Whittaker Graduation and the Hodrick-Prescott Filter

Enrique de Alba[1] Sergio Gómez[1]

[1]Instituto Nacional de Estadística y Geografía (INEGI)
México

ASTIN Mexico Colloquium

October 2012

Introduction

Both the Whittaker-Henderson and the H-P methods require the specification of a particular constant, the smoothing constant, usually identified as λ.

Introduction

Both the Whittaker-Henderson and the H-P methods require the specification of a particular constant, the smoothing constant, usually identified as λ.

- The specification of λ is arbitrary

Introduction

Both the Whittaker-Henderson and the H-P methods require the specification of a particular constant, the smoothing constant, usually identified as λ.

- The specification of λ is arbitrary
- The guidelines for choosing tend to be ad-hoc

Introduction

Both the Whittaker-Henderson and the H-P methods require the specification of a particular constant, the smoothing constant, usually identified as λ.

- The specification of λ is arbitrary
- The guidelines for choosing tend to be ad-hoc
- **We present a Bayesian approach using MCMC methods**

The Hodrick-Prescott Method

The Hodrick-Prescott Method

- Procedure used for Economic Time Series

The Hodrick-Prescott Method

- Procedure used for Economic Time Series
- This approach is usually applied to series that have been previously seasonally adjusted

The Hodrick-Prescott Method

- Procedure used for Economic Time Series
- This approach is usually applied to series that have been previously seasonally adjusted
- The series consist in a cyclical and a growth component

Hodrick-Prescott model

$$y_i = g_i + c_i \qquad i = 1 \dots T \tag{1}$$

Hodrick-Prescott model

$$y_i = g_i + c_i \qquad i = 1 \dots T \tag{1}$$

- $\sum_{i=3}^{N} (\Delta^2 g_i)^2$, the sum of the squares of the second difference is used as a smoothness measure for the g_i path

Hodrick-Prescott model

$$y_i = g_i + c_i \qquad i = 1 \dots T \tag{1}$$

- $\sum_{i=3}^{N} (\Delta^2 g_i)^2$, the sum of the squares of the second difference is used as a smoothness measure for the g_i path
- The c_i components are deviations from g_i

Hodrick-Prescott model

$$y_i = g_i + c_i \qquad i = 1 \dots T \tag{1}$$

- $\sum_{i=3}^{N} (\Delta^2 g_i)^2$, the sum of the squares of the second difference is used as a smoothness measure for the g_i path
- The c_i components are deviations from g_i
- Over long time periods, the average of the c_i components is expected to be near zero

Optimization problem

$$\min_{\{g_i\}} \left\{ \sum_{i=1}^{N} (y_i - g_i)^2 + \lambda \sum_{i=3}^{N} \left[(g_i - g_{i-1}) - (g_{i-1} - g_{i-2}) \right]^2 \right\} \qquad (2)$$

λ is an arbitrary positive number that penalizes variability in the growth component series.

Optimization problem

$$\min_{\{g_i\}} \left\{ \sum_{i=1}^{N} (y_i - g_i)^2 + \lambda \sum_{i=3}^{N} \left[(g_i - g_{i-1}) - (g_{i-1} - g_{i-2}) \right]^2 \right\} \qquad (2)$$

λ is an arbitrary positive number that penalizes variability in the growth component series.

- What could be a possible interpretation for λ?

Optimization problem

$$\min_{\{g_i\}} \left\{ \sum_{i=1}^{N} (y_i - g_i)^2 + \lambda \sum_{i=3}^{N} \left[(g_i - g_{i-1}) - (g_{i-1} - g_{i-2}) \right]^2 \right\} \qquad (2)$$

λ is an arbitrary positive number that penalizes variability in the growth component series.

- What could be a possible interpretation for λ?
- Is there an objective way to choose a value?

Hodrick and Prescott suggested:

Hodrick and Prescott suggested:

- For monthly data, a value of **14400**,

Hodrick and Prescott suggested:

- For monthly data, a value of **14400**,
- for quarterly data, **1600**, and

Hodrick and Prescott suggested:

- For monthly data, a value of **14400**,
- for quarterly data, **1600**, and
- for annual data, **100**.

Bayesian Graduation

Verrall (1993) shows Whittaker graduation is equivalent to a dynamic regression analysis.

Quantity to be graduated: θ_i , and crude estimate: x_i (the observed value).

$$W(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \left(x_i - \theta_i\right)^2 + \lambda \sum_{i=z+1}^{n} \left(\Delta^z \theta_i\right)^2 \tag{3}$$

w_i is a weight function, λ is the smoothing parameter and $\boldsymbol{\theta}$ is the vector of all the θ_i.

Viewed from a Bayesian perspective

The likelihood of $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \theta_i)$$

The additional information that θ_i are expected to follow a smooth progression is incorporated imposing a prior distribution on $\boldsymbol{\theta}$ proportional to

$$\exp\left(-\frac{\lambda}{2} S(\boldsymbol{\theta})\right)$$

$S(\boldsymbol{\theta})$ is the smoothness penalty of the Whittaker graduation approach with $z = 2$, i.e. $S(\boldsymbol{\theta}) = \sum_{i=3}^{n} (\Delta^2 \theta_i)^2$

The posterior distribution

$$\prod_{i=1}^{n} f(x_i;\theta_i) \times \exp\left(-\frac{\lambda}{2}S(\boldsymbol{\theta})\right)$$

Assuming $x_i|\theta_i \sim N\left(\theta_i, \frac{\sigma^2}{w_i}\right)$

Log-posterior distribution (penalized log-likelihood) proportional to

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i(x_i-\theta_i)^2 - \frac{\lambda}{2}S(\boldsymbol{\theta})$$

Performing a Whittaker graduation is equivalent to maximizing the penalised likelihood function (the posterior distribution), when the crude (observed) values have a normal distribution.

Bayesian formulation similar to the one given by Akaike (1980).

Probability distribution function for series

$$f(y|\kappa, a) = \left(\frac{1}{2\pi}\right)^{N/2} \kappa^{N/2} \exp\left(-\frac{\kappa}{2}\|y - a\|^2\right) \tag{4}$$

Probability distribution function for prior trend

$$f(a|R, \delta, a_0) = \left(\frac{1}{2\pi}\right)^{M/2} \delta^{M/2}|R|^{1/2} \exp\left(-\frac{\delta}{2}(a - a_0)^T R(a - a_0)\right) \tag{5}$$

Gamma priors to κ and δ . Prior precision matrix $R = \delta D^T D$.

$$Da = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & \dots & 1 & -2 & 1 \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 - a_1 \\ a_3 - 2a_2 - a_1 \\ \vdots \\ a_N - 2a_{N-1} - a_{N-2} \end{bmatrix}$$

The corresponding penalty parameter is $\lambda = \delta/\kappa$

Index to measure the proportion of P in $(P+Q)^{-1}$, where P and Q are $N \times N$ positive definite matrices (Guerrero et al. (2001)). In our notation:

Percentage of smoothness

$$S(\lambda; N) = 1 - \operatorname{tr}\left[(I_N + \lambda K_2' K_2)^{-1}\right] / N \tag{6}$$

$K_2' K_2$ determines the precision matrix of the trend, K_2 fixed.

Using precision matrices δR and κI_N.

$$p = \operatorname{tr}\left(\delta R(\kappa I_N + \delta R)^{-1}\right)/N$$
$$= 1 - \operatorname{tr}\left[\left(I_N + \frac{\delta}{\kappa}R\right)^{-1}\right]/N$$

$\delta R =$ prior precision matrix

$\delta R + \kappa I_N =$ precision matrix for $a|y, \delta, \kappa$.

Using precision matrices δR and κI_N.

$$p = \operatorname{tr}\left(\delta R(\kappa I_N + \delta R)^{-1}\right)/N$$
$$= 1 - \operatorname{tr}\left[\left(I_N + \frac{\delta}{\kappa}R\right)^{-1}\right]/N$$

$\delta R =$ prior precision matrix

$\delta R + \kappa I_N =$ precision matrix for $a|y, \delta, \kappa$.

- Proportion of prior precision matrix relative to posterior precision matrix (given δ and κ)

Using precision matrices δR and κI_N.

$$p = \text{tr}\left(\delta R(\kappa I_N + \delta R)^{-1}\right)/N$$
$$= 1 - \text{tr}\left[\left(I_N + \frac{\delta}{\kappa}R\right)^{-1}\right]/N$$

$\delta R =$ prior precision matrix

$\delta R + \kappa I_N =$ precision matrix for $a|y, \delta, \kappa$.

- Proportion of prior precision matrix relative to posterior precision matrix (given δ and κ)
- Directly related to the smoothness of the resulting trend

An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

- Logarithms of the original data.

An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

- Logarithms of the original data.
- Non-informative prior distributions for δ and κ are used:

  $\delta \sim \mathrm{Ga}(1, 0.01)$, $\kappa \sim \mathrm{Ga}(1, 0.001)$.

An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

- Logarithms of the original data.
- Non-informative prior distributions for δ and κ are used:

  $\delta \sim \mathrm{Ga}(1, 0.01)$, $\kappa \sim \mathrm{Ga}(1, 0.001)$.
- Prior information of the series, a_0: a linear trend.

