ETC3250: Regression

Semester 1, 2019

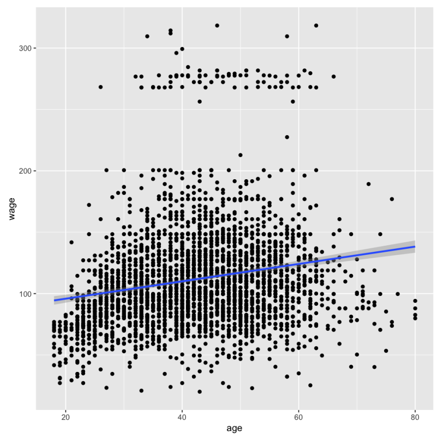
Professor Di Cook

Econometrics and Business Statistics
Monash University

Week 2 (b)

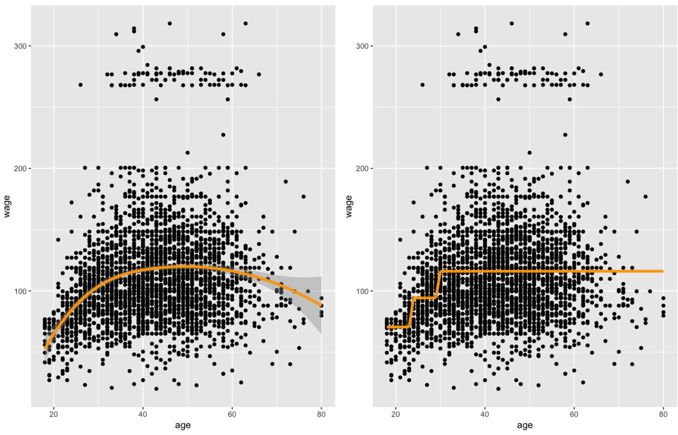
Outline

- Moving beyond linearity



Outline

- Moving beyond linearity
  - splines and GAMs



Outline

- Moving beyond linearity
  - splines and GAMs
  - and more

The truth is rarely linear, but often the linearity assumption is good enough.

When it's not ...

- polynomials,
- step functions,
- splines,
- local regression, and
- generalized additive models

offer a lot of flexibility, without losing the ease and interpretability of linear models.

Outline

- Moving beyond linearity
- polynomials
  - basis functions

Instead of fitting a linear model (in x), we fit the model

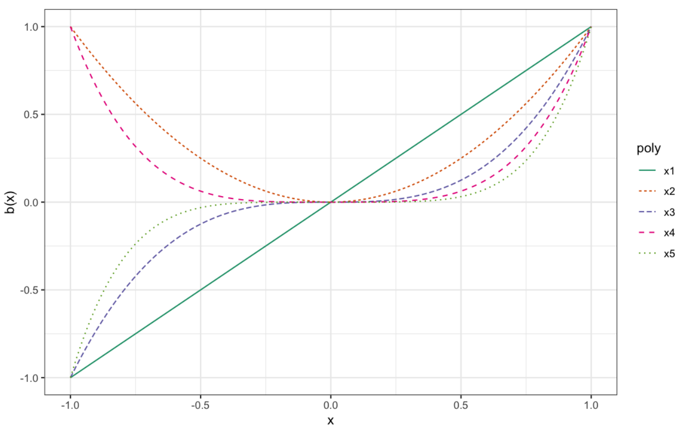
$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_K b_K(x_i) + e_i,$$

where $b_1(X), b_2(X), \ldots, b_K(X)$ are a family of functions or transformations that can be applied to a variable x, and $i = 1, \ldots, n$.

- Polynomial regression: $b_k(x_i) = x_i^k$
- Piecewise constant functions: $b_k(x_i) = I(c_k \le x_i \le c_{k+1})$

Outline

- Moving beyond linearity
- polynomials
  - basis functions



x1 $= x$, x2= x^2, x3= x^3, x4= x^4, x5= x^5

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots

Knots: $\kappa_1, \ldots, \kappa_K$.

A spline is a continuous function $f(x)$ consisting of polynomials between each consecutive pair of "knots" $x = \kappa_j$ and $x = \kappa_{j+1}$.

- Parameters constrained so that $f(x)$ is continuous.
- Further constraints imposed to give continuous derivatives.

