Comparison of global wheat prices with wheat flour Producer Price Index

The **Wheat Flour Producer Price Index** depends strongly on the **global wheat prices**. The higher the prices of wheat, the higher the wheat flour producer price index. We empirically investigate their relationship with each other.

The two time series downloaded from the Federal Reserve Bank of St. Louis website https://fred.stlouisfed.org/ are the *Global Price of Wheat (PWHEAMTUSDM)* and the *Producer Price Index by Commodity: Processed Foods and Feeds: Wheat Flour (WPU02120301)*. In this project, We work with the seasonally adjusted quarterly time series. Number of observations is $n = 123$, where the observation period covers 31 years, from 1st quarter 1990 to 4th quarter 2020. We denote the variable **Price of Wheat** by X and the variable **Wheat Flour Producer Price Index** by Y.

In the first figure, we can clearly see that the two variables (both original (a) and logarithmized (b)) have a very similar pattern, indicating that they have a relationship.

```
wheat = read.csv("PWHEAMTUSDM.csv",na.strings="null")
wheat = wheat[2:(nrow(wheat)),]
X <- as.numeric(wheat[,2]); T <- as.Date(wheat[,1])
flour = read.csv("WPU02120301.csv",na.strings="null")
flour = flour[2:(nrow(flour)), ]
Y <- flour[,2] # gleicher Zeitraum!
n <- length(T)

par(mfrow=c(2,1),mar=c(2,2,0.5,2),pch=20)
plot(T,X,col="blue", type = 'l'); lines(T,Y, col = 'red')
mtext(" (a)",side=4,las=1)
y <- log(Y); x <- log(X)
plot(T,x,col="blue", type = 'l'); lines(T,y,col="red")
mtext(" (b)",side=4,las=1)
```

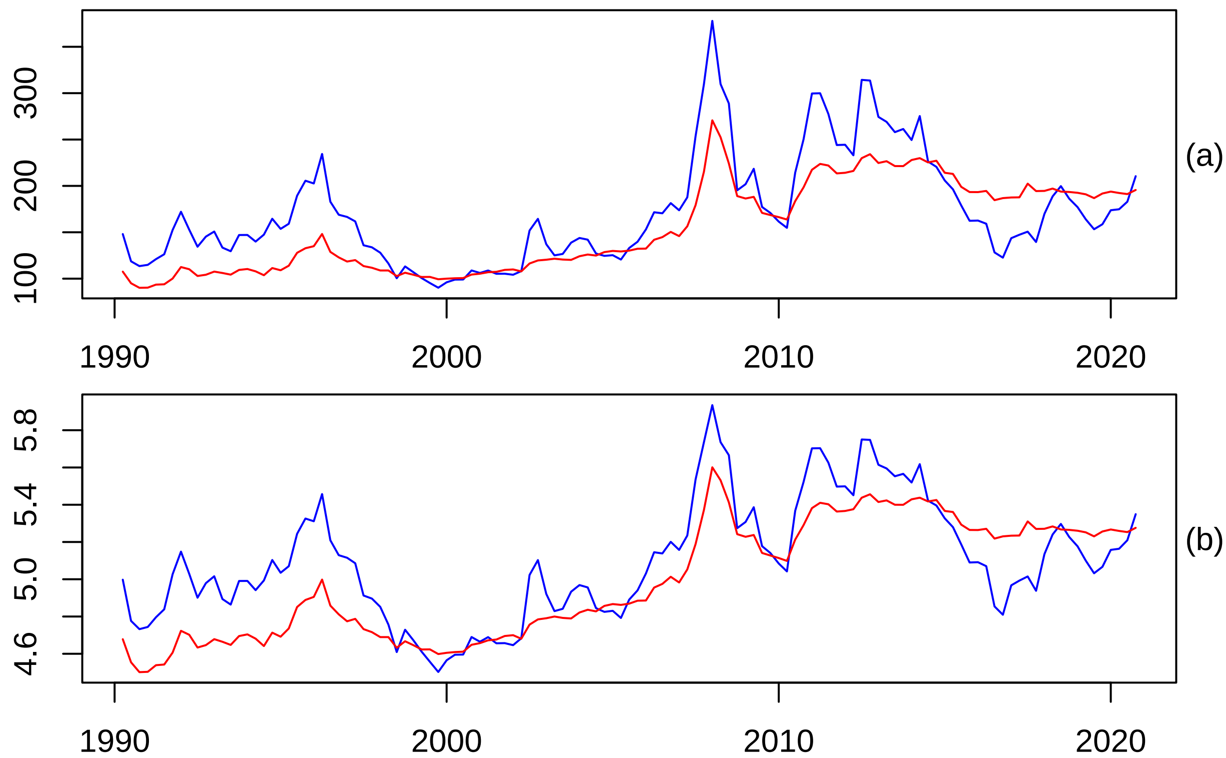

Figure 1: (a) X - wheat (blue) and Y - wheat flour (red); (b) log(X) (blue) and log(Y) (red)

The second figure shows us that the two logarithmized time series have a linear trend. However, the growth can be described very badly only by the linear trend. This can be seen well with the deviations from the linear trend (c - d). They do not look stationary, which is also confirmed by the autocorrelations of the trend residuals (e - f), because the peaks fall very slowly and thus show the autocorrelation.

```r
par(mfcol=c(3,2),mar=c(2,2,0.5,2),pch=20)
plot(T,x,col="blue"); h <- 1:n; H <- lm(x~h)
lines(T,H$fitted.values,lwd=2); mtext(" (a)",side=4,las=1)
xres <- H$residuals; plot(T,xres); mtext(" (c)",side=4,las=1)
acf(xres); mtext(" (e)",side=4,las=1)
plot(T,y,col="red"); H <- lm(y~h)
lines(T,H$fitted.values,lwd=2); mtext(" (b)",side=4,las=1)
yres <- H$residuals; plot(T,yres); mtext(" (d)",side=4,las=1)
acf(yres); mtext(" (f)",side=4,las=1)
```

