k beantwortet die Frage: „Wie stark hängt ein Wert mit dem Wert k Schritte vorher zusammen?“

In der Abbildung links unsere einfache Beispielzeitreihe, in der rechten Abbildung sieht man die berechnete Autokorrelation zwischen den einzelnen Zeitpunkten:





Normiert man die Autokovarianz (teilt durch die Varianz), erhält man die **Autokorrelation** — einen Wert zwischen −1 und +1, der einfacher zu interpretieren ist. Dazu mehr in Block 4.

Die drei Bedingungen für Stationarität

Eine Zeitreihe heißt **(schwach) stationär**, wenn drei Bedingungen gleichzeitig erfüllt sind:

Bedingung	Bedeutung	Gegenbeispiel
1. Konstanter Mittelwert	$E(x_t)$ ändert sich nicht über die Zeit	Aufsteigender Trend: der Mittelwert verschiebt sich nach oben
2. Konstante Varianz	$Var(x_t)$ ändert sich nicht über die Zeit	Wachsende Amplitude: die Streuung nimmt mit der Zeit zu
3. Konstante Autokovarianz	$Cov(x_t, x_{t+k})$ hängt nur von k ab, nicht von t	Die Korrelation zwischen benachbarten Werten ändert sich je nach Position in der Reihe

Anschaulich: Eine stationäre Zeitreihe „vergisst", wo sie in der Zeit steht. Egal, welchen Ausschnitt man betrachtet — die statistischen Eigenschaften sehen gleich aus.

> 📌 **Merksatz:** Stationarität bedeutet: Der Mittelwert, die Varianz und die Autokovarianz einer Zeitreihe verändern sich nicht über die Zeit. ARIMA-Modelle setzen Stationarität voraus — ohne sie ist keine zuverlässige Prognose möglich.

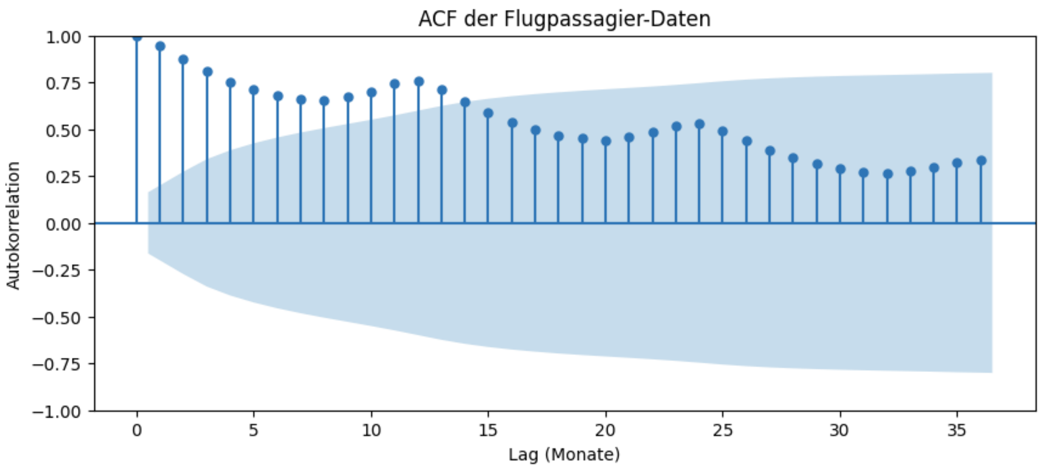
4) ACF: Autokorrelation sichtbar machen

Wir haben gerade gelernt, dass die Autokovarianz bei Stationarität nur vom Lag abhängen darf. Die **Autokorrelationsfunktion (ACF)** macht genau diese Abhängigkeiten sichtbar.

Was zeigt die ACF?

Die ACF plottet für jeden Lag k (1, 2, 3, ...) den Autokorrelationskoeffizienten — also die normierte Autokovarianz. Der Wert liegt zwischen −1 und +1:

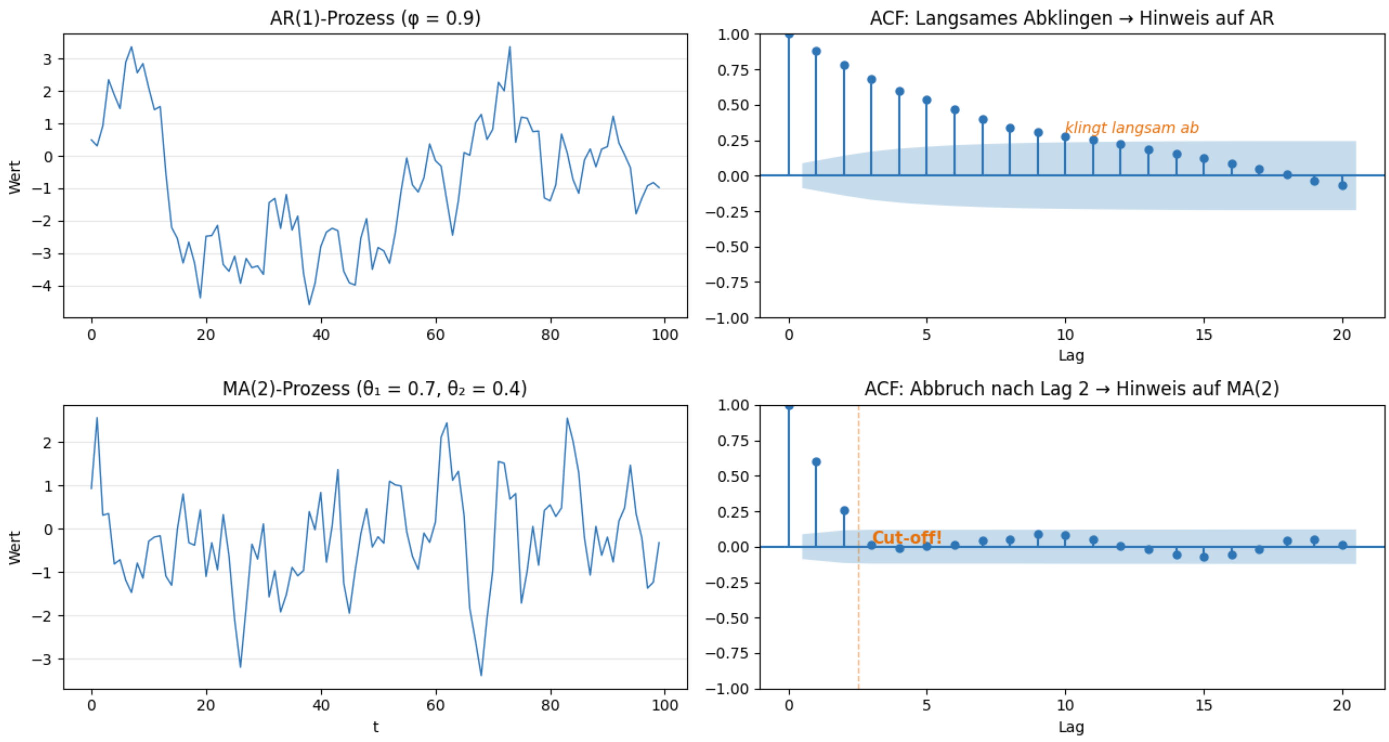
+1 = perfekte positive Korrelation (wenn x_t hoch, ist x_{t-k} auch hoch). **0** = keine Korrelation. **−1** = perfekte negative Korrelation.



Wie liest man einen ACF-Plot?

Ein ACF-Plot zeigt vertikale Balken für jeden Lag. Zusätzlich gibt es ein Konfidenzband (oft blau schattiert). Balken, die über das Band hinausragen, gelten als **statistisch signifikant**.

Es gibt zwei typische Muster, die für die spätere Modellwahl entscheidend sind:



Muster im ACF-Plot	Was es bedeutet	Hinweis auf ...
Langsames Abklingen	Die Korrelation nimmt mit wachsendem Lag nur langsam ab	AR-Prozess (autoregressive Abhängigkeit)
Abrupter Abbruch nach Lag q	Bis Lag q signifikant, danach plötzlich nicht mehr	MA(q)-Prozess (q = letzter signifikanter Lag)
Periodische Spitzen	Hohe Korrelation bei Lags 12, 24, 36 ...	Saisonalität mit Periode 12

⚠ **Hinweis:** Auf die eigentliche Berechnung der ACF-Koeffizienten von Hand verzichten wir hier. Wichtig ist, den Plot lesen und interpretieren zu können. Die Berechnung übernehmen Werkzeuge wie Python (statsmodels), R oder Orange Data Mining.

📌 **Merksatz:** Die ACF zeigt, wie stark ein Zeitreihenwert mit seinen vergangenen Werten korreliert. Sie macht Muster und saisonale Effekte sichtbar und gibt uns einen ersten Hinweis auf die Art des Modells (AR oder MA).

5) Differenzierung: Stationarität herstellen (das I in ARIMA)

Die meisten realen Zeitreihen sind **nicht** stationär — sie haben einen Trend. Die Lösung: **Differenzierung**. Wir bilden die Differenz aufeinanderfolgender Werte.

Differenzierung 1. Ordnung (d = 1)

$$x'_t = x_t - x_{t-1}$$

Wir ziehen von jedem Wert seinen Vorgänger ab. Der Trend wird dadurch entfernt, denn Trends verschwinden bei der Differenzbildung.

Beispielrechnung mit unserer Zeitreihe

t	1	2	3	4	5	6	7	8
x_t	2	5	4	7	6	9	8	11

Differenzierung Schritt für Schritt:

Schritt	Rechnung	Ergebnis x'_t
t = 2	5 − 2	3
t = 3	4 − 5	−1
t = 4	7 − 4	3
t = 5	6 − 7	−1
t = 6	9 − 6	3
t = 7	8 − 9	−1
t = 8	11 − 8	3

Die differenzierte Reihe lautet: **3, −1, 3, −1, 3, −1, 3**

Beobachtung: Der ansteigende Trend ist verschwunden. Die differenzierte Reihe schwankt um einen Mittelwert (hier: ca. 1,3) mit konstanter Varianz. Das sieht stationär aus!

Differenzierung 2. Ordnung (d = 2)

Falls eine einfache Differenzierung nicht ausreicht (z. B. bei einem parabolischen Trend), kann man **nochmals differenzieren**:

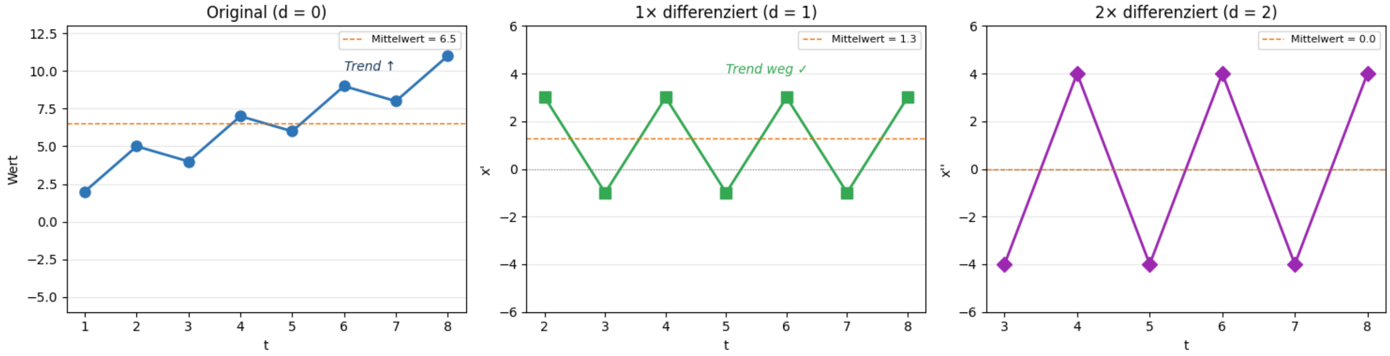
$$x''_t = x'_t - x'_{t-1}$$

In unserem Beispiel (x': 3, −1, 3, −1, 3, −1, 3):

Schritt	Rechnung	Ergebnis x''_t
t = 3	−1 − 3	−4
t = 4	3 − (−1)	4
t = 5	−1 − 3	−4
t = 6	3 − (−1)	4
t = 7	−1 − 3	−4
t = 8	3 − (−1)	4

