RENT ANALYSIS BEYOND CONDITIONAL EXPECTATION – HOW TO ESTIMATE THE SPATIAL DISTRIBUTION OF QUOTED RENTS BY USING GEOGRAPHICALLY WEIGHTED REGRESSION

Anne Seehase, M. Sc., Amt für Statistik, Landeshauptstadt Magdeburg, Germany

July 1, 2016

Scorus-Conference 2016 – Lisbon

CONTENT

MOTIVATION

- 57 percent of all households in Germany are living in a rented accommodation (Statistisches Bundesamt, Wiesbaden 2015, effective 2013).
- Most areas with a high population density are embossed by tight and dynamically developing rental markets.
- It is expected that the increasing number of households caused by the demographic transition and urbanization leads to increasing rents also in cities like Magdeburg, which right now are characterized by quite a relaxed rental market.
- Monitoring of the developments at the rental markets are relevant for urban planning and social welfare.

ALTERNATIVE DATA SOURCES

ALTERNATIVE DATA SOURCES

- Traditional data collection to rental analysis focus asset rents:
  $\rightarrow$ information of the common (average) level of residental cost over a longer time-period
  $\rightarrow$ insufficient conclusions for the market price by re-letting in very dynamic markets
- Alternative data sources make it possible to focus actual rent prices by re-letting:
  $\rightarrow$ information about the current market situation and development

Advantages

- In-time-data-collection of actual offers
- Easy availability of a high number of rental property offers, e.g. via online platforms or the advertisement section in local newspapers
- Easy and cheap data-collection by the dint of
  - Web-crawling-algorithms
  - Web-APIs
  - Database of commercial service providers

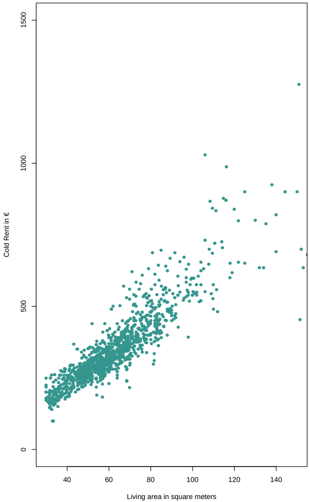
Disadvantages

- Rental offers prices may differ from the real price by re-letting.
- Data are biased by the interest of supplier.
- The researcher has no bearing on the characteristics covered in the rental property offer.
- Very good housing supplies are likely to be re-hired without a publication on an online platform.

METHODOLOGY

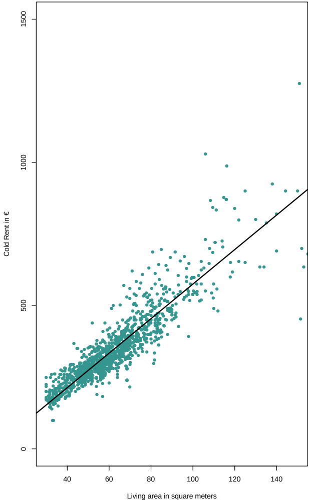
HEDONIC PRICE REGRESSION

- Background: The rental price of a property is determined by the individual characteristics of the flat and the characteristics of the surrounding neighborhood.
- Method of hedonic modeling: Estimation of the prices of the individual characteristic by multiple regression-models
- Enables the estimation of the conditional expectation by a special constellation of characteristics.
- It will be possible to estimate the expected price of a hypothetical average flat and to compare this price depending on the location of the flat or the date of the offer.



© Department of Statistics, Landeshauptstadt Magdeburg
Data: Empirica−Systeme

- Background: The rental price of a property is determined by the individual characteristics of the flat and the characteristics of the surrounding neighborhood.
- Method of hedonic modeling: Estimation of the prices of the individual characteristic by multiple regression-models
- Enables the estimation of the conditional expectation by a special constellation of characteristics.
- It will be possible to estimate the expected price of a hypothetical average flat and to compare this price depending on the location of the flat or the date of the offer.



