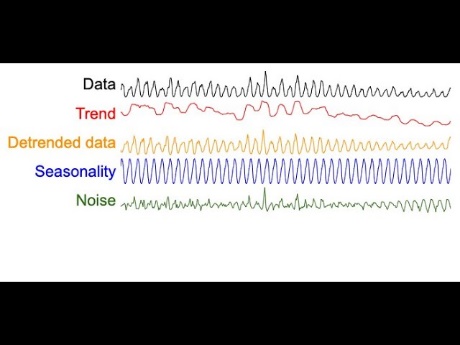
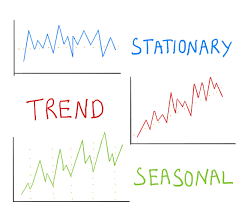
# Introduction and basic concepts



Anders ausgedrückt: Ist die Zeitreihe stationär? Eine stationäre Zeitreihe hat konstante Mittelwerte, Varianzen und keine Abhängigkeit der Statistik von der Zeit.

Gibt es einen Trend? Ein Trend ist eine langfristige Veränderung in den Daten, wie z. B. ein kontinuierlicher Anstieg oder Abfall über die Zeit.

Saisonale Effekte sind wiederkehrende Muster oder Schwankungen in der Zeitreihe, die regelmäßig auftreten (z. B. monatliche oder jährliche Zyklen)

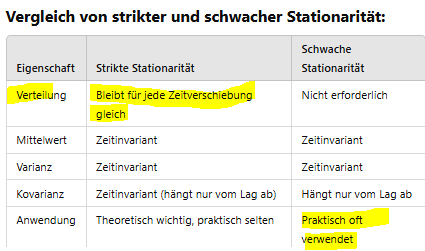
Ein **stationäres Muster** zeigt keine Trends oder Saisonalitäten.

Univariate Zeitreihe:Eine Zeitreihe, bei der nur eine Größe über die Zeit beobachtet wird.

Multivariate Zeitreihe: Falls wir an mehreren Größen interessiert sind, wird eine Sammlung von Zufallsvariablen betrachtet:

Mittelwert: : Varianz: Covariance(Autokovarianz misst, wie stark die Werte der Zeitreihe zu zwei verschiedenen Zeitpunkten miteinander zusammenhängen.) COV: : Autocorrelation: liegt zwischen -1 und 1.

**White Noise**: Zeitreihe wird als White Noise bezeichnet, wenn die Variablen

unabhängig und identisch verteilt (i.i.d.) sind,und der Mittelwert sowie die Varianz zeitlich konstant sind.

|  |  |  |
| --- | --- | --- |
| Modell with Trend | Moving Average | Non Stationary |
| Nicht stationär | stationär | Nicht stationär |
| Mittelwert hängt von Trend ab | Mittel und Varianz sind konstant | Mittel und Varianz ändert mit Zeit |
|  |  |  |

# Box-Jenkins Modelling: ARIMA Models

ARIMA hilft, zukünftige Punkte in einer Zeitreihe vorherzusagen, indem es: Abhängigkeiten zwischen Beobachtungen (AR: Autoregressiv) nutzt.

Daten durch Differenzierung stationär macht (I: Integriert).Fehler der Vergangenheit glättet (MA: Moving Average).

|  |  |  |
| --- | --- | --- |
| AR (Autoregressiv): Modelliert den aktuellen Wert basierend auf vorherigen Werten.  Beispiel: AR(1) hängt vom Wert des letzten Tages ab. | I (Integriert): Entfernt Trends durch Differenzierung.  Beispiel: Subtrahiere den Wert von gestern vom heutigen. | MA (Moving Average): Modelliert Abhängigkeiten durch vorherige Fehler.  Beispiel: Nutze Fehler der letzten zwei Tage, um den heutigen Wert zu glätten. |
|  |  |  |
| p: Anzahl der autoregressiven (AR) Terme | d: Anzahl der Differenzierungen (Integrated) | q: Anzahl der Moving-Average (MA) Terme. |
| Beispiel: p=2 bedeutet, dass der heutige Wert von den letzten 2 Werten abhängt. | Beispiel: d=1 bedeutet, dass einmal angewendet wird, um Trends zu entfernen. | Beispiel: q=1 bedeutet, dass der heutige Wert durch den Fehler des letzten Tages beeinflusst wird. |
| Plotten Sie die Zeitreihe: Wenn ein Trend sichtbar ist, könnte d≥1d \geq 1d≥1 nötig sein. | Dickey-Fuller-Test (ADF-Test): Ein statistischer Test, um Stationarität zu überprüfen. Wenn die Zeitreihe nicht stationär ist, erhöhen Sie d | Verwenden Sie die Autokorrelationsfunktion (ACF), um die Anzahl der gleitenden Mittelwerte zu bestimmen.  Ein starker Abfall in der ACF zeigt den Wert für q |
| stationär  d=0 und q=0 | grenzstationär  p=0 und d=0d = 0d=0 | nicht stationär  Werte explodieren.  p=0 und q=0 |

Schritte zum Aufbau eines ARIMA-Modells:

1. Zeitreihe stationär machen:

Verwenden Sie Differenzierung, um Trends oder Saisonalität zu entfernen.