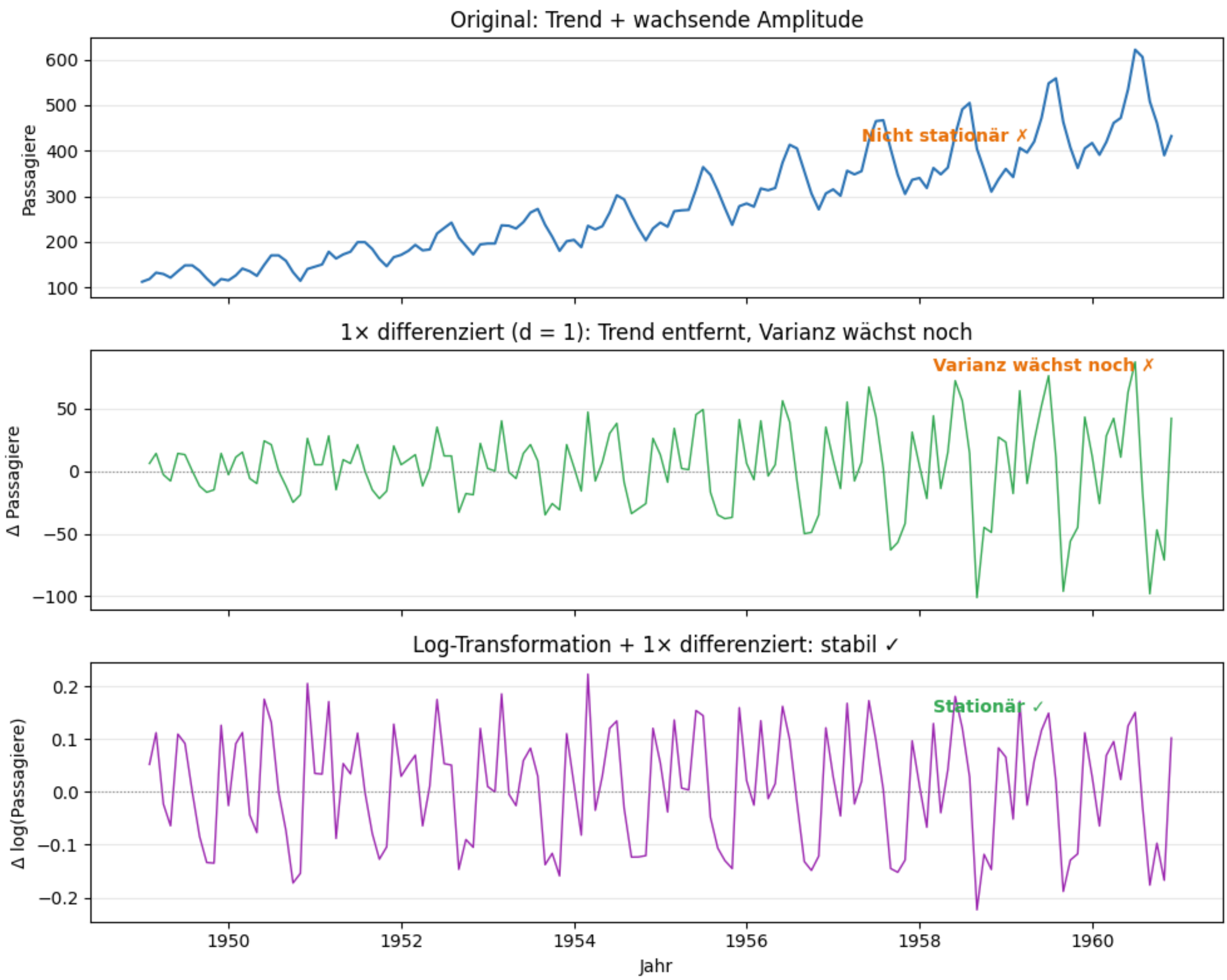
Ergebnis: **−4, 4, −4, 4, −4, 4** — ein perfekt alternierendes Muster. In der Praxis bedeutet d = 2, dass auch ein quadratischer Trend entfernt wird.

Hier die Visualisierungen der beiden Differenzierungen.



> ⚠ **Hinweis:** In der Praxis reicht fast immer d = 1. Selten braucht man d = 2, und d ≥ 3 kommt praktisch nie vor. Wenn nach zweimaligem Differenzieren keine Stationarität erreicht ist, liegt das Problem vermutlich woanders (z. B. nicht-konstante Varianz → Log-Transformation).

Differenzierung Air-Passengers: Log-Transformation wegen wachsender Amplitude



Der Parameter d in ARIMA(p,d,q)

Der Buchstabe **I** in ARIMA steht für „Integrated“ (integriert). Das bezieht sich darauf, dass die originale Reihe durch Aufsummierung (Integration) der differenzierten Reihe wiederhergestellt werden kann. Der Parameter **d** gibt an, wie oft differenziert werden muss, bis die Reihe stationär ist.

> 📌 **Merksatz:** Differenzierung entfernt Trends durch Bildung der Differenz aufeinanderfolgender Werte. Der Parameter d in ARIMA(p,d,q) gibt die Anzahl der nötigen Differenzierungen an. Meistens gilt: d = 1.

6) Autoregression: Vergangenheit als Prognose (das AR in ARIMA)

Die ACF hat uns gezeigt, dass in vielen Zeitreihen eine echte Abhängigkeit zwischen aufeinanderfolgenden Werten besteht. Die **Autoregression (AR)** nutzt diese Abhängigkeit für Prognosen.

Die Grundidee

Ein AR-Modell sagt: „Der nächste Wert hängt von den vorherigen Werten ab.“ Das ist wie ein Pendel mit Trägheit — der aktuelle Zustand wird stark vom letzten Zustand beeinflusst.

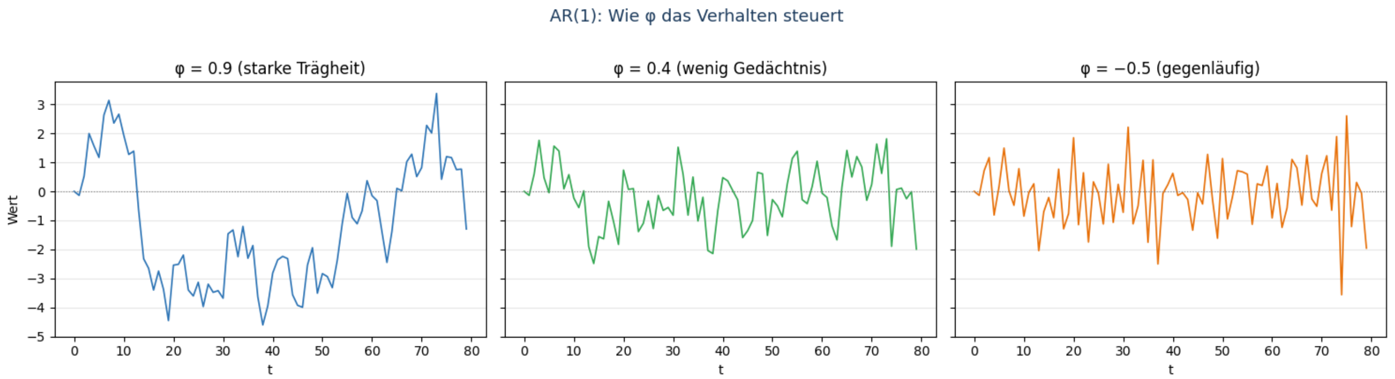
AR(1) — das einfachste Modell

$$x_t = \varphi \cdot x_{t-1} + \varepsilon_t$$

Dabei ist **φ** (phi) der Autoregressive Koeffizient und $\boldsymbol{\varepsilon_t}$ ein zufälliger Fehler (Rauschen). Der aktuelle Wert ist also eine gewichtete Version des vorherigen Wertes plus etwas Zufall.

Interpretation von φ

Wert von φ	Bedeutung
φ nahe 1 (z. B. 0,9)	Starke Trägheit: Der aktuelle Wert liegt nah am vorherigen. Die Reihe ändert sich langsam.
φ nahe 0 (z. B. 0,1)	Wenig Gedächtnis: Der vorherige Wert hat kaum Einfluss. Die Reihe springt.
φ negativ (z. B. −0,5)	Gegenläufig: Auf einen hohen Wert folgt tendenziell ein niedriger.
|φ| ≥ 1	Nicht stationär! Die Reihe explodiert oder oszilliert immer stärker.



Exkurs: Warum „wandert“ eine AR(1)-Zeitreihe bei hohem φ?

Auf den ersten Blick ist die AR(1)-Formel einfach: Der nächste Wert ist ein Bruchteil des letzten plus ein zufälliger Schock. Aber das erklärt noch nicht, warum eine AR(1)-Reihe mit $\varphi = 0{,}9$ wie ein Aktienkurs aussieht — mit langen Phasen, in denen sie in eine Richtung driftet. Das Geheimnis liegt in der **Aufstapelung vergangener Schocks**.

Die Rekursion aufrollen

Statt immer nur einen Schritt zurückzuschauen, setzen wir die AR(1)-Formel wiederholt in sich selbst ein:

$$x_1 = \varphi \cdot x_0 + \varepsilon_1$$

$$x_2 = \varphi \cdot x_1 + \varepsilon_2 = \varphi^2 \cdot x_0 + \varphi \cdot \varepsilon_1 + \varepsilon_2$$

$$x_3 = \varphi^3 \cdot x_0 + \varphi^2 \cdot \varepsilon_1 + \varphi \cdot \varepsilon_2 + \varepsilon_3$$

Das allgemeine Muster lautet:

$$x_t = \varepsilon_t + \varphi \cdot \varepsilon_{t-1} + \varphi^2 \cdot \varepsilon_{t-2} + \varphi^3 \cdot \varepsilon_{t-3} + \ldots$$

Der aktuelle Wert ist eine gewichtete Summe **aller** bisherigen Schocks. Die Gewichte sind die Potenzen von φ. Und genau diese Gewichte machen den Unterschied:

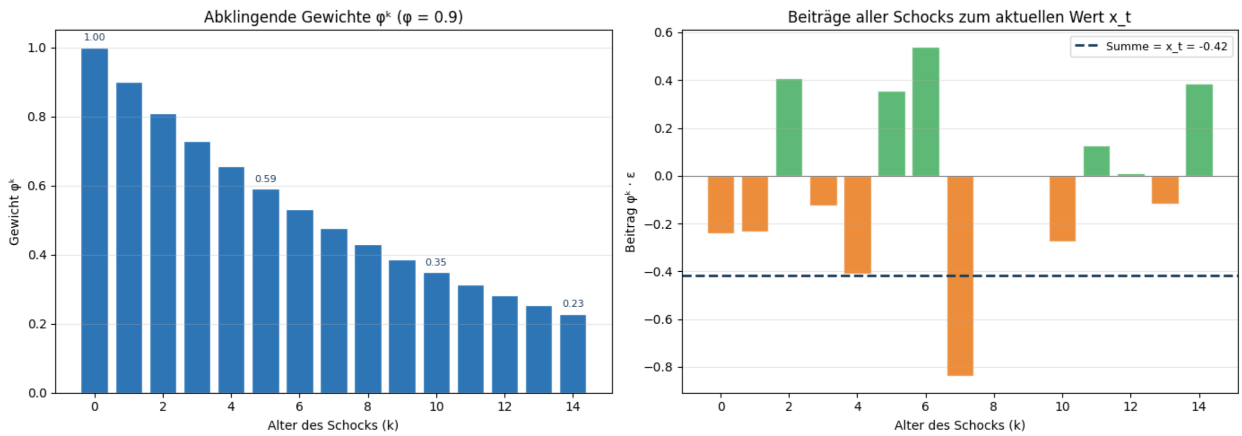
Alter des Schocks	Gewicht bei φ = 0,9	Gewicht bei φ = 0,4
Aktuell (ε_t)	1,000	1,000
1 Schritt her	0,900	0,400
2 Schritte her	0,810	0,160
3 Schritte her	0,729	0,064
5 Schritte her	0,590	0,010
10 Schritte her	**0,349**	**0,000**
20 Schritte her	**0,122**	≈ 0

Bei $\varphi = 0{,}4$ ist ein Schock nach drei Schritten praktisch vergessen (Gewicht 0,064). Bei $\varphi = 0{,}9$ wirkt derselbe Schock nach zehn Schritten noch mit einem Drittel seiner Stärke nach.

Warum die Reihe „wandert“: Die Aufstapelung

Zufällig kommen manchmal drei, vier positive Schocks hintereinander. Jeder einzelne wäre harmlos. Aber bei $\varphi = 0{,}9$ wirken die alten noch nach — sie **stapeln sich auf**. Die Reihe driftet vom Mittelwert weg. Und weil die Gewichte so langsam abklingen, dauert es lange, bis dieser aufgestapelte Effekt abgebaut ist — selbst wenn danach neutrale oder negative Schocks kommen.

Das ist wie ein Tanker auf See: Jede Welle (Schock) gibt nur einen kleinen Stoß, aber weil der Tanker so träge ist (φ nahe 1), richten ihn mehrere Wellen aus derselben Richtung nachhaltig aus. Die Kursänderung dauert — auch wenn der Wind sich längst gedreht hat.



Der Grenzfall: φ = 1 (Random Walk)

Bei φ = 1,0 werden die Gewichte alle gleich 1 — nichts klingt mehr ab. Jeder Schock, der jemals passiert ist, wirkt ungedämpft bis heute. Das ist ein **Random Walk**: Die Reihe „erinnert" sich an alles und kehrt nie zum Mittelwert zurück. Genau deshalb ist φ = 1 die Grenze zur **Nicht-Stationarität**.

