

# Statistical Modeling and Analysis of Fertility in China

Yu Zhang

Tianjin University of Commerce, Tianjin, China

## Abstract

Fertility is one of the key indicators in demographic research. In recent years, fertility in China has continued to decline, and population change has attracted widespread attention. This paper applies five models—the Hadwiger model, the Gamma model, Model I and Model II proposed by Peristera and Kostaki, and the Adjusted Error Model—to fit national age-specific fertility rate (ASFR) data for China in 2000, 2010, and 2020. The fitting performance of these models is compared in order to identify the models most suitable for China's fertility pattern and to provide a scientific basis for fertility-related policy formulation and population regulation. The results show that the Gamma model and Model I yield the smallest residual sum of squares and provide the best fit, indicating that they are well suited to China's fertility pattern.

## Keywords

Age-specific fertility rate; Hadwiger model; Gamma model; Model I; Model II; Adjusted Error Model.

## 1. Introduction

China is one of the most populous countries in the world, and population-related issues have long been a major concern of the Chinese government. For a country with a large population base, fertility is directly related to the sustainability of future socioeconomic development. Over the past several decades, China has experienced rapid urbanization and industrialization, and these processes have had a profound impact on fertility.

The early implementation of the family planning policy led to a rapid decline in fertility from the high levels observed in the 1970s. As a result, China's population age structure has gradually shifted toward population aging and low fertility, with the total fertility rate falling below the level required for population replacement. According to the Seventh National Population Census, China's total fertility rate in 2020 was only 1.3. This indicates that fertility has fallen below the national warning threshold and that population development has entered a critical turning point.

The age-specific fertility rate (ASFR) describes fertility behavior across different age groups and is defined as the number of births per 1,000 women in a given age group. ASFR reflects changes in fertility behavior and reproductive patterns within a country or region. Through the analysis of ASFR, it is possible to identify fertility patterns such as early or delayed childbearing. Consequently, fertility research has long been a major focus of government agencies and scholars, and one important research direction is the mathematical modeling of fertility.

In the domestic literature, Huang Rongqing and Qi Xin (1990) found that ASFR follows a lognormal distribution[1]. Chen Hui and Liu Xiaoxian (2003) showed that the lognormal and Weibull distributions provide relatively good fits to cumulative fertility rates[2]. Li Changwei and Chen Yun (2017) argued that a combined lognormal-Weibull model can be used to forecast ASFR with relatively high predictive accuracy[3]. In comparison, international research in this area is more extensive. Hadwiger (1940) first proposed the Hadwiger model and applied it to fertility modeling using Swiss population data[4]. Coale and Trussell (1974) suggested that fertility can be expressed as the product of marriage rates, natural fertility, and deviations from

natural fertility[5]. Hoem et al. (1981) analyzed Danish ASFR data from 1962 to 1971 using cubic splines, the Hadwiger model, the Coale–Trussell function, the Gamma distribution, the Beta distribution, and the Gompertz curve, finding that cubic splines performed best, followed by the Coale–Trussell function and the Gamma distribution[6]. Chandola et al. (1999) found that existing models perform poorly when fertility at younger ages is relatively high and recommended a two-component Hadwiger model[7]. Peristera and Kostaki (2007) argued that existing models lack sufficient flexibility to capture new fertility patterns and proposed two new models, referred to as Model I and Model II[8].

A review of the literature indicates two main limitations in existing fertility modeling studies for China: a limited range of modeling approaches and relatively early study periods. To address these limitations, this paper applies the Hadwiger model, the Gamma model, Model I, Model II, and the Adjusted Error Model to ASFR data from 2000, 2010, and 2020. Model performance is evaluated using the residual sum of squares and the Akaike Information Criterion (AIC), with the aim of identifying the models that best describe China’s fertility pattern. This study contributes to a deeper understanding of population change in China, provides empirical support for policy formulation, and offers reference value for fertility research in other countries and regions.

## 2. Analysis of Changes in Fertility Patterns in China

With continuous social and economic development, fertility has remained a key demographic indicator of interest. In China, fertility has undergone substantial changes alongside rapid economic growth and policy adjustments. The family planning policy, introduced in the 1970s to control population size and alleviate pressure on economic and social development, has evolved over time, including the introduction of the “only-child-for-both-parents” policy and the universal two-child policy. These policy changes have also influenced fertility behavior, prompting increasing attention to fertility modeling from a quantitative perspective.

