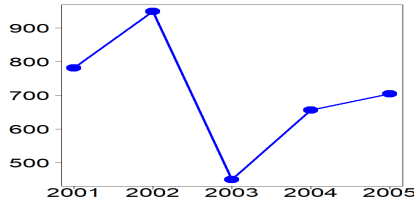


Jährlicher Niederschlag in Regensburg (von HISTALP):

2001	2002	2003	2004	2005
781 mm	949 mm	450 mm	656 mm	704 mm



Wir interpretieren die fünf Messwerte als Realisierung einer Folge

$$y_1, y_2, y_3, y_4, y_5$$

von fünf Zufallsvariablen

$$y_t: \Omega \rightarrow \mathbb{R}, t = 1, \dots, 5.$$

Für jedes  $\omega \in \Omega$  erhält man eine andere Realisierung

$$y_1(\omega), y_2(\omega), y_3(\omega), y_4(\omega), y_5(\omega).$$

Die Folge der fünf Zufallsvariablen ist eine Zeitreihe.

Zeitreihe der Länge  $n$ :

$$y_1, y_2, y_3, \dots, y_n$$

$y_1$  ... Beobachtung zum Zeitpunkt 1 (oder in Periode 1)

$y_2$  ... Beobachtung zum Zeitpunkt 2 (oder in Periode 2)

$y_3$  ... Beobachtung zum Zeitpunkt 3 (oder in Periode 3)

⋮

$y_n$  ... Beobachtung zum Zeitpunkt  $n$  (oder in Periode  $n$ )

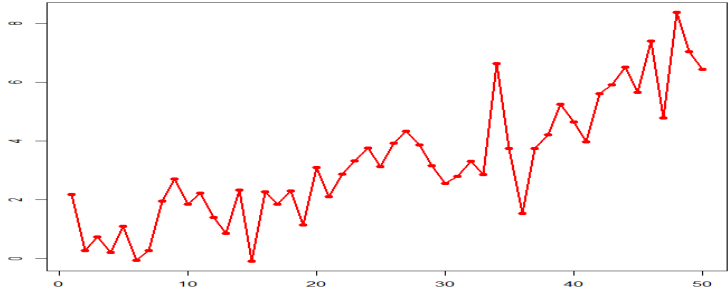
Idealisierend nehmen wir an, dass sich Zeitreihen immer in die unendliche Zukunft und in die unendliche Vergangenheit erstrecken, auch wenn in der Praxis natürlich nur ein endlicher Ausschnitt beobachtbar ist:

$$\dots, y_{-1}, y_0, y_1, y_2, y_3, \dots, y_n, y_{n+1}, y_{n+2}, \dots$$

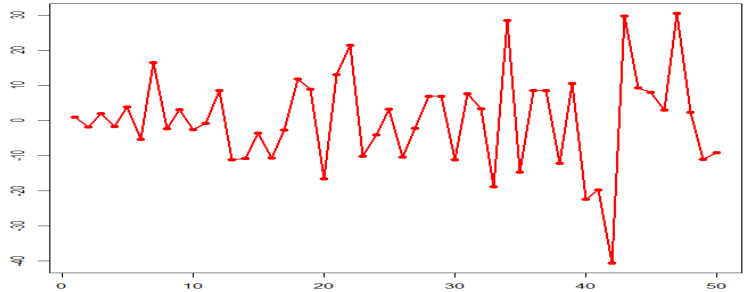
Bei einer stationären Zeitreihe haben alle Beobachtungen  $y_t$  denselben Erwartungswert und die Kovarianz zwischen zwei Beobachtungen  $y_t$  und  $y_{t-k}$  hängt nur von ihrem zeitlichen Abstand  $k$  ab, d.h.,

- (i)  $E(y_t) = \mu \quad \forall t \in \mathbb{Z}$ ,
- (ii)  $Cov(y_t, y_{t-k}) = \gamma(k) \quad \forall t, k \in \mathbb{Z}$ .