φ-Wert	Abklingverhalten	Charakter	Stationär?
0,4	Schnell (nach ~3 Schritten weg)	Springende, nervöse Reihe	Ja ✓
0,9	Langsam (nach ~20 Schritten weg)	Träge, wandernde Reihe	Ja, aber knapp ✓
1,0	Gar nicht (ewiges Gedächtnis)	Random Walk	Nein ✗

> ⚠ **Hinweis:** Und hier schließt sich der Kreis zur ACF: Die langsam abklingende ACF bei AR-Prozessen hat genau dieselbe Ursache. Weil alte Schocks noch nachwirken, korreliert der aktuelle Wert auch mit weit zurückliegenden Werten — die ACF klingt so langsam ab wie die Gewichte $φ^k$.

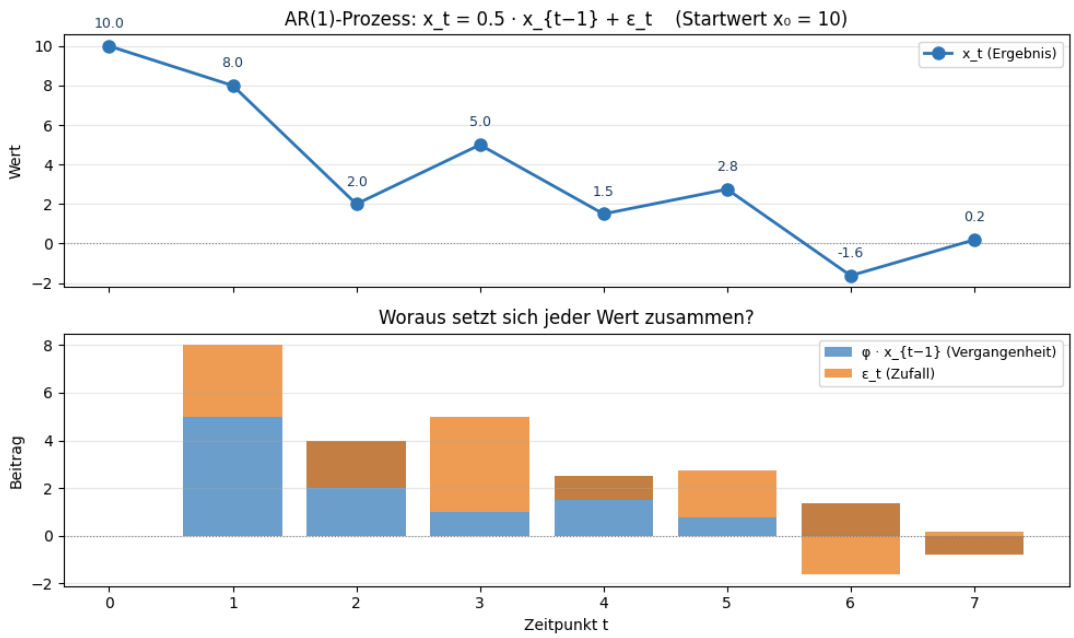
> 📌 **Merksatz:** Ein AR(1)-Prozess ist eine gewichtete Summe aller vergangenen Schocks mit exponentiell abklingenden Gewichten $φ^k$. Je näher φ an 1 liegt, desto länger wirken Schocks nach, desto träger wandert die Reihe — und desto langsamer klingt die ACF ab. Bei φ = 1 ist die Grenze zur Nicht-Stationarität erreicht.

Beispielrechnung: AR(1) mit φ = 0,5

Gegeben: Startwert x_0 = 10, Koeffizient φ = 0,5, und zufällige Fehler $ε_1$ = 3, $ε_2$ = −2, $ε_3$ = 4, $ε_4$ = −1.

Zeitpunkt	Rechnung	Ergebnis
t = 1	0,5 × 10 + 3 = 5 + 3	x_1 = 8
t = 2	0,5 × 8 + (−2) = 4 − 2	x_2 = 2
t = 3	0,5 × 2 + 4 = 1 + 4	x_3 = 5
t = 4	0,5 × 5 + (−1) = 2,5 − 1	x_4 = 1,5

Beobachtung: Jeder neue Wert „erbt" die Hälfte seines Vorgängers und bekommt dann noch einen zufälligen Stoß dazu. Das Gedächtnis ist kurz: Nach wenigen Schritten hat ein Wert kaum noch Einfluss.



AR(p) — mehrere Schritte zurück

$$x_t = \varphi_1 \cdot x_{t-1} + \varphi_2 \cdot x_{t-2} + \dots + \varphi_p \cdot x_{t-p} + \varepsilon_t$$

Bei AR(2) hängt der aktuelle Wert von den **zwei** vorherigen ab. Bei AR(p) von den **p** vorherigen. Der Parameter **p** in ARIMA(p,d,q) gibt genau diese Tiefe an.

> 📌 **Merksatz:** Bei der Autoregression AR(p) wird der aktuelle Wert als gewichtete Summe der p vorherigen **Werte** plus Rauschen modelliert. Je größer |φ|, desto stärker die Trägheit der Reihe.

7) Moving Average: die Nachwirkung von Schocks (MA in ARIMA)

Während AR auf vergangene **Werte** schaut,
schaut MA auf vergangene **Fehler**. Das ist ein fundamentaler Unterschied.

Die Grundidee: Die klingende Glocke

Stellen Sie sich eine Glocke vor, die einen Schlag erhält. Der Klang hallt nach — beim ersten Zeitpunkt laut, dann immer leiser. Ein MA-Modell beschreibt genau dieses Nachhallen: Ein zufälliger Schock („Fehler") wirkt nicht nur im Moment seines Auftretens, sondern klingt über die nächsten Zeitschritte nach.

MA(1) — das einfachste Modell

$$x_t = \mu + \varepsilon_t + \theta \cdot \varepsilon_{t-1}$$

Dabei ist **μ** der Mittelwert der Reihe, $\boldsymbol{\varepsilon_t}$ der aktuelle Fehler (Schock) und **θ** (theta) der Moving-Average-Koeffizient, der bestimmt, wie stark der vorherige Fehler nachwirkt.

Der entscheidende Unterschied zu AR

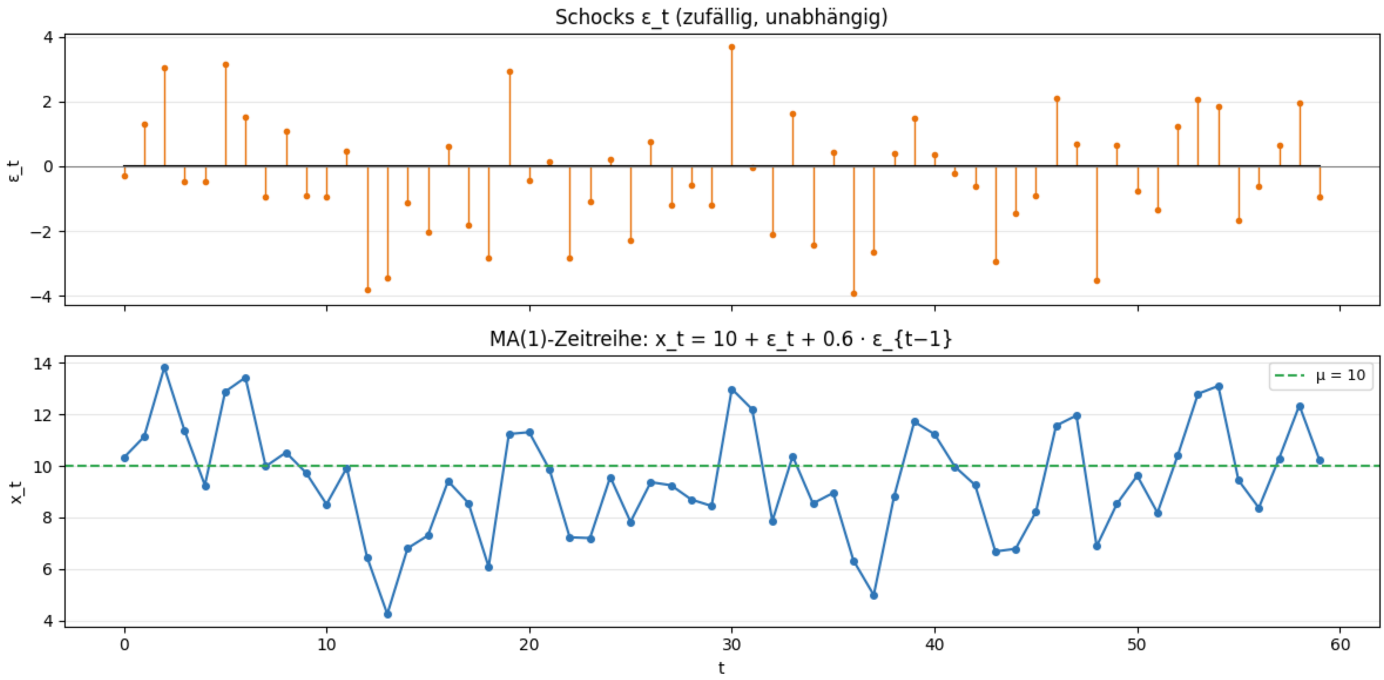
	AR: Regression auf Werte	MA: Regression auf Fehler
Formel	$x_t = \varphi \cdot x_{t-1} + \varepsilon_t$	$x_t = \mu + \varepsilon_t + \theta \cdot \varepsilon_{t-1}$
Schaut auf ...	... vergangene **Werte** der Reihe	... vergangene **Fehler** (Residuen)
Gedächtnis	Lang — klingt langsam ab	Kurz — bricht nach q Schritten ab
Metapher	Pendel mit Trägheit	Glocke, die nachhallt

Beispielrechnung: MA(1) mit μ = 10, θ = 0,6

Gegeben: Mittelwert μ = 10, Koeffizient θ = 0,6, und zufällige Fehler $\varepsilon_0 = 0$, $\varepsilon_1 = 3$, $\varepsilon_2 = -2$, $\varepsilon_3 = 4$.

Zeitpunkt	Rechnung	Ergebnis
t = 1	10 + 3 + 0,6 × 0 = 10 + 3 + 0	$x_1 = 13$
t = 2	10 + (−2) + 0,6 × 3 = 10 − 2 + 1,8	$x_2 = 9{,}8$
t = 3	10 + 4 + 0,6 × (−2) = 10 + 4 − 1,2	$x_3 = 12{,}8$

Beobachtung bei t = 2: Der positive Schock von t = 1 ($\varepsilon_1 = 3$) hallt noch nach (Beitrag: +1,8), obwohl der neue Schock negativ ist. Bei t = 3 hallt der negative Schock von t = 2 nach. Das Gedächtnis reicht aber immer nur einen Schritt zurück — typisch für MA(1).

