

Dynamische Systeme und Zeitreihenanalyse

Beschreiben von Zeitreihen *Kapitel 9*

Statistik und Mathematik – WU Wien

Michael Hauser

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Beschreiben von Zeitreihen – 9 – p.0??

Lernziele

- Beschreiben von Zeitreihen
- Graphische Darstellungen
- Univariate und bivariate Maßzahlen
- Transformationen
- Maße für die Prognosegüte

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Beschreiben von Zeitreihen – 9 – p.1??

Graphische Darstellung

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Beschreiben von Zeitreihen – 9 – p.2??

Zeitreihen-Plot

Der Pfad der Reihen wird im **Zeitreihen-Plot** visualisiert.
Hier werden

- Trends
- Zyklen
- Saison
- Irreguläres Verhalten (keine Regelmäßigkeit erkennbar)
- Ausreißer, atypische Beobachtungen

meist sichtbar.

4 Muster

- **Horizontales Muster:** Daten fluktuieren um ein konstantes Mittel.
Bsp: Renditen
- **Saisonales Muster:** Daten fluktuieren mit fixer Periodizität.
Bsp: Nüchtigungen im Tourismus (Monatsdaten)
- **Zyklisches Muster:** Daten fluktuieren in Zyklen unregelmäßiger Länge und Amplitude.
Bsp: Konjunkturzyklen
- **Trend:**
Daten steigen oder fallen über eine längere Periode.
Bsp: BIP-real

Überlagerung der Strukturen

- Ausgeprägte Strukturen machen die Wahl eines geeigneten Modells leicht.
- In kurzen Reihen sind einzelne Muster nicht zuverlässig erkennbar.
- Die Unterscheidung zwischen Trend und langen Zyklen ist allgemein ein schwieriges Problem.

Histogramm

Das **Histogramm** dient zur graphischen Beschreibung der Verteilung einer Variablen. Es gibt im Wesentlichen darüber Auskunft, wie die Variable streut und welche Form die Dichte hat.

Nur wenn die x_t eine stationäre Struktur aufweisen, ie alle x_t zumindest das gleiche Mittel und die gleiche Varianz aufweisen, ist das Histogramm brauchbar.

Bsp: Liegt ein positiver Trend vor, wird das Histogramm auseinandergezogen und zeigt je nach Zeitabschnitt ein anderes Niveau.

Streudiagramm, Zeitreihen-Plot

Hier werden zwei Variablen gegeneinander aufgetragen. Im Zusammenhang mit der Regressionsanalyse haben wir von einer Punktwolke, *scatter plot*, gesprochen.

Man kann daraus den Korrelationskoeffizienten abschätzen und sich überlegen, wie eine Regressionsgerade hineinzulegen wäre.

Der **Zeitreihen-Plot** ist ein spezielles *Streudiagramm*:

Zeit × Variable

Einfache Maßzahlen

Mittelwert

Die Reihe werde mit $\{y_t\}$, $t = 1, \dots, T$ bezeichnet.

Mittelwert, \bar{y} , μ

In der *Stichprobe*:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

In der *Grundgesamtheit*, für eine Zufallsvariable y , sofern die Erwartung existiert,

$$\mu = E(y)$$

Ein *Schätzwert* für μ ist \bar{y} :

$$\hat{\mu} = \bar{y}$$

Varianz

Varianz, s^2 , σ^2

In der *Stichprobe*:

Die Varianz ist die mittlere Abweichungsquadratsumme, mittlere Summe der quadrierten Abweichungen der einzelnen Beobachtungen vom Mittelwert.

$$s^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

oder die korrigierte Stichprobenvarianz

$$s^2 = \frac{1}{T-1} \sum_{t=1}^T (y_t - \bar{y})^2$$

Varianz (Fs)

In der *Grundgesamtheit*, für eine Zufallsvariable y , sofern die Varianz existiert:

$$\sigma^2 = V(y) = E(y - \mu)^2$$

Ein *Schätzwert* für σ^2 ist s^2 :

$$\hat{\sigma}^2 = s^2$$

Wendet man den Verschiebungssatz an:

$$\sigma^2 = V(y) = E(y - \mu)^2 = E(y^2) - \mu^2$$

Standardabweichung

Standardabweichung, s, σ

Die Standardabweichung ist die Wurzel aus der Varianz.

$$s = \sqrt{s^2}$$

bzw

$$\sigma = \sqrt{V(y)}$$

Ein *Schätzwert* für σ ist s :

$$\hat{\sigma} = s$$

Kovarianz

Seien $\{x_t\}$ und $\{y_t\}$ zwei Reihen, $t = 1, \dots, T$.