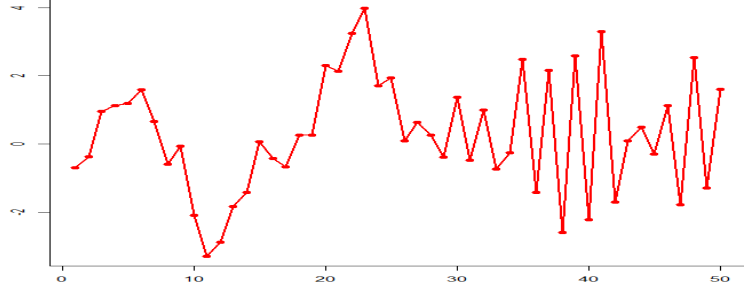
**Realisierungen nichtstationärer Zeitreihen:**



Der Erwartungswert ist nicht konstant ist, weil es einen wachsenden Trend gibt, d.h.,  $E(y_s) < E(y_t)$ , falls  $s < t$ .

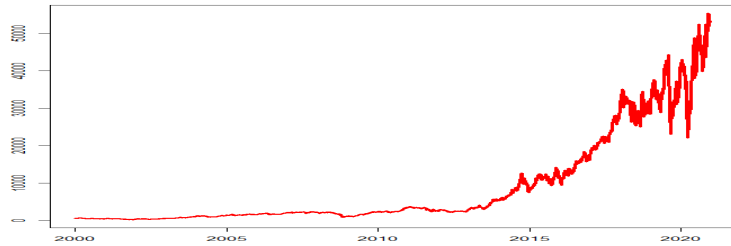


Die Varianz  $\gamma(0)$  ist nicht konstant, sondern wächst.

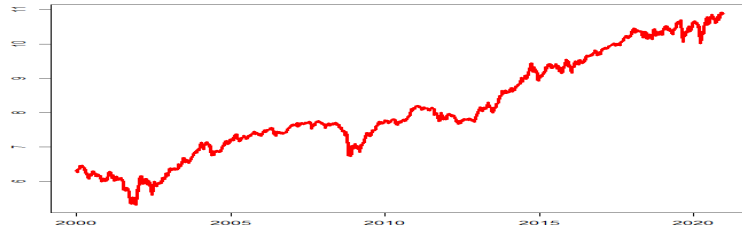


Aufeinanderfolgende Beobachtungen sind zuerst positiv korreliert ( $\gamma(1) > 0$ ) und später negativ ( $\gamma(1) < 0$ ).

Der argentinische Aktienindex Merval (von Yahoo Finance) weist einen ansteigenden Trend auf. Auch die Schwankungen um diesen Trend werden immer größer.

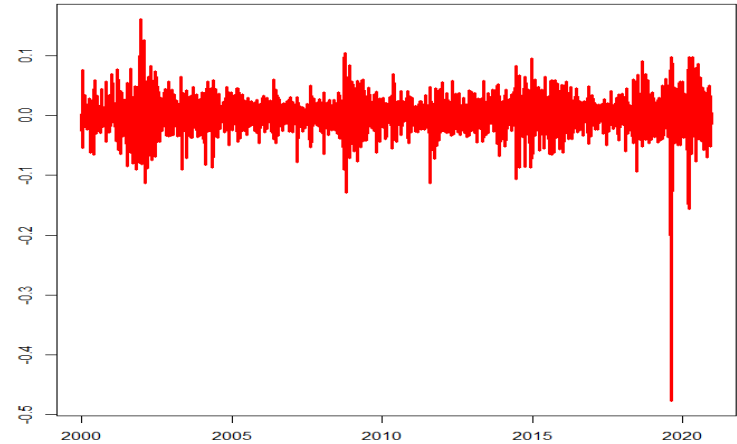


Durch Logarithmieren erhält man eine neue Zeitreihe, die statt eines annähernd exponentiellen Trends einen annähernd linearen Trend aufweist. Die Schwankungen um diesen linearen Trend bleiben annähernd gleich groß.