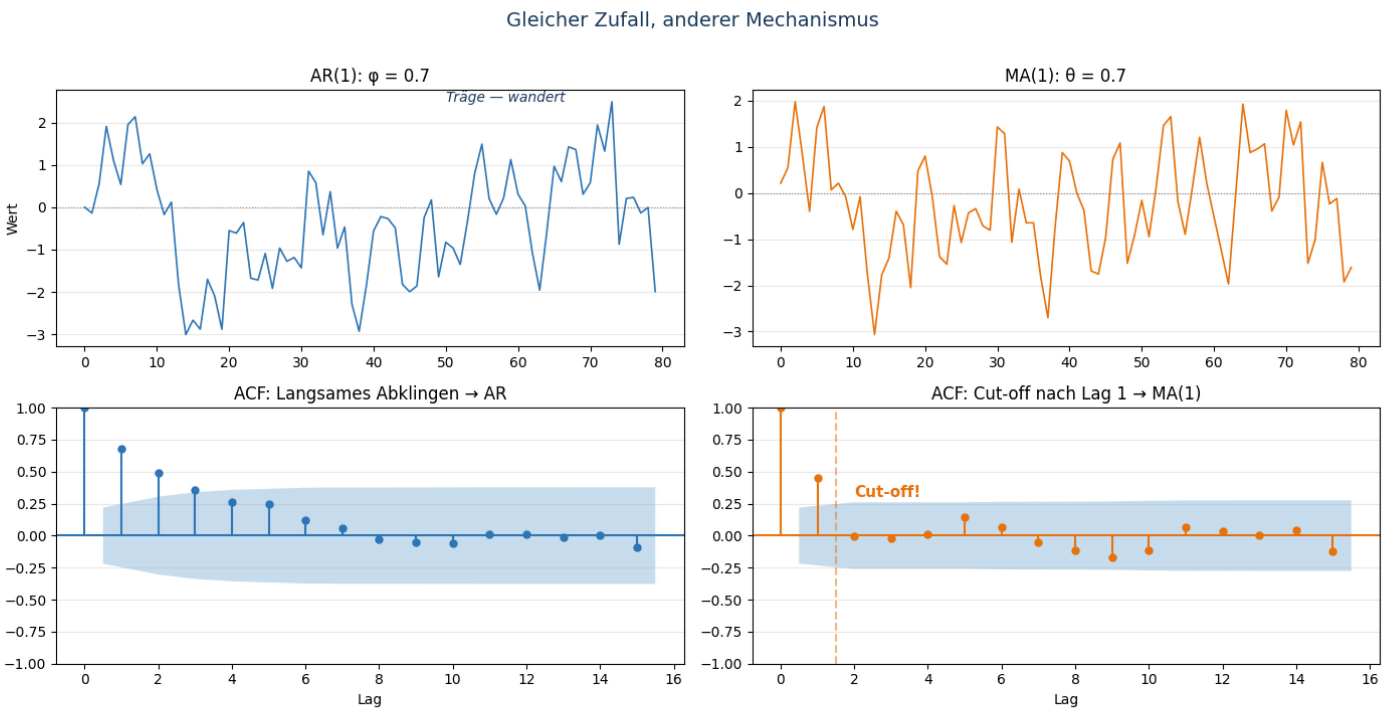


MA(q) — mehrere Schritte nachhallen

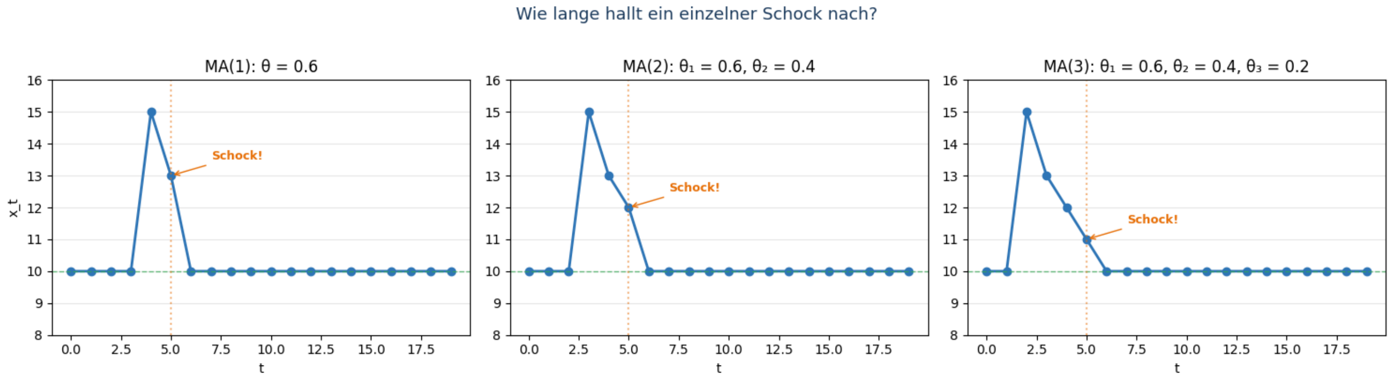
$$x_t = \mu + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \varepsilon_{t-2} + \ldots + \theta_n \cdot \varepsilon_{t-n}$$

Der Parameter **q** in ARIMA(p,d,q) gibt an, über wie viele vergangene Fehler sich die Nachwirkung erstreckt.

> 📌 **Merksatz:** Beim Moving-Average-Modell MA(q) wird der aktuelle Wert aus dem Mittelwert plus dem aktuellen Fehler plus den gewichteten q vergangenen **Fehlern** berechnet. AR schaut auf vergangene Werte, MA auf vergangene Fehler — das ist der Schlüsselunterschied.



Ein einzelner isolierter Schock trifft auf MA(1), MA(2) und MA(3). Man sieht direkt, wie die Glocke bei MA(1) nur einen Zeitschritt nachklingt, bei MA(2) zwei, bei MA(3) drei — und dann ist Ruhe. Das macht den „Cut-off" in der ACF visuell greifbar.



8) PACF und Modellwahl: p und q bestimmen

Wir kennen nun AR und MA als zwei verschiedene Modellierungsansätze. Aber wie entscheiden wir, welchen wir brauchen — und mit welcher Ordnung? Dafür gibt es neben der ACF ein zweites Werkzeug: die **partielle Autokorrelationsfunktion (PACF)**.

ACF vs. PACF

Die ACF aus Block 4 zeigt die **gesamte** Korrelation zwischen x_t und x_{t-k}. Das Problem: Bei Lag 3 enthält die ACF auch die **indirekten** Effekte, die über Lag 1 und Lag 2 laufen. Ein Wert kann stark mit seinem Vorgänger korrelieren, der wiederum stark mit seinem Vorgänger korreliert — und so entsteht auch bei Lag 3 eine hohe Korrelation, obwohl es keinen **direkten** Effekt gibt.

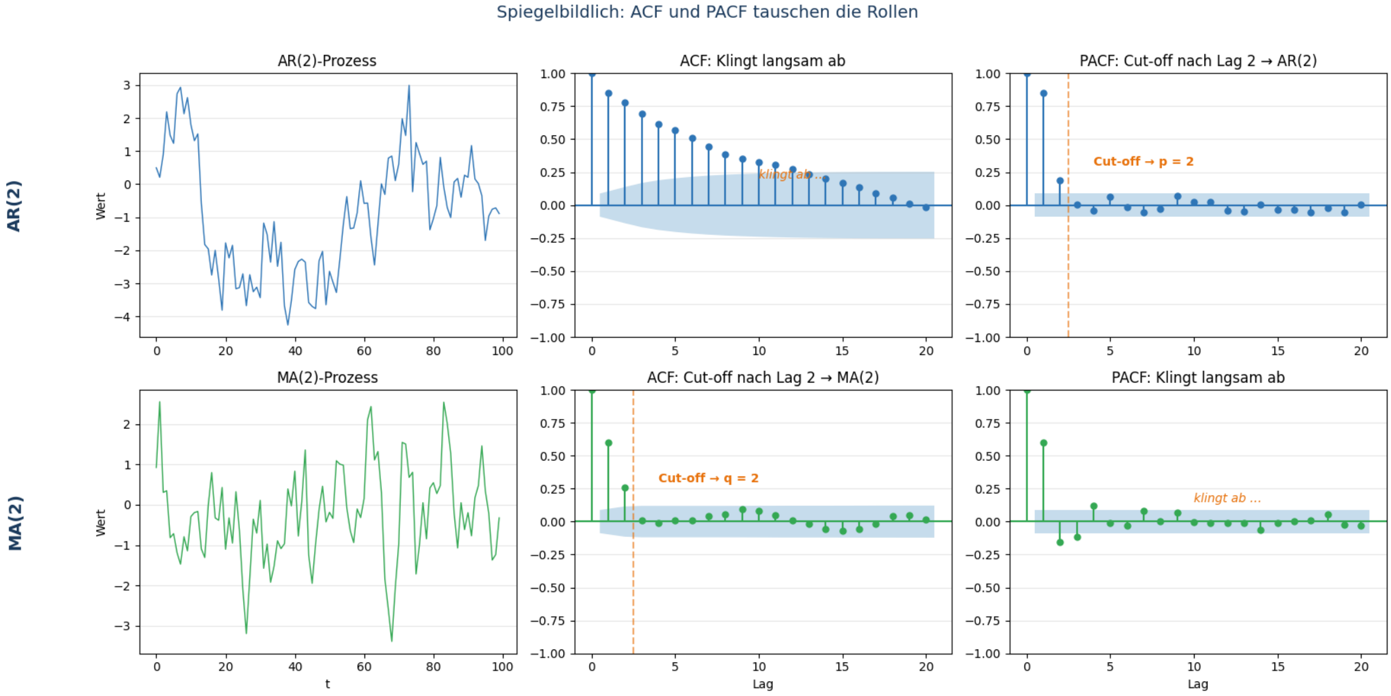
Die PACF filtert diese indirekten Effekte heraus. Sie zeigt nur die **direkte** Korrelation zwischen x_t und x_{t-k}, nachdem der Einfluss aller dazwischenliegenden Werte herausgerechnet wurde.

> 📐 **Beispiel:** Stellen Sie sich eine Stille-Post-Kette vor. Die ACF misst, wie ähnlich die Nachricht am Anfang und am Ende ist (Gesamteffekt). Die PACF misst nur, wie gut eine einzelne Person an ihren direkten Nachbarn weitergibt (direkter Effekt).

Die Entscheidungstabelle

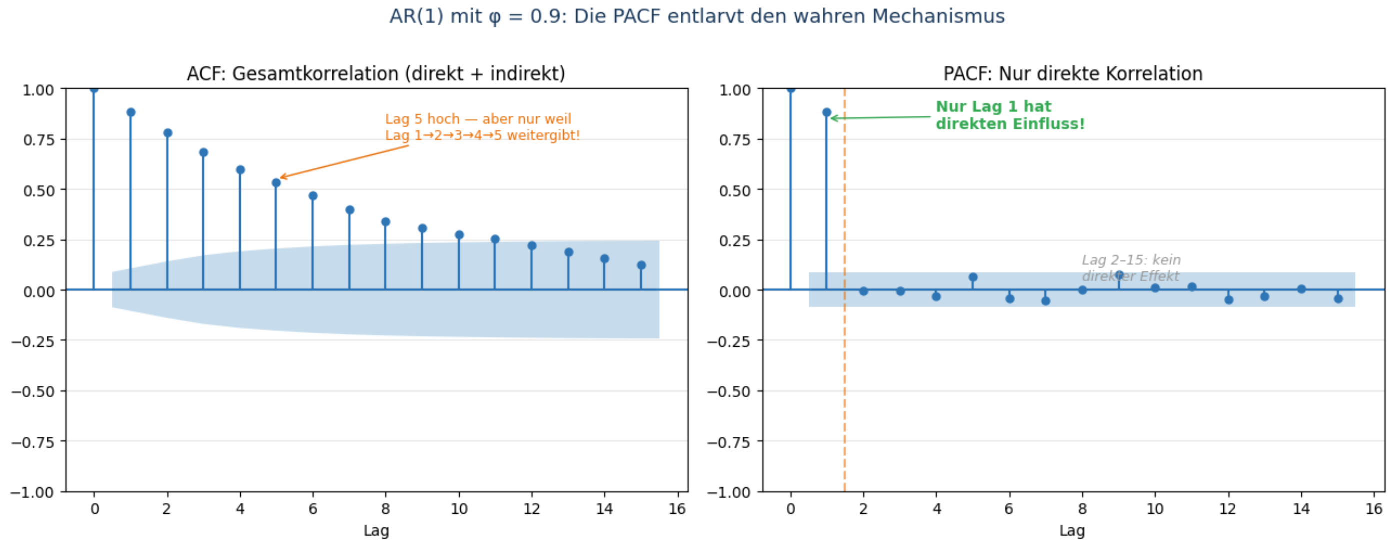
Das Zusammenspiel von ACF und PACF ermöglicht eine systematische Modellwahl:

ACF-Verhalten	PACF-Verhalten	Modell	Interpretation
Klingt langsam ab	Bricht nach Lag p ab	AR(p)	p Lags haben direkten Einfluss
Bricht nach Lag q ab	Klingt langsam ab	MA(q)	Schocks hallen q Schritte nach
Beide klingen langsam ab	Beide klingen langsam ab	ARMA(p,q)	Kombination beider Effekte



Wie liest man „bricht ab“?

„Abbruch“ (englisch: Cut-off) bedeutet: Bis zu einem bestimmten Lag sind die Werte signifikant (Balken ragen über das Konfidenzband), danach sind sie schlagartig nicht mehr signifikant. „Langsames Abklingen“ (englisch: Tailing off) bedeutet: Die Werte werden zwar kleiner, bleiben aber noch über längere Lags signifikant.



Die ACF zeigt hohe Korrelation bei Lag 5, 6, 7 ... und man könnte denken, da gibt es überall direkte Abhängigkeiten. Die PACF daneben zeigt: Nein, nur Lag 1 zählt wirklich — alles andere ist Stille Post.

> 📌 **Merksatz:** Die PACF zeigt die direkte Korrelation pro Lag (ohne indirekte Effekte). Zusammen mit der ACF ergibt sich eine klare Entscheidungsregel: ACF bricht ab → MA. PACF bricht ab → AR. Beide klingen ab → ARMA.

9) ARIMA(p,d,q): der Baukasten

Jetzt fügen wir alle Puzzleteile zusammen. ARIMA ist kein einzelner Algorithmus, sondern ein **Baukastensystem** aus drei Komponenten, die wir einzeln kennengelernt haben.

