

# Over-dispersion and Outlier Handling

Iain Currie

Heriot-Watt University

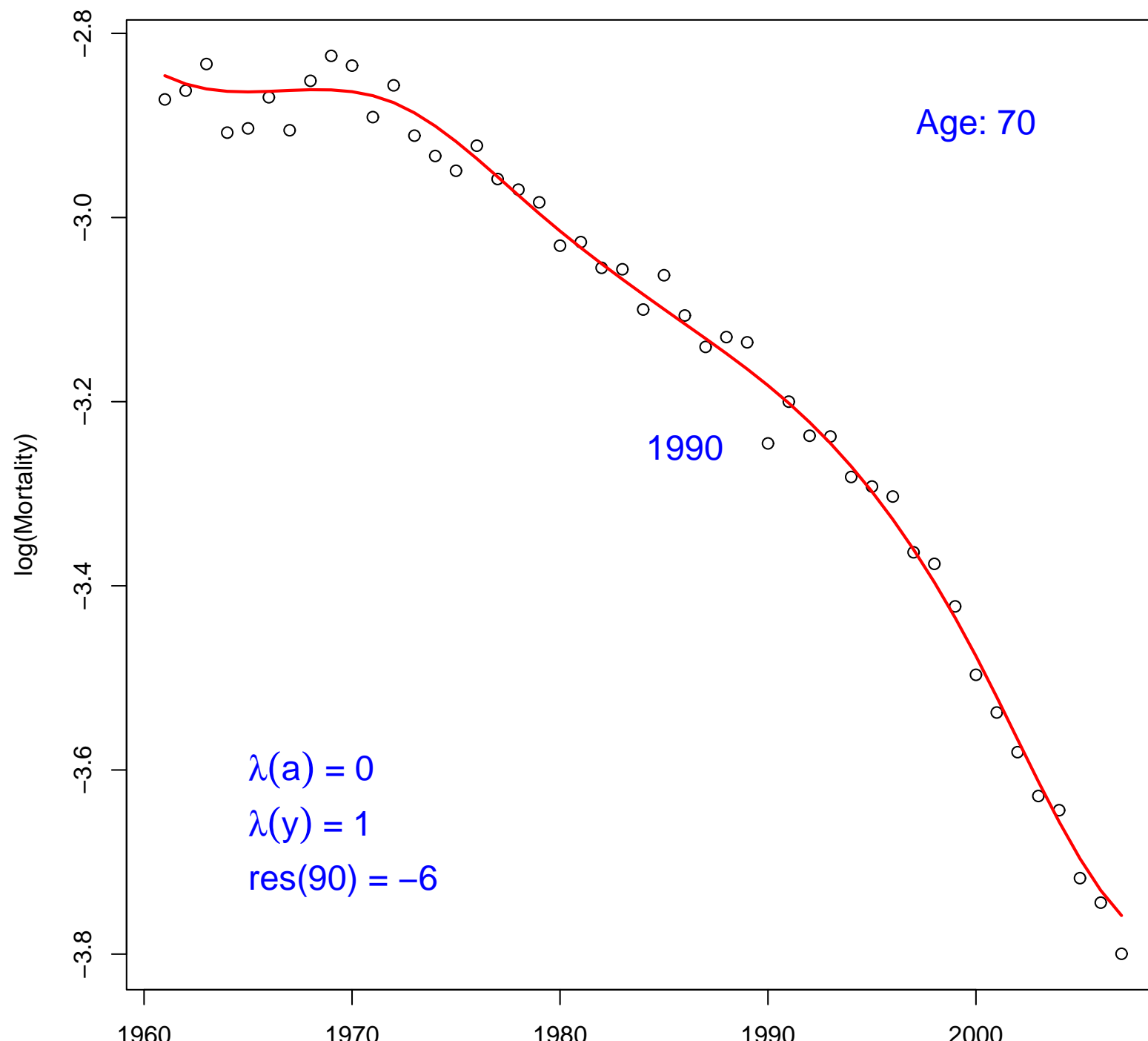
&



London  
18th May 2010

- **Data:** ONS, age 40-90, year 1961-2007.
- **Model:** 2d AP P-splines.