### 2.1. Population Change in China

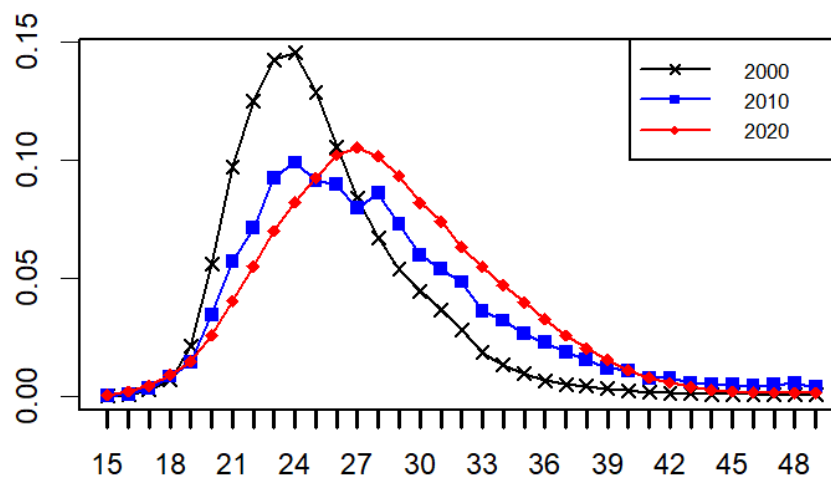
**Table 1.** Crude Birth Rate, Crude Death Rate, and Natural Population Growth Rate in China, 1978–2020 (‰)

Year	Crude Birth Rate	Crude Death Rate	Natural Population Growth Rate
1978	18.25	6.25	12
1979	17.82	6.21	11.61
1980	18.21	6.34	11.87
1981	20.91	6.36	14.55
1982	22.28	6.6	15.68
1983	20.91	6.9	13.29
1984	19.9	6.82	13.08
1985	21.04	6.78	14.26
1986	22.43	6.86	15.57
...	...	...	...
2015	11.99	7.07	4.93
2016	13.57	7.04	6.53
2017	12.64	7.06	5.58
2018	10.86	7.08	3.78
2019	10.41	7.09	3.32
2020	8.52	7.07	1.45

To examine population change in China, data on the crude birth rate, crude death rate, and natural population growth rate from 1978 to 2020 are drawn from the *China Population and Employment Statistical Yearbook 2022*. The results show a clear downward trend in the birth rate, which declined from 18.25‰ in 1978 to 8.52‰ in 2020, indicating a gradual reduction in fertility levels. In contrast, the death rate increased steadily from 6.25‰ to 7.07‰ over the same period. The natural population growth rate, defined as the difference between the birth rate and the death rate, declined from 12‰ in 1978 to 1.45‰ in 2020. This decline is mainly attributable to decreasing fertility and increasing mortality, indicating a slowdown in natural population growth in China.

## 2.2. Changes in Age-Specific Fertility Rates in China

The number of women of childbearing age increased from 347 million in 2000 to 371 million in 2010, before declining to 322 million in 2020. ASFR data for 2000, 2010, and 2020 (Figure 1) show that China's fertility pattern has changed significantly over the past two decades.



**Figure 1.** Trends in National Age-Specific Fertility Rates in 2000, 2010, and 2020

Overall fertility levels have declined, particularly among women under the age of 25, for whom the decrease is especially pronounced. Among women aged 20 and above, the location of the fertility peak has shifted noticeably. In 2000, fertility peaked at around age 24, whereas in 2020 the peak occurred at approximately age 28, indicating a clear postponement of childbearing. Fertility rates declined sharply among women under age 20, while fertility among women aged 28 and above increased. These patterns suggest that fertility in China is increasingly concentrated at older ages. A comparison of ASFRs across 2000, 2010, and 2020 shows a clear rightward shift of the fertility curve, indicating delayed childbearing, a longer reproductive span, and declining fertility levels.

## 3. Overview of Fertility Models

With social development and rising living standards, fertility has remained a topic of sustained interest. Over the past several decades, fertility in China has experienced considerable fluctuations. To better understand fertility dynamics and predict future trends, many scholars have proposed and applied various fertility models. This section introduces the fertility models considered in this study.