An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

- Logarithms of the original data.
- Non-informative prior distributions for δ and κ are used:

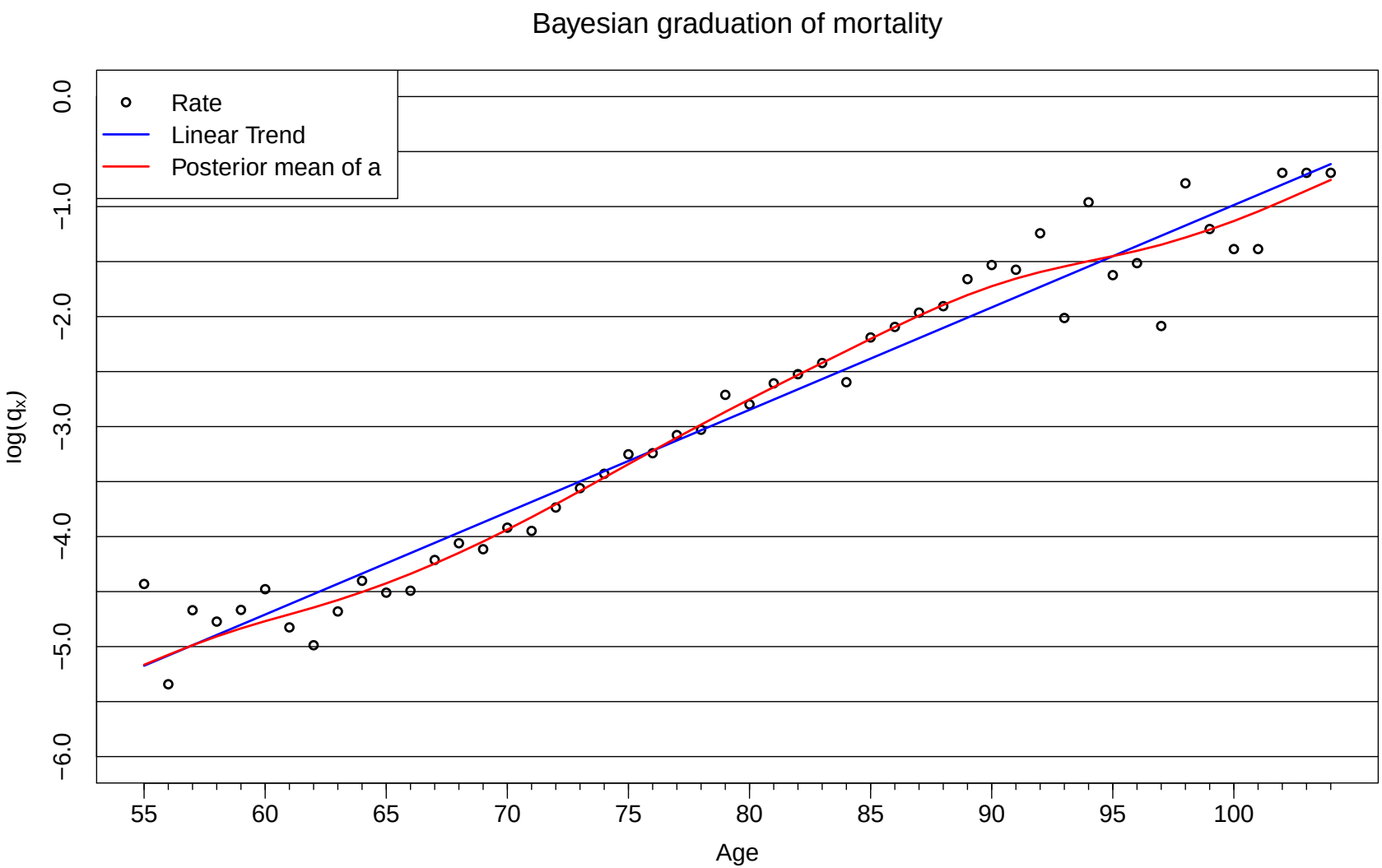
  $\delta \sim \mathrm{Ga}(1, 0.01)$, $\kappa \sim \mathrm{Ga}(1, 0.001)$.
- Prior information of the series, a_0: a linear trend.
- With WinBUGS generating 50,000 samples.

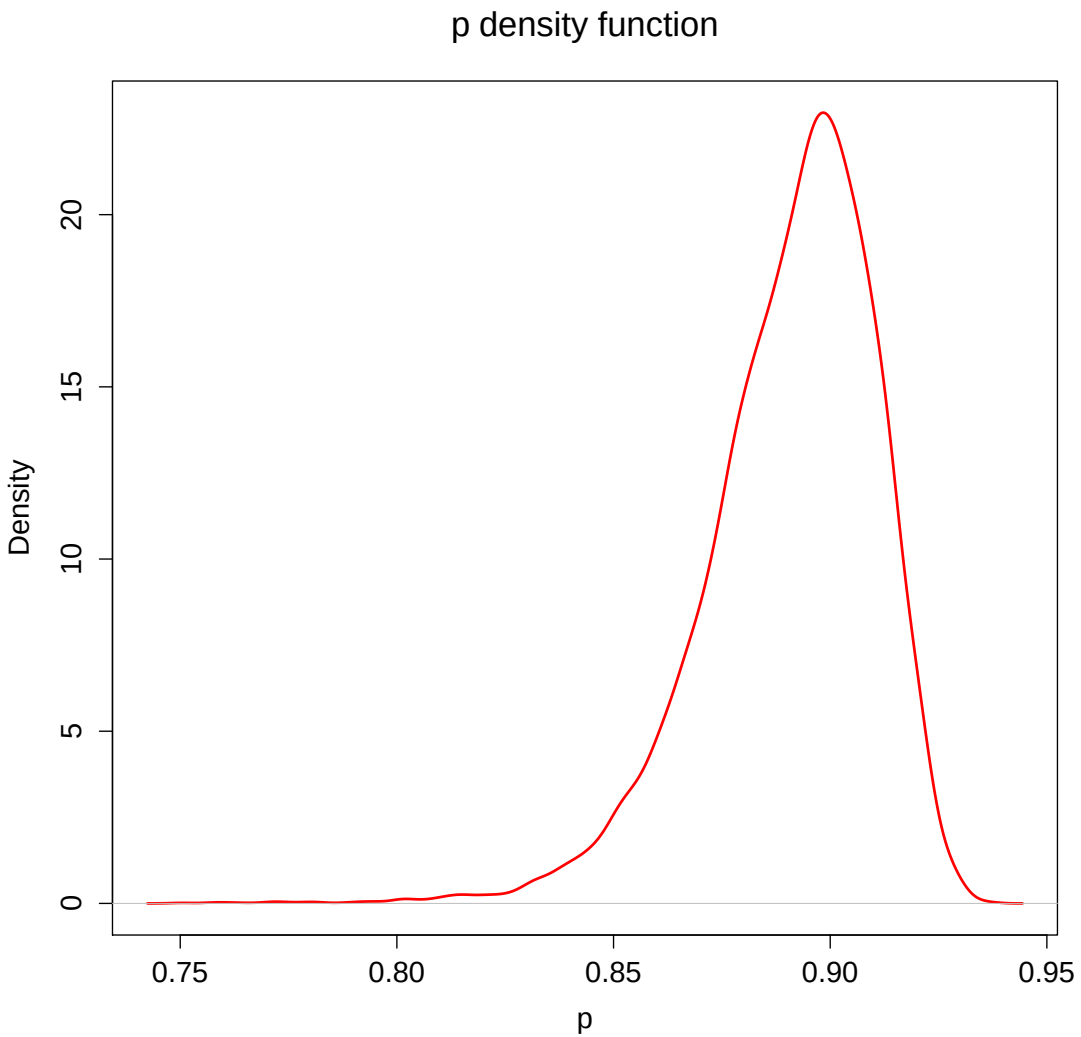
An application

1971 Group Annuity Mortality Table for Female Lives, London (1985).

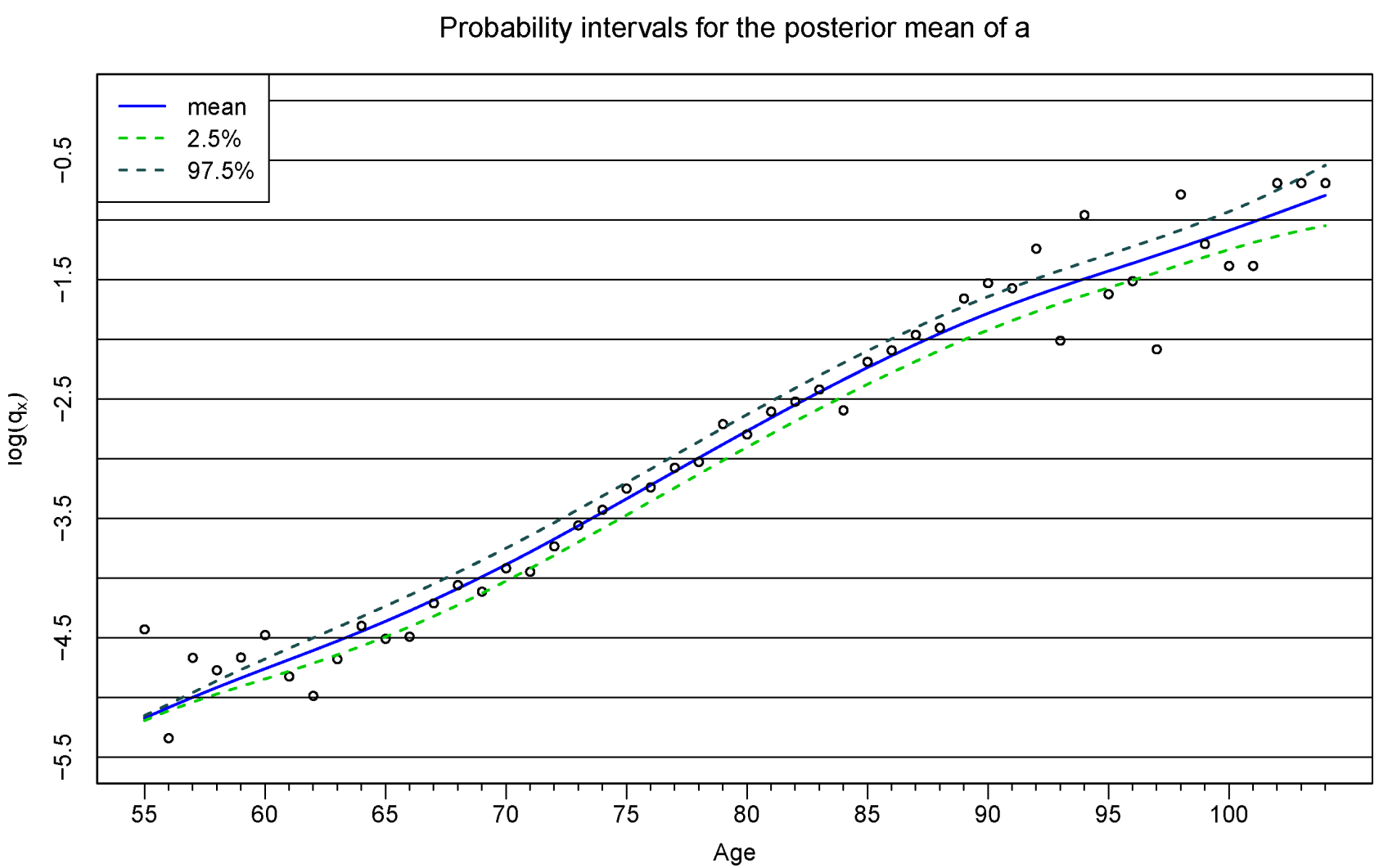
- Logarithms of the original data.
- Non-informative prior distributions for δ and κ are used:

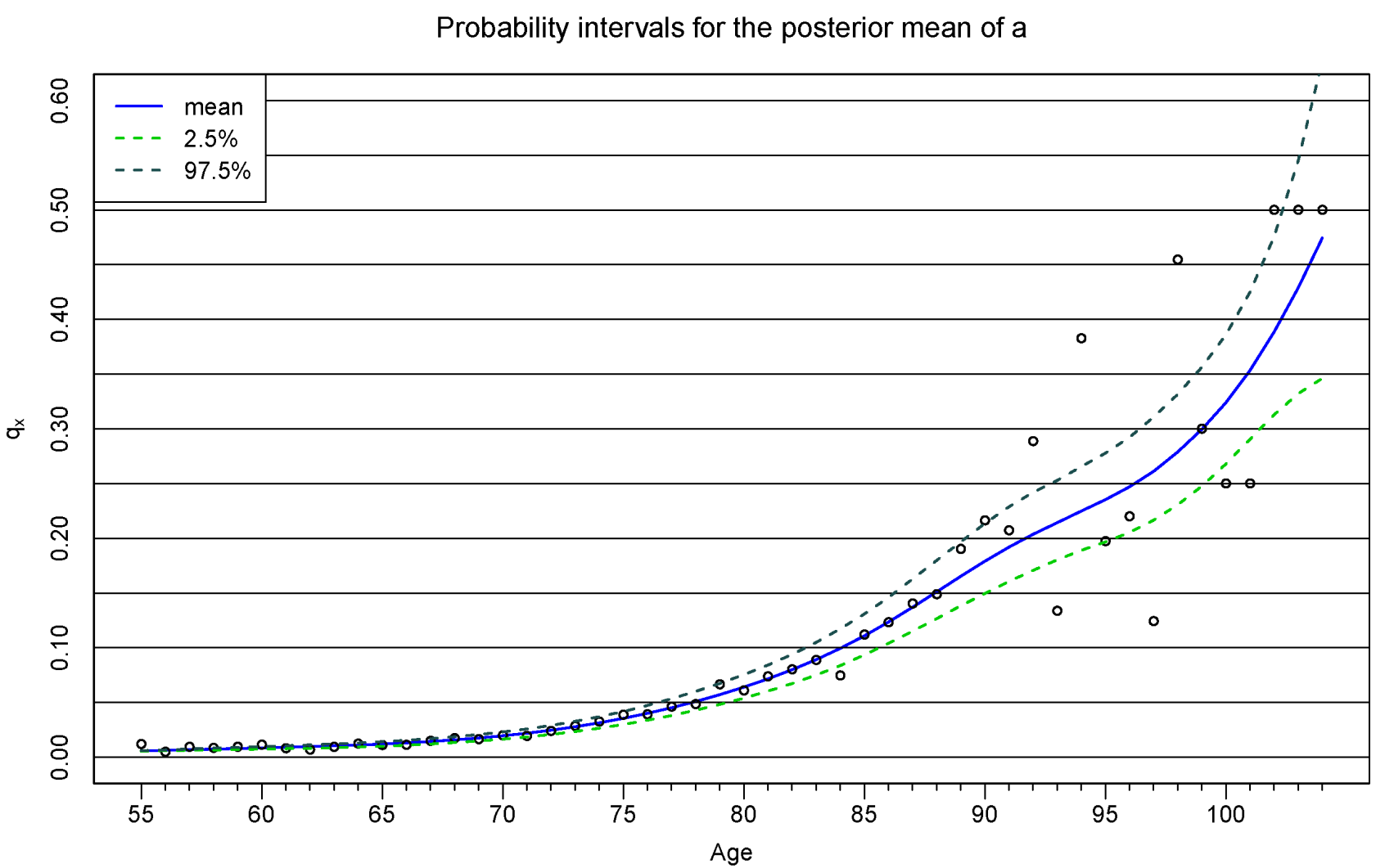
  $\delta \sim \mathrm{Ga}(1, 0.01)$, $\kappa \sim \mathrm{Ga}(1, 0.001)$.
- Prior information of the series, a_0: a linear trend.
- With WinBUGS generating 50,000 samples.
- Probability intervals for the trend.


Bayesian graduation of mortality
Rate
Linear Trend
Posterior mean of a
log(qx)
0.0
−1.0
−2.0
−3.0
−4.0
−5.0
−6.0
55
60
65
70
75
80
85
90
95
100
Age



Summary	
mean	0.891
sd	0.0198
2.5%	0.845
25.0%	0.88
50.0%	0.894
75.0%	0.905
97.5%	0.921


Probability intervals for the posterior mean of a
mean
2.5%
97.5%
log(q_x)
−0.5
−1.5
−2.5
−3.5
−4.5
−5.5
55
60
65
70
75
80
85
90
95
100
Age


Probability intervals for the posterior mean of a
mean
2.5%
97.5%
q_x
0.00
0.10
0.20
0.30
0.40
0.50
0.60
55
60
65
70
75
80
85
90
95
100
Age

Model comparison

To be comparable to that obtained assigning a value to λ:

Model comparison

To be comparable to that obtained assigning a value to λ:

- "Standardize" the observed data $\Longrightarrow z_i = y_i/\sigma_y = y_i\kappa^{1/2}$

Model comparison

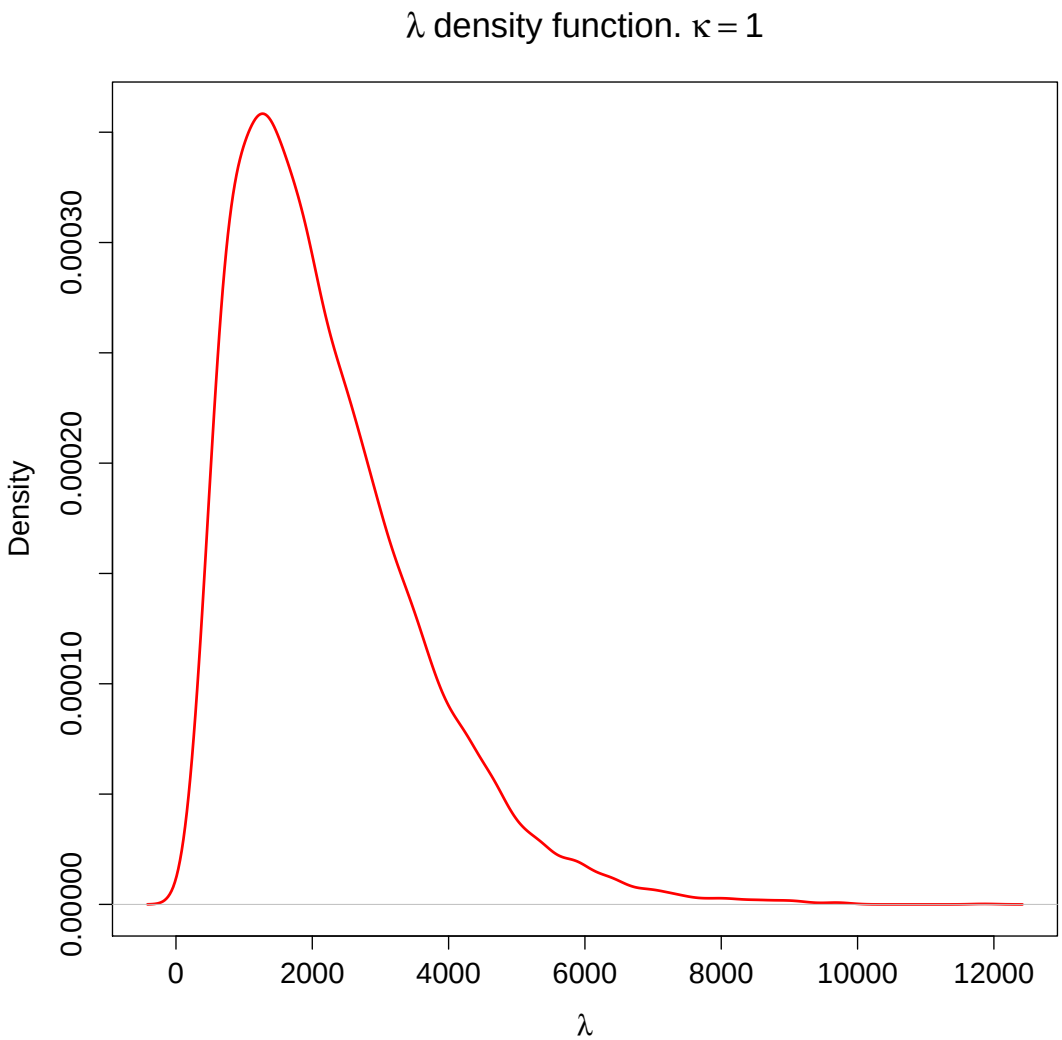
To be comparable to that obtained assigning a value to λ:

- "Standardize" the observed data $\Longrightarrow z_i = y_i/\sigma_y = y_i\kappa^{1/2}$
- Then $y \sim N(b, I_N)$ and $b \sim N(b_0, \lambda R)$, where $\lambda = \delta/\kappa$

