Chapter 8

Nonparametric Regression

Data $\Rightarrow$ distribution.

$\rightarrow$ assume some family, for example, $N(\mu, \sigma^2) \Rightarrow \hat{\mu}, \hat{\sigma}^2$

$\rightarrow$ without any distributional assumptions. $\Rightarrow$ form from the data.

Nonparametric regression

explanatory

dependent

$$\{(X_i, Y_i)\},\ i = 1, \ldots, n; \quad \boldsymbol{X} \in \mathbb{R}^{J+1}, Y \in \mathbb{R}$$

- Engel curve: X = net-income, Y = expenditure

$$Y = m(X) + \varepsilon$$

how to model it without specifying the functional form.

- CHARN model: time series of the form

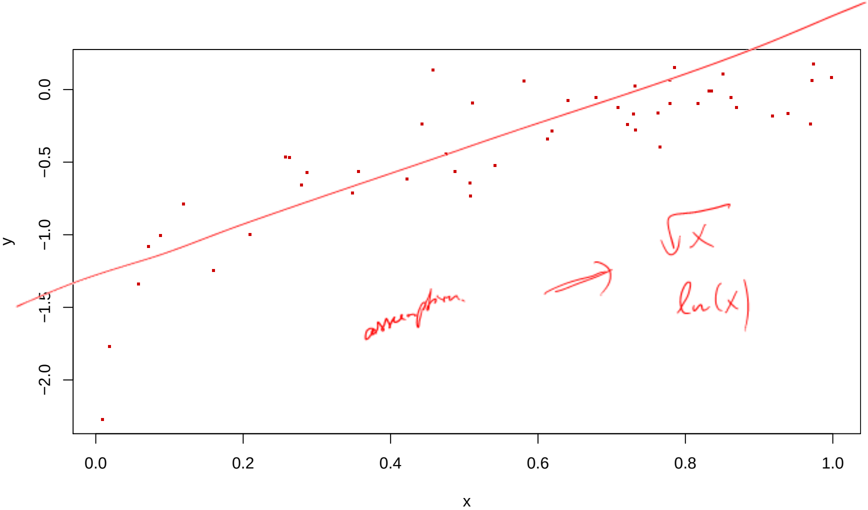
return on investment.

$$Y_t = \underbrace{m(Y_{t-1})}_{\text{some function of last period return}} + \underbrace{\sigma(Y_{t-1})\xi_t}_{\text{risk which depends on the last period return.}}$$

estimate m & σ without fixing the form.

Universität Augsburg

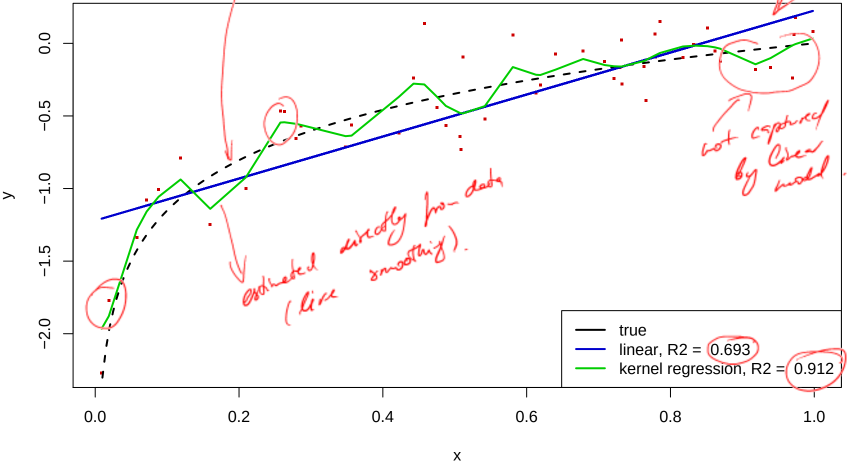
Let us have a sample of size $n = 50$ from an unknown model.



The data is simulated from the model

$$\underbrace{y_i = 0.5\ln(x_i) + 0.2u_i}, \qquad u_i \sim N(0,1).$$

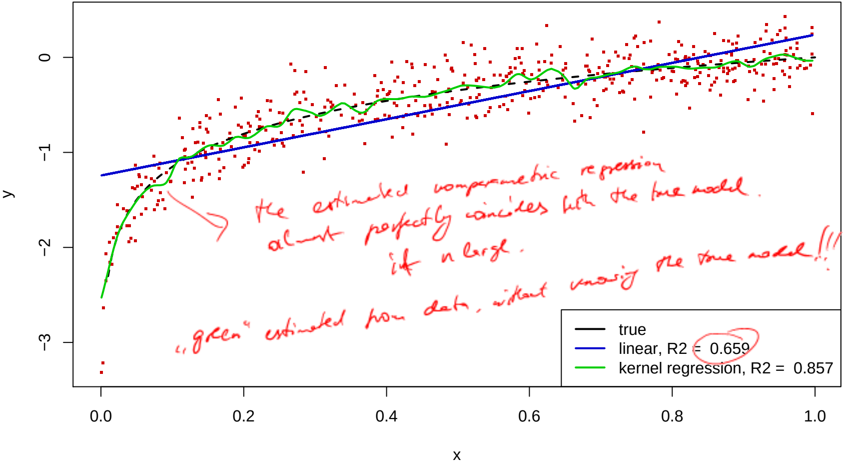
LR.



estimated directly from data (like smoothing).

not captured by linear model.

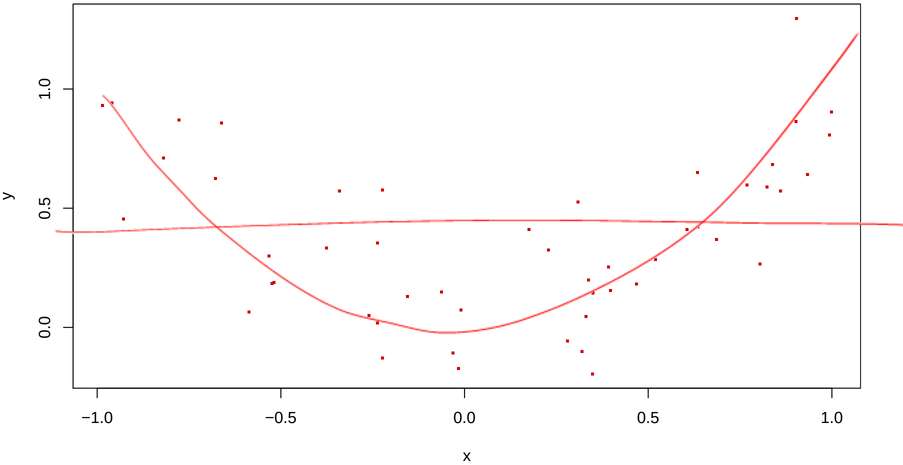
For $n = 500$



the estimated nonparametric regression almost perfectly coincides with the true model. if n large.

„green" estimated from data, without knowing the true model!!!

Another example.