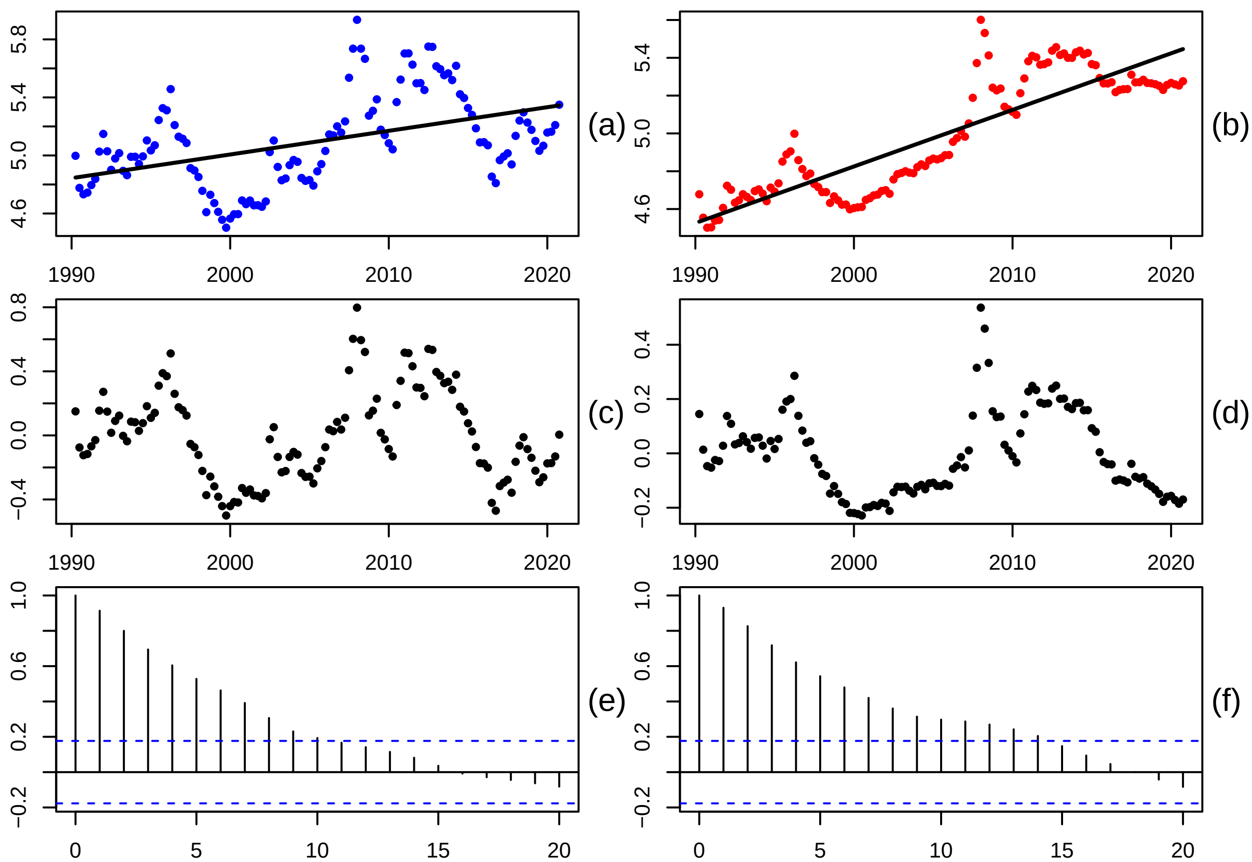

Figure 2: (a),(b) X and Y each with linear trend; (c),(d) deviations from linear trend; (e),(f) autocorrelations of trend residuals

In the third figure it can be seen just as well that the time series are not stationary. This can be well shown by the different periodograms, in that they rise very steeply at a low frequency. The log periodograms of the trend residuals of $log(X)$ and $log(Y)$ also have a much steeper slope than the comparison lines with a slope of -1 (e - f). Therefore, we can assume that the time series are not stationary.

```
par(mfcol=c(3,2),mar=c(2,2,0.5,2),pch=20); h <- 2*pi
H <- spec.pgram(x,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (a)",4,las=1)
H <- spec.pgram(xres,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (c)",4,las=1)
plot(log(h*H$freq[1:20]),log(H$spec[1:20]/h),type="o",lwd=2)
mtext(" (e)",4,las=1)
for (c in (-15):(-4)) abline(a=c,b=-1,col="gray")
H <- spec.pgram(y,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (b)",4,las=1)
H <- spec.pgram(yres,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (d)",4,las=1)
plot(log(h*H$freq[1:20]),log(H$spec[1:20]/h),type="o",lwd=2)
mtext(" (f)",4,las=1)
for (c in (-15):(-4)) abline(a=c,b=-1,col="gray")
```

