Estimation of Total Fertility Rate in Smaller Domain Using Female's Current Age

Brijesh P. Singh
Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi

Introduction

Fertility is the primary engine of global population change (Gerland et al. 2014). Reduction in fertility is argued to be essential for sustainable development including female education, child and maternal mortality, gender equality, and reproductive health (Abel et al. 2016). Total fertility rate (TFR) is a completed measure of fertility that is commonly used in demographic and development research. These include the direct method of calculating TFR requires data on births disaggregated by the age of the female which are generally not available at the local level. As such, different indirect methods of estimating TFR have been suggested in the literature. These include Brass P/F Ratio technique (Brass 1975), Own Children method (Cho et al. 1986), and several regression methods (Bogue and Palmore 1964; Palmore 1978; Gunasekaran and Palmore 1984; and Rele 1967 and 1987; Pacheco and Engracia 1985; Rao 1987; Hanenberg1983). Coale and Demeny(1967) developed a formula (TFR= P_3^2/P_2) to estimate the total fertility rate, where P_2 and P_3 represent mean births to females of age group (20-24) and (25-29). Singh et al. (2020) have used the proportion of females having birth in the last five years prior to the date of survey date as the predictor variable, suggested the error like recall lapse in count that is lesser than the time variable (open birth interval more than five years) as taken by Yadava and Kumar (2002). In addition, the relationship between fertility and social indices can differ over time and over populations, making indirect methods error prone when applied outside the context from which regression coefficients were derived (Tuchfeld et al. 1974, Hauer et al. 2013, Schmertmann & Hauer 2019). Moreover, instead of using the expectation of life at birth, Palmore (1978) method uses IMR as the measure of mortality. In this paper, we develop a regression method which is based on the age distribution of currently married females in the reproductive age group (15-49 years). The advantage of the method being proposed here is that the data pertaining to the age distribution of currently married reproductive age females can be culled out from the records maintained by the grassroots level health and family welfare services providers. As such, the method can be used for estimating TFR at the local level which, then, can become the basis for planning for the delivery of family welfare services and monitoring fertility transition at the grassroots level.

Development of the Model

The model proposed for estimating TFR is based on establishing an empirical relationship between TFR and the age distribution of currently married females in the reproductive age group. The age distribution of currently married females in the reproductive age group can be characterised in terms of the first four moments of the age distribution. In the present exercise, we have used coefficient of variation (*CV*), involving first and second moments; skewness (*Sk*), involving third and second moments; and kurtosis (*Ku*), involving fourth and second moments, to characterise the age distribution of currently married females in the reproductive age group. Using these three indicators of the age distribution, we have examined the empirical relationship between TFR and the age distribution of females on the basis of the following seven models:

Model 1 $\Rightarrow TFR = f(CV, Sk, Ku)$

Model 2 $\Rightarrow TFR = f(CV, Sk)$ **Model 3** $\Rightarrow TFR = f(CV, Ku)$

Model 4 $\Rightarrow TFR = f(Sk, Ku)$ **Model 5** $\Rightarrow TFR = f(CV)$
Model 6 $\Rightarrow TFR = f(Sk)$ **Model 7** $\Rightarrow TFR = f(Ku)$

In order to test the cross-validity prediction power of different models, we have used the method proposed by Herzberg (1969). The cross-validity prediction power of the model is calculated 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 observations, p is the number of explanatory or independent variables in the model and c is the correlation coefficient between predicted and observed value of the dependent variable or TFR. Moreover, standard adjustment has been made in the coefficient of determination to compensate for the subjective effects of further sampling. Moreover, we have also estimated shrinkage which is the reduction in the effects of sampling variation. 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 formula given as

Shrinkage= $|\rho_v^2 - r^2|$

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

Source of Data

In order to examine the empirical relationship between TFR and the age distribution of currently married females in reproductive age, we have used the available from the fourth round of the National Family Health Survey (NFHS) which was conducted during 2015-16.

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	TFR =44.708(*CV*)-3.699(*Sk*)+7.724(*Ku*) -3.149	0.900	0.870	0.181
2	TFR =31.885(*CV*)+0.759(*Sk*) -8.352	0.797	0.760	0.246
3	TFR =28.937(*CV*)+2.768(*Ku*) -4.247	0.854	0.817	0.215
4	TFR =2.191(*Sk*)+2.402(*Ku*)+4.181	0.653	0.590	0.322
5	TFR=38.717(*CV*)-10.381	0.787	0.769	0.241
6	TFR =3.624(*Sk*)+1.350	0.640	0.610	0.313
7	TFR =6.559(*Ku*)+9.157	0.609	0.576	0.327

CV=Coefficient of variation, *Sk*=Skewness and *Ku*=Kurtosis

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	0.928	0.742	0.788	0.112	0.888
2	0.868	0.968	0.708	0.089	0.911