### 3.1. Hadwiger Model

The Hadwiger model is a mathematical model used to describe population age structure and fertility patterns. Hadwiger (1940) first proposed the model to describe reproductive processes in biological populations and provided its functional form. The model was later applied to Swiss population data, demonstrating its effectiveness in describing reproductive behavior.

The Hadwiger function is given by

$$f(x) = \frac{ab}{c} \left(\frac{c}{x}\right)^2 \exp \left\{ -b^2 \left( \frac{c}{x} + \frac{x}{c} - 2 \right) \right\} \quad (1)$$

Where  $x$  denotes the age of the mother at childbirth, and  $a, b$  and  $c$  are unknown parameters to be estimated. Chandola et al. (1999) noted that these parameters have demographic interpretations: parameter  $a$  is related to total fertility, and parameter  $b$  is related to the mean age at childbearing, parameter  $c$  determines the height of the ASFR curve and parameter  $\frac{ab}{c}$  is associated with the maximum fertility level.

The Hadwiger model has a wide range of applicability. Compared with more complex fertility models, it has a relatively simple parameter structure and has shown good accuracy in fitting and forecasting fertility data. As a result, it has been widely used in fertility research.

### 3.2. Gamma Model

The Gamma model is a commonly used approach for fitting fertility curves. Hoem et al. (1981) successfully applied the Gamma model to Danish fertility data, demonstrated its suitability, and proposed methods for parameter estimation. They also used the model for forecasting and simulation, examining the sensitivity of fertility to different factors and analyzing the causes of fertility change, thereby promoting the application of the Gamma model in fertility research.

The Gamma function is given by

$$f(x) = R \frac{1}{\Gamma(b)c^b} (x - d)^{b-1} \exp \left\{ -((x - d)/c) \right\}, x > d \quad (2)$$

Where,  $d$  represents the lower bound of childbearing age, the parameter  $R$  determines the fertility level, and the parameters  $b, c$  have no direct demographic interpretation. Hoem et al. have substituted these by the mode  $m$ , the mean  $\mu$  and the variance  $\sigma^2$  of the density, for  $c = \mu - m$  and  $b = \frac{\mu - d}{c} = \frac{\sigma^2}{c^2}$ .

### 3.3. Model I

Peristera and Kostaki (2007) observed that fertility among young women in developed countries such as the United Kingdom, Ireland, and the United States has increased in recent years, and that existing fertility models cannot adequately describe this pattern. They also found that in countries with distorted fertility patterns, first-birth fertility among young women exhibits a pronounced hump that is stronger than that of total fertility. To address these issues, Peristera and Kostaki proposed two models, referred to as Model I and Model II, which can be adjusted to capture both historical and modern distorted ASFR patterns.

To address the problem of elevated fertility at young ages, Peristera and Kostaki proposed the following formulation, referred to here as Model I:

$$f(x) = \begin{cases} c_1 \exp \left\{ -\left(\frac{x-\mu}{\sigma_{11}}\right)^2 \right\}, & x \leq \mu \\ c_1 \exp \left\{ -\left(\frac{x-\mu}{\sigma_{12}}\right)^2 \right\}, & x \geq \mu \end{cases} \tag{3}$$

where  $f(x)$  denotes the fertility rate at age  $x$ , and  $c_1, \mu, \sigma$  are parameters to be estimated. Parameter  $c_1$  reflects the baseline fertility level and is related to total fertility, parameter  $\mu$  determines the location of the distribution, and parameters  $\sigma_{11}$  and  $\sigma_{12}$  reflect the dispersion of fertility before and after the peak.

### 3.4. Model II

Model II proposed by Peristera and Kostaki (2007) is given by

$$f(x) = c_1 \exp \left[ -\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \right] + c_2 \exp \left[ -\left(\frac{x-\mu_2}{\sigma_2}\right)^2 \right] \tag{4}$$

where  $f(x)$  denotes the fertility rate at age  $x$ ,  $c_1, c_2, \mu_1, \mu_2, \sigma_1$  and  $\sigma_2$  are parameters to be estimated.

These parameters describe different features of the fertility curve. Parameters  $c_1, c_2$  represent fertility levels at two peaks, parameters  $\mu_1, \mu_2$  represent the mean ages at childbearing for two subpopulations, and parameters  $\sigma_1, \sigma_2$  represent the variances associated with the two peaks. These parameters can be used to compare fertility curves across different populations or time periods.

