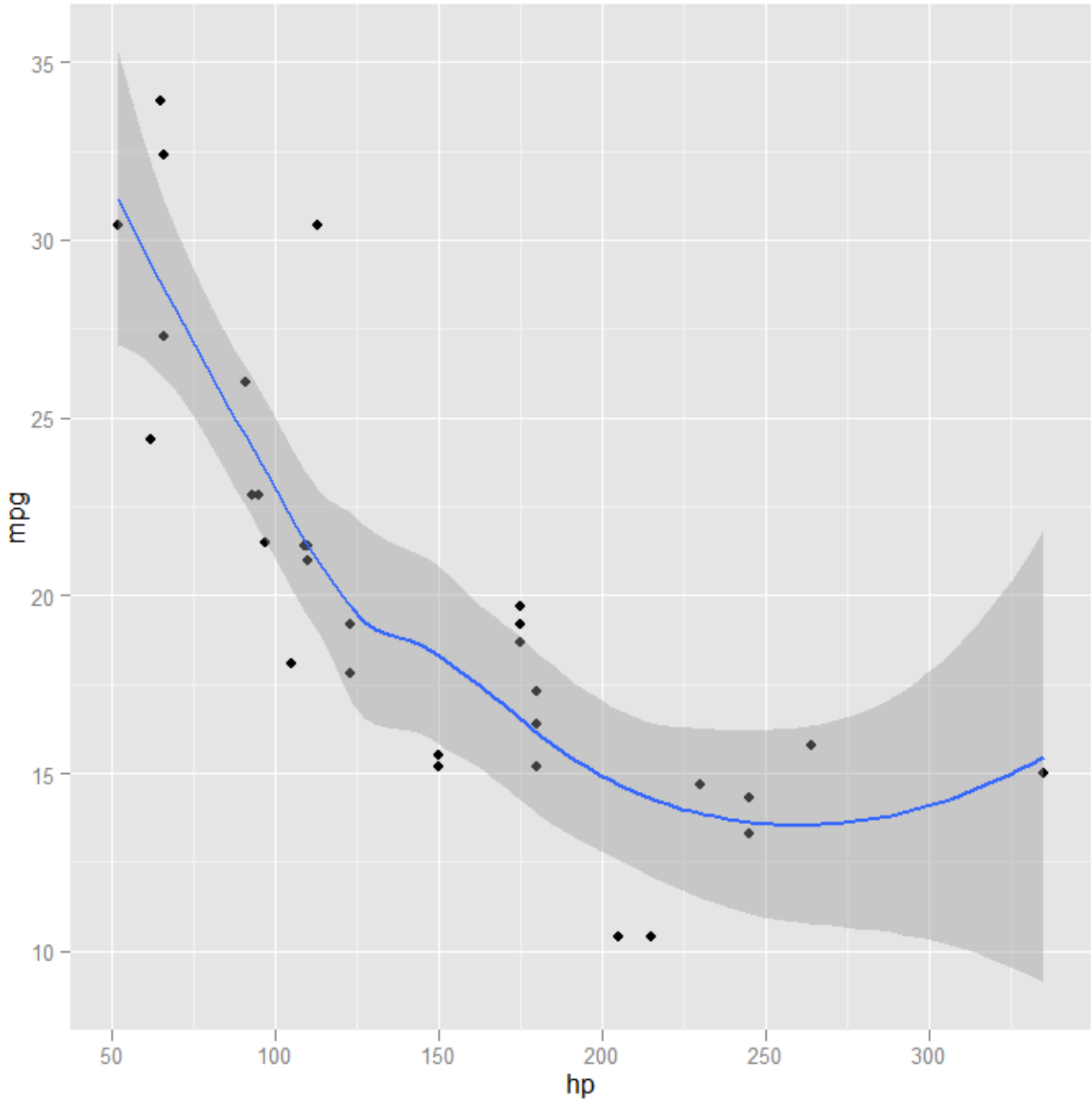
LOWESS LOESS

- **LOESS** and **LOWESS** (locally weighted scatterplot smoothing) are two strongly related non-parametric regression methods that combine multiple regression models in a k-nearest-neighbor-based meta-model.
- "LOESS" is a later generalization of LOWESS; although it is not a true initialism, it may be understood as standing for "**LOcal regrESSion**".

