ABSTRACT

Total fertility rate (TFR) is the most acceptable and widely used measure of current fertility. Since TFR is based on age-specific fertility rate which required the total number of births in different age groups as well as age of female. When the population is illiterate or older then the information on age may have some recall bias, misreporting digit preference etc., thus in this situation TFR may departed from the actual. Therefore, need some indirect methodology which enables us to have an idea about the estimation of TFR. In this study an attempt has been made to identify some predictors that the explain TFR and try to suggest the best combination of predictors to get estimate of TFR. The methodology used in this study is essentially based on the regression technique. The identification and acceptance of possible predictors are based on the coefficient of determination. The data for the major states of India from National Family Health Survey (NFHS 4) is used for the analysis.

Keywords: Total fertility rate; predictors; regression; coefficient of determination.
1. INTRODUCTION

Fertility is considered one of the most important factors in the study of population dynamics. It refers to the actual reproductive outcomes and significantly affected by many demographic, socio-economic, cultural and biological factors. Total fertility rate (TFR) is a synthetic measure of fertility that is independent on age structure of population and is best single measure to compare fertility across the population. TFR represents the average number of children that would be born to a woman till the end of her childbearing capability if she was to pass through all her reproductive years conforming to the age-specific fertility rates (ASFR) of given year. Crude birth rate is simple of all indices for measuring the 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. Since TFR is based on ASFR, the fact that computation of various ASFR for different age's needs to knowledge about ages of all the females included in the study and therefore, the question on age which subject to various errors like recall lapse, age heaping, etc. is occurred. In any survey the data for developing countries, however, carefully planned and executed, are subject to large errors such as error or omission of events, error in identification of time period in which event occurred. Hence it is quite difficult to estimate the actual fertility of population accurately through the use of TFR especially when age misreporting is common. Further ASFR related to births to women in one year and chance for a woman to bear a child in single year is very low, therefore, a comparatively large sample size is required to compute ASFR for desired degree of accuracy. Thus in such case, certain indirect method of computation of TFR is required and adopted which may not need such defective data and may also be so designed to possess the computational ease and accuracy of results.

2. LITERATURE REVIEW

The demographers have given supremacy to understanding and define the fertility behaviour through statistical techniques in proper manner and also given priority to study the differentials and determinants of fertility through various mathematical and statistical models. A large number of indirect techniques have been proposed by researchers to estimate the TFR using exploratory variables. For the estimation of total fertility rate, Brass [1] suggested a P/F ratio method for estimating fertility and its advancement has been done studied Hobcraft et al. [2]. After that Cho et al. [3] have suggested own child method which contains reverse survival technique (15 years) for estimating age specific fertility rate (ASFR) from cross-sectional survey. Furthermore stable population method has been used by Rele [4] for estimating TFR’s. With the use of sample registration system some modification has been done by Swamy et al. [5]. To overcome the difficulties present in the above mentioned methods some regression technique has been used indirect estimation of TFR.

According to Coale and Demeny [6] have developed a formula (TFR= \( P_2^2/P_3 \)) 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) and further it was modified by Yadava and Tiwari [7] by taking \( P_2^2/P_3 \) and percentage of current contraceptive users jointly as predictors. Another modification has been done by Gupta et al. [8] considering situation of current time point and estimated TFR has been obtained by \( P_2^2/P_3 \) as a predictor variable, where \( P_3 \) and \( P_4 \) are mean births to females of age groups (25-29) and (30-34), respectively. Yadava and Kumar [9] have estimated TFR using percentage of currently married women having open birth interval greater than equal to 60 months. Further, Yadava et al. [10] proposed another predictor which is the weighted average of proportions of different birth orders and estimated the TFR. Jain [11] has used Contraceptive prevalence rate (CPR) to estimate total fertility rate of any population. Mauldin and Ross [12], Jain [11] have used CPR to predict TFR and Singh et al. [13] modified this model by taking the combination of CPR and sterility as a predictor variable to predict TFR of any population.

Some Demographers have used proportion of 3rd order births for estimation of total fertility rate. Gunasekaran and Palmore [14] suggested few regression models to provide estimates of fertility levels indirectly. Singh et al. [15] have used the proportion of women having birth in last five years before the survey date as predictor variable, suggesting the error like recall lapse in count that is lesser than the time variable as taken by Yadava and Kumar [9]. These predictor variables give quite reliable prediction for TFR, but there are a number of factors which affected fertility negatively and positively as well. Rai et al.
[16] used some variables and the concept of Singh et al. [13] for predicting TFR by using ridge regression. It has been observed that there are various variables which affect the TFR. 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 study an attempt has been made to identify various parameters which affect the TFR and tried to develop the regression model with minimum number of predictors which explain maximum variability in TFR.

