Lückenbearbeitung

Zeitreihenanalyse

Fehlende Werte in Zeitreihen ersetzen: Verfahren, Bedienung, Beispiel

Deutsch

Wolfgang Böhme

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

8. November 2026

Dokumentationen dieser Webseite wurden mit Unterstützung der KI Claude (Anthropic) entwickelt.

Inhaltsverzeichnis

1 Einführung **2**

2 Daten laden **2**
2.1 Datenformat 2
2.2 Beispieldaten 2
2.3 Frequenzerkennung 3

3 Die sechs Verfahren **3**
3.1 Lineare Interpolation 3
3.2 Letzter bekannter Wert (LOCF) 3
3.3 Nächster bekannter Wert (NOCB) 3
3.4 Wert einer vorherigen Periode (Lag) 4
3.5 Mittelwert und Median 4

4 Bedienung **4**

5 Durchgerechnetes Beispiel: Aktie_A **5**

6 Hinweise zur Wahl des Verfahrens **6**

1 Einführung

Viele Verfahren der Zeitreihenanalyse – Transformationen, Filter, Modellschätzungen, Strukturbruchtests – setzen lückenlose Reihen voraus. Reale Daten enthalten dagegen häufig fehlende Werte: Feiertage, Meldeverzögerungen, Übertragungsfehler oder Umstellungen in der Statistik. Die *Lückenbearbeitung* ersetzt fehlende Werte nach sechs gängigen Verfahren, zeigt die Ergebnisse aller Verfahren nebeneinander und exportiert die bereinigten Reihen zusammen mit den Originaldaten.

Das Werkzeug arbeitet auf den **Originaldaten** und unabhängig von den übrigen Zeitreihen-Werkzeugen. Es läuft vollständig im Browser; nur die Diagrammbibliothek Chart.js wird beim Öffnen einmalig aus dem Internet geladen.

2 Daten laden

2.1 Datenformat

Geladen wird eine CSV-Datei über „Datei auswählen“:

```
Datum,Aktie_A,Krypto_B,Index_C
2024-01-01,100.5,1000,2500
2024-01-02,102.3,1050,2520
2024-01-03,,980,2510
```

- **Spalte 1** enthält die Zeit (Datum, Jahr oder beliebige Bezeichnung), **ab Spalte 2** folgen die Wertspalten; jede Wertspalte ist eine Zeitreihe.
- Trennzeichen: Semikolon, falls die Kopfzeile ein Semikolon enthält, sonst Komma. Bei Semikolon-Dateien ist das Dezimalkomma erlaubt (`102,3`).
- Als **fehlend** gelten leere Felder, `NaN` und nicht numerische Einträge.
- Zeilen ohne Eintrag in der Zeitspalte werden ignoriert; Anführungszeichen werden entfernt.

Auch Dateien aus dem Zeitreihen-Viewer (Spalte `year` zuerst) lassen sich direkt laden.

2.2 Beispieldaten

„Beispieldaten laden“ lädt 20 Tageswerte (1.–20. Januar 2024) dreier fiktiver Reihen mit insgesamt sechs Lücken:

Tabelle 1: Lücken in den Beispieldaten

Reihe	Lücken	Datum
Aktie_A	3	03.01., 10.01., 17.01.
Krypto_B	2	05.01., 12.01.
Index_C	1	07.01.

2.3 Frequenzerkennung

Das Tool versucht, die Zeitspalte als Datum zu lesen (Formate `JJJJ-MM-TT`, `TT.MM.JJJJ` oder Monatsnamen). Aus dem Median der Abstände zwischen aufeinanderfolgenden Zeitpunkten (in Tagen) bestimmt es die Frequenz:

Tabelle 2: Frequenzerkennung

Median-Abstand m (Tage)	Frequenz
$m \leq 1.5$	täglich
$1.5 < m \leq 9$	wöchentlich
$9 < m \leq 45$	monatlich
$45 < m \leq 200$	quartalsweise
$m > 200$	jährlich

Lässt sich die Zeitspalte nicht als Datum lesen (z. B. reine Jahreszahlen), gilt die Frequenz als „unbekannt (Index)“ mit Schrittweite 1. Die erkannte Frequenz wird nur für die Voreinstellungen des Lag-Verfahrens benötigt.

