Vol.10.Issue.1.2022 (Jan-Mar)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com

Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

ESTIMATING TOTAL FERTILITY RATE USING INFORMATION ON PROPORTION OF FEMALES OF DIFFERENT REPRODUCTIVE AGE GROUPS

BRIJESH P. SINGH, ABHAY K. TIWARI & VAISHALI PATEL*

Department of Statistics, Institute of Science, Banaras Hindu University

Varanasi-221005, Uttar Pradesh, India.

*Coresponding author

DOI:10.33329/bomsr.10.1.1

ABSTRACT

Total Fertility Rate (TFR) is the most acceptable and widely used index of current situation of fertility. There are several direct and indirect methodologies are available to obtain TFR. Indirect methodologies are popularly used because TFR is based on age specific fertility rate which required the total number of births for different age groups of female and age of females also. It would be very difficult when the target population is illiterate or older. In this case we observe errors in age of females due to recall bias, misreporting or digit preference. In this study we have discussed a method for obtaining TFR given by Coale-Demeny (1967) and it relevancy in the present time with some modification. We have also tried to show the stability of the model using appropriate statistical tool.

Keywords: Total fertility rate; Coale-Demeny; regression; coefficient of determination.

Introduction

Total fertility rate (TFR) is a synthetic measure of fertility that is independent of age structure of population and is best single measure to compare fertility across the population. Crude birth rate is simplest among all the indices for measuring fertility, but it suffers from some drawbacks as it is affected by age and sex composition of population for which it is computed. TFR overcomes this drawback that is why it is the most accepted and widely used index of current fertility but TFR has its some limitations also. The demographic data in developing countries are erroneous, however, carefully planned and executed. Therefore, it is quite difficult to estimate the actual fertility level of population through the use of conventional TFR formula. In situations where conventional vital statistics are lacking or thought to be inadequate, techniques that use indirect evidence from cross-sectional reports or retrospective data are useful in the estimation of fertility parameters.

The job of demographers is to understand and define the fertility behaviour through statistical techniques in a proper manner and also given concern to study the differentials and determinants of fertility. A large number of indirect techniques have been proposed by researchers to estimate the TFR. For the estimation of TFR, Brass (1968) suggested a P/F ratio method for estimating fertility and its advancement has been done studied Hobcraft et al. (1982). After that Cho et al. (1986) have suggested own child method which contains reverse survival technique (15 years) for estimating age specific fertility rate (ASFR). Moreover, stable population method has been used by Rele (1967) for estimating TFR. With the use of sample registration system some modification has been done by Swamy et al. (1992). To overcome the difficulties present in the above mentioned methods some regression technique has been used indirect estimation of TFR.

Mauldin and Ross (1991), Jain (1997) have used contraceptive prevalence rate (CPR) to predict TFR and Singh et al. (2012) modified this model by taking the combination of CPR and sterility as a predictor variable to predict TFR and birth averted due to contraception. Birth interval reflects the reproductive behavior of a woman. Open birth interval is the interval from the date of last live birth to the date of survey. Srinivasan (1980) took mean OBI as predictor in predicting GMFR. But in calculating mean OBI the large intervals affect the mean in two ways: firstly the chance of error due to recall lapse is more, secondly very large intervals will themselves tend to increase the mean value. Therefore keeping this into mind, Yadava and Kumar (2002) have estimated TFR using percentage of currently married women having OBI greater than equal to 60 months.

Further, Yadava et al. (2009) proposed another predictor which is the weighted average of proportions of different birth orders and estimated the TFR. Singh et al. (2020) consider discrete variable rather than time variable that has a less chance of getting this type of error. They used proportion of women having birth in last five years before the survey date (PWBL5Y) as predictor variable and found more than 95 percent explanation in TFR in India and various states. Tiwari et al. (2020) used proportion of childless women in last 5 years before the survey date to explain TFR. In this paper they have shown there is no significant effect of contraceptive use in the reduction of TFR further. These predictor variables give quite reliable prediction for TFR but there are a number of factors which affected fertility negatively and positively as well. Most of the demographers developed models with single predictors or combination of two with coefficient of determination up to 0.9 or more to predict the TFR. In this paper an attempt has been explore the procedure given by Coale-Demeny (1967) and it relevancy in the present time with some modification. We have also tried to show the stability of the model using appropriate statistical tool.