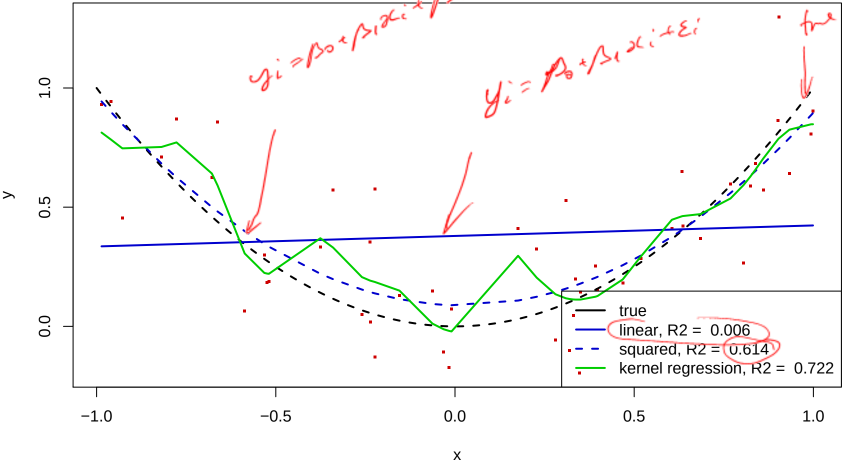
The data is simulated from the model

$$y_i = x_i^2 + 0.2u_i, \qquad u_i \sim N(0,1).$$

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$

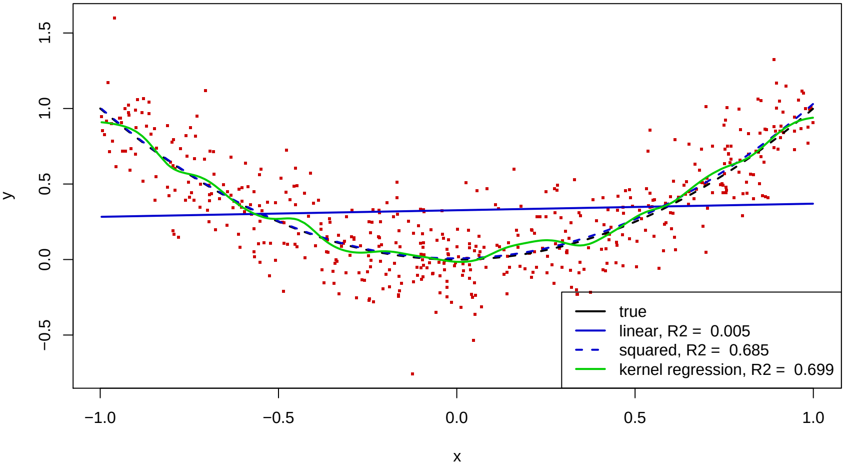
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

true

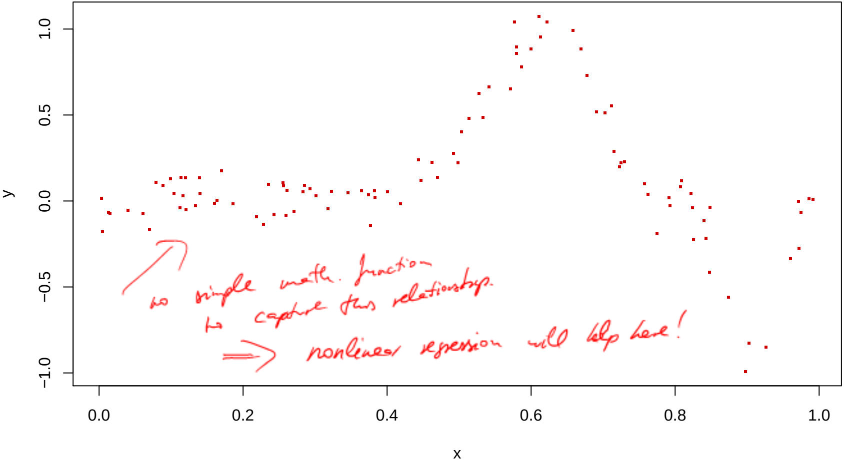


Universität Augsburg

... and with $n = 500$.

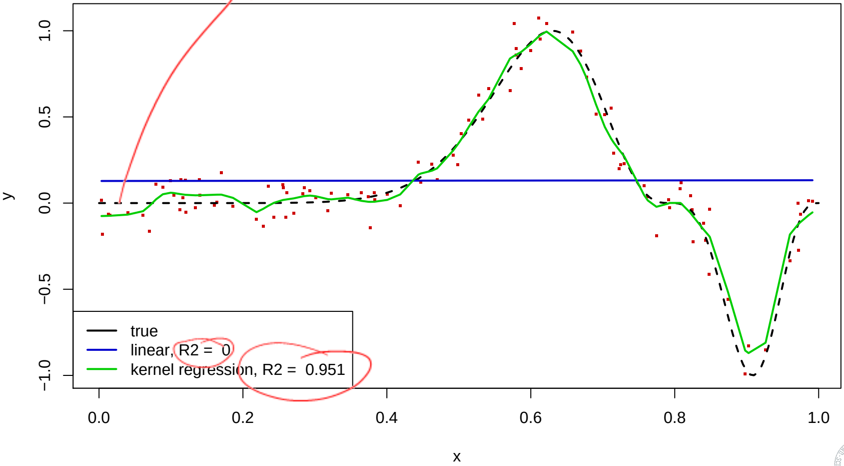


Hopeless example ...



... simulated from

$$y_i = \{sin(2\pi x_i^3)\}^3 + 0.1u_i, \qquad u_i \sim N(0,1).$$



1st: nonparametric estimator of distribution.
2nd: nonparametric estimator of regression.

Kernel density estimator

First: the procedure requires a non-parametric estimator of a density.

Here: DAX30 returns, 20 years of daily data, 5217 observations with normal and t-densities $\rightsquigarrow$ poor fit in the middle and in the tails

many returns, close to 0.

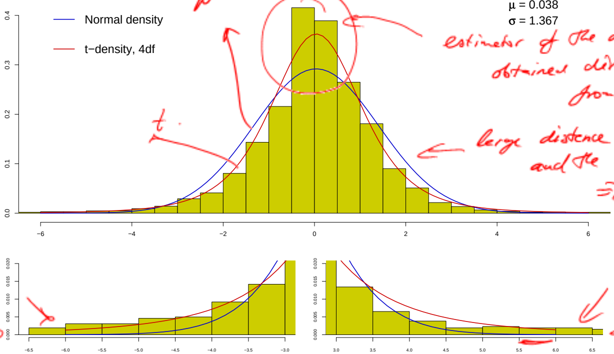
N

t.

estimator of the distribution obtained directly from data.

large distance between N & t and the histogram.

⇒ N & t are not appropriate.



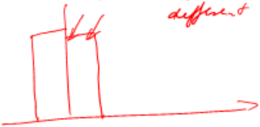
< 0

large losses

many large returns >> large profits

Universität Augsburg

the same probabilities, but different x-values



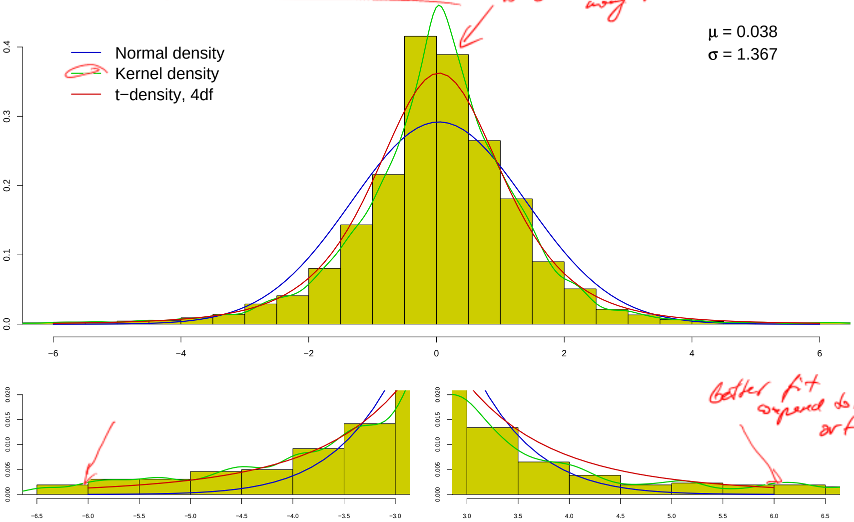
Drawbacks of the histogram

- constant over intervals, step function
- results depend strongly on origin
- binwidth choice ⟹ width of rectangles?
- (slow rate of convergence)

starting point

... and with a kernel density estimator

is high to reflect many returns which are close to 0.



