Financial Times Series

Lecture 7

Regime Switching

- We sometimes talk about bull and bear markets
- By bull market we mean a market where prices are increasing and where volatility is relatively low
- By bear market we mean a market where prices are decreasing and where volatility is relatively high

Regime Switching

- So if we would like to model log-returns under the assumption that the market switches between the bull and bear states we let

$$r_t = \begin{cases} \mu_{bull} + \sigma_{bull}\varepsilon_t, \ s_t = bull \\ \mu_{bear} + \sigma_{bear}\varepsilon_t, s_t = bear \end{cases}$$

- Here s_t denotes the state of the market at time t and ε_t is i.i.d $N(0,1)$

Regime Switching

- Typically we cannot tell what state the market is in but using some filtering techniques and assumptions on the switching mechanism we can estimate switching probabilities, drifts and volatilities

- We may assume that the switching between the states is Markovian, i.e. the probability of switching from one state to another depends only on which state the market is in a the time point of switching and not on in which states the market has been in or returns at previous time points

$$P(s_t|s_{t-1}) = P(s_t|s_{t-1}, s_{t-2}, \dots, r_{t-1}, r_{t-2}, \dots)$$

Regime Switching

- If we denote the market states 0 and 1 we let

$$p_{00} = P(s_t = 0|s_{t-1} = 0)$$
$$p_{01} = P(s_t = 1|s_{t-1} = 0)$$
$$p_{10} = P(s_t = 0|s_{t-1} = 1)$$
$$p_{11} = P(s_t = 1|s_{t-1} = 1)$$

- Clearly $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$

Estimation

- So given a series of observations $\{r_1, \dots, r_T\}$ we want to estimate the parameters $\theta = \{p_{00}, p_{11}, \mu_0, \mu_1, \sigma_0, \sigma_1\}$

- Since we cannot tell which state we are in at a given time point it is not obvious how the estimations can be done but we may formally write the likelihood function

$$L(\theta) = f(r_1|\theta)f(r_2|\theta, r_1)\cdots f(r_T|\theta, r_1, \dots, r_{T-1})$$

where f is the density of a $N(\mu_s, {\sigma_s}^2)$ random variable

Estimation

- So the contribution of r_t to the log-likelihood is

$$\log f(r_t|\theta, r_1, \dots, r_{t-1})$$

- Using conditional probabilities and the Markov property, we can write (exercise)

$$f(s_t, s_{t-1}, r_t|\theta, r_1, \dots, r_{t-1})$$
$$= f(s_{t-1}|\theta, r_1, \dots, r_{t-1}) f(s_t|s_{t-1}, \theta) f(r_t|s_t, \theta)$$

Estimation

- Above $f(s_t|s_{t-1},\theta)$ is the switching probability and

$$f(r_t|s_t,\theta) = \frac{1}{\sqrt{2\pi}\sigma_{s_t}} exp\left\{-\frac{1}{2}\left(\frac{r_t-\mu_{s_t}}{\sigma_{s_t}}\right)^2\right\}$$

- The function $f(s_{t-1}|\theta, r_1, \dots, r_{t-1})$ is given by

$$\frac{f(s_{t-1},s_{t-2}=0,r_{t-1}|\theta,r_1,\dots,r_{t-2})+f(s_{t-1},s_{t-2}=1,r_{t-1}|\theta,r_1,\dots,r_{t-2})}{f(r_{t-1}|\theta,r_1,\dots,r_{t-2})}$$

Estimation

- So we get

$$f(r_t|\theta, r_1, \dots, r_{t-1}) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(s_t = i, s_{t-1} = j, r_t|\theta, r_1, \dots, r_{t-1})$$

- To start the recursion we may let

$$f(s_1 = 0, r_1|\theta) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma_0} exp\left\{-\frac{1}{2}\left(\frac{r_1 - \mu_0}{\sigma_0}\right)^2\right\}$$
$$f(s_1 = 1, r_1|\theta) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma_1} exp\left\{-\frac{1}{2}\left(\frac{r_1 - \mu_1}{\sigma_1}\right)^2\right\}$$

