A two-component regression model for the number of births

Gillian Heller
Macquarie University

ViCBiostat
22 Sept 2016

The data

- Rural women's health clinic in western Fiji
- Cross-sectional study 2013-14
- n = 5,136 women aged ≥ 18 years

iTaukei (native Fijians)	48%
Fijians of Indian Descent (FID)	52%

The data

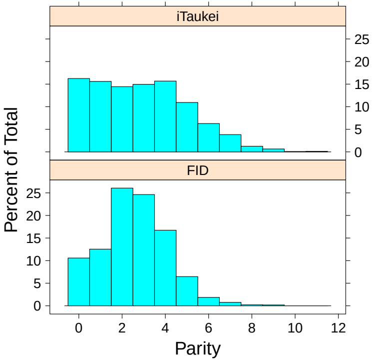
- Rural women's health clinic in western Fiji
- Cross-sectional study 2013-14
- $n = 5{,}136$ women aged ≥ 18 years

iTaukei (native Fijians)	48%
Fijians of Indian Descent (FID)	52%

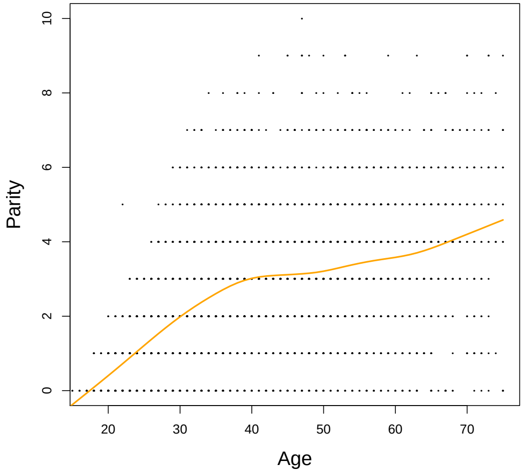
- demographic, clinical variables observed
- two ethnic groups compared

Number of births (parity)

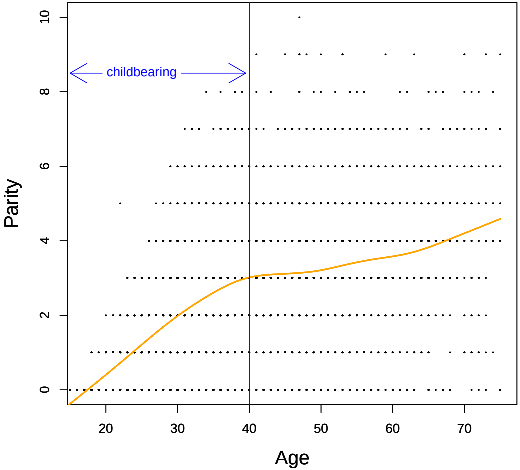
- Parity compared across ethnic groups



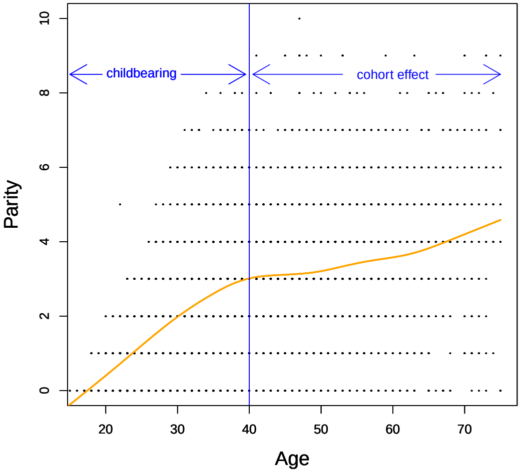
Relationship with age



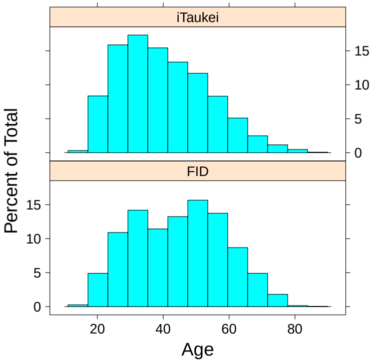
Relationship with age



Relationship with age



Relationship with age



Childbearing effect

Relationship of parity with age

Childbearing effect

Relationship of parity with age

- Starts at zero
- Increases monotonically during the childbearing years
- Levels off once fertility ceases ($\simeq$ 40 years)