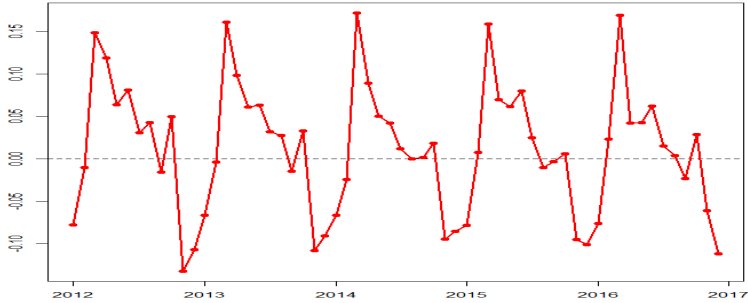


Da bei einer Geraden die Steigung überall gleich groß ist, verschwindet ein linearer Trend, wenn man statt der ursprünglichen Beobachtungen die Zuwächse gegenüber den Vorperioden betrachtet.

Wenn man die Differenzen  $y_t = \log(P_t) - \log(P_{t-1})$  der logarithmierten Merval-Kurse plottet, ist tatsächlich kein Trend zu erkennen. Auffällig sind nur vereinzelte, für Finanzdaten typische Ausreißer.

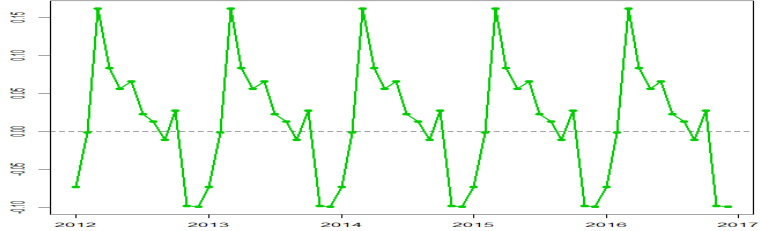


In den logarithmierten und differenzierten U.S. Wohnbauausgaben (vom U.S. Census Bureau via FRED) ist immer noch ein Trend zu erkennen, allerdings kein wachsender oder ein fallender, sondern ein zyklischer.

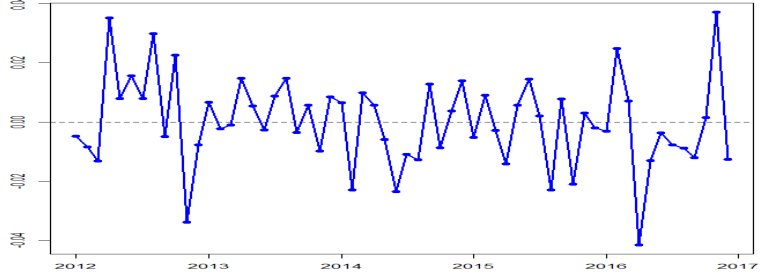


Mit zunehmender saisonaler Erwärmung von März bis Juni steigen die Wohnbauausgaben, erkennbar an den großen positiven Werten in diesen Monaten. Mit zunehmender Abkühlung von November bis Jänner sinken die Wohnbauausgaben, erkennbar an den großen negativen Werten in diesen Monaten. Dieses saisonale Muster ist nicht zufällig, sondern wiederholt sich jedes Jahr. Der Erwartungswert ist also nicht konstant, sondern hängt vom jeweiligen Monat ab.

Das saisonale Muster kann z.B. geschätzt werden durch Mitteln aller Jänner-Werte, aller Februar-Werte usw.



Für die saisonbereinigte Zeitreihe (blau), die wir durch Subtraktion des geschätzten saisonalen Musters (grün) von den logarithmierten und differenzierten U.S. Wohnbauausgaben (rot) erhalten, ist die Annahme eines konstanten Erwartungswerts plausibel.



