



OPEN

## Experiments in modeling recent Indian fertility pattern

Ujjaval Srivastava<sup>1,2</sup>✉, Kaushalendra Kumar Singh<sup>1</sup>, Anjali Pandey<sup>3</sup> & Neeraj Narayan<sup>1</sup>

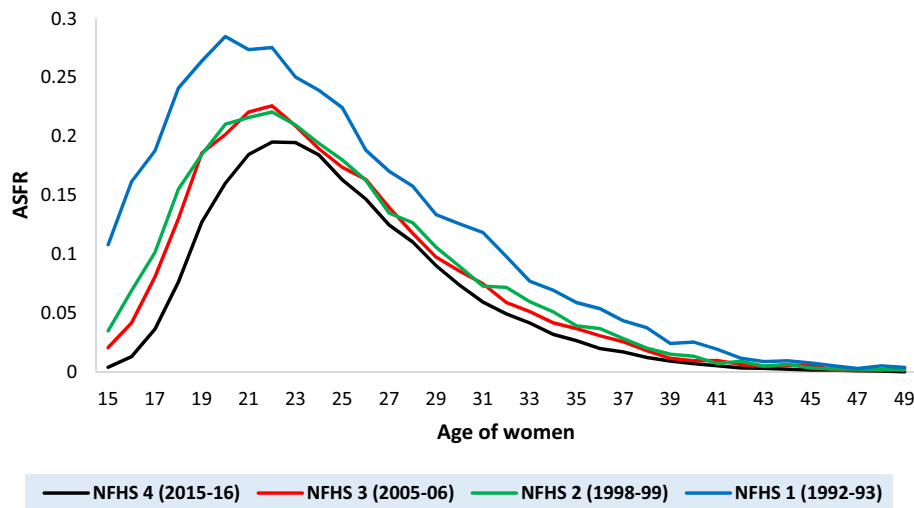
Modelling is a well-established concept for understanding the typical shape and pattern of age-specific fertility. The distribution of India's age-specific fertility rate (ASFR) is unimodal and positively skewed and is distinct from the ASFR of the developed countries. The existing models (P-K model, Gompertz model, Skew-normal model and G-P model considered here) that were developed, based on the experiences of the developed countries, failed to fit the single-year age-specific fertility pattern for India as a whole and for the six selected states. Our study has proposed four flexible models, to capture the diverse age pattern of fertility, observed in the Indian states. The proposed models were compared in three ways; among themselves, with the original models and with the popular Hadwiger model. The parameters of these proposed models were estimated through the Non-Linear Least Squares Method. To find the model with best fit, we used the corrected version of Akaike's Information Criterion (AICc). Optimization of the four original models was successfully done. When the model was fitted to the empirical data of the 4th round of the National Family Health Survey conducted in 2015–2016, the results of this study showed that all the four proposed models outperform their corresponding original models and the Hadwiger model. When comparison among the proposed models was done, the Modified Gompertz Model provided the best fit for India, Uttar Pradesh and Gujarat. Whereas, the Modified P-K model gave the best fit for West Bengal, Tripura and Karnataka. The Modified G-P model is the most suitable model for Punjab. Although our proposed models illustrated the fitting of ASFR for India as a whole and the selected six states only, it provides an important tool for the policymakers and the government authorities to project fertility rates and to understand the fertility transitions in India and various other states.

### Abbreviations

NFHS	National Family Health Survey
DHS	Demographic Health Survey
IIPS	International Institute of Population Sciences
MoHFW	Ministry of Health and Family Welfare
GFR	General fertility rate
TFR	Total fertility rate
ASFR	Age-specific fertility rate
SSE	Sum of square due to error
AIC	Akaike's information criterion
AICc	Corrected version of Akaike's information criterion
LOWESS	Locally estimated scatterplot smoothing

The optimal way of unfolding the fertility and determining its curve through modeling, has been a crucial issue in the demographers' research. Modelling fertility behaviour is not only practicable in estimating fertility indicators such as the General fertility rate (GFR), Total fertility rate (TFR), Age-specific fertility rate (ASFR), but also beneficial for forecasting and fertility projection that facilitate the population projections. In literature, there are a variety of parametric fertility models to explain the behaviour of the fertility pattern. The class of models with a certain built-in probability density function includes the Hadwiger model<sup>1,2</sup>, Hadwiger mixture model<sup>3</sup>, Coale-Trussell model<sup>4–6</sup>, Beta and Gamma models<sup>6</sup>, Pearson Type 1 curve<sup>7,8</sup> and Type III curves<sup>9</sup>, P-K model and P-K mixture model<sup>10</sup>, Flexible Generalised skew-Normal model<sup>11</sup>, The scaled Weibull mixture model<sup>12</sup>, logistic model<sup>13</sup>, Skew logistic model<sup>7,14</sup>. Other researchers have designed their parametric models using mathematical

<sup>1</sup>Department of Statistics, Banaras Hindu University, Varanasi, India. <sup>2</sup>Ministry of Tourism, Government of India, New Delhi, India. <sup>3</sup>Department of Biostatistics (Main), All India Institute of Medical Sciences, New Delhi, India. ✉email: [ujjavalsri777@gmail.com](mailto:ujjavalsri777@gmail.com)



**Figure 1.** The age-specific fertility rate (ASFR) of India as per different National Family Health Survey (NFHS) rounds.

functions. These include the Gompertz model<sup>15–21</sup> and the adjusted error model<sup>22</sup>. All the fertility models have their advantages and work as best fit under different conditions.

Over the years, it has been observed that the shape of the ASFR curve of the developing and the developed countries has changed significantly. India has a moderate level of fertility and an Early-peak type fertility schedule (when peak fertility is observed among women of the age group 20–24 years<sup>23</sup>). A rightward shift in the age of extreme fertility and a downward shift in the magnitude of extreme fertility has been observed in Indian ASFR (Fig. 1). The contribution of the adolescents' fertility (fertility among women aged 15–19 years) in the total fertility dropped from 24 percent (NFHS-1<sup>24</sup>) to 12 percent (NFHS-4<sup>27</sup>). However, the overall structure of the Indian fertility schedule is unaltered over time. It is pertinent to mention that the pattern and the level of fertility in the developed countries are very different from the developing countries like India. In recent years, ASFR patterns of developed countries like the United Kingdom, the United States, and Ireland experienced one more bulge for the early age fertility apart from the extreme fertility. Chandola et al.<sup>3</sup> conjectured that the differences in the fertility curve of the modern population might be linked to a mixture of two sub-populations (marital and non-marital), having different fertility patterns. Peristera and Kostaki<sup>10</sup> also found this aberration pattern in early births in the other European countries. To account for such heterogeneity, Peristera and Kostaki<sup>10</sup>, fit their model on order-specific fertility rates. For the United States population, they fit separate models for the Black and the White population. Mixed models can be used to model this fertility pattern. Hadwiger mixture model<sup>3</sup>, The P-K mixed model<sup>10</sup>, the flexible generalized skew-normal model<sup>11</sup>, and the scaled Weibull mixed model<sup>12</sup> will provide a better fit for these type of fertility patterns.

The other classes of fertility models are non-parametric. Polynomial model<sup>28–30</sup>, Spline model<sup>31–35</sup> and the other models<sup>36</sup> are non-parametric models of the age-specific fertility schedules. Although polynomial models are simple and easy to handle, they are not quite popular due to the non-interpretability of their parameters. On the other hand, the splines model requires the estimation of more number of parameters, which is unclear.

In recent literature, there have been only a handful of studies<sup>7,28,29</sup>, that attempt to fit the age-specific fertility pattern in the Indian context. To the best of our knowledge, no study has ever tried to fit a single-year, age-specific Indian fertility pattern, through parametric models. The existing parametric models that were developed based on the experiences of the developed countries, failed to fit the single-year ASFR of India as a whole and the six selected states. This provides a rationale for the present study. This study proposes four flexible models to capture the diverse fertility pattern observed in the Indian states. Evaluation of these models is done by fitting the ASFR curve of the six selected Indian states and the performance of each model is also measured by applying these models on the real data.