Kovarianz, $\widehat{Cov}(x, y) = cov(x, y), cov_{xy}$

In der *Stichprobe*:

$$cov_{xy} = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})$$

bzw korrigiert

$$cov_{xy} = \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})$$

Kovarianz (Fs), Verschiebungssatz

In der *Grundgesamtheit*, für zwei Zufallsvariable x und y , für die Erwartung und Varianz existieren:

$$Cov(x, y) = E(x - \mu_x)(y - \mu_y)$$

Der **Verschiebungssatz** auf die Kovarianz angewendet liefert

$$Cov(x, y) = E(x - \mu_x)(y - \mu_y) = E(x \cdot y) - \mu_x \cdot \mu_y$$

Korrelation

Korrelationskoeffizient, $\hat{\rho}_{xy} = r_{xy}$, ρ_{xy}

In der *Stichprobe*:

$$r_{xy} = \frac{cov_{xy}}{s_x s_y} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}$$

Die zugehörige graphische Darstellung ist das **Streudiagramm**.

In der *Grundgesamtheit*, für zwei Zufallsvariable x und y , für die Erwartung und Varianz existieren:

$$\rho_{xy} = \text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{V(x)} \sqrt{V(y)}}$$

Der Korrelationskoeffizient liegt immer zwischen -1 und 1 :

$$-1 \leq \rho_{xy} \leq 1 \quad -1 \leq r_{xy} \leq 1$$

Korrelogramm

Autokorrelation k -ter Ordnung

Autokorrelation k -ter Ordnung misst die Korrelation zwischen Werten einer Reihe, die k Perioden voneinander entfernt sind.

Gegeben sei die Reihe $\{y_t\}$ mit $t = 1, \dots, T$.

Die Werte y_{t-k} und y_t sind k Perioden voneinander entfernt. Wir sagen y_{t-k} ist der Wert von y_t um k Perioden **verzögert**.

In der *Stichprobe* ist der **Autokorrelationskoeffizient k -ter Ordnung** definiert als

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Fasst man die Autokorrelationen der Ordnungen $0, 1, 2, 3, \dots$ zusammen, erhält man die **Stichprobenautokorrelationsfunktion**, **SACF**, bzw **Korrelogramm** mit $r_0 = 1$: $1, r_1, r_2, \dots$

Beispiel - Korrelogramm

Beispiel: Berechnung des Autokorrelationskoeffizienten

Wir berechnen den Autokorrelationskoeffizienten 1. Ordnung für die australische Bierproduktion. Die Formel für $k = 1$ ist

$$r_1 = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Das Mittel ist mit $T = 56$: $\bar{y} = 8361/56 = 149.30$

Die Autokovarianz für Lag 1

$$\text{cov}(y_t, y_{t-1}) = 8893.51/56 = 158.80$$

und die Autokorrelation 1. Ordnung

$$r_1 = \frac{8893.51}{21135.84} = 0.421$$

Beispiel - Tabelle - SACF zu beer2

(1)	(2)	(3)	(4)	(5)	(6)	(7)
t	y_t	y_{t-1}	$(y_t - \bar{y})$	$(y_{t-1} - \bar{y})$	$(y_t - \bar{y})^2$	$(y_t - \bar{y}) \cdot (y_{t-1} - \bar{y})$
1	164	-	14.70	-	215.99	-
2	148	164	-1.30	14.70	1.70	-19.16
3	152	148	2.70	-1.30	7.27	-3.51
4	144	152	-5.30	2.70	28.13	-14.30
⋮		⋮		⋮		⋮
55	119	130	-30.30	-19.30	918.31	584.97
56	153	119	3.70	-30.30	13.66	-112.01
Σ	8361				21135.84	8893.51

Interpretation von r_{xy}

Wir betrachten zwei Reihen x_t und y_t und die bivariate Regression y_t auf x_t .

$$y_t = a + b x_t + \epsilon_t$$

ϵ_t ist der Störterm.