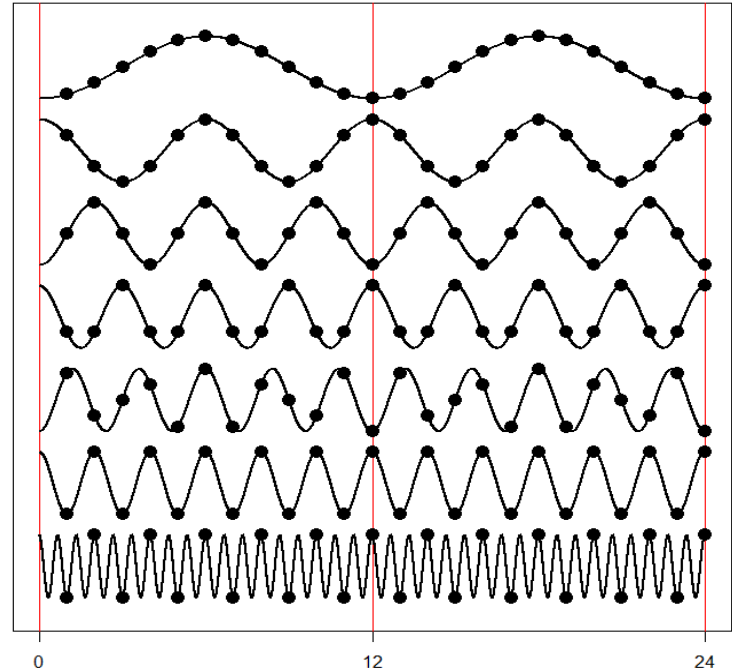
Ein saisonales Muster kann auch durch Anpassung einer periodischen Funktion an die Daten geschätzt werden. Eine einfache, aber trotzdem flexible periodische Funktion ist das Sinusoid

$$g(t) = R \sin(\omega t + \phi).$$

Die Amplitude  $R$  bestimmt den Abstand zwischen dem Minimum und dem Maximum des Sinusoids, die Frequenz  $\omega$  bestimmt, wie viele Minima und Maxima es in einer gegebenen Zeitspanne gibt, und die Phase bestimmt die Lage der Minima und Maxima.

Das Sinusoid mit der Frequenz  $\omega = 2\pi/12$  ist periodisch mit der Periode 12. Es hat in einem Zeitintervall der Länge 12 ein Minimum und ein Maximum. Das Sinusoid mit der doppelten Frequenz  $2 \cdot (2\pi/12)$  ist auch periodisch mit der Periode 12. Es hat in einem Zeitintervall der Länge 12 zwei Minima und zwei Maxima. Das Sinusoid mit der dreifachen Frequenz  $3 \cdot (2\pi/12)$  ist auch periodisch mit der Periode 12. Es hat in einem Zeitintervall der Länge 12 drei Minima und drei Maxima ...

Im Fall von Monatsdaten stoßen wir allerdings bei sechs Gipfeln pro Jahr bereits an die Grenzen der Auflösung.