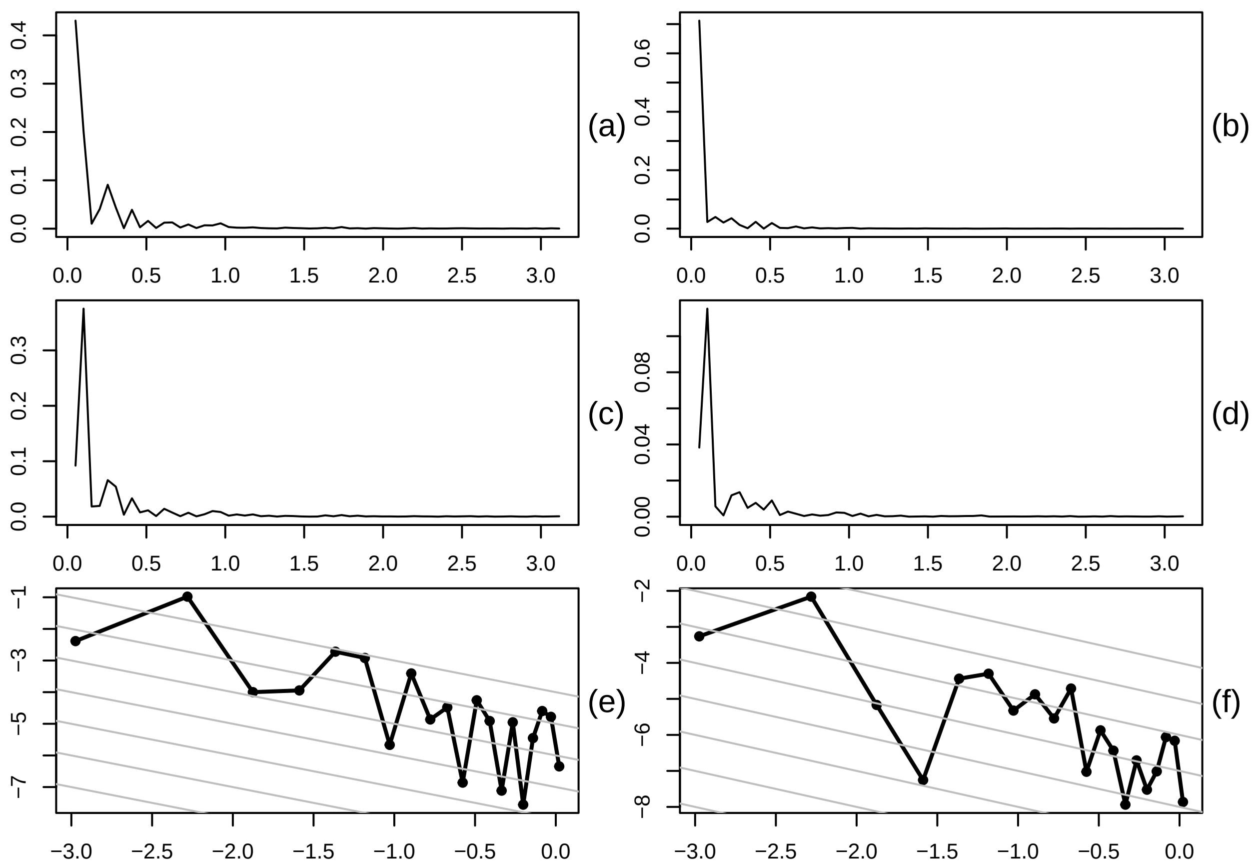

Figure 3: (a),(b) periodograms of log(X), log(Y); (c),(d) periodograms of trend residuals; (e),(f) log periodograms vs. log Fourier frequencies 1-20

To achieve stationarity we find a "one-time" differences. Figures a-b show us that the differenced time series are very close to zero now. Since we have achieved stationarity, we now see no higher peaks at frequency 0 in the periodograms.

Figures c-f show us the estimators for the underlying spectral densities. We see here that compared to the high frequency vibrations, the low and medium frequency vibrations contribute much more to the variance of the time series. This can be seen much better in the case of the Wheat Flour Producer Price Index.

```r
dx <- x[2:n]-x[1:(n-1)]; dy <- y[2:n]-y[1:(n-1)]
par(mfcol=c(3,2),mar=c(2,2,0.5,2),pch=20); h <- 2*pi
plot(T[2:n],dx,type="l"); mtext(" (a)",4,las=1)
H <- spec.pgram(dx,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (c)",4,las=1)
HH <- spec.pgram(dx,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE,spans=5)
lines(h*HH$freq,HH$spec/h,col="blue",lwd=2)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (e)",4,las=1)
HH <- spec.pgram(dx,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE,spans=10)
lines(h*HH$freq,HH$spec/h,col="blue",lwd=2)
plot(T[2:n],dy,type="l"); mtext(" (b)",4,las=1)
H <- spec.pgram(dy,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (d)",4,las=1)
HH <- spec.pgram(dy,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE,spans=10)
lines(h*HH$freq,HH$spec/h,col="red",lwd=2)
plot(h*H$freq,H$spec/h,type="l"); mtext(" (f)",4,las=1)
HH <- spec.pgram(dy,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE,spans=25)
lines(h*HH$freq,HH$spec/h,col="red",lwd=2)
```