2.1 Identification of Predictors

To establish the mathematical relationship between dependent and independent variables, the choice of predictors plays a significant role. The purpose of regression model is to estimate phenomena with possible high accuracy and optimum number of independent variable. An incomplete choice of predictors may provide vague results and sometimes misleading conclusions. The basic requirement to choose exploratory variables is that there should be high correlation between dependent and independent variables and data on independent variables should be easily obtained. This study attempts to identify various predictors from literature review which are mentioned below:

2.1.1 Woman age at marriage (AGM)

Age at marriage influencing the fertility, since most of the births take place within marriage in many traditional societies like India. The relationship between marriage and fertility suggests that women who marry at younger age produce more children than women marry late and the fecundity of women sharply rises as in increase in age at marriage whereas proportion of temporary sterility decreases [17], thus AGM is negatively associated with TFR.

2.1.2 Age at first birth (AFB)

An early age at first birth can have a negative effect on occupational attainment, marital stability, asset accumulation, and on the woman’s health, as well as a positive effect on the spacing of subsequent children and on completed family size [18,19,20,21,22,23]. This variable is negatively correlated with TFR.

2.1.3 Age at last birth (ALB)

The findings of Varea [24] shows that the final family size is independent of marital duration, but is correlated with maternal age at first and last births. Women with completed families, those with early age at marriage stopped childbearing about 10 years before reaching menopause, while women who married later continued to bear children until the end of their reproductive age. Therefore, ALB is positively correlated with TFR.

2.1.4 Reproductive span (RS)

Reproductive span is nothing but difference between age at first birth and last birth i.e. it is the duration in which a woman gives her total births. It is highly associated with fertility. If the reproductive span of a woman is larger than other then it is expected the number of birth of this woman will be more than other.

2.1.5 Proportion of contraceptive users (CU)

Individuals who practice contraception intends to avoid pregnancy, therefore, it is not surprising that the contraceptive use in population is negatively and causally related to fertility. Typically, TFR is around six to seven births per woman in countries with no contraceptive use while fertility is about two births per woman in countries in which the contraceptive prevalence rate (CPR) among women in union is around 75% [25]. Hence CU is negatively associated with TFR.

2.1.6 Proportion of childless women in last five years (PCL)

Proportion of childless women is negatively associated with TFR that means as childless women increases in last five years prior to the survey, TFR decreases. In this study ‘proportion of childless women in last five year from survey date’ is considered as predictor to predict TFR.

2.1.7 Proportion of women having 3+ order births (BO3)

Birth order analysis is important in understanding trends and differentials in fertility and its indirect implications to population growth [10]. Thus, birth order data may be a good choice for predictor for TFR, thus BO3 is positively correlated with TFR.

2.1.8 Open birth interval (OBI)

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 [26] took mean OBI as predictor in
3.1 Application of the Methodologies

The regression analysis technique has been used to estimate the total fertility rate. Mathematical relationship between dependent and independent variables using the concept of a linear regression and have tried to establish the relationship between TFR and different variables identified considering eighteen major states of India as the units of observation. Parameters in the model have been estimated by least square method.

First of all have obtained the quantitative measure of relationship of all predictors with TFR. It is observed that the some predictors are positively correlated and some are negatively correlated with TFR. Table 1, shows the correlation coefficient along with its $p$-value between all predictors with TFR. Age at last birth (ALB), three plus birth order (BO3), proportion of women having OBI five year or more (OBI), infant mortality rate (IMR), weighted average of proportions of different birth orders (PBW) and proportion of women having births in last five years (PWBL5Y) are positively correlated with TFR that means as value of predictors increases TFR also increases. On the other hand age at marriage (AGM), age at first birth (AFB), reproductive span (RS), proportion of contraceptive user (CU), proportion of childless women in last five years (PCL), and open birth interval (OBI) are negatively correlated that means as value of predictors increases TFR decreases. Some of the predictors (RS, PBW, PCL, OBI and PWBL5Y) are highly correlated, the values of correlation coefficient of these predictors with TFR is greater than 0.90.

Table 1 shows that the AGM is positively associated with AFB and negatively associated with BO3 it indicate that as AGM increases birth order decreases. Age at first birth positively associated with age at last birth means those women start childbearing earlier, also stop early. Age at last birth increases with OBI decreases but the proportion of women having birth in last five years increases. Reproductive span is positively associated with proportion of contraceptive users that means as reproductive span increases contraceptive users also increases. CU is inversely related to three plus birth order and directly related to OBI. PCL is positively associated with OBI that means as OBI increases proportion of childless women in last five years from survey date also increases. Three plus birth order is inversely related with OBI It indicates BO3 increases in opposite direction with OBI. Proportion of women having OBI five year or more’ is highly associated with PWBL5Y. The variable ‘PWBL5Y’ proposed by Singh et al.

\[ B = p_1 + 2p_2 + 3p_3 + \ldots \]

Where, $p_i$ is the proportion of $i^{th}$ order births in the given period. PBW is positively correlated with TFR.