Childbearing effect

Relationship of parity with age

- Starts at zero
- Increases monotonically during the childbearing years
- Levels off once fertility ceases ($\simeq$ 40 years)
- S-curve
- We will model this part of the curve parametrically

Generalized logistic function

$$y(x) = \frac{k}{(1 + e^{-bx})^w}$$

- lower asymptote $= 0$
- k = upper asymptote
- b, w are slope and shift parameters
- $k = b = w = 1$ gives the standard logistic function

Generalized logistic function

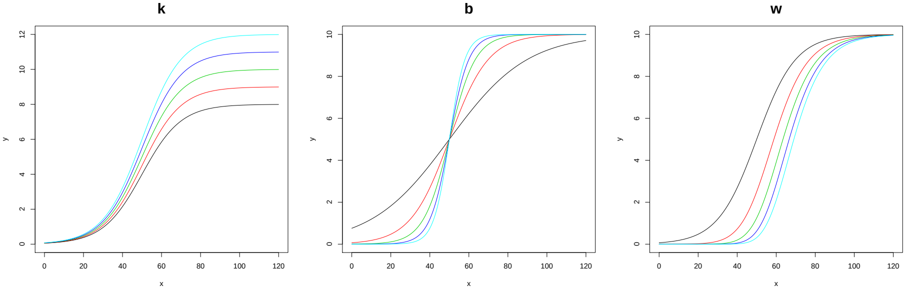
$$y(x) = \frac{k}{(1 + e^{-bx})^{w}}$$

- lower asymptote $= 0$
- k = upper asymptote
- b, w are slope and shift parameters
- $k = b = w = 1$ gives the standard logistic function

Generalized logistic function

$$y(x) = \frac{k}{(1 + e^{-bx})^{w}}$$

- lower asymptote $= 0$
- k = upper asymptote
- b, w are slope and shift parameters
- $k = b = w = 1$ gives the standard logistic function



Cohort effect

- Changes in parity after the childbearing years due to a cohort effect
- appears to be a postive trend, but
- we don't want to put a parametric model on this trend
- use a smooth term

The regression model

Negative binomial regression model for parity (y):

$$y \,|\, x \sim \mathsf{NB}(\mu, \sigma)$$

$$\mu = \frac{k}{(1 + e^{-bx})^{w}} + s(x)$$

- x = age, $k > 0$, $b > 0$, $w > 0$

The regression model

Negative binomial regression model for parity (y):

$$y \mid x \sim \mathsf{NB}(\mu, \sigma)$$

$$\mu = \frac{k}{(1 + e^{-bx})^{w}} + s(x)$$

- x = age, $k > 0$, $b > 0$, $w > 0$
- for stability of computation we impose the constraint $s(x) \geq 0$
- $s(x)$ modelled with exponentiated B-splines.

The regression model

Negative binomial regression model for parity (y):

$$y \,|\, x \sim \mathsf{NB}(\mu, \sigma)$$

$$\mu = \frac{k}{(1 + e^{-bx})^{w}} + s(x)$$

- x = age, $k > 0$, $b > 0$, $w > 0$
- for stability of computation we impose the constraint $s(x) \geq 0$
- $s(x)$ modelled with exponentiated B-splines.
- Could call this a semiparametric model *but* it is not semiparametric in the usual sense.

The regression model

- A two-stage iterative procedure for likelihood maximization is used:

  1. parameters of the growth curve are estimated for fixed spline;
  2. parameters of the spline are estimated for fixed growth curve.

- R function `optim`

The regression model

- A two-stage iterative procedure for likelihood maximization is used:

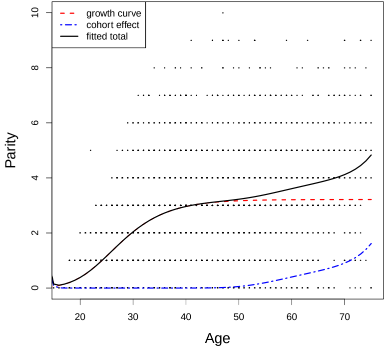
  1. parameters of the growth curve are estimated for fixed spline;
  2. parameters of the spline are estimated for fixed growth curve.

