

Einfaches Regressionsmodell:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

Beobachtungen: $(y_1, x_1), \dots, (y_n, x_n)$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Autoregressives (AR) Modell der Ordnung 1:

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t$$

Beobachtungen: y_1, \dots, y_n

$$\begin{pmatrix} y_2 \\ \vdots \\ y_n \end{pmatrix} = \phi_0 + \phi_1 \begin{pmatrix} y_1 \\ \vdots \\ y_{n-1} \end{pmatrix} + \begin{pmatrix} u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Autoregressives Modell der Ordnung 2:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

Beobachtungen: y_1, \dots, y_n

$$\begin{pmatrix} y_3 \\ \vdots \\ y_n \end{pmatrix} = \phi_0 + \phi_1 \begin{pmatrix} y_2 \\ \vdots \\ y_{n-1} \end{pmatrix} + \phi_2 \begin{pmatrix} y_1 \\ \vdots \\ y_{n-2} \end{pmatrix} + \begin{pmatrix} u_3 \\ \vdots \\ u_n \end{pmatrix}$$

Zufallsstichprobe: $(y_1, x_1), \dots, (y_n, x_n)$

Theoretische Kovarianz:

$$\gamma = E(x_t - Ex_t)(y_t - Ey_t)$$

Empirische Kovarianz:

$$\hat{\gamma} = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})$$

Stationäre Zeitreihe: y_1, \dots, y_n

Theoretische Autokovarianz zum Lag 1:

$$\gamma(1) = E(y_t - Ey_t)(y_{t-1} - Ey_{t-1})$$

Empirische Autokovarianz zum Lag 1:

$$\hat{\gamma}(1) = \frac{1}{n} \sum_{t=2}^n (y_t - \bar{y})(y_{t-1} - \bar{y})$$

Theoretische Autokovarianz zum Lag k :

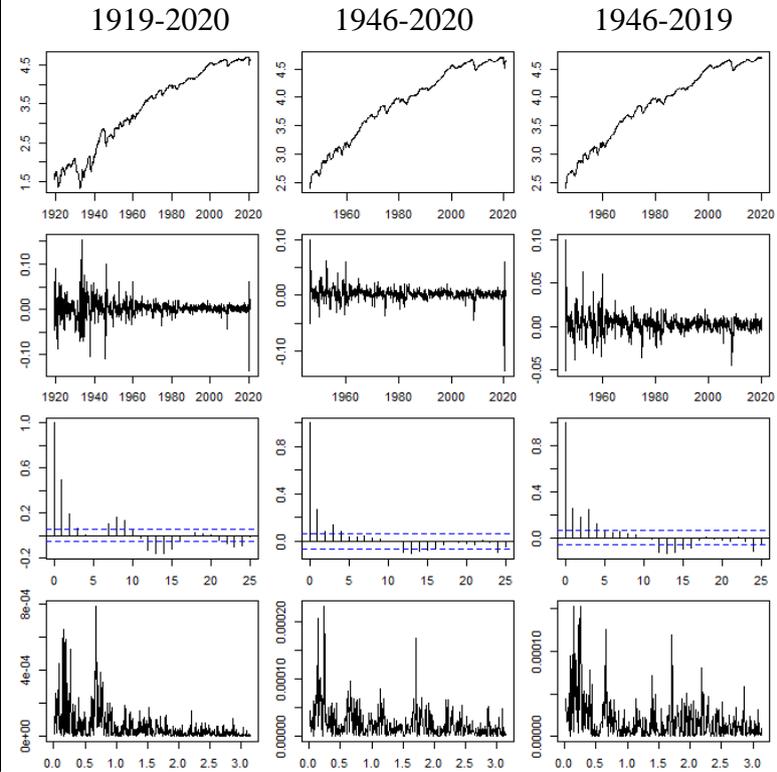
$$\begin{aligned} \gamma(k) &= E(y_t - Ey_t)(y_{t-k} - Ey_{t-k}) \\ &= E(y_t - Ey_t)(y_{t+k} - Ey_{t+k}) = \gamma(-k) \end{aligned}$$