Outline

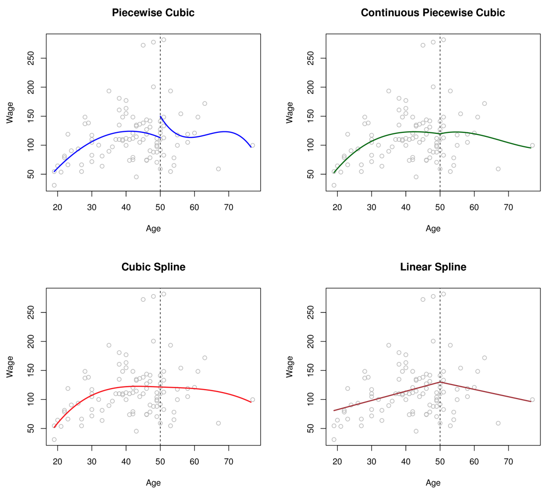
- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly

Piecewise cubic polynomial with a single knot at a point c:

$$\hat{y}_i = \left\{ \begin{array}{ll} \beta_{01} + \beta_{11} x_i + \beta_{21} x_i^2 + \beta_{31}^2 & if\, x_i < c \\ \beta_{02} + \beta_{12} x_i + \beta_{22} x_i^2 + \beta_{32}^2 & if\, x_i \geq c \end{array} \right\}$$

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly



(Chapter7/7.3.pdf)

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns

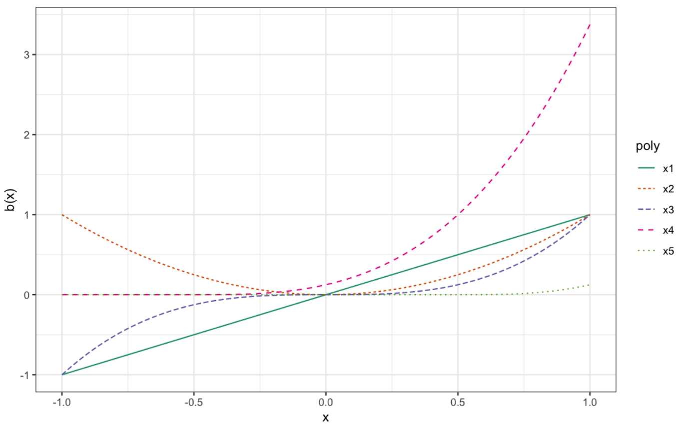
- Truncated power basis
- Predictors: x, ..., x^p, $(x-\kappa_1)_+^p$, ..., $(x-\kappa_K)_+^p$

Then the regression is piecewise order- p polynomials.

- $p-1$ continuous derivatives.
- Usually choose $p=1$ or $p=3$.
- $p+K+1$ degrees of freedom

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns



x1 $= x$, x2= x^2, x3= x^3, x4= $(x+0.5)_+^3$, x5= $(x-0.5)_+^3$

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural

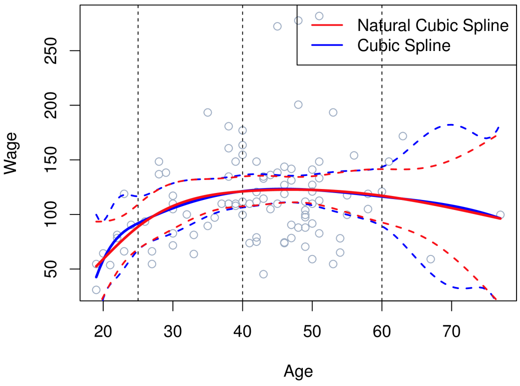
- Splines based on truncated power bases have high variance at the outer range of the predictors.
- Natural splines are similar, but have additional boundary constraints: the function is linear at the boundaries. This reduces the variance.

Degrees of freedom $\text{df} = K$.

Create predictors using `ns` function in R (automatically chooses knots given `df`).

Outline

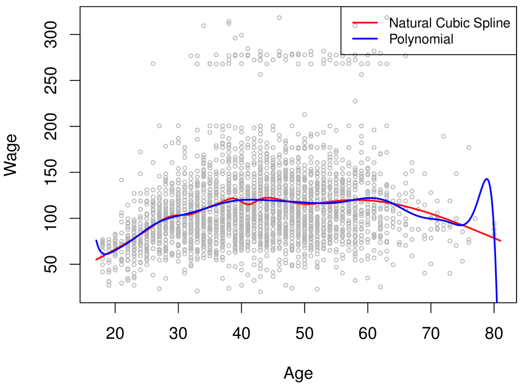
- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural



(Chapter7/7.4.pdf)