Ziel ist es, die Zeitreihe stationär zu machen (konstanter Mittelwert und Varianz).

1. Bestimmen der Parameter ppp, ddd, qqq:

ppp (AR-Term):

Verwenden Sie die partielle Autokorrelationsfunktion (PACF), um die Anzahl der autoregressiven Terme zu bestimmen.

Ein starker Abfall in der PACF zeigt den Wert für ppp.

ddd (Differenzierung):

Bestimmen Sie die Anzahl der Differenzierungen, die nötig sind, um die Daten stationär zu machen.

Überprüfen Sie dies mit Tests wie dem Augmented Dickey-Fuller-Test (ADF).

qqq (MA-Term):

Verwenden Sie die Autokorrelationsfunktion (ACF), um die Anzahl der gleitenden Mittelwerte zu bestimmen.

Ein starker Abfall in der ACF zeigt den Wert für qqq.

1. Modell anpassen:
2. Prognosen erstellen:

# Model Estimation

Der Backward Operator (B) wird in der Zeitreihenanalyse verwendet, um Indizes einer Zeitreihe zu verschieben, Differenzen zu berechnen und Modelle wie ARIMA einfacher darzustellen. Der Backward Operator kann als Matrix dargestellt werden, die eine Zeitreihe "verschiebt". Forward (F) ist das Gegenteil. autoregressive Modell der Ordnung p Backshift Operators (B) wird das Modell vereinfacht geschrieben:

Ein AR(2) Modell: wird zu

Parameter (𝜙,𝜃) für ein gegebenes ARIMA-Modell zu schätzen. ϕ = Der autoregressive Parameter. Θ = Moving-Average-Parameter.

**Drift or Trend entfernen**

Drift und Trend machen ARIMA-Modelle flexibler für reale Daten, die nicht nur zufällige Schwankungen zeigen, sondern auch systematische Änderungen aufweisen.

μ: Driftparameter, der eine konstante Verschiebung in der Zeitreihe modelliert. Zweck: Dieses Modell passt sich Zeitreihen an, die einen konstanten Anstieg oder Abfall im Mittelwert über die Zeit zeigen.

= ARIMA mit drift : Drift in der Praxis = Mittelwert

Zweck: Dieses Modell passt sich Zeitreihen an, die einen systematischen linearen Trend über die Zeit enthalten. Ein ARIMA-Modell mit Trend ist nützlich für Zeitreihen, die einen linearen Trend zeigen und nicht vollständig stationär sind.

Berechnen: Berechnen t sind alle Zeitindizes, y sind alle Werte

Die Yule-Walker-Gleichungen sind eine analytische Methode, um AR-Koeffizienten basierend auf der Autokorrelationsstruktur zu schätzen.

Das Verfahren ist besonders effizient für AR(p)-Modelle und nutzt ein einfaches lineares Gleichungssystem.

AR(2): ϕ**:** Autoregressive Koeffizienten werden gesucht. Da Cov() = Cov()+ Cov()

# Model Diagnostics

|  |  |  |
| --- | --- | --- |
| Residuen = Unterschied zwischen den beobachteten Werten der Zeitreihe und den durch das Modell vorhergesagten Werten  Wenn ARIMA-Modell korrekt, sollten die Residuen White Noise sein: Zufällig verteilt ohne erkennbare Muster. Einen Mittelwert von 0 haben. Eine konstante Varianz besitzen. | Schritt 1 :Plotten der Residiums    sollten keine sichtbaren Muster oder Trends aufweisen | 2. Schritt: QQ-Plot der Residuen  Wenn die Punkte entlang der Diagonalen liegen, sind die Residuen normal verteilt |

|  |  |  |
| --- | --- | --- |
| Schritt 3: Residuen vs. vorhergesagte Werte  Ziel: Überprüfung, ob die Residuen von den vorhergesagten Werten abhängen.  Die Punkte sollten zufällig verteilt sein, ohne sichtbare Muster oder Trends. Ein systematisches Muster deutet darauf hin, dass das Modell die Daten nicht vollständig erklären kann. | Schritt4: Autokorrelationsfunktion (ACF) der Residuen misst die lineare Abhängigkeit der Residuen. von ihren verzögerten (gelagerten) Werten  bei verschiedenen Verzögerungen 𝑘.-  Wenn nicht Null: q zu klein | Schritt 5: partielle Autokorrelationsfunktion (ACF of the residuals. Verzögerungen  Wenn nicht Null: p zu klein |