Better fit compared to N dist.

Kernel Density Estimation

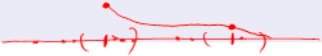
KDE as a generalization of the histogram

Idea of the histogram:

$$\frac{1}{n \cdot \text{interval length}} \#\{\text{obs. that fall into a small interval } \text{CONTAINING } x\}$$

Idea of the kernel density:

$$\frac{1}{n \cdot \text{interval length}} \#\{\text{obs. that fall into a small interval } \text{AROUND } x\}$$



indicator function.

$$\begin{aligned}\widehat{f}_h(x) &= \frac{1}{2hn} \sum_{i=1}^{n} I\left(\left|\frac{x - X_i}{h}\right| \leq 1\right) \\ &= \frac{1}{2hn} \#\{X_i \text{ in interval around } x\}\end{aligned}$$

kernel density estimate (KDE)

bell-shaped

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

with kernel function $K(u)$

Required properties of kernels

the square under K = 1.

- $K(\bullet)$ is a density function:

$$\int_{-\infty}^{\infty} K(u)du = 1 \quad \text{and} \quad K(u) \geq 0$$

- $K(\bullet)$ is symmetric:

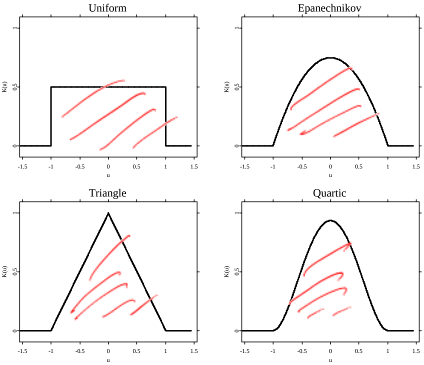
$$\int_{-\infty}^{\infty} uK(u)du = 0$$

Universität Augsburg

Different Kernel Functions

Kernel	$K(u)$	
Uniform	$\frac{1}{2}I(\lvert u\rvert \leq 1)$	
Triangle	$(1-\lvert u\rvert)I(\lvert u\rvert \leq 1)$	theoretical properties.
Epanechnikov	$\frac{3}{4}(1-u^2)I(\lvert u\rvert \leq 1)$	
Quartic	$\frac{15}{16}(1-u^2)^2 I(\lvert u\rvert \leq 1)$	
Triweight	$\frac{35}{32}(1-u^2)^3 I(\lvert u\rvert \leq 1)$	
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$	$\Rightarrow$ because of the normal.
Cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}u)I(\lvert u\rvert \leq 1)$	

Table: Kernel functions



every is symmetric

square = 1.

Figure: Some kernel functions: Uniform (top left), Triangle (bottom left), Epanechnikov (top right), Quartic (bottom right)

Example: Construction of the KDE

consider the KDE using a Gaussian kernel

$$\begin{aligned}\widehat{f}_h(x) &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\\ &= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\end{aligned}$$

here we have

the point where the density must be estimated

$$u = \frac{x-X_i}{h}$$

data point (sample)

bandwidth

Universität Augsburg

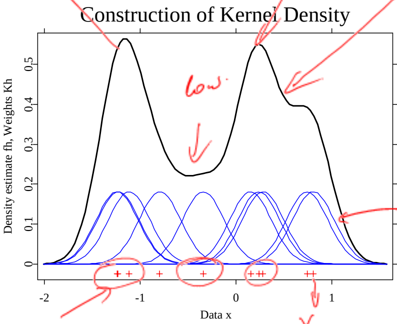
many points ⇒ high.

high

sum of all blue curves.

⇒ our $\hat{f}(x)$

⇒ it is smooth in contrary to the histogram!



low.

X_1.

for every X_i we have this bell-shaped curve

Figure: Kernel density estimate as a sum of bumps

$$X_1 \Rightarrow \frac{1}{n\cdot h\sqrt{2\pi}} \exp\left(\frac{x-X_1}{h}\right)^2 \Rightarrow$$

Plotting kernel estimators in R

bandwidth „method" K functions.

```
density(x, bw = "nrd0", adjust = 1,
        kernel = c("gaussian", "epanechnikov", "rectangular",
                   "triangular", "biweight", "cosine", "optcosine"),
        weights = NULL, window = kernel, width, give.Rkern = FALSE,
        n = 512, from, to, cut = 3, na.rm = FALSE, ...)
```

```
Value:
x: the 'n' coordinates of the points where the density is
          estimated.
y: the estimated density values.  These will be non-negative,
          but can be zero.
bw: the bandwidth used.
```


Universität Augsburg

```
> x = rnorm(100)
> k = density(x)
> k
Data: x (100 obs.);     Bandwidth 'bw' = 0.3029
       x                  y
 Min.   :-3.4975   Min.   :0.000175
 1st Qu.:-1.7048   1st Qu.:0.023764
 Median : 0.0879   Median :0.061481
 Mean   : 0.0879   Mean   :0.139314
 3rd Qu.: 1.8806   3rd Qu.:0.248836
 Max.   : 3.6733   Max.   :0.443788
> k$x
  [1] -3.497498092 -3.483465207 -3.469432322 -3.455399437 -3.441366553
  [6] -3.427333668 -3.413300783 -3.399267898 -3.385235013 -3.371202128
  ...
> k$y
  [1] 0.0002050555 0.0002360509 0.0002706135 0.0003111481 0.0003565894
  [6] 0.0004070709 0.0004629191 0.0005277777 0.0005996858 0.0006789000
  ...
```

100 obs from $N(0,1)$

grid on the X-axis

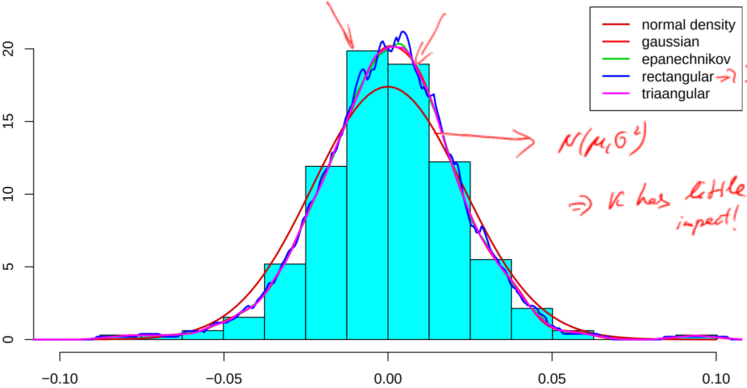
$\hat{f}(\text{grid})$

Universität Augsburg

Kernel estimator of the density of DJ30 returns with different kernels and “optimal” bandwidths.

histogram

all KDE ⇒ similar
⇒ except the blue.



⇒ Indicator

$N(\mu, \sigma^2)$

⇒ K has little impact!