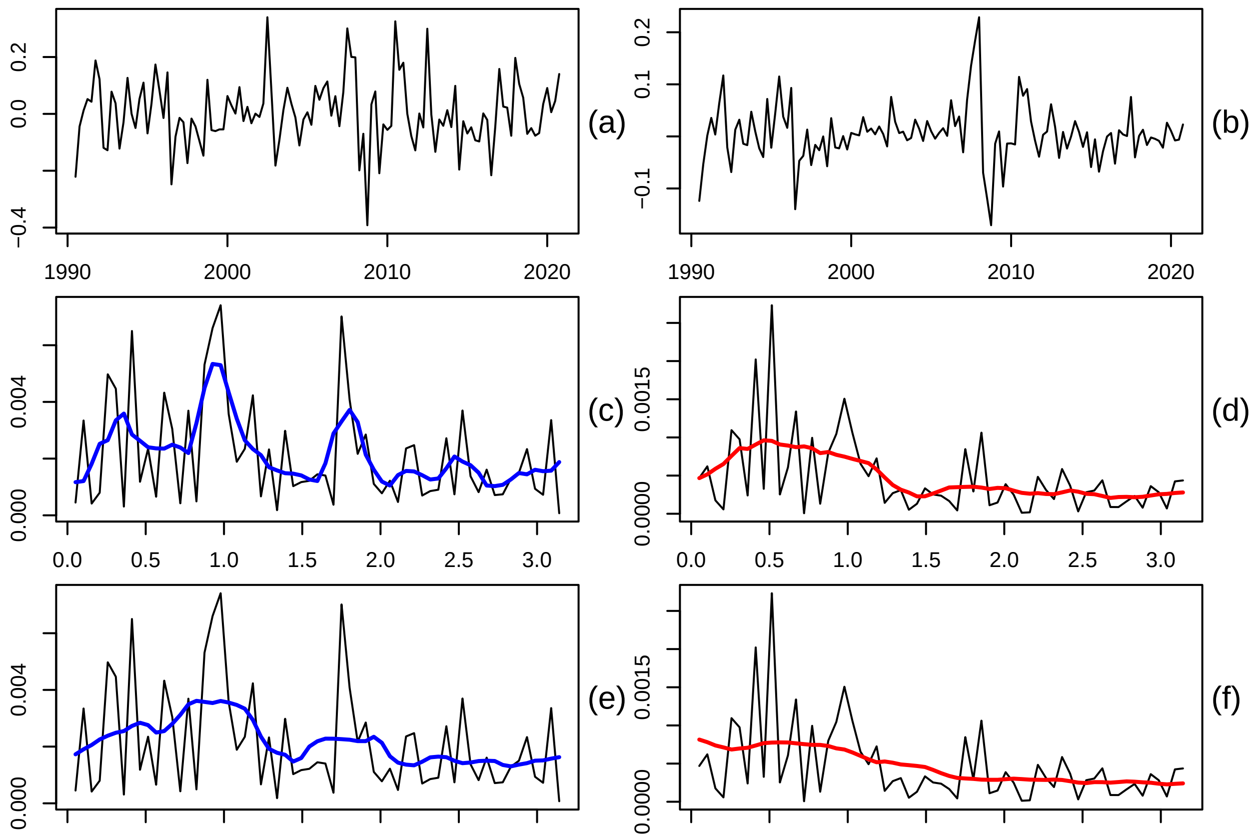

Figure 4: (a),(b) Differences of log(X), log(Y); (c),(d) Smoothed periodograms of the differences; (e), (f) Stronger smoothing

Figure 5 shows us the cross-correlations that can help us examine the relationships between the two differenced time series. In the top left and bottom right, respectively, we see the autocorrelations of $\mathbf{X}$ ($\hat{\rho}_{xx}(j) = \hat{\rho}_{xx}(-j)$) and of $\mathbf{Y}$ ($\hat{\rho}_{yy}(j) = \hat{\rho}_{yy}(-j)$). In the top right and bottom left, the cross-correlations ($\hat{\rho}_{xy}(j) = \hat{\rho}_{yx}(-j)$) are shown for $j \geq 0$ and for $j \leq 0$, respectively.

We see here that for wheat prices there are no significant lags, that is, prices for different years are uncorrelated. However, for the wheat flour price index, we see that there is a large value at least for lag 1, which means that the price index from the next quartal depends on the price index from the current quartal.

In the case of $\hat{\rho}_{xy}(j)$, we can say that there is a correlation between x_t and y_t (because of the large value at lag 0). At the negative lag of -1, we also see a significant coefficient, which may indicate that x_t has an effect on y_{t+1}.

```r
par(mai = c(0.4, 0.4, 0.4, 0.4))
D <- cbind(dx,dy); D <- ts(D); acf(D, ylab = '', xlab = '', ylim = c(-0.2, 1), main = '')
```

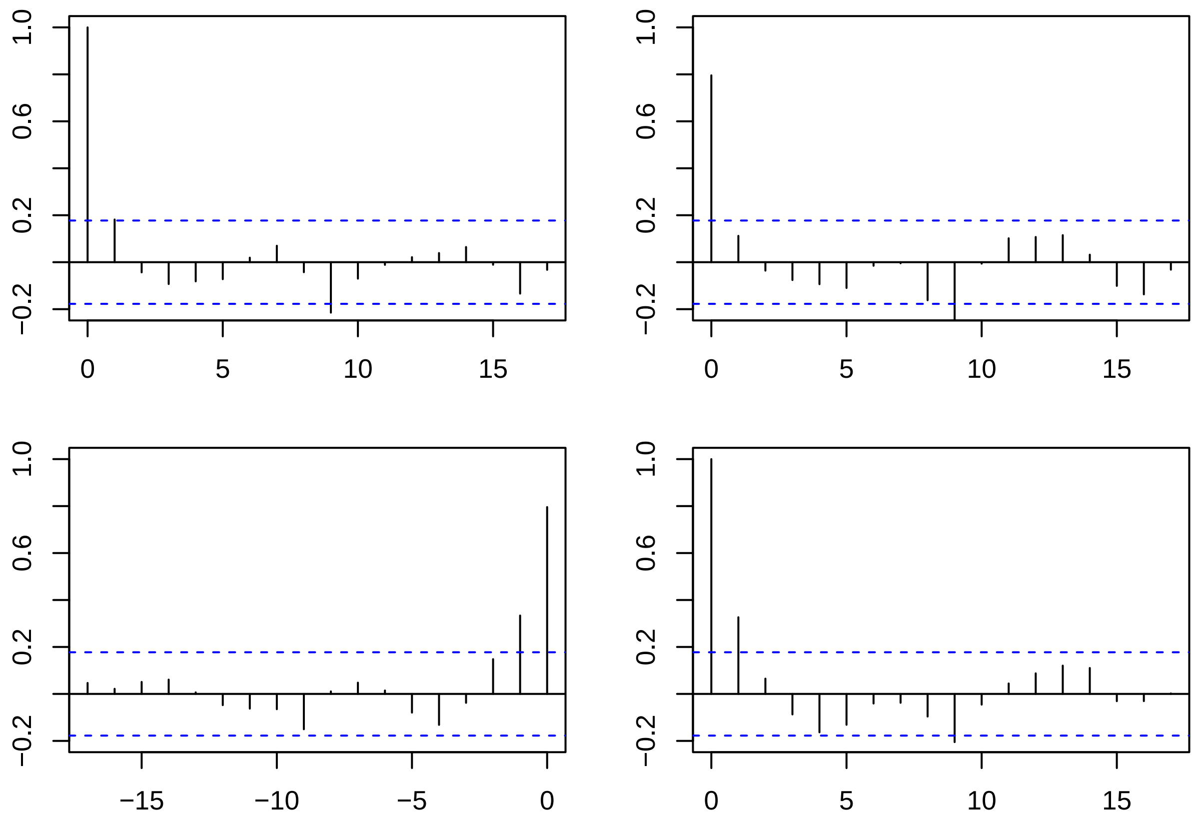

Figure 5: Empirical cross-correlations

In order to get a more precise statement about the cross-correlations, we try to remove the autocorrelations from the two time series by performing the pre-whitening. We do this using the 2nd order ARMA model.