3	0.900	0.859	0.775	0.079	0.921
4	0.796	1.167	0.566	0.087	0.913
5	0.881	0.993	0.733	0.057	0.943
6	0.783	1.335	0.541	0.099	0.901
7	0.770	1.206	0.518	0.092	0.908

Results and Discussion

Table 2 shows the relationship between TFR and different indicators characterizing the current age distribution of females. In all models, the regression coefficients are statistically significant but the coefficient of determination is the highest in Model 1. The standard error of r^2 is also the lowest in Model 1. On the other hand, table 3presents the correlation between observed and predicted values of TFR (c^2), root mean square error (RMSE), Cross-validity prediction power ρ_v^2 and Stability of r^2 for different models. The value of c^2 is maximum for Model 1 whereas the root mean square error is the lowest for Model 1. Cross validity prediction power ρ_v^2 also suggests that Model 1 is the most powerful of the seven models. However, the shrinkage is the lowest in case of model 5 whereas the stability is high in all the seven models. Model 1 shows highest r^2 and ρ_v^2 and the lowest RMSE but shrinkage of the model is high so that its stability is quite low. Ranking all the seven models in terms of five parameters r^2, RMSE, ρ_v^2, shrinkage of r^2 and stability of r^2 the mean rank score has been found to be the lowest in case of model 5 which means that model 5 best estimates TFR among the seven models. A comparison of the observed and estimated values of TFR based on different models is given in table 3. The predicted value of TFR is close to the observed values in case of all models. The percent difference between observed and estimated TFR is also shown in the table. For model 1, 2, 3, 4, 6 and 7, more than half of the states show over estimation in TFR and for Model 5, only 8 states show over estimation. An interesting result has been observed that all models provide under estimated TFR for Bihar and over estimated TFR for Uttarakhand for NFHS-4. Perhaps a possible reason may be the variability in the age distribution of females.

Conclusion

This paper proposes a simple method of estimating TFR from the current age distribution of females of childbearing age. The empirical validity of different models suggests that Model 5 involving the first two moments of the current age distribution of females of reproductive age is the most appropriate to estimate TFR. This model is used for comparison of estimate of TFR of all districts in Uttar Pradesh and also found suitable for further use.

Table 3. Observed and Predicted TFR through Various Regression Models for Some States of IndiaNFHS-4

States	Observed TFR	Estimated TFR by													
		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
		estimate	% Diff.	estimate	% Diff.	estimate	% Diff.	estimate	% Diff.	estimate	% Diff.	estimate	% Diff.	estimate	% Diff.
Andhra Pradesh	1.83	1.58	13.72	1.63	10.87	1.62	11.29	1.78	2.97	1.62	11.40	1.82	0.49	1.81	1.04
Bihar	3.41	3.12	8.64	2.75	19.36	2.88	15.49	2.68	21.30	2.75	19.37	2.73	20.03	2.79	18.04
Chhattisgarh	2.23	2.54	-13.69	2.55	-14.44	2.54	-14.02	2.41	-7.89	2.56	-14.58	2.55	-14.17	2.34	-4.74
Gujarat	2.03	1.82	10.17	2.00	1.65	1.90	6.41	1.84	9.40	2.03	0.08	1.97	3.15	1.75	14.02
Haryana	2.05	2.22	-8.44	2.11	-2.88	2.25	-9.83	2.50	-22.05	2.02	1.60	2.55	-24.19	2.60	-26.73
Jharkhand	2.55	2.68	-5.27	2.45	3.89	2.58	-1.00	2.57	-0.78	2.41	5.34	2.62	-2.68	2.66	-4.45
Karnataka	1.80	1.82	-0.85	1.77	1.65	1.81	-0.29	1.94	-7.52	1.75	2.58	1.97	-9.23	2.01	-11.54
Kerala	1.56	1.66	-6.20	1.51	3.36	1.45	7.14	1.30	16.49	1.61	-3.19	1.28	18.11	1.42	9.14
Madhya Pradesh	2.32	2.33	-0.31	2.50	-7.80	2.43	-4.62	2.29	1.29	2.51	-8.26	2.47	-6.61	2.14	7.81
Maharashtra	1.87	2.03	-8.38	2.15	-15.20	2.06	-9.99	1.91	-1.98	2.20	-17.72	2.04	-9.01	1.81	3.16
Odisha	2.05	2.00	2.57	2.18	-6.35	2.09	-1.83	2.00	2.60	2.21	-7.57	2.15	-4.75	1.88	8.46
Punjab	1.62	1.76	-8.53	1.64	-1.27	1.72	-6.30	1.94	-19.59	1.61	0.91	1.93	-19.13	2.07	-27.98
Rajasthan	2.40	2.50	-4.06	2.52	-4.86	2.55	-6.06	2.52	-4.99	2.48	-3.53	2.65	-10.61	2.47	-2.78
Tamil Nadu	1.70	1.77	-3.89	1.65	2.95	1.65	3.10	1.64	3.41	1.69	0.57	1.64	3.53	1.75	-2.67
Telangana	1.78	1.67	6.37	1.83	-2.58	1.86	-4.27	2.16	-21.27	1.75	1.83	2.26	-26.74	2.14	-20.16
Uttarakhand	2.07	2.38	-15.02	2.45	-18.36	2.42	-16.70	2.29	-10.73	2.46	-18.80	2.44	-17.74	2.20	-6.50
Uttar Pradesh	2.74	2.90	-5.82	2.92	-6.51	2.90	-5.95	2.65	3.16	2.93	-6.80	2.84	-3.50	2.53	7.58
West Bengal	1.77	1.67	5.84	1.83	-3.16	1.86	-4.86	2.16	-21.96	1.75	1.28	2.26	-27.46	2.14	-20.84
INDIA	**2.18**	**2.22**	**-1.63**	**2.29**	**-4.83**	**2.26**	**-3.63**	**2.20**	**-1.03**	**2.29**	**-4.90**	**2.33**	**-6.81**	**2.14**	**1.89**