© Department of Statistics, Landeshauptstadt Magdeburg
Data: Empirica−Systeme

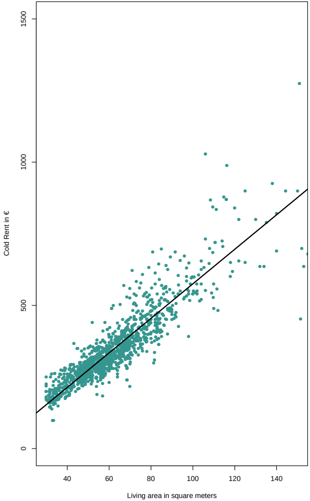
- The quantile regression enables the analysis of the conditional quantile to the quantile value $\tau \in (0,1)$ of a dependent variable in conjunction with a set of explanatory variables (Koenker and Basset (1978)).

Quantile Regression Model for $\tau \in (0,1)$

$$\begin{aligned} Y_i \quad &= \beta_0(\tau) + \beta_1(\tau)X_{1i} + \ldots + \beta_{(p-1)}(\tau)X_{(p-1)i} + U_i(\tau) \\ &= X_i^T \boldsymbol{\beta}(\tau) + U_i(\tau) \qquad i = 1, \ldots, n \end{aligned}$$

Assumption: $F_{U_i(\tau)}(0) = \tau \Rightarrow Q_Y(\tau | X_i = x_i) = x_i^T \boldsymbol{\beta}(\tau)$, $U_i(\tau)$ for $i = 1, \ldots, n$ are independent.

Y_i		...dependent variable
$X_i =$	$(1, X_{1i}, \ldots, X_{(p-1)i})$	...independent variables
$\boldsymbol{\beta}(\tau) =$	$(\beta_0(\tau), \ldots, \beta_{(p-1)}(\tau))$	...parameters depending on τ
$U_i(\tau)$		...perturbation depending on τ



© Department of Statistics, Landeshauptstadt Magdeburg
Data: Empirica−Systeme

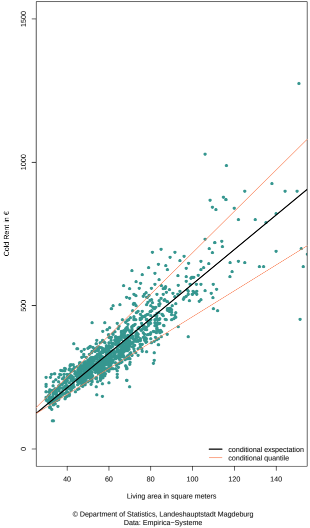
- The quantile regression enables the analysis of the conditional quantile to the quantile value $\tau \in (0,1)$ of a dependent variable in conjunction with a set of explanatory variables (Koenker and Basset (1978)).

Quantile Regression Model for $\tau \in (0,1)$

$$\begin{aligned} Y_i &= \beta_0(\tau) + \beta_1(\tau)X_{1i} + \ldots + \beta_{(p-1)}(\tau)X_{(p-1)i} + U_i(\tau) \\ &= X_i^T \boldsymbol{\beta}(\tau) + U_i(\tau) \qquad i = 1, \ldots, n \end{aligned}$$

Assumption: $F_{U_i(\tau)}(0) = \tau \Rightarrow Q_Y(\tau | X_i = x_i) = x_i^T \boldsymbol{\beta}(\tau)$, $U_i(\tau)$ for $i = 1, \ldots, n$ are independent.