Figure 6 shows us that pre-whitening has an effect in both time series.

```
par(mfcol=c(1,2),mar=c(2,2,0.5,2),pch=20); h <- 2*pi
H <- spec.pgram(dx,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,cumsum(H$spec)/sum(H$spec),type="l"); mtext(" (a)",4,las=1)
H <- arima(dx,order=c(2,0,0),include.mean=TRUE,transform.pars=TRUE)
dx.res <- H$residuals
H <- spec.pgram(dx.res,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
lines(h*H$freq,cumsum(H$spec)/sum(H$spec),col="red")
H <- spec.pgram(dy,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
plot(h*H$freq,cumsum(H$spec)/sum(H$spec),type="l"); mtext(" (b)",4,las=1)
H <- arima(dy,order=c(2,0,0),include.mean=TRUE,transform.pars=TRUE)
dy.res <- H$residuals
H <- spec.pgram(dy.res,taper=0,detrend=FALSE,fast=FALSE,plot=FALSE)
lines(h*H$freq,cumsum(H$spec)/sum(H$spec),col="red")
```

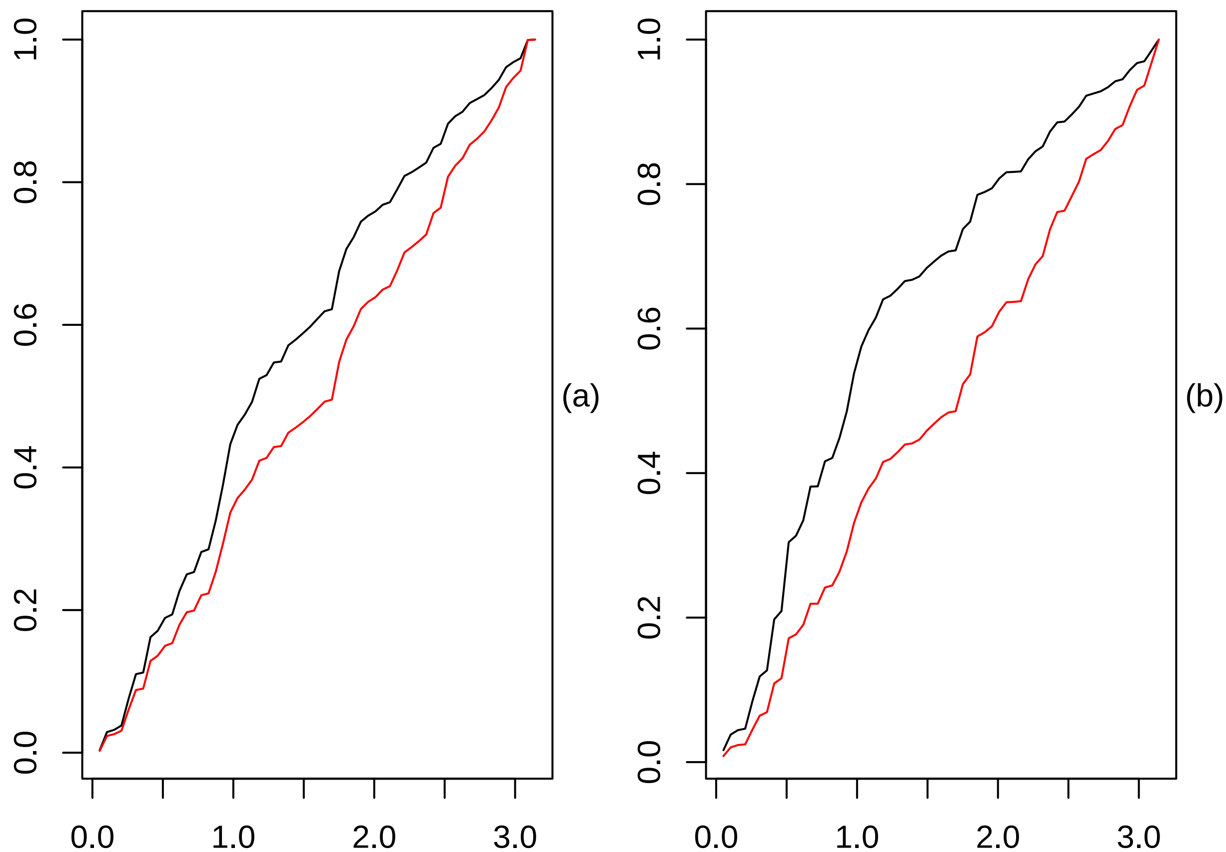

Figure 6: Effect of prewhitening using an AR model of order 2 (red) on the normalized cumulative periodogram (black) of the differences in log wheat price (a) and log wheat flour price index (b).