## Data and methodology

The present study used the data from the National Family Health Survey (NFHS-4), conducted in 2015–2016. The NFHS is a national representative cross-sectional survey conducted for the fourth time in India, as part of The Demographic Health Survey (DHS) Program. Previously, three rounds (1992–1993<sup>24</sup>, 1998–1999<sup>25</sup> and 2004–2005<sup>26</sup>) have been successfully conducted. The International Institute of Population Sciences (IIPS) conducts the NFHS, with the collaborative efforts of a large group of organizations under the stewardship of the Ministry of Health and Family Welfare (MoHFW), Government of India. NFHS provides major indicators like mortality, fertility, maternal and child health indicators, family planning, child nutrition, domestic violence, morbidity estimates for several diseases etc. at the national and the state level. This is for the first time, that the NFHS-4 has given estimates of major indicators for 640 districts of India (as per the 2011 Census of India). NFHS-4 also provides information on several new and emerging issues, including insurance coverage, use of

Region	Name of states	Randomly selected states
North	Haryana, Himachal Pradesh, Jammu and Kashmir, Punjab, Uttarakhand	Punjab
Central	Chattisgarh, Madhya Pradesh, Uttar Pradesh	Uttar Pradesh
East	Bihar, Jharkhand, Odisha, West Bengal	West Bengal
Northeast	Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura	Tripura
West	Rajasthan, Gujarat, Maharashtra, Goa	Gujarat
South	Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana	Karnataka

**Table 1.** Regional classification of Indian states.

mosquito nets for malaria prevention, abortion, HIV testing during antenatal care, ownership of physical and economic assets by women, and domestic violence during pregnancy. The study sample for NFHS-4 is based on the stratified two-stage sampling of households. A total of 6,99,686 eligible women of the age-group 15–49 years and 1,12,122 eligible men of the age-group 15–54 years, were successfully interviewed using the Computer Assisted Personal Interviewing method. For more details regarding the survey design and the questionnaires, one can refer to the NFHS-4 report<sup>27</sup>.

Amongst all the fertility indicators, the ASFR secures a crucial place because of its role in forecasting the population by the cohort component projection method. Though, the ASFR estimated from the empirical data is a good portrayal of the existing fertility behaviour of the population, using it for the projection of future population requires an afterthought. Owing to its ability, to pragmatically translate the assumptions about the future fertility behaviour, modeling of ASFR is of considerable interest to the demographers. If someone wants to analyse the mean age of childbearing or wants to know the proportion of births by mothers belonging to the adolescent age group (15–19) years or by mothers over the age of 35 years, information cannot be directly obtained through ASFR summary measure. There is also much information about the variance, kurtosis, and skewness of the ASFR curve of a particular region. Modelling would be necessary for those conditions. Modeling of age-specific fertility is also necessary to understand the phenomenon of age specific fertility transition among different countries. The foremost aim of any model building is to extract all possible information from the available data and to give an accurate picture of all the known and the unknown facets of the phenomena under study<sup>22</sup>. But, the approximation errors of the fertility model, serve as the primary drivers of deviation in the forecasted population<sup>37</sup>. Hence, it is necessary to identify the correct model for fertility schedule.

In order to simulate the fertility schedule, information on the birth histories of women in the age group of 15–49, is extracted from the NFHS-4. Although 5-year ASFR is reported in the national as well as state reports of NFHS-4, we have computed single-year ASFR for the modelling fertility pattern. The reference period for the calculation of single-year ASFR is 3 years before the survey, i.e., only the births that have occurred within 3 years, before the survey date was utilized to compute single-year ASFR. A total of 6,98,185 eligible women were considered in our study. For the calculation of all India and state level single-year ASFR, we used STATA `tfr2` module as described by Schoumaker<sup>38</sup>.

To find the true nature of the fertility schedule, it is better to choose a single year ASFR than 5-year grouped ASFR. However, single-year ASFR suffers from some fluctuations and some digit preferences at certain ages. To detect the correct trend of fertility patterns suffering from the problem of fluctuations, we have applied the smoothing technique. Here, we used *loess* (*locally estimated scatterplot smoothing*) techniques to smoothen the scatterplot. The advantage of loess smoothing is that it is a bit more flexible than the other smoothing techniques. It does not assume any particular mathematical form, to smoothen the data but allows them to discover the desired form from data itself. We have implemented `loess()` function in R with 0.2 as the value of the smoothing parameter (20% smoothening). The reason for choosing the parameter equal to 0.2 is that, it is neither too small to eliminate fluctuations in the data nor too large to over-smoothen the scatter plot. In this study, we try to fit the fertility pattern of India and the 6 selected Indian states with four proposed models. For this purpose, based on the socio-cultural similarities, we have classified the Indian states into six geographical regions, depending on their locations; north, central, east, northeast, west and south. And from each of these six regions, one representative state was selected randomly (using random number generation). The region wise classification of states and the 6 selected states from their respective region is given in Table 1. The age-specific fertility pattern of India and the six selected states (Punjab, Uttar Pradesh, West Bengal, Tripura, Gujarat and Karnataka) are obtained from NFHS-4.

In order to evaluate our proposed model in comparison to the Hadwiger model, we have fitted Hadwiger model to the empirical fertility of India and the selected states and then compared the estimates. The functional form of the Hadwiger model is as follows:

$$f(x, a, b, c) = \frac{ab}{c} \left(\frac{c}{x}\right)^{\frac{3}{2}} \exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}$$

Here,  $f(x)$  is the ASFR at age  $x$  and  $a, b, c$  are the three parameters. Chandola et al.<sup>3</sup> has discussed the demographic interpretation of these parameters. Parameter  $a$  is related to the total fertility,  $b$  is associated with the height of the ASFR curve and parameter  $c$  is associated with the mean age of childbearing whereas the term  $\frac{ab}{c}$  is associated with the extent of extreme fertility.

**Proposed models.** *Model 1 (Modified P-K Model).* The first model is an extension of the work done by Peristera and Kostaki<sup>10</sup>, in which a constant parameter  $a$  is added to capture the age specific fertility in the later ages (45–49 age-group). The original P-K Model<sup>10</sup> is as follows:

$$f(x, b, \mu, \sigma_1, \sigma_2) = \begin{cases} b \exp\left(\frac{x-\mu}{\sigma_1}\right)^2 & \text{if } x \leq \mu \\ b \exp\left(\frac{x-\mu}{\sigma_2}\right)^2 & \text{if } x > \mu \end{cases}$$

Here  $f(\cdot)$  is the ASFR of women of age  $x$ .  $b$  is the peaked value of fertility,  $\mu$  is the age at the time of peaked fertility,  $\sigma_1$  and  $\sigma_2$  depicts the spread of the ASFR curve, before and after the peaked fertility respectively. A Similar concept was used by Gayawan et al.<sup>22</sup> to fit the ASFR of some African countries and it was referred to as the *Adjusted Error Model (AEM)*. In this model, a constant term is added to the complementary error function.

Our proposed model is defined as follows:

$$f(x, a, b, \mu, \sigma_1, \sigma_2) = \begin{cases} a + b \exp\left(\frac{x-\mu}{\sigma_1}\right)^2 & \text{if } x \leq \mu \\ a + b \exp\left(\frac{x-\mu}{\sigma_2}\right)^2 & \text{if } x > \mu \end{cases}$$

In our model,  $a$  handles fertility in the late ages of the reproductive span. All the other parameters have the same interpretation as explained above. If we put  $a = 0$  in the Modified P-K Model, it reduces to a model proposed by Peristera and Kostaki<sup>10</sup>.