Die Methode der Kleinsten Quadrate liefert eine Schätzung für die Parameter a , b und für die Fehler ϵ_t .

$$y_t = \hat{a} + \hat{b} x_t + \hat{\epsilon}_t$$

mit

$$\hat{b} = \frac{\hat{\sigma}_y}{\hat{\sigma}_x} \hat{\rho}_{xy}, \quad \hat{a} = \hat{\mu}_y - \hat{b} \hat{\mu}_x$$

Interpretation von $r_1: x_t = y_{t-1}$

In der Einführung in die Statistik wird \hat{b} als $\hat{b} = \widehat{Cov}(x, y) / \hat{\sigma}_x^2$ angegeben.

Nun ist aber $\rho = Cov(x, y) / (\sigma_x \sigma_y)$, bzw aufgelöst nach der Kovarianz $Cov(x, y) = \rho \sigma_x \sigma_y$. Einsetzen und Kürzen ergibt die obige Beziehung für \hat{b} : $\hat{b} = (\hat{\sigma}_y / \hat{\sigma}_x) \hat{\rho}_{xy}$

Im Fall $x_t = y_{t-1}$ folgt unter der Annahme $V(x_t) = \hat{\sigma}_x^2 = const$:
 $\sigma_x = \sigma_y$ und $\rho_{xy} = \rho(1)$ ist der erste Autokorrelationskoeffizient.
Damit ist $\hat{b} = \hat{\rho}(1)$.

Die lineare Beziehung zwischen y_t und y_{t-1} lautet daher

$$y_t = a + \rho(1) y_{t-1} + \epsilon_t$$

Interpretation von $r_k, k = 1, 2, \dots$

Wenn $\rho(1) \neq 0$ gilt, macht die Regression Sinn und y_{t-1} kann zur Prognose von y_t verwendet werden.

Die Modellierbarkeit von y_t wird auch im Streudiagramm $y_{t-1} \times y_t$ sichtbar.

Verallgemeinerung für $r_k, k = 1, 2, 3, \dots$:

Analog für den Fall $x_t = y_{t-k}$:

$$y_t = a + \rho(k) y_{t-k} + \epsilon_t$$

Es sind die Streudiagramme $y_{t-k} \times y_t$ zu inspizieren.

Für mehrere verzögerte Variable auf der rechten Seite der Regression sind die Zusammenhänge etwas komplizierter.

Transformationen

Ziel der Transformation

Reihen können mit Standardverfahren modelliert werden, wenn sie

- einen linearen Trend, oder
- konstantes Mittel und
- konstante Varianz

besitzen.

Exponentielles Wachstum: $y_t = \alpha \exp(\beta t)$

Exponentielles Wachstum wird durch Logarithmierung linearisiert.

$$x_t = \log(y_t) = \tilde{\alpha} + \beta t$$

(In der engl. Literatur ist \log iA der natürliche Logarithmus.)

Bsp: Aktienindizes, reales BIP

Varianzstabilisierung

Allgemeiner als die Logarithmierung ist die **Box-Cox Transformation**. Sie wird zur Varianzstabilisierung verwendet.

Steigt (selten fällt) die Varianz einer Reihe mit steigendem Mittel, so kann eine konstante Varianz durch die geeignete Wahl von p erreicht werden.

$$y_t^{BC} = \begin{cases} -y_t^p & p < 0 \\ \log(y_t) & p = 0 \\ y_t^p & p > 0 \end{cases}$$

Varianzstabilisierung bewirkt für

- $p > 1$: Die Varianz für große y_t Werte wird größer.
- $p = 1$: Reihe bleibt unverändert.
- $p < 1$: Die Varianz für große y_t Werte wird kleiner.
- $p = 0$: Geeignete Transformation für exponentielles Wachstum.
- $p < 0$: Die Varianz für große y_t Werte wird noch stärker reduziert.