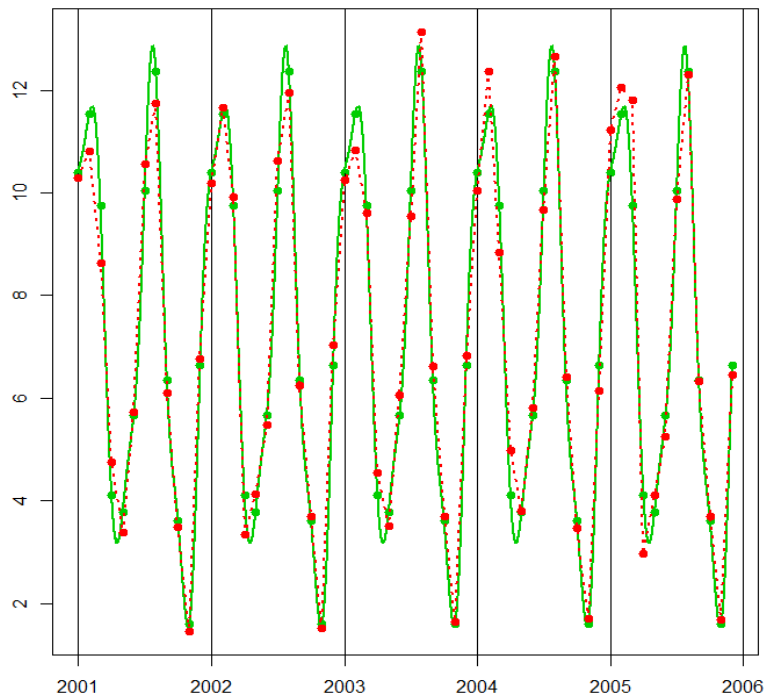


Im Fall von komplexeren saisonalen Mustern in monatlichen Zeitreihen kann man bis zu sechs Sinusoide mit den Frequenzen  $(2\pi/12) \cdot k$  und den Perioden  $12/k$ ,  $k = 1, \dots, 6$ , kombinieren. Für  $k = 6$  erhalten wir die höchstmögliche Frequenz  $\pi$  und die kürzestmögliche Periode 2. Mehr als ein Minimum und ein Maximum können wir in zwei Monaten nicht beobachten.

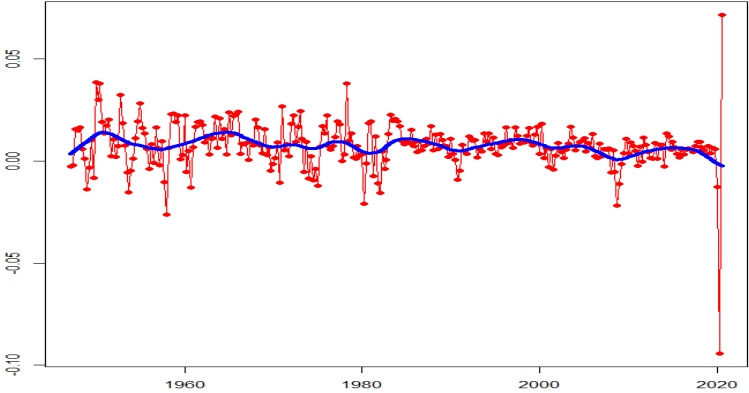
Die Anzahl von Ausländer-Nächtigungen in Österreich (von OeNB) weist zwei Gipfel pro Jahr auf, einen im Sommer und einen im Winter. Mit Hilfe des Modells

$$y_t = \mu + \sum_{k=1}^6 R_k \sin\left(\frac{2\pi k}{12} t + \phi_k\right) + u_t$$

kann das komplexe saisonale Muster in dieser Zeitreihe (rote Kurve) recht gut beschrieben werden (grüne Kurve).



Nach dem Logarithmieren und Differenzieren sieht das saisonbereinigte reale U.S. Bruttoinlandsprodukt (vom U.S. Bureau of Economic Analysis via FRED) halbwegs stationär aus, wenn man von der geringeren Volatilität seit den frühen 80er Jahren (Great Moderation) absieht. Auffällig sind auch die durch die Corona-Krise verursachten Ausreißer am Ende der Zeitreihe (2. und 3. Quartal 2020). Diese zeigen, dass die Normalverteilungsannahme nicht nur für Finanzzeitreihen, sondern auch für makroökonomische Zeitreihen unplausibel ist.