*Model 2 (Modified Gompertz Model).* The second model that is proposed is an extension of Gompertz model, in which we have added one scale parameter  $\gamma$ . The functional form of Gompertz model is taken from Hoem et al.<sup>6</sup> and Mathivanan et al.<sup>39</sup>, which is given as follows:

$$f(x, \alpha, \beta, m) = \left(\frac{\alpha}{\beta}\right) \exp\left[-\left(\frac{x-m}{\beta}\right) - \alpha \exp\left\{-\left(\frac{x-m}{\beta}\right)\right\}\right]$$

The three parameter Gompertz model is not quite flexible. Murphy and Nagur<sup>20</sup> showed that it has fixed kurtosis and therefore, it does not fit well to India as well as to the states' ASFR curves. After inclusion of scale parameter  $\gamma$ , the curve fits very well. The proposed model is defined below:

$$f(x, \alpha, \beta, \gamma, m) = \left(\frac{\alpha}{\beta}\right)^\gamma \exp\left[-\left(\frac{x-m}{\beta}\right) - \alpha \exp\left\{-\left(\frac{x-m}{\beta}\right)\right\}\right]$$

$m$  is defined as the lowest marriage age of women in the population. Our model will reduce to Gompertz model as soon as we put  $\gamma = 1$  in the above model.

*Model 3 (Modified Skew Normal Model).* The third model presented here is a modification of the generalized skew normal function. The original skew normal distribution is defined as follows:

$$f(x, \lambda, \sigma^2, \delta) = 2\sigma^{-1} \phi\left(\frac{x-\lambda}{\sigma}\right) \Phi\left\{\delta\left(\frac{x-\lambda}{\sigma}\right)\right\}$$

where  $\phi$  and  $\Phi$  are the density and the distribution function of the standard normal distribution respectively.  $\lambda$  is the location parameter and  $\sigma$  is the scale parameter.  $\delta$  is the skewness parameter. The detailed property of this function was first studied by Azzalini<sup>40</sup>. In the modified version of this model, the skewness of the ASFR curve is controlled by  $\delta$  and the two scale parameters  $s_1$  and  $s_2$ . The proposed model is defined as follows:

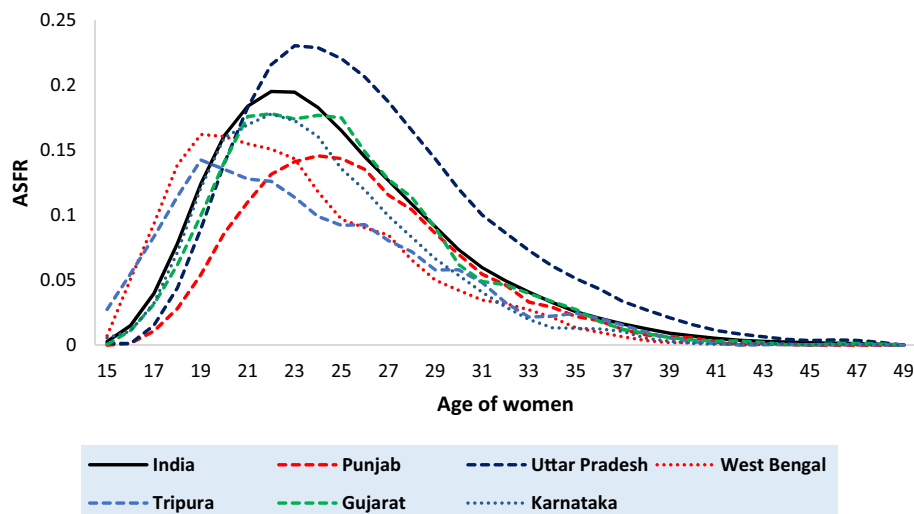
$$f(x, \theta, \lambda, s_1, s_2, \delta) = \begin{cases} \theta \exp\left(-\frac{1}{2}\left(\frac{x-\lambda}{s_1}\right)^2\right) \Phi\left(\frac{\delta x - \lambda}{s_1}\right) & \text{if } x \leq \lambda \\ \theta \exp\left(-\frac{1}{2}\left(\frac{x-\lambda}{s_2}\right)^2\right) \Phi\left(\frac{\delta x - \lambda}{s_2}\right) & \text{if } x > \lambda \end{cases}$$

Here,  $\theta$  reflects the peaked value of the age specific fertility,  $\lambda$  being the age at which the peaked fertility is being achieved.  $s_1$  and  $s_2$  represents the deviation of the ASFR curve, before and after the peaked fertility respectively.

*Model 4 (Modified G-P Model):.* The fourth model that we have proposed is a variation of the behavioral model, suggested by Gupta and Pasupuleti<sup>41</sup>. The original form of G-P Model<sup>24</sup> is as follows:

$$f(x, p, q, r) = px^q \left(1 - \frac{x}{49}\right)^r$$

Here,  $f(\cdot)$  is the ASFR at age  $x$ , parameter  $q$  being responsible for the amplification in the level of the fertility, due to the 'exposure to marriage and sexual union'. Parameter  $r$  is responsible for the decline in the level of fertility due to the 'birth control measures and the biological inability to reproduce'. This model overestimates the early age specific fertility and underestimates peak age specific fertility. Hence, we have introduced a constant parameter



**Figure 2.** Age-specific fertility rate (ASFR) of India and the 6 selected states as per Fourth round of National Family Health Survey (NFHS).

s, to capture the late age specific fertility and to make few scale and location modifications to the variable  $x$ . The proposed model is expressed as follows:

$$f(x, p, q, r, s) = s + p \left( \frac{x-15}{49} \right)^q \left( 1 - \frac{(x-15)}{49} \right)^r$$

**Estimation of parameters of the models.** For the estimation of the proposed models, we have the applied Non-Linear Least Squares Method. Under this method, we have minimized the sum of squares due to error (SSE), with respect to the parameters of the model. SSE function is given by:

$$SSE = \sum_x \left( f(x) - \widehat{f(x)} \right)^2$$

Here  $f(x)$  is the observed ASFR at age  $x$  and  $\widehat{f(x)}$  is the estimated ASFR from the proposed models. We have used `nlminb()` function in R, for non-linear minimization of SSE function. `nlminb()` function estimates the parameters, along with their asymptotic standard errors. In this study, we did not report the asymptotic standard errors as per the convention<sup>10</sup>.

**Performance of the models.** While modelling the ASFR, it is necessary to check the performance of different models. Usually, one can plot the predicted values and the observed values of the ASFR from the model and then choose the best fit among all the models. In quantitative terms, SSE should be calculated for each model and the one with the least SSE will be the best fit model. However, this will not take care of the model complexity, resulting from the inclusion of more parameters. Hence in order to keep the model fit and to consider its complexity, we have used corrected Akaike's Information Criterion (AICc)<sup>42</sup>, to select the best performing model in this study. AICc is Akaike's Information Criterion (AIC)<sup>43</sup> adjusted for small sample sizes. AIC for a particular model adjusts the SSE, with the number of parameters and the number of observations in the model. The formula of AICc is as follows:

$$AIC = 2k + n \ln \left( \frac{SSE}{n-k} \right)$$

$$AICc = AIC + \frac{2k(k+1)}{(n-k-1)}$$

Here,  $k$  denotes the number of parameters present in the model and  $n$  denotes the number of observations. As  $n \rightarrow \infty$ , the results of the AICc converge to AIC. The criteria is to choose the model with the least value of AICc, as the best performing model.