Universität Augsburg

```
z1=density(rdj, bw="nrd0",  kernel="epanechnikov")

pdf("ch2_hist_dax.pdf", width=9, height=5, onefile=FALSE);
truehist(rdj,  prob=TRUE, xlab="",xlim=c(-0.07,0.07), col="grey80");
matplot(z1$x,z1$y, add=T, type="l", lty=1, col=2, lwd=2);
dev.off()
```

Gaussian kernel for the density of DJ30 returns with different adjusted “optimal” bandwidths.

0.5 of the optimal bandwidth

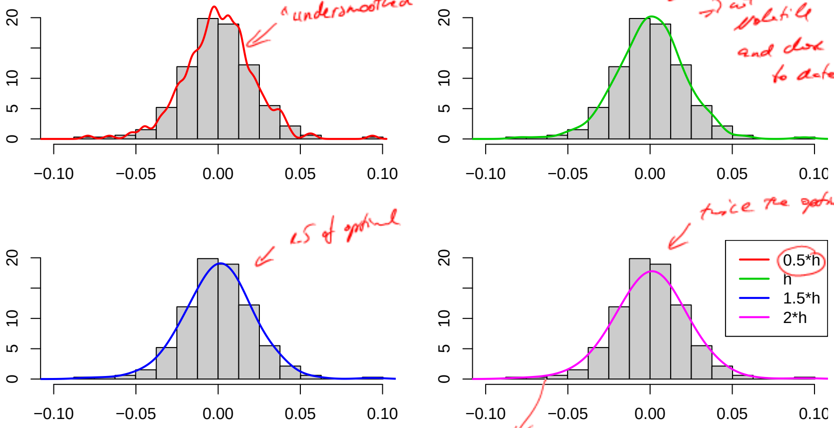
“undersmoothed“

optimal

⇒ not too volatile and close to data.

1.5 of optimal

twice the optimal.

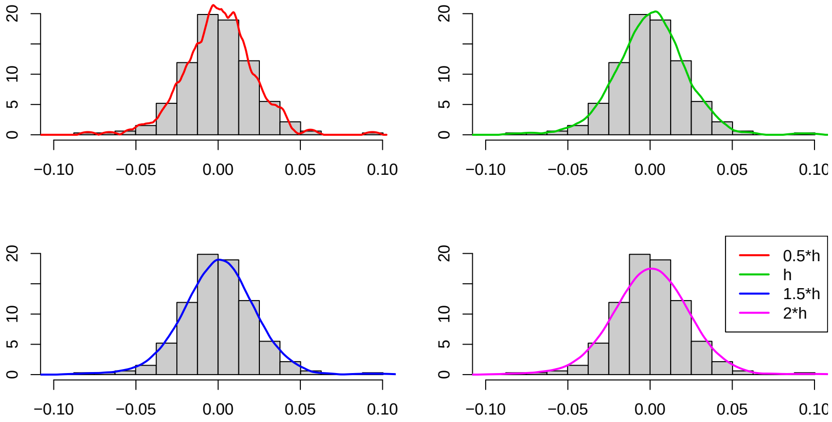


starts to deviate from data

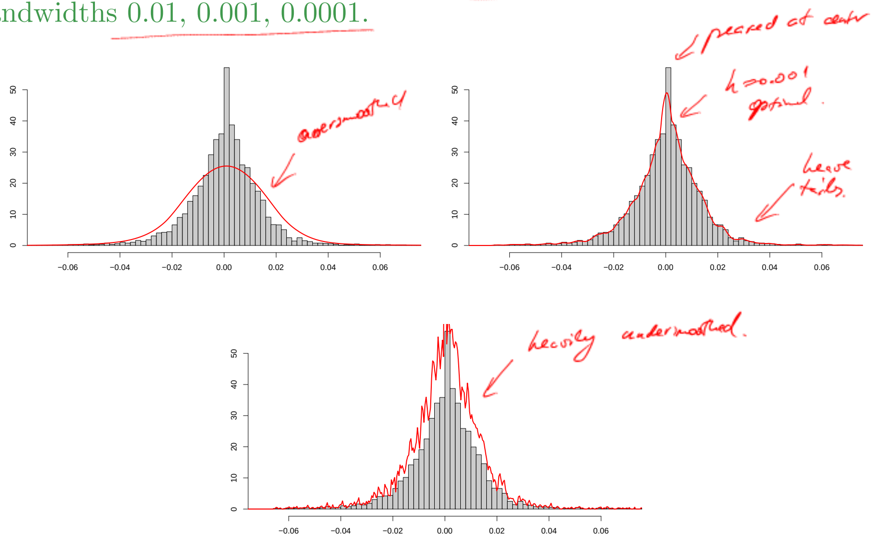
“oversmoothed“.

Epanechnikov kernel for the density of DJ30 returns with different adjusted "optimal" bandwidths.

almost no visible changes if we move from Gaussian to Epan

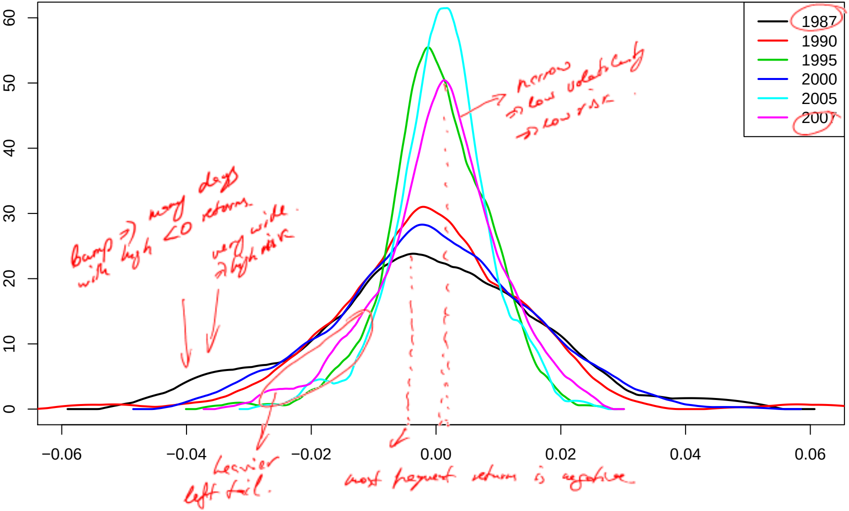


Epanechnikov kernel for the density of DAX30 returns with the bandwidths 0.01, 0.001, 0.0001.