- R function `optim`

- stable results obtained by initially setting the curve $s(x)$ to zero, and allowing the growth curve to dominate the solution in the region of growth due to childbearing.

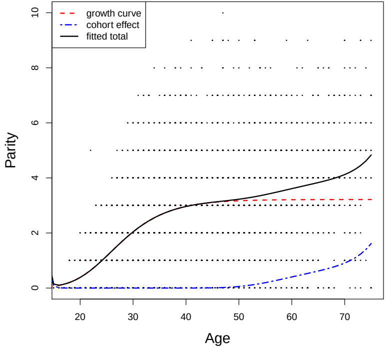
Results

- The model was initially implemented on all women.



Results

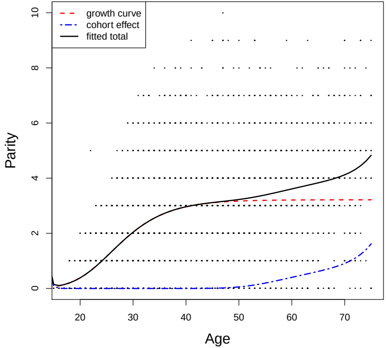
- The model was initially implemented on all women.



- growth curve flattens off around the age of 40
- upper asymptote $\hat{k} = 3.22$

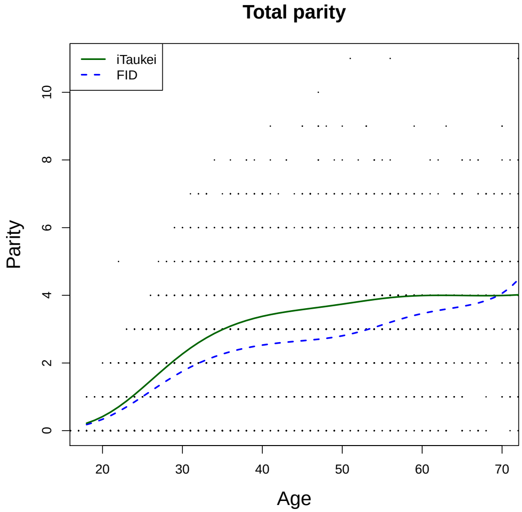
Results

- The model was initially implemented on all women.

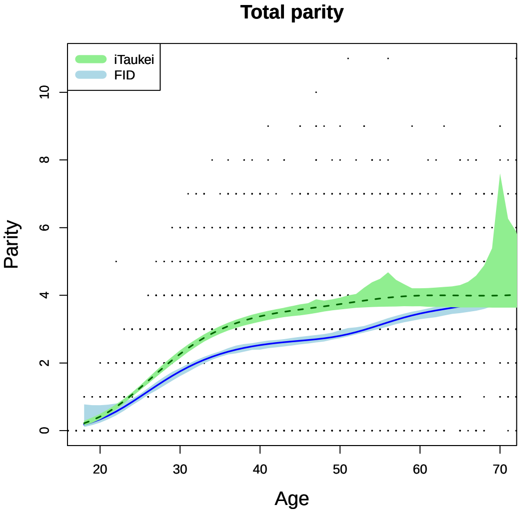


- growth curve flattens off around the age of 40
- upper asymptote $\hat{k} = 3.22$
- cohort effect becomes active around the age of 50
- women currently $\geq$ 50 years were bearing children at a time when rates of birth were higher than in the current cohort of childbearing women