LOWESS LOESS

- LOESS and LOWESS thus build on "classical" methods, such as linear and nonlinear least squares regression. They address situations in which the classical procedures do not perform well or cannot be effectively applied without undue labor.
- LOESS combines much of the simplicity of linear least squares regression with the flexibility of nonlinear regression.
- It does this by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point.
- In fact, one of the chief attractions of this method is that the data analyst is not required to specify a global function of any form to fit a model to the data, only to fit segments of the data.

Loess



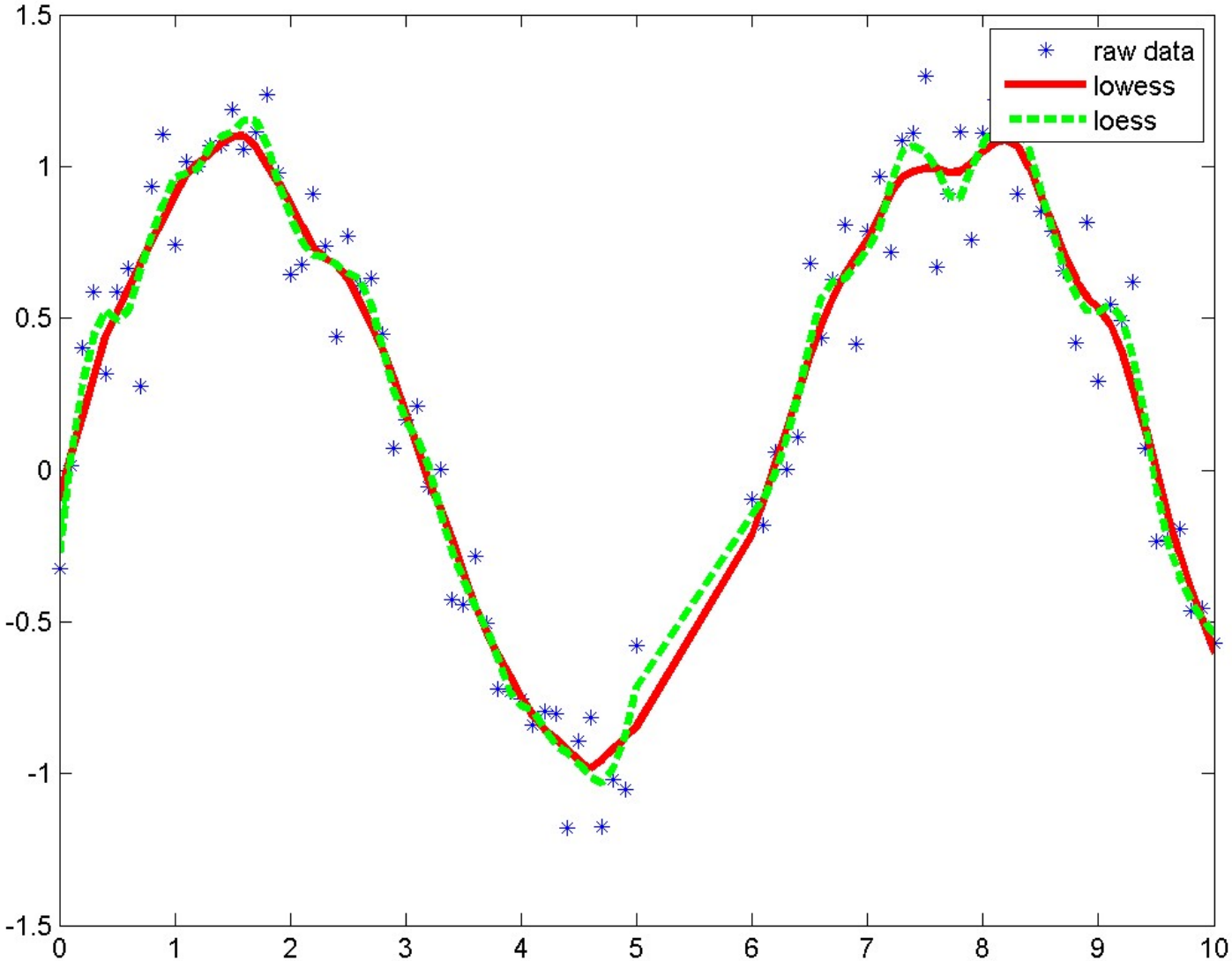


Figure:

5. Ridge Regression

- Ridge Regression is a technique used when the data suffers from **multicollinearity** (independent variables are highly correlated).
- In multicollinearity, even though the least squares estimates (OLS) are unbiased, their variances are large which deviates the observed value far from the true value.
- By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors.

Above, we saw the equation for linear regression. Remember? It can be represented as:

$$y = a + b * x$$

This equation also has an error term. The complete equation becomes:

```
y=a+b*x+e (error term),  [error term is the value needed to
=> y=a+y= a+ b1x1+ b2x2+....+e, for multiple independent va
```

- In a linear equation, prediction errors can be decomposed into two sub components.
- First is due to the biased and second is due to the variance.
- Prediction error can occur due to any one of these two or both components. Here, well discuss about the error caused due to variance.

Ridge Regression

Ridge regression solves the multicollinearity problem through shrinkage parameter λ (lambda). Look at the equation below.
Ridge
In this equation, we have two components. First one is least square term and other one is lambda of the summation of 2 (beta- square) where is the coefficient. This is added to least square term in order to shrink the parameter to have a very low variance.

Ridge Regression

Important Points:

- The assumptions of this regression is same as least squared regression except normality is not to be assumed
- It shrinks the value of coefficients but doesnt reaches zero, which suggests no feature selection feature
- This is a regularization method and uses l2 regularization.

Regularization

- In simple terms, regularization is tuning or selecting the preferred level of model complexity so your models are better at predicting (generalizing).
