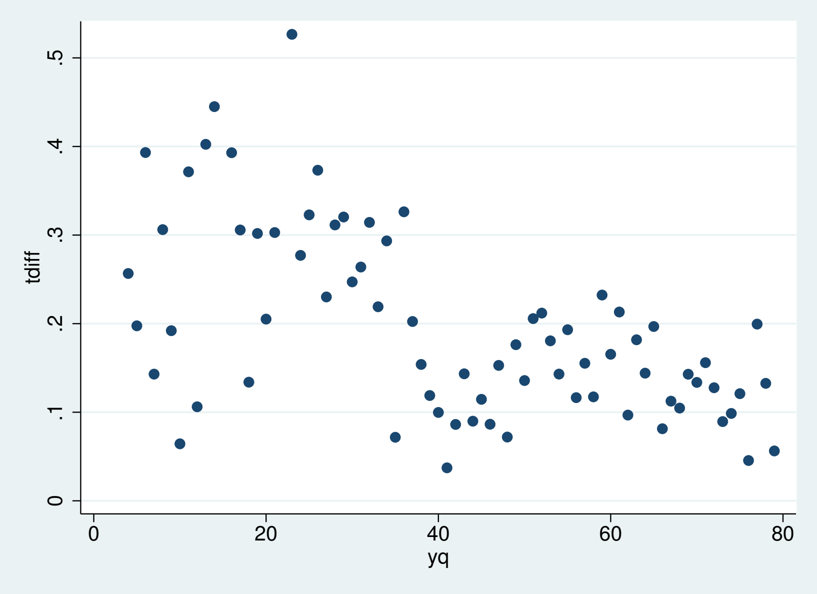
Smoothing Splines

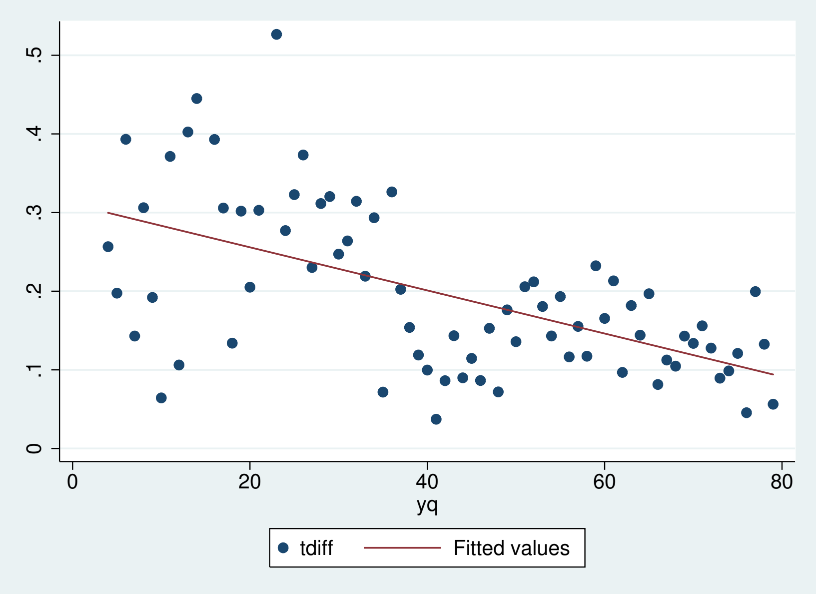
Florian Szücs

December 2010

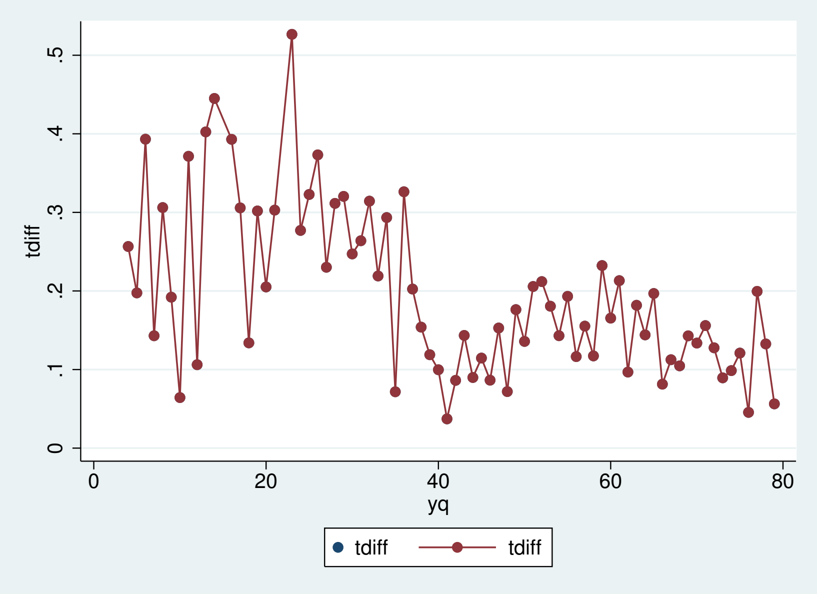
Motivation



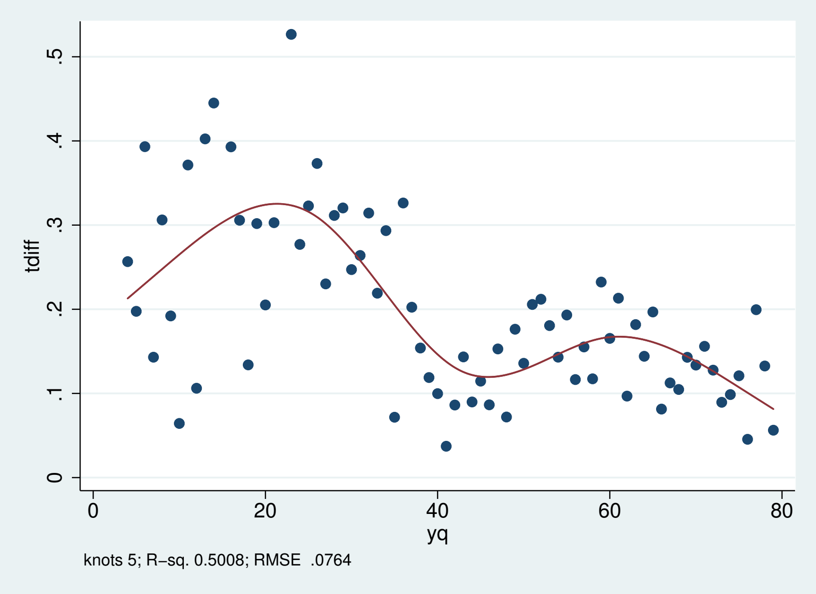
Motivation



Motivation



Motivation



Motivation

- 'roughness penalty approach' aims at
  - good data fit
  - minimal curve fluctuation

  $\rightarrow$ tradeoff
- importance of both goals can be weighted
- basic approach and algorithm developed in Reinsch (1967)

Formulation

- Quantification of roughness
  $\rightarrow \int_{x_1}^{x_n} g''(x)^2 dx$
  where g is any twice-differentiable function on $[x_1, x_n]$