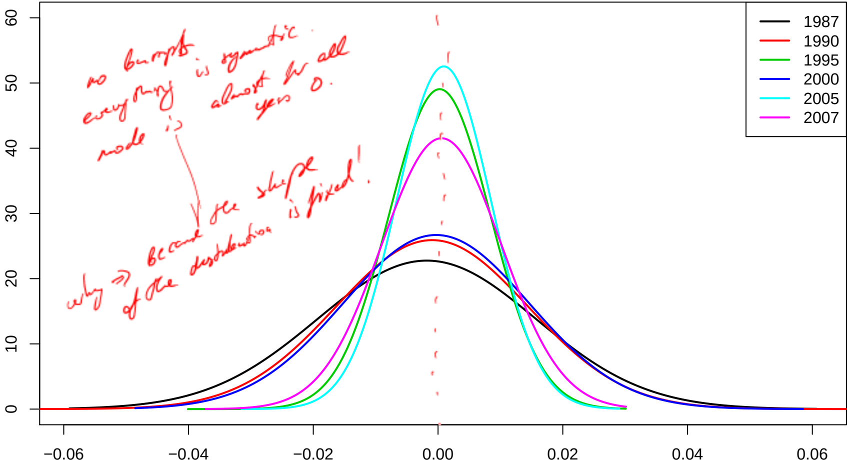
Universität Augsburg

Example: time-variation of the distribution of DAX returns (Epanechnikov kernel)



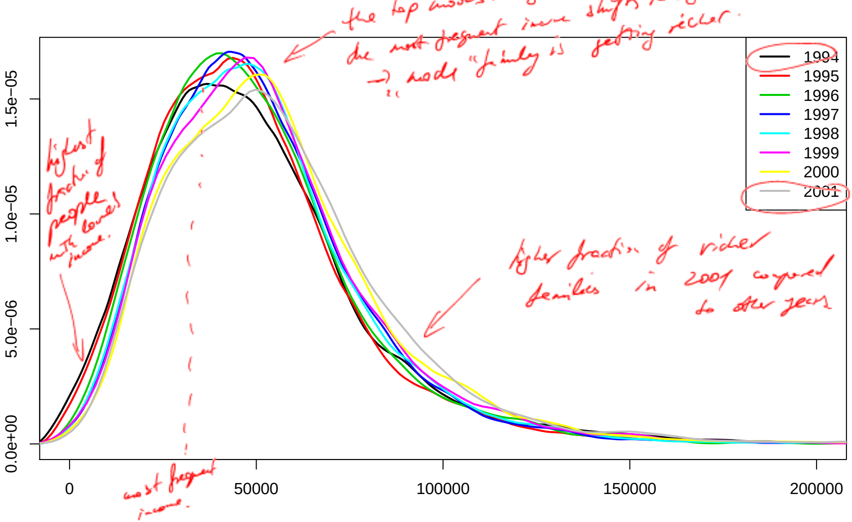
Universität Augsburg

Example: time-variation of the distribution of DAX returns (normal density)

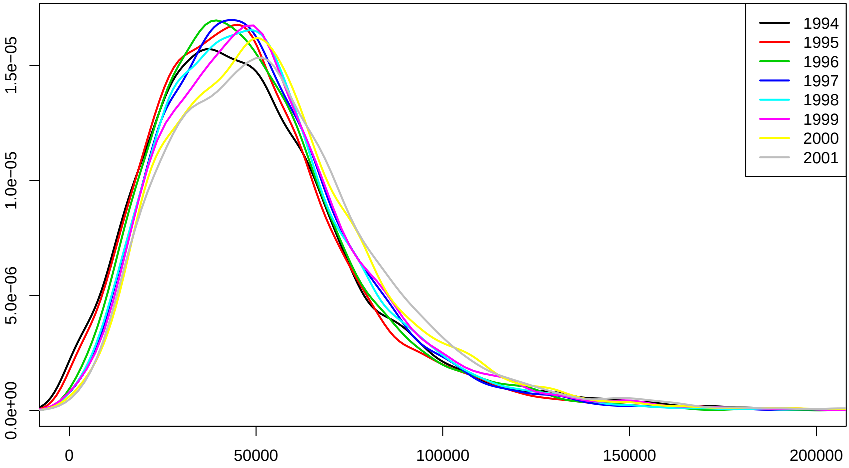


Universität Augsburg

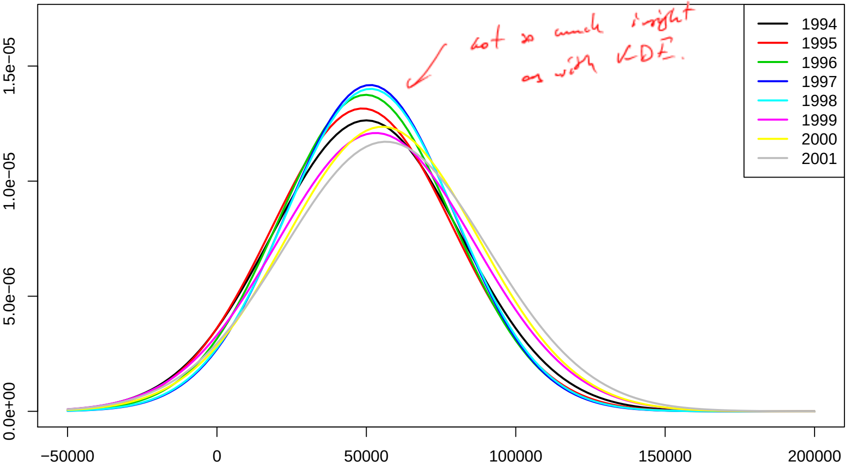
European Community Household Panel for the period 1994-2001. Data for Germany, net household income, ca. 48000 observations and Epanechnikov kernel.



European Community Household Panel for the period 1994-2001. Data for Germany, net household income, ca. 48000 observations and gaussian kernel.



European Community Household Panel for the period 1994-2001. Data for Germany, net household income, ca. 48000 observations and gaussian density .



not so much insight as with KDE.

(Asymptotic) statistical properties of KDE bias of the kernel density estimator

$E(\hat{\theta}) - \theta = 0$ if unbiased

$$
\begin{aligned}
Bias\left\{\widehat{f}_h(x)\right\} &= E\left[\widehat{f}_h(x)\right] - f(x) \\
&= E\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right] - f(x) \\
&\approx \frac{h^2}{2} f''(x)\mu_2(K) \quad \text{for } h \to 0
\end{aligned}
$$

Bias disappears if bandwidth is getting smaller!!!

$\Rightarrow$ characteristic of K

$\Rightarrow$ characteristic of the true density

where $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u)du$.

$$
\begin{aligned}
Var\left[\widehat{f}_h(x)\right] &= Var\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right] \\
&\approx \frac{1}{nh}\|K\|_2^2 f(x) \quad \text{for } nh \to \infty
\end{aligned}
$$

$\rightarrow$ true density.

$\Rightarrow$ characteristic of K

$\Rightarrow$ h increases $\Rightarrow$ Var $\searrow$.

sample size.

where $\|K\|_2^2 = \int_{-\infty}^{\infty} \{K(u)\}^2 du$.

trade off: $h \nearrow \Rightarrow Var \searrow$ if $h \searrow \Rightarrow Bias \searrow$] h_{opt} - ?

Universität Augsburg

How to choose the bandwidth for the KDE?

find the bandwidth which minimizes the $MISE$

$$
\begin{aligned}
\underbrace{MISE}_{\text{mean integrated squared error}}\left\{\widehat{f}_h(x)\right\} &= E\left[\int_{-\infty}^{\infty}\{\widehat{f}_h(x)-f(x)\}^2dx\right] \\
&= \int_{-\infty}^{\infty} MSE[\widehat{f}_h(x)]dx \\
&\approx \underbrace{\frac{1}{nh}\|K\|_2^2+\frac{h^4}{4}\mu_2(K)^2\|f''\|_2^2}_{\to \text{ min w.r.t. } h} = AMISE\left\{\widehat{f}_h(x)\right\}
\end{aligned}
$$

known / fixed.

and thus

$$
\boxed{h_{opt}} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2\mu_2(K)^2 n}\right)^{1/5} \sim n^{-1/5}
$$

are determined by K

Problematic: to estimate f you need h, but to compute h you need f!

