Calibrated Spline Graduation of Age-Group Fertility Rates

UN Expert Group Meeting
Methods for the WPP 2021+

April 2020

Carl Schmertmann
Florida State University
Tallahassee, Florida USA

Objective

For age-group fertility data $\{{}_n\boldsymbol{F}_x\}$
find a continuous fertility schedule $\boldsymbol{f(x)}$ that

1. Matches age-group data*

2. Looks like schedules from a large calibration database*

* = cannot <u>*exactly*</u> meet both objectives

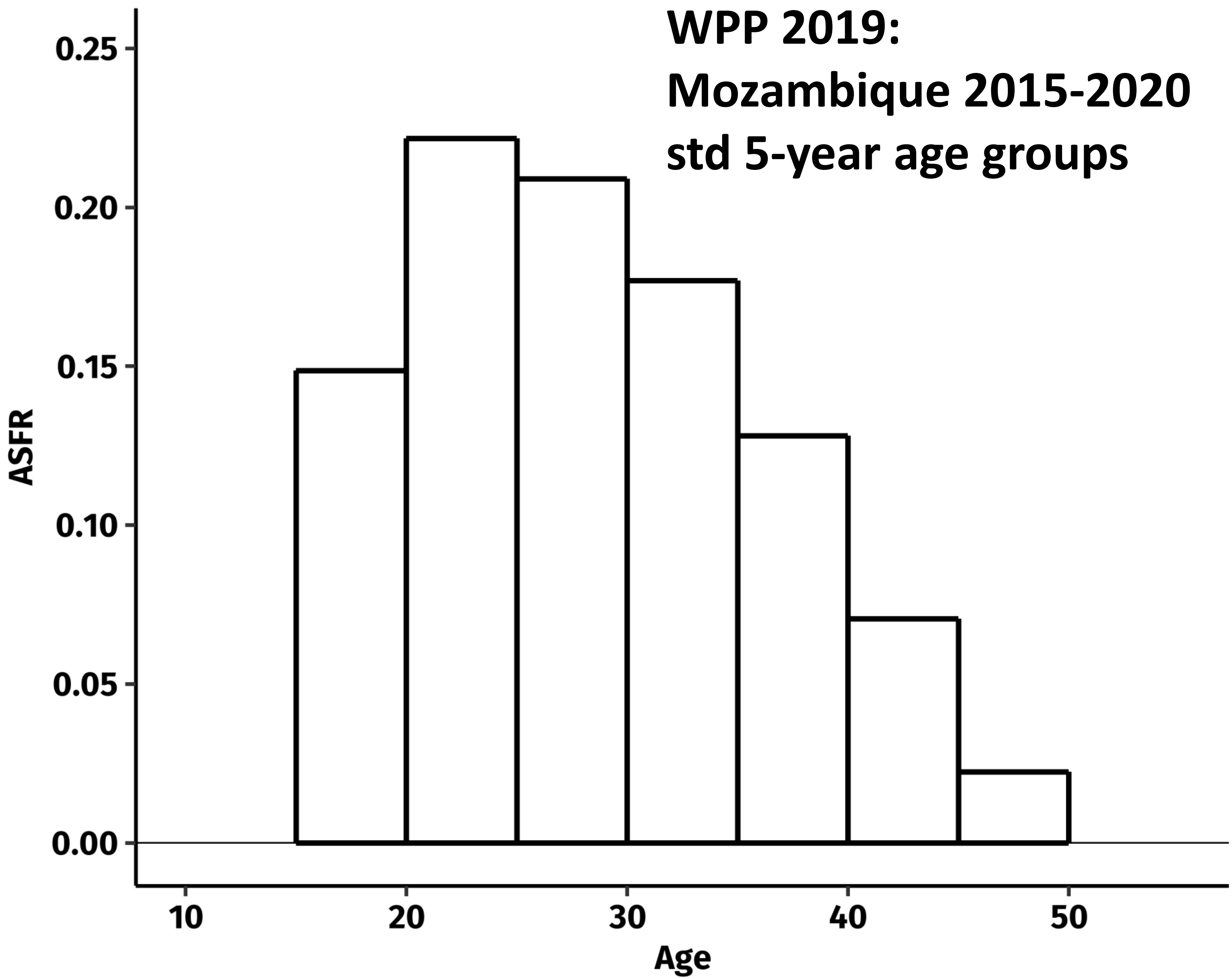
Principles

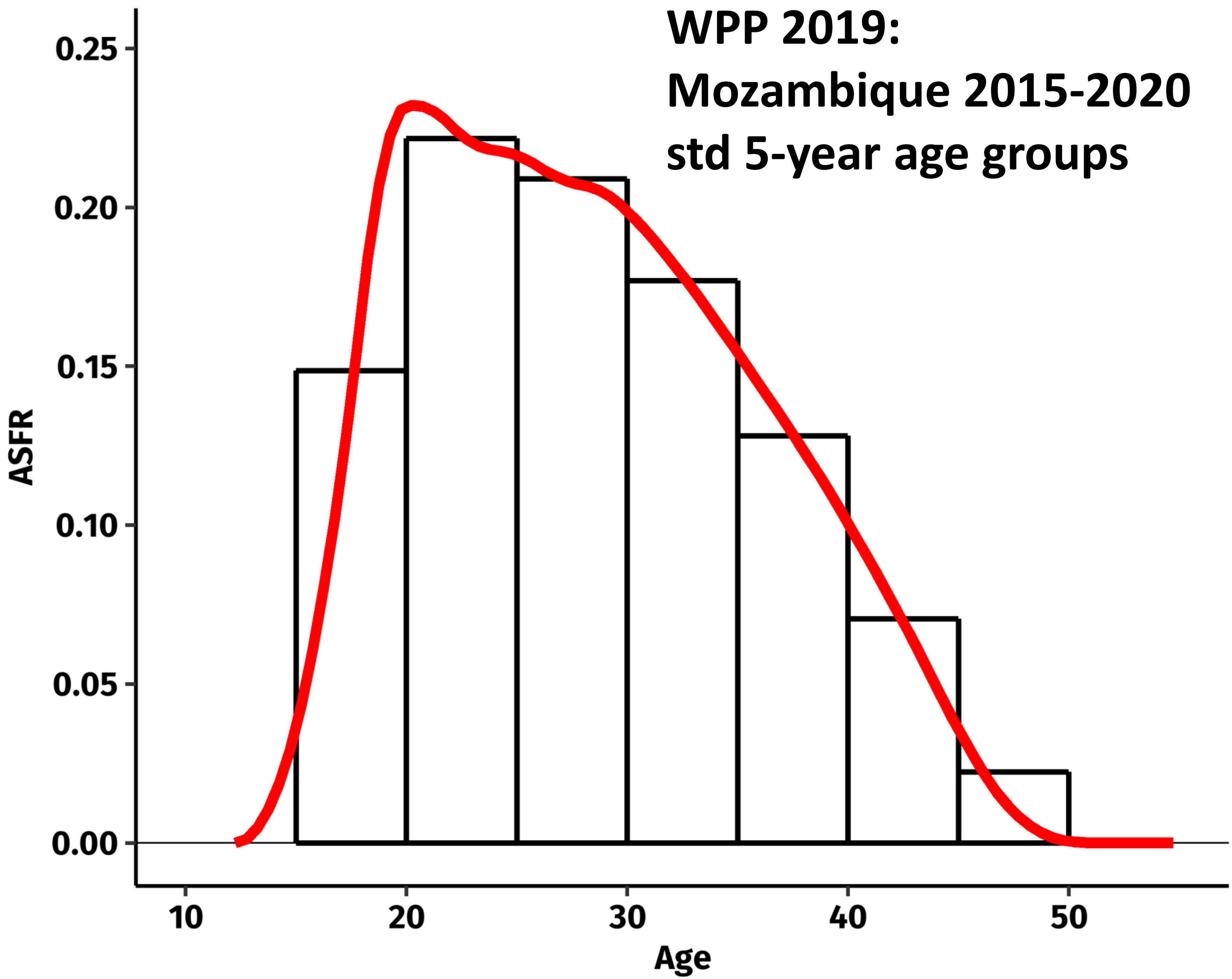
Construct $f(x)$ such that it can be estimated