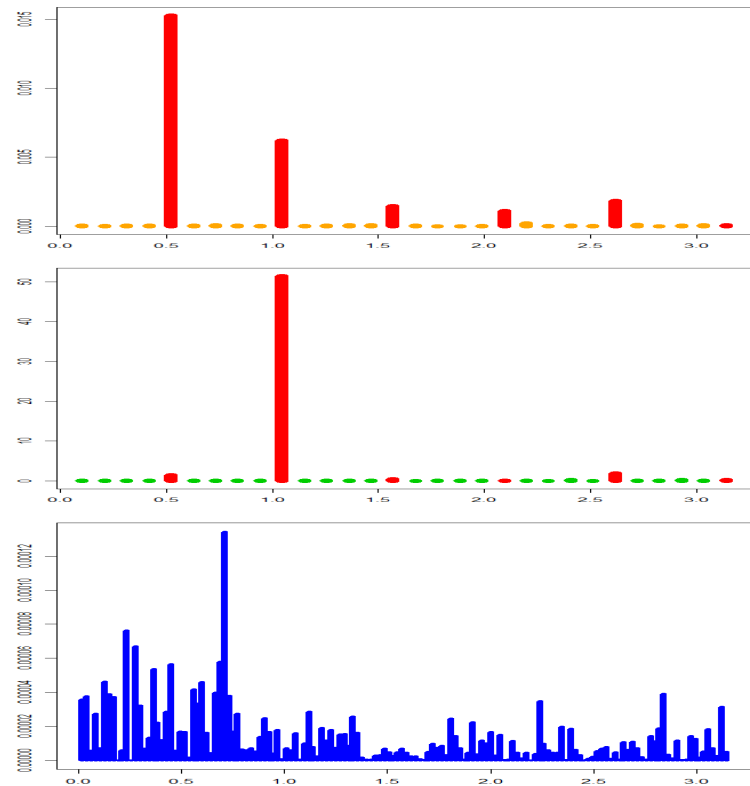


Glättet man diese Zeitreihe, so erkennt man Schwingungen unterschiedlicher Dauer, die man als Konjunkturzyklen interpretiert. Das Ergebnis der Glättung (blaue Kurve) hängt nur unwesentlich davon ab, wie man glättet (z.B. durch Mitteln benachbarter Beobachtungen), dafür aber sehr stark vom Grad der Glättung (z.B. der Anzahl der Beobachtungen, über die man mittelt).

Da es sich bei den Konjunkturzyklen nicht um strikt periodische Schwingungen handelt, kann ein einzelnes Sinusoid nur einen sehr kleinen Teil der Gesamtvarianz erklären. Die geschätzte Amplitude wird also selbst dann sehr klein sein, wenn die Periode des Sinusoids mit der Länge eines Konjunkturzyklus übereinstimmt, und noch viel kleiner, wenn das nicht der Fall ist.

Bei einer Spektralanalyse passen wir viele Sinusoide mit unterschiedlichen Frequenzen bzw. Perioden an eine Zeitreihe an und vergleichen die geschätzten Amplituden. Wir beschränken uns dabei auf Fourier-Frequenzen im Intervall  $(0, \pi]$ . Die erste Fourier-Frequenz  $\omega_1 = 2\pi/n$  entspricht der Periode  $n$ , die zweite Fourier-Frequenz  $\omega_2 = 2 \cdot \omega_1$  entspricht der Periode  $n/2$ , die dritte Fourier-Frequenz  $\omega_3 = 3 \cdot \omega_1$  entspricht der Periode  $n/3$  usw. Die Anzahl der Fourier-Frequenzen hängt von der Länge  $n$  der Zeitreihe ab. Das Periodogramm ordnet jeder Fourier-Frequenz die quadrierte geschätzte Amplitude zu (bis auf einen unbedeutenden Skalierungsfaktor).