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural

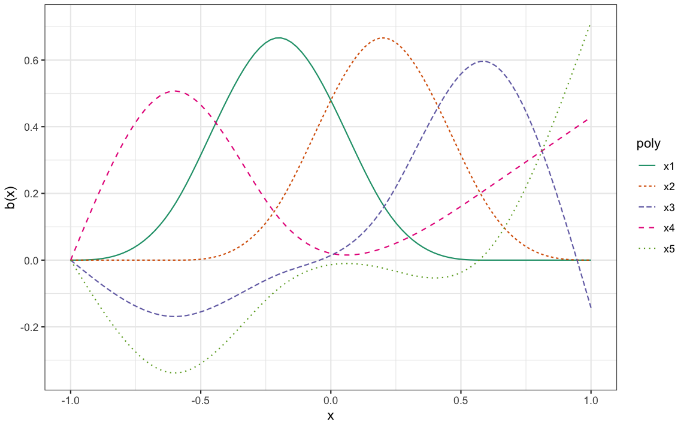


(Chapter7/7.7.pdf)

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural

Natural cubic splines



Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

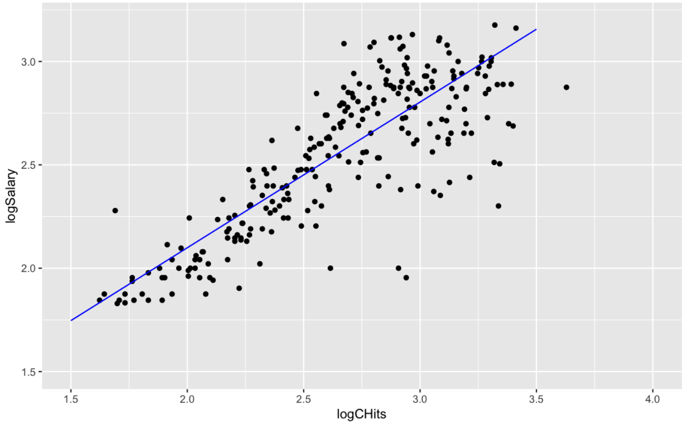
Knot placement

- Strategy 1: specify `df` (equivalently K) and let `ns` place them at appropriate quantiles of the observed x.
- Strategy 2: choose K and their locations.

Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

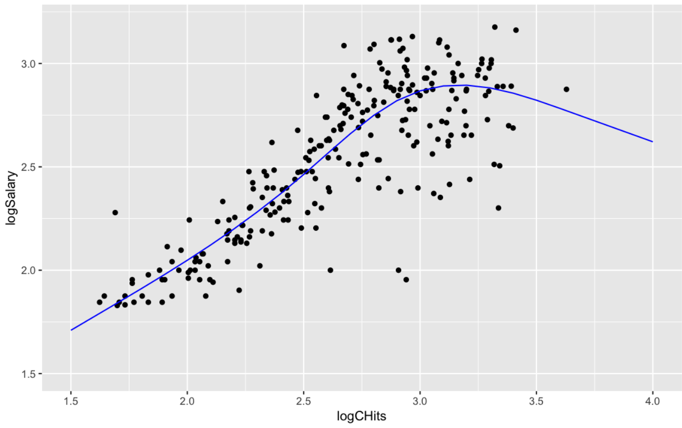
DF = 2 (linear fit)



Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

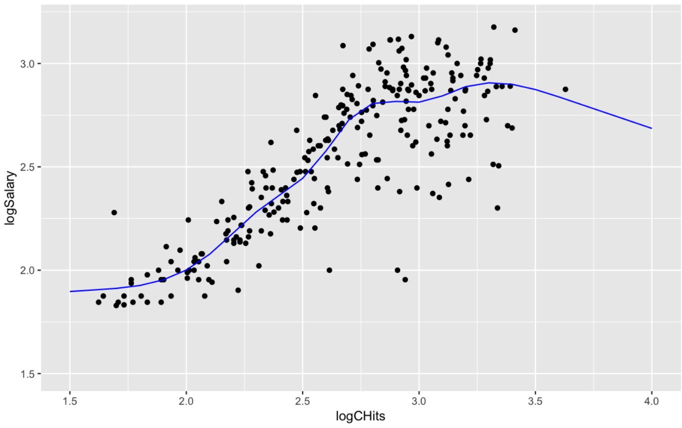
DF = 3



Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

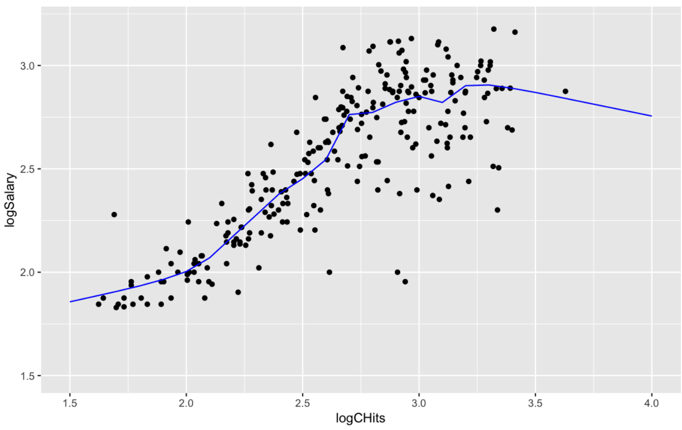
DF = 8



Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

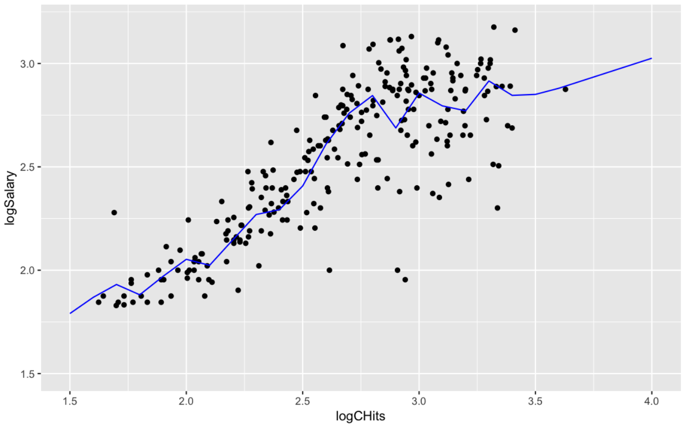
DF = 15



Outline

- Moving beyond linearity
- polynomials
- splines
  - knots
  - piecewise poly
  - basis fns
  - natural
  - knots

DF = 50



Outline

- Moving beyond linearity
- polynomials
- splines
- Generalised additive models (GAMs)
  - Curse of dimensionality

Why is it hard to fit models of the form

$$y = f(x_1, x_2, \dots, x_p) + e?$$

- Data is very sparse in high-dimensional space.
- Model assumes p-way interactions which are hard to estimate.

Outline

- Moving beyond linearity
- polynomials
- splines
- Generalised additive models (GAMs)
  - Curse of dimensionality
  - Additive functions

$$y_i = \beta_0 + f_1(x_{i,1}) + f_2(x_{i,2}) + \ldots + f_p(x_{p,1}) + e_i$$

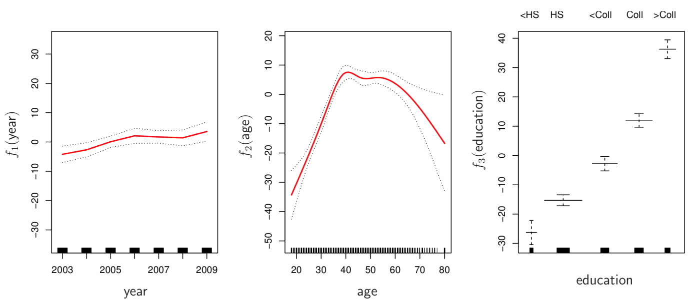
where each f is a smooth univariate function.

Allows for flexible nonlinearities in several variables, but retains the additive structure of linear models.

Outline

- Moving beyond linearity
- polynomials
- splines
- Generalised additive models (GAMs)
  - Curse of dimensionality
  - Additive functions

$$wage = \beta_0 + f_1(year) + f_2(age) + f_3(education) + \varepsilon$$



(Chapter7/7.11.pdf)

Outline

- Moving beyond linearity
- polynomials
- splines
- Generalised additive models (GAMs)
  - Curse of dimensionality
  - Additive functions
  - Generalisation

- Can fit a GAM simply using, e.g. natural splines:
- Coefficients not that interesting; fitted functions are.
- Use `plot.gam` from `gam` package.
- Can mix terms --- some linear, some nonlinear --- and use `anova()` to compare models.
- GAMs are additive, although low-order interactions can be included in a natural way using, e.g. bivariate smoothers or interactions of the form `ns(age,df=5):ns(year,df=5)`.

Outline

- Moving beyond linearity
- polynomials
- splines
- Generalised additive models (GAMs)
  - Curse of dimensionality
  - Additive functions
  - Generalisation
  - Interaction

- Additive models assume no interactions.
- Add bivariate smooths for two-way interactions.
- Graphically check for interactions using faceting.

👩‍💻 Made by a human with a computer

Slides at https://monba.dicook.org.

Code and data at https://github.com/dicook/Business_Analytics.

Created using R Markdown with flair by **xaringan**, and **kunoichi** (female ninja) style.