3 Die sechs Verfahren

Sei $y_1, \ldots, y_n$ die Reihe und y_i ein fehlender Wert.

3.1 Lineare Interpolation

Liegt die Lücke zwischen den bekannten Werten y_s (vorher) und y_e (nachher), gilt

$$\hat{y}_i = y_s + (y_e - y_s)\,\frac{i - s}{e - s}, \qquad s < i < e. \tag{1}$$

Mehrere aufeinanderfolgende Lücken werden gleichmäßig auf der Verbindungsgeraden verteilt. Lücken am **Anfang** der Reihe erhalten den ersten bekannten Wert, Lücken am **Ende** den letzten bekannten Wert. Die Interpolation unterstellt eine gleichmäßige Entwicklung zwischen den Nachbarwerten und ist bei trendbehafteten Reihen meist die beste einfache Wahl.

3.2 Letzter bekannter Wert (LOCF)

Last Observation Carried Forward: $\hat{y}_i = y_s$ mit s = letzter bekannter Zeitpunkt vor i. Typisch für Kurse und Bestände, die sich bis zur nächsten Beobachtung nicht ändern (z. B. Wochenenden). Lücken am Anfang der Reihe bleiben leer, da kein Vorgänger existiert.

3.3 Nächster bekannter Wert (NOCB)

Next Observation Carried Backward: $\hat{y}_i = y_e$ mit e = nächster bekannter Zeitpunkt nach i. Lücken am Ende der Reihe bleiben leer.

3.4 Wert einer vorherigen Periode (Lag)

$\hat{y}_i = y_{i-L}$, also der Wert L Zeitpunkte zuvor. Damit lassen sich saisonale Muster übernehmen (z. B. Vorwoche bei Tagesdaten mit Wochenmuster, Vorjahr bei Monatsdaten). Über die Auswahl „Periode“ wird L aus der erkannten Frequenz berechnet:

$$L = \max\left(1,\ \text{round}\left(\frac{\text{Periodenlänge in Tagen}}{m}\right)\right) \tag{2}$$

mit den Periodenlängen 1 (Vortag), 7 (Vorwoche), 30.44 (Vormonat) und 365.25 (Vorjahr); „Manuell“ erlaubt die freie Eingabe von L. Ist y_{i-L} selbst fehlend oder liegt $i - L$ vor Beginn der Reihe, greift ersatzweise LOCF und danach NOCB – die Reihe wird also immer vollständig gefüllt.

3.5 Mittelwert und Median

Alle Lücken erhalten denselben Wert: das arithmetische Mittel bzw. den Median aller bekannten Werte der Reihe. Diese Verfahren ignorieren den zeitlichen Verlauf und eignen sich nur für Reihen ohne Trend, die um ein festes Niveau schwanken. Der Median ist robuster gegenüber Ausreißern.

4 Bedienung

1. **Zeitreihe auswählen** – Auswahlliste oben.
2. **Verfahren wählen** – bei „Wert einer vorherigen Periode“ zusätzlich Periode bzw. Lag. Darunter stehen die erkannte Frequenz und die Zahl der Lücken.
3. **Vorschau** – die Grafik zeigt die Originalreihe mit Lücken (grau) und die bearbeitete Reihe (weinrot).
4. **Vergleich aller Methoden** – eine zweite Grafik markiert an jeder Lücke die Werte aller sechs Verfahren; das gewählte Verfahren ist als Raute hervorgehoben.
5. **Tabelle** – für jede Lücke die Werte aller Verfahren, das gewählte Verfahren fett.
6. **„Als neue Zeitreihe übernehmen“** – fügt die bereinigte Reihe als neue Spalte hinzu, z. B. „Aktie_A (bereinigt: interpoliert)“. Die Originalreihe bleibt erhalten; mehrere Verfahren können nacheinander übernommen werden.
7. **„Als CSV exportieren“** – speichert Zeitspalte, Originalreihen und alle übernommenen Reihen in einer Datei `lueckenbearbeitung_export_JJJJ-MM-TT.csv` (Komma-getrennt, bis zu sechs Nachkommastellen, noch vorhandene Lücken als leere Felder). Die Datei hat dasselbe Format wie ein Upload.