Die Rücktransformation ist

$$y_t = \begin{cases} (-y_t^{BC})^{1/p} & p < 0 \\ \exp(y_t^{BC}) & p = 0 \\ (y_t^{BC})^{1/p} & p > 0 \end{cases}$$

Prognose

Prognose

Das **Prognoseproblem** ist:

Wir beobachten Daten $y_t, t = 1, \dots, T$, und wollen wissen, wie sie sich in der Zukunft, $y_{T+\tau}, \tau > 0$, für die Zeitpunkte $(T+1), (T+2), \dots, (T+\tau), \dots$ entwickeln.

Allgemein unterscheidet man

- **Punkt- und**
- **Intervallprognosen,**

ähnlich wie Punkt- und Intervallschätzer.

Einfache Prognosen

Angenommen unsere Reihe weist

- ein *horizontales Muster* auf.
Eine einfache (konstante) Prognose ist das Mittel der Reihe fortzuschreiben. Modell: $y_t = \mu + \epsilon_t, \hat{\mu} = \bar{y}$

$$\{y_1, y_2, \dots, y_T\} \quad \{\hat{y}_{T+1} = \bar{y}, \hat{y}_{T+2} = \bar{y}, \dots\}$$

- einen *linearen (deterministischen) Trend* auf.
Der Trend wird zur Prognose fortgeschrieben:

$$y_t = \hat{a} + \hat{b}t + \hat{\epsilon}_t \quad t = 1, \dots, T \quad \hat{y}_{T+\tau} = \hat{a} + \hat{b}(T+\tau) \quad \tau = 1, 2, \dots$$

Die prognostizierten Fehler, $\hat{\epsilon}_{T+\tau}$, werden Null gesetzt.

Einfache Prognosen (Fs)

Angenommen unsere Reihe weist

- ein *random walk* Muster auf.

Die einfachste Prognose, auch **naïve Prognose** genannt, ist hier der zuletzt beobachtete Wert.

Modell: $y_t = y_{t-1} + \epsilon_t$

$$\{y_1, y_2, \dots, y_T\} \quad \{\hat{y}_{T+1} = y_T, \hat{y}_{T+2} = y_T, \dots\}$$

τ -Schritt Prognose, τ -Schritt Prognosefehler

$\hat{y}_{T+\tau}$ heißt **τ -Schritt Prognose**, die Prognose für τ Schritte in die Zukunft.

Der **τ -Schritt Prognosefehler**, $e_{T+\tau}$, ist für $\tau = 1, 2, \dots$

$$e_{T+\tau} = y_{T+\tau} - \hat{y}_{T+\tau}$$

$y_{T+\tau} \dots$ zukünftige, noch nicht beobachtete Werte.

Ziel ist es, den Prognosefehler so klein wie möglich zu machen.

in-sample Prognose

Man unterscheidet in-sample und out-of-sample Prognosen.

- **in-sample** Prognosen:

Für in-sample Prognosen wird ein Modell für $t = 1, \dots, T$ geschätzt, die Parameter festgehalten und für die Zeitpunkte der Schätzperiode Prognosen berechnet.

Sie dient zur Überprüfung der Anpassung eines geschätzten Modells an die Daten.

out-of-sample Prognose

- **out-of-sample Prognosen:**
 - Für out-of-sample Prognosen wird eine Schätzperiode vorgegeben, zB $t = 1, \dots, T$, und Werte außerhalb der Schätzperiode prognostiziert.
ZB für $T + \tau$, $\tau = 1, \dots$, also 1-Schritt, 2-Schritt, etc Prognosen.
 - Out-of-sample Prognosen werden oft so durchgeführt, daß die Länge der Schätzperiode festgehalten und diese Periode über die Zeit hinweg verschoben wird. Man nennt das **moving window**.
Den **Prognosehorizont**, zB die 5-Schritt Prognose, hält man fest und bewegt die Prognoseperiode mit.

ex post und ex ante Prognosen

Nach der Verfügbarkeit dritter Variabler kann unterschieden werden:

- **ex post Prognose:** Die Werte der zu erklärenden Variablen können bekannt oder unbekannt sein. Die Werte der dritten (exogenen) Variablen sind bekannt.
- **ex ante Prognose:** Eine Prognose für zukünftige Zeitpunkte, für die keine Werte bekannt sind.