## Results

**Age-specific fertility schedule of India and six selected states.** Figure 2 shows the single year ASFR curves for India and the six selected states (Punjab, Uttar Pradesh, West Bengal, Tripura, Gujarat and Karnataka). All the ASFR curves are unimodal and skewed towards the right (positively skewed), i.e. age-specific

Model	Region	SSE	AICc
P-K Model (K = 4)	India	0.00056	-372.633
	Punjab	0.00032	<b>-392.736</b>
	Uttar Pradesh	0.00085	-358.236
	West Bengal	0.00061	<b>-369.945</b>
	Tripura	0.00061	<b>-370.204</b>
	Gujarat	0.00091	-356.094
	Karnataka	0.00031	<b>-393.476</b>
Modified P-K Model (K = 5)	India	0.00028	<b>-393.008</b>
	Punjab	0.00046	-366.255
	Uttar Pradesh	0.00032	<b>-389.052</b>
	West Bengal	0.00074	-359.33
	Tripura	0.00066	-366.825
	Gujarat	0.00066	<b>-366.374</b>
	Karnataka	0.00086	-353.947

**Table 2.** Minimized sum of square (SSE) and AICc values under P-K Model and Modified P-K Model for India and the 6 selected states. The least value of AICc is written in the bold case for each state.

fertility rises fast in the initial years, attains its peak in the middle ages (19–24) and then falls slowly towards the higher age of the reproductive period. The selected states have different levels of fertility at various ages. Hence, the Total Fertility Rate (TFR) also differs. Here, Uttar Pradesh has the highest level of TFR (2.75) followed by Gujarat (2.03), Karnataka (1.79), West Bengal (1.76), Tripura (1.68) and Punjab (1.60). The value of the peak fertility and its attainment age also varies over selected states. The ASFR curve of India attains its peak fertility (0.195) at the age of 22. Among the selected states, Uttar Pradesh reaches its peak fertility (0.230) at the age of 23, followed by Karnataka (0.178 at the age of 22), West Bengal (0.162 at the age of 19), Punjab (0.146 at the age of 24) and Tripura (0.143 at the age of 19). However, the peak of the fertility pattern of Gujarat remains somewhat constant between 21 to 25 years. Tripura has the highest age-specific fertility at the early ages, whereas, Punjab has the lowest level of fertility at the early ages.

Uttar Pradesh has the highest level of fertility in the later ages (45–49 years), while Punjab has the lowest level of fertility in the later ages too. It is also observed from the figure, that although Karnataka and West Bengal have approximately the same level of TFR, their ASFR is very much different in shape. From the age of 15–20 years, West Bengal has a higher level of fertility than that of Karnataka; in the later ages, Karnataka has a higher level of fertility than that of West Bengal. The rate of increase from the start of the curve to its peak is highest for West Bengal and least for Punjab.

**Performance of the proposed models with respect to the original models.** Table 2 displays the performance of the Modified P-K Model vs. P-K model in terms of SSE and AICc values for India and the six selected states. The P-K model has 4 parameters, whereas the Modified P-K Model has five parameters. By comparing SSE and AICc values of both the models, we have found that the Modified P-K Model is working better only for India, Uttar Pradesh and Gujarat. The SSE values of both the models of Tripura are approximately the same. But the P-K Model is preferred over the Modified P-K Model, because it has less number of parameters. Likewise, Table 3 shows the performance of the Modified Gompertz Model vs. Gompertz model, w.r.t their SSE and AICc values. In comparison to the Gompertz model, both the SSE and AICc values under the Modified Gompertz Model are smaller. It means that the Modified Gompertz Model is more flexible and is able to describe the empirical fertility schedules of India and the 6 selected states. Tables 4 and 5 shows the performance of the Modified Skew Normal Model and the Modified G-P Model vs. Skew-Normal model and G-P Model, respectively. The Modified Skew Normal Model has the least SSE and AICc values, as compared to the Skew-Normal model for India and all the other selected states (Table 4), i.e. the Modified Skew Normal Model outperforms the Skew-Normal model, in recreating the fertility schedules. From the results in Table 5, it can be seen that the Modified G-P Model has the least SSE and AICc value, as compared to the G-P model in India and all the six selected states. It means that the Modified G-P Model provides better results than the G-P model, for fitting the fertility patterns of India and the selected states. Figures B1 to B4 in the Appendix B, show the observed and the fitted age-specific fertility pattern, under the proposed model and the existing model under study. We have shown these graphs only for Indian fertility schedules.

Tables A1 to A7 in the Appendix A present the observed and the fitted estimates of fertility pattern, under each of the four proposed models and Hadwiger model. The tables also present the number of parameters, estimated value of TFR, the values of SSE, the values of AICc and the value of coefficient of determination ( $R^2$ ) for each model. The least value of AICc is written in bold case for each case. Figures 3, 4, 5, 6, 7, 8 and 9 show the smoothed observed and fitted curve, under the proposed models and the Hadwiger model for India, Punjab, Uttar Pradesh, West Bengal, Tripura, Gujarat and Karnataka respectively.

India's fertility pattern is best predicted by the Modified Gompertz Model, as it provides the best fit among all the other models under comparison (Table A1 and Fig. 3). The Modified Gompertz Model also gives the best fit for the fertility pattern of Uttar Pradesh (Table A3 and Fig. 5) and Gujarat (Table A6 and Fig. 8). The

Model	Region	SSE	AICc
Gompertz Model (K = 3)	India	0.0684	-208.41
	Punjab	0.01907	-253.122
	Uttar Pradesh	0.14701	-181.63
	West Bengal	0.0304	-236.79
	Tripura	0.02212	-247.924
	Gujarat	0.05463	-216.278
	Karnataka	0.03432	-232.544
Modified Gompertz Model (K = 4)	India	0.00034	<b>-390.9</b>
	Punjab	0.00017	<b>-413.885</b>
	Uttar Pradesh	0.00022	<b>-405.284</b>
	West Bengal	0.00054	<b>-374.085</b>
	Tripura	0.00155	<b>-337.176</b>
	Gujarat	0.0016	<b>-336.202</b>
	Karnataka	0.00075	<b>-362.856</b>

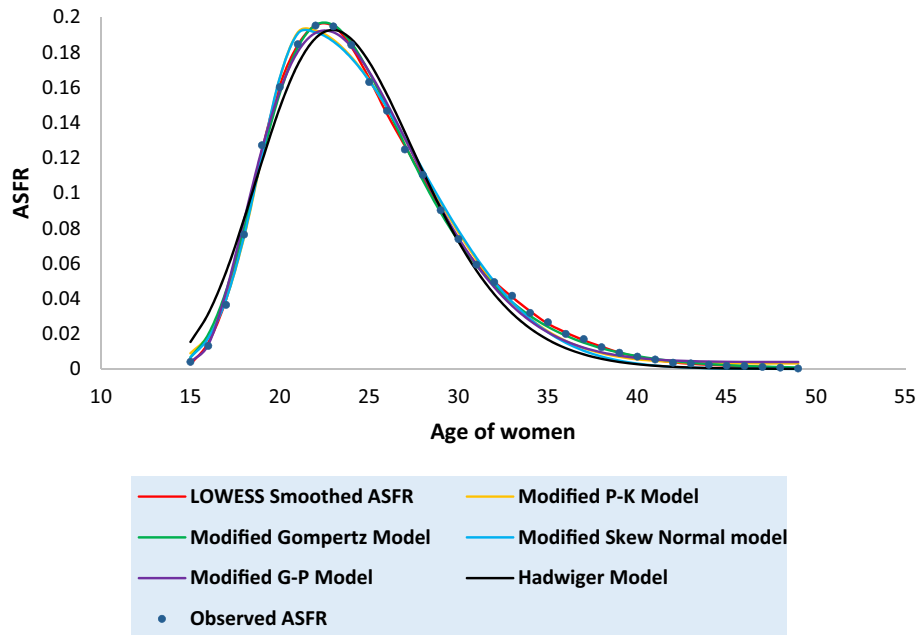
**Table 3.** Minimized sum of square (SSE) and AICc values under Gompertz Model and Modified Gompertz model for India and the 6 selected states. The least value of AICc is written in the bold case for each state.