Empirische Autokovarianz zum Lag k :

$$\begin{aligned} \hat{\gamma}(k) &= \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}) \\ &= \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) = \hat{\gamma}(-k) \end{aligned}$$

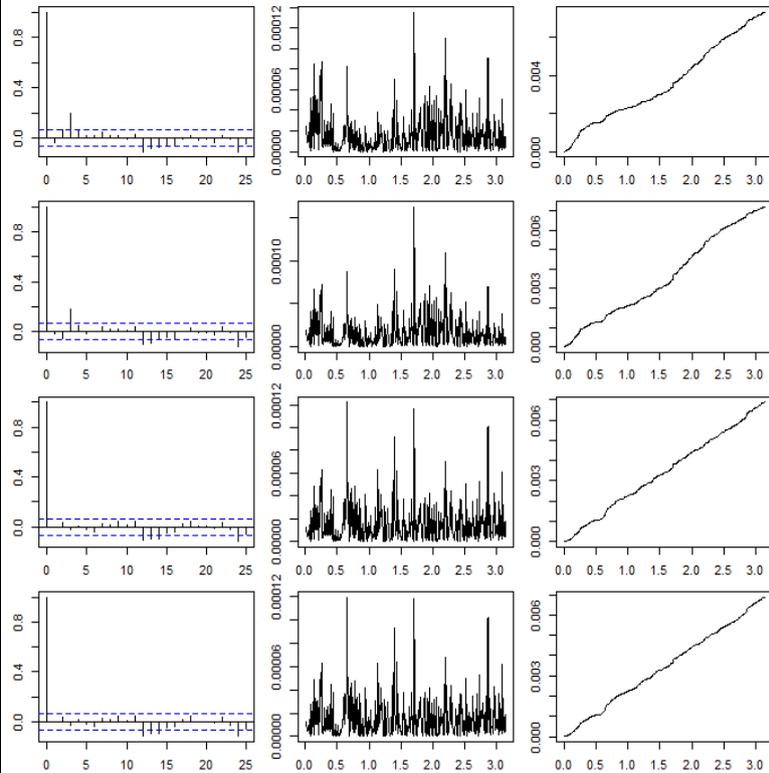
Für einen realen, saisonbereinigten Index der industriellen Produktion in den U.S.A (von FRED) plotten wir die logarithmierten Indexzahlen $Y_t = \log(IP_t)$, $t = 0, \dots, n$, ihre Differenzen $y_t = Y_t - Y_{t-1}$, $t = 1, \dots, n$, sowie die empirischen Autokorrelationen $\hat{\gamma}(k)/\hat{\gamma}(0)$, $k = 1, \dots, 25$, und das Periodogramm $I(\omega_k)$, $0 < \omega_k \leq \pi$, der Differenzen.

Wenn man die insbesondere im Plot der Differenzen gut erkennbaren Perioden hoher Volatilität vor 1946 und nach 2019 weglässt, wird das Periodogramm deutlich „weißer“. Es ist dann viel flacher und enthält kaum noch nennenswerte Charakteristiken, wenn man von der durch lange Konjunkturzyklen bedingten Häufung großer Periodogrammordinaten im niedrigfrequenten Bereich absieht.



Wenn wir autoregressive Modelle wachsender Ordnung (von $p = 1$ bis $p = 4$) an die approximativen Wachstumsraten y_1, \dots, y_n der industriellen Produktion in den U.S.A. anpassen und die empirischen Autokorrelationen, das Periodogramm sowie das kumulative Periodogramm der Residuen plotten, dann sehen wir, dass auffällige Muster sukzessive verschwinden (in der Abbildung von oben nach unten).

Beispielsweise gibt es den großen positiven Autokorrelationskoeffizienten beim Lag 3 nur bis zur Ordnung $p = 2$ (zweite Zeile), ebenso den Cluster großer Periodogrammordinaten zwischen den Frequenzen 1.7 und 2.2, der auch im kumulativen Periodogramm zu erkennen ist (durch einen deutlich steileren Anstieg in diesem Bereich).