Estimation

- This gives

$$f(r_1|\theta) = f(s_1 = 0, r_1|\theta) + f(s_1 = 1, r_1|\theta)$$

- and

$$f(s_1 = 0|\theta, r_1) = \frac{f(s_1 = 0, r_1|\theta)}{f(r_1|\theta)}$$

$$f(s_1 = 1|\theta, r_1) = \frac{f(s_1 = 1, r_1|\theta)}{f(r_1|\theta)}$$

Estimation

- Next we get

$$f(r_2|\theta, r_1) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(s_1 = i, s_2 = j, r_2|\theta, r_1)$$

- where

$$f(s_1 = i, s_2 = j, r_2|\theta, r_1)$$
$$= f(s_1 = i|\theta, r_1)p_{ji}\frac{1}{\sqrt{2\pi}\sigma_j}exp\left\{-\frac{1}{2}\left(\frac{r_2 - \mu_j}{\sigma_j}\right)^2\right\}$$

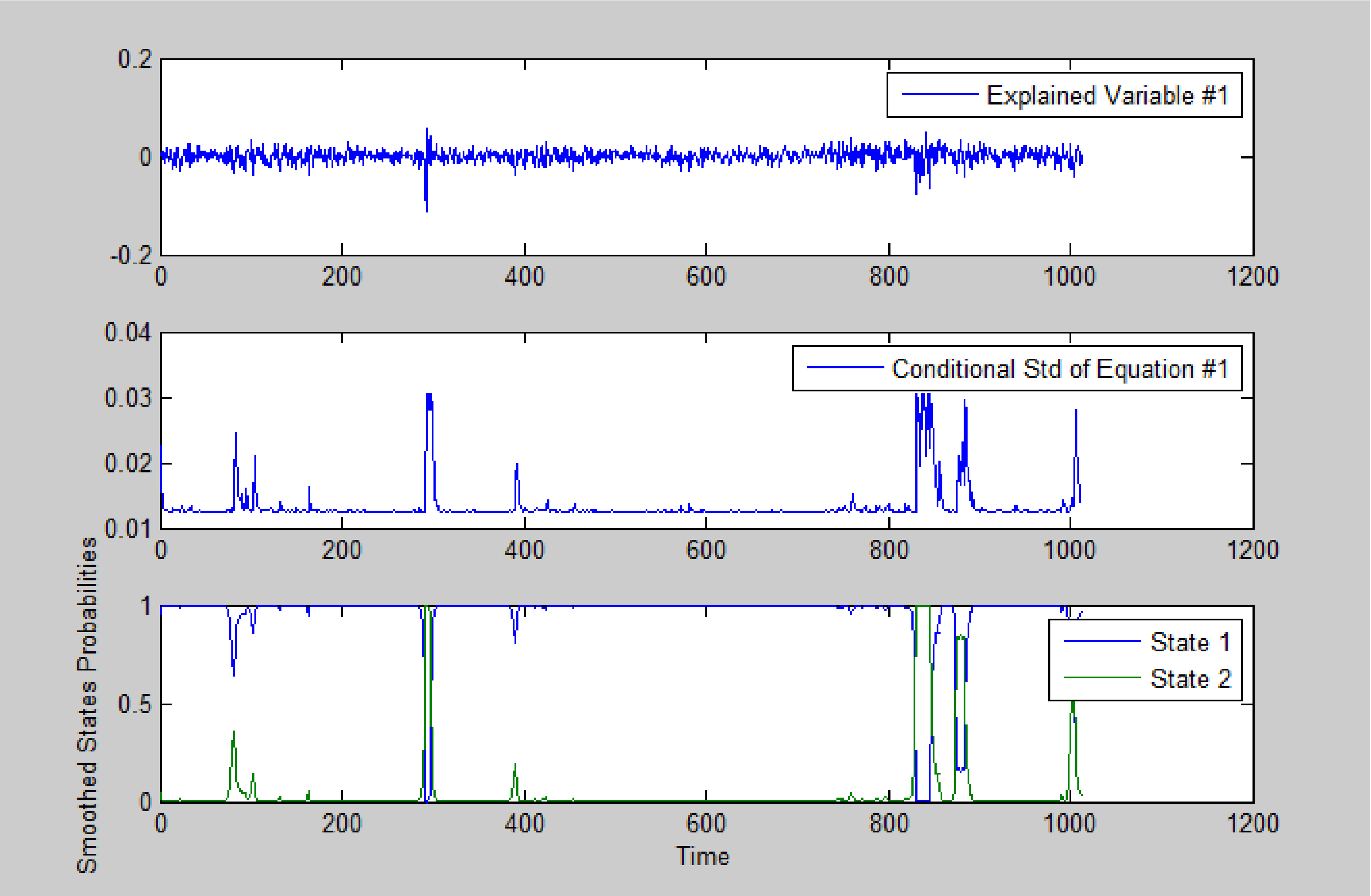
- And so forth…

Estimation

- So, the maximization of the likelihood-function cannot be done manually but using standard routines like fminsearch or fmincon in matlab we can find our parameter estimates

- There is a matlab package by Marcelo Perlin called MS_Regress available online (for free) that does the parameter estimation and then some…

For the ^N225



Explained Variable #1
0.2
0
-0.2
0
200
400
600
800
1000
1200
Conditional Std of Equation #1
0.04
0.03
0.02
0.01
Smoothed States Probabilities
State 1
State 2
1
0.5
0
Time

For the ^N225

- By smoothed probabilities in the above plot we mean

$$f(s_t|\theta, r_1, \dots, r_T)$$

- The parameter estimates are

$$p_{bull,bull} = 0.99, p_{bear,bear} = 0.89,$$
$$\mu_{bull} = 0.0007, \mu_{bear} = -0.0086$$
$$\sigma_{bull} = 0.00015, \sigma_{bear} = 0.0011$$

- As expected the bear volatility is higher than the bull volatility

Non-parametric models

- What if we do not make any distribution assumptions?
- Assume that $\{r_t\}$ and $\{x_t\}$ are two time series for which we want to explore their relationship
- Maybe it is possible to fit a model

$$r_t = m(x_t) + a_t$$

where m is some smooth function to be estimated from the data

Non-parametric models

- If we had independent observerations $(r_1, \ldots, r_T)$ for a fixed $x_t = x$ the we could write

$$\frac{\sum_{t=1}^{T} r_t}{T} = m(x) + \frac{\sum_{t=1}^{T} a_t}{T}$$

- For a sufficiently large T the mean of the noise (LLN) will be close to zero so in this case