„Neue Bearbeitung starten“ verwirft alle Daten und kehrt zur Ladeansicht zurück.

5 Durchgerechnetes Beispiel: Aktie_A

Die Reihe Aktie_A hat 20 Tageswerte, davon 17 bekannt und 3 fehlend. Erkannte Frequenz: täglich ($m = 1$).

Tabelle 3: Ausschnitt der Reihe Aktie_A

Datum	02.01.	**03.01.**	04.01.	09.01.	**10.01.**	11.01.	16.01.	**17.01.**	18.01.
Wert	102.3	–	105.2	106.2	–	107.9	117.4	–	119.6

Lineare Interpolation am 03.01.: $102.3 + (105.2 - 102.3) \cdot \frac{1}{2} = 103.75$; am 10.01.: $(106.2 + 107.9)/2 = 107.05$; am 17.01.: $(117.4 + 119.6)/2 = 118.50$.

Mittelwert der 17 bekannten Werte: $1869.0/17 = 109.94$; **Median**: 108.80 (9. Wert der geordneten Reihe).

Tabelle 4: Ersatzwerte aller Verfahren für Aktie_A

Datum	Linear	LOCF	NOCB	Lag (1)	Mittelwert	Median
03.01.	**103.75**	102.30	105.20	102.30	109.94	108.80
10.01.	**107.05**	106.20	107.90	106.20	109.94	108.80
17.01.	**118.50**	117.40	119.60	117.40	109.94	108.80

Lag mit Vorwoche: Bei täglicher Frequenz ergibt die Voreinstellung „Vorwoche“ $L = \text{round}(7/1) = 7$. Für den 03.01. liegt der Wert eine Woche zuvor vor Beginn der Reihe, für den 10.01. und 17.01. ist der Vorwochenwert (03.01. bzw. 10.01.) selbst fehlend. In allen drei Fällen greift daher LOCF, und das Ergebnis stimmt mit der LOCF-Spalte überein – ein Beispiel dafür, dass das Lag-Verfahren ausreichend lange Reihen ohne gehäufte Lücken voraussetzt.

Bewertung: Aktie_A hat einen deutlichen Aufwärtstrend (von 100.5 auf 122.8). Mittelwert und Median setzen in der Anfangsphase viel zu hohe und am Ende zu niedrige Werte ein (am 03.01. rund 5–8 Einheiten über den Nachbarwerten) und verzerren damit Renditen und Volatilität. LOCF und NOCB übernehmen jeweils einen Nachbarwert und erzeugen eine Nullrendite bzw. eine verdoppelte Rendite an einem Tag. Die lineare Interpolation liegt zwischen den Nachbarwerten und verteilt die Veränderung gleichmäßig – für diese Reihe die plausibelste Wahl.

6 Hinweise zur Wahl des Verfahrens

Tabelle 5: Empfehlungen

Situation	Empfohlenes Verfahren
Trend, einzelne Lücken	lineare Interpolation
Kurse, Bestände (Wert gilt bis zur nächsten Meldung)	LOCF
Nachträglich gemeldete Werte, die rückwirkend gelten	NOCB
Ausgeprägte Saison (Woche, Jahr)	Lag mit passender Periode
Stationäre Reihe ohne Trend	Mittelwert; bei Ausreißern Median
Lange Lücken (viele Perioden)	keines der Verfahren ist verlässlich – Zeitraum kürzen oder modellbasiert schätzen (z. B. Zeitreihenmodellierung)

Wichtig: Ersetzte Werte sind keine Beobachtungen. Sie glätten die Reihe (Interpolation), erzeugen künstliche Stillstände (LOCF/NOCB) oder verringern die Streuung (Mittelwert/Median). Bei vielen Lücken sollten Ergebnisse späterer Analysen mit mehreren Verfahren auf Robustheit geprüft werden; der Methodenvergleich im Tool zeigt, wie stark die Ersatzwerte voneinander abweichen.