Coale-Demeny procedure

The shape of the fertility schedules in populations that utilize little birth control differs primarily in the way in which fertility first rises from the start of childbearing to the ages where fertility is a maximum, and relatively much less in the way fertility declines after the peak is reached. This greater relative variability in the early part of fertility schedules results from the fact that the rise of fertility with age is strongly affected by customs and institutions governing the establishment of sexual unions-strongly affected, that is to say, by the age pattern of nuptiality in societies where formal marriage is a principal determinant of cohabitation.

These considerations suggest the hypothesis that the ratio of the average parity of women at the end of child-bearing to the average parity of a younger group (say women 25-30) is closely related to the relative parity of women early and late in their twenties. The reasoning behind the hypothesis is as follows: *(a)* if the average parity of women 25-30 is an unusually large multiple of the average parity of women 20-25. *(b)* On the other hand, an unusually low ratio of parity at 25-30 to parity at 20-25 indicates that high rates of child-bearing began early, that an unusually small fraction of total fertility occurs in the later years of child-bearing, and that the ratio of final average parity to the average at 25-30 is unusually low. Suppose that the average number of children ever born (average parity) to women l5-20 is designated P_1, to women 20-24 P_2, and so on, until P_7 designates the average parity of women 45-50, Suppose the average parity of women reaching age 50 assumed to be the upper limit of child bearing is designated TFR.

This procedure permits the estimation of the average number of children born to women by the end of their reproductive life. If P_2 and P_3 are average parities for women in age groups 20-25 years and 25-30 years respectively, the relationship between the ratios $\frac{TFR}{P_3}$ and $\frac{P_3}{P_2}$ is very close. Total fertility, estimated from average parities of younger women, is represented by $TFR = \frac{P_3^2}{P_2}$.

The age pattern of fertility should also conform to the typical age relationship found in populations practicing little birth control. The procedure is expected to give reasonable total fertility estimates for the Gilbert and Ellice Islands. It was, however, found not to be robust when it was applied to West African data (Brass, 1975).

An alternative formula proposed by Brass and Rachad (1979) also estimates total fertility from average parities of younger women. If P_4 is the average parity for women aged 30-34, then $TFR = P_2\left(\frac{P_4}{P_3}\right)^4$.

The analysis of the above methods suggested disparities between current fertility data and information on children ever born reported by the same women. Coale-Demeny procedure is relatively simple and needs only on lifetime fertility data which is likely to be reported more accurately. This estimate total fertility from the average parities of younger women, because in Indian traditional settings where birth control is not widely practiced and rise and fall of fertility with age is strongly affected by social customs, the shape of the early part of the fertility is dominated by the age pattern of entry into marriage. These suggest that the ratio of the average parity of a younger, say 25-30 years age group, is closely related to relative parity of women in their early and late twenties. Hence $\frac{P_3^2}{P_2}$ is appropriate for estimating TFR when older women are likely to under report the number of children ever borne and the younger to report parity accurately. The Coale-Demeny procedure depicted data from a very small age range as a basis for extrapolation to all reproductive years. The Brass-Rachad procedure used a wider range of ages. In that, respect the use of a wider age range in the Brass-Rachad procedure seems to amplify errors in the basic data. It may be logical to conclude that the Coale-Demeny's procedures less affected by reporting errors since it use more reliable average parities. Yadava and Tiwari (2007) modified Coale-Demeny (1967) by taking $\frac{P_3^2}{P_2}$ and percentage of current contraceptive users jointly as predictors. Another modification has been done by Gupta et al. (2014) considering situation of current time point in which the age at marriage is increasing and the fertility schedule is shifting towards higher ages and estimated TFR has been obtained using $\frac{P_4^2}{P_3}$ as a predictor variable, where P_3 and P_4 are mean births to females of age groups 25-30 and 30-35 respectively.

Source of data

In this study, the data has been taken from the National Family Health Survey (NFHS). From the mid-1990s, the Ministry of Health and Family Welfare, Government of India has been developing on the country's National Family Health Survey (NFHS) to monitor and evaluate the family planning and reproductive and child health programs both national level and individual states. For this study authors have taken data on above mentioned variables from NFHS-4.