- If you don't do this your models may be too complex and overfit or too simple and underfit, either way giving poor predictions.

Regularization

- If you least-squares fit a complex model to a small set of training data you will probably overfit, this is the most common situation.
- The optimal complexity of the model depends on the sort of process you are modeling and the quality of the data, so there is no a-priori correct complexity of a model.

Regularization

To regularize you need 2 things:

- A way of testing how good your models are at prediction, for example using cross-validation or a set of validation data (you can't use the fitting error for this).
- A tuning parameter which lets you change the complexity or smoothness of the model, or a selection of models of differing complexity/smoothness.

Regularization

Basically you adjust the complexity parameter (or change the model) and find the value which gives the best model predictions. Note that the optimized regularization error will not be an accurate estimate of the overall prediction error so after regularization you will finally have to use an additional validation dataset or perform some additional statistical analysis to get an unbiased prediction error.
An alternative to using (cross-)validation testing is to use Bayesian Priors or other methods to penalize complexity or non-smoothness, but these require more statistical sophistication and knowledge of the problem and model features.

6. Lasso Regression

- Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients.
- In addition, it is capable of reducing the variability and improving the accuracy of linear regression models.
- Look at the equation below: LassoLasso regression differs from ridge regression in a way that it uses absolute values in the penalty function, instead of squares.

Lasso Regression

In statistics and statistical and machine learning, lasso (least absolute shrinkage and selection operator) (also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. It was introduced by Robert Tibshirani in 1996 based on Leo Breimans Nonnegative Garrote.[1][2]

Lasso Regression

lasso was originally formulated for least squares models and this simple case reveals a substantial amount about the behavior of the estimator, including its relationship to ridge regression and best subset selection and the connections between lasso coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates need not be unique if covariates are collinear.