$$\widehat{m}(x) = \frac{\sum_{t=1}^{T} r_t}{T}$$

Non-parametric models

- In financial applications we will typically not have data as above. Rather we will have pairs of observations

$$\{(r_1, x_1), \dots, (r_T, x_T)\}$$

- But if the function m is sufficiently smooth then a value of r_t for which $x_t \approx x$ will still give a good approximation of $m(x)$

- For a value of r_t for which x_t is not close to x will give a less accurate approximation of $m(x)$

Non-parametric models

- So instead of a simple average, we use a weighted average

$$\widehat{m}(x) = \frac{1}{T}\sum_{t=1}^{T} w_t(x) r_t$$

where the weights $w_t(x)$ are large for those r_t with x_t close to x and weights are small for r_t with x_t not close to x

Non-parametric models

- Above we assume that $\sum_{t=1}^{T} w_t(x) = T$

- One may also treat $\frac{1}{T}$ as part of the weights and work under the assumption $\sum_{t=1}^{T} w_t(x) = 1$

- A construction like the one at hand where we will let the weights depend on a choice of measure for the distance between x_t and x and the size of the weights depends on the distance may be referred to as a local weighted average

Kernel regression

- One way of finding appropriate weights is to use kernels $K(x)$ which are typically probability density functions

$$K(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} K(x)dx = 1$$

- For flexibility we will allow scaling of the kernel using a ”bandwidth” h

$$K_h(x) = \frac{1}{h}K(x/h), \int_{-\infty}^{\infty} K_h(x)dx = 1$$

- The weights may be defined as

$$w_t(x) = \frac{K_h(x - x_t)}{\sum_{t=1}^{T} K_h(x - x_t)}$$

Nadaraya-Watson

- The N-W kernel estimator (N-W 1964) is given by

$$\widehat{m}(x) = \sum_{t=1}^{T} w_t(x) r_t = \frac{\sum_{t=1}^{T} K_h(x - x_t) r_t}{\sum_{t=1}^{T} K_h(x - x_t)}$$

- The choice of kernel is often (Gaussian)

$$K_h(x) = \frac{1}{h\sqrt{2\pi}} exp\left\{-\frac{x^2}{2h^2}\right\}$$

or (Epanechnikov)

$$K_h(x) = \frac{3}{4h}\left(1 - \frac{x^2}{h^2}\right)\mathbf{1}\{|x| \leq h\}$$

What does the bandwidth do?