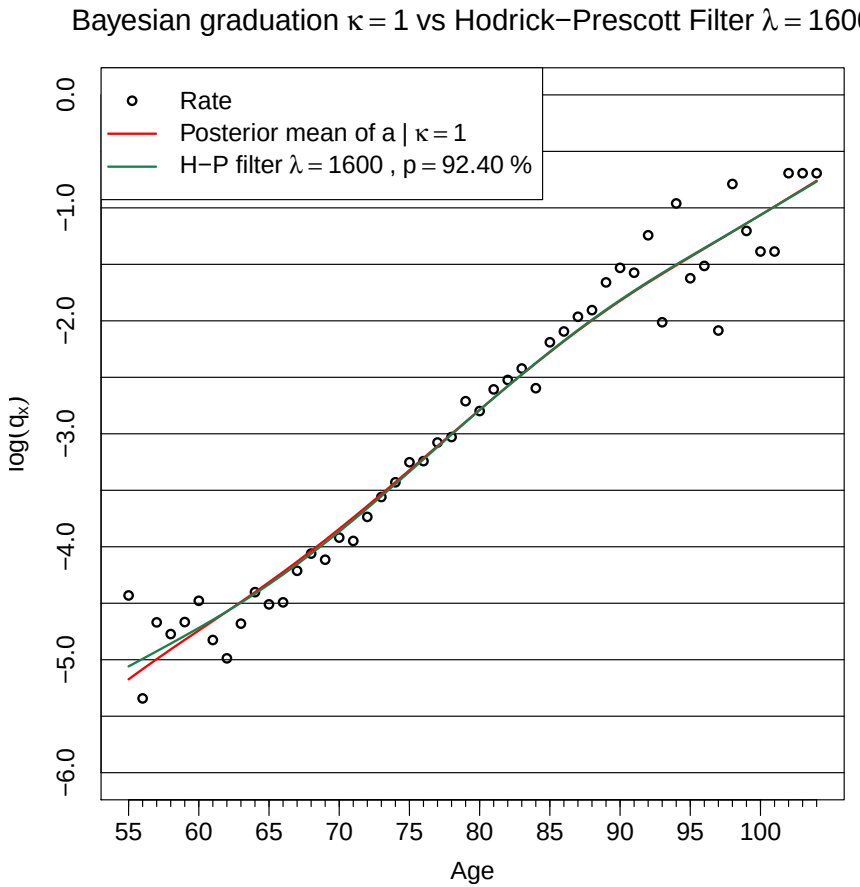
Model comparison

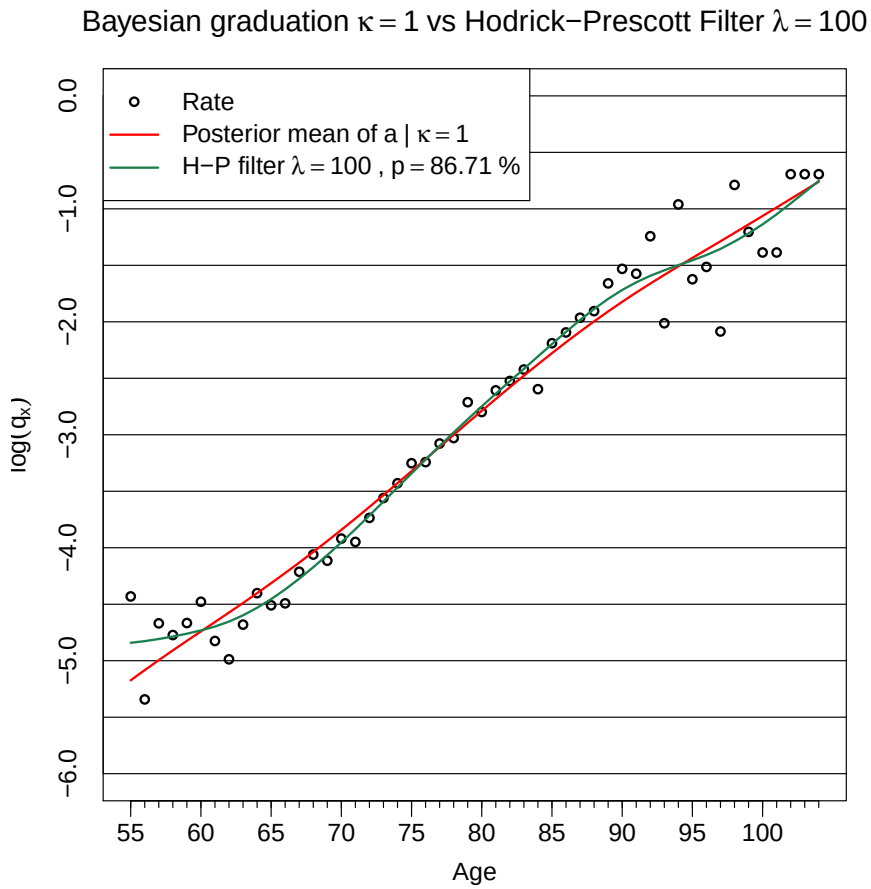
To be comparable to that obtained assigning a value to λ:

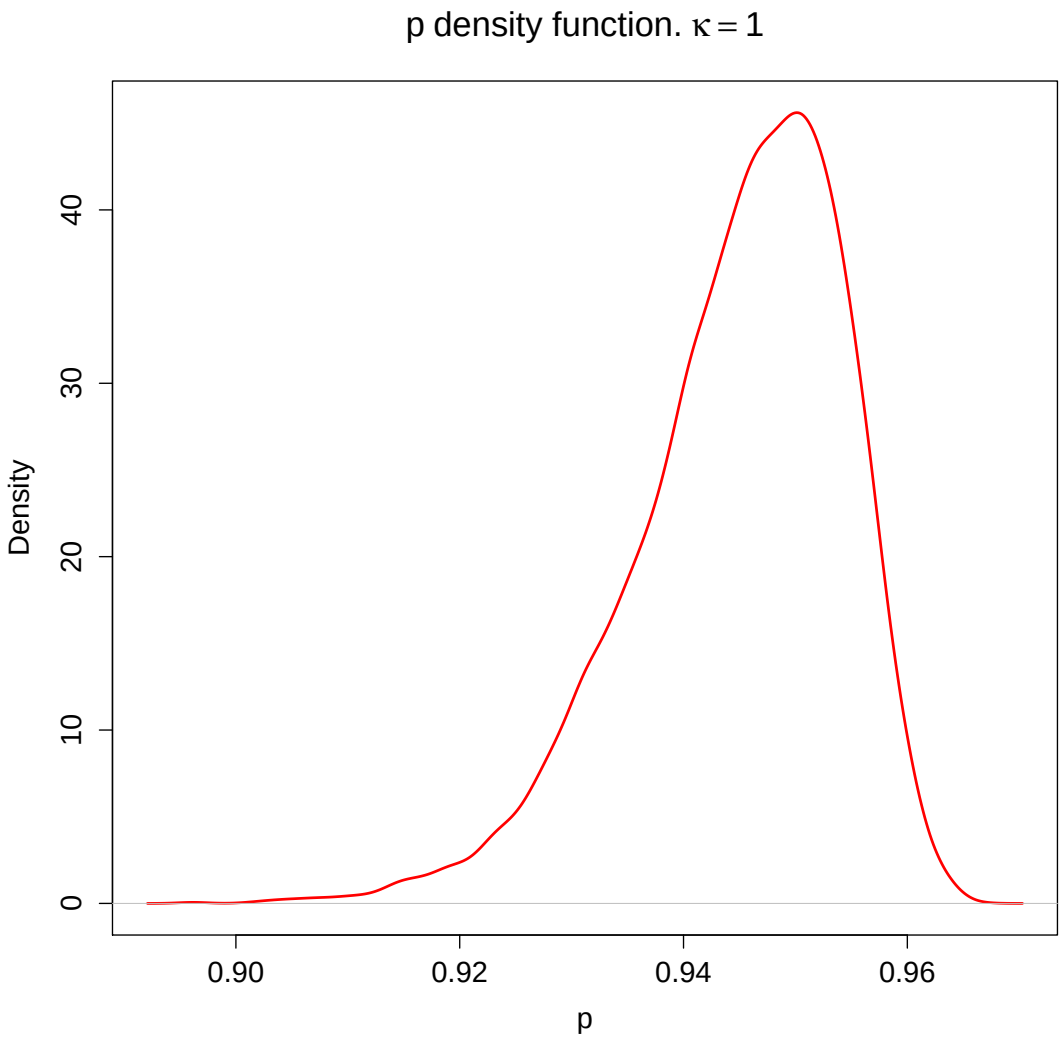
- "Standardize" the observed data $\Longrightarrow z_i = y_i/\sigma_y = y_i\kappa^{1/2}$
- Then $y \sim N(b, I_N)$ and $b \sim N(b_0, \lambda R)$, where $\lambda = \delta/\kappa$
- If $\kappa = 1$ in WinBUGS then δ is directly comparable to λ.



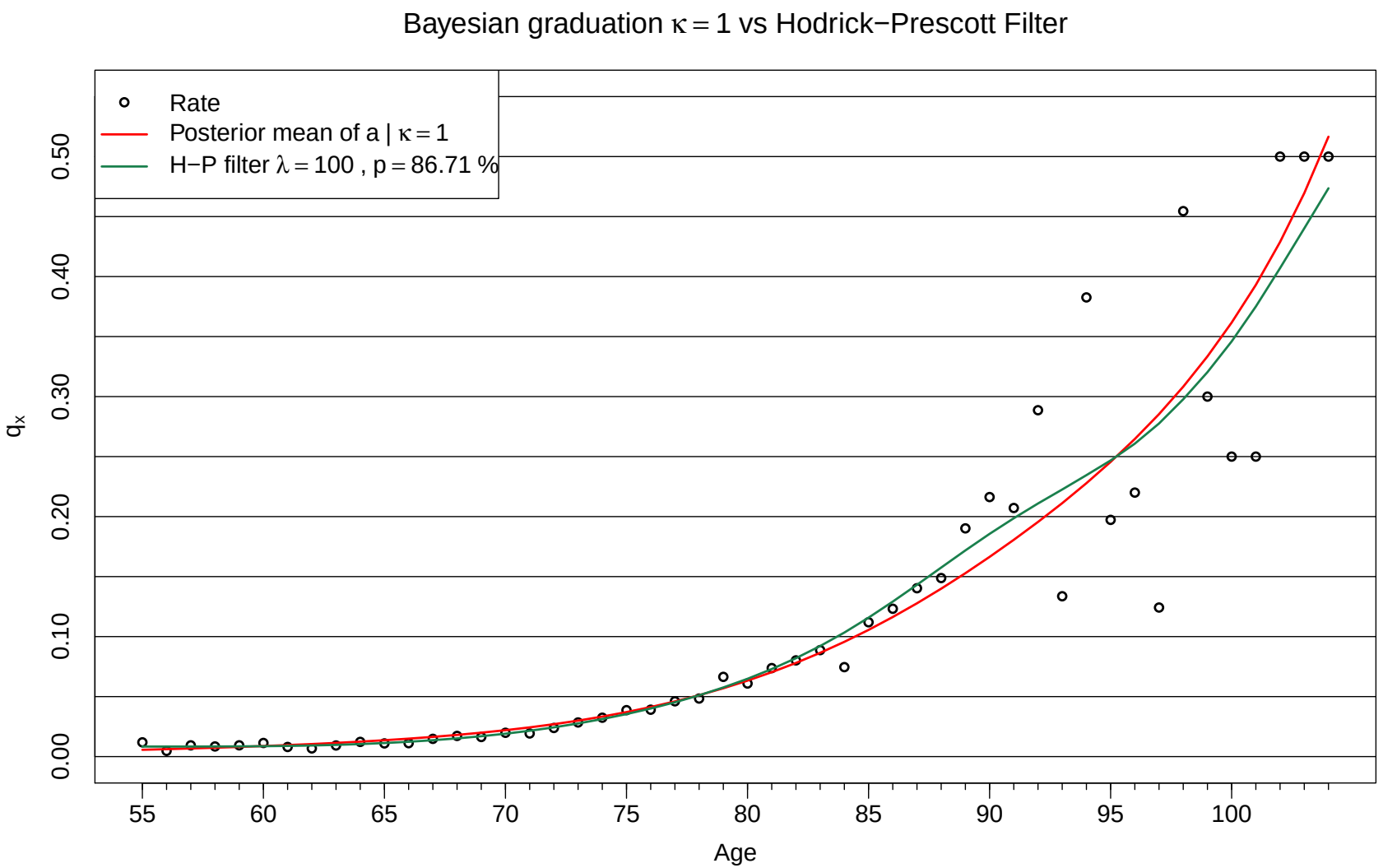
Summary	
mean	2215
sd	1382
2.5%	462
25.0%	1187
50.0%	1906
75.0%	2938
97.5%	5645