Table 1. Regression Models, r^2 (Coefficient of Determination), Adjusted r^2 and Standard Error

Model	Mathematical form	r^2	Adjusted r^2	Standard error
1	$0.410\left(\frac{P_3^2}{P_2}\right)+0.384$	0.558	0.530	0.317
2	$0.588\left(\frac{P_4^2}{P_3}\right)+0.157$	0.779	0.765	0.224
3	$0.621\left(\frac{P_5^2}{P_4}\right)+0.128$	0.858	0.849	0.180
4	$0.162\left(\frac{P_3^3}{P_2}\right)+0.884$	0.840	0.830	0.190
5	$0.132\left(\frac{P_4^3}{P_3}\right)+1.014$	0.880	0.872	0.165
6	$0.112\left(\frac{P_5^3}{P_4}\right)+1.079$	0.894	0.887	0.156
7	$0.061\left(\frac{P_3^4}{P_2}\right)+1.261$	0.893	0.887	0.156

Predictor of all the models is highly significant and constant term of model 4, 5, 6 and 7 is also highly significant however constant term of model 1, 2 and 3 is insignificant.

Model validation of the fitted model

Cross validity prediction power

It is necessary to find out the estimate that a predictive model will perform in practice or to know how much the proposed model is stable over population. In this respect an appropriate technique known as cross validity prediction power (CVPP) given by Herzberg (1969) have been utilized which is given as

$$\rho_v^2 = 1 - \frac{(n^2-1)(n-2)(1-c^2)}{n(n-p-1)(n-p-2)}$$

Where *n* is the number of cases, *p* is the number of explanatory variables in the model and *c* is the correlation coefficient between predicted and observed value of the dependent variable TFR. It is well known in the regression analysis that a fitted relationship performs less well on a new data set than on the data set that is used for fitting (Everitt, 2002), so that the value of the coefficient of determination, particularly, 'shrinks'. Shrinkage is separate from the standard adjustment made in the coefficient of determination. The shrinkage of the model is estimated by the following formula:

Shrinkage=$|\rho_v^2 - r^2|$ where r^2 is the coefficient of determination. Finally, we have also calculated the stability of the model which is equal to (1-Shrinkage) which implies that the lower the shrinkage the more stable the model.

Table 2. Correlation between Observed and Predicted Value of TFR (c^2)

RMSE, ρ_v^2 and Stability of r^2

Model	c^2	**RMSE**	ρ_v^2	**Shrinkage of r^2**	**Stability of r^2**
1	0.746	0.292	0.320	0.238	0.762
2	0.882	0.206	0.659	0.120	0.880
3	0.925	0.166	0.779	0.079	0.921
4	0.915	0.177	0.750	0.090	0.910
5	0.938	0.152	0.816	0.064	0.936
6	0.945	0.143	0.836	0.058	0.942
7	0.943	0.147	0.830	0.063	0.937

Results and discussion

In the present modern time where women's education and autonomy level is increasing, also their reproductive knowledge is growing; we observe a shift in fertility schedule towards higher ages. Therefore the procedure given by Coale-Demeny (1967) needs modification. In this study we have tried to develop a relationship between TFR and different combination of average parities. In some states of India the age at marriage is considerably high thus we have considered average parities of ages from 20-40 i.e. P_2, P_3, P_4 and P_5.

We have calculated cross validity power prediction (CVPP) for all models considered in this study. The value of CVPP for all models given in Table 2 and it is maximum for Model 6 and the stability of model is obtain as (1-Shrinkage) and the estimate of stability is also maximum for Model 6. Therefore this model is more stable for the population than other models considered in this study. Although the Model 7 is equally good as Model 6, therefore we can use it also. The observed and estimated values of TFR for all the major states of India using all proposed regression models are given in Table 3. A critical review of the results given in Table 1, we observe that the constant term of the model 1, 2, and model 3 is small and insignificant; means in the absence of the predictors, models provide a very small value of dependent variable, while the constant term of the model 4, 5, 6, and model 7 is significant and plays an important role. The value of coefficient of determination is more than 80 percent for model 3 to 7 and maximum for model 6 and 7. The standard error of model 1 and 2 is high as compared to model 3 to 7 and it is lowest for model 6 and 7.