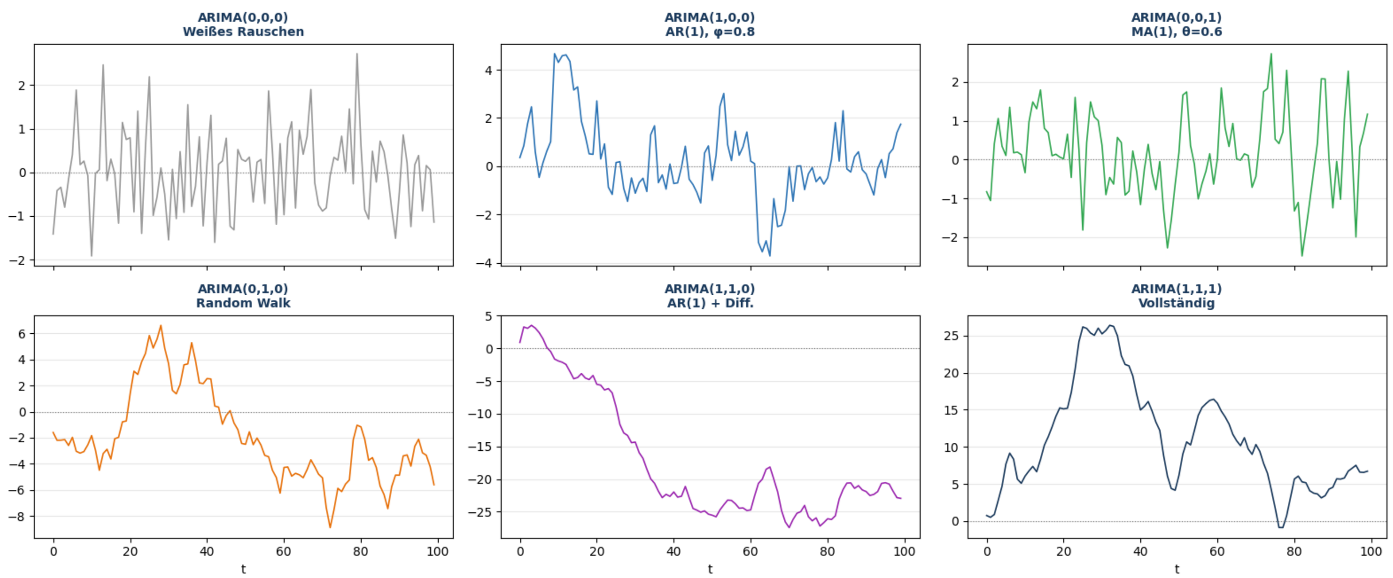
Die drei Buchstaben

Buchstabe	Steht für	Parameter	Funktion	Block
AR	AutoRegressive	p	Wie viele vergangene **Werte** fließen ein?	6
I	Integrated	d	Wie oft muss differenziert werden?	5
MA	Moving Average	q	Wie viele vergangene **Fehler** fließen ein?	7

Bekannte Spezialfälle

Notation	Bedeutung	Beschreibung
ARIMA(1,0,0)	= AR(1)	Reines AR-Modell 1. Ordnung, Reihe ist bereits stationär
ARIMA(0,0,1)	= MA(1)	Reines MA-Modell 1. Ordnung, Reihe ist bereits stationär
ARIMA(0,1,0)	= Random Walk	Nur Differenzierung, keine AR/MA-Modellierung
ARIMA(1,1,0)	= AR(1) mit Diff.	Ein Schritt differenziert, dann AR(1) auf die Differenzen
ARIMA(1,1,1)	= vollständig	Differenzierung + Autoregression + Moving Average
ARIMA(0,0,0)	= Weißes Rauschen	Keine Struktur — nur Zufall

Der ARIMA-Zoo: Jede Kombination hat ihren eigenen Charakter



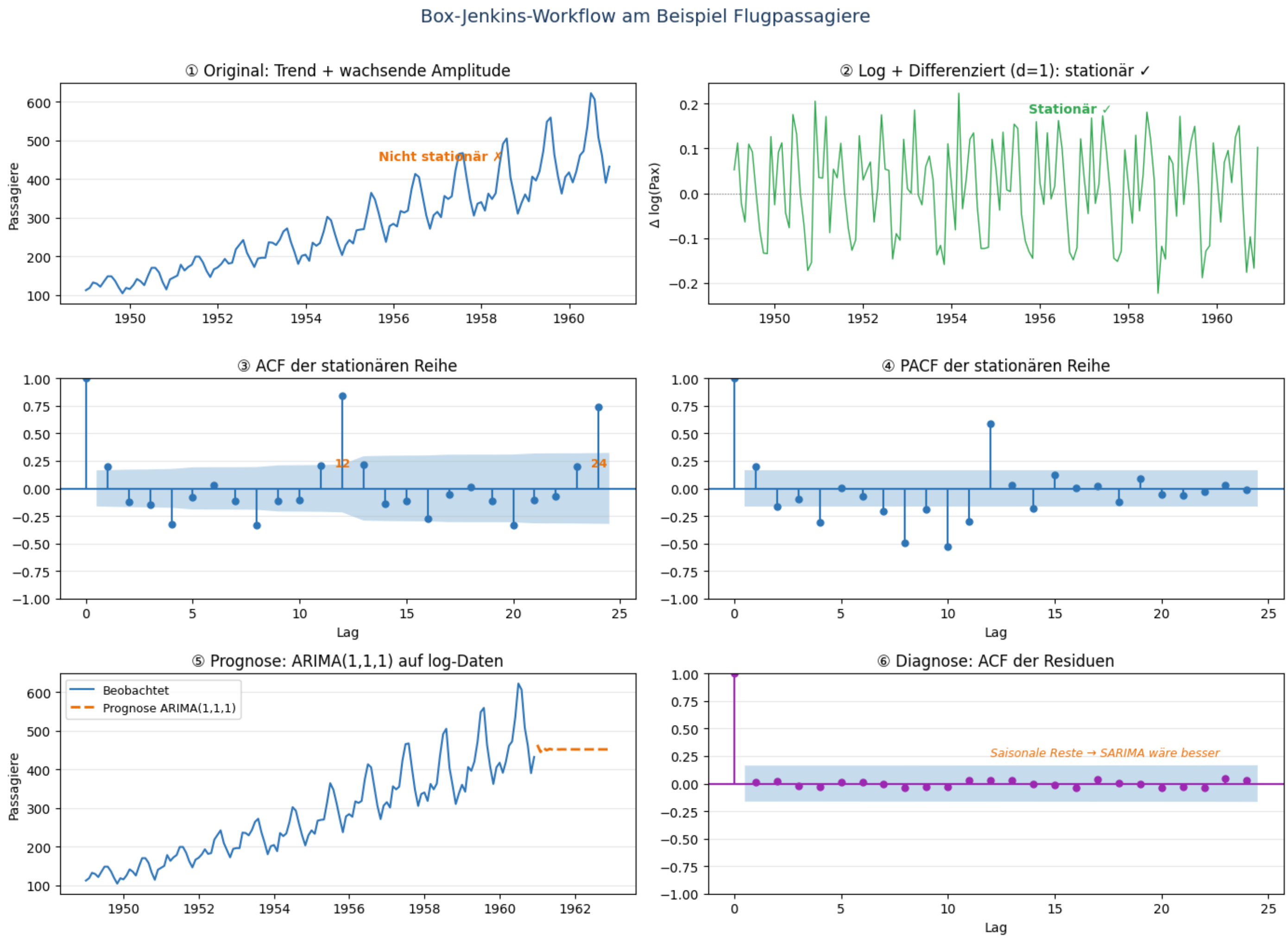
Der Arbeitsablauf in der Praxis

Die systematische Vorgehensweise (oft **Box-Jenkins-Methode** genannt) umfasst drei Schritte:

1. Identifikation: Zeitreihe plotten, auf Trend/Saisonalität prüfen. Wenn nötig differenzieren (d bestimmen). ACF und PACF der stationären Reihe analysieren, um p und q zu bestimmen.

2. Schätzung: Das gewählte ARIMA(p,d,q)-Modell auf die Daten anpassen. Die Koeffizienten φ und θ werden vom Algorithmus geschätzt (z. B. per Maximum-Likelihood).

3. Diagnose: Die Residuen des Modells prüfen. Sind sie reines Rauschen (keine Muster, keine Autokorrelation)? Wenn ja, ist das Modell gut. Wenn nein: zurück zu Schritt 1 mit einem anderen Modell.



> ⚠ **Hinweis:** In der Praxis gibt es auch automatische Verfahren wie auto.arima (R) oder pmdarima (Python), die verschiedene (p,d,q)-Kombinationen durchprobieren und die beste wählen. Das ersetzt aber nicht das Verständnis der Bausteine!

Ausblick: Wohin geht die Reise?

ARIMA ist die Basis — darauf aufbauend gibt es wichtige Erweiterungen:

SARIMA (Seasonal ARIMA): Ergänzt saisonale Parameter. Notation: $\text{ARIMA}(p,d,q)(P,D,Q)_t$ — wobei die großen Buchstaben die saisonalen Gegenstücke sind und m die Periode.

Holt-Winters: Ein alternatives Verfahren (aus den Vorlesungsfolien bekannt), das Niveau, Trend und Saisonalität rekursiv schätzt. Besonders gut geeignet für Prognosen mit veränderlicher Saisonalität.

Exponentielles Glätten: Eine Familie von Verfahren, zu der auch Holt-Winters gehört. Neuere Beobachtungen werden stärker gewichtet als ältere.

> 📌 **Merksatz:** ARIMA(p,d,q) kombiniert drei Bausteine: Autoregression (p Schritte zurück auf Werte), Differenzierung (d-mal für Stationarität) und Moving Average (q Schritte zurück auf Fehler). Der Box-Jenkins-Ablauf: Identifizieren → Schätzen → Diagnostizieren.

Schnellreferenz: ACF + PACF → Modellwahl

ACF	PACF	→ Modell
Klingt langsam ab	Cut-off nach p	AR(p)
Cut-off nach q	Klingt langsam ab	MA(q)
Beide klingen ab	Beide klingen ab	ARMA(p,q)

10) Prognose: Was ARIMA leisten kann — und was nicht

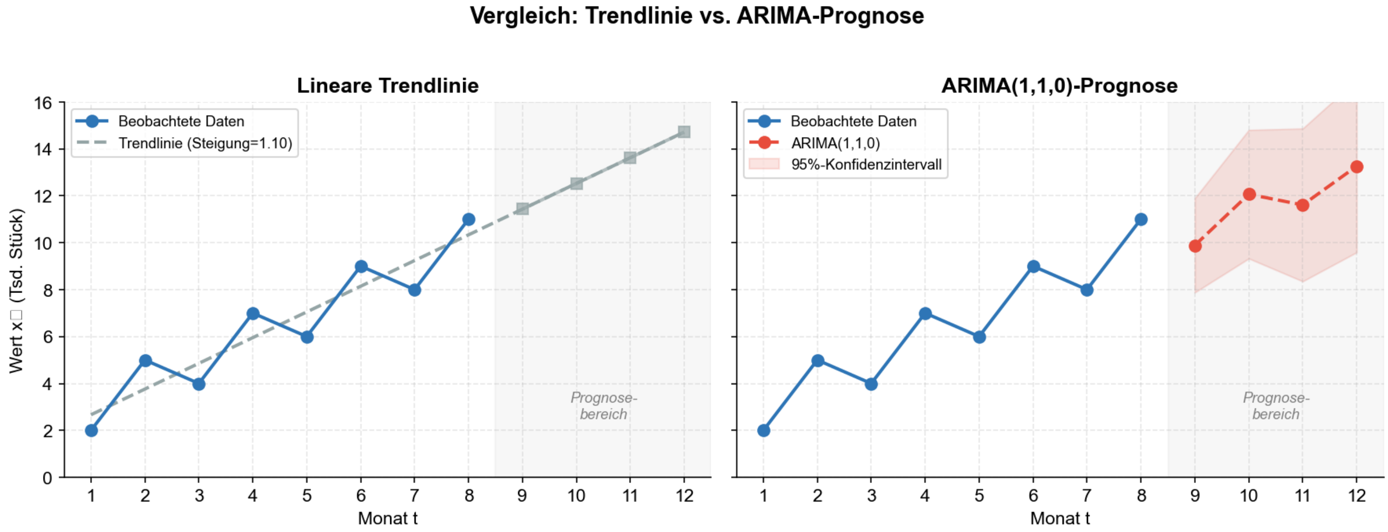
Wir haben den gesamten ARIMA-Baukasten kennengelernt: Dekomposition, Stationarität, Differenzierung, AR, MA, ACF, PACF und die Box-Jenkins-Methode. Aber eine zentrale Frage steht noch aus: **Wozu das Ganze? Welchen konkreten Wert hat ein ARIMA-Modell für die Vorhersage künftigen Verhaltens?**

Was ARIMA über die Zukunft sagen kann

Ein ARIMA-Modell macht keine *kausalen* Aussagen. Es weiß nicht, *warum* der Umsatz steigt — nur *dass* er es tut und *wie* sich die jüngste Dynamik anfühlt. Die Prognosekraft entsteht dadurch, dass ARIMA drei verschiedene Arten von Struktur in den Daten erkennt und für die Zukunft fortschreibt:

Komponente	Was sie erfasst	Prognose-Beitrag
AR (Trägheit)	Der aktuelle Wert hängt von vergangenen Werten ab (Pendel-Metapher)	Extrapoliert die kurzfristige Drift: „Wenn wir gerade steigen, steigen wir vermutlich noch ein Stück weiter.“
I (Differenzierung)	Ein Trend liegt vor, der die statistischen Eigenschaften verändert	Trennt die Grundrichtung von der Dynamik, damit die Prognose nicht auf verschobenen Grundlagen fußt
MA (Schock-Nachhall)	Vergangene Überraschungen wirken noch nach (Glocken-Metapher)	Berücksichtigt, ob ein jüngster Ausreißer noch nachwirkt oder bereits abgeklungen ist

Zusammen ergibt sich eine Vorhersage, die deutlich reicher ist als eine simple Trendlinie: ARIMA berücksichtigt die **kurzfristige Dynamik** (Schwingungen, Trägheit, Nachhalleffekte), die eine Trendlinie schlicht ignoriert.