### 3.5. Adjusted Error Model

Gayawan et al. (2010) noted that most existing fertility models focus on developed countries, with relatively limited attention to other regions[9]. To address this gap, they proposed a flexible parametric model capable of describing ASFR patterns in African countries. Because the model incorporates an error function, it is referred to as the Adjusted Error Model. This model can fit ASFR curves of different shapes, particularly those with two or more peaks, and is suitable for analyzing fertility patterns across different populations.

The Adjusted Error Model is specified as

$$f(x|a, b, \mu, s_1, s_2) = \begin{cases} \operatorname{Erfc} \left[ \left(\frac{x-\mu}{s_1}\right)^2 \right], & x \leq \mu \\ \operatorname{Erfc} \left[ \left(\frac{x-\mu}{s_2}\right)^2 \right], & x > \mu \end{cases} \tag{5}$$

where  $\operatorname{erfc}()$  denotes the complementary error function,  $f(x|a, b, \mu, s_1, s_2)$  is the fertility rate at age  $x$ ,  $b$  is the maximum fertility level,  $\mu$  is the mean age at childbearing, and  $s_1, s_2$  represent the dispersion of childbearing ages around their respective reference points. The constant term approximates fertility in the 45–49 age group.

Model parameters are estimated by minimizing the residual sum of squares  $SSQ = \sum_x (f_x - \hat{f}_x)^2$ , where  $f_x$  denotes the observed fertility rate and  $\hat{f}_x$  denotes the fitted value.

## 4. Modeling Analysis of Fertility in China

To capture recent changes in fertility in China, the five models are applied to ASFR data for 2000, 2010, and 2020 to identify fertility patterns. Parameter estimation is conducted using R software, and the estimated parameters are reported in Tables 2–5.

**Table 2.** Parameter Estimates of the Hadwiger Model and the Gamma Model

	Hadwiger Model				Gamma Model		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>R</i>	<i>b</i>	<i>c</i>	<i>d</i>
2000	0.66	5.08	24.59	1.18	7.72	1.28	15.00
2010	0.65	3.73	27.01	1.16	5.40	2.30	15.00
2020	0.72	3.80	28.40	1.29	10.43	1.63	11.43

**Table 3.** Parameter Estimates of Model I

	Model I			
	<i>c</i> <sub>1</sub>	$\mu$	$\sigma_{11}$	$\sigma_{12}$
2000	0.13	22.64	0.99	1.92
2010	0.09	23.58	1.25	2.33
2020	0.10	26.21	1.66	2.02

**Table 4.** Parameter Estimates of Model II

	Model II					
	<i>c</i> <sub>1</sub>	<i>m</i> <sub>1</sub>	$\sigma_1$	<i>c</i> <sub>2</sub>	<i>m</i> <sub>2</sub>	$\sigma_2$
2000	1.84	0.07	20.14	37.78	0.07	0.12
2010	3.78	0.32	19.92	37.19	0.01	0.11
2020	0.10	1.01	27.62	36.04	7.05	0.01

**Table 5.** Parameter Estimates of the Adjusted Error Model

	AEM				
	<i>a</i>	<i>b</i>	$\mu$	<i>s</i> <sub>1</sub>	<i>s</i> <sub>2</sub>
2000	0.004	0.131	22.648	3.480	8.579
2010	-0.133	0.198	23.585	13.473	40.001
2020	-0.195	0.267	25.475	18.559	40.339

The Akaike Information Criterion (AIC) is a commonly used measure of model goodness of fit that balances model complexity and flexibility. Tables 6–8 report the residual sum of squares, number of parameters, and AIC values for each model in 2000, 2010, and 2020.

**Table 6.** SSQ, Number of Parameters, and AIC Values of the Five Models for 2000

	Hadwiger Model	Gamma Model	Model I	Model II	AEM
SSQ× 10 <sup>5</sup>	150.1	60.9	59.4	11454.3	131.8
K	3	4	4	6	5
AIC	-116.446	-202.695	-119.512	-123.088	-341.146

**Table 7.** SSQ, Number of Parameters, and AIC Values of the Five Models for 2010

	Hadwiger Model	Gamma Model	Model I	Model II	AEM
SSQ× 10 <sup>5</sup>	102.4	47.1	34.9	7693.4	1083.6
K	3	4	4	6	5
AIC	-138.069	-168.75	-141.314	-157.621	-267.412