$\|f''\| \Rightarrow$ normal density $\Rightarrow$ „Silverman's rule of thumb"

Multivariate KDE

d-dimensional data and d-dimensional kernel

$$\boldsymbol{X} = (X_1, \dots, X_d)^\top, \quad \mathcal{K} : \mathbb{R}^d \to \mathbb{R}_+$$

- multivariate kernel density estimator (simple)

$$\widehat{f}_h(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^d} \mathcal{K}\left(\frac{x_i - X_i}{h}\right)$$

  each component is scaled equally.

- multivariate kernel density estimator (more general)

multivariate kernel $\mathcal{K}$.

$$\widehat{f}_h(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_1 \cdot \ldots \cdot h_d} \mathcal{K}\left(\frac{x_1 - X_1}{h_1}, \dots, \frac{x_d - X_d}{h_d}\right)$$

individual bandwidths for each dimension.

Universität Augsburg

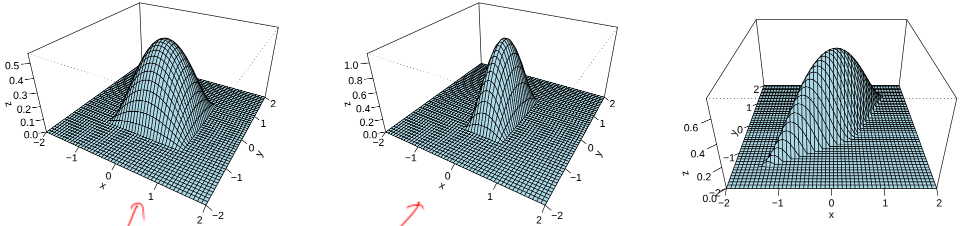
- multivariate kernel density estimator (most general)

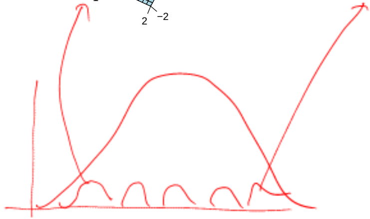
$$\widehat{f}_{\mathbf{H}}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\det(\mathbf{H})}\mathcal{K}\left\{\mathbf{H}^{-1}(\boldsymbol{x}-\boldsymbol{X}_i)\right\}$$

where $\mathbf{H}$ is a (symmetric) bandwidth matrix.
Each component is scaled separately, correlation between components can be handled.

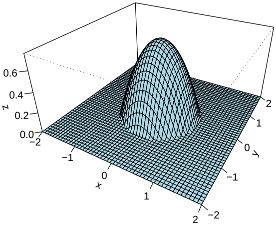
Example: product Epanechnikov kernel with bandwidth matrices

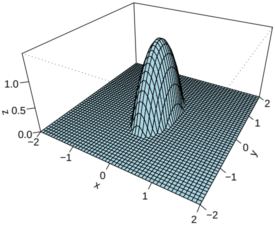
$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

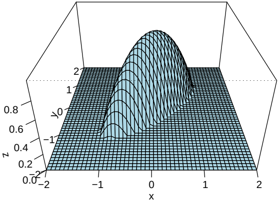




Example: radially symmetric Epanechnikov kernel with bandwidth matrices $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{H} = \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{H} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$







Universität Augsburg

Example: bandwidth matrix $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

more weights in the angles of the distribution

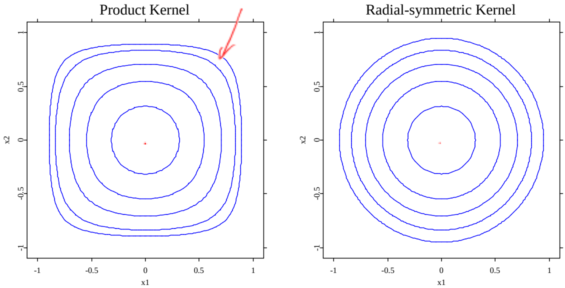


Figure: Contours from bivariate product (left) and bivariate radially symmetric (right) Epanechnikov kernel

Example: bandwidth matrix $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$

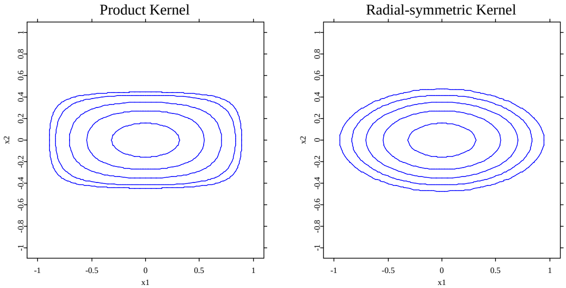


Figure: Contours from bivariate product (left) and bivariate radially symmetric (right) Epanechnikov kernel

Example: bandwidth matrix $\mathbf{H} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{1/2}$

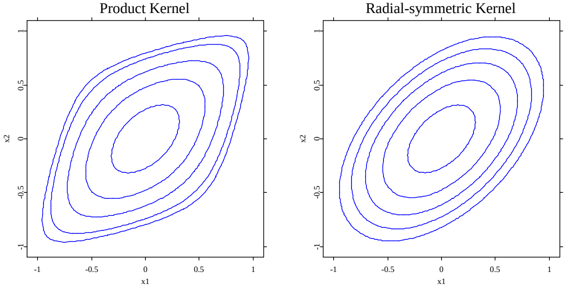


Figure: Contours from bivariate product (left) and bivariate radially symmetric (right) Epanechnikov kernel

Kernel properties

$\mathcal{K}$ is a density function

$$\int_{\mathbb{R}^d} \mathcal{K}(\boldsymbol{u})d\boldsymbol{u} = 1,\ \mathcal{K}(\boldsymbol{u}) \geq 0$$

$\mathcal{K}$ is symmetric

$$\int_{\mathbb{R}^d} \boldsymbol{u}\mathcal{K}(\boldsymbol{u})d\boldsymbol{u} = 0_d$$

Italian GDP growth panel for 21 regions covering the period 1951-1998 (millions of Lire, 1990=base). There are 1008 observations in total.

```
data("Italy")
fhat = npcdens(gdp ~ year, tol=0.1, ftol=0.1, data=Italy)
summary(fhat)
plot(fhat, view="fixed", main="", theta=300, phi=50)
```

⇒ 2-dim. data

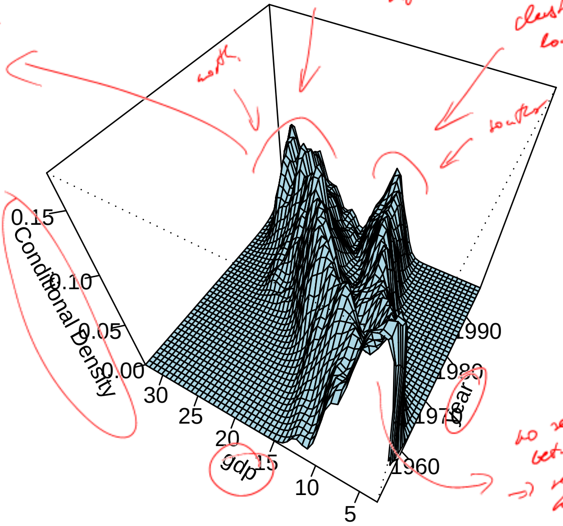
```
Conditional Density Data: 1008 training points, in 2 variable(s)
(1 dependent variable(s), and 1 explanatory variable(s))
                               gdp
Dep. Var. Bandwidth(s): 0.697371
                               year
Exp. Var. Bandwidth(s): 0.6725248

Bandwidth Type: Fixed
Log Likelihood: -2550.493
```

bimodal density
⇒ no way to model it
parametrically!