1. from any age grouping $\{{}_nF_x\}$

2. for ages outside available $\{{}_nF_x\}$
   e.g. <15 or >50

3. using simple spreadsheet arithmetic

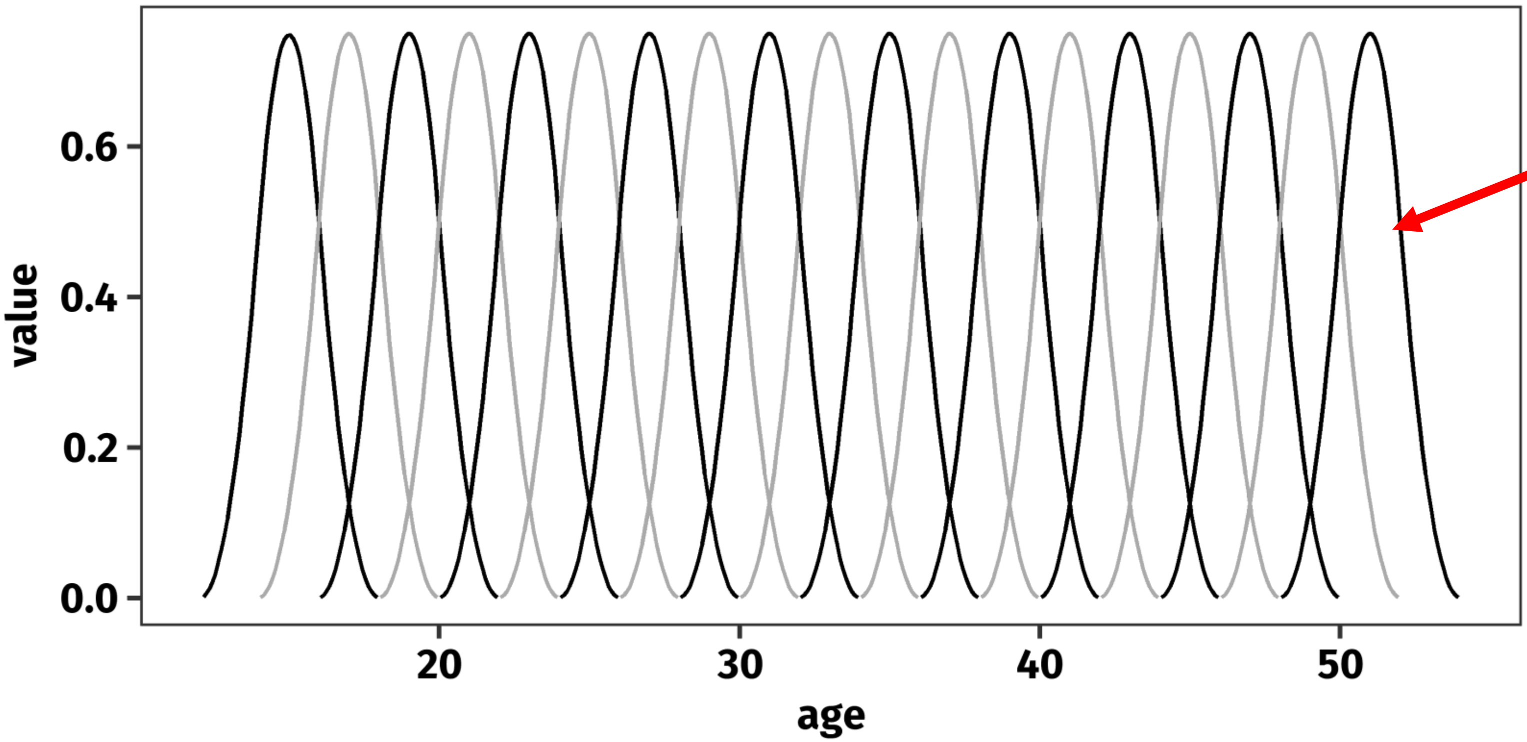

WPP 2019:
Mozambique 2015-2020
std 5-year age groups
ASFR
0.25
0.20
0.15
0.10
0.05
0.00
10
20
30
40
50
Age


WPP 2019:
Mozambique 2015-2020
std 5-year age groups
ASFR
0.25
0.20
0.15
0.10
0.05
0.00
10
20
30
40
50
Age

Main Ideas

$f(x)$ is a continuous, high-dimensional B-spline

$$f(x) = \sum_k \theta_k \, b_k(x)$$



$f(x)$ = **wtd sum** of these 19 functions

Main Ideas

Discretize over a fine grid of A ages such as

$(x_1 \ldots x_A) \quad = \quad (12.25, 12.75, \ldots, 54.75)$

$$\underset{A \times 1}{\boldsymbol{f}} = \begin{bmatrix} \boldsymbol{f_1} \\ \vdots \\ \boldsymbol{f_A} \end{bmatrix} = \underset{A \times 19}{\boldsymbol{B}} \; \underset{19 \times 1}{\boldsymbol{\theta}}$$