- If we use the Epanechnikov kernel we get

$$\widehat{m}(x) = \frac{\sum_{t=1}^{T} K_h(x - x_t) r_t}{\sum_{t=1}^{T} K_h(x - x_t)} = \frac{\sum_{t=1}^{T} \left(1 - \frac{(x - x_t)^2}{h^2}\right) \mathbf{1}\{|x - x_t| \leq h\} r_t}{\sum_{t=1}^{T} \left(1 - \frac{(x - x_t)^2}{h^2}\right) \mathbf{1}\{|x - x_t| \leq h\}}$$

- If $h \to \infty$

$$\widehat{m}(x) \to \frac{1}{T} \sum_{t=1}^{T} r_t$$

and if $h \to 0$

$$\widehat{m}(x) \to r_t$$

where r_t is the observation for which $|x - x_t|$ is smallest within the sample

Bandwidth Selection

- Fan and Yao (2003) suggest

$$h = 1.06sT^{-1/5}$$

for the Gaussian Kernel and

$$h = 2.34sT^{-1/5}$$

for the Epanechnikov kernel where s is the sample standard deviation of $\{x_t\}$ which is assumed stationary

Cross Validation

- Let

$$\widehat{m}_{h,j}(x_j) = \frac{1}{T-1}\sum_{t\neq j} w_t(x_j)y_t$$

which is an estimate of y_j where the weights sum to $T-1$.

- Also let

$$CV(h) = \frac{1}{T}\sum_{j=1}^{T}\left[y_j - \widehat{m}_{h,j}(x_j)\right]^2 W(x_j)$$

where $W(\cdot)$ is anonnegative weight function satisfying $\sum_{j=1}^{T} W(x_j) = T$

Cross Validation

- The function $CV(h)$ is called the cross-validation function since it validates the ability of the smoother m to predict $\{y_t\}$

- The weight function W may be chosen to downweight certain observations if necessary but $W\left(x_j\right) = 1$ is often sufficient

Cross Validation

- It is an exercise to show that

$$CV(h) = \frac{1}{T}\sum_{j=1}^{T}\left[y_j - \widehat{m}_{h,j}(x_j)\right]^2$$

$$= \frac{1}{T}\sum_{j=1}^{T}\left[y_j - \widehat{m}(x_j)\right]^2 \Big/ \left(1 - \frac{K_h(0)}{\sum_{i=1}^{T} K_h(x_j - x_i)}\right)^2$$

N-W Volatility Estimation

- Assume that $\{r_t\}$ is our log-return series that has been centered at zero so that