a cluster of
regions with
higher gdp

cluster with
lower gdp.



no separation
between the regions
⇒ region very
homogenous.

x_1 x_2

Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

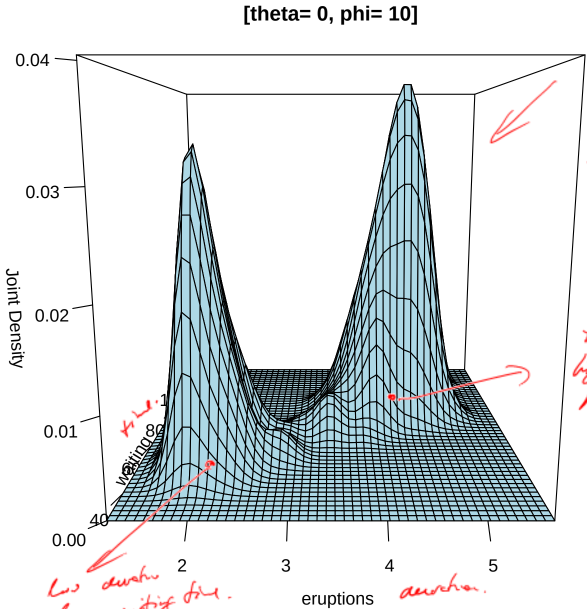
```
library("datasets");library("np");
data("faithful"); attach(faithful);

bw = npudensbw(dat=faithful)
summary(bw)

npplot(bws=bw, xtrim=-0.2)
detach("faithful")
```

```
Data (272 observations, 2 variable(s)):
Bandwidth Selection Method: Maximum Likelihood Cross-Validation
Bandwidth Type: Fixed
Objective Function Value: 4.197484 (achieved on multistart 1)

Var. Name: eruptions Bandwidth: 0.1470088 Scale Factor: 0.327852
Var. Name: waiting   Bandwidth: 2.925438  Scale Factor: 0.5477395
```



time

no chances to model parametrically

=> there are no popular bimodal densities.

two clusters high duration and high waiting

low duration low waiting time.

duration.

Summary for KDE

- The KDE does not depend on the starting points of the classes.
- The resulting estimator is a smooth and continuous function.
- The estimator heavily depends on the bandwidth. ⇒ h_{opt}
- The estimator is robust to different choices of the kernel function.
  ⇒ Gauss, Epane
- The estimator is biased in general.
- Decreasing bandwidth implies smaller bias, but larger variance.

→ two slides with formulas.

h_{opt} ⇒ trade-off (as for usual parameters)

LR! $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \Rightarrow$ we approx. m by a linear function

$E(Y_i | X_i) = \beta_0 + \beta_1 X_i$

Univariate nonparametric regression

Here: estimate $m(\cdot)$ from data, without fixing it.

model

on analogy $E(Y_i | X_i = x_i)$

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

$m(\bullet)$ smooth regression function, ε_i i.i.d. error terms with $E\varepsilon_i = 0$

we aim to estimate the conditional expectation of Y given $X = x$

cond. expect — leads to a conditional density function. — joint density bivariate.

$$m(x) = E(Y|X = x) = \int y\, f(y|x)\, dy = \int y \frac{f(x, y)}{f_X(x)}\, dy$$

marginal density of X.

where $f(x, y)$ denotes the joint density of (X, Y) and $f_X(x)$ the marginal density of X

$$= \frac{\left(\int y \hat{f}(x,y)\, dy\right)}{\hat{f}_X(x)}$$

$\hat{f}$ and $\hat{f}_X$ can be estimated by KDE $\Rightarrow \hat{m}$ without any distributional of functional assumptions!!!

$\boxed{E(Y) \underset{\text{Def}}{=} \int y f(y)\, dy}$ → density of y

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ Def of conditional probability

Universität Augsburg

Nadaraya-Watson Estimator

idea: (X_i, Y_i) have a joint pdf, so we can estimate $m(\bullet)$ by a multivariate kernel estimator

KDE.

estimator for $f(x,y)$

$$\widehat{f}_{h,\widetilde{h}}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{1}{h}K\left(\frac{x-X_i}{h}\right)}_{\text{part for } X} \underbrace{\frac{1}{\widetilde{h}}K\left(\frac{y-Y_i}{\widetilde{h}}\right)}_{\text{for } y.}$$

and therefore $\int y\, \widehat{f}_{h,\widetilde{h}}(x,y)dy = \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)Y_i$ ⇒ what is left from the integral.

resulting estimator:

$\hat{f}(x,y)$

no functional assumptions!!!

estimated regression function

$$\widehat{m}_h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x-X_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x-X_i)} = \frac{\widehat{r}_h(x)}{\widehat{f}_h(x)}$$

$K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$

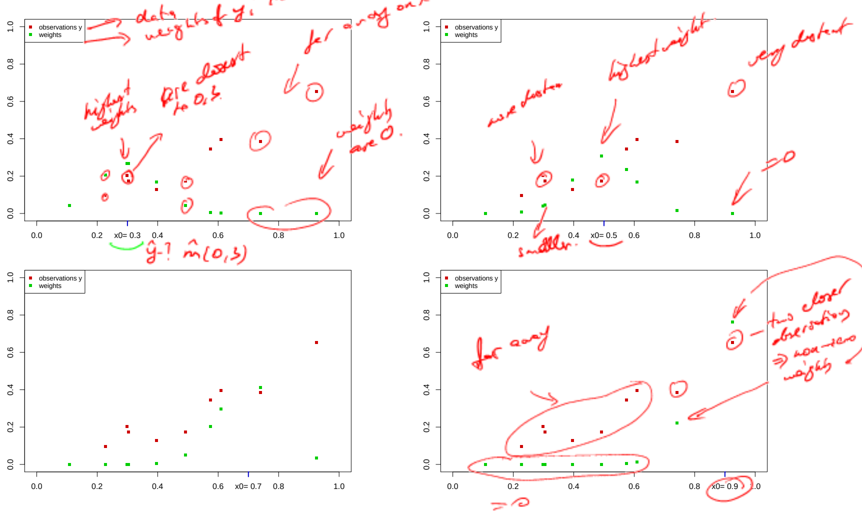
$f_X(x) \to$ KDE

green points on the next slide.

Observations and weights W_{hi} for $h = 0.1$, Gaussian kernel and different values of $x = x_0$.

into NW-formula.

on X-axis.



Universität Augsburg

Example: happiness

The happiness of nations (measured as average happiness of the citizens) depends of the average income per capita.