Main Ideas

Observed rates for groups are averages of the detailed rates

$$\{{}_{n}F_{x}\} = \boldsymbol{G} \; \boldsymbol{f} \; = \; \boldsymbol{G} \; \boldsymbol{B} \; \boldsymbol{\theta}$$

7 x 1 **7 x A** **A x 1** **7 x A** **A x 19** **19 x 1**

→ 19 parameters to fit **7 targets**

→ needs *regularization*

Any particular θ implies spline fitting errors

$$\varepsilon_{fit} = \{{}_{n}F_{x}\} - G \quad B \quad \theta$$

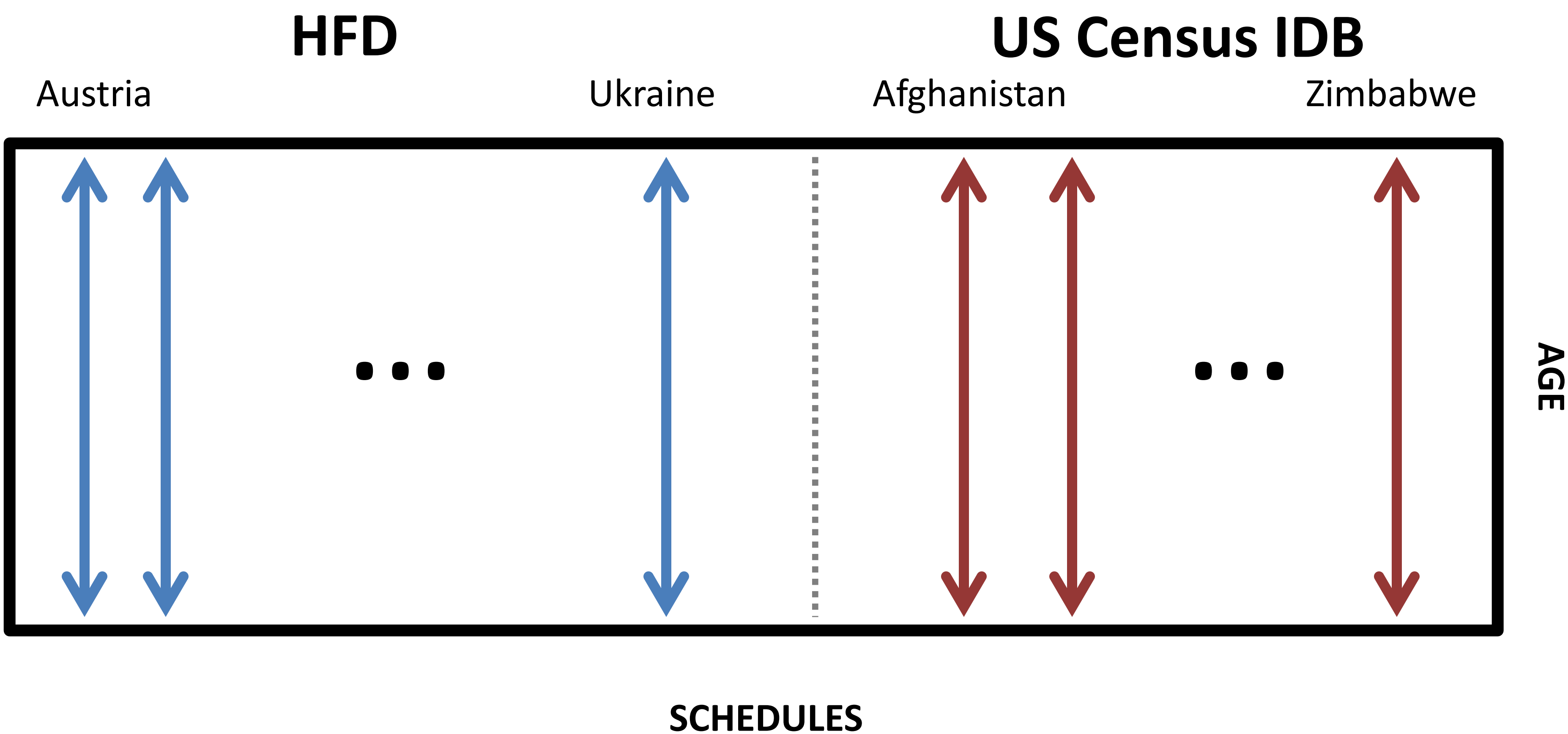
7 x 1   7 x 1   7 x A   A x 19   19 x 1

Lower SSE → better θ