ONS data: 2-d AP model



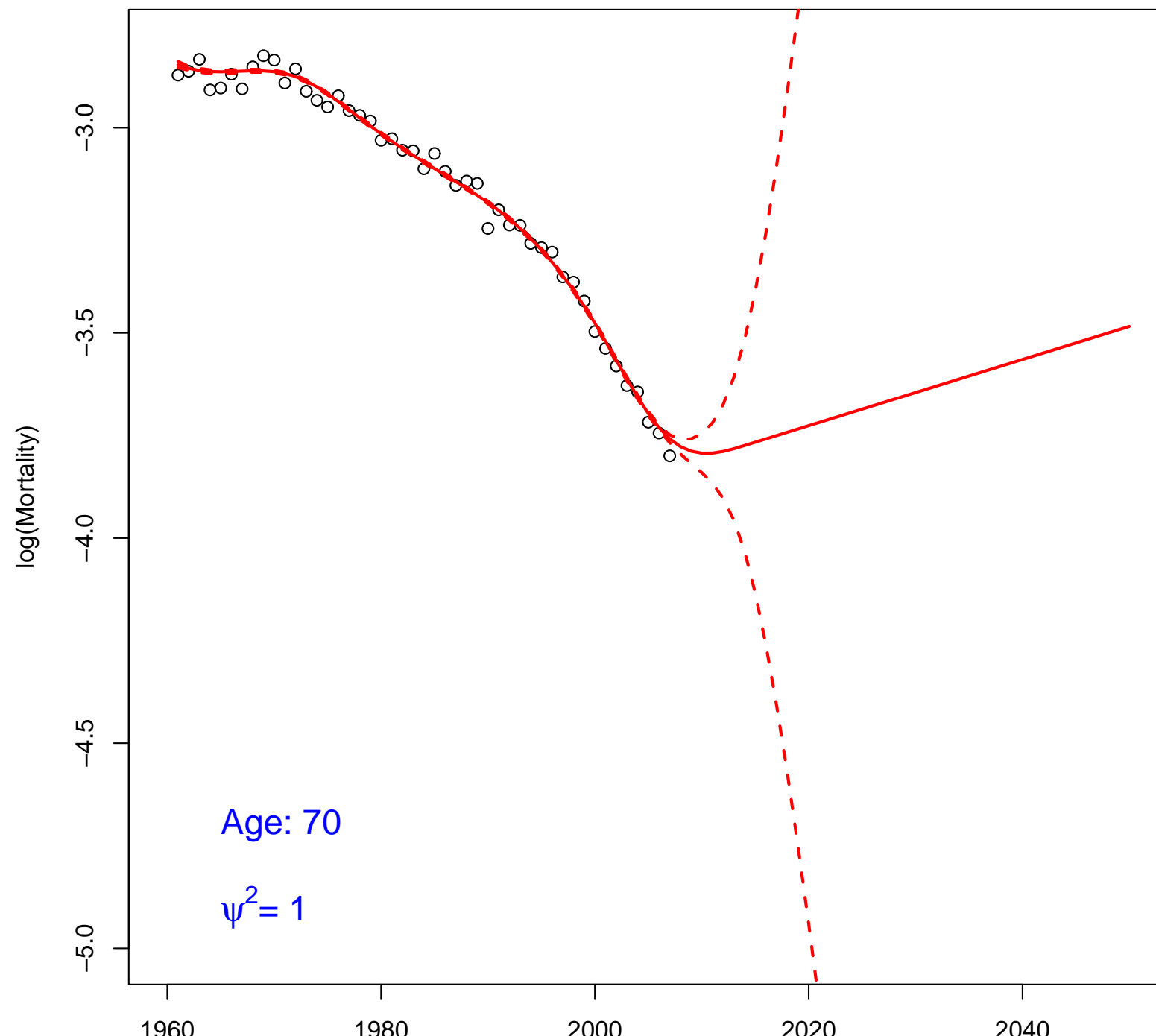
## Problems with 2d AP

- Small smoothing parameters ( $\lambda_a = 0$ ,  $\lambda_y = 1$ )  $\Rightarrow$  Volatile forecasts.
- Large residuals. Example, at age 70,
  - $r_{90} = -6$ ,
  - 8 of 47 residuals  $> 3$ .

**Overdispersion** arises from

- heterogeneity
- duplicates

Forecast to 2050: ONS data



## Modelling Overdispersion

Let  $D$  be the number of claims/deaths at each age and year of death,  $E$  be the corresponding central exposure and  $\mu$  the force of mortality.