Package 'glmnet'

April 12, 2015

Type Package

Title Lasso and Elastic-Net Regularized Generalized Linear Models

Version 2.0-2

Date 2015-4-11

Author Jerome Friedman, Trevor Hastie, Noah Simon, Rob Tibshirani

Maintainer Trevor Hastie <hastie@stanford.edu>

Depends Matrix (>= 1.0-6), utils, foreach

Suggests survival, knitr, lars

Description Extremely efficient procedures for fitting the entire lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models, Poisson regression and the Cox model. Two recent additions are the multiple-response Gaussian, and the grouped multinomial. The algorithm uses cyclical coordinate descent in a pathwise fashion, as described in the paper linked to via the URL below.

6. Lasso Regression

- This leads to penalizing (or equivalently constraining the sum of the absolute values of the estimates) values which causes some of the parameter estimates to turn out exactly zero.
- Larger the penalty applied, further the estimates get shrunk towards absolute zero.
- This results to variable selection out of given n variables.

6. Lasso Regression

Important Points:

- The assumptions of this regression is same as least squared regression except normality is not to be assumed
- It shrinks coefficients to zero (exactly zero), which certainly helps in feature selection
- This is a regularization method and uses l1 regularization
- If group of predictors are highly correlated, lasso picks only one of them and shrinks the others to zero

Elastic Net

- ElasticNet is hybrid of Lasso and Ridge Regression techniques.
- It is trained with L1 and L2 prior as regularizer. Elastic-net is useful when there are multiple features which are correlated.
- Lasso is likely to pick one of these at random, while elastic-net is likely to pick both.

7. ElasticNet Regression

- A practical advantage of trading-off between Lasso and Ridge is that, it allows Elastic-Net to inherit some of Ridges stability under rotation.

7. ElasticNet Regression

Important Points:
It encourages group effect in case of highly correlated variables
There are no limitations on the number of selected variables It can suffer with double shrinkage Beyond these 7 most commonly used regression techniques, you can also look at other models like Bayesian, Ecological and Robust regression.

Package 'elasticnet'

February 19, 2015

Version 1.1

Date 2012-06-25

Title Elastic-Net for Sparse Estimation and Sparse PCA

Author Hui Zou <hzou@stat.umn.edu> and Trevor Hastie <hastie@stanford.edu>

Maintainer Hui Zou <hzou@stat.umn.edu>

Depends R (>= 2.10), lars

Description This package provides functions for fitting the entire solution path of the Elastic-Net and also provides functions for estimating sparse Principal Components. The Lasso solution paths can be computed by the same function. First version: 2005-10.

License GPL (>= 2)