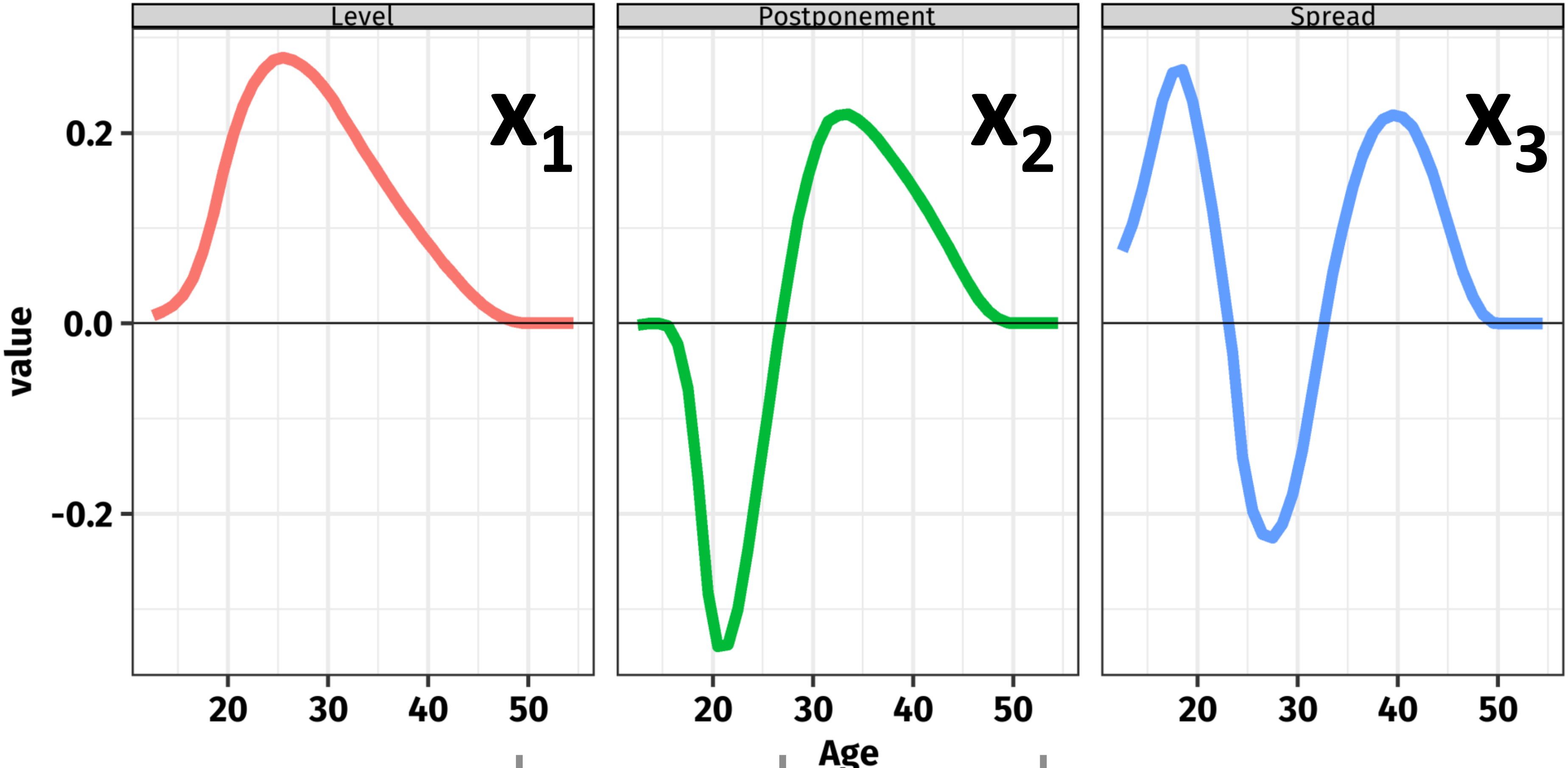
$$SSE_{fit}(\theta) = \varepsilon'_{fit}\varepsilon_{fit}$$

Singular Value Decomposition

find principal components of single-year age schedules over *CALIBRATION DATA*



f(x) well approx. by three components



$$f = x_1\ \beta_1 + x_2\ \beta_2 + x_3\ \beta_3 + \varepsilon$$

$$= X\ \beta + \varepsilon$$

← near-zero "shape residuals"

Any particular θ implies spline "shape residuals"

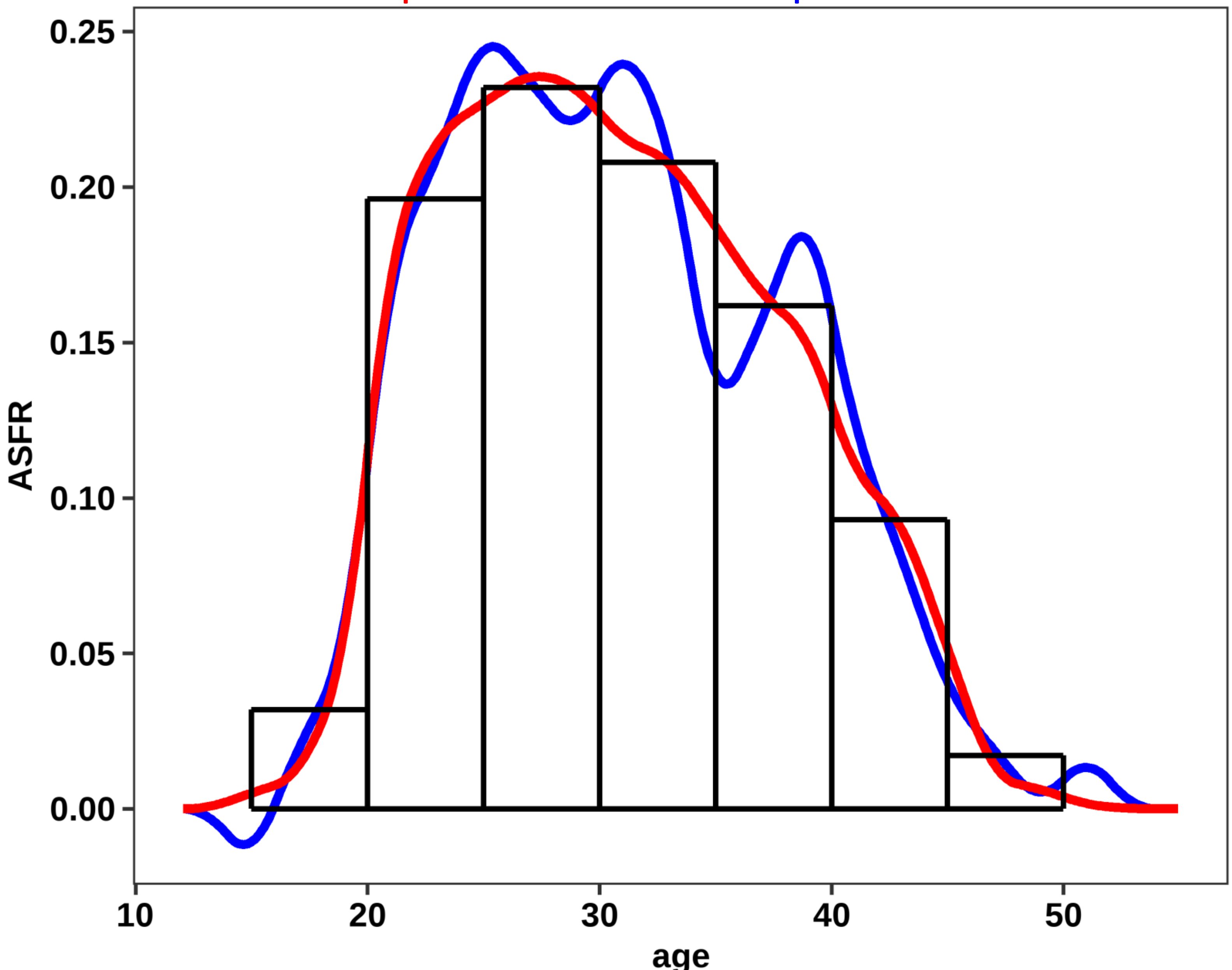
$$\varepsilon_{shape} = f - Proj_X(f)$$
$$= M_X\ f$$
$$= M_X\ B\ \theta$$