Overfitting tritt auf, wenn ein Modell zu komplex ist (z. B. zu viele Parameter 𝑝,𝑞 ), sodass es nicht nur die Daten beschreibt, sondern auch Zufälligkeiten und Rauschen in den Daten modelliert. Ein überangepasstes Modell kann instabil werden und unzuverlässige Schätzungen liefern.

# Time Series Decomposition

|  |  |  |
| --- | --- | --- |
| Additives Modell:  Die Zeitreihe wird in Trend  saisonale Effekte und zufällige Schwankungen zerlegt. Verwendung: Geeignet, wenn die Saisonalität unabhängig vom Trend ist. | Multiplikatives Modell:  Die Komponenten werden multiplikativ kombiniert, wodurch sie miteinander interagieren.  Verwendung: Geeignet, wenn die Saisonalität mit dem Trend skaliert (z. B. größere Schwankungen bei steigendem Trend). | Pseudo-additives Modell:  Verwendung: Für komplexe Zeitreihen, die weder rein additiv noch rein multiplikativ sind. |

Die Saisonalität ist oft K-periodisch (z. B. mit einer Periode von 12 Monaten).

Saisonalität wird durch ACF-Plots sichtbar gemacht, und wiederkehrende Muster geben die Periode an.

**Saisonbereinigte Zeitreihen:**

Entfernen wiederkehrender saisonaler Muster, um langfristige Trends und zufällige Schwankungen besser zu analysieren.

Prognose: Ein saisonbereinigtes Modell (z. B. ARIMA) ist oft einfacher und effektiver für Vorhersagen.

|  |  |
| --- | --- |
| **Schätzung des Trends**  Zuerst wird der langfristige Trend geschätzt, z. B. mit einem (Moving Averages) oder einem Modell wie ARIMA. | **Schätzung der Saisonalität**  saisonalen Effekte über eine komplette Periode hinweg im Durchschnitt neutralisiert werden. |
| K-Moving Average (K-MA)  K muss ungerade sein. m: Anzahl der Perioden auf jeder Seite des aktuellen Wertes t.    **Regression Smoothing**  Flexibilität: Kann kurzfristige Trends besser erkennen,  Genauigkeit: Beschreibt die Dynamik der Zeitreihe präziser als einfache gleitende Durchschnitte, insbesondere bei Daten mit sich ändernden Trends. | **Double Moving Average**  Zweites glätten der Ergebnisse  **Weighted Moving Average (WMA)**  Hinzufügen von Gewicht a. Die Summe aller a =1 skaliert  Fexibler als der einfache gleitende Durchschnitt, da er Trends in den Daten besser berücksichtigt.  **Regression Smoothing**  Anstelle eines einfachen Durchschnitts wird der Trend durch eine lineare Funktion lokal angenähert:  Geschätzter lokaler Achsenabschnitt: a(t)  Geschätzte lokale Steigung b(t)  Die geglättete Zeitreihe ergibt sich dann als: |

# Multicatiate TS & Causality

Bei multivariaten Zeitreihen werden mehrere Variablen (p Zeitreihen) gleichzeitig beobachtet, die möglicherweise voneinander abhängen.

Vector Autoregressive Model (VAR)

Erweiterung autoregressiver Modelle (AR-Modelle) auf multivariate Zeitreihen. Jede Zeitreihe wird als Linearkombination ihrer eigenen vergangenen Werte und der vergangenen Werte anderer Zeitreihen modelliert. Verwendung von VAR-Modellen: Analyse der Kausalität zwischen Variablen. Prognosen für multivariate Zeitreihen. mUntersuchung der Dynamik und Interaktionen zwischen Variablen. Die Schätzung von Φ erfolgt typischerweise durch das **Kleinste-Quadrate-Verfahren (Least Squares)** oder durch **Maximum-Likelihood-Schätzung**.

Die Zeitreihen müssen stationär gemacht werden (z. B. durch Differenzierung), bevor kausale Beziehungen untersucht werden.

Kreuzkorrelation (Cross-Correlation)

Die Kreuzkorrelation misst den Zusammenhang zwischen zwei Zeitreihen bei verschiedenen Verzögerungen (k= Lags).