- Classical solution:  $D \sim \mathcal{P}(E\mu)$ . Then

$$E(D) = Var(D) = E\mu$$

- Quasi-likelihood:  $D \sim$  quasi-Poisson with mean  $E\mu$ . Then

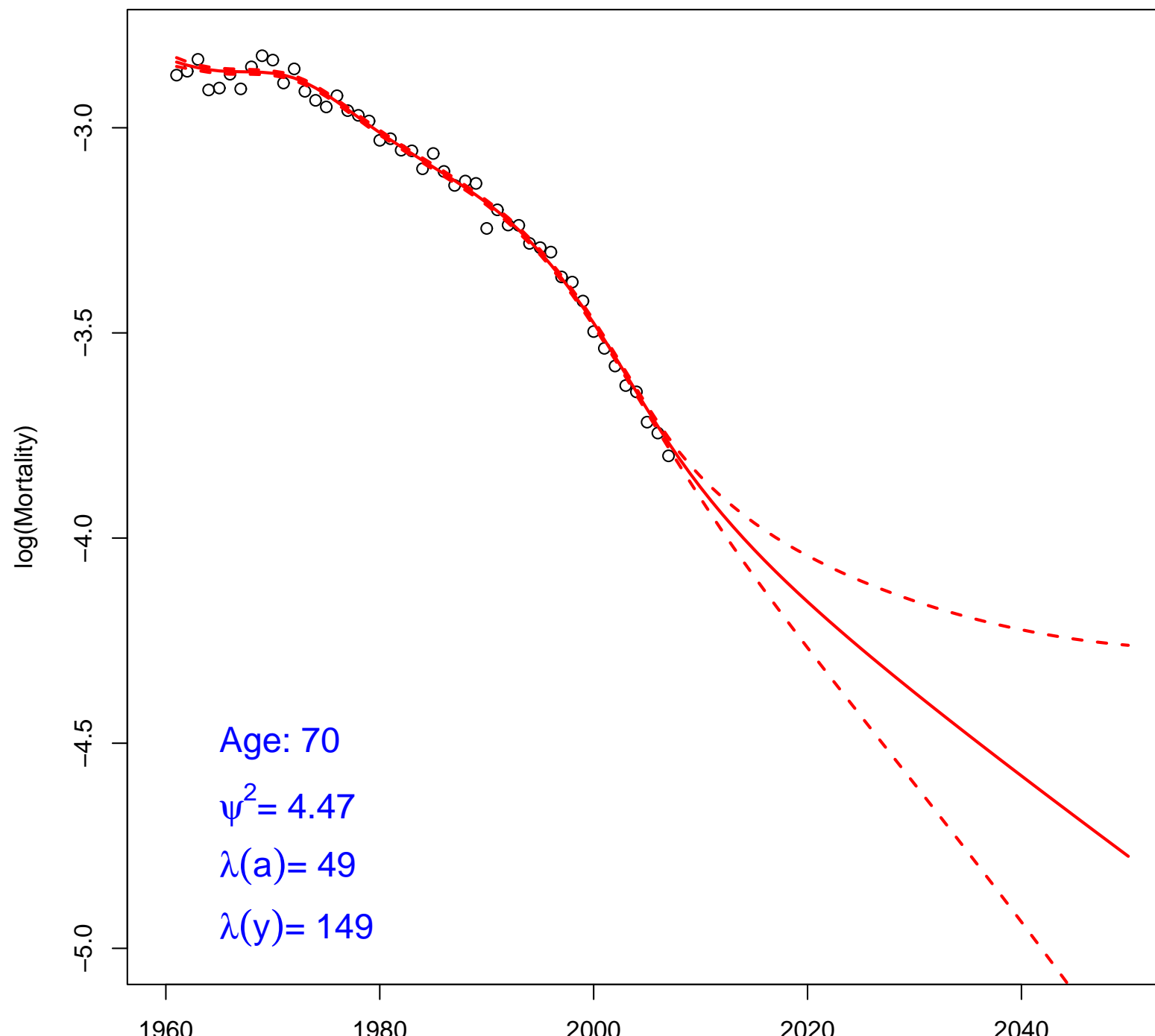
$$E(D) = E\mu, \quad Var(D) = \Psi^2 E\mu$$

where  $\Psi^2$  is the overdispersion parameter.