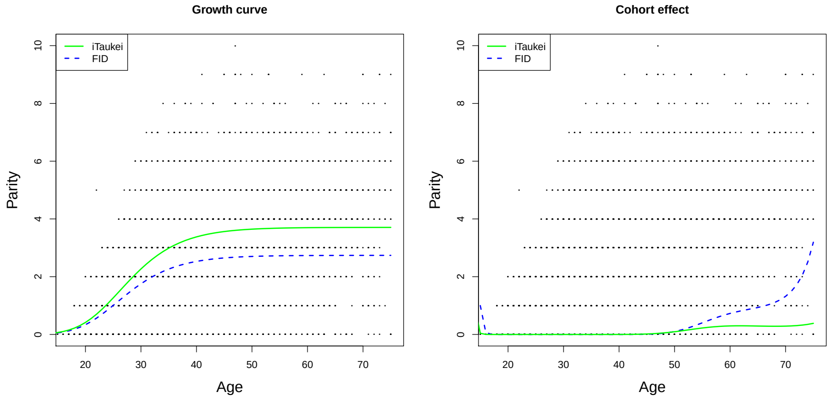
Ethnicity effects



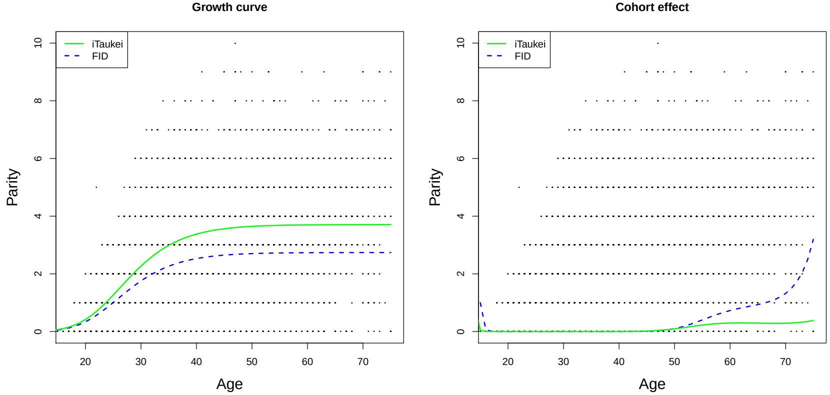
Ethnicity effects



Ethnicity effects



Ethnicity effects



$$\hat{k} = \begin{cases} 3.71 & \text{iTaukei} \\ 2.74 & \text{FID} \end{cases}$$

Ethnicity effects

- Strong cohort effect for FIDs after age 50 years
  - reduction in family size due to increased availability of contraception, better health care

Ethnicity effects

- Strong cohort effect for FIDs after age 50 years
  - reduction in family size due to increased availability of contraception, better health care
- Very small (no?) cohort effect for iTaukei
  - lack of access to health care, contraception

Ethnicity effects

- Strong cohort effect for FIDs after age 50 years
  - reduction in family size due to increased availability of contraception, better health care
- Very small (no?) cohort effect for iTaukei
  - lack of access to health care, contraception
- For the oldest women in our sample, mean parity for both ethnic groups was around 4 births.

Ethnicity effects

- Strong cohort effect for FIDs after age 50 years
  - reduction in family size due to increased availability of contraception, better health care
- Very small (no?) cohort effect for iTaukei
  - lack of access to health care, contraception
- For the oldest women in our sample, mean parity for both ethnic groups was around 4 births.
  - FID: mean parity has decreased to 2.74 births in the current cohort of childbearing women,
  - iTaukei: higher parity persists at 3.71 births

Ethnicity effects

- Strong cohort effect for FIDs after age 50 years
  - reduction in family size due to increased availability of contraception, better health care
- Very small (no?) cohort effect for iTaukei
  - lack of access to health care, contraception
- For the oldest women in our sample, mean parity for both ethnic groups was around 4 births.
  - FID: mean parity has decreased to 2.74 births in the current cohort of childbearing women,
  - iTaukei: higher parity persists at 3.71 births
- Seniloli (1992): ".. fertility is levelling off among Fijians and consistently declining among the Indians in Fiji".

More on the S-curve

- Any S-curve could have been used

More on the S-curve

- Any S-curve could have been used
- All CDFs (of continuous distributions) are S-curves
- This opens up a huge range of possibilities

More on the S-curve

- Any S-curve could have been used
- All CDFs (of continuous distributions) are S-curves
- This opens up a huge range of possibilities

- Derivative of the CDF: density function

More on the S-curve

- Any S-curve could have been used
- All CDFs (of continuous distributions) are S-curves
- This opens up a huge range of possibilities

- Derivative of the CDF: density function
- Parity is cumulative
- Derivative of growth curve gives incidence of births (approximately)