Table 2 represents the correlation between observed and predicted value of TFR, Root mean square error (RMSE), Cross-validity prediction power, Shrinkage of r^2 and Stability of r^2 for all seven models. The value of correlation coefficient is high for all models except models 1 and 2 and it is maximum for model 6. Root mean square error is low for all models except model 1 and 2 and minimum for model 6. Cross-validity prediction power is low for model 1 whereas high for model 5, 6, and 7 and it is maximum for model 6. Since lower shrinkage provides more stability to the model therefore models 3, 4, 5, 6 and 7 are more stable then model 1 and 2 and also model 6 is highly stable among all.

Table 3 shows the observed and predicted TFR from different models and also the percentage error between observed and predicted values. Estimated TFR through model 3 to 7 are much close to observed TFR rather than the model 1 and 2. The value of percentage error through model 1 is range from -38 to 24 and through model 2 is range from -23 to 15 while for models 3, 4, 5, 6, and 7 the percentage error is range from -15 to 15. In Table 3, the negative values of percentage error indicate over estimate and positive values indicate under estimate of TFR. We have seen from the Table 3 that the models 1 to 5 have given ten under estimated TFR and nine over estimated TFR which is similar to the model 6 with eleven under-estimated and eight over-estimated while model 7 behave different and having six under-estimated and thirteen over-estimated values. Estimation through the model 1 has given poor estimate for the states Kerala and Bihar and through model 2 the poor estimate has given for state Punjab.

Conclusion

Coale-Demeny procedure using only information on average parities for women in their twenties and not suitable for the present situation. The possible reason may be the late age at marriage and adequate gaps between the births. Thus the assumption of shifting of fertility schedule is seems to be appropriate. Analysis indicates that the ratio of cubic term rather than square term of the average parities of the preceding age group to the present age group explains TFR better.

Table 3. Observed and Predicted TFR from different Models for India and Major States

States	Observed	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
		Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error
Andhra Pradesh	1.83	1.72	6.07	1.60	12.76	1.60	12.49	1.82	0.54	1.69	7.65	1.67	8.85	1.88	-2.84
Bihar	3.41	2.58	24.44	2.89	15.30	2.96	13.10	2.99	12.29	3.08	9.57	3.08	9.73	3.18	6.83
Chhattisgarh	2.33	2.16	7.18	2.15	7.87	2.27	2.70	2.12	8.84	2.11	9.45	2.20	5.78	2.08	10.78
Gujarat	2.03	1.96	3.24	2.07	-1.93	2.06	-1.30	1.93	4.94	2.02	0.51	2.01	0.85	1.92	5.63
Haryana	2.05	2.19	-6.72	2.08	-1.57	2.07	-0.99	2.14	-4.46	2.06	-0.34	2.04	0.64	2.09	-1.97
Jharkhand	2.55	2.17	14.72	2.36	7.30	2.39	6.39	2.32	8.91	2.39	6.34	2.37	7.15	2.35	7.65
Karnataka	1.80	1.82	-0.87	1.79	0.30	1.73	4.05	1.79	0.45	1.79	0.32	1.75	2.79	1.81	-0.32
Kerala	1.56	2.16	-38.63	1.77	-13.32	1.52	2.69	1.71	-9.66	1.67	-6.76	1.58	-1.27	1.62	-4.14

States	Observed	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
		Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error	Estimate	% error
Madhya Pradesh	2.32	2.27	2.17	2.39	-3.02	2.45	-5.52	2.41	-4.04	2.42	-4.29	2.43	-4.63	2.44	-4.98
Maharashtra	1.87	1.86	0.65	1.90	-1.84	1.86	0.54	1.89	-0.98	1.91	-1.90	1.86	0.45	1.91	-2.06
Odisha	2.05	1.84	10.40	1.99	2.86	2.00	2.27	1.75	14.45	1.91	6.70	1.94	5.20	1.75	14.45
Punjab	1.62	1.93	-19.08	2.00	-23.51	1.86	-14.86	1.65	-1.90	1.84	-13.49	1.81	-11.86	1.62	-0.07
Rajasthan	2.40	2.52	-5.01	2.40	0.20	2.38	0.71	2.63	-9.40	2.43	-1.16	2.37	1.11	2.60	-8.52
Tamil Nadu	1.70	1.78	-4.87	1.62	4.82	1.62	4.77	1.70	0.03	1.64	3.28	1.65	2.73	1.71	-0.64
Uttar Pradesh	2.74	2.95	-7.61	2.98	-8.73	2.97	-8.50	2.95	-7.82	3.01	-9.73	3.02	-10.17	2.84	-3.74
Uttarakhand	2.07	2.39	-15.45	2.43	-17.49	2.35	-13.56	2.25	-8.72	2.34	-12.98	2.29	-10.87	2.14	-3.48
West Bengal	1.77	1.62	8.23	1.73	2.47	1.87	-5.45	1.70	3.68	1.76	0.47	1.84	-3.93	1.77	-0.28
Telangana	1.78	1.90	-7.02	1.67	6.38	1.83	-2.61	1.98	-11.25	1.75	1.68	1.82	-2.36	2.01	-12.85
India	2.18	2.10	3.90	2.24	-2.78	2.25	-3.36	2.07	5.04	2.18	-0.22	2.20	-0.73	2.04	6.47
Under estimate			10		10		10		10		10		11		6
Over estimate			9		9		9		9		9		8		13