Data: 62 nations

? $\Rightarrow$ Nadaraya-Watson to estimate m.

$$\text{happiness} = m(\text{income}) + \varepsilon.$$

Nonparametric regression in R

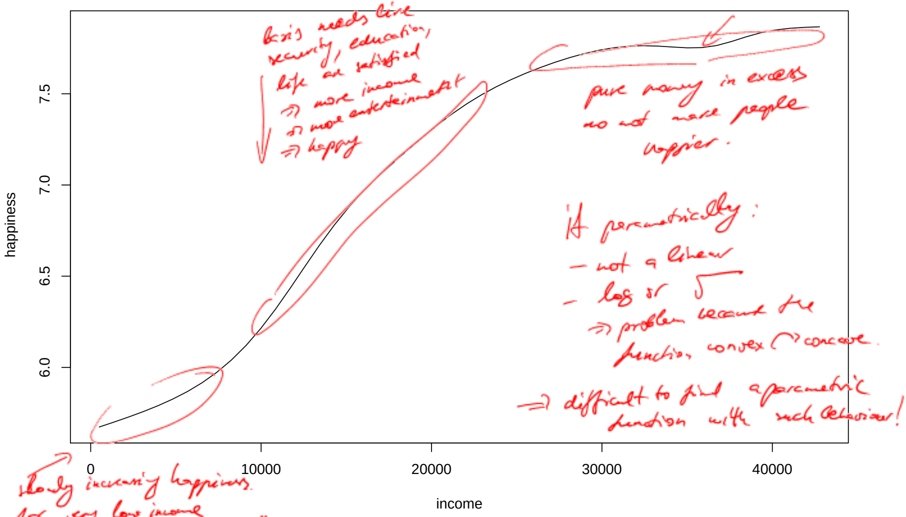
```
library("np")
all = read.table("ch3_happiness.txt");
happiness = all[,1];
income = all[,2];

bw = npregbw(formula=happiness ~ income, lt="lc")
model = npreg(bws=bw);

npplot(bw, type="l");
```

one function to determine the bandwidth

one function to compute the estimator.



steady increasing happiness for very low income (other factors like „security" make people happy)

Universität Augsburg

Example: wage in Canada

Canadian cross-section wage data consisting of a random sample taken from the 1971 Canadian Census Public Use Tapes for male individuals having common education (Grade 13). There are n = 205 observations in total, and 2 variables, the logarithm of the individual's wage (logwage) and their age (age).

to guarantee >0 forecasts and make residuals symmetric

$$\underbrace{log(\text{wage})} = m(\text{age}) + \varepsilon$$

Universität Augsburg

$logwage = \beta_0 + \beta_1 \cdot wage + \varepsilon_i$

```
data("cps71")
model.par = lm(logwage ~ age , data = cps71)
summary(model.par)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.021902   0.144705   89.99  < 2e-16 ***
age          0.012046   0.003554    3.39  0.00084 ***

Residual standard error: 0.6206 on 203 degrees of freedom
Multiple R-squared: 0.05357,    Adjusted R-squared: 0.04891
F-statistic: 11.49 on 1 and 203 DF,  p-value: 0.0008407
```

$logwage = \beta_0 + \beta_1 \cdot age + \beta_2 age^2 + \varepsilon_i$

```
model.par2 = lm(logwage ~ age+I(age^2), data = cps71)
summary(model.par2

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.0419773  0.4559986  22.022  < 2e-16 ***
age          0.1731310  0.0238317   7.265 7.96e-12 ***
I(age^2)    -0.0019771  0.0002898  -6.822 1.02e-10 ***

Residual standard error: 0.5608 on 202 degrees of freedom
Multiple R-squared: 0.2308,     Adjusted R-squared: 0.2232
F-statistic:  30.3 on 2 and 202 DF,  p-value: 3.103e-12
```

young

middle

salary

economic interpretable

$\beta_2 < 0$

$\beta_2 > 0$

Cross-validation for bandwidth

```
model.np = npreg(logwage ~ age, regtype = "lc", bwmethod = "cv.aic",
                 data = cps71)
summary(model.np)
```

```
Regression Data: 205 training points, in 1 variable(s)
                  age
Bandwidth(s): 1.551218

Kernel Regression Estimator: Local-Constant
Bandwidth Type: Fixed
Residual standard error: 0.2750934
R-squared: 0.3261299
```

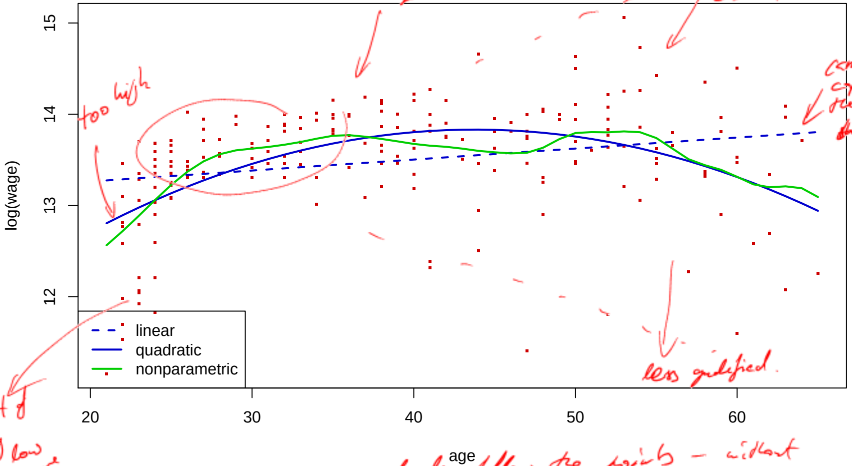
$\Rightarrow$ 10% than for parabola

the highest income for mid-age people

CEO/managerial

too high

cannot capture the relationship



a lot of young low with income

less qualified.

⇒ green – closely follows the points – without any parametric or distributional assumptions!

Example: options

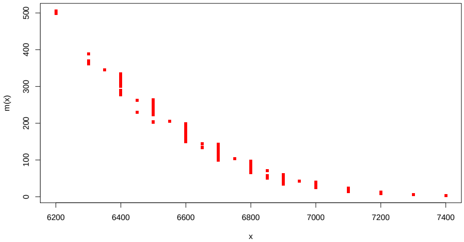
Consider call options on the DAX-certificates. The price of the option $C = C_t(S, K, \tau, r, \sigma^2)$ at time point t depends in a complex way on the asset price S, strike price K, time to maturity τ, risk-free rate r and the volatility σ^2.

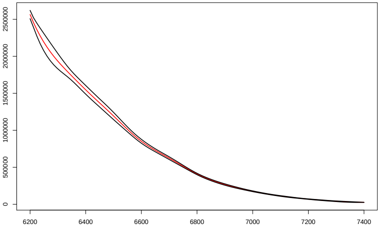
Data:

Call options prices on 17.01.2001 with 1 month to maturity and with different strike prices.

$$C_i = m(K_i) + \varepsilon_i.$$

The NW estimation implies the following behaviour.





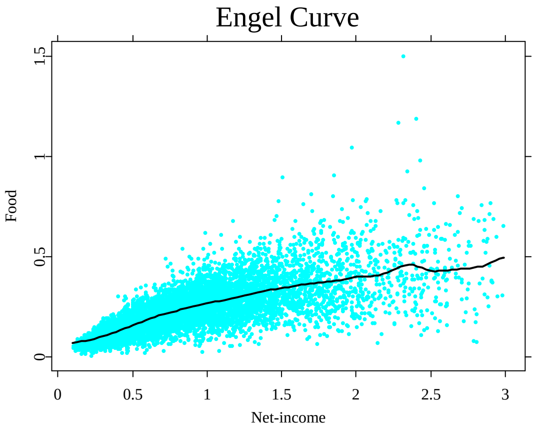


Figure: Nadaraya-Watson kernel regression, $h = 0.2$, U.K. Family Expenditure Survey 1973

KDE: h_{opt} by minimizing MISE ⇒ Silverman's rule of thumb.

Bandwidth selection: Cross Validation

separate estimation and validation by using leave-one-out estimators

out-of-sample prediction for the ith observation

min w.r.t. h,

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \widehat{m}_{h,-i}(X_i)\}^2 w(X_i)$$

weighted CV.

estimated m function, if we drop the i-th observation

minimizing gives $\widehat{h}_{CV}$

out-of-sample forecasts

Important: not really correct for time series data!

Universität Augsburg