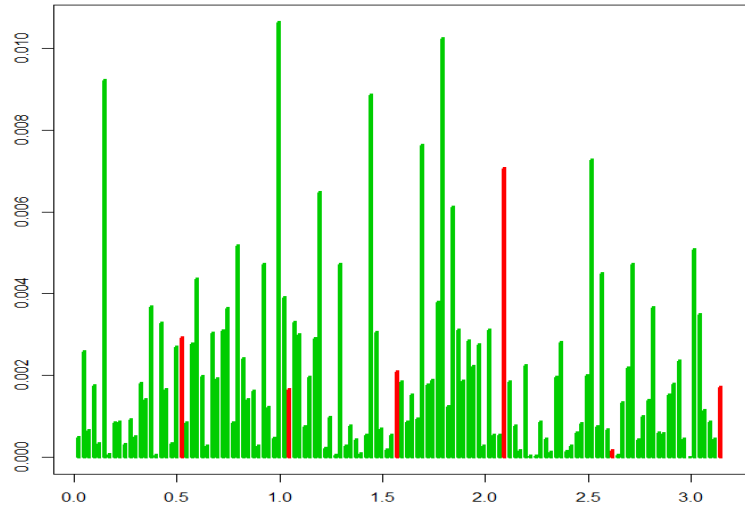
Im Periodogramm der diff. log U.S. Wohnbauausgaben (gelb) erkennt man Spitzen bei den saisonalen Frequenzen  $2\pi k/12$ ,  $k = 1, \dots, 6$  (rot), analog auch im Periodogramm der Ausländer-Nächtigungen in Ö (grün). Hingegen gibt es im Periodogramm des diff. log U.S. BIP exkl. 2020 (blau) keine isolierten Spitzen, also keine Hinweise auf strikte Periodizitäten. Allerdings sind hier wegen der Konjunkturzyklen die Amplituden im niedrigfrequenten Bereich im Allgemeinen größer als im höherfrequenten Bereich.





Bei den U.S. Wohnbauausgaben und den Ausländer-Nachtungen in Österreich reicht schon ein Blick auf die Plots der Zeitreihen, um die saisonalen Muster zu erkennen. In anderen Fällen, wo es nicht von vornherein klar ist, ob es strikte Periodizitäten gibt, kann die Spektralanalyse helfen.

Im Folgenden betrachten wir den Merval und untersuchen, ob es ein saisonales Muster in den Log>Returns gibt (d.h. in den Differenzen der logarithmierten Merval-Kurse). Wir verwenden Monatsdaten statt Tagesdaten, weil die jährliche Anzahl der Handelstage nicht konstant ist. Um zu erreichen, dass die sechs saisonalen Frequenzen  $2\pi/12$ ,  $2 \cdot 2\pi/12$ , ...,  $6 \cdot 2\pi/12$  mit den Perioden 12, 6, ..., 2 auch Fourier-Frequenzen sind und somit im Periodogramm aufscheinen, müssen wir eine durch 12 teilbare Zeitreihenlänge  $n$  wählen. Analog müsste im Fall von Quartalsdaten  $n$  durch 4 teilbar sein, damit die beiden saisonalen Frequenzen  $2\pi/4 = \pi/2$  und  $2 \cdot 2\pi/4 = \pi$  mit den Perioden 4 und 2 Fourier-Frequenzen sind.



Im Periodogramm der monatlichen Log>Returns erkennen wir nur eine größere Periodogrammordinate bei einer saisonalen Frequenz, nämlich bei der vierten. Die restlichen fünf sind unauffällig. Insgesamt ist das zu wenig. Wenn wir einen formalen statistischen Test ausführen, beispielsweise mit der Summe aller sechs Ordinaten als Teststatistik, werden wir wohl kein signifikantes Testergebnis erhalten.

Anstatt im Periodogramm nach größeren Werten bei einzelnen Frequenzen (diskrete Muster) zu suchen, können wir auch nach Clustern von größeren Werten (stetige Muster) suchen. Oft sind solche Cluster im kumulativen Periodogramm leichter zu entdecken,

Dass die Periodogrammordinaten der täglichen Log>Returns des Merval (grün) in der Nähe der Frequenz 0 ein bisschen größer sind, erkennt man (mit etwas Phantasie und viel gutem Willen) im kumulativen Periodogramm (rot) an einem etwas steileren Anstieg in diesem Bereich.

Für einen formalen statistischen Test könnte man beispielsweise die Summe der Periodogrammordinaten bei den niedrigsten 10%, 15% oder 20% der Fourier-Frequenzen als Teststatistik verwenden.