Incidence of births

Generalized logistic function

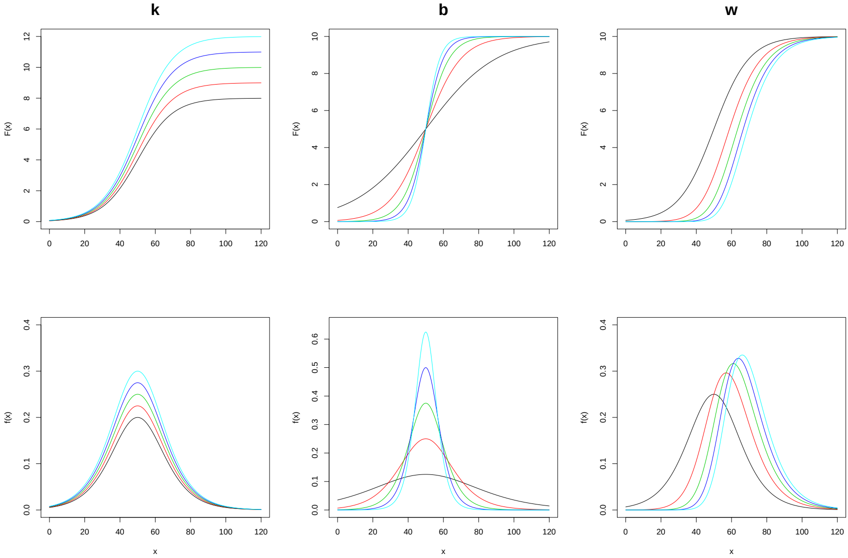
$$F(x) = \frac{1}{(1 + e^{-bx})^{w}}$$

$$F'(x) = f(x) = \frac{b\,w\,e^{-bx}}{(1 + e^{-bx})^{w+1}}$$

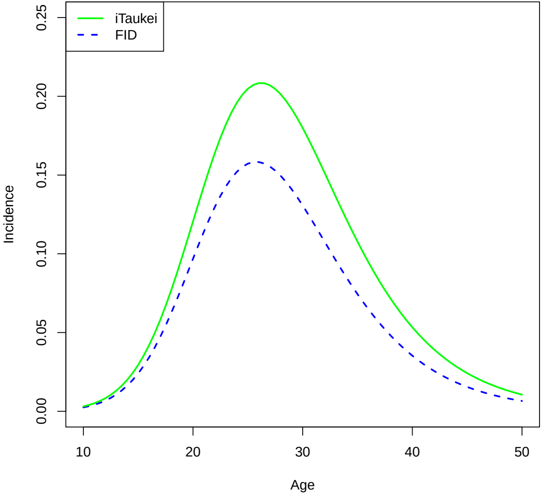
Type I generalized

logistic distribution

Incidence of births



Incidence of births: Fiji data



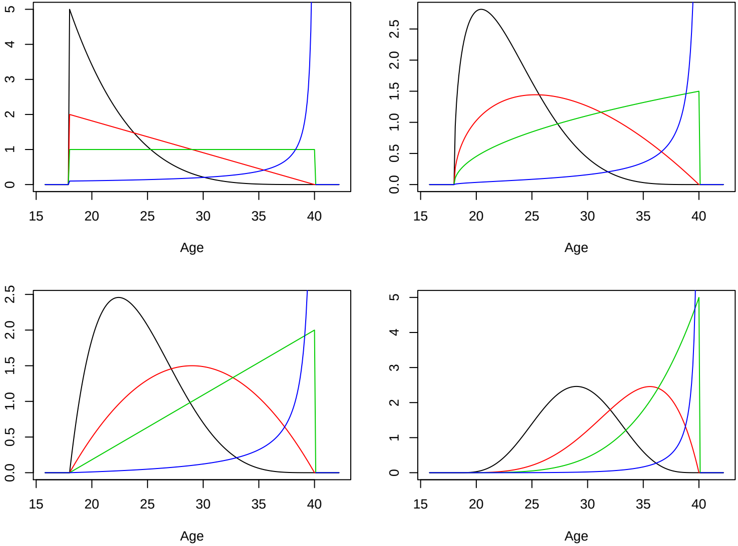
Incidence of births

- We can pick any continuous density to describe the incidence of births
- Symmetry seems an unnecessary assumption

Incidence of births

- We can pick any continuous density to describe the incidence of births
- Symmetry seems an unnecessary assumption
- Beta distribution has desirable features:
  - bounded range;
  - left- or right-skewed.