$$E[r^2{}_t] = \sigma^2{}_t$$

- We also assume that $r^2{}_t = \sigma^2{}_t + \varepsilon_t$ where ε_t is WN

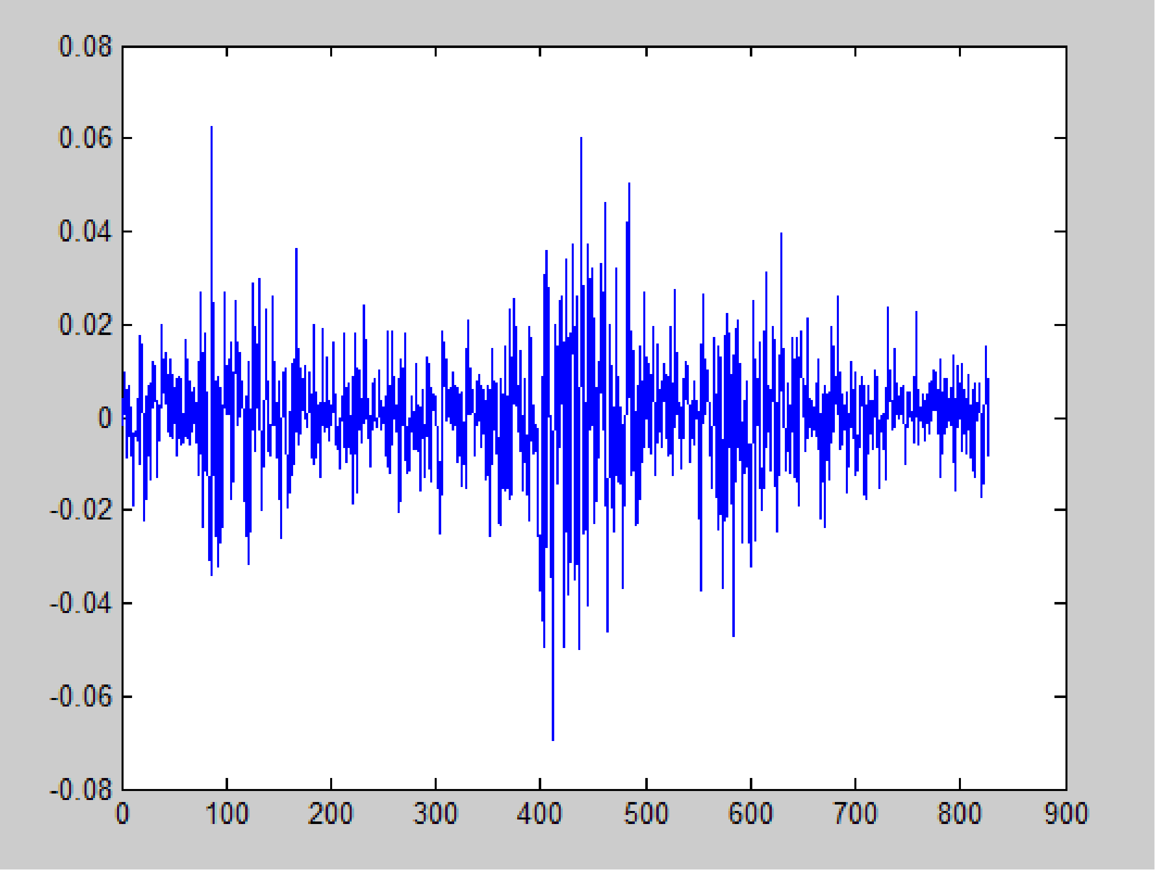
N-W Volatility Estimation

- We may then use the N–W kernel estimator

$$\hat{\sigma}^2{}_t = \frac{\sum_{i=1}^{t-1} K_h(t-i) r^2{}_i}{\sum_{i=1}^{t-1} K_h(t-i)}$$

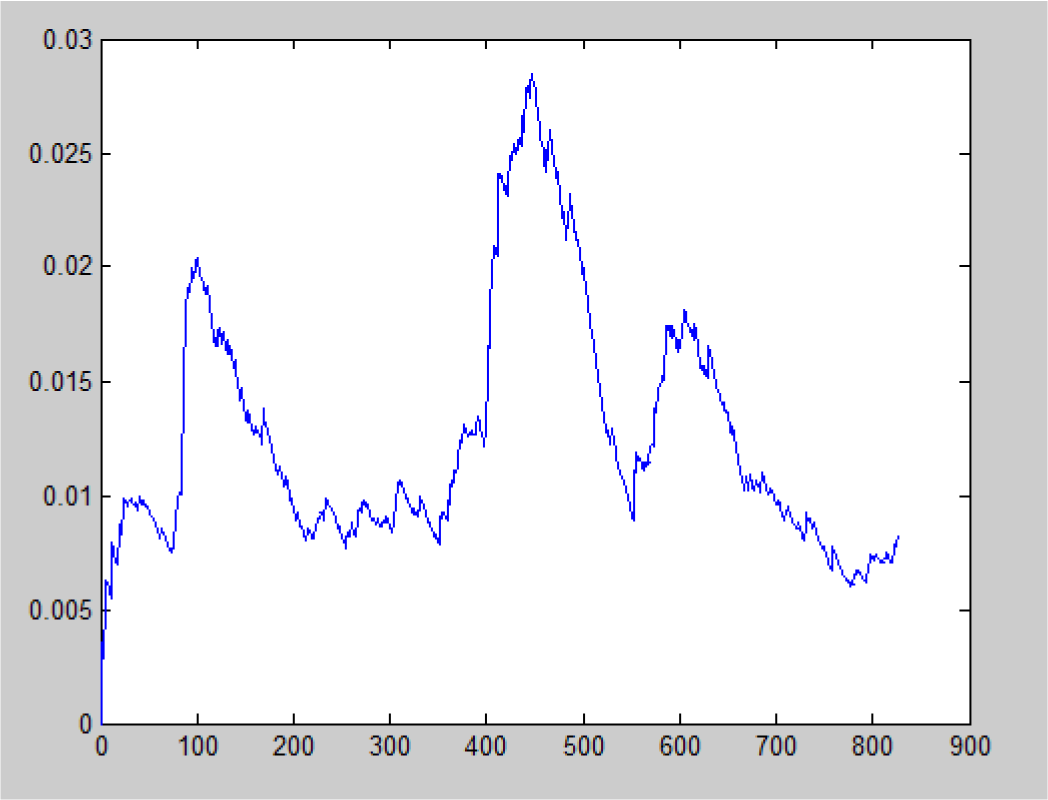
- Below we use Gaussian kernels for OMXS30 data