Figure 5 changes accordingly if we use the residual series (see Figure 7). Here we see that there is no longer any autocorrelation in the two time series. And if we look at the cross correlations, we see that there is only one correlation between x_t and y_t. x_t no longer has an influence on y_{t+1}.

```
par(mai = c(0.4, 0.4, 0.4, 0.4))
D <- cbind(dx.res,dy.res); D <- ts(D); acf(D, ylab = '', xlab = '', ylim = c(-0.2, 1), main = '')
```

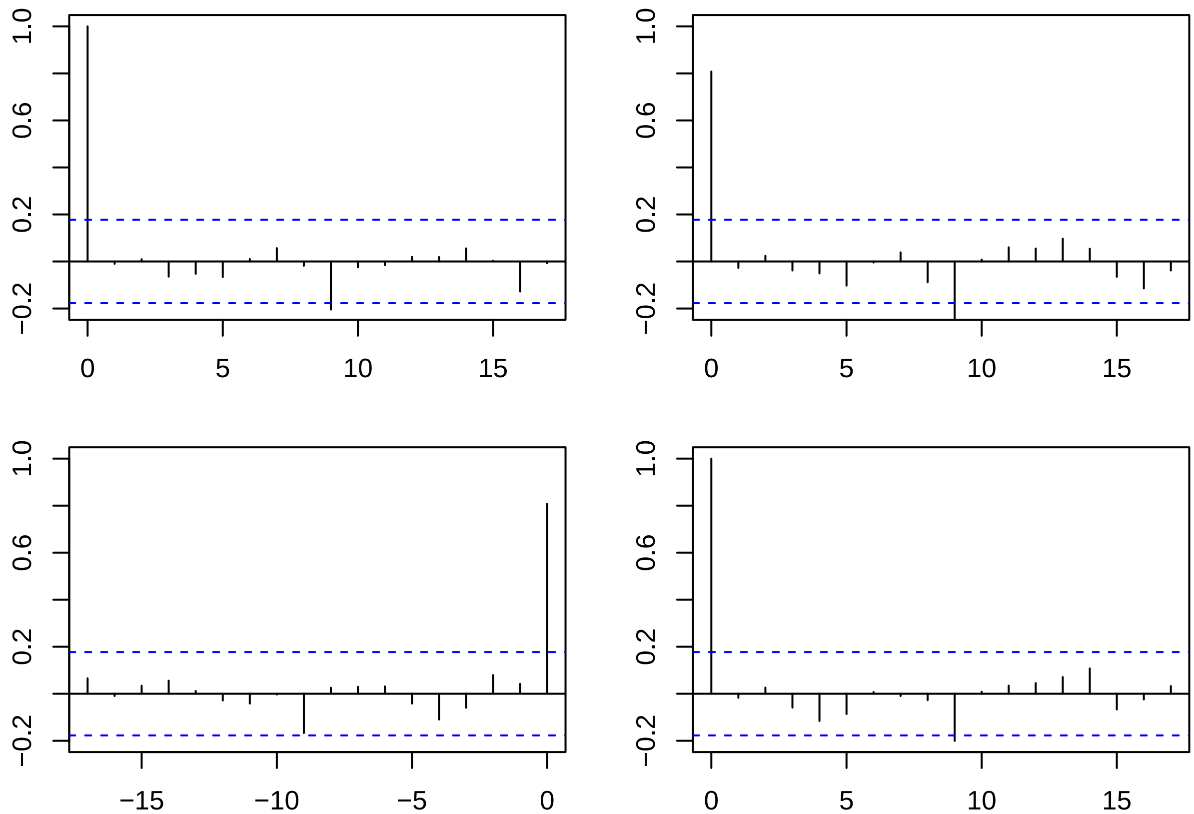

Figure 7: Empirical cross-correlations in the residual series.

Figure 8a) shows us the empirical squared coherence for two different smoothings. We see that it is large at all frequencies (see y values of 0.5 - 0.8), indicating that there is a linear relationship between the two time series at all frequencies.

We now interpret the phase spectrum and consider all frequencies because there is a linear relationship everywhere. However, we unfortunately find no evidence of such relationship in the Figure 8b) because the empirical phase spectrum is practically constant. It could mean that there is no relationship between the previous or future values of x on y and reversed notch.

```
par(mfcol=c(2,1),mar=c(2,2,0.5,2),pch=20); h <- 2*pi
H1 <- spec.pgram(D,taper=0,detrend=F,fast=F,plot=F,spans=20)
plot(H1$freq*h,H1$coh,type="l",col="orange3")
H2 <- spec.pgram(D,taper=0,detrend=F,fast=F,plot=F,spans=30)
lines(H2$freq*h,H2$coh,col="darkgreen"); mtext(" (a)",4,las=1)
plot(H1$freq*h,H1$phase,type="l",col="orange3")
lines(H2$freq*h,H2$phase,col="darkgreen"); mtext(" (b)",4,las=1)
```