Beispielrechnung: Ein-Schritt-Prognose mit unserer Zeitreihe

Zurück zu unserer Beispielzeitreihe mit 8 Werten. Wir haben in Block 5 gesehen, dass die differenzierte Reihe (d = 1) das Muster **3, −1, 3, −1, 3, −1, 3** ergibt. Nun nehmen wir vereinfachend ein **AR(1)-Modell auf die differenzierte Reihe** an — also ein ARIMA(1,1,0).

Angenommen, die Schätzung ergibt $\varphi = -0{,}8$ (der stark negative Wert spiegelt das Alternieren wider) und der Mittelwert der differenzierten Reihe ist $\mu \approx 1{,}3$.

Prognose für den nächsten Differenzwert:

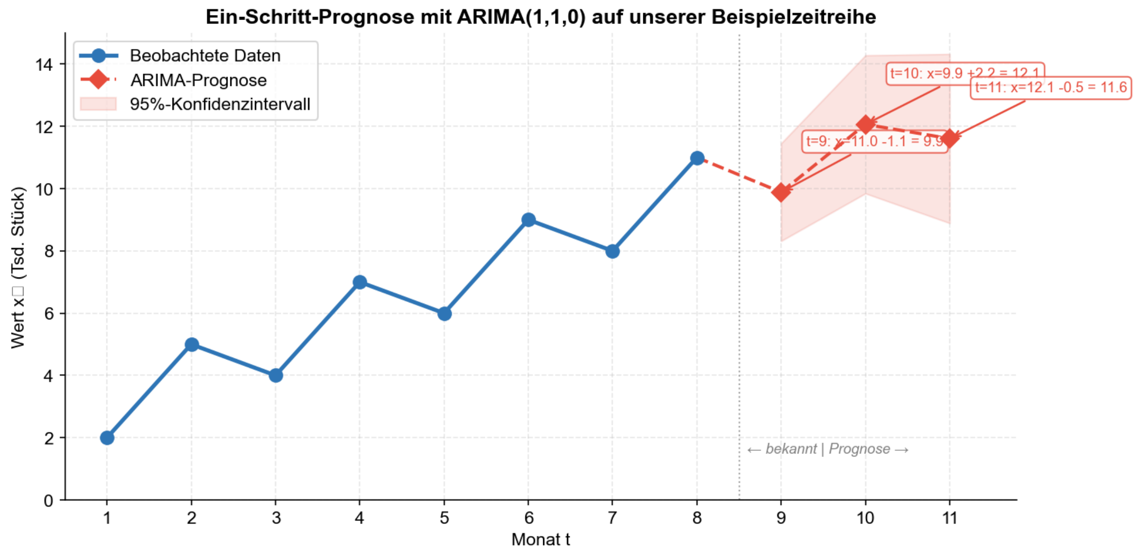
$$x'_9 = \mu + \phi \cdot x'_8 = 1{,}3 + (-0{,}8) \cdot 3 = 1{,}3 - 2{,}4 = -1{,}1$$

Rück-Transformation (Differenz rückgängig machen):

$$x_9 = x_8 + x'_9 = 11 + (-1{,}1) = 9{,}9$$

Die ARIMA-Prognose für Monat 9 lautet also **≈ 9,9 Tausend Stück**. Das Modell hat erkannt, dass nach einem hohen Differenzwert (+3) typischerweise ein niedrigerer folgt — es bildet das Zickzack-Muster ab.

Eine naive Trendlinie würde dagegen einfach den linearen Aufwärtstrend fortschreiben und käme auf ca. 12–13 — deutlich höher, weil sie das Alternieren ignoriert.

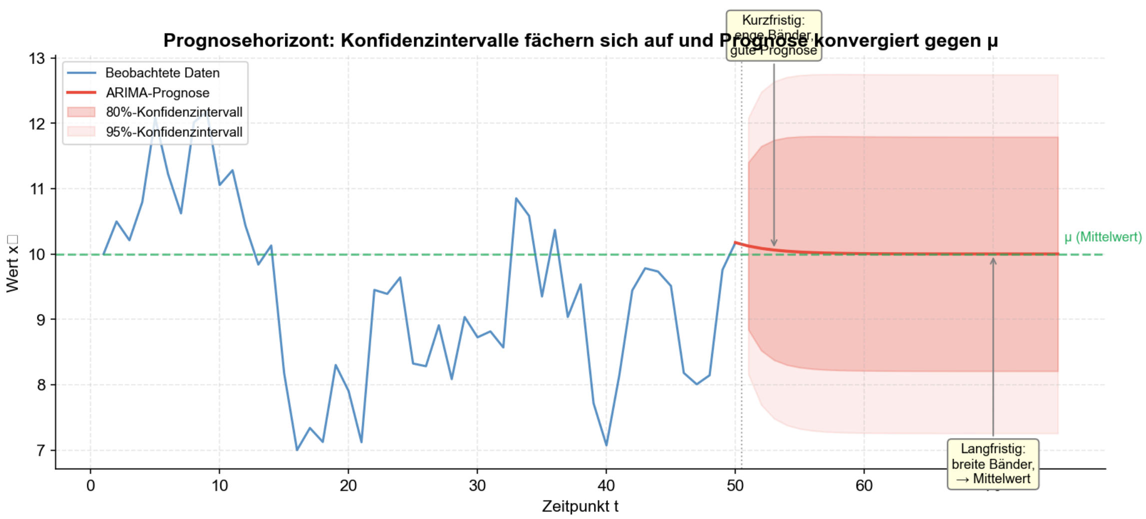


Der Prognosehorizont: Warum ARIMA kurzfristig stark ist

Ein zentrales Merkmal von ARIMA-Prognosen ist, dass sie mit zunehmendem Horizont an Schärfe verlieren. Das liegt an der Natur der AR- und MA-Komponenten:

Horizont	AR-Beitrag	MA-Beitrag	Prognosequalität
1 Schritt	Nutzt den letzten bekannten Wert direkt	Nutzt den letzten bekannten Fehler	Am besten — maximale Information
2–5 Schritte	Nutzt eigene Prognosen als Input („auf Sand gebaut“)	Fehler werden auf 0 gesetzt (unbekannt)	Konfidenzintervall wächst schnell
Viele Schritte	Konvergiert gegen den Mittelwert	Kein Beitrag mehr	Prognose ≈ Mittelwert ± große Unsicherheit

Anschaulich: Das ARIMA-Modell „vergisst“ mit jedem Prognoseschritt ein Stück Vergangenheit. Nach genügend Schritten sagt es nur noch: „Ich erwarte den langfristigen Durchschnitt.“ Das ist kein Defekt — es ist **ehrliche Bescheidenheit**. Das Modell gibt zu, wo sein Wissen endet.



Einordnung: ARIMA vs. andere Ansätze

Um den Wert von ARIMA zu verstehen, hilft ein Vergleich mit einfacheren und komplexeren Alternativen:

Methode	Idee	Stärke	Schwäche
Naive Prognose	Morgen = Heute	Braucht kein Modell	Ignoriert jede Dynamik
Lineare Trendlinie	Gerade durch die Daten legen	Einfach, schön interpretierbar	Ignoriert Schwingungen und Schock-Effekte
ARIMA	Trägheit + Schock-Nachhall + Trend-Bereinigung	Erfasst kurzfristige Dynamik; formale Modellwahl	Nur univariat; keine externen Ursachen
Expon. Glätten / Holt-Winters	Gewichtete Mittel mit Trend + Saison	Ähnlich gut; einfache Implementierung	Weniger diagnostische Werkzeuge (ACF/PACF)
Machine-Learning (LSTM, Transformer)	Lernt beliebig komplexe Muster	Kann nichtlineare Zusammenhänge finden	Braucht viele Daten; „Black Box“; Überanpassung

ARIMA sitzt in einer produktiven Mitte: Es ist **mächtiger als einfache Trendverfahren** (weil es Dynamik modelliert), aber **transparenter als Black-Box-Modelle** (weil man p, d, q interpretieren kann und die Residuen diagnostisch prüfbar sind).

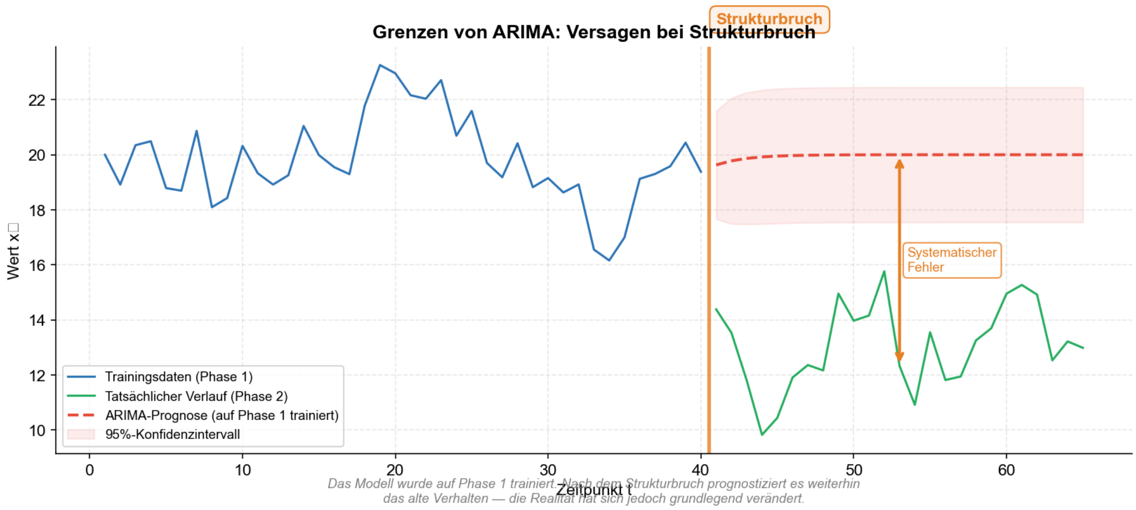
Grenzen: Wann ARIMA an seine Grenzen stößt

Ehrlichkeit über die Grenzen eines Modells ist Teil guter Modellierung. ARIMA hat drei strukturelle Limitationen:

1. Keine externen Treiber: ARIMA kennt nur die Zeitreihe selbst. Es kann nicht wissen, dass ein Umsatzsprung durch eine Marketing-Kampagne verursacht wurde. Für solche Fälle gibt es Erweiterungen wie ARIMAX (ARIMA mit exogenen Variablen) oder VAR (Vektorautoregression).

2. Linearität: ARIMA modelliert lineare Zusammenhänge zwischen vergangenen und zukünftigen Werten. Nichtlineare Dynamiken (z. B. Schwellwerteffekte, plötzliche Regimewechsel) werden nicht erfasst. Hier können GARCH-Modelle oder Machine-Learning-Ansätze ergänzen.

3. Stationaritäts-Annahme: Selbst nach Differenzierung muss die transformierte Reihe stationär sein. Wenn sich die Grundstruktur der Daten verändert (z. B. durch einen Pandemie-Schock), kann das Modell versagen, weil die „Spielregeln“ plötzlich andere sind.



Zusammenfassung: Wozu ARIMA?

ARIMA ist kein Orakel. Es beantwortet eine sehr spezifische, aber praktisch enorm relevante Frage:

> ***„Was passiert in den nächsten Schritten, wenn sich nichts Grundsätzliches ändert?“***

Diese Frage stellt sich täglich in der Praxis — bei Lagerbestandsplanung, Personalplanung, Budgetierung, Energiebedarfsprognosen oder Qualitätskontrolle. Und genau für diese Frage liefert ARIMA ein **transparentes, mathematisch fundiertes und diagnostisch überprüfbares** Werkzeug.

> 📌 **Merksatz:** ARIMA prognostiziert, indem es die innere Mechanik einer Zeitreihe — Trägheit (AR), Trend (I) und Schock-Nachhall (MA) — quantifiziert und fortschreibt. Die Prognose ist kurzfristig am stärksten und wird mit zunehmendem Horizont ehrlich unsicherer. ARIMA kennt keine Ursachen, nur Muster — aber genau diese Muster sind für kurzfristige Vorhersagen oft der wichtigste Hebel.

> ⚠ **Hinweis:** In der Praxis wird ARIMA häufig als Baseline-Modell eingesetzt: Bevor man komplexe Machine-Learning-Modelle baut, prüft man, ob ein einfaches ARIMA bereits gute Ergebnisse liefert. Wenn ja, ist das ML-Modell den Mehraufwand oft nicht wert. Wenn nein, hat man eine saubere Messlatte, gegen die man das komplexere Modell benchmarken kann.