Um die Konstante ϕ_0 aus dem autoregressiven Modell entfernen zu können, nehmen wir im Folgenden an, dass der Erwartungswert μ der zu analysierenden stationären Zeitreihe praktisch gleich null ist (z.B. im Fall von täglichen Aktien>Returns oder Intraday>Returns) oder die Zeitreihe bereits mittelwertbereinigt wurde.

Das AR(p)-Modell

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t$$

erklärt die Beobachtung y_t bis auf den Fehler u_t . Ist dieser Fehler mit den weiter zurückliegenden Beobachtungen $y_{t-(p+1)}$, $y_{t-(p+2)}$ usw. korreliert, dann ist es sinnvoll, die Ordnung des AR-Modells zu erhöhen.

Alternativ kann man versuchen, den Fehler zu verkleinern, indem man vergangene Fehler in das Modell aufnimmt. Man erhält dann ein autoregressives Moving Average (ARMA) Modell der Form

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q},$$

das das AR-Modell (für $q = 0$) und das MA-Modell (für $p = 0$) als Spezialfälle enthält.

Wendet man den Lag-Operator L auf eine Zeitreihe y an, dann erhält man eine neue Zeitreihe $z = Ly$, die der ursprünglichen Zeitreihe um eine Zeiteinheit hinterherhinkt:

$$z_t = Ly_t = y_{t-1}$$

Analog:

$$L^2 y_t = LLy_t = y_{t-2}$$

$$L^3 y_t = LLLy_t = y_{t-3}$$

⋮

Mit Hilfe des Lag-Operators schreiben wir ein ARMA-Modell auch als

$$y_t = \phi_1 Ly_t + \dots + \phi_p L^p y_t + u_t + \theta_1 Lu_t + \dots + \theta_q L^q u_t$$

bzw. als

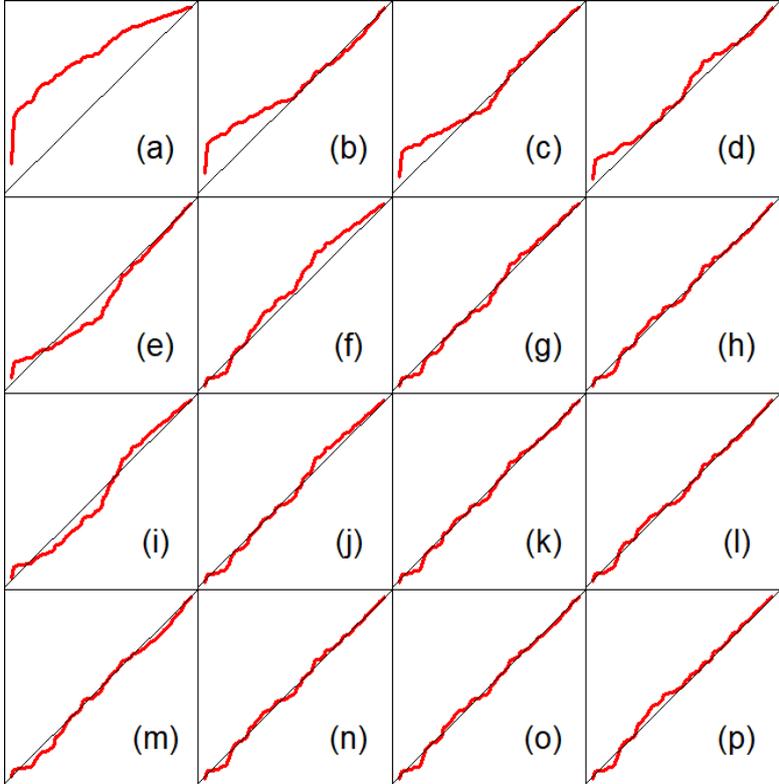
$$y_t - \phi_1 Ly_t - \dots - \phi_p L^p y_t = u_t + \theta_1 Lu_t + \dots + \theta_q L^q u_t$$

und abgekürzt als

$$(1 - \phi_1 L - \dots - \phi_p L^p) y_t = (1 + \theta_1 L + \dots + \theta_q L^q) u_t.$$

Die Parameter eines ARMA-Modells können nicht so ohne Weiteres mit OLS geschätzt werden, weil nur die y_t beobachtbar sind, aber nicht die u_t . Eine simple Lösung für dieses Problem besteht darin, zuerst ein AR-Modell höherer Ordnung mit OLS zu schätzen und die Residuen \hat{u}_t dieses Modells dann als Proxys für die u_t zu verwenden. Vorzuziehen ist aber die Verwendung einer effizienteren Schätzmethode wie z.B. ML.