Lower SSE → better shape → better θ

$$SSE_{shape}(\theta) = \varepsilon'_{shape}\ \Omega^{-1}\ \varepsilon_{shape}$$

↑ Estimated covar. matrix from calibration data

Same fit. SSE_{shape}=91, SSE_{shape} = 1050



Objective Function:

select θ to minimize

$$Q(\theta) = \lambda \cdot \underline{SSE_{fit}(\theta)} + \underline{SSE_{shape}(\theta)}$$

quadratic in θ, includes $\{{}_nF_x\}$

quadratic in θ

Unique, closed-form solution

$$\theta^* = \mathbf{C}\ \{{}_nF_x\}$$

$$f^* = B\,\theta^* = \mathbf{K}\ \{{}_nF_x\}$$

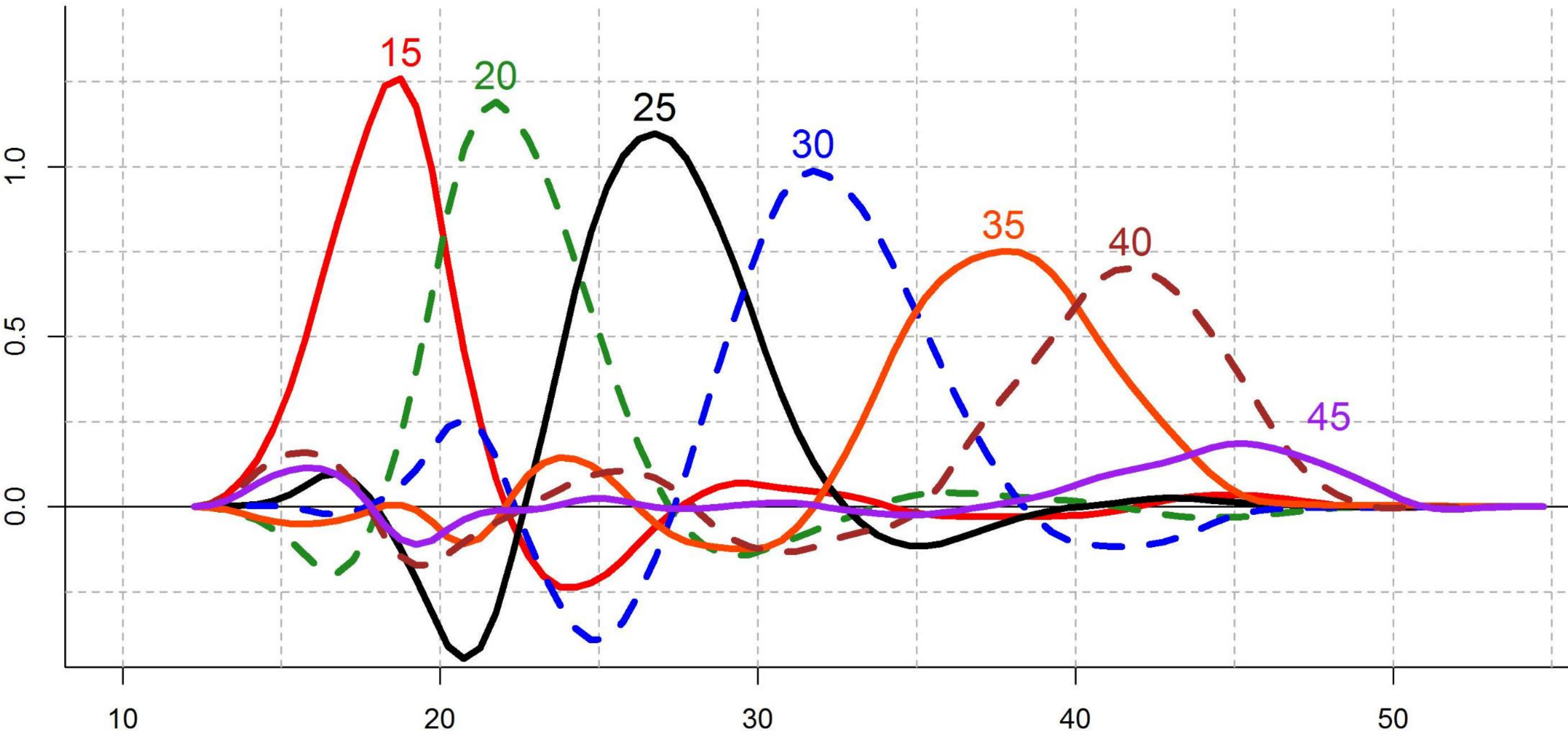
A x 1 **A x 7** **7 x 1**

A simple final product

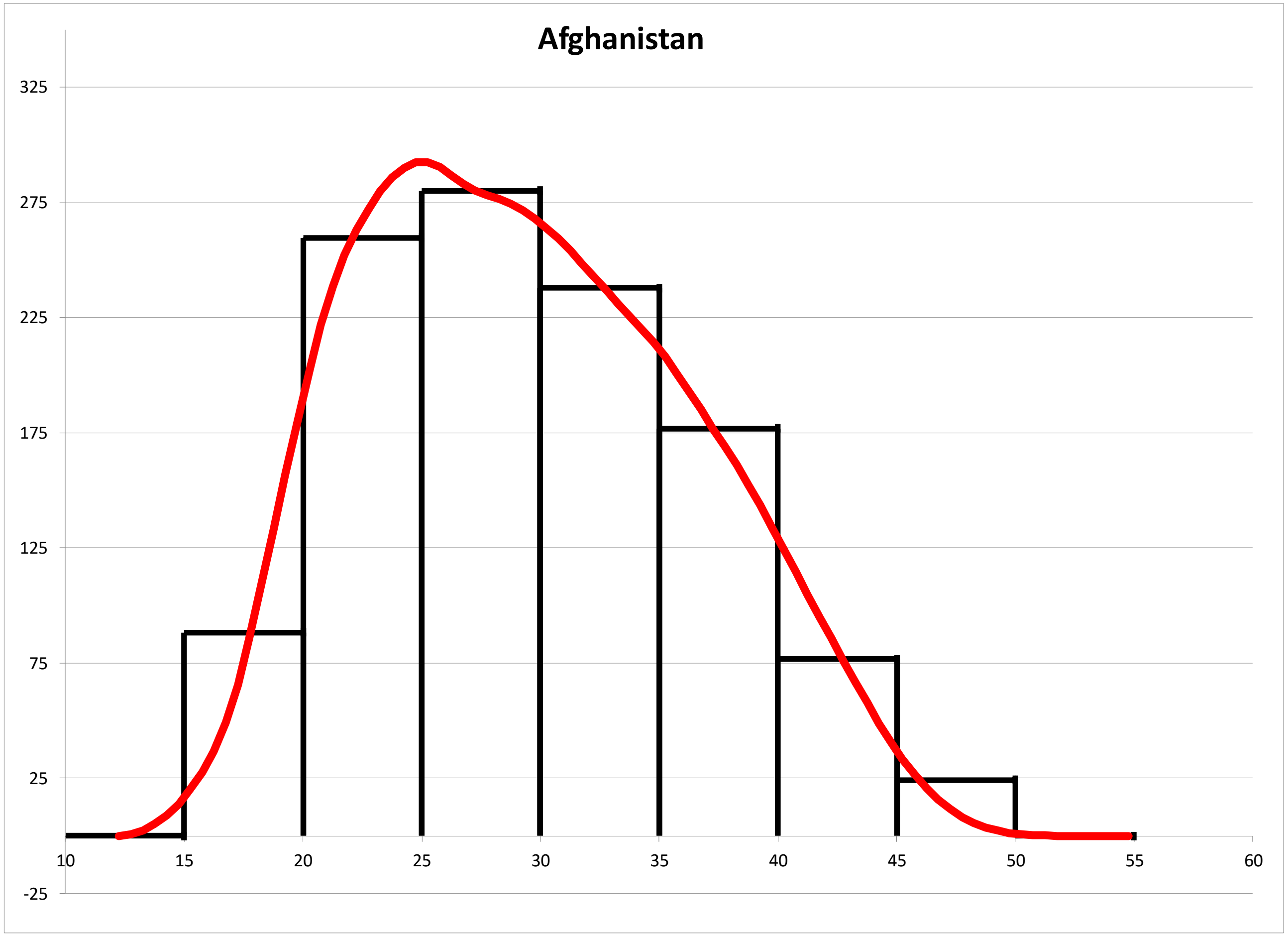
- Fitted schedule is a wtd sum of **K**'s columns
- Weights = observed ${}_nF_x$ values

$$\mathbf{f} = \mathbf{k}_1 \cdot {}_5F_{15} \ldots + \mathbf{k}_7 \cdot {}_5F_{45}$$