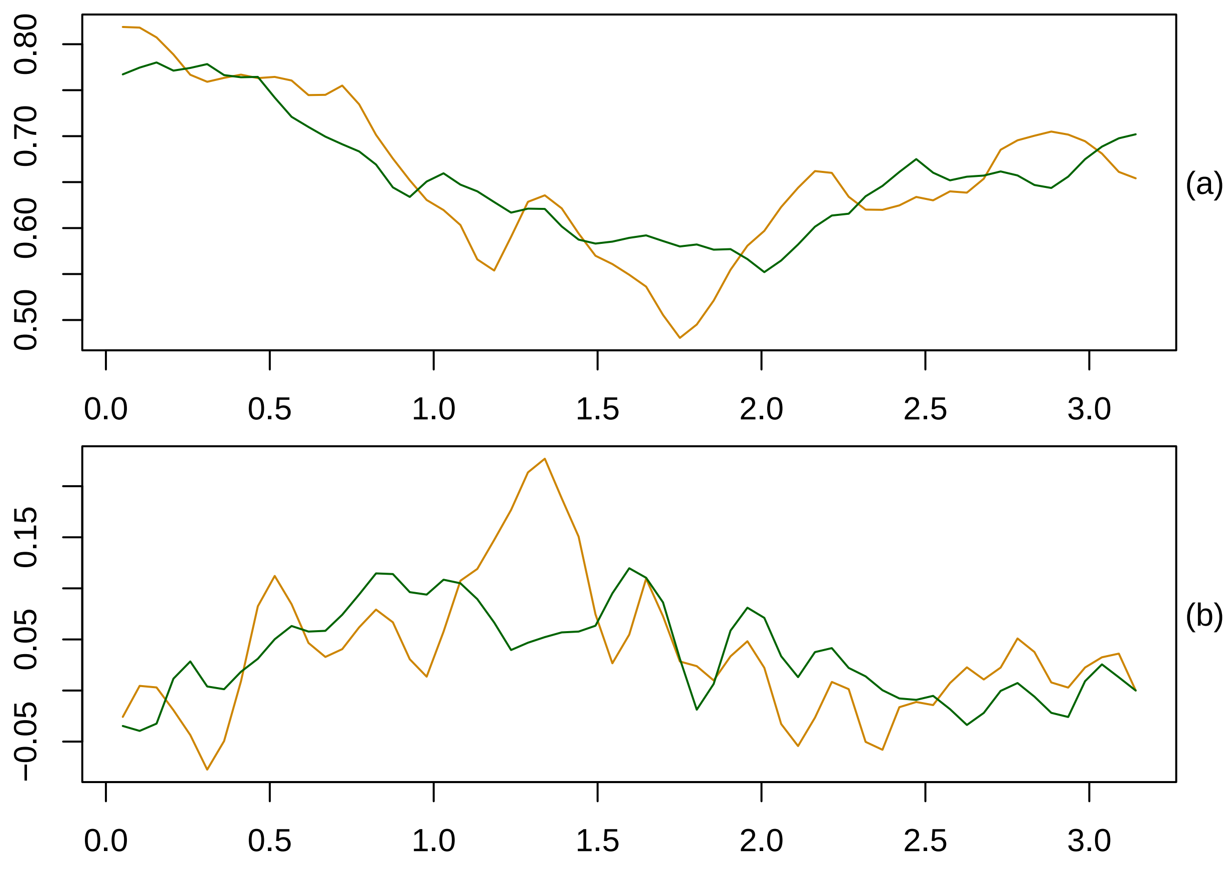

Figure 8: a) Squared coherence; b) Phase spectrum in each case for two different smoothings (green: stronger, orange: weaker).

```
# AR-Spektrum eines vektorautoregressiven Prozesses
var.spec <- function(fr,AR.p) {
# fr ... Frequenzen, AR.p ... AR(p) Modell geschÃ¤tzt mit R-Funktion ar
  nf <- length(fr); p <- AR.p$order
  sigma <- AR.p$var.pred; k <- length(sigma[1,])
```

```
  Id <- diag(1,nrow=k,ncol=k) # identity matrix
  sp <- array(dim=c(nf,k,k))
  for (w in 1:nf) {
    A <- Id
    for (l in 1:p) A <- A-AR.p$ar[l,,]*exp(-1i*fr[w]*l)
      A <- solve(A) # inverse of A
      sp[w,,] <- A%*%sigma%*%t(Conj(A)) }
  return(sp/(2*pi)) }
fr <- H1$freq*(2*pi); AR.3 <- ar(D,order.max=3,aic=FALSE,demean=TRUE)
AR.1 <- ar(D,order.max=3,aic=FALSE,demean=TRUE)
AR.2 <- ar(D,order.max=6,aic=FALSE,demean=TRUE)
sp.1 <- var.spec(fr,AR.1); sp.2 <- var.spec(fr,AR.2)
```

When estimating the squared coherence and the phase spectrum with parametric methods (AR models) (see figure 9.c-d), similar results are obtained as with non-parametric methods (see figure 8.a-b).

Figures 9a) and 9b) show us additionally the estimated cross spectrum and estimated quadrature spectrum, respectively. We see that the estimated cross spectrum is positive in the range we are interested in (whole range). But the estimated quadrature spectrum takes relatively small values there. Therefore we cannot say anything about the sign.

```
par(mfcol=c(2,2),mar=c(2,2,0.5,2),pch=20);
plot(fr,Re(sp.1[,1,2]),type="l",col="darkgreen"); abline(h=0,lty=2)
lines(fr,Re(sp.2[,1,2]),col="orange3"); mtext(" (a)",4,las=1)
plot(fr,-Im(sp.1[,1,2]),type="l",col="darkgreen"); abline(h=0,lty=2)
lines(fr,-Im(sp.2[,1,2]),col="orange3"); mtext(" (b)",4,las=1)
plot(fr,Mod(sp.1[,1,2])^2/(sp.1[,1,1]*sp.1[,2,2]),type="l",col="darkgreen")
lines(fr,Mod(sp.2[,1,2])^2/(sp.2[,1,1]*sp.2[,2,2]),col="orange3")
mtext(" (c)",4,las=1)
plot(fr,Arg(sp.1[,1,2]),type="l",col="darkgreen"); mtext(" (d)",4,las=1)
lines(fr,Arg(sp.2[,1,2]),col="orange3")
```