Y_i			. . .dependent variable
X_i	$=$	$(1, X_{1i}, \ldots, X_{(p-1)i})$	. . .independent variables
$\boldsymbol{\beta}(\tau)$	$=$	$(\beta_0(\tau), \ldots, \beta_{(p-1)}(\tau))$	. . .parameters depending on τ
$U_i(\tau)$			. . .perturbation depending on τ



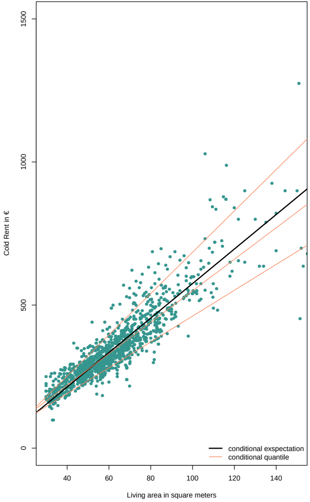
- The quantile regression enables the analysis of the conditional quantile to the quantile value $\tau \in (0,1)$ of a dependent variable in conjunction with a set of explanatory variables (Koenker and Basset (1978)).

Quantile Regression Model for $\tau \in (0,1)$

$$\begin{aligned} Y_i &= \beta_0(\tau) + \beta_1(\tau) X_{1i} + \ldots + \beta_{(p-1)}(\tau) X_{(p-1)i} + U_i(\tau) \\ &= X_i^T \boldsymbol{\beta}(\tau) + U_i(\tau) \qquad i = 1, \ldots, n \end{aligned}$$

Assumption: $F_{U_i(\tau)}(0) = \tau \Rightarrow Q_Y(\tau | X_i = x_i) = x_i^T \boldsymbol{\beta}(\tau)$, $U_i(\tau)$ for $i = 1, \ldots, n$ are independent.

Y_i		. . .dependent variable
$X_i =$	$(1, X_{1i}, \ldots, X_{(p-1)i})$	. . .independent variables
$\boldsymbol{\beta}(\tau) =$	$(\beta_0(\tau), \ldots, \beta_{(p-1)}(\tau))$	. . .parameters depending on τ
$U_i(\tau)$		. . .perturbation depending on τ



© Department of Statistics, Landeshauptstadt Magdeburg
Data: Empirica−Systeme

PRINCIPAL OF THE GEOGRAPHICALLY WEIGHTED REGRESSION

- Toblers first law of geographic: "Everything is related to everything else, but near things are more related than distant things." (Tobler (1970))
- The principal of the geographically weighted regression (GWR) is based on the ideas of Fotheringham et al. (2002). The geografical weighted extension of QR-Modell is based on the description from Chen et al. (2014) and McMillen (2013).
- The basic idea of the geographical weighting is a local estimation of the model at special target-points which are denoted by their geographical position $\mathbf{s}_i = (lat_i, lon_i)$.

Model of the Geografically Weighted Quantile Regression (GWQR)

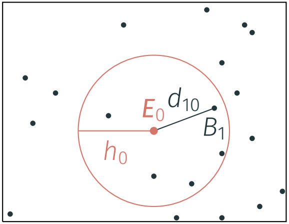
$$Y_i = X_i^T \boldsymbol{\beta}(\tau; \mathbf{s}_i) + U_i(\tau; \mathbf{s}_i), \quad i = 1, \dots, n$$

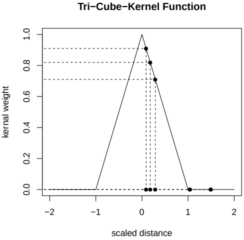
Assumption: $F_{U_i(\tau;\mathbf{s}_i)}(0) = \tau \Rightarrow Q_Y(\tau; \mathbf{s}_i | X_i = x_i) = x_i^T \boldsymbol{\beta}(\tau; \mathbf{s}_i)$, $U_i(\tau; \mathbf{s}_i)$ is independent.