Zusammenfassung — Cheat Sheet

Konzept	**Kernaussage**
Zeitreihe	Zeitlich geordnete Messwerte. Die Reihenfolge trägt Information.
Dekomposition	Zeitreihe = Trend + Saisonalität + Residuum (additiv oder multiplikativ)
Stationarität	Konstanter Mittelwert, konstante Varianz, Autokovarianz hängt nur vom Lag ab
ACF	Gesamtkorrelation über Lags. Cut-off nach q → MA(q)
Differenzierung (I)	Entfernt Trend durch Differenzbildung. d = Anzahl der Differenzierungen
Autoregression (AR)	Aktueller Wert = gewichtete Summe vergangener **Werte** + Rauschen
Moving Average (MA)	Aktueller Wert = Mittelwert + gewichtete Summe vergangener **Fehler**
PACF	Direkte Korrelation pro Lag. Cut-off nach p → AR(p)
ARIMA(p,d,q)	p = AR-Tiefe, d = Differenzierungen, q = MA-Tiefe
Prognose-Idee	*ARIMA extrapoliert die kurzfristige Dynamik (Trägheit + Nachhall), nicht bloß den Trend*
Prognosehorizont	*Kurzfristig stark, langfristig konvergiert die Vorhersage gegen den Mittelwert*
Konfidenzintervall	*Die Unsicherheit fächert sich mit jedem Schritt auf — das Modell zeigt ehrlich, wo sein Wissen endet*
Baseline-Rolle	*ARIMA dient oft als Referenz: Jedes komplexere Modell muss ARIMA erst schlagen*
Grenzen	*Keine Kausalität, nur lineare Muster, versagt bei Strukturbrüchen*

11 SARIMA: Saisonalität modellieren

Dieses Kapitel ist eine Erweiterung des ARIMA-Leitfadens. Es setzt die Blöcke 1–10 voraus und baut direkt auf dem *Ausblick* in Block 9 auf, der SARIMA bereits als nächsten Schritt angekündigt hat.
Als durchgängiges Beispiel verwenden wir die **Air-Passengers-Zeitreihe**, die uns aus den vorherigen Blöcken bereits bekannt ist — monatliche Flugpassagierzahlen von 1949 bis 1960 mit 144 Datenpunkten. Sie zeigt genau die beiden Phänomene, die ARIMA allein nicht greifen kann: einen steigenden Trend *und* ein klares saisonales Muster (Sommer-Hochs, Winter-Tiefs).

Warum reicht ARIMA nicht?
In Block 5 haben wir gesehen, dass die **erste Differenzierung (d = 1)** den Trend entfernt. Aber schauen wir uns an, was bei den Air Passengers danach übrig bleibt:

Schritt	Was passiert	Was bleibt übrig
Original-Reihe	Trend + Saisonalität + Rauschen	Alles
Nach d = 1	Trend ist weg	Saisonalität + Rauschen
Nach d = 2	Auch quadratischer Trend weg	Saisonalität + Rauschen (stärker verrauscht)

Das Problem: Egal wie oft wir differenzieren — die saisonalen Spitzen bei Lag 12, 24, 36 im ACF-Plot bleiben bestehen. Die gewöhnliche Differenzierung vergleicht einen Monat mit dem Vormonat. Aber Saisonalität bedeutet, dass der Juli 1957 dem Juli 1956 ähnelt — nicht dem Juni 1957.
Wir brauchen also ein Werkzeug, das auf einer zweiten Ebene arbeitet: nicht Schritt für Schritt, sondern **Saison für Saison**. Genau das leistet SARIMA.

> 📌 **Merksatz:** Die gewöhnliche Differenzierung (d) entfernt den Trend, aber nicht die Saisonalität. Dafür brauchen wir eine saisonale Differenzierung, die Werte mit dem gleichen Saisonpunkt des Vorjahres vergleicht.

Die SARIMA-Notation: ARIMA(p,d,q)(P,D,Q)m
SARIMA erweitert ARIMA um eine **zweite Parameterschicht**. Die Notation liest sich so:

Ebene	Param.	Bedeutung	Kapitel
Nicht-saisonal	p	AR-Ordnung (Abhängigkeit von vergangenen Werten)	Block 6
Nicht-saisonal	d	Anzahl gewöhnlicher Differenzierungen	Block 5
Nicht-saisonal	q	MA-Ordnung (Nachhall vergangener Fehler)	Block 7
Saisonal	**P**	Saisonale AR-Ordnung (Abhängigkeit vom gleichen Monat des Vorjahres)	**NEU**
Saisonal	**D**	Anzahl saisonaler Differenzierungen	**NEU**
Saisonal	**Q**	Saisonale MA-Ordnung (Nachhall saisonaler Fehler)	**NEU**
Periode	**m**	Länge eines Saisonzyklus (z. B. 12 für Monatsdaten, 4 für Quartalsdaten)	**NEU**

Die Kernidee: Die drei neuen Parameter P, D, Q tun auf der saisonalen Ebene (also bei den Lags m, 2m, 3m, ...) exakt dasselbe, was p, d, q auf der Einzelschritt-Ebene tun. Es ist derselbe Baukasten — nur auf einer langsameren Zeitskala.

Metapher: Die zwei Uhren
Stellen Sie sich eine Uhr mit zwei Zeigern vor. Der *Minutenzeiger* (p, d, q) erfasst die kurzfristige Dynamik von Monat zu Monat — Trägheit, kleine Schocks, lokale Trends. Der *Stundenzeiger* (P, D, Q) erfasst die langsame, wiederkehrende Dynamik von Saison zu Saison — warum der Juli immer ein Hoch hat und der Januar ein Tief. SARIMA liest beide Zeiger gleichzeitig ab.

> 📌 **Merksatz:** SARIMA = ARIMA + saisonale Schicht. Die Notation ARIMA(p,d,q)(P,D,Q)m enthält zwei Baukästen: einen für die kurzfristige Dynamik (Kleinbuchstaben) und einen für die saisonale Dynamik (Großbuchstaben). Der Index m gibt die Saisonperiode an.

Saisonale Differenzierung (D): Saisonalität entfernen
In Block 5 haben wir die gewöhnliche Differenzierung kennengelernt. Die **saisonale Differenzierung** funktioniert analog, aber mit einem Sprung über eine ganze Saison:

$$x^*_t = x_t - x_{t-m}$$

Bei monatlichen Daten (m = 12) zieht man also vom Juli 1957 den Juli 1956 ab, vom August 1957 den August 1956, und so weiter. Dadurch wird die regelmäßige saisonale Schwankung eliminiert.
Beispielrechnung mit Air Passengers
Hier ein Ausschnitt der Air-Passengers-Daten (Passagierzahlen in Tausend):

Monat	Jan	Feb	Mär	Apr	Mai	Jun
1949	112	118	132	129	121	135
1950	115	126	141	135	125	149

Saisonale Differenzierung (m = 12): Wir ziehen von jedem Monat 1950 den gleichen Monat 1949 ab:

Schritt	Rechnung	Ergebnis x*t
Jan 1950	115 − 112	3
Feb 1950	126 − 118	8
Mär 1950	141 − 132	9
Apr 1950	135 − 129	6
Mai 1950	125 − 121	4
Jun 1950	149 − 135	14

Interpretation: Die saisonal differenzierte Reihe zeigt die *Veränderung gegenüber dem Vorjahresmonat*. Der Wert 3 für Januar bedeutet: Im Januar 1950 flogen 3.000 Passagiere mehr als im Januar 1949. Die wiederkehrende saisonale Grundform (Sommer hoch, Winter tief) ist herausgefiltert — übrig bleibt das Wachstum von Jahr zu Jahr.
Die Kombination: d und D gemeinsam In der Praxis wendet man häufig beide Differenzierungen an:

Reihenfolge	Operation	Was wird entfernt
1. Saisonale Diff. (D = 1)	x*t = xt − xt-12	Das saisonale Muster
2. Gewöhnliche Diff. (d = 1)	x'*t = x*t − x*t-1	Der verbleibende Trend

Nach beiden Schritten sollte die Reihe stationär sein — sowohl der Trend als auch die Saisonalität sind entfernt. Für die Air Passengers (mit Log-Transformation wegen wachsender Amplitude) ist d = 1 und D = 1 die Standardwahl.

⚠ **Hinweis:** Die Reihenfolge der Differenzierungen spielt mathematisch keine Rolle — das Ergebnis ist dasselbe. In der Praxis beginnt man oft mit der saisonalen Differenzierung (D), weil man dann am ACF-Plot direkt sieht, ob der Trend bereits verschwunden ist oder noch eine gewöhnliche Differenzierung (d) nötig ist.

📌 **Merksatz:** Die saisonale Differenzierung $x^*_t = x_t - x_{t-m}$ vergleicht jeden Zeitpunkt mit dem gleichen Saisonpunkt der Vorperiode. D gibt die Anzahl saisonaler Differenzierungen an — in der Praxis fast immer D = 0 oder D = 1.

Saisonale AR- und MA-Komponenten (P und Q)
So wie p und q die kurzfristige Dynamik beschreiben, modellieren P und Q die **saisonale Dynamik** — also die Abhängigkeiten zwischen gleichen Saisonpunkten über die Jahre hinweg.
Saisonaler AR-Term (P)
Ein saisonaler AR(1)-Term mit m = 12 besagt: Der Juli dieses Jahres hängt vom Juli letzten Jahres ab. Die Formel lautet:

$$x_t = \Phi \cdot x_{t-12} + \varepsilon_t$$

Φ (großes Phi) ist der saisonale AR-Koeffizient. Die Logik ist identisch zum gewöhnlichen AR aus Block 6 — nur der Zeitsprung ist größer: statt eines Schritts zurück schauen wir eine ganze Saison zurück.
Zurück zur Pendel-Metapher: Wenn das gewöhnliche AR ein Pendel mit Trägheit von Monat zu Monat ist, dann ist das saisonale AR ein Pendel, das sich an den gleichen Monat des Vorjahres erinnert. Hoher Sommer-Umsatz letztes Jahr? Dann vermutlich auch dieses Jahr.
Saisonaler MA-Term (Q)
Ein saisonaler MA(1)-Term besagt: Ein Überraschungseffekt vom Juli letzten Jahres wirkt noch auf den Juli dieses Jahres nach. Die Formel:

$$x_t = \varepsilon_t + \Theta \cdot \varepsilon_{t-12}$$

Θ (großes Theta) ist der saisonale MA-Koeffizient. Auch hier gilt die Glocken-Metapher aus Block 7 — nur hallt die Glocke nicht einen Zeitschritt nach, sondern eine ganze Saison.

	Nicht-saisonal (Kleinbuchstaben)	Saisonal (Großbuchstaben)
AR	$\varphi \cdot x_{t-1}$ → Vormonat beeinflusst heute	$\Phi \cdot x_{t-12}$ → Gleicher Monat letztes Jahr beeinflusst heute
Differenzierung	$x_t - x_{t-1}$ → Trend entfernen	$x_t - x_{t-12}$ → Saisonalität entfernen
MA	$\theta \cdot \varepsilon_{t-1}$ → Schock von gestern hallt nach	$\Theta \cdot \varepsilon_{t-12}$ → Saisonaler Schock hallt nach

📌 **Merksatz:** P und Q funktionieren exakt wie p und q — nur auf der saisonalen Zeitskala. Saisonales AR (P) modelliert die Trägheit zwischen gleichen Saisonpunkten, saisonales MA (Q) das Nachhallen saisonaler Schocks. Die Konzepte aus Block 6 und 7 werden wiederverwendet, nicht ersetzt.

ACF und PACF bei saisonalen Daten lesen

In Block 8 haben wir gelernt, ACF und PACF für die Modellwahl zu nutzen. Bei saisonalen Daten müssen wir den Plot auf **zwei Ebenen** lesen:

Ebene	Relevante Lags	Was wir ablesen	Bestimmt
Nicht-saisonal	1, 2, 3, ...	Kurzfristige Abhängigkeiten (Monat zu Monat)	p und q
Saisonal	m, 2m, 3m, ... (z. B. 12, 24, 36)	Saisonale Abhängigkeiten (Jahr zu Jahr)	P und Q

Praktisch bedeutet das: Wir schauen den ACF-Plot zweimal an. Einmal die ersten Lags (1–11) für die kurzfristige Ebene. Dann die Vielfachen von 12 (12, 24, 36) für die saisonale Ebene. Dieselbe Logik — Cut-off vs. langsames Abklingen — gilt auf beiden Ebenen getrennt.

Erweiterte Entscheidungstabelle

Ebene	ACF-Verhalten	PACF-Verhalten	→ Modellwahl
Nicht-saisonal (Lags 1, 2, 3 ...)	Klingt langsam ab	Cut-off nach Lag p	→ AR(p): p setzen
Nicht-saisonal (Lags 1, 2, 3 ...)	Cut-off nach Lag q	Klingt langsam ab	→ MA(q): q setzen
Saisonal (Lags 12, 24, 36 ...)	Klingt langsam ab	Cut-off nach Lag m·P	→ **Saisonales AR: P setzen**
Saisonal (Lags 12, 24, 36 ...)	Cut-off nach Lag m·Q	Klingt langsam ab	→ **Saisonales MA: Q setzen**

Typisches Muster bei Air Passengers (nach Log-Transformation und Differenzierung mit d = 1, D = 1): Die ACF zeigt einen signifikanten Spike bei Lag 12, dann Abbruch → das deutet auf Q = 1. Die PACF zeigt ebenfalls einen Spike bei Lag 12, klingt dann ab. Auf der nicht-saisonalen Ebene zeigt die ACF einen Spike bei Lag 1, dann Abbruch → q = 1. Ein häufig gewähltes Modell ist daher ARIMA(0,1,1)(0,1,1)12.