Model	Region	SSE	AICc
Skew Normal Model (K = 3)	India	0.00449	-303.731
	Punjab	0.02743	-240.39
	Uttar Pradesh	0.00241	-325.443
	West Bengal	0.01739	-256.334
	Tripura	0.01009	-275.384
	Gujarat	0.00968	-276.829
	Karnataka	0.02038	-250.795
Modified Skew Normal Model (K = 5)	India	0.00029	<b>-392.215</b>
	Punjab	0.00056	<b>-368.75</b>
	Uttar Pradesh	0.00032	<b>-388.852</b>
	West Bengal	0.00085	<b>-354.352</b>
	Tripura	0.00061	<b>-36.061</b>
	Gujarat	0.00061	<b>-36.321</b>
	Karnataka	0.00084	<b>-354.882</b>

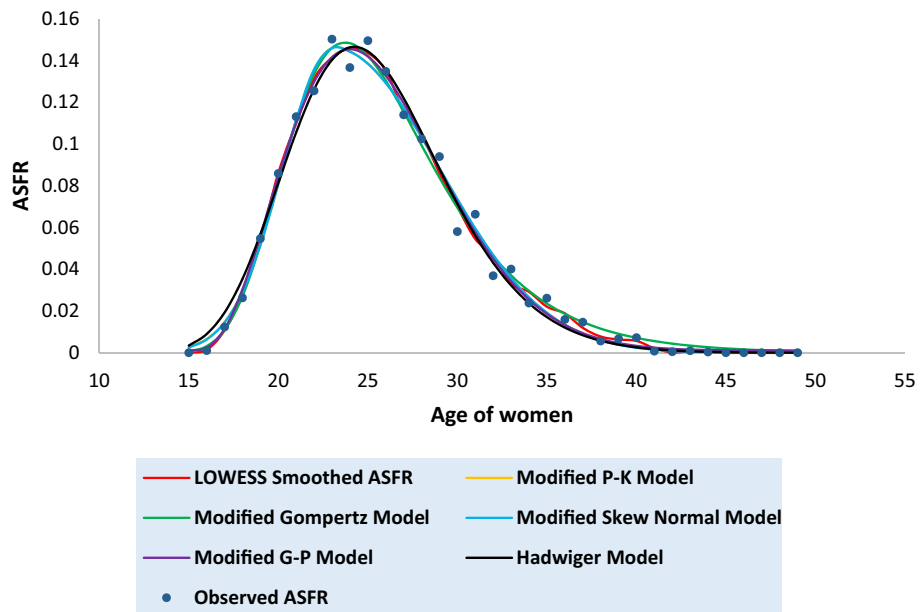
**Table 4.** Minimized sum of square (SSE) and AICc values under Skew Normal Model and Modified Skew Normal Model for India and 6 selected states. The least value of AICc is written in the bold case for each state.

Model	Region	SSE	AICc
G-P Model (K = 3)	India	0.00718	-287.307
	Punjab	0.01829	-254.579
	Uttar Pradesh	0.01326	-265.841
	West Bengal	0.00963	-277.032
	Tripura	0.00472	-302.004
	Gujarat	0.00788	-284.046
	Karnataka	0.00721	-287.162
Modified G-P Model (K = 4)	India	0.00038	<b>-386.121</b>
	Punjab	0.00039	<b>-385.583</b>
	Uttar Pradesh	0.00015	<b>-419.288</b>
	West Bengal	0.00159	<b>-336.338</b>
	Tripura	0.00056	<b>-372.685</b>
	Gujarat	0.00153	<b>-337.806</b>
	Karnataka	0.00081	<b>-359.998</b>

**Table 5.** Minimized sum of square (SSE) and AICc values under G-P Model and Modified G-P Model for India and 6 selected states. The least value of AICc is written in the bold case for each state.

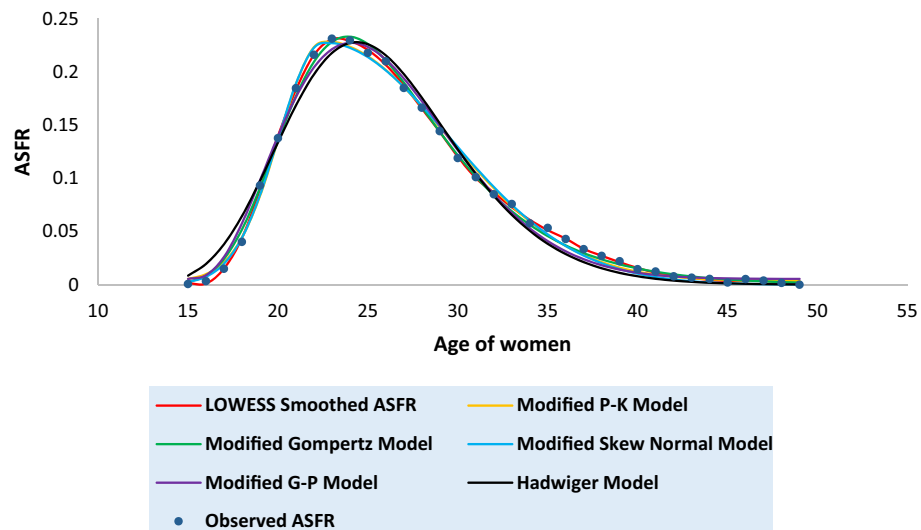


**Figure 3.** Observed and expected age-specific fertility rate (ASFR) for India under different proposed models and Hadwiger model.

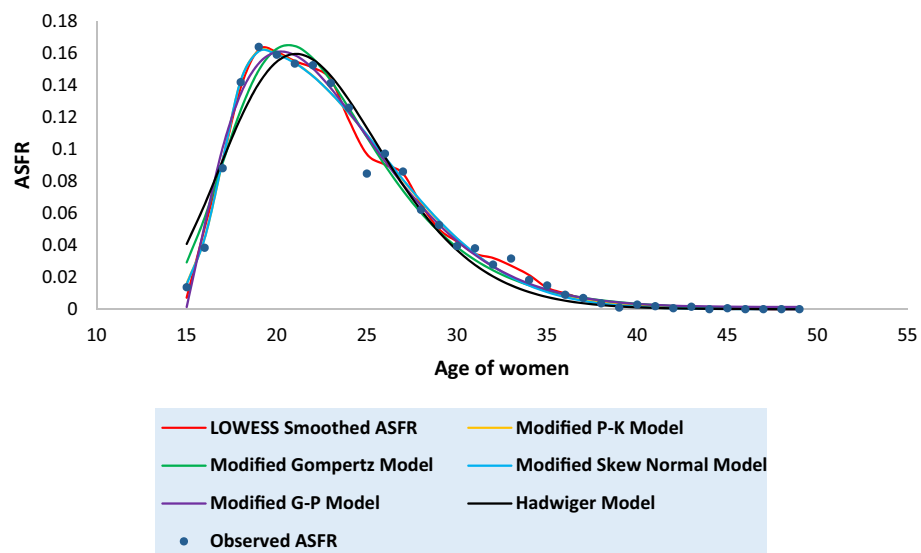


**Figure 4.** Observed and expected age-specific fertility rate (ASFR) for Punjab under different proposed models and Hadwiger model.

Modified P-K Model is superior to the other models under comparison, for the fertility schedules of West Bengal (Table A4 and Fig. 6), Tripura (Table A5 and Fig. 7) and Karnataka (Table A7 and Fig. 9). On the other hand, the fertility pattern of Punjab is best fitted by the Modified G-P Model amongst all the other models (Table A2 and Fig. 4). From Tables A1 to A7, we observe that the estimated values of single-year ASFR under the Modified P-K Model and the Modified Skew Normal Model are quite close. Both have five parameters. However, the AICc values of the Modified P-K Model are slightly smaller than that of the Modified Skew Normal Model for India and for each of the selected states. Further, Tables 6, 7, 8, 9 and 10 display the estimates of parameters of the Modified P-K Model, the Modified Gompertz Model, the Modified Skew Normal Model, the Modified G-P Model and Hadwiger model respectively.

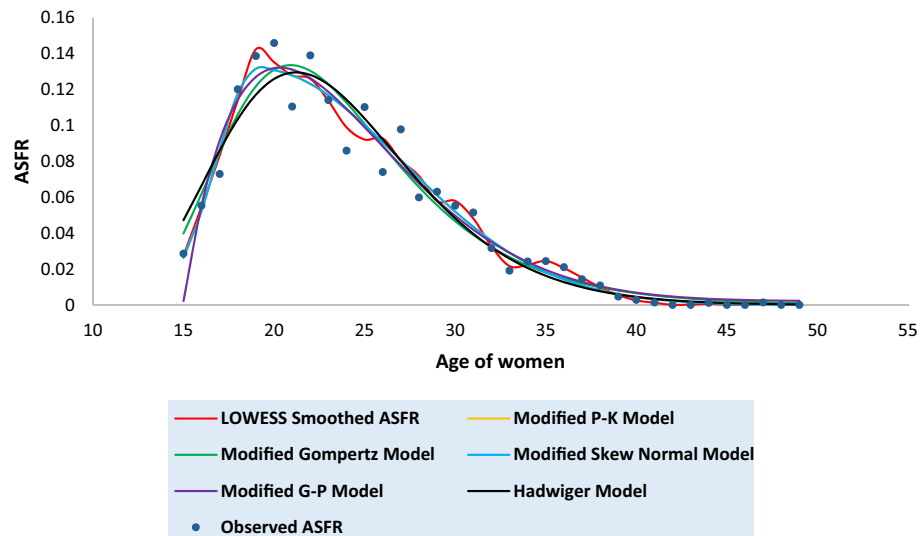


**Figure 5.** Observed and expected age-specific fertility rate (ASFR) for Uttar Pradesh under different proposed models and Hadwiger model.