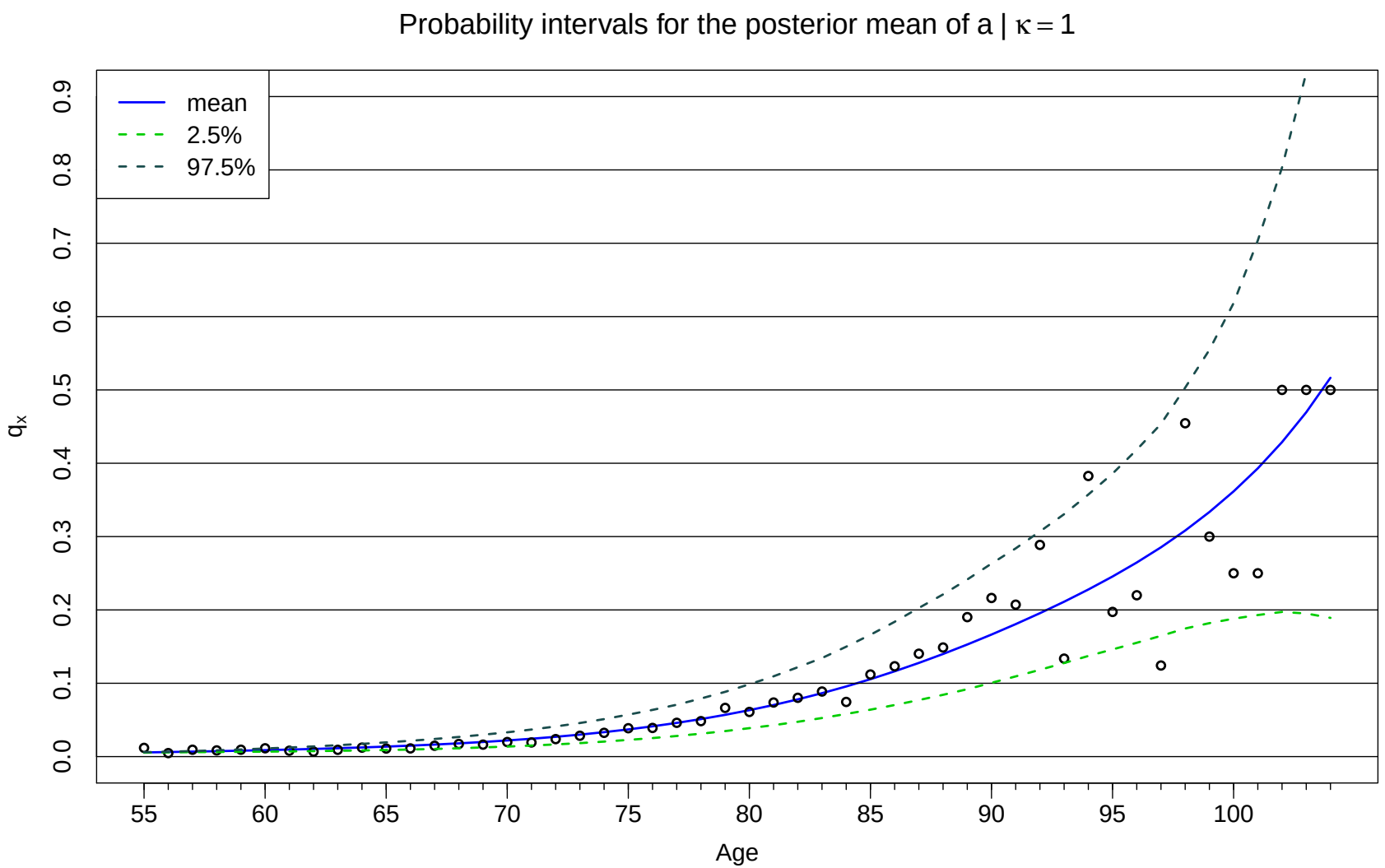

Bayesian graduation κ = 1 vs Hodrick−Prescott Filter λ = 1600
Rate
Posterior mean of a | κ = 1
H−P filter λ = 1600 , p = 92.40 %
log(qx)
0.0
−1.0
−2.0
−3.0
−4.0
−5.0
−6.0
55 60 65 70 75 80 85 90 95 100
Age


Bayesian graduation κ = 1 vs Hodrick−Prescott Filter λ = 100
Rate
Posterior mean of a | κ = 1
H−P filter λ = 100 , p = 86.71 %
log(qx)
0.0
−1.0
−2.0
−3.0
−4.0
−5.0
−6.0
55 60 65 70 75 80 85 90 95 100
Age



Summary	
mean	0.945
sd	0.00941
2.5%	0.923
25.0%	0.94
50.0%	0.946
75.0%	0.952
97.5%	0.959


Bayesian graduation κ = 1 vs Hodrick−Prescott Filter
Rate
Posterior mean of a | κ = 1
H−P filter λ = 100 , p = 86.71 %
q_x
0.00
0.10
0.20
0.30
0.40
0.50
55
60
65
70
75
80
85
90
95
100
Age


Probability intervals for the posterior mean of a | κ = 1
mean
2.5%
97.5%
q_x
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
55
60
65
70
75
80
85
90
95
100
Age

Alternative approach

Choose the prior distributions of δ and κ so that the prior distribution of $\lambda = \delta/\kappa$ has an expected value of 100, which is the recommended level for annual data. Let:

Alternative approach

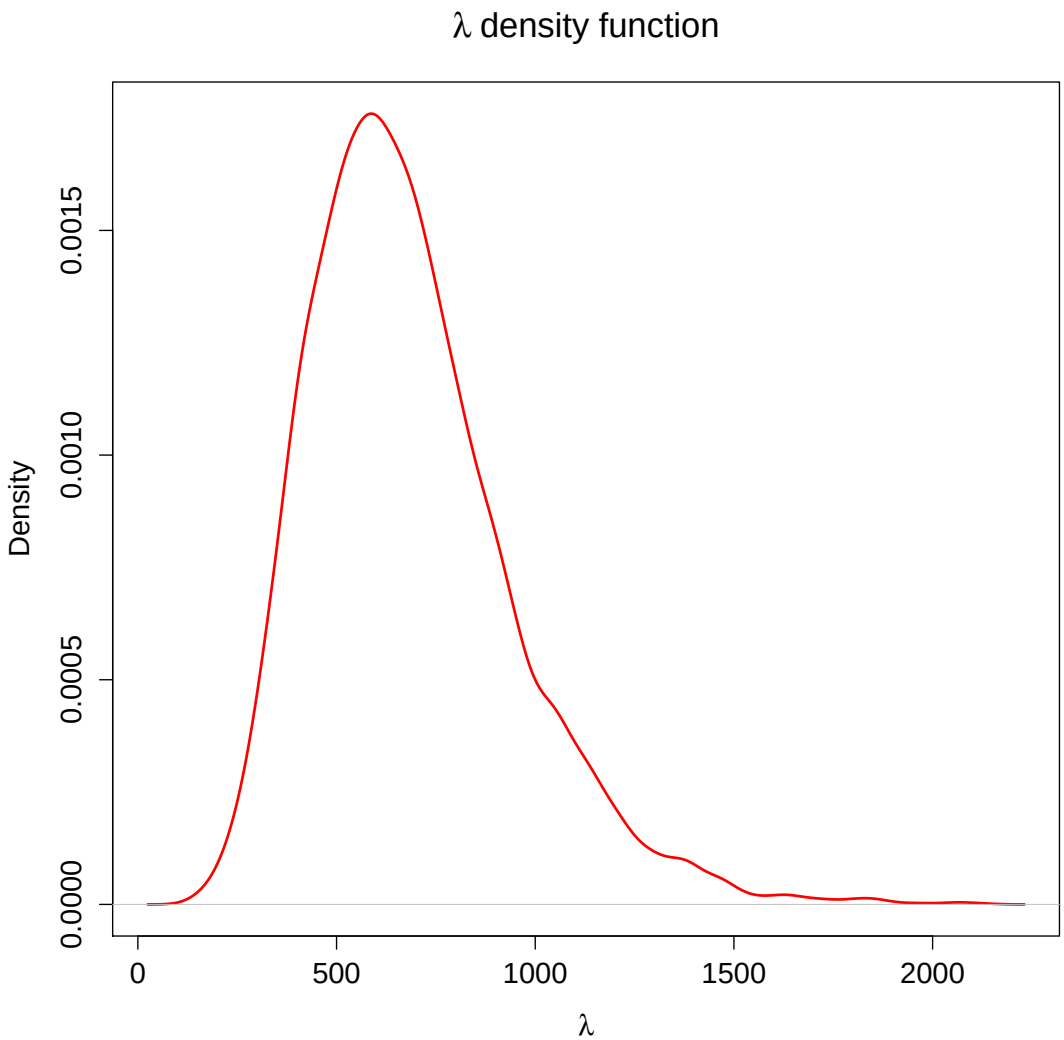
Choose the prior distributions of δ and κ so that the prior distribution of $\lambda = \delta/\kappa$ has an expected value of 100, which is the recommended level for annual data. Let:

- $\delta \sim \mathrm{Ga}(10, 0.001)$

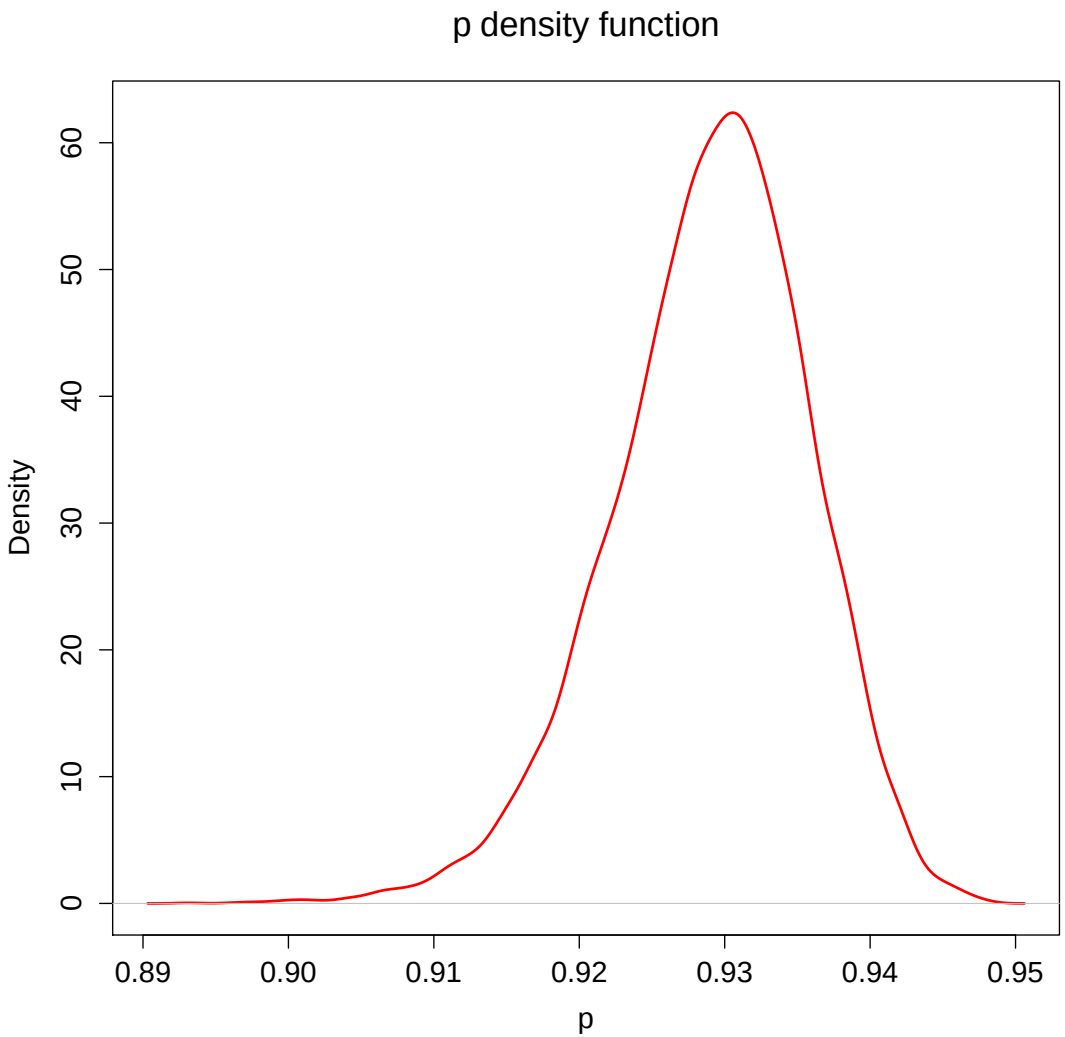
Alternative approach

Choose the prior distributions of δ and κ so that the prior distribution of $\lambda = \delta/\kappa$ has an expected value of 100, which is the recommended level for annual data. Let:

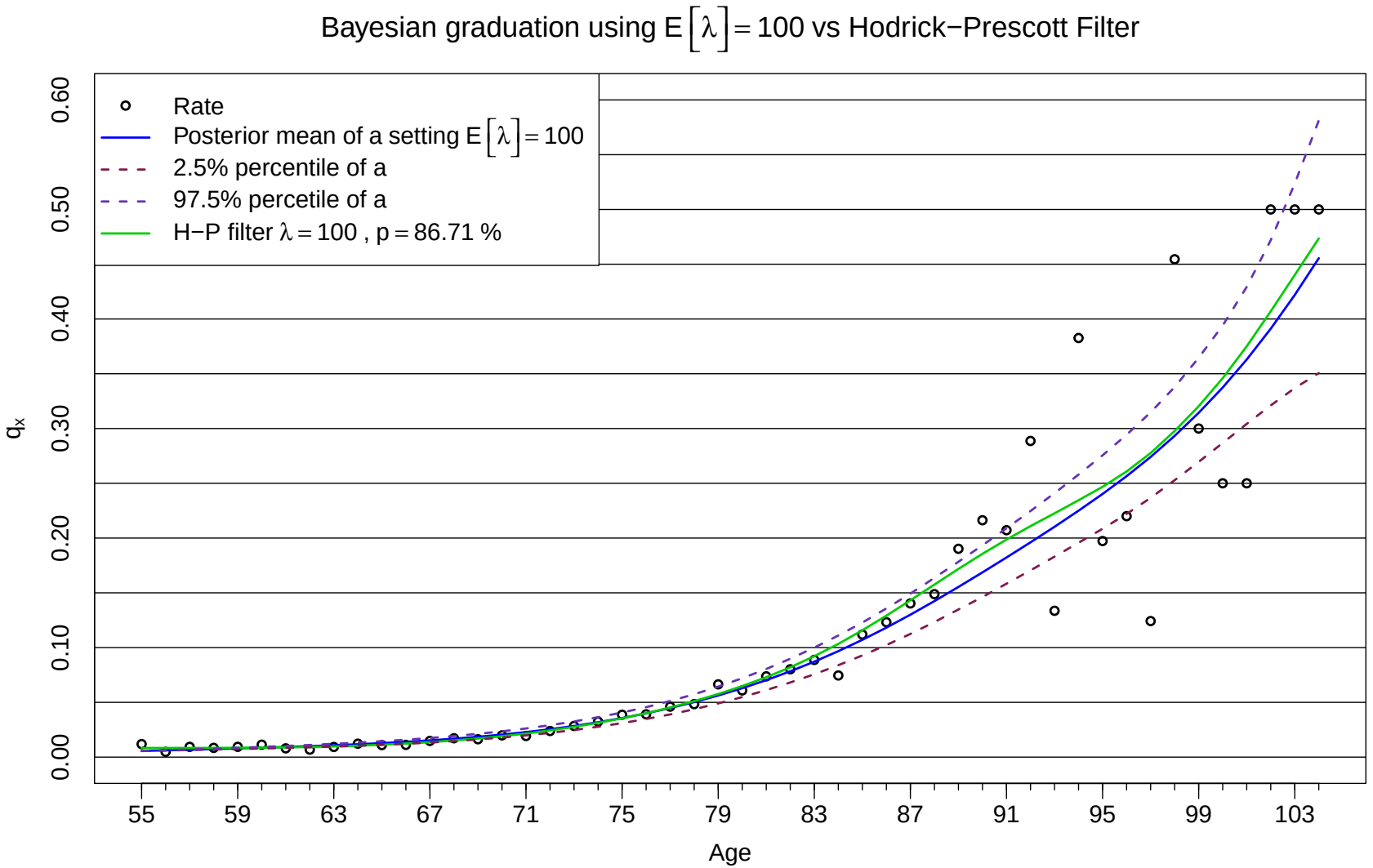
- $\delta \sim \mathrm{Ga}(10, 0.001)$
- $\kappa \sim \mathrm{Ga}(2, 0.01)$



Summary	
mean	684
sd	256
2.5%	302
25.0%	503
50.0%	647
75.0%	822
97.5%	1291



Summary	
mean	0.929
sd	0.00676
2.5%	0.914
25.0%	0.925
50.0%	0.929
75.0%	0.934
97.5%	0.941


Bayesian graduation using E[λ] = 100 vs Hodrick−Prescott Filter
Rate
Posterior mean of a setting E[λ] = 100
2.5% percentile of a
97.5% percetile of a
H−P filter λ = 100 , p = 86.71 %
q_x
0.00
0.10
0.20
0.30
0.40
0.50
0.60
55
59
63
67
71
75
79
83
87
91
95
99
103
Age

Final Remarks

Final Remarks

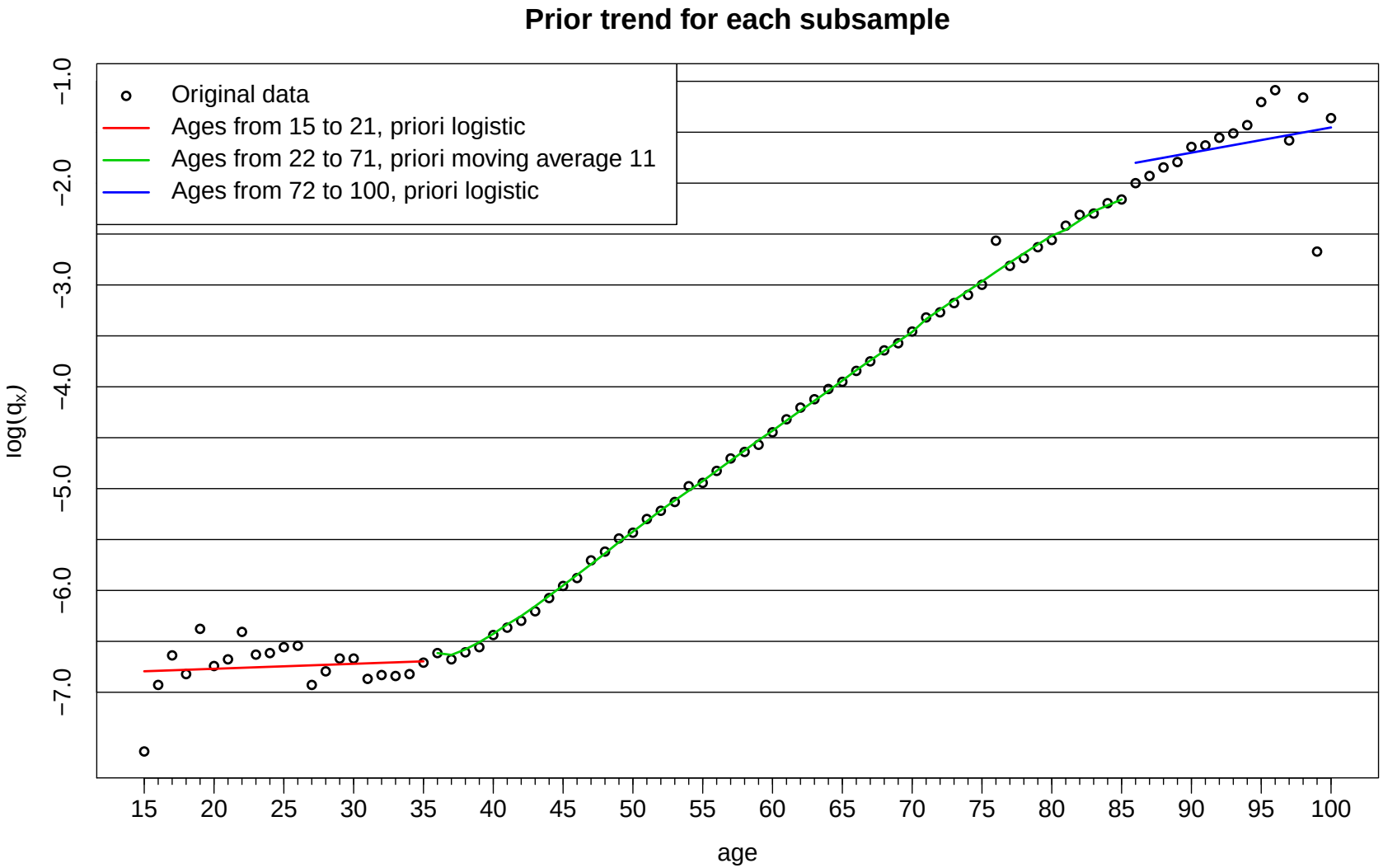
- Bayesian graduation approach provides a natural way for obtaining the smoothing constant