OMXS30 returns (100104-130412)

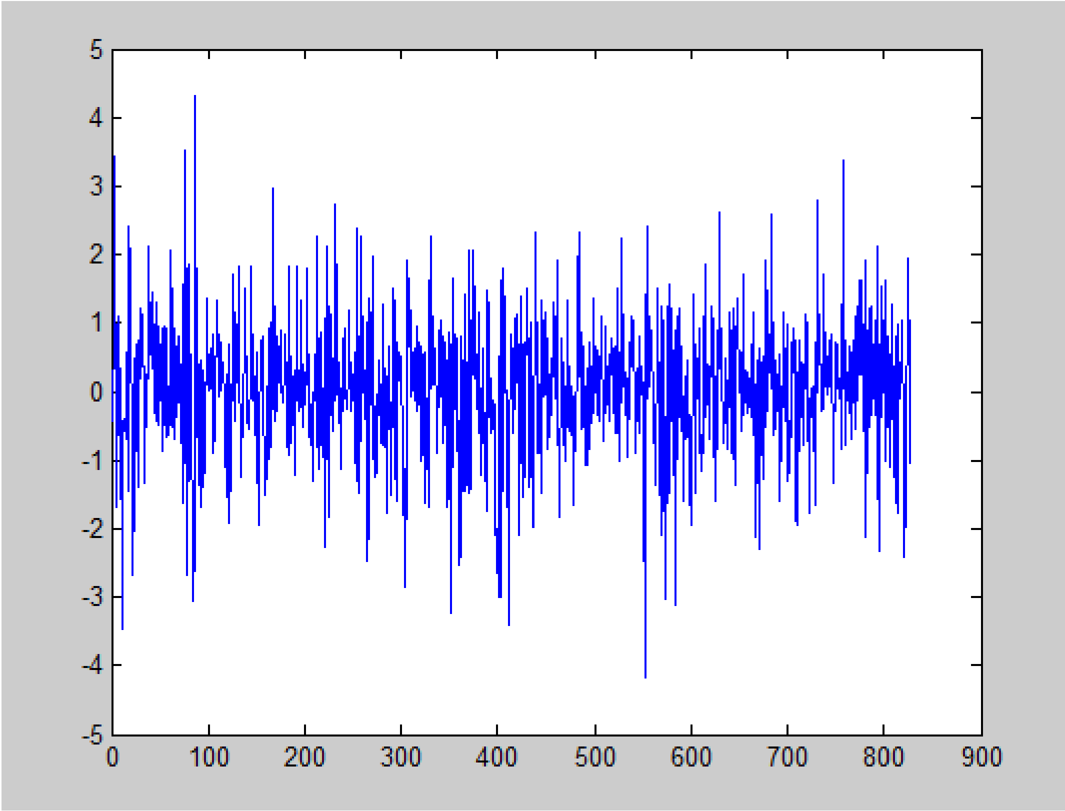


OMXS30 N-W volatility estimates

- CV gives $h = 23.26$ and volatility estimates:

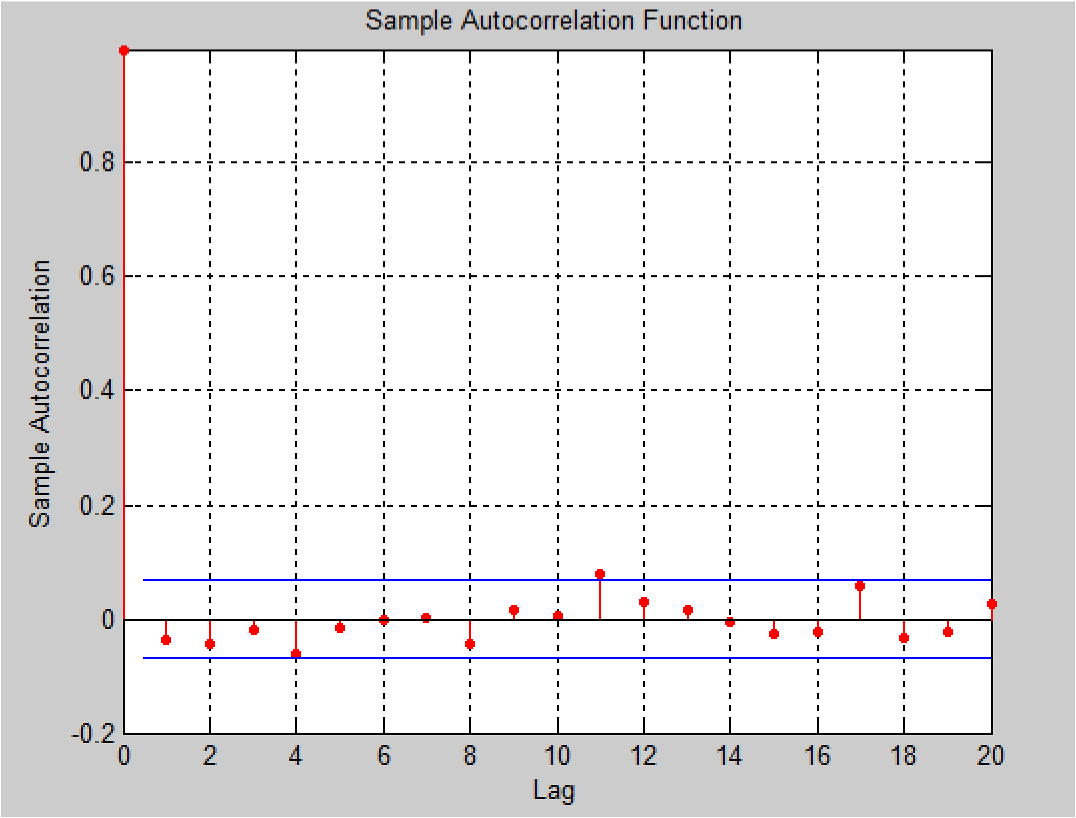


OMXS30 N-W volatility devolatized returns



The Ljung-Box null hypothesis of no autocorrelations cannot be rejected at 5% level $p=0.4433$