## Refit with 2d AP and overdispersion

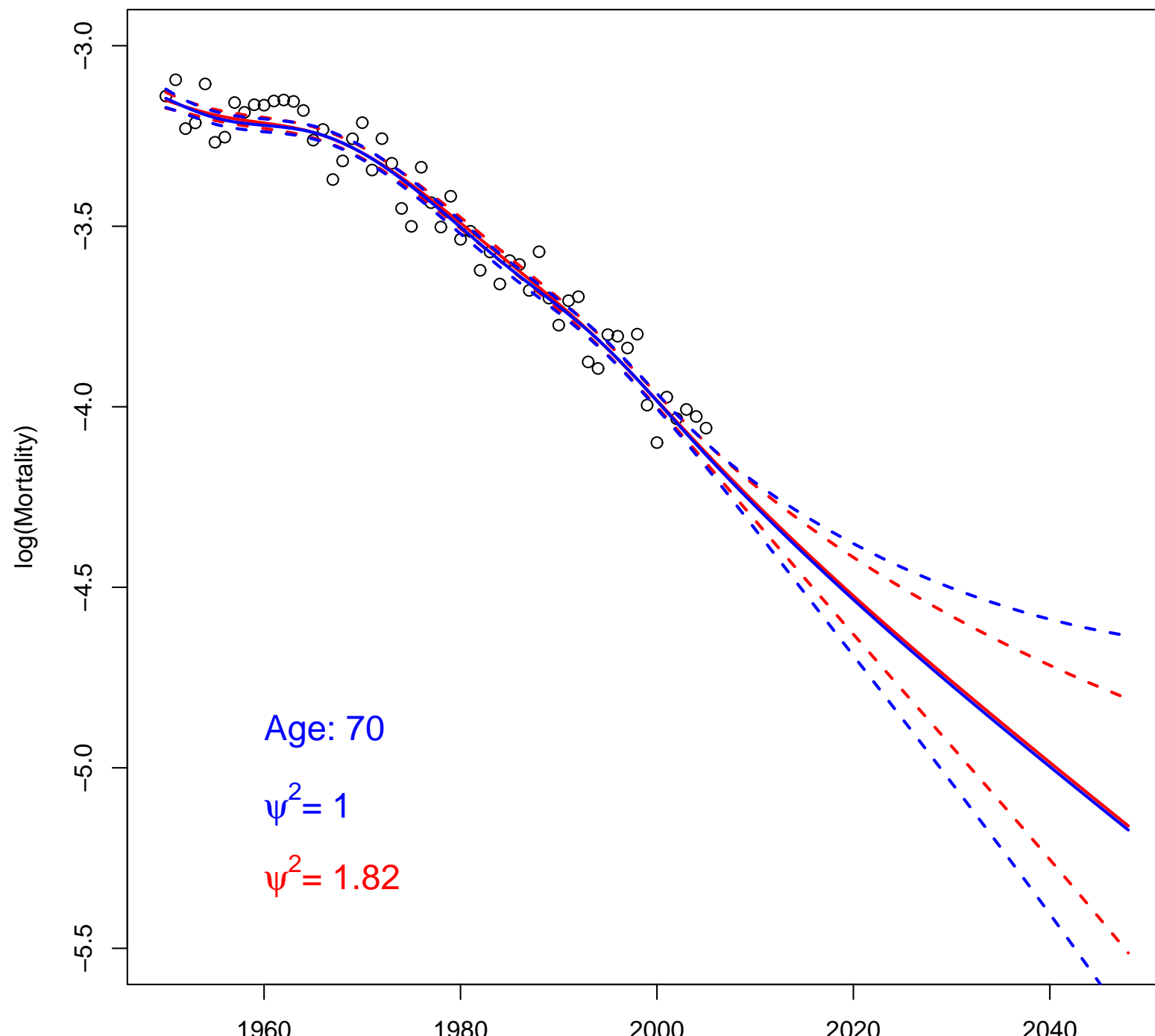
- $\Psi^2 = 4.47$
- Smoothing parameters ( $\lambda_a = 49$ ,  $\lambda_y = 149$ )  $\Rightarrow$  Stable forecasts with narrower CIs.
- Smaller residuals. Example, at age 70,
  - $r_{90} = -2.85$ ,
  - 2 of 47 residuals  $> 2$ .

Forecast to 2050: ONS data





Forecast to 2048: CMI data



## Outlier handling

- Overdispersion reduces the size of residuals but what if a **small number** of large residuals remain?
- We can weight such rogue points out of the data.
- Should be used with care - the model may not fit!
- Will further stabilize forecasts.

## References

- Longevity blogs: See *Over-dispersion (Dec 9, 2009)* and *Over-dispersion (reprise for actuaries) (Jan 3, 2010)* in [www.longevity.co.uk/site/informationmatrix/](http://www.longevity.co.uk/site/informationmatrix/)