Columns of **K** matrix (Calibrated splines)



Examples (change windows, Carl!)

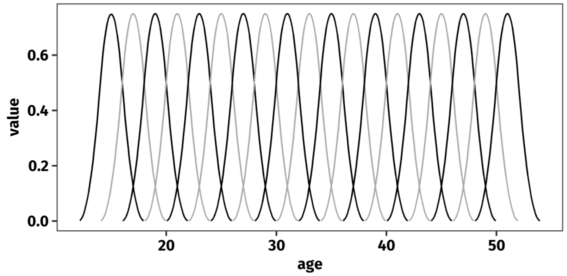


Issues/Problems

- Graduated schedule does not *exactly* match age-group rates (unless **λ** $\rightarrow \infty$)

- Graduated rates occasionally have (very!) small negative values at highest and lowest ages

WPP Modeling Choices



- Number of spline knots for **B**
- Order of splines for **B** (quadratic, cubic, …)
- Age grid ($x_1 \ldots x_A$) for discretizing f(x)
- CALIBRATION DATA (HFD, HFC, smoothed WPP,…)
- Number of principal shape components (3,4,…)
- Relative weight on fitting errors **λ** (should vary with sample size, data quality)
- Add a smoothing penalty? (e.g. squared 2^{nd} diffs in rates)

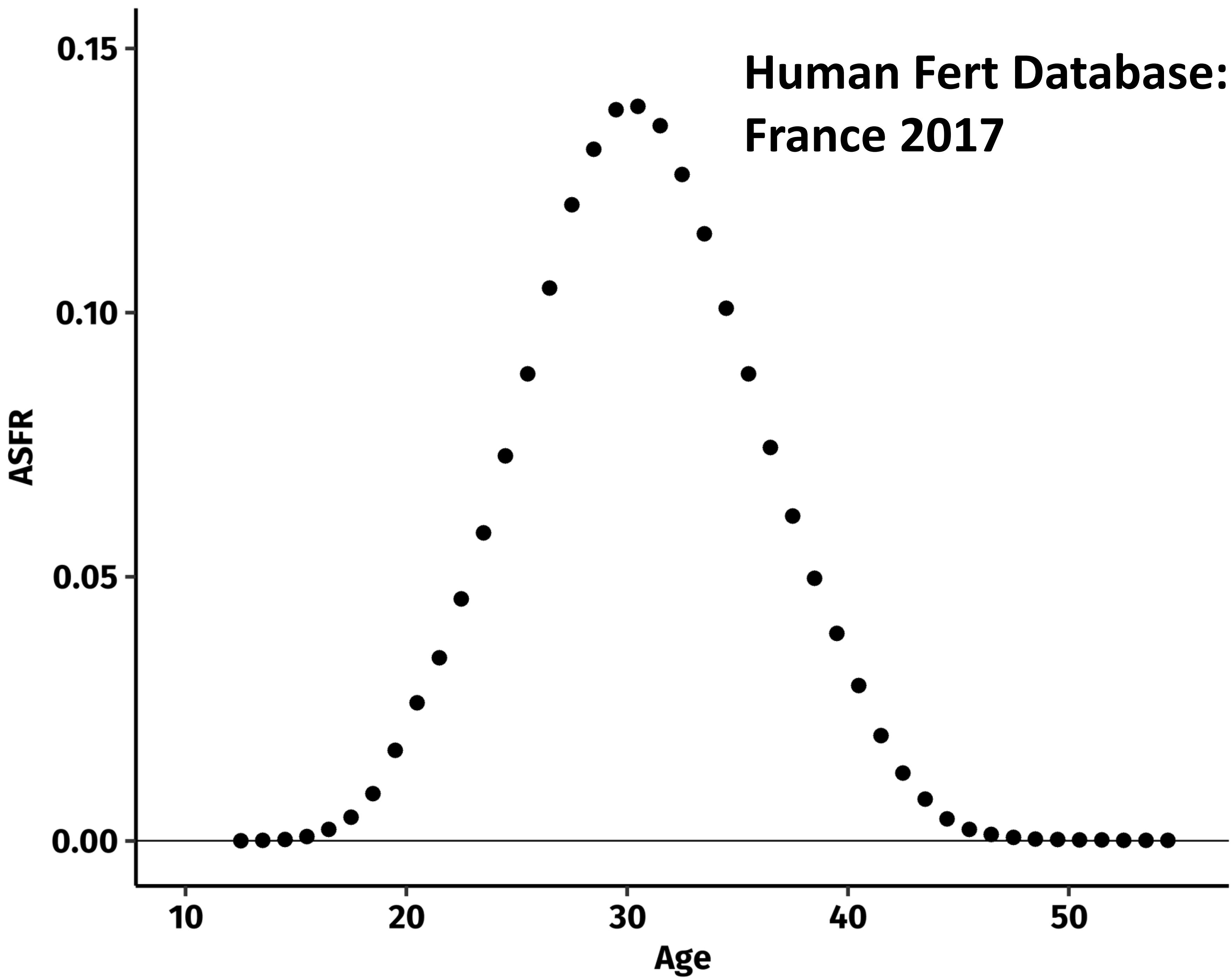
THANKS!