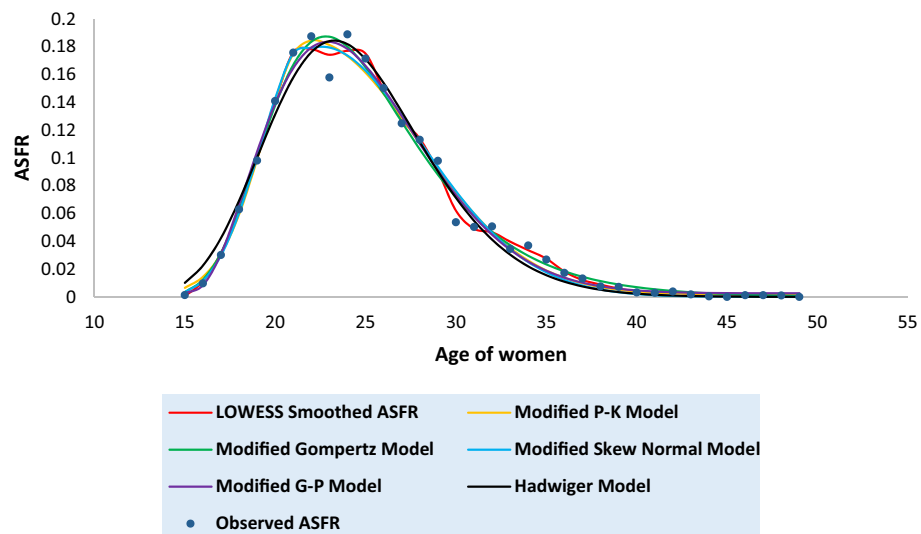


**Figure 6.** Observed and expected age-specific fertility rate (ASFR) for West Bengal under different proposed models and Hadwiger model.

We have also explored the possible demographic interpretation of the new parameters added in the proposed models. There is a strong association between the later ages (45–49 years) fertility and the constant parameter  $a$  of the Modified P-K Model. The correlation coefficient between these two ( $r=0.857$ ) are significant. We have also reviewed the demographic interpretation of the parameters that were already present in the previously existing models. The parameter  $b$  in the Modified P-K Model is significantly related to the extent of peak fertility ( $r=0.988$ ). Parameter  $\mu$  in the Modified P-K Model is related to the women's age at extreme fertility, whereas  $\sigma_1$  and  $\sigma_2$  describes the spread of the ASFR curve before and after the peaked fertility respectively. All the fertility schedules are skewed towards the right. For the Modified Gompertz Model, we find that parameter  $\gamma$  is strongly related to the extent of extreme fertility and the term is significantly correlated ( $r=0.998$ ) to the total fertility rate. Notwithstanding, we could not test for the association of parameter  $m$  with the lowest age of childbearing in the Modified Gompertz Model. There is a great resemblance between the parameters  $\theta$  and  $\lambda$  from the Modified Skew Normal Model and  $b$  and  $\mu$  from the Modified P-K Model. Both the estimates are quite close and hence have similar interpretations.  $\theta$  is strongly associated with level of extreme fertility ( $r=0.985$ ) whereas  $\lambda$  is related to the age at modal fertility. The parameters  $s_1$  and  $s_2$  of the Modified Skew Normal Model represents the spread of the ASFR curve. Estimates of the parameter  $\delta$  in each case was found to be positive under the Modified Skew Normal Model. It means that ASFR curve is positively skewed. The parameter  $s$  in the Modified G-P



**Figure 7.** Observed and expected age-specific fertility rate (ASFR) for Tripura under different proposed models and Hadwiger model.



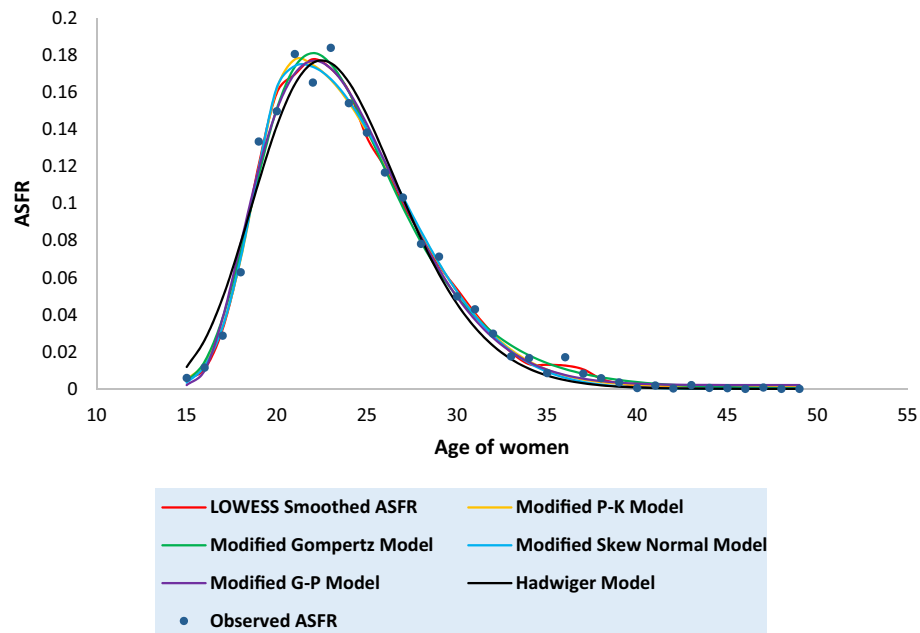
**Figure 8.** Observed and expected age-specific fertility rate (ASFR) for Gujarat under different proposed models and Hadwiger model.

Model is found to be significantly correlated ( $r = 0.969$ ) with the level of fertility in the later ages (45–49 years). Parameter  $q$  is strongly associated with the age of women at extreme fertility ( $r = 0.968$ ). However, we could not find reasonable interpretations for the parameters  $p$  and  $r$  of the Modified G-P Model. Further, parameter  $a$  in Hadwiger model is significantly associated with the total fertility ( $r = 0.997$ ). The other parameter  $c$  in Hadwiger model is closely related to the mean age of childbearing. However, the interpretation of the parameter  $b$  in the Hadwiger model as given by Chandola et al.<sup>10</sup> could not be justified in terms of the height of the fertility curve.

## Discussion

India is a vast country comprising 28 states and 8 union territories, with wide demographic and socio-cultural diversity. Many of the Indian states have higher population than several countries of the world. Comparing the fertility schedules of the Indian states is an extensive exercise. In comparison to the previous studies<sup>7,29</sup> done on the modelling of Indian fertility schedule, our study paid more attention to model with the single-year ASFR of India and the six states. Our study is also distinct from the Pandey et al.<sup>28</sup>, as it proposes four parametric models for single-year ASFR. Pandey et al.<sup>28</sup> also fitted the data of single-year ASFR through polynomial functions.

Results of our study conclude that the Modified P-K Model performs better than P-K model only if the tail of the fertility curve has some amount of fertility. P-K model provides best fit for the countries with non-enhanced



**Figure 9.** Observed and expected age-specific fertility rate (ASFR) for Karnataka under different proposed models and Hadwiger model.