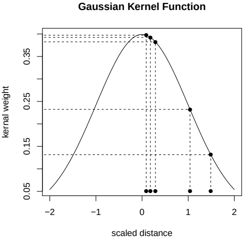
Y_i	... dependent variable	X_i	$= (1, X_{1i}, \dots, X_{(p-1)i})$	... independent variables
$U_i(\tau, \mathbf{s}_i)$	... perturbation depending on τ at location $\mathbf{s}_i$	$\boldsymbol{\beta}(\tau; \mathbf{s}_i)$	$= (\beta_0(\tau, \mathbf{s}_i), \dots, \beta_{(p-1)}(\tau; \mathbf{s}_i))$	... parameters depending on τ at location $\mathbf{s}_i$

Basic Principle

- Nearby observations will be integrated with a higher weight.
- The weights are denoted by a kernel function $K(\frac{d_{i0}}{h})_{|i=1,...,n}$ of the scaled distance d_{i0} of the observation i to the target-point by the bandwidth h.
- Fixed kernel weighting routine
- Adaptive kernel weighting routine







EMPIRICAL APPLICATION

Data

- Rental offers from the *empirica systema AG* database
- Published between January 1, 2012 and the March 31, 2016
- Subset of the row data by
  - geocoded offers with a location precision fewer than 10 meters
  - accommodations with a living area of 30 to 250 squares meters
- 18442 offers are included into the research

Methodical Specification

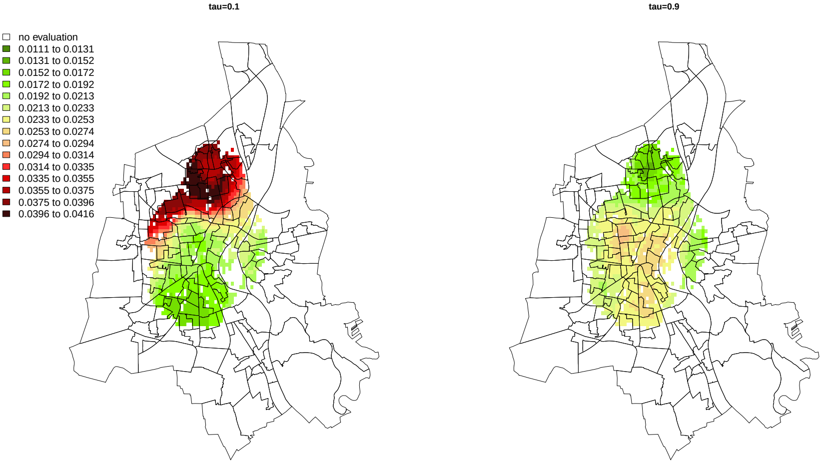
- Semi-log model; dependent variable: log of price per square meter
- First step: Estimation a global global OLS-model and global QR-models by the quantile values $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$
- Seccond step: Estimation the local GWR-models and GWQR-models by the quantile values $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$
  - Spatial weighting by a gaussian kernel-function and an adaptive bandwith
    - target-points: location of each observation
    - mapping: target points are the center points of a grid over the town area which covered with buildings (adaptive bandwidth have to be under 1 km)