Carl Schmertmann

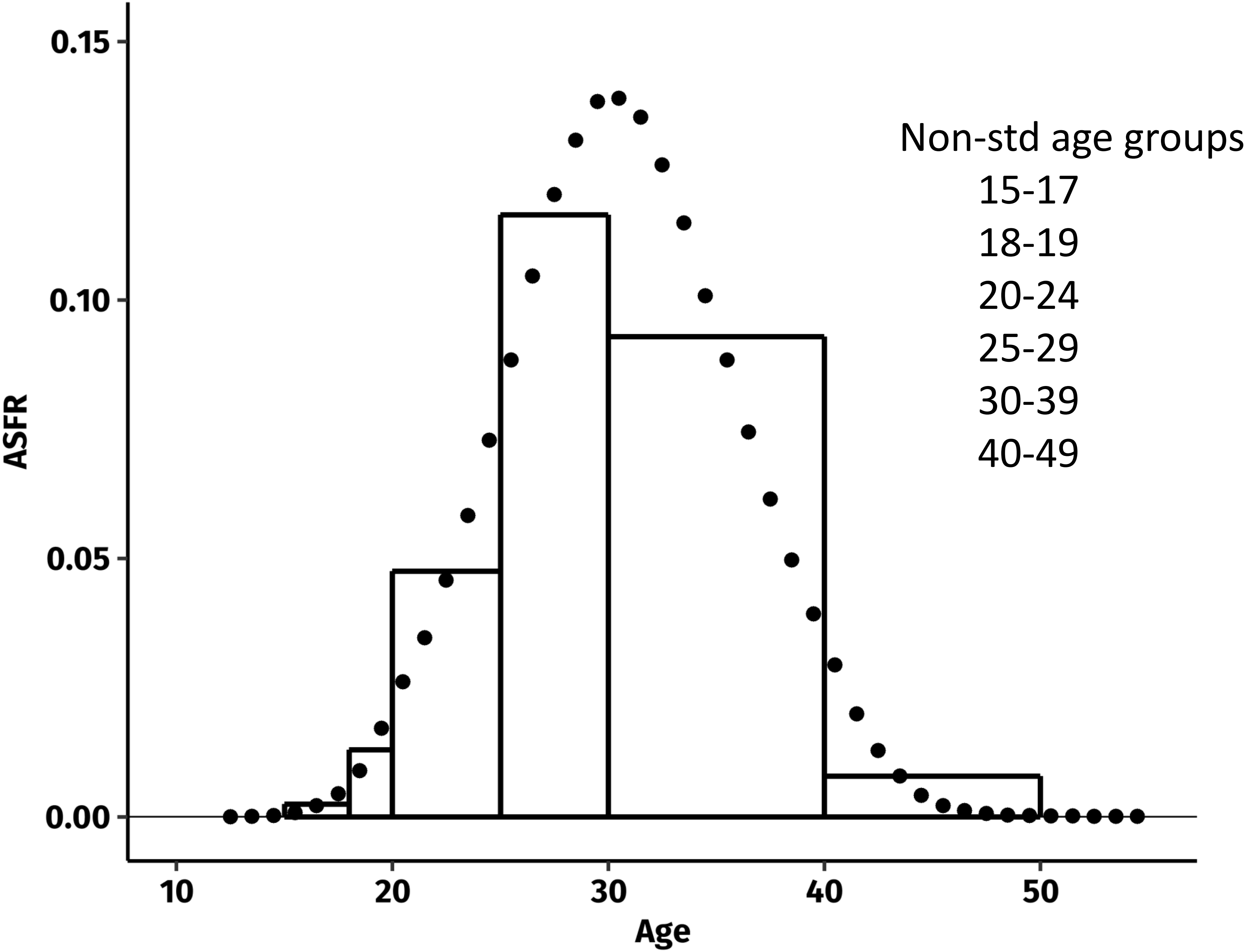
Article at tinyurl.com/fertility-splines