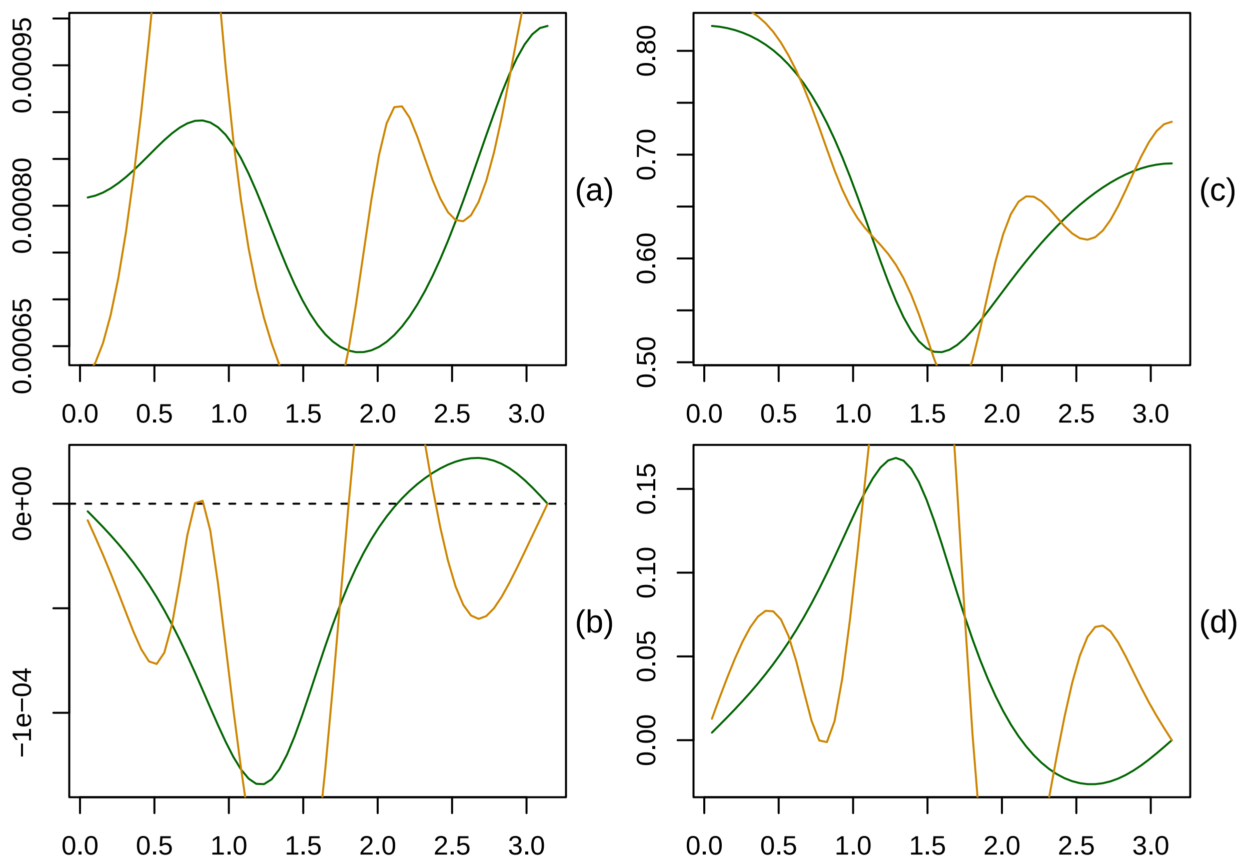

Figure 9: Parametric cross-spectral analysis with bivariate AR models of order p = 3 (green) and p = 6 (orange).

In the previous graphs, we have seen that the periodograms for the two time series in the region of frequency 0 have a steep slope, which could be an indication of non-stationarity. Now we run a unit-root test for our logarithmized data to test for stationarity. From the results for the two time series, we see clearly that no matter how many lags we take, we cannot reject the null hyphotesis that says there is a stochastic trend present in the time series. That is why we can do the cointegration analysis.

```
h <- read.csv("PWHEAMTUSDM.csv",na.strings="null"); h = h[2:(nrow(h)),]; X <- as.numeric(h[,2]); T <- as
h <- read.csv("WPU02120301.csv",na.strings="null"); h = h[2:(nrow(h)), ]; Y <- h[,2]
n <- length(T); n1 <- n-1; y <- log(Y); x <- log(X)
O1 <- 1:n; O2 <- O1^2; O3 <- O1^3
dx <- x[2:n]-x[1:n1]; dy <- y[2:n]-y[1:n1]; T1 <- T[-1]
```

```
library(tseries)
```

1) ADF test for wheat prices.

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
adf.test(x, alternative = 'stationary', k = 0)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -2.3336, Lag order = 0, p-value = 0.4379
## alternative hypothesis: stationary
```

```
adf.test(x, alternative = 'stationary', k = trunc((length(x)-1)^(1/3))) # 4, r recommend
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -2.4092, Lag order = 4, p-value = 0.4066
## alternative hypothesis: stationary
```

```
adf.test(x, alternative = 'stationary', k = 8)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -2.5105, Lag order = 8, p-value = 0.3645
## alternative hypothesis: stationary
```

```
library(tseries)
adf.test(y, alternative = 'stationary', k = 0)
```

2) ADF test for wheat flour price indices.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -1.8982, Lag order = 0, p-value = 0.6189
## alternative hypothesis: stationary
```

```
adf.test(y, alternative = 'stationary', k = trunc((length(y)-1)^(1/3))) # 4, r recommend
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -2.0555, Lag order = 4, p-value = 0.5535
## alternative hypothesis: stationary
```

```
adf.test(y, alternative = 'stationary', k = 8)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -1.7807, Lag order = 8, p-value = 0.6677
## alternative hypothesis: stationary
```

Now we perform a cointegration test. With the help of this we can test whether the two time series are cointegrated. The null hyphotesis ($r = 0$) says here that there is no cointegrated relationship between the two. This hyphotesis can be discarded here. Therefore, we perform the hyphotesis of $r <= 0$. This hyphotesis cannot be rejected here. So we can reason that there is a cointegrated relation here.

```
library(urca)
h <- ca.jo(ts(cbind(x,y)),K=10,ecdet="const")
 # with constant and lag order=10
summary(h)
```

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegra
##
## Eigenvalues (lambda):
## [1] 1.058009e-01 3.795762e-02 1.110223e-16
```

```
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 1 |  4.37  7.52  9.24 12.97
## r = 0  | 12.64 13.75 15.67 20.20
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##                x.l10      y.l10  constant
## x.l10     1.00000000  1.0000000  1.000000
## y.l10    -0.95819484 -0.5945367 -3.177024
## constant -0.05080313 -2.1928113 10.481612
##
## Weights W:
## (This is the loading matrix)
##
##          x.l10       y.l10      constant
## x.d 0.02967492 -0.12276533 -1.247544e-15
## y.d 0.04758851 -0.03585508  4.137989e-16
```