Umgang mit Scheinzusammenhängen: Modelle mit rohen Daten können zu falschen Schlussfolgerungen führen.

Differenzierung: Entfernt Trends und saisonale Muster, wodurch echte Zusammenhänge identifiziert werden können. Cointegration:Für nichtstationäre Zeitreihen, die langfristig miteinander verbunden sind (z. B. Wechselkurse und Zinssätze).

**Granger Causality Test:** vergangenen Werte von X zusätzliche Informationen liefern, um Y besser vorherzusagen, als dies nur durch die vergangenen Werte von Y selbst möglich wäre.

**F-Test** vergleicht die Vorhersagekraft eines eingeschränkten und eines erweiterten Modells.Die Nullhypothese wird verworfen, wenn die Hinzunahme von X die Vorhersage von Y signifikant verbessert. Die Berechnung der F-Statistik basiert auf den Residual Sum of Squares (RSS) beider Modelle.

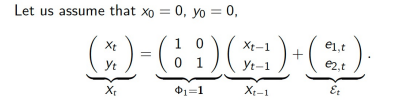
# Cointegration

Nichtstationäre Zeitreihen: Kointegration bezieht sich auf Zeitreihen, die individuell nichtstationär sind, d. h., sie zeigen Trends oder sich verändernde statistische Eigenschaften (z. B. Mittelwert und Varianz ändern sich im Zeitverlauf).

Eine **lineare Kombination** dieser nichtstationären Zeitreihen kann **stationär** sein.

Langfristige Beziehung: Kointegration deutet darauf hin, dass die Zeitreihen eine langfristige Gleichgewichtsbeziehung haben.

Fehlerkorrektur: Wenn Kointegration vorliegt, tendieren kurzfristige Abweichungen von der langfristigen Beziehung dazu, sich selbst zu korrigieren.

Wenn dann gilt Dies impliziert, dass x und y nicht unabhängig sind und möglicherweise kointegriert sind (sie folgen einer gemeinsamen Gleichgewichtsbeziehung).

Problemstellung:

(Einkommen) und (Konsum) sind beide nichtstationäre Zeitreihen (sie besitzen Einheitswurzeln).

Lösung durch Differenzierung Um Stationarität zu erreichen, können die Differenzen der Zeitreihen betrachtet werden:

Veränderung des Konsums, hat einen Zusammenhang (Steigung) mit dem Einkommen. Obwohl diese Methode statistisch korrekt ist, liefert sie ökonomisch oft keine bedeutungsvolle Interpretation, da sie langfristige Gleichgewichtsbeziehungen zwischen den Variablen ignoriert.

**Error Correction Model (ECM):**

**Engle-Granger-Tests** Überprüfung, ob zwei nichtstationäre Zeitreihen X und Y eine kointegrierte Beziehung besitzen.

Wenn die Residuen einer linearen Kombination dieser Zeitreihen stationär sind, besteht eine langfristige Gleichgewichtsbeziehung.

Nichtstationarität von beiden Reihen mit Augmented Dickey-Fuller-Test (ADF)

**lineare Regression** durch

**Test der Residuen auf Stationarität**: Verwende dazu erneut den **ADF-Test**.

**Ergebnis:** Wenn die Residuen stationär sind, sind x und y **kointegriert**.

Geringe Teststärke:Der Test tendiert dazu, die Hypothese der Kointegration zu oft abzulehnen. wegen empfindlichkeit auf kleine Abweichungen und Verzerrungen.

Begrenzung auf zwei Zeitreihen: Für mehr als zwei Zeitreihen ist der Johansen-Test eine bessere Wahl.

# Heteroskedasticity

Heteroskedastizität tritt auf, wenn die **Varianz der Fehlerterme** einer Zeitreihe nicht konstant ist.

Dies ist ein häufiges Phänomen in finanziellen Zeitreihen, bei denen Perioden hoher Volatilität (Schwankungen) oft von weiteren Perioden hoher Volatilität gefolgt werden.

Volatility Clustering:In der Zeitreihe (oben) sind Phasen mit hoher Volatilität (große Schwankungen) und niedriger Volatilität sichtbar.

Dieses Verhalten wird als Volatility Clustering bezeichnet, ein charakteristisches Merkmal heteroskedastischer Zeitreihen.

# Time Series Topics

* 2 Seiten
* 50 Fragen zu TS in Chatgpt
* Taschenrechner testen wegen Varianz

<https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data>

<https://www.kaggle.com/datasets/matthieugimbert/french-bakery-daily-sales>