SEMI-LOG QR/GWQR-MODEL FOR $\tau = 0.5$

	GWQR mean	$\tau = 0.5$ min	1st quantile	median	3rd quantile	max	QR Std.	Sign.
Intercept	1.69	1.55	1.66	1.68	1.72	1.91	1.658	***
log(living area)	-0.03	-0.07	-0.03	-0.02	-0.02	0.004	-0.02	***
elevator	0.02	-0.05	-0.00	0.01	0.04	0.11	0.01	*
balcony/terrace	0.03	-0.02	0.02	0.03	0.04	0.05	0.02	***
parking	0.06	0.02	0.05	0.06	0.08	0.11	0.07	***
kitchen	0.04	-0.01	0.03	0.04	0.05	0.07	0.05	***
social housing	-0.04	-0.10	-0.04	-0.04	-0.03	-0.00	-0.03	***
good condition	-0.00	-0.03	-0.00	0.00	0.01	0.01	0.00	
bad condition	-0.08	-0.17	-0.11	-0.08	-0.05	0.02	-0.10	***
time	0.02	0.01	0.02	0.02	0.02	0.03	0.02	***
	GWR						OLS	
Intercept	1.63	1.49	1.58	1.61	1.65	1.90	1.67	***
log(living area)	-0.01	-0.07	-0.02	-0.01	-0.00	0.02	-0.00	
elevator	0.04	-0.01	0.02	0.03	0.06	0.11	0.03	***
balcony/terrace	0.03	-0.01	0.01	0.03	0.04	0.06	0.02	***
parking	0.06	0.03	0.05	0.06	0.07	0.11	0.06	***
kitchen	0.05	0.01	0.04	0.05	0.06	0.08	0.06	***
social housing	-0.06	-0.11	-0.07	-0.06	-0.05	-0.01	-0.05	***
good condition	0.01	-0.02	0.01	0.01	0.01	0.02	0.01	***
bad condition	-0.07	-0.14	-0.1	-0.07	-0.05	0.03	-0.08	***
time	0.02	0.02	0.02	0.02	0.02	0.03	-0.02	***

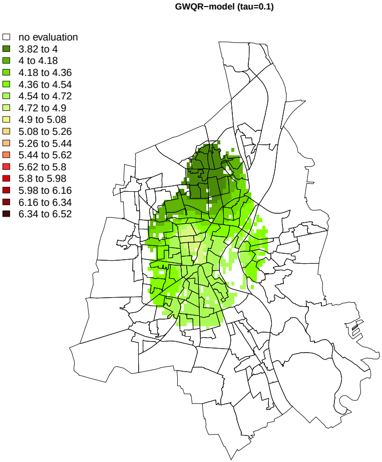
Significance levels [0,001]'***';(0.001,0.01]'**';(0.01;0.05]'*';(0.5,0.1]'°',(0.1,1]' '

SPATIAL HETEROGENEITY OF THE TIME EFFECT (MAGDEBURG)



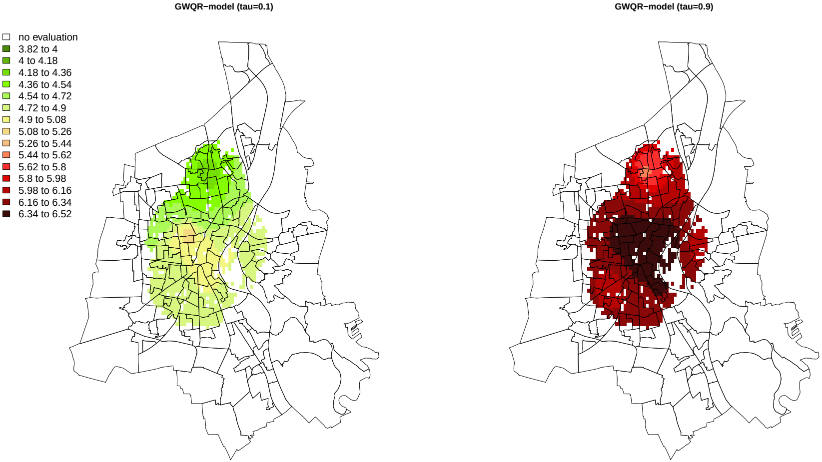
Data: Empirica−Systeme

PREDICTION OF THE COLD RENT PER SQUARE METER (IN €)



Flat characteristics: 61.8 square meters, balcony, good condition effective July 1,2012

