Re: Deseasonalization and detrending of Keeling curve

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2007−01/msg00339.html

- *From*: dave@xxxxxxxxxxx
- *Date*: 21 Jan 2007 19:40:47 −0800

David Winsemius wrote:

> "Pekka Jarvela" <pekkajarvela@xxxxxxxxx> wrote in
> news:1169326747.707212.147920@xxxxxxxxxxxxxxxxxxxxxxxxxxxxx:
>
> > CO2 concentrations in atmosphere measured at Mauna Loa in Hawaii are known as Keeling curve, http://en.wikipedia.org/wiki/Keeling_curve You can get the update date from http://scrippsco2.ucsd.edu/data/in_situ_co2/mlo_in_situ_record.txt In this page it is said that
> >
> > "The "detrended" data is seasonally adjusted by removing a 4−harmonic fit with a linear gain factor. The "fit" is based on a stiff spline plus 4−harmonic functions with linear gain."
> >
> > 1. Is detrending fitting a line y = ax + b to data and then subtracting this line from data?
> > 2. What does "removing a 4−harmonic fit with a linear gain factor" mean? Has this something to do with Fourier analysis?
>
> http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1190&context=sio
>
> "The number of harmonics refers to a portion of the fitting function which involves sinusoidal terms with a fundamental period of one year plus higher order Fourier components. Thus, 2 harmonics indicates that terms with periods of 1 year and 6 months were fit, 4 harmonics indicates additional terms with periods of 4 and 3 months."
>
> See also
> http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1110&context=sio

This data set can be adequately modeled as a ARIMA Model of the

follwing form.

Rather than assume a particular deterministiv form, the data
autocorrelative structure can be examined which yields Gaussian
Residuals while pointing to anaomalies that didn't followthe paradigm
....suggesting unusual events or readings..

MODEL STAGE: 888 25EST 1

MODEL STATISTICS AND EQUATION FOR THE CURRENT EQUATION (DETAILS FOLLOW).

Estimation/Diagnostic Checking for Variable Y C02

: NEWLY IDENTIFIED VARIABLE X1 I~P00064 1964/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X2 I~P00160 1972/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X3 I~P00046 1962/ 10
PULSE
: NEWLY IDENTIFIED VARIABLE X4 I~P00509 2001/ 5
PULSE
: NEWLY IDENTIFIED VARIABLE X5 I~P00448 1996/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X6 I~P00266 1981/ 2
PULSE
: NEWLY IDENTIFIED VARIABLE X7 I~P00376 1990/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X8 I~P00506 2001/ 2
PULSE
: NEWLY IDENTIFIED VARIABLE X9 I~P00081 1965/ 9
PULSE
: NEWLY IDENTIFIED VARIABLE X10 I~P00079 1965/ 7
PULSE
: NEWLY IDENTIFIED VARIABLE X11 I~P00577 2007/ 1
PULSE
: NEWLY IDENTIFIED VARIABLE X12 I~P00350 1988/ 2
PULSE

Number of Residuals (R) =n 564

Number of Degrees of Freedom =n−m 549

Residual Mean =Sum R / n −.121718E−03

Sum of Squares =Sum R**2 66.4928

Variance var=SOS/(n) .117895

Adjusted Variance =SOS/(n−m) .121116

Standard Deviation =SQRT(Adj Var) .348018

Standard Error of the Mean =Standard Dev/ .148530E−01

Mean / its Standard Error =Mean/SEM −.819486E−02

Mean Absolute Deviation =Sum(ABS(R))/n .280075

AIC Value (Uses var) =nln +2m −1175.81

SBC Value (Uses var) =nln +m*lnn −1110.78

BIC Value (Uses var) =see Wei p153 −103.425

R Square = .999691

Durbin−Watson Statistic =[A−A(T−1)]**2/A**2 1.99054

D−W STATISTIC SUGGESTS NO SIGNIFICANT AUTOCORRELATION for lag1.

THE DURBIN−WATSON STATISTIC IS VALID ONLY FOR MODELS THAT HAVE A WHITE NOISE
ERROR TERM AND NO LAGS OF THE Y SERIES. OTHERWISE IT IS INVALID.
IN THIS CASE THE TEST IS INVALID.