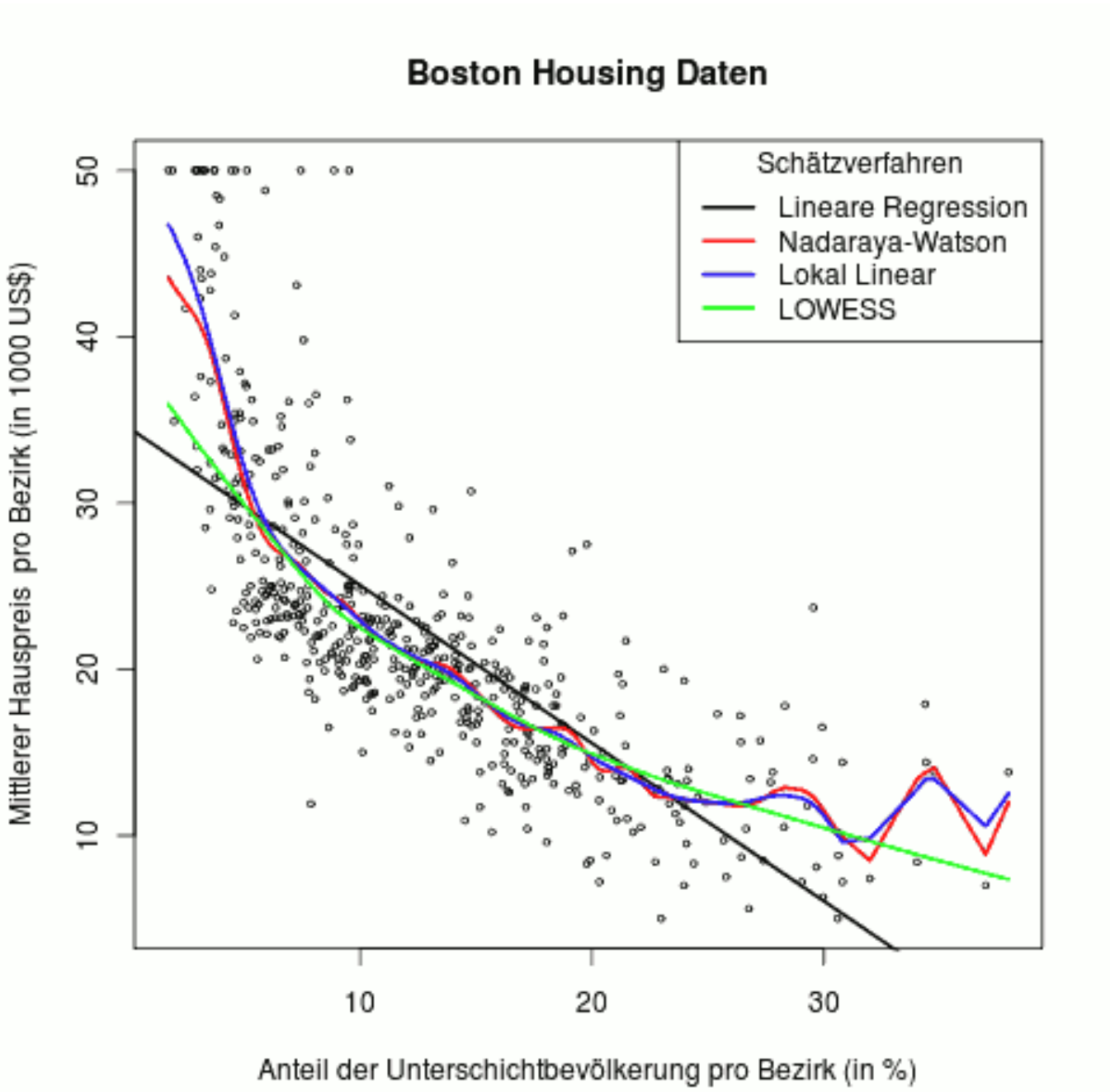
URL http://www.stat.umn.edu/~hzou

Repository CRAN

Date/Publication 2012-06-28 08:57:54

NeedsCompilation no

Boston Data Set



- Quantile regression is a type of regression analysis used in statistics and econometrics. Whereas the method of least squares results in estimates that approximate the conditional mean of the response variable given certain values of the predictor variables, quantile regression aims at estimating either the conditional median or other quantiles of the response variable.
- R offers several packages that implement quantile regression, most notably quantreg by Roger Koenker,but also gbm,and quantregForest.

Quantile Regression

Package 'quantreg'

August 31, 2015

Title Quantile Regression

Description Estimation and inference methods for models of conditional quantiles: Linear and nonlinear parametric and non-parametric (total variation penalized) models for conditional quantiles of a univariate response and several methods for handling censored survival data. Portfolio selection methods based on expected shortfall risk are also included.