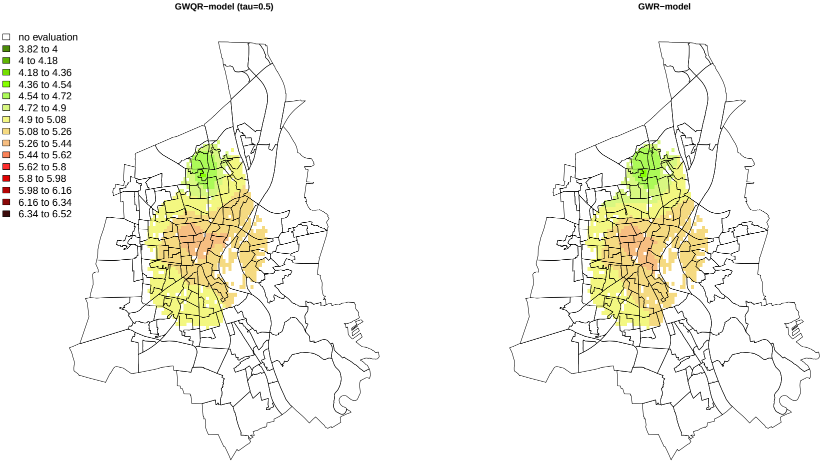
PREDICTION OF THE COLD RENT PER SQUARE METER (IN €)



Flat characteristics: 61.8 square meters, balcony, good condition effective July 1,2015

Data: Empirica−Systeme

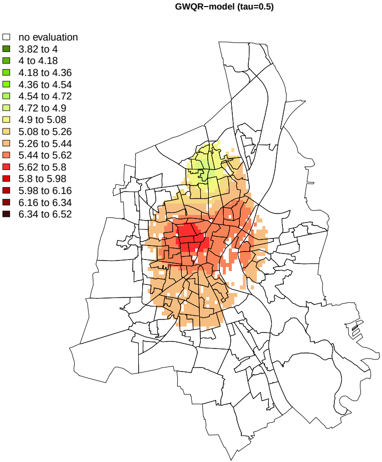
PREDICTION OF THE COLD RENT PER SQUARE METER (IN €)



Flat characteristics: 61.8 square meters, balcony, good condition effective July 1,2012

Data: Empirica−Systeme

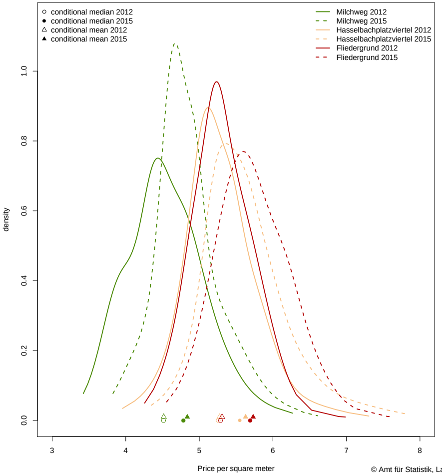
PREDICTION OF THE COLD RENT PER SQUARE METER (IN €)

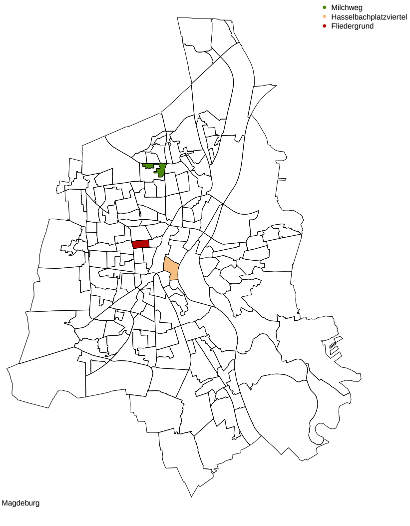


Flat characteristics: 61.8 square meters, balcony, good condition effective July 1,2015

Data: Empirica−Systeme

PREDICTION OF THE COLD RENT PER SQUARE METER (IN €)





CONCLUSION

CONCLUSION

- Geocoded online-data are a substantial data-source for the analysis of rental data.
  - Enables a lot of analysis in detail; high potential for geostatistical analysis
  - Data-restriction has to be kept in mind.
- Quantile regression allows a detailled view at the conditional distribution of the variable of interest.
  - Main advantages: rubustness aigainst outlier
  - Disadvantages: local estimation could lead to quantile-crossing
- The extension of the geographical weighting allows to consider the location of interest.
- Time trends could be analyzed considering the geographical position.
  - Indicator for the development in special districts (social segmentation or heterogenity).
- GWQR-Model based on local estimation.
  - computationally intensive
  - difficults in modell-selection

Thank you for your attention!