Beta densities



Childbearing model based on Beta CDF

The S-curve we use is

$$y(x) = k \cdot F(x \,|\, \alpha, \beta, \ell, u)$$

where $F(\cdot)$ is the CDF of the Beta distribution defined on (ℓ, u):

Childbearing model based on Beta CDF

The S-curve we use is

$$y(x) = k \cdot F(x \mid \alpha, \beta, \ell, u)$$

where $F(\cdot)$ is the CDF of the Beta distribution defined on (ℓ, u):

$$f(x) = \frac{1}{B(\alpha,\beta)(u-\ell)^{\alpha+\beta-1}}(x-\ell)^{\alpha-1}(u-x)^{\beta-1} \qquad \ell < x < u$$

Childbearing model based on Beta CDF

The S-curve we use is

$$y(x) = k \cdot F(x \,|\, \alpha, \beta, \ell, u)$$

where $F(\cdot)$ is the CDF of the Beta distribution defined on (ℓ, u):

$$f(x) = \frac{1}{B(\alpha,\beta)(u-\ell)^{\alpha+\beta-1}}(x-\ell)^{\alpha-1}(u-x)^{\beta-1} \qquad \ell < x < u$$

We expect the interval (ℓ, u) to be the childbearing years, i.e. something like (18, 40).

Childbearing model based on Beta CDF

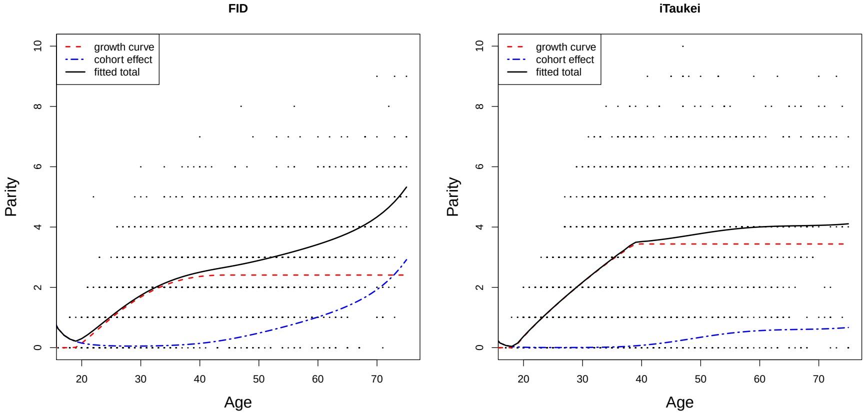
Parameters in childbearing component of model: $k, \alpha, \beta, \ell, u$

Childbearing model based on Beta CDF

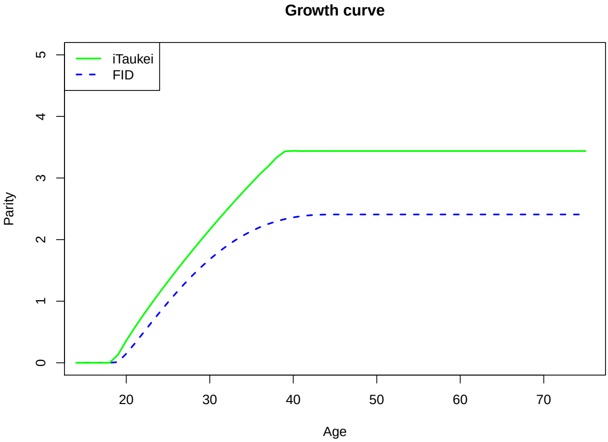
Parameters in childbearing component of model: $k, \alpha, \beta, \ell, u$

Parameter	iTaukei	FID
$\hat{k}$	3.4	2.4
$\hat{\alpha}$	0.9	1.2
$\hat{\beta}$	1.1	2.6
$\hat{\ell}$	18.5	18.9
$\hat{u}$	39.0	45.0