Wenn wir ARMA-Modelle bis zur Ordnung $(p, q) = (3,3)$ an die jährlichen Temperaturmittel in Tromsø (von 1880 bis 2019) anpassen und die kumulativen Periodogramme der Residuen plotten, sehen wir, dass für einen annähernd linearen Anstieg (bzw. für annähernd unkorrelierte Residuen) mindestens drei Parameter nötig sind, z.B. (g) ARMA(1,2), (j) ARMA(2,1), (m) AR(3). Das einzige ganz schlechte Modell mit mehr als zwei Parametern ist (d) MA(3). Im Plot entspricht die AR-Ordnung p der Zeilenzahl minus 1 und die MA-Ordnung q der Spaltenzahl minus 1. Die Verwendung größerer Modelle mit mehr als drei Parametern bringt keine wesentliche Verbesserung.



Modellwahl mittels Minimierung der Summe der Fehlerquadrate oder der negativen Log-Likelihood L ist nicht sinnvoll, weil man damit zwangsläufig immer beim größten Modell landet. Mit einem zusätzlichen Parameter erreicht man zwar immer eine Verbesserung der Anpassung, gleichzeitig steigt aber auch die Varianz. Diesem Trade-off zwischen Bias und Varianz kann man am besten Rechnung tragen, wenn man ein möglichst kleines Modell wählt, das aber noch alle wesentlichen Charakteristiken der Daten erklärt. Bei dieser schwierigen Entscheidung kann man zwar die beliebten Modellwahlkriterien

$$\text{AIC} = -2L + 2(p + q + 1)$$

und

$$\text{BIC} = -2L + (p + q + 1) \log(n)$$

zu Rate ziehen, blind vertrauen sollte man ihnen aber nicht. BIC bestraft einen zusätzlichen Parameter mit $\log(n)$, AIC hingegen nur mit 2, wählt also typischerweise ein größeres Modell. Das ist auch bei den Temperaturdaten der Fall. Hier wählt BIC ARMA(1,1) und AIC ARMA(1,2), allerdings nur knapp vor ARMA(1,1) und ARMA(2,1).

Wenn wir nur AR-Modelle betrachten (1. Spalte) bzw. nur MA-Modelle (1. Zeile), dann sehen wir, dass laut AIC und BIC jeweils die größten Modelle die besten sind. Das könnte ein Hinweis darauf sein, dass die maximale Modellordnung ($p = 3$ bzw. $q = 3$) zu klein gewählt wurde.

		$q = 0$	$q = 1$	$q = 2$	$q = 3$
AIC	$p = 0$	406.4871	376.0429	372.8178	366.6935
	$p = 1$	362.7260	341.5743	341.5619	342.5085
	$p = 2$	359.6461	341.9721	342.7773	344.3453
	$p = 3$	351.6526	342.6683	344.4861	344.9871
BIC	$p = 0$	409.4216	381.9119	381.6212	378.4314
	$p = 1$	368.5950	350.3777	353.2998	357.1809
	$p = 2$	368.4495	353.7100	357.4496	361.9522
	$p = 3$	363.3904	357.3407	362.0930	365.5284

Unsere visuelle Analyse der kumulativen Periodogramme hat ergeben, dass ARMA(2,1), ARMA(1,2) und AR(3) (in den Tabellen rot) die kleinsten Modelle sind, die annähernd unkorrelierte Residuen produzieren. Bezüglich der ersten beiden Modelle gibt es eine Übereinstimmung mit AIC.