Prognoseintervall

Sind die Fehler ε_t bzw e_t normalverteilt, so kann man auch zB 95% **Prognoseintervalle** für $y_{T+\tau}$ angeben:

$$[\hat{y}_{T+\tau} - 1.96 \hat{\sigma}_e(\tau), \hat{y}_{T+\tau} + 1.96 \hat{\sigma}_e(\tau)]$$

$\hat{\sigma}_e(\tau)$ ist die Standardabweichung des τ -Schritt Prognosefehlers.
Wie man diese bestimmt, werden wir später sehen.

Maße für die Prognosegüte

Mean Error, Mean Square Error

Hat man 2 verschiedene Modelle zur Prognose zur Auswahl, will man wissen, welches der beiden die besseren Prognosen liefert. k bezeichnet die maximale Anzahl der Prognoseschritte.

Mean Error, ME

$$ME = \frac{1}{k} \sum_{\tau=1}^k (y_{T+\tau} - \hat{y}_{T+\tau})$$

Der Mean Error misst den Bias.

Mean Square Error, MSE

$$MSE = \frac{1}{k} \sum_{\tau=1}^k (y_{T+\tau} - \hat{y}_{T+\tau})^2$$

Er ist analog zur Fehlerquadratsumme konstruiert.

Root Mean Square Error, Mean Absolute Error

Root Mean Square Error, Mean Absolute Error

$$RMSE = \sqrt{MSE}$$

Der Root Mean Square Error verhält sich zum Mean Square Error wie die Standardabweichung zur Varianz.

Mean Absolute Error, MAE

$$MAE = \frac{1}{k} \sum_{\tau=1}^k |y_{T+\tau} - \hat{y}_{T+\tau}|$$

Der Mean Absolute Error misst die absoluten Abweichungen, nicht die quadrierten.

Mean Absolute Percent Error

Alle 4 Kriterien können auch als Percent Error definiert werden. ZB:

Mean Absolute Percent Error, MAPE

$$MAPE = \frac{100}{k} \sum_{\tau=1}^k \left| \frac{y_{T+\tau} - \hat{y}_{T+\tau}}{y_{T+\tau}} \right|$$

Hier wird der Fehler in Relation zum realisierten Wert gesetzt.
"Große" Fehler bei "großen" y_t -Werten werden genauso bewertet wie "kleine" Fehler bei "kleinen" Realisationen.

Mittleres Maß für die Prognosegüte

Wenn das selbe Modell für mehrere Reihen, $y_t^{(i)}$, $i = 1, \dots, n$, auf deren Prognosegüte überprüft werden soll, hält man die Schrittlänge τ fest und mittelt über die Werte $\hat{y}_{T+\tau}^{(i)}$, $i = 1, \dots, n$.
ZB

$$MSE_{\tau} = \frac{1}{n} \sum_{i=1}^n [y_{T+\tau}^{(i)} - \hat{y}_{T+\tau}^{(i)}]^2$$

Prognoseexperimente

- **Problem:** Gegeben ist ein bekannter Prozess. Suche unter verschiedenen Modellen das beste heraus.
Lösung: Monte Carlo Simulationen in denen Realisationen des (datengenerierenden) Prozesses erzeugt werden. Für zB 1000 oder 10.000 Reihen werden die Modelle geschätzt und die Prognosen mit den Realisationen verglichen. Dazu kann man einen fixen zukünftigen Zeitpunkt heranziehen. ZB $T + 1$.
- **Problem:** Es liegen viele verschiedene Datensätze vor. Welches von 2 Modellen liefert durchschnittlich die besten Prognoseergebnisse?