**Table 8.** SSQ, Number of Parameters, and AIC Values of the Five Models for 2020

	Hadwiger Model	Gamma Model	Model I	Model II	AEM
SSQ× 10 <sup>5</sup>	9.4	8.5	13.4	86.0	1126.7
K	3	4	4	6	5
AIC	-130.238	-161.619	-133.108	-160.231	-266.048

At the national level, the Gamma model and Model I provide the best fit to the 2000 ASFR data, with residual sums of squares substantially smaller than those of the other models. Although the Adjusted Error Model yields the smallest AIC, it involves more parameters than the Gamma model and Model I. As a result, the Gamma model and Model I provide a more appropriate description of China's ASFR in 2000. For 2010, Model I performs best, with the smallest residual sum of squares and fewer parameters than the Adjusted Error Model. For 2020, the Gamma model yields the smallest residual sum of squares and provides the best fit.

Model II and the Adjusted Error Model are not suitable for China's ASFR in any of the three years. Model II was designed to fit fertility patterns characterized by relatively high fertility at young ages, while the Adjusted Error Model is more appropriate for African fertility patterns, where nonmarital fertility plays an important role. In China, the legal minimum age at marriage for women is 20, and cultural norms further limit nonmarital childbearing, resulting in generally low fertility at young ages. The Hadwiger model is more suitable for smooth fertility curves and does not adequately capture China's fertility pattern.

## 5. Conclusions and Recommendations

### 5.1. Conclusions

The analysis of age-specific fertility rates provides a comprehensive understanding of fertility conditions in a country or region, including fertility behavior across age groups, the timing of fertility peaks, and trends in delayed childbearing. By fitting five fertility models to China's national ASFR data for 2000, 2010, and 2020, this study compares their performance and identifies the most suitable models. The results show that the Gamma model and Model I are well suited to China's fertility pattern. Factors such as age at childbearing, reproductive span, and birth spacing have significant effects on fertility. Overall, China's fertility pattern is characterized by delayed childbearing, an extended reproductive period, and declining fertility levels. These factors should be considered in the formulation of fertility and population policies.

### 5.2. Recommendations

China's total fertility rate has continued to decline. Based on the findings of this study, the following recommendations are proposed to address the current situation of extremely low fertility:

From a national perspective, policies should aim to lower the age at first marriage and encourage childbearing at younger ages. Comparisons of ASFRs across 2000, 2010, and 2020 show a clear rightward shift in the fertility curve, indicating delayed childbearing, a longer reproductive span, and declining fertility. Women at younger ages generally have better physiological conditions for pregnancy and childbirth. Lowering the age at first marriage may not only improve biological readiness for childbearing but also extend the reproductive period, providing more time for family planning and childbearing decisions.

## References

- [1] Huang, R. Q., and Qi, X. (1990). A study on models of first marriage and first birth in China. *Population Science of China*, (4), 6-13. [in Chinese]

- [2] Chen, H., and Liu, X. X. (2003).An exploration of probabilistic models of female fertility patterns. *Chinese Journal of Health Statistics*, (4), 17–20. [in Chinese]
- [3] Li, C. W., and Chen, Y. (2017).A study of probabilistic models for age-specific fertility rates. *Mathematics in Practice and Theory*, 47(12), 234–238. [in Chinese]
- [4] Hadwiger H. Eine analytische reproduktionsfunktion für biologische gesamtheiten[J]. *Scandinavian Actuarial Journal*, 1940, 1940(3-4): 101-113.
- [5] Coale A J, Trussell T J. Model fertility schedules: variations in the age structure of childbearing in human populations[J]. *Population index*, 1974: 185-258.
- [6] Hoem J M, Madsen D, Nielsen J L, et al. Experiments in modelling recent Danish fertility curves[J]. *Demography*, 1981, 18(2): 231-244.
- [7] Chandola T, Coleman D A, Hiorns R W. Recent European fertility patterns: Fitting curves to 'distorted' distributions[J]. *Population Studies*, 1999, 53(3): 317-329.
- [8] Peristera P, Kostaki A. Modeling fertility in modern populations[J]. *Demographic Research*, 2007, 16: 141-194.
- [9] Gayawan E, Adebayo S B, Ipinoyomi R A, et al. Modeling fertility curves in Africa[J]. *Demographic Research*, 2010, 22: 211-236.