FORECASTING WITH FINAL MODEL

MODEL COMPONENT LAG COEFF STANDARD P
T
(BOP) ERROR VALUE
VALUE

Differencing 12

1CONSTANT .120 .294E−01 .0001

4.07
2Autoregressive−Factor # 1 1 .916 .194E−01 .0000
47.23
3Moving Average−Factor # 2 1 .210 .481E−01 .0000
4.36

INPUT SERIES X1 I~P00064 1964/ 4 PULSE

Differencing 12

4Omega (input) −Factor # 3 0 −1.68 .198 .0000
−8.49

INPUT SERIES X2 I~P00160 1972/ 4 PULSE

Differencing 12

5Omega (input) −Factor # 4 0 .901 .197 .0000
4.57

INPUT SERIES X3 I~P00046 1962/ 10 PULSE

Differencing 12

6Omega (input) −Factor # 5 0 −.547 .198 .0058
−2.77

INPUT SERIES X4 I~P00509 2001/ 5 PULSE

Differencing 12

7Omega (input) −Factor # 6 0 .627 .197 .0015
3.18

INPUT SERIES X5 I~P00448 1996/ 4 PULSE

Differencing 12

8Omega (input) −Factor # 7 0 −.715 .198 .0003
−3.62

INPUT SERIES X6 I~P00266 1981/ 2 PULSE

Differencing 12

9Omega (input) −Factor # 8 0 .507 .197 .0104
2.57

INPUT SERIES X7 I~P00376 1990/ 4 PULSE

Differencing 12

10Omega (input) −Factor # 9 0 −.646 .197 .0011
−3.28

INPUT SERIES X8 I~P00506 2001/ 2 PULSE

Differencing 12

11Omega (input) −Factor # 10 0 .426 .198 .0317
2.15

INPUT SERIES X9 I~P00081 1965/ 9 PULSE

Differencing 12

12Omega (input) −Factor # 11 0 .554 .198 .0052
2.80

INPUT SERIES X 10 I~P00079 1965/ 7 PULSE

Differencing 12

13Omega (input) −Factor # 12 0 .554 .198 .0053
2.80

INPUT SERIES X 11 I~P00577 2007/ 1 PULSE

14Omega (input) −Factor # 13 0 −1.08 .344 .0018
−3.14

INPUT SERIES X 12 I~P00350 1988/ 2 PULSE

15Omega (input) −Factor # 14 0 .713 .278 .0107
2.56

MODEL STATISTICS AND EQUATION FOR THE CURRENT EQUATION (DETAILS FOLLOW).

Estimation/Diagnostic Checking for Variable Y C02

: NEWLY IDENTIFIED VARIABLE X1 I~P00064 1964/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X2 I~P00160 1972/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X3 I~P00046 1962/ 10
PULSE
: NEWLY IDENTIFIED VARIABLE X4 I~P00509 2001/ 5
PULSE
: NEWLY IDENTIFIED VARIABLE X5 I~P00448 1996/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X6 I~P00266 1981/ 2
PULSE
: NEWLY IDENTIFIED VARIABLE X7 I~P00376 1990/ 4
PULSE
: NEWLY IDENTIFIED VARIABLE X8 I~P00506 2001/ 2
PULSE
: NEWLY IDENTIFIED VARIABLE X9 I~P00081 1965/ 9
PULSE
: NEWLY IDENTIFIED VARIABLE X10 I~P00079 1965/ 7
PULSE

: NEWLY IDENTIFIED VARIABLE X11 I~P00577 2007/ 1
PULSE
: NEWLY IDENTIFIED VARIABLE X12 I~P00350 1988/ 2
PULSE

Number of Residuals (R) =n 564

Number of Degrees of Freedom =n−m 549

Residual Mean =Sum R / n −.121718E−03

Sum of Squares =Sum R**2 66.4928

Variance var=SOS/(n) .117895

Adjusted Variance =SOS/(n−m) .121116

Standard Deviation =SQRT(Adj Var) .348018

Standard Error of the Mean =Standard Dev/ .148530E−01

Mean / its Standard Error =Mean/SEM −.819486E−02

Mean Absolute Deviation =Sum(ABS(R))/n .280075

AIC Value (Uses var) =nln +2m −1175.81

SBC Value (Uses var) =nln +m*lnn −1110.78

BIC Value (Uses var) =see Wei p153 −103.425

R Square = .999691

Durbin−Watson Statistic =[A−A(T−1)]**2/A**2 1.99054

D−W STATISTIC SUGGESTS NO SIGNIFICANT AUTOCORRELATION for lag1.

Hope this helps ..

Dave Reilly
Automatic Forecasting Systems
http://www.autobox.com

.