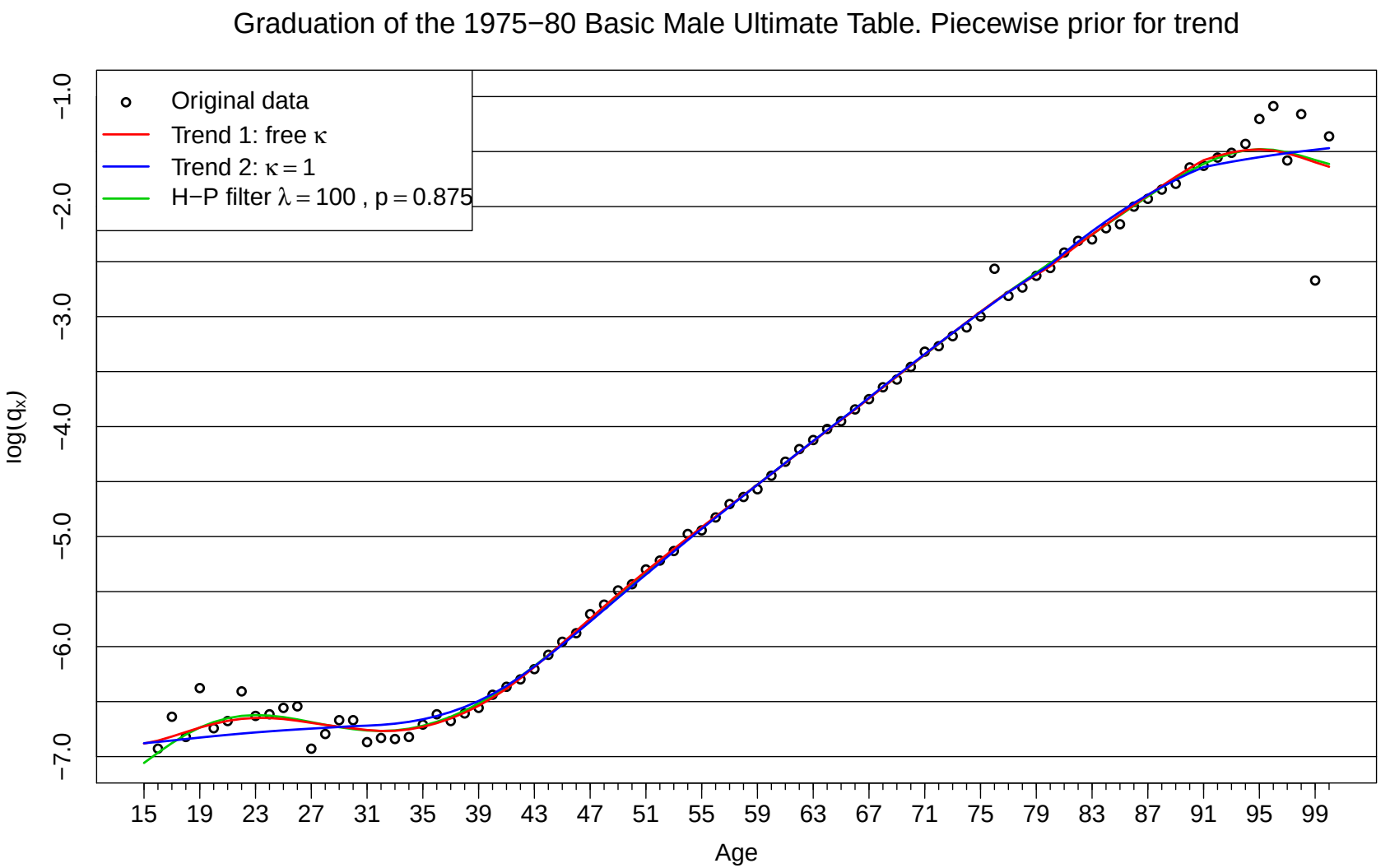
Final Remarks

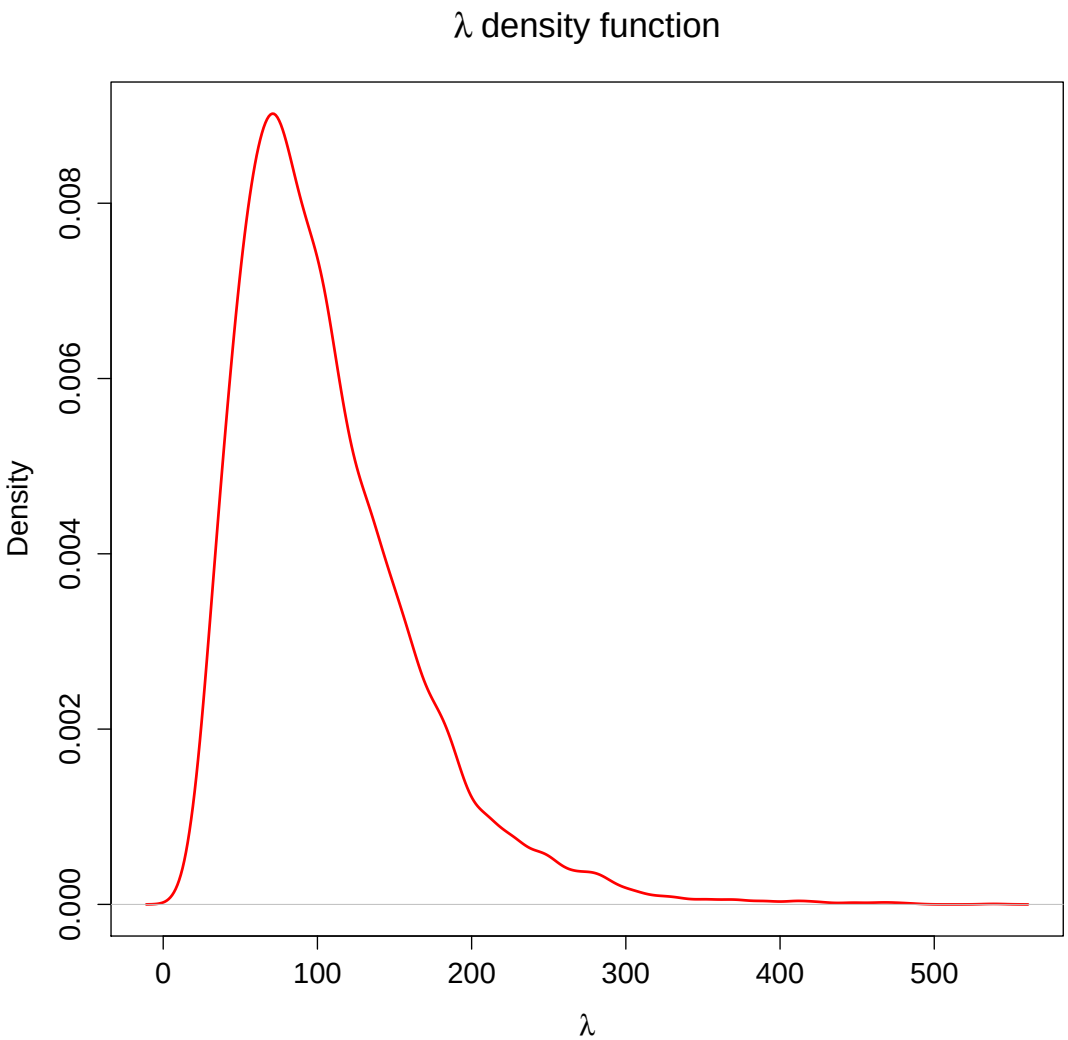
- Bayesian graduation approach provides a natural way for obtaining the smoothing constant
- Wealth of additional probabilistic information