References

[1]. Brass, W. (1968). Methods of analysis and estimation, In: The Demography of Tropical Africa. Edited by W. Brass et al., Princeton University Press, Princeton.

[2]. Brass, W. (1975). Methods for Estimating Fertility and Mortality from Limited and Defective Data, Laboratories for Population Statistics, Chapel Hill, NC, University of North Carolina.

[3]. Brass, W. and Rachad, H. (1979). Evaluation of Levels and Trends in Fertility. WFS data: Bangladesh. Unpublished Manuscript.

[4]. Cho, Lee-Jay, Retherford, R. D. and Choe, M. K. (1986). The Own-Children Method of Fertility Estimation. Honolulu: University of Hawaii Press.

[5]. Coale, A., and Demeny, P. (1967). Methods of estimating basic demographic measures from incomplete data, manuals on methods of estimating population, Manual 4, New York: United Nations, Department of Economics and Social Affairs.

[6]. Everitt, B.S. (2002). Cambridge Dictionary of Statistics, 2nd Edition. OUP.

[7]. Jain, A. (1997). Consistency between Contraceptive Use and Fertility in India. *Demography India,* 26(1) 19-36.

[8]. Gupta, Kushagra, Singh, Brijesh P. and Singh, K.K. (2014). Estimation of Total Fertility Rates in India using Indirect Techniques.*Journal of National Academy of Mathematics*, *28, 21-28*.

[9]. Herzberg, P. A. (1969). The parameters of cross validation. *Psychometrika Monograph Supplement*, 16. 34(2), 1-70.

[10]. Mauldin, W. P. and Ross, J. A. (1991). Family planning programmes: Efforts and Results, 1982-1989. *Studies in family planning,* 22 (6), 350-367.

[11]. Rele, J. R. (1967). Fertility analysis through extension of stable population concepts. Berkeley: Institute of International Studies, University of California. (Ph.D. dissertation).

[12]. Singh, K. K., Singh, Brijesh. P., and Gupta, K. (2012). Estimation of total fertility rate and birth averted due to contraception: regression approach. *International Journal Of Statistics And Applications*. 2(5), 47-55.

[13]. Singh, Brijesh. P., Singh, N., and Singh, S. (2020). Estimation of total fertility rate: an indirect approach using auxiliary information. *Journal of the Social Sciences.* 48(3),789-798.

[14]. Srinivasan, K. (1980). Birth interval analysis in fertility surveys. *International Statistical Institute.*

[15]. Swamy, V. S., Saxena, A. K., Palmore James A., Mishra, Vinod, Rele, J. R. and Luther Norman Y. (1992). RGI: Evaluating the sample registration system using indirect estimates of fertility and mortality, New Delhi: *Registrar General of India; Occasional Paper*.

[16]. Tiwari, A. K., Singh, Brijesh P. & Patel, V. (2020). Retrospective Study of Investigation of Possible Predictors for Total Fertility Rate in India. *Journal of Scientific Research & Reports*, 26(9), 111-119.

[17]. Yadava, R. C. & Kumar, A. (2002). On an Indirect Estimation of Total Fertility Rate from Open Birth Interval*. Demography India,* 31(2), 211-222.

[18]. Yadava, R. C., Tiwari, A. K. (2007). A modified procedure to estimate total fertility rate. *Journal of Empirical Research in Social Science*2(1), 82-87.

[19]. Yadava, R. C., Tiwari, A. K., and Sharma, S. S. (2009). Indirect measurements of total fertility rate. *The Journal of Family Welfare*, *55*(2), 70-73.