REFERENCES

- Chen, Vivian Y.; Deng, Wen-Shuenn; Yang, Tse-Chuan; Matthews, Stephan: Geographically Weighted Quantile Regression (GWQR): An Application to U.S. Mortality Data. In: Geographical Analysis 44(2) 2012, pp. 134-150
- Fotheringham, A. Stewart ; Brunsdon, Chris ; Charlton, Martin: Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley 2002
- Koenker, Roger: Quantile regression, Cambridge university press 2005
- Koenker, Roger; Basset Jr., Gilbert: Regression quantiles. In: Econometrica: journal of the Econometric Society, 1978, pp. 33-50
- McMillen, Daniel P.: Quantile regression for spatial data. Berlin, Springer, 2013
- Tobler, Waldo R.: A Computer Movie Simulating Urban Growth in the Detroit Region. In: Economic Geography, Vol. 46,1999, pp. 234-240

QUANTILE REGRESSION (QR)

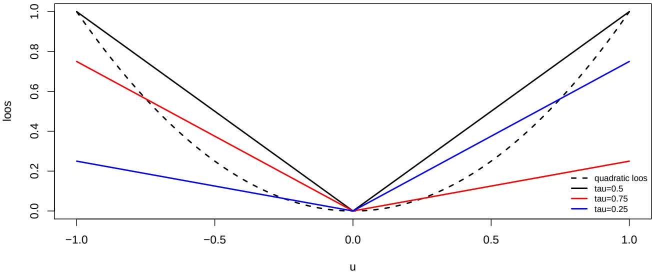
- In case of empirical analysis, the unknown parameter $\boldsymbol{\beta}(\tau)$ has to be estimated.
- Solution of the minimization problem:

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} R_\tau(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$$

- The loos-function ρ_τ is defined by:

$$\rho_\tau(u) = (u)(\tau - I(u < 0)) = \begin{cases} \tau \cdot u & , \quad \text{for} \quad u \leq 0 \\ u \cdot u & , \quad \text{for} \quad u < 0 \end{cases}$$

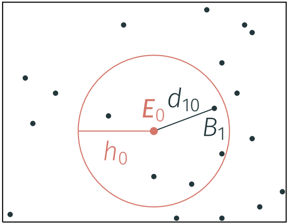
- $R_\tau(\boldsymbol{\beta})$ is not differentiable. $\rightarrow$ The minimization-problem has to be solved by methods of linear programming.

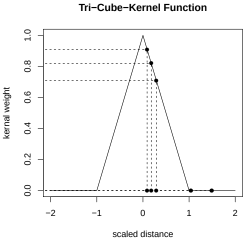


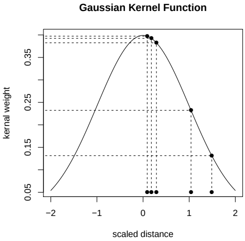
PRINCIPAL OF THE GEOGRAPHICALLY WEIGHTED REGRESSION

Basic Principle

- Nearby observations will be integrated with a higher weight.
- The weights are denoted by a kernel function $K(\frac{d_{i0}}{h})_{|i=1,\dots,n}$ of the scaled distance d_{i0} of the observation i to the target-point by the bandwidth h.
- Fixed kernel weighting routine
- Adaptive kernel weighting routine







Minimization Problem

$$\hat{\boldsymbol{\beta}}(\tau; \mathbf{s}_0) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} (\rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}(\mathbf{s}_0)) \cdot K(\frac{d_{i0}}{h}))$$