> 📌 **Merksatz:** Bei saisonalen Daten liest man ACF und PACF auf zwei Ebenen: die kurzen Lags (1, 2, 3, ...) für p und q, die saisonalen Lags (m, 2m, 3m, ...) für P und Q. Die Entscheidungslogik aus Block 8 gilt auf beiden Ebenen identisch.

Bekannte SARIMA-Spezialfälle

Notation	Kurzname	Beschreibung
ARIMA(0,0,0)(0,0,0)12	Weißes Rauschen	Keine Struktur — weder kurzfristig noch saisonal
ARIMA(0,1,1)(0,1,1)12	Airline Model	Der Klassiker für monatliche Daten mit Trend und Saisonalität (nach Box & Jenkins, entwickelt an den Air-Passengers-Daten!)
ARIMA(1,0,0)(1,0,0)12	Saisonales AR	Kurzfristiges AR(1) + saisonales AR(1), Reihe ist stationär
ARIMA(0,1,0)(0,1,0)12	Saisonaler Random Walk	Doppelte Differenzierung ohne AR/MA-Modellierung
ARIMA(1,1,1)(1,1,1)12	Volles Modell	Alle Komponenten aktiv — selten nötig, aber flexibel

⚠ **Hinweis:** Das Airline Model ARIMA(0,1,1)(0,1,1)12 ist eines der erfolgreichsten statistischen Modelle überhaupt. Es wurde 1970 von Box und Jenkins an genau den Air-Passengers-Daten entwickelt und ist bis heute ein bewährter Ausgangspunkt für monatliche Daten mit Trend und Saisonalität.

Der erweiterte Arbeitsablauf (Box-Jenkins für SARIMA)
Der Ablauf aus Block 9 wird um die saisonale Ebene ergänzt:

Schritt	Aktion	Ergebnis
1. Visualisieren	Zeitreihe plotten, saisonale Muster identifizieren, Periode m bestimmen	m festlegen (z. B. m = 12)
2. Stabilisieren	Falls Amplitude wächst: Log-Transformation	Konstante Varianz
3. Saisonal diff.	x*t = xt − xt-m (falls saisonales Muster vorhanden)	D bestimmen (meist 0 oder 1)
4. Gewöhnlich diff.	x’t = xt − xt-1 (falls Trend noch vorhanden)	d bestimmen (meist 0 oder 1)
5. ACF/PACF	Nicht-saisonale Lags → p, q. Saisonale Lags → P, Q	Vollständige Parameterwahl
6. Modell schätzen	ARIMA(p,d,q)(P,D,Q)m auf die Daten fitten	Koeffizienten φ, θ, Φ, Θ
7. Residuen prüfen	ACF der Residuen: Keine signifikanten Lags bei 1,2,3... UND bei 12,24,36...	Modell validiert oder Rücklauf

Der entscheidende Unterschied zu Block 9: In Schritt 7 müssen wir die Residuen auf **beiden Ebenen** prüfen. Ein gutes SARIMA-Modell produziert Residuen, die weder kurzfristige noch saisonale Autokorrelation zeigen.

⚠ **Hinweis:** Auch bei SARIMA gibt es automatische Verfahren: auto.arima (R) und pmdarima (Python) probieren verschiedene (p,d,q)(P,D,Q)m-Kombinationen durch. Für monatliche Daten mit klarer Saisonalität ist ARIMA(0,1,1)(0,1,1)12 (das Airline Model) ein bewährter erster Versuch.

Zusammenfassung: ARIMA vs. SARIMA

Aspekt	ARIMA(p,d,q)	SARIMA(p,d,q)(P,D,Q)m
Erfasst Trend?	Ja, über Differenzierung (d)	Ja, identisch
Erfasst kurzfristige Dynamik?	Ja (AR + MA)	Ja, identisch
Erfasst Saisonalität?	Nein	Ja — über P, D, Q und m
Anzahl Parameter	3 (p, d, q)	7 (p, d, q, P, D, Q, m)
Neue Konzepte?	—	Nur die saisonale Differenzierung (D). P und Q sind Wiederverwendung bekannter Konzepte.
Typische Anwendung	Daten ohne saisonales Muster	Monatliche, quartalsweise oder wöchentliche Daten mit wiederkehrendem Muster

📌 **Merksatz:** SARIMA ist kein neues Modell, sondern ARIMA mit einer saisonalen Erweiterung. Die drei neuen Parameter P, D, Q wiederholen auf der Saison-Ebene (Lags m, 2m, 3m, ...) exakt das, was p, d, q auf der Einzelschritt-Ebene tun. Der einzige wirklich neue Baustein ist die saisonale Differenzierung xt − xt-m. Wer ARIMA verstanden hat, versteht auch SARIMA — es ist derselbe Baukasten, nur mit einer zweiten Uhr.

Schnellreferenz: SARIMA-Erweiterung zum Cheat Sheet

Konzept	Kernaussage
Saisonale Diff. (D)	xt − xt-m entfernt das saisonale Muster, indem jeder Wert mit dem gleichen Saisonpunkt verglichen wird
Saisonales AR (P)	Trägheit auf der Saison-Ebene: Der Juli beeinflusst den nächsten Juli
Saisonales MA (Q)	Saisonale Schock-Nachwirkung: Ein Überraschungseffekt im Juli hallt bis zum nächsten Juli nach
Periode m	Länge eines vollen Saisonzyklus (12 = monatlich, 4 = quartalsweise, 7 = täglich mit Wochenmuster)
Airline Model	ARIMA(0,1,1)(0,1,1)12: Standard für monatliche Daten mit Trend und Saison
ACF/PACF (saisonal)	Gleiche Logik wie bei p/q, aber auf den Lags m, 2m, 3m ablesen: bestimmt P, Q
Modellwahl-Tipp	Bei klarer monatlicher Saisonalität: Starte mit dem Airline Model als Baseline

12 Addon: Der ADF-Test (Augmented Dickey-Fuller)

Grundverständnis und praktischer Einsatz in der Zeitreihenanalyse

1 Das Ausgangsproblem: Stationarität

Viele Verfahren der Zeitreihenanalyse (z. B. ARIMA-Modelle) setzen voraus, dass die Daten **stationär** sind. Das bedeutet vereinfacht: Die Zeitreihe schwankt um einen festen Mittelwert und verhält sich statistisch über die Zeit gleich.

Zur Erinnerung (Leitfaden Block 3) müssen dafür drei Bedingungen erfüllt sein:

Bedingung	Was heißt das?	Gegenbeispiel
Konstanter Mittelwert	Die Reihe pendelt immer um denselben Wert	Aufsteigender Trend
Konstante Varianz	Die Schwankungsbreite bleibt gleich	Immer größere Ausräge
Konstante Autokovarianz	Der Zusammenhang zwischen aufeinanderfolgenden Werten ändert sich nicht	Sich ändernde Korrelation

Ein **Random Walk** dagegen erbt den vorigen Wert komplett und addiert nur zufälliges Rauschen. Schocks verpuffen nicht, sie häufen sich an – die Reihe wandert. Man sagt: sie hat eine **Einheitswurzel (Unit Root)**.

> **Kernfrage:** Gegeben eine Zeitreihe – ist sie stationär, oder hat sie eine Einheitswurzel? Der ADF-Test beantwortet genau diese Frage.

2 Was macht der ADF-Test?

Der ADF-Test prüft statistisch, ob eine Zeitreihe stationär ist oder nicht. Dabei funktioniert er wie ein Hypothesentest, den Sie schon aus der Statistik kennen:

	Hypothese	Bedeutung
H_0	Einheitswurzel vorhanden	**Nicht stationär** (die „schlechte Nachricht“)
H_1	Keine Einheitswurzel	**Stationär** (die „gute Nachricht“)

> **Achtung – umgekehrte Logik!** H_0 ist hier die „schlechte“ Annahme (nicht stationär). Ein kleiner p-Wert ist deshalb ein ***gutes*** Ergebnis, weil er H_0 verwirft und Stationarität nahelegt.

Was passiert intern? Der Test berechnet eine Teststatistik und vergleicht sie mit speziellen kritischen Werten (der sogenannten Dickey-Fuller-Verteilung). Daraus ergibt sich ein p-Wert. Die Python-Funktion *adfuller()* erledigt das alles automatisch – Sie müssen die Berechnung nicht selbst durchführen.

Wichtig: Der ADF-Test verwendet *nicht* die normale t-Verteilung, sondern eine eigene, nach links verschobene Verteilung. Die Funktion *adfuller()* berücksichtigt das automatisch. Deshalb sollten Sie immer die Bibliotheksfunktion nutzen und nicht manuell mit t-Werten vergleichen.

2.1 Einstellungen für den Test

Der ADF-Test kann optional berücksichtigen, ob die Zeitreihe um einen Mittelwert oder um einen Trend schwankt. Das steuert man über den Parameter *regression*:

Python-Parameter	Wann verwenden?
regression='c'	Standardeinstellung – passt für die meisten Fälle (Reihe schwankt um einen Mittelwert)
regression='ct'	Wenn die Reihe einen sichtbaren linearen Trend hat
regression='n'	Nur wenn die Reihe offensichtlich um 0 schwankt (selten)

> **Analogie:** Stellen Sie sich vor, Sie wollen testen, ob ein Kind Fieber hat. Aber das Kind war draußen in der Sonne, und die Temperatur steigt allein deswegen. Wenn Sie den Sonneneffekt nicht rausrechnen, denken Sie fälschlicherweise „Fieber!“. Der Parameter *'ct'* rechnet solche Trendeffekte raus.

3 Ablauf in der Praxis

3.1 Schritt für Schritt

Schritt	Aktion
1	Zeitreihe plotten. Visuell prüfen: Wandert sie? Hat sie einen Trend?
2	ADF-Test durchführen: adfuller(x, regression='c')
3	p-Wert ablesen
4a	$p < 0{,}05$ → H_0 verworfen → Reihe ist stationär → **direkt modellieren**
4b	$p \geq 0{,}05$ → Reihe differenzieren (Δx bilden) und bei Schritt 2 wiederholen

3.2 Python-Code

```
from statsmodels.tsa.stattools import adfuller

ergebnis = adfuller(x, regression='c')

adf_statistik = ergebnis[0]   # z.B. -2.26
p_wert        = ergebnis[1]   # z.B.  0.184
krit_werte    = ergebnis[4]   # kritische Werte
```

> **Entscheidungsregel:** $p < 0{,}05$ → stationär. So einfach ist es in der Praxis.

4 Wozu differenziert man?

Wenn der ADF-Test sagt „nicht stationär“, muss man die Reihe **differenzieren** (vgl. Leitfaden Block 5). Dabei wird statt der Originalwerte die **Veränderung** von Zeitschritt zu Zeitschritt betrachtet: $\Delta x = x_t - x_{t-1}$.

Aber warum? Ein reiner Random Walk wäre tatsächlich nicht vorhersagbar. Reale Zeitreihen sind aber selten rein zufällig. Typischerweise überlagern sich verschiedene Komponenten (vgl. Leitfaden Block 2, Dekomposition):

Komponente	Beispiel	Stationär?
Niveau (Level)	Umsatz liegt bei ca. 100k €	Kann wandern → oft nicht stationär
Trend	Umsatz steigt jedes Jahr um 5 %	Systematischer Anstieg → nicht stationär
Saisonalität	Dezember-Umsatz immer höher	Wiederkehrend → vorhersagbar
Rauschen	Zufällige Schwankungen	Stationär

Differenzieren entfernt das wandernde Niveau und legt die **Veränderungsmuster** frei. „Im Dezember steigt der Umsatz immer um ca. X“ ist eine Aussage über die Veränderung Δx, nicht über den Absolutwert x.

> **Merksatz:** Differenzieren ist kein Selbstzweck, sondern der Schritt, der die modellierbaren Muster sichtbar macht. Der Parameter *d* in ARIMA(p,d,q) gibt die Anzahl der nötigen Differenzierungen an (Leitfaden Block 9). Der ADF-Test sagt Ihnen, ob dieser Schritt nötig ist.

4.1 Entscheidungstabelle

Situation	Was tun?
ADF sagt „stationär“	Direkt modellieren (z. B. ARMA)
ADF sagt „Einheitswurzel“	Differenzieren (Δx bilden), erneut testen
Differenzierte Reihe ist stationär	Auf Δx modellieren → ARIMA-Modell
Auch nach Differenzierung nicht stationär	Nochmals differenzieren (selten nötig, $d \geq 3$ kommt praktisch nie vor)

5 Zusammenfassung

Der ADF-Test in drei Sätzen

1. Der ADF-Test prüft, ob eine Zeitreihe stationär ist oder eine Einheitswurzel hat.

2. Ist der p-Wert < 0,05, wird H_0 verworfen und die Reihe gilt als stationär.

3. Ist die Reihe nicht stationär, wird differenziert und erneut getestet. Das Ergebnis bestimmt den Parameter d im ARIMA-Modell.

Hinweis: Das begleitende Jupyter-Notebook zeigt alle Schritte mit konkreten Daten und Python-Code.