Region	Modified P-K Model				
	<i>b</i>	$\mu$	$\sigma_1$	$\sigma_2$	<i>a</i>
India	0.1902	21.3593	3.4055	8.9152	0.00307
Punjab	0.1457	23.0807	3.9941	8.3289	0.00039
Uttar Pradesh	0.2252	22.5390	3.4817	9.7071	0.00338
West Bengal	0.1605	18.8874	2.5272	9.6826	0.00104
Tripura	0.1315	19.0725	3.1982	11.300	0.00032
Gujarat	0.1827	21.8334	3.5650	8.5596	0.00190
Karnataka	0.1764	20.9251	3.0144	8.1487	0.00144

**Table 6.** Estimates of parameters of Modified P-K Model for India and 6 selected states.

Region	Modified Gompertz Model			
	$\alpha$	$\beta$	$\gamma$	<i>m</i>
India	4.3619	4.0885	13.1039	16.4648
Punjab	4.2900	4.0528	9.67184	17.8301
Uttar Pradesh	4.7330	4.3582	13.316	17.0962
West Bengal	4.1874	3.9694	11.7965	14.966
Tripura	5.1183	4.7769	8.99602	13.1106
Gujarat	4.2689	4.0249	13.2025	16.9462
Karnataka	3.8511	3.698	13.2892	17.0609

**Table 7.** Estimates of parameters of Modified Gompertz Model for India and 6 selected states.

early-age fertility<sup>10</sup>. P-K model performed better than Gamma model, Hadwiger model and quadratic spline model for the majority of countries (Norway, Denmark, Italy, Greece and Belgium) examined<sup>10</sup>. The Modified P-K Model generally fits well to the fertility schedules with a low level of fertility. It has provided the best fit to the fertility pattern of West Bengal (TFR = 1.8), Tripura (TFR = 1.7) and Karnataka (TFR = 1.8). The AICc value of the Modified Gompertz Model for Karnataka is quite close to that of the Modified P-K Model. The common

Region	Modified Skew Normal Model				
	$\lambda$	$s_1$	$\delta$	$s_2$	$\theta$
India	21.3305	2.4382	2.8960	6.4817	0.1924
Punjab	23.0761	2.8307	2.4989	5.9163	0.1460
Uttar Pradesh	22.5175	2.4965	2.9761	7.0375	0.2275
West Bengal	18.8777	1.7939	3.3029	6.9201	0.1612
Tripura	19.0693	2.2651	3.5001	8.0211	0.1318
Gujarat	21.9311	2.6643	1.4815	6.0231	0.1864
Karnataka	20.9544	2.1678	1.5237	5.7689	0.1793

**Table 8.** Estimates of parameters of Modified Skew Normal Model for India and 6 selected states.

Region	Modified G-P Model			
	$p$	$q$	$r$	$s$
India	165.321	2.426	13.3747	0.00392
Punjab	837.923	3.3364	14.8441	0.00110
Uttar Pradesh	492.428	2.9784	13.0964	0.00544
West Bengal	9.42817	1.2826	10.7123	0.00141
Tripura	2.98192	0.9902	8.12041	0.00223
Gujarat	321.995	2.7327	14.2034	0.00263
Karnataka	359.998	2.6692	15.7946	0.00202

**Table 9.** Estimates of parameters of Modified G-P Model for India and 6 selected states.

Region	Hadwiger Model		
	$a$	$b$	$c$
India	1.2141	3.6940	24.2482
Punjab	0.9044	3.9764	25.4145
Uttar Pradesh	1.5322	3.6763	25.7625
West Bengal	1.0012	3.4104	22.4573
Tripura	0.9832	2.8627	23.2429
Gujarat	1.1338	3.8335	24.4840
Karnataka	1.0023	4.0238	23.5473

**Table 10.** Estimates of parameters of Hadwiger Model for India and 6 selected states.

feature of the fertility curve of West Bengal and Tripura is that, both have a high level of fertility in the early ages (15–19) and the Modified P-K Model gives a considerable amount of fertility in the early ages.

While an experiment of Hoem et al.<sup>6</sup> to fit Gompertz curve to Danish fertility schedules for the calendar years 1962–1971 was found to be unpromising, the Modified Gompertz Model provided best fit for India (moderate level TFR = 2.2), Uttar Pradesh (High-level TFR = 2.8) and Gujarat (Low-level TFR = 2.0), i.e. it can reproduce all types of fertility schedules, where the total fertility is greater or equal to two. However, it is observed that it works better for those fertility curves, where the decline in the level of fertility after the peak is not so rapid.

Mazzucco and Scarpa showed that the skew-normal density is quite good for modelling the unimodal fertility curve. Skew-normal density has outperformed Gamma, Hadwiger and P-K model for the United States fertility data of 1963<sup>41</sup>. Our study has confirmed that the Modified Skew Normal Model has worked better than the original skew-normal model for India and all the other six states. Gupta and Pasupuleti have examined the applicability of G-P model on the cohort fertility schedules of India, Finland, Sweden, Netherlands, Switzerland and USA and the period fertility schedules of Bhutan, Iran, Somoa and Algeria<sup>41</sup>. The results of their study concluded that the fit was inferior to Gamma and Hadwiger model, but superior to the Gompertz model. However, the differences were marginal<sup>41</sup>. The Modified G-P Model provided better fitting than that of G-P model for the fertility schedule of India and all the six states considered in our study.

Although Punjab also has a low level of total fertility, the Modified G-P Model has provided the best fit. The reason behind the failure of the Modified P-K Model, for the best fit of Punjab might be that, the ASFR curve of Punjab has a shallow level of fertility in the early and the later ages. The inclusion of constant parameters in the Modified P-K Model and the Modified G-P Model, gives more weights of fertility in the tail of the fertility curve,

when compared to the model without these constant parameters. Figures of the empirical and the observed fertility patterns, show that the Hadwiger model overestimates the early and the age-specific fertility around the peak. Though the attempt to fit the proposed models to the recent single-year ASFR curve has met some success, there is great scope of work for other researchers to find the suitability of these models for grouped and order-specific ASFR and the past fertility patterns of India and its varied states.

Most of the studies<sup>7,10,16,22</sup> have either used AIC or residual sum of squares criterion to evaluate the performance of the proposed models. However, our study has used the corrected version of AIC criterion (AICc). Burnham and Anderson<sup>44</sup> suggested that, AICc should be preferred over AIC when the ratio of the number of observations ( $n$ ) and the number of parameters ( $k$ ) is less than 40. In our study, the number of observations is 35 and the number of parameters is either 3, 4 or 5. So in each case, the ratio will be less than 40. Hence, the use of AICc is justified here. This also makes our study distinct from other studies.

It has been a subject of great discussion, whether the period or cohort ASFR should be adopted for the illustration of the model fit. Our results were based on the period fertility only. However, there is a lot of scope for researchers to fit our proposed models on cohort fertility. We have considered period data, since period data is easily available for analytical purpose and it also enables us to evaluate and examine the influence of recent demographic events and targeted interventions. On the other hand, cohort data either suffers from attrition or difficulty to reconstruct. Additionally, cohort data is concerned with the long term development, as cohort trends would be observable only if it is followed up for a longer period of time.

## Conclusion

The models that we have proposed are extensions to P-K Model, Gompertz model, generalized skew-normal function and Gupta and Pasupuleti model. These proposed models were compared among themselves and with their previously existing forms. Results of this study suggests, that all the proposed models are quite flexible and are able to produce a close fit to the variable type of the observed fertility pattern. It is pertinent to mention, that all the proposed models are more efficient than the original models considered in this article and the Hadwiger model, in terms of the goodness of fit for fertility schedules of India and its states. The demographic interpretation of the parameters of the model is also explored in this study. It is evident from the demographic interpretation of the parameters that the demographic profile of the Indian states is much more diversified. Although our proposed models have illustrated the fitting of fertility schedules of India on the whole and for the six states only, it was intended to provide an important channel, for the policymakers and the government authorities, to project the fertility rates and to understand the fertility transitions in India and its various states.

## Availability of data and materials

The datasets analysed in our study are available in <https://dhsprogram.com/data/available-datasets.cfm>.