Table 4. Estimated TFR for all districts of India and 95% Confidence Interval using Model 5

States	Districts	Estimated TFR	95% Confidence Interval	
			Lower	Upper
India	**Total**	**2.29**	**2.13**	**2.44**
Uttar Pradesh	Saharanpur	2.69	2.45	2.93
	Muzaffarnagar	2.74	2.48	3.00
	Bijnor	3.09	2.73	3.45
	Moradabad	3.20	2.81	3.59
	Rampur	3.19	2.80	3.58
	Jyotiba Phule Nagar	3.24	2.84	3.65
	Meerut	2.72	2.47	2.97
	Baghpat	2.92	2.61	3.23
	Ghaziabad	2.64	2.41	2.87
	Gautam Buddha Nagar	2.06	1.92	2.20
	Bulandshahr	2.94	2.63	3.25
	Aligarh	2.90	2.60	3.20
	Mahamaya Nagar	2.84	2.56	3.12
	Mathura	3.22	2.82	3.61
	Agra	2.92	2.61	3.23
	Firozabad	3.26	2.85	3.66
	Mainpuri	3.23	2.83	3.63
	Budaun	3.12	2.75	3.49
	Bareilly	3.34	2.91	3.78
	Pilibhit	2.95	2.64	3.27
	Shahjahanpur	2.99	2.66	3.32
	Kheri	2.62	2.40	2.85
	Sitapur	2.62	2.40	2.85
	Hardoi	2.81	2.53	3.08
	Unnao	3.14	2.77	3.52
	Lucknow	2.33	2.17	2.50
	Rae Bareli	3.43	2.97	3.89
	Farrukhabad	2.90	2.59	3.20
	Kannauj	2.94	2.63	3.25
	Etawah	2.87	2.58	3.17
	Auraiya	2.38	2.21	2.56
	Kanpur Dehat	2.53	2.33	2.74
	Kanpur Nagar	2.32	2.16	2.49
	Jalaun	2.56	2.35	2.77
	Jhansi	2.67	2.43	2.91
	Lalitpur	2.40	2.23	2.58
	Hamirpur	2.93	2.62	3.24
	Mahoba	3.29	2.87	3.71
	Banda	2.89	2.59	3.19
	Chitrakoot	3.38	2.93	3.83
	Fatehpur	2.90	2.60	3.20
	Pratapgarh	3.15	2.78	3.53
	Kaushambi	3.37	2.93	3.82
	Allahabad	2.78	2.51	3.04
	Bara Banki	2.76	2.50	3.02
	Faizabad	3.25	2.85	3.66
	Ambedkar Nagar	3.07	2.72	3.42
	Sultanpur	3.42	2.96	3.88
	Bahraich	2.22	2.07	2.36
	Shrawasti	2.27	2.12	2.43
	Balrampur	2.71	2.46	2.95
	Gonda	2.96	2.64	3.27
	Siddharth Nagar	3.08	2.73	3.44
	Basti	3.12	2.75	3.48
	Sant Kabir Nagar	3.24	2.84	3.65
	Mahrajganj	3.51	3.02	3.99
	Gorakhpur	2.98	2.65	3.30
	Kushinagar	3.37	2.93	3.81
	Deoria	3.02	2.68	3.36
	Azamgarh	3.19	2.80	3.57
	Mau	3.30	2.88	3.73
	Ballia	3.23	2.83	3.63
	Jaunpur	2.95	2.64	3.27
	Ghazipur	3.22	2.82	3.62
	Chandauli	2.89	2.59	3.19
	Varanasi	2.60	2.38	2.82
	Sant Rvidas Nagar (Bhadohi)	2.87	2.58	3.16
	Mirzapur	2.72	2.47	2.97
	Sonbhadra	2.55	2.34	2.76
	Etah	2.99	2.66	3.32
	Kanshiram Nagar	3.19	2.80	3.58
	Total	**2.93**	**2.62**	**3.24**