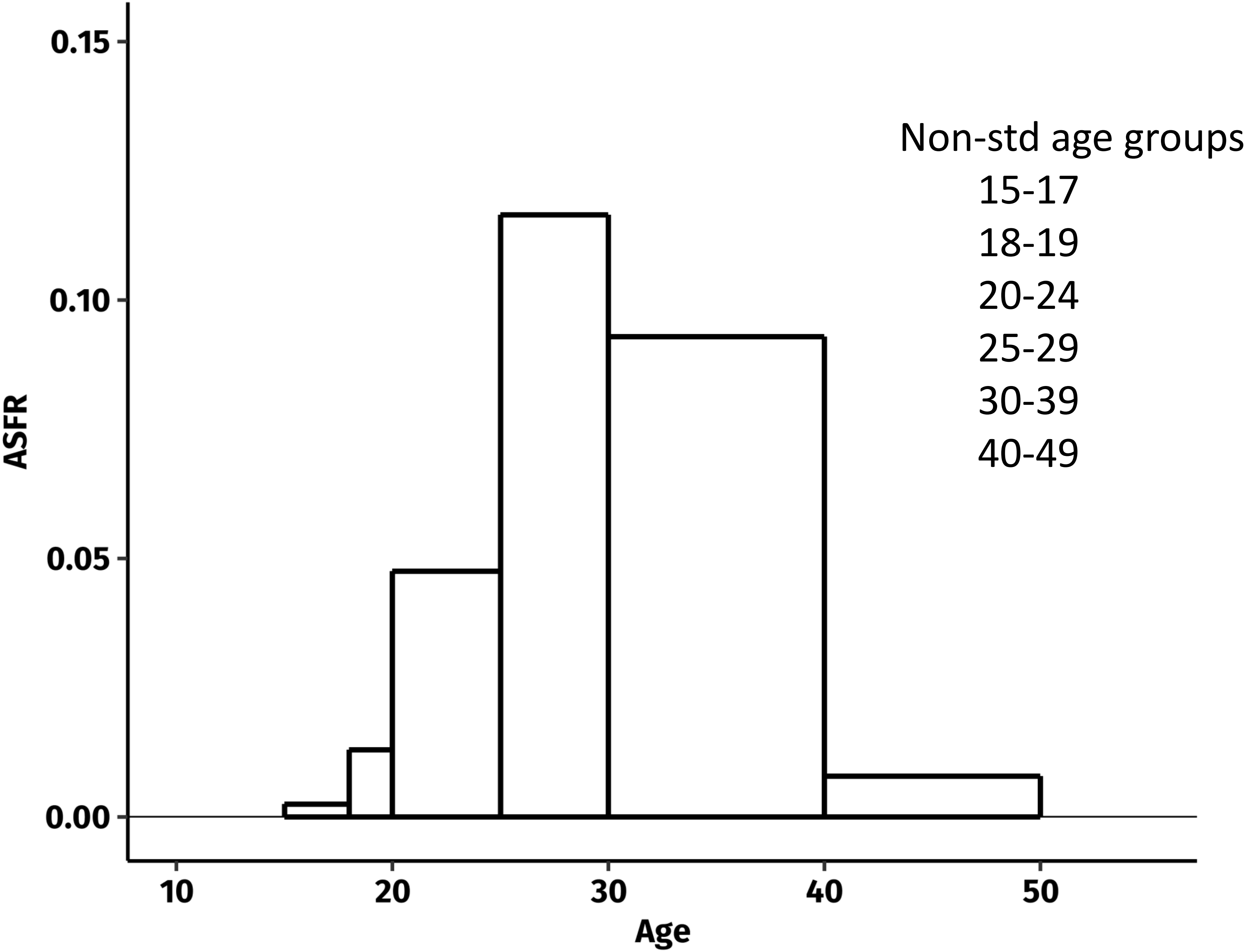
Data/code at tinyurl.com/fertility-splines-replication

Non-standard age groups
(same procedure, different **G** matrix)

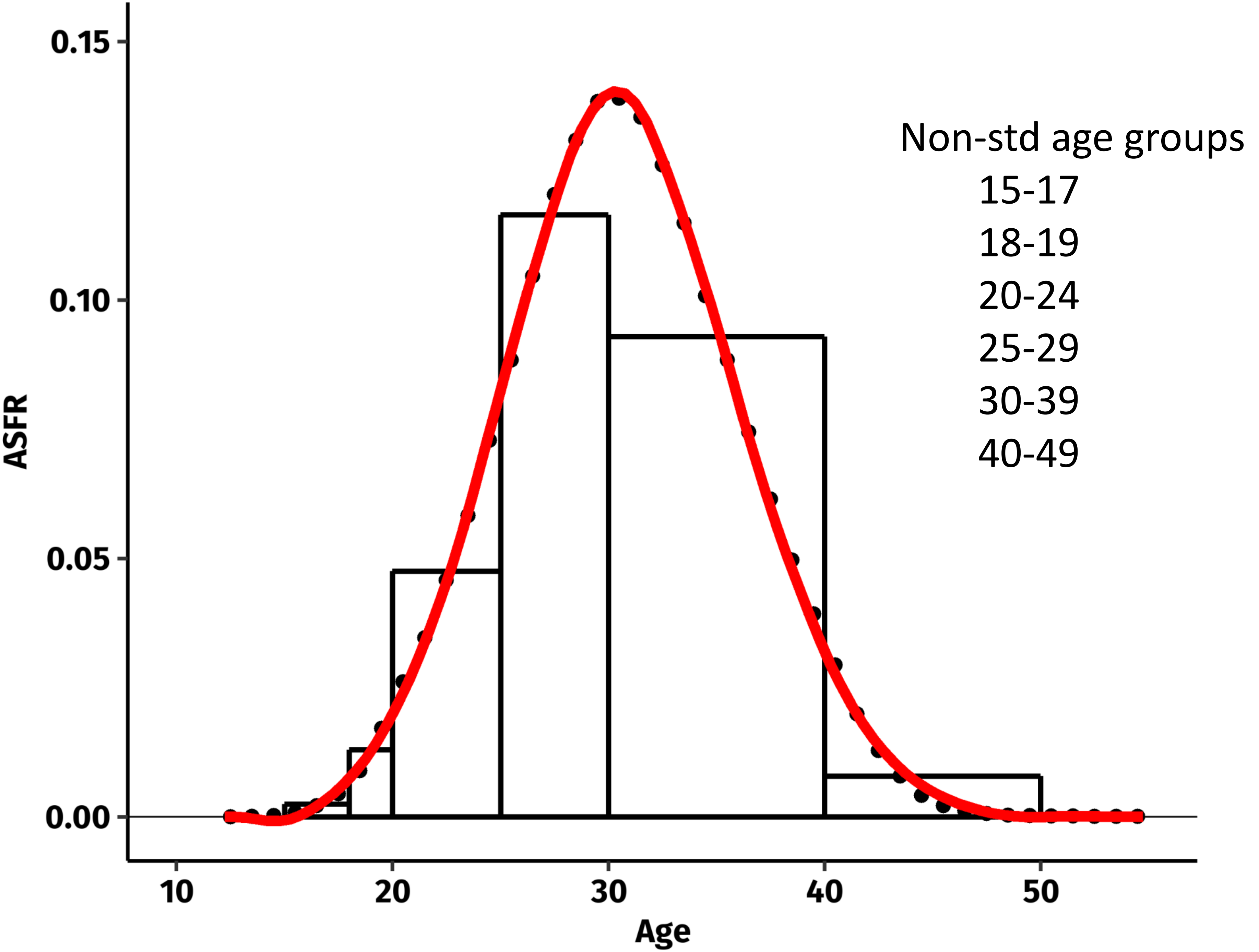


Human Fert Database:
France 2017
ASFR
0.15
0.10
0.05
0.00
10
20
30
40
50
Age


0.15
0.10
0.05
0.00
ASFR
10
20
30
40
50
Age
Non-std age groups
15-17
18-19
20-24
25-29
30-39
40-49



ASFR
0.15
0.10
0.05
0.00
10
20
30
40
50
Age
Non-std age groups
15-17
18-19
20-24
25-29
30-39
40-49



Non-std age groups
15-17
18-19
20-24
25-29
30-39
40-49
0.15
0.10
0.05
0.00
ASFR
10
20
30
40
50
Age