Ist y stationär und erfüllt die AR(1)-Gleichung

$$y_t = \phi y_{t-1} + u_t,$$

wobei u weißes Rauschen ist und $|\phi| < 1$, dann gilt

$$\begin{aligned} y_t &= \phi y_{t-1} + u_t \\ &= \phi(\phi y_{t-2} + u_{t-1}) + u_t \\ &= \phi^2 y_{t-2} + \phi u_{t-1} + u_t \\ &= \phi^2(\phi y_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t \\ &= \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\ &\quad \vdots \\ &= \sum_{k=0}^{\infty} \phi^k u_{t-k}. \end{aligned}$$

Die Varianz von y ist gegeben durch:

$$\begin{aligned} \gamma(0) &= \text{var}(\sum_{k=0}^{\infty} \phi^k u_{t-k}) = \sum_{k=0}^{\infty} \text{var}(\phi^k u_{t-k}) \\ &= \sum_{k=0}^{\infty} \phi^{2k} \text{var}(u_{t-k}) = \sigma^2 \sum_{k=0}^{\infty} (\phi^2)^k \\ &= \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

Allgemein sind die Autokovarianzen von y gegeben durch

$$\gamma(k) = \frac{\sigma^2 \phi^{|k|}}{1-\phi^2}$$

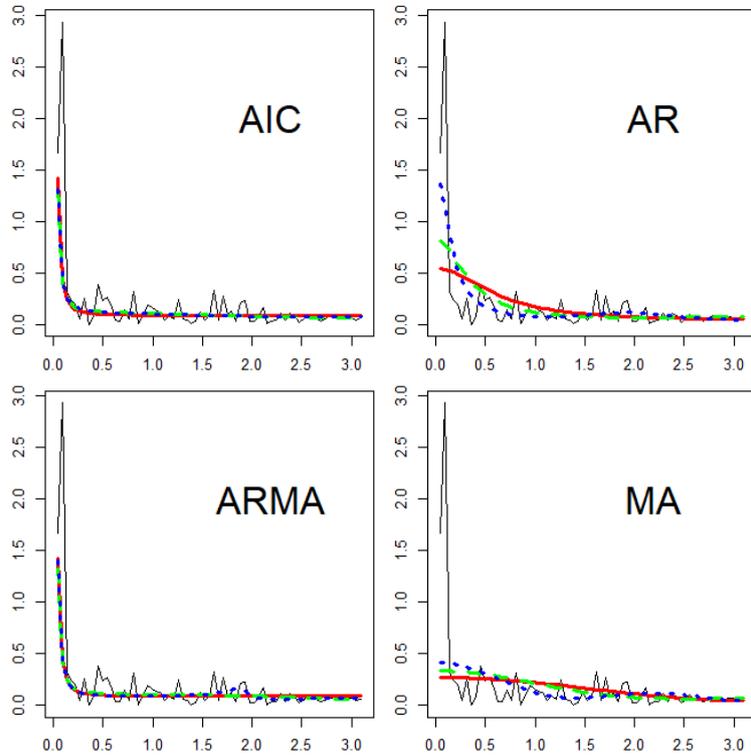
und die Spektraldichte von y durch

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} (\sum_{k=0}^{\infty} \phi^k e^{i\omega k} + \sum_{k=0}^{\infty} \phi^k e^{-i\omega k} - 1) \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \left(\frac{1}{1-\phi e^{i\omega}} + \frac{1}{1-\phi e^{-i\omega}} - 1 \right) \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \frac{1-\phi e^{-i\omega} + 1-\phi e^{i\omega} - (1-\phi e^{i\omega})(1-\phi e^{-i\omega})}{(1-\phi e^{i\omega})(1-\phi e^{-i\omega})} \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \frac{1-\phi^2}{(1-\phi e^{i\omega})(1-\phi e^{-i\omega})} \\ &= \frac{\sigma^2}{2\pi} \left| \frac{1}{1-\phi e^{-i\omega}} \right|^2. \end{aligned}$$

ARMA-Spektraldichte:

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{1 + \theta_1 e^{-i\omega 1} \dots + \theta_q e^{-i\omega q}}{1 - \phi_1 e^{-i\omega 1} \dots - \phi_p e^{-i\omega p}} \right|^2$$

Die geschätzten Spektraldichten der drei von AIC favorisierten Modelle ARMA(1,2), ARMA(1,1) und ARMA(2,1) unterscheiden sich kaum. Alle steigen steil zur Frequenz 0 hin an. Es ändert sich auch nichts, wenn man die Ordnung wachsen lässt, von ARMA(1,1) über ARMA(2,2) bis ARMA(3,3). Lediglich reine AR Modelle und reine MA Modelle sind mit dem steilen Anstieg überfordert. Mit wachsendem p bzw. q (1: rot; 2: grün, 3: blau) nimmt die Steilheit zwar zu, für eine im niedrigfrequenten Bereich zufriedenstellende Anpassung an das Periodogramm würde man aber eine wesentlich größere Modellordnung benötigen (insbesondere für MA).



Ein steiler Anstieg des Periodogramms zur Frequenz 0 hin kann ein Hinweis auf einen Pol der Spektraldichte an der Stelle 0 sein. Im Fall eines Pols hängt die Stationarität von der Steilheit des Anstiegs ab. Ist der Anstieg so steil, dass die Spektraldichte nicht integrierbar ist, dann existiert auch die Varianz $\gamma(0)$ nicht und die Zeitreihe kann nicht stationär sein.

$$\begin{aligned}
 & \int_{-\pi}^{\pi} f(\omega) d\omega \\
 &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} d\omega \\
 &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \int_{-\pi}^{\pi} e^{-i\omega k} d\omega \\
 &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \int_{-\pi}^{\pi} (\cos(\omega k) - i \sin(\omega k)) d\omega \\
 &= \gamma(0) \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos(0) - i \sin(0)) d\omega \\
 &= \gamma(0) \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 d\omega \\
 &= \gamma(0)
 \end{aligned}$$

Um abschätzen zu können, ob eine Zeitreihe stationär ist, vergleichen wir ihr Periodogramm in der Nähe der Frequenz 0 mit der einfachen Funktion

$$f(\omega) = \omega^{-b},$$

die integrierbar ist, wenn $b < 1$, und nicht integrierbar ist, wenn $b \geq 1$.

Von besonderem Interesse ist der nichtintegrierbare Grenzfall $b = 1$. Da sich die Steilheit von Kurven am einfachsten im linearen Fall beurteilen lässt, vergleichen wir aber nicht das Periodogramm

$$\omega_k \rightarrow I(\omega_k)$$

mit der Funktion

$$\omega \rightarrow \omega^{-1},$$

sondern

$$\log(\omega_k) \rightarrow \log(I(\omega_k))$$

mit

$$\log(\omega) \rightarrow \log(\omega^{-1}) = -\log(\omega)$$

und beachten, dass die letztere Funktion Steigung -1 hat.

Wir plotten nun das log Periodogramm der Temperaturzeitreihe gegen die log Frequenz und schauen, ob die negative Steigung in der Nähe der Frequenz 0 mindestens so klein ist wie die von Vergleichslinien, die alle die Steigung -1 haben. Die Abbildung zeigt, dass dies tatsächlich der Fall ist. Wir interpretieren dieses Ergebnis als Evidenz dafür, dass die Temperatur nicht stationär ist, und somit auch als Evidenz für Global Warming.

Differenziert man eine Zeitreihe durch Anwendung des Differenzenoperators $\Delta = 1 - L$, dann ändert sich die Spektraldichte um den multiplikativen Faktor

$$|1 - e^{-i\omega}|^2 \sim \omega^2.$$

Plottet man also das log Periodogramm der differenzierten Zeitreihe gegen die log Frequenz, dann erkennt man den Effekt des Differenzierens an einer um 2 vergrößerten Steigung in der Nähe der Frequenz 0. Die Vergleichslinien zur Beurteilung der Stationarität der ursprünglichen Zeitreihe müssen also Steigung 1 haben.