- Deviation from datapoints
  $\rightarrow \sum_{i=1}^{n} (y_i - g(x_i))^2$

Formulation

- solution can be obtained by minimizing:

$$L = \int_{x_1}^{x_n} g''(x)^2 dx + \alpha\left(\sum_{i=1}^{n}(y_i - g(x_i))^2\right) \tag{1}$$

- solve for $\hat{g} = arg\ min\ L$

Formulation

solution of (1) imposes that

- $f''''(x) = 0$
- f, f', f'' continuous
- f''' discontinuous

$\rightarrow$ 'composition of cubic parabolas'

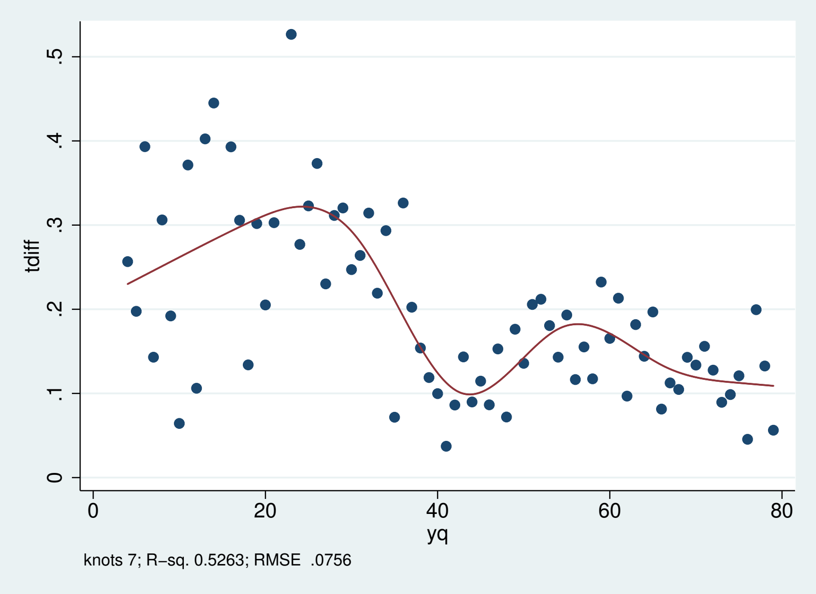
Implementation

- Reinsch implements algorithm for ALGOL
- implementations for Stata, R, Matlab, Mathematica, gretl etc.

Choice of α

- personal choice
- choice by algorithm

Large value of α



Small value of α