Version 5.19

Maintainer Roger Koenker <rkoenker@illinois.edu>

Repository CRAN

Depends R (>= 2.6), stats, SparseM

Imports methods, graphics, Matrix, MatrixModels

Suggests tripack, akima, MASS, survival, rgl, logspline, nor1mix, Formula, zoo

License GPL (>= 2)

Quantile Regression

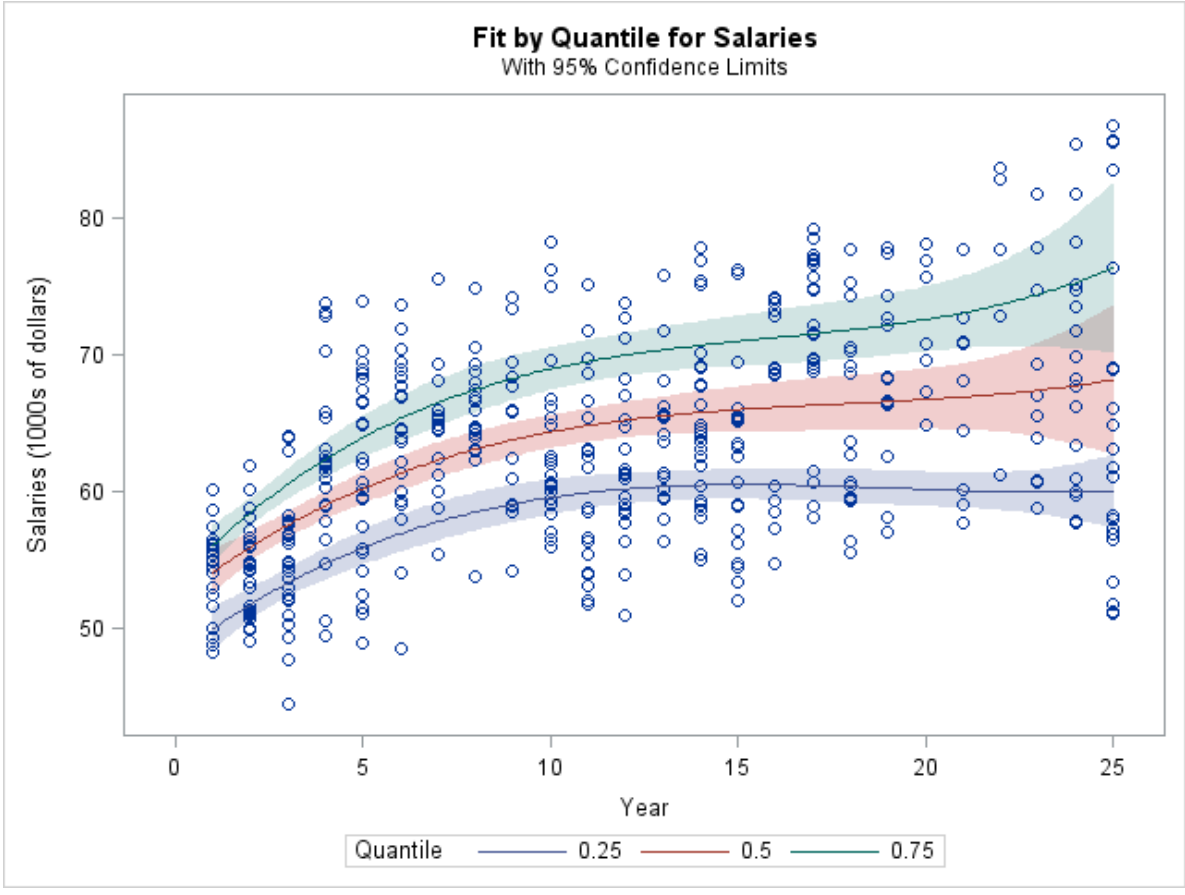


Figure:

Kernel Regression

Nadaraya-Watson Kernel Regression

Nadaraya 1964 and Watson 1964 proposed to estimate m as a locally weighted average, using a kernel as a weighting function.

The Nadaraya-Watson estimator is:

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)}$$

where K is a kernel with a bandwidth h. The fraction is a weighting term with sum 1.