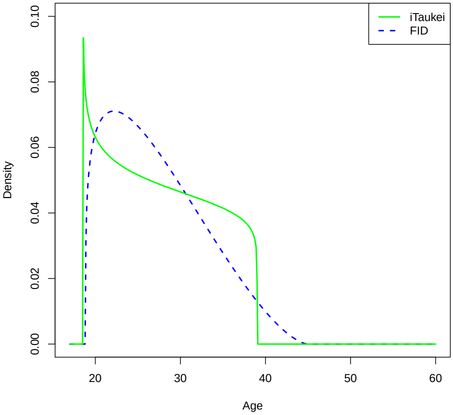
Childbearing model based on Beta CDF



Childbearing model based on Beta CDF



Childbearing model based on Beta CDF



1970 US census data

- Retherford & Cho (1978) report 1970 US census data
- parity by age for women aged 15-50 years
- $n \simeq 500{,}000$

1970 US census data

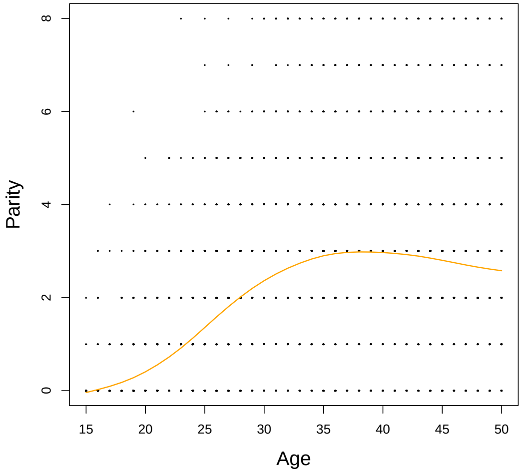
- Retherford & Cho (1978) report 1970 US census data
- parity by age for women aged 15-50 years
- $n \simeq 500{,}000$
- parity ≥ 8 reported as "8+"

1970 US census data

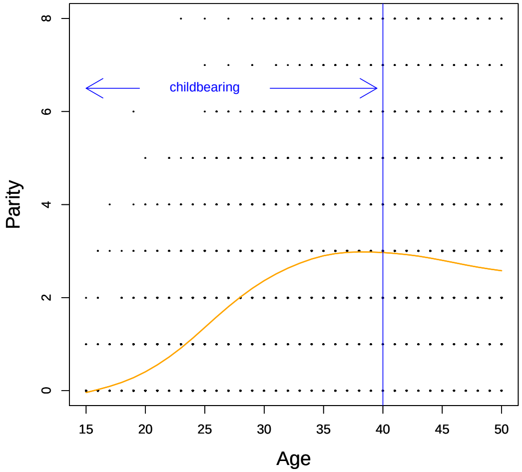
TABLE B.2. *Women by single years of age and parity, United States 1970 census, one per cent public use sample*

	Parity								
Age	0	1	2	3	4	5	6	7	8+
15	19416	221	63	0	0	0	0	0	0
16	18297	544	90	31	0	0	0	0	0
17	17297	1105	167	34	17	0	0	0	0
18	16252	1812	350	58	21	7	0	0	0
19	14359	2627	656	138	33	9	11	0	0
20	12834	3432	1114	264	71	28	10	9	0
21	10727	4128	1672	417	111	32	7	5	1
22	9560	4568	2451	734	199	43	10	8	10
23	7914	4682	3213	989	272	80	30	14	29
24	5269	3555	3115	1113	383	145	37	19	15
25	4554	3358	3654	1543	552	202	59	24	25
26	3850	3105	4105	1982	757	262	114	24	35
27	3310	2940	4436	2399	1043	392	126	51	40
28	2618	2276	3828	2394	1111	458	155	88	38
29	2094	1886	3576	2705	1259	506	224	89	82
30	1981	1709	3327	2838	1490	696	328	124	123
31	1623	1482	2975	2745	1563	702	342	168	122
32	1565	1327	2917	2697	1679	813	391	166	174
33	1377	1198	2668	2587	1719	881	363	198	192
34	1366	1209	2607	2558	1697	905	464	250	263
35	1311	1137	2625	2616	1718	980	486	286	304
36	1303	1058	2358	2442	1710	946	541	302	333
37	1396	1076	2439	2511	1833	1016	559	294	358
38	1317	1153	2463	2458	1748	952	535	300	378
39	1437	1183	2595	2516	1739	1038	531	315	430
40	1493	1283	2646	2553	1801	1020	582	303	533
41	1495	1331	2662	2484	1652	914	556	299	453
42	1585	1380	2757	2467	1740	938	553	321	485