Received: 30 June 2020; Accepted: 8 March 2021

Published online: 23 March 2021

## References

- Gilje, E. Fitting curves to age-specific fertility rates: Some examples. *Stat. Rev. Swed. Natl. Cent. Bureau Stat.* **3**, 118–134 (1969).
- Hadwiger, H. Eine analytische Reproduktionsfunktion für biologische Gesamtheiten. *Scand. Actuar. J.* **1940**, 101–113 (1940).
- Chandola, T., Coleman, D. A. & Hiorns, R. W. Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Popul. Stud.* **53**, 317–329 (1999).
- Coale, A. J. & Trussell, T. J. Model fertility schedules: Variations in the age structure of childbearing in human populations. *Popul. Index* **40**, 185 (1974).
- Coale, A. J. & Trussell, T. J. Technical note: Finding the two parameters that specify a model schedule of marital fertility. *Popul. Index* **44**, 203 (1978).
- Hoem, J. M. *et al.* Experiments in modelling recent Danish fertility curves. *Demography* **18**, 231 (1981).
- Mishra, R., Singh, K. K. & Singh, A. A model for age-specific fertility rate pattern of India using skew-logistic distribution function. *Am. J. Theor. Appl. Stat.* **6**, 32 (2017).
- Romaniuk, A. A three parameter model for birth projections. *Popul. Stud.* **27**, 467–478 (1973).
- Islam, M. N. & Mallick, S. A. On the use of a truncated Pearsonian Type III curve in fertility estimation. *Dhaka Univ. Stud. Part B Sci.* **35**, 23–32 (1987).
- Peristera, P. & Kostaki, A. Modeling fertility in modern populations. *Demogr. Res.* **16**, 141–194 (2007).
- Mazzuco, S. & Scarpa, B. Fitting age-specific fertility rates by a flexible generalized skew normal probability density function. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **178**, 187–203 (2014).
- Bermúdez, S., Blanquero, R., Hernández, J. A. & Planelles, J. A new parametric model for fitting fertility curves. *Popul. Stud.* **6**, 297–310 (2012).
- Rueda-Sabater, C. & Alvarez-Esteban, P. C. The analysis of age-specific fertility patterns via logistic models. *J. Appl. Stat.* **35**, 1053–1070 (2008).
- Asili, S., Rezaei, S. & Najjar, L. Using skew-logistic probability density function as a model for age-specific fertility rate pattern. *Biomed. Res. Int.* **2014**, 1–5 (2014).
- Brass, W. Assessment of the validity of fertility trend estimates from maternity histories. *Regional Workshop on Techniques of Analysis of WFS Data* 50–57 (1980).
- Brass, W. The use of the Gompertz relational model to estimate fertility. *International Population Conference, Manila (IUSSP)* **3**, 345–362 (1981).
- Farid, S. M. On the pattern of cohort fertility. *Popul. Stud.* **27**, 159–168 (1973).
- Goldstein, J. R. A behavioral Gompertz model for cohort fertility schedules in low and moderate fertility populations (2010). <https://doi.org/10.4054/mpidr-wp-2010-021>
- Gompertz, B. On the nature of the function expressive of the law of human mortality. *Math. Demogr. Biomath.* **6**, 279–282. [https://doi.org/10.1007/978-3-642-81046-6\\_30](https://doi.org/10.1007/978-3-642-81046-6_30) (1977).
- Murphy, E. M. & Nagnur, D. N. A Gompertz fit that fits: Applications to Canadian fertility patterns. *Demography* **9**, 35 (1972).
- Pasupuleti, S. S. R. & Pathak, P. Special form of Gompertz model and its application. *Genus* **6**, 95–125 (2011).

22. Gayawan, E., Adebayo, S. B., Ipinyomi, R. A. & Oyejola, B. A. Modeling fertility curves in Africa. *Demogr. Res.* **22**, 211–236 (2010).
23. United Nations. *Conditions and trends of fertility in the world*, Population Bulletin No. 7, ST/SO A/Series N/7. (New York: United Nations Department of Economic and Social Affairs) (1976).
24. NFHS-1. *National Family Health Survey, India*. <http://rchiips.org/NFHS/india1.shtml> (Accessed 9th April 2020)
25. NFHS-2. *National Family Health Survey, India*. <http://rchiips.org/NFHS/india2.shtml> (Accessed 9th April 2020)
26. NFHS-3. *National Family Health Survey, India*. [http://rchiips.org/NFHS/NFHS-3%20Data/VOL-1/India\\_volume\\_I\\_corrected\\_17oct08.pdf](http://rchiips.org/NFHS/NFHS-3%20Data/VOL-1/India_volume_I_corrected_17oct08.pdf) (Accessed 9th April 2020)
27. NFHS-4. *National Family Health Survey, India*. <http://rchiips.org/nfhs/pdf/NFHS4/India.pdf> (Accessed 9th April 2020)
28. Pandey, R. & Kaur, C. Modelling fertility schedules of India. *Can. Stud. Popul.* **46**, 47–60 (2019).
29. Singh, B. P., Gupta, K. & Singh, K. K. Analysis of fertility pattern through mathematical curves. *Am. J. Theor. Appl. Stat.* **4**, 64 (2015).
30. Brass, W. The graduation of fertility distributions by polynomial functions. *Popul. Stud.* **14**, 148–162 (1960).
31. Hoem, J. M. & Rennermalm, B. *On the Statistical Theory of Graduation by Splines* (Laboratory of Actuarial Mathematics, University of Copenhagen, 1977).
32. Mcneil, D. R., Trussell, T. J. & Turner, J. C. Spline interpolation of demographic data. *Demography* **14**, 245 (1977).
33. Schmertmann, C. A system of model fertility schedules with graphically intuitive parameters. *Demogr. Res.* **9**, 81–110 (2003).
34. Schmertmann, C. P. Calibrated spline estimation of detailed fertility schedules from abridged data<sup>1</sup>. *Revista Brasileira de Estudos de População* **31**, 291–307 (2014).
35. Schmertmann, C. Quadratic spline fits by nonlinear least squares. *Demogr. Res.* **12**, 105–106 (2005).
36. Kostaki, A., Moguerza, J., Olivares, A. & Psarakis, S. Graduating the age-specific fertility pattern using Support Vector Machines. *Demogr. Res.* **20**, 599–622 (2009).
37. Alho, J. M. The magnitude of error due to different vital processes in population forecasts. *Int. J. Forecast.* **8**, 301–314 (1992).
38. Schoumaker, B. A Stata module for computing fertility rates and TFRs from birth histories: tfr2. *Demogr. Res.* **28**, 1093–1144 (2013).
39. Mathivanan, N. M. N., Ghani, P. A. & Md. Ghani, N. A. Tracing mathematical function of age specific fertility rate in Peninsular Malaysia. *Indones. J. Electr. Eng. Comput. Sci.* **9**, 637 (2018).
40. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171–178 (1985).
41. Gupta, A. & Pasupuleti, S. S. R. A new behavioural model for fertility schedules. *J. Appl. Stat.* **40**, 1921–1930 (2013).
42. Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **76**, 297–307 (1989).
43. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
44. Burnham, K. P. & Anderson, D. R. Multimodel Inference. *Sociol. Methods Res.* **33**, 261–304 (2004).

## Acknowledgements

All the authors wish to render profound sense of gratitude to Ms. Amelia Betsy Christopher for her insightful comments and suggestions for the improvement in the language of the manuscript. The authors would like to thank all the reviewers and the editor for their valuable comments and thoughtful suggestions, which has helped us to improve the quality of our article.

## Author contributions

U.S. has proposed the problem, designed the study, analysed the data and had done the literature writing. K.K.S. and A.P. have revised the manuscript critically, for the intellectual content and gave the final approval of the version to be published. N.N. has reviewed the literature and conceptualized the problem. All the authors have read and have approved the final manuscript. The views expressed by the Ujjaval Srivastava (US) in this paper are his own and not that of the Government of India.

## Funding

The authors have not received any specific funding for this study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-85959-z>.

**Correspondence** and requests for materials should be addressed to U.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021