Final Remarks

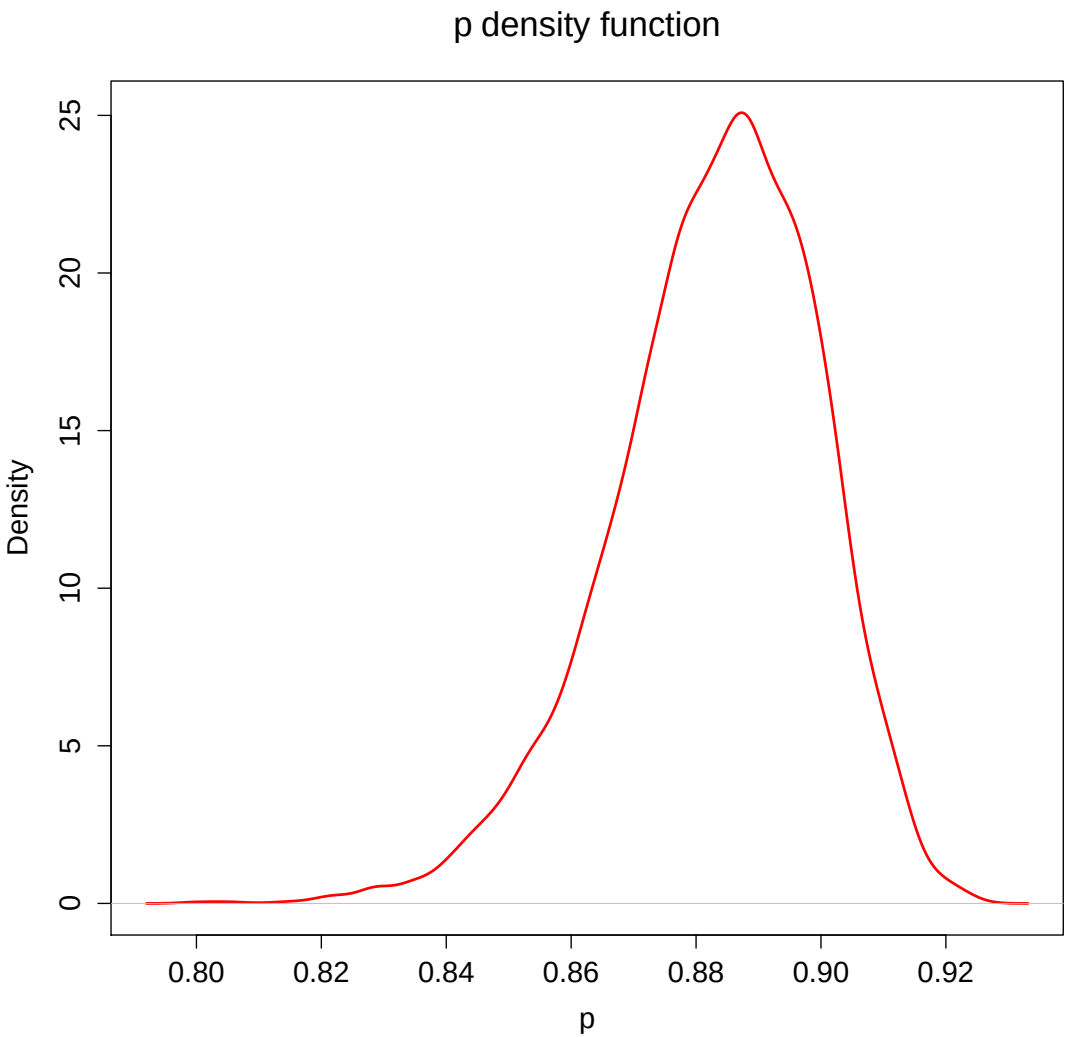
- Bayesian graduation approach provides a natural way for obtaining the smoothing constant
- Wealth of additional probabilistic information
- Using some previously given value, as it is done in H-P, has many implications for the results (trend, confidence limits, etc.) depending on sample size, and the data themselves.


Prior trend for each subsample
Original data
Ages from 15 to 21, priori logistic
Ages from 22 to 71, priori moving average 11
Ages from 72 to 100, priori logistic
log(qx)
−1.0
−2.0
−3.0
−4.0
−5.0
−6.0
−7.0
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
age

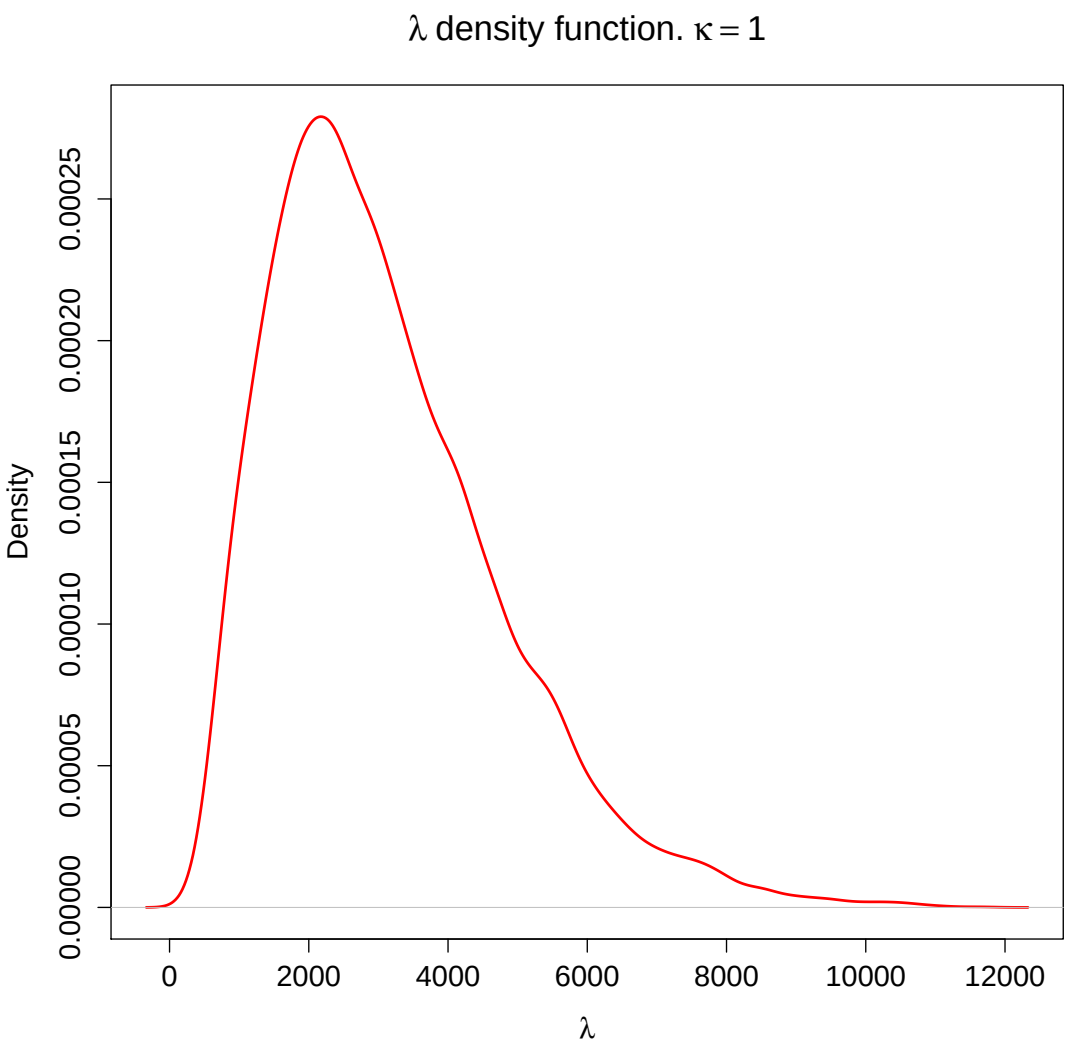

Graduation of the 1975−80 Basic Male Ultimate Table. Piecewise prior for trend
Original data
Trend 1: free κ
Trend 2: κ = 1
H−P filter λ = 100 , p = 0.875
log(qx)
−1.0
−2.0
−3.0
−4.0
−5.0
−6.0
−7.0
15 19 23 27 31 35 39 43 47 51 55 59 63 67 71 75 79 83 87 91 95 99
Age



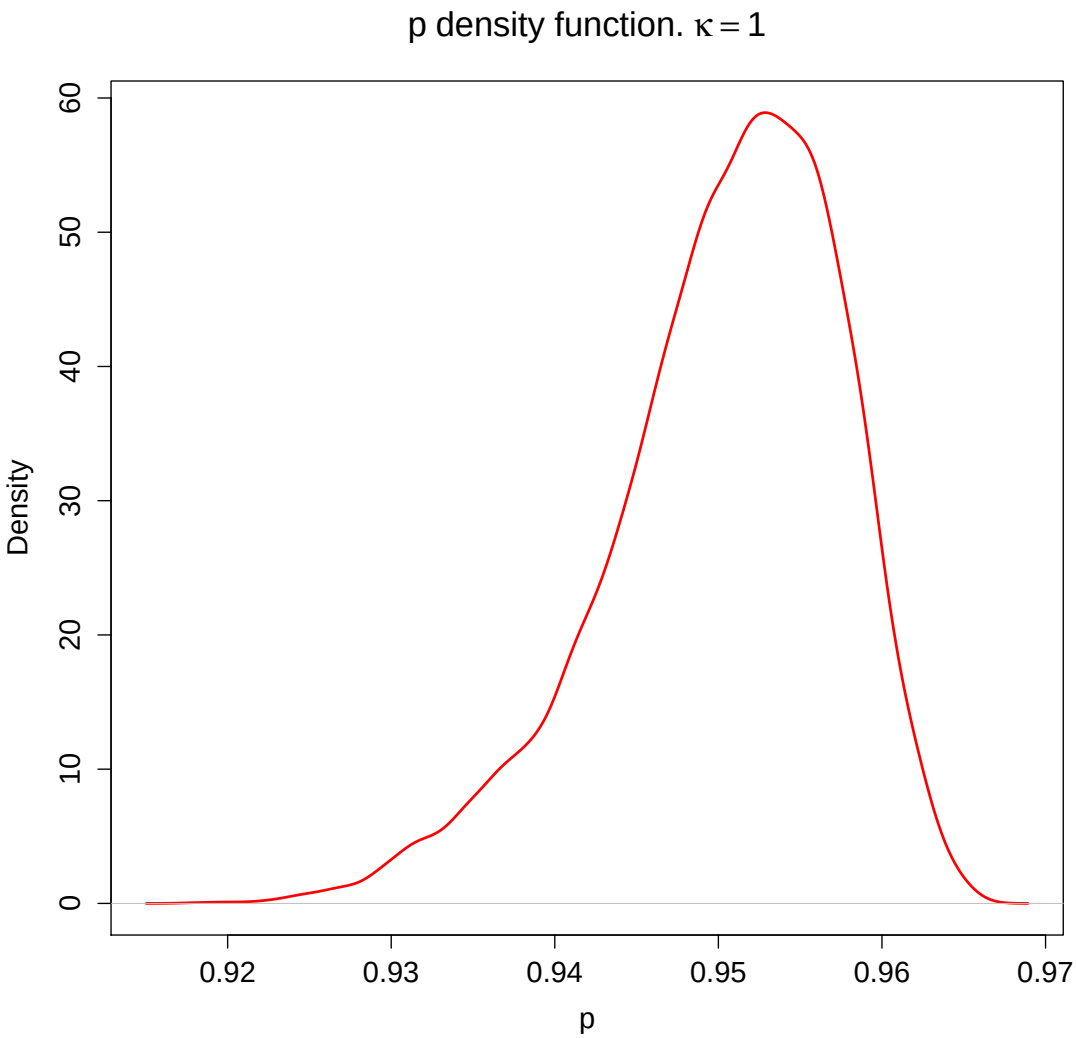
Summary	
mean	106
sd	58.6
2.5%	30.9
25.0%	64.4
50.0%	93
75.0%	134
97.5%	254



Summary	
mean	0.883
sd	0.0165
2.5%	0.846
25.0%	0.873
50.0%	0.884
75.0%	0.895
97.5%	0.911



Summary	
mean	3112
sd	1650
2.5%	816
25.0%	1886
50.0%	2808
75.0%	4047
97.5%	7153



Summary	
mean	0.95
sd	0.00715
2.5%	0.934
25.0%	0.946
50.0%	0.951
75.0%	0.956
97.5%	0.962