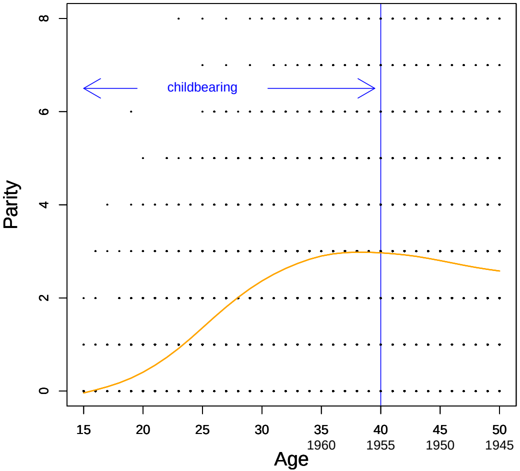
1970 US census data



1970 US census data

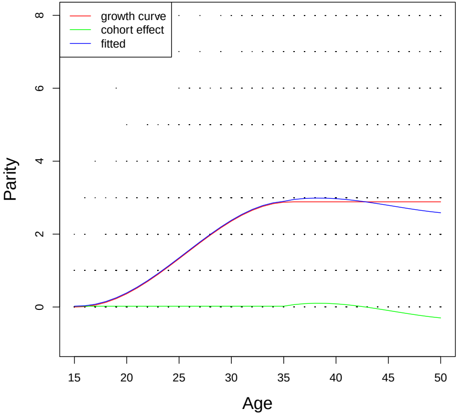


1970 US census data



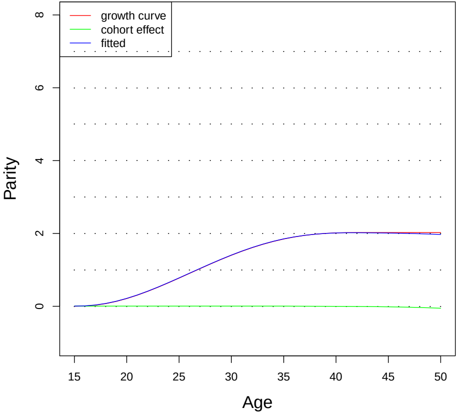
1970 US census data

Beta model



$$\hat{k} = 2.89$$

US 2006 data

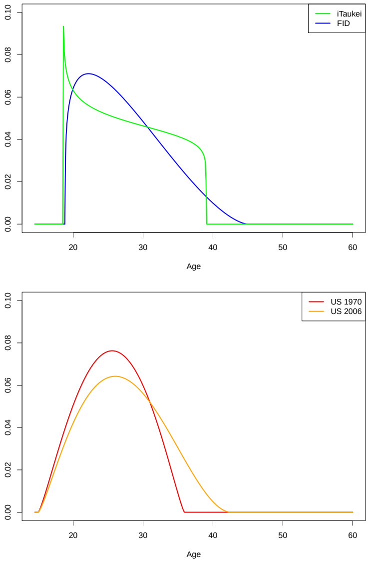


$$\hat{k} = 2.02$$

Comparing birth incidences over studies

	Fiji		US	
Parameter	iTaukei	FID	1970	2006
$\hat{k}$	3.4	2.4	2.9	2.0
$\hat{\alpha}$	0.9	1.2	2.2	2.3
$\hat{\beta}$	1.1	2.6	2.2	3.0
$\hat{\ell}$	18.5	18.9	15.0	15.0
$\hat{u}$	39.0	45.0	35.9	42.6

Comparing birth incidences over studies



References

Seniloli, K. (1994). Fertility and family planning in Fiji. *Espace, populations, sociétiés*, 12(2), 237-244.

Retherford, R. D., & Cho, L. J. (1978). Age-parity-specific birth rates and birth probabilities from census or survey data on own children. *Population Studies*, 32(3), 567-581.