Autocorrelation for devolatized returns



Another application

- Remember the ARIMA example for the Swedish GDP

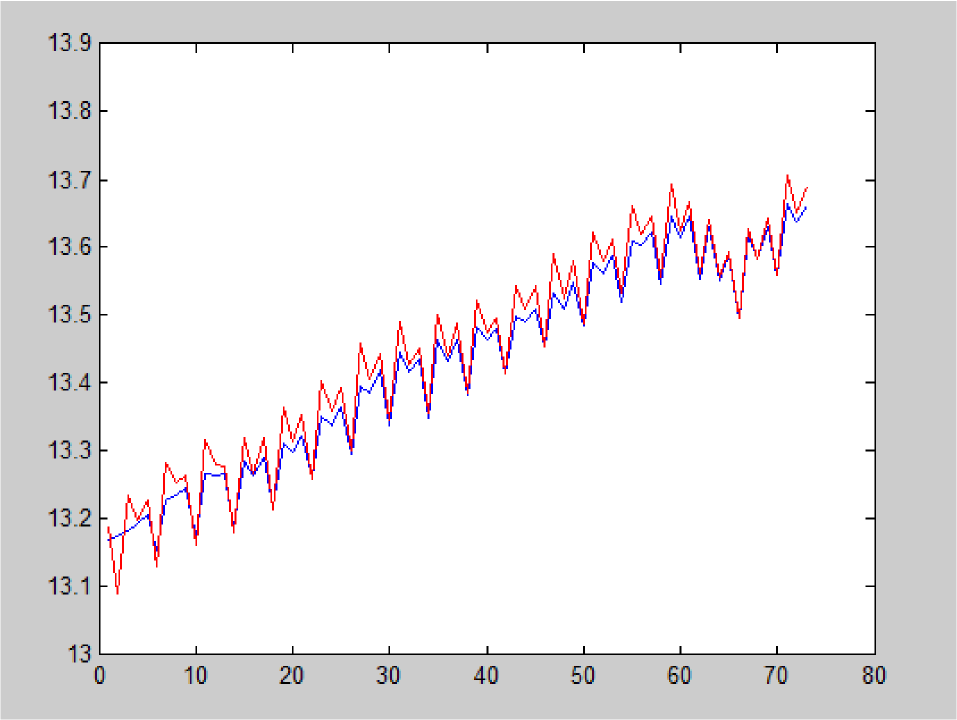
- What if we try to use N-W to model log GDP p_t as

$$p_t = m(p_{t-1}) + a_t$$

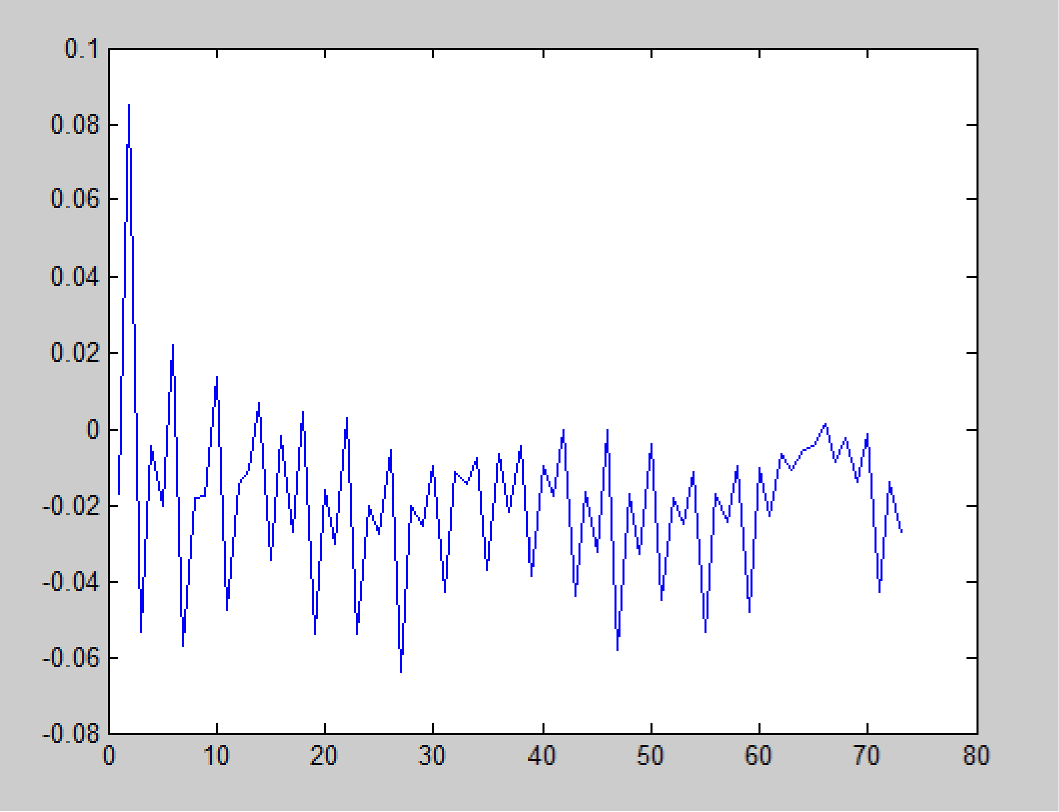
where a_t is WN

Log BNP

- Using CV we find $h = 0.0367$ to be the optimal threshold for the Gaussian kernel and



Residuals



- The seasonality seen using the ARIMA is still left, but we know how to deal with that