2.1.9 Infant mortality rate (IMR)

IMR is positively associated with TFR, but the direction of the relation between IMR and TFR is not clear i.e. whether TFR regulate the IMR or IMR regulate the TFR. Singh et al., [27] suggest that IMR does not Granger Cause of TFR where as TFR Granger Cause of IMR that means TFR can be predicted by IMR but IMR cannot be predicted by TFR.

2.1.10 Weighted average of proportions of different birth orders (PBW)

PBW is another predictor proposed by Yadava et al., [10] to predict TFR which is defined as:

\[ PBW = p_1 + 2p_2 + 3p_3 + \ldots \]

Where, $p_i$ is the proportion of $i^{th}$ order births in the given period. PBW is positively correlated with TFR.

3. 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.

3.1 Application of the Methodologies

The regression analysis technique has been used to estimate the total fertility rate...
Table 1. Correlation matrix of various predictors (NFHS 4 Data)

<table>
<thead>
<tr>
<th></th>
<th>TFR</th>
<th>AGM</th>
<th>AFB</th>
<th>ALB</th>
<th>RS</th>
<th>CU</th>
<th>PCL</th>
<th>BO3</th>
<th>OBI</th>
<th>IMR</th>
<th>PBW</th>
<th>PWBL5Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFR</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGM</td>
<td>-0.619 (0.006)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFB</td>
<td>-0.277 (0.266)</td>
<td>0.877 (0.000)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALB</td>
<td>0.576 (0.012)</td>
<td>0.158 (0.531)</td>
<td>0.556 (0.017)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS</td>
<td>-0.906 (0.000)</td>
<td>0.429 (0.075)</td>
<td>0.031 (0.904)</td>
<td>-0.778 (0.000)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU</td>
<td>-0.764 (0.000)</td>
<td>0.292 (0.240)</td>
<td>0.079 (0.754)</td>
<td>-0.522 (0.000)</td>
<td>0.647 (0.004)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCL</td>
<td>-0.960 (0.000)</td>
<td>0.666 (0.003)</td>
<td>0.351 (0.154)</td>
<td>-0.500 (0.035)</td>
<td>0.913 (0.002)</td>
<td>0.685 (0.004)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BO3</td>
<td>0.785 (0.000)</td>
<td>-0.671 (0.002)</td>
<td>-0.368 (0.133)</td>
<td>0.379 (0.121)</td>
<td>-0.705 (0.001)</td>
<td>-0.426 (0.078)</td>
<td>-0.763 (0.000)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBI</td>
<td>0.957 (0.000)</td>
<td>-0.566 (0.014)</td>
<td>-0.204 (0.416)</td>
<td>0.661 (0.003)</td>
<td>-0.972 (0.000)</td>
<td>-0.687 (0.002)</td>
<td>-0.968 (0.000)</td>
<td>0.777 (0.000)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>0.761 (0.000)</td>
<td>-0.641 (0.004)</td>
<td>-0.348 (0.157)</td>
<td>0.462 (0.054)</td>
<td>-0.749 (0.000)</td>
<td>-0.382 (0.002)</td>
<td>-0.742 (0.000)</td>
<td>0.789 (0.000)</td>
<td>0.810 (0.000)</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBW</td>
<td>0.950 (0.000)</td>
<td>-0.556 (0.016)</td>
<td>-0.161 (0.524)</td>
<td>0.695 (0.001)</td>
<td>-0.954 (0.000)</td>
<td>-0.674 (0.002)</td>
<td>-0.907 (0.000)</td>
<td>0.801 (0.000)</td>
<td>0.962 (0.000)</td>
<td>0.843 (0.000)</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>PWBL5Y</td>
<td>0.960 (0.000)</td>
<td>-0.666 (0.003)</td>
<td>-0.351 (0.154)</td>
<td>0.500 (0.035)</td>
<td>-0.913 (0.000)</td>
<td>-0.685 (0.002)</td>
<td>-1.000 (0.000)</td>
<td>0.763 (0.000)</td>
<td>0.968 (0.000)</td>
<td>0.742 (0.000)</td>
<td>0.907 (0.000)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p-value is given in the parenthesis*
[15] is highly correlated with TFR and explained more variability as compared to other variables, in short it is sufficient itself to explain TFR. Therefore, in search of some other predictors other than this, which have capacity to explain as much as possible.

4. DEVELOPMENT OF MODEL

In this study an attempt has been done to develop some model for estimating TFR using some new predictor variable. Table 2 provides the model summary in which Model 1 include OBI(1) (proportion of women having OBI five years or more) as a predictor variable and it explains 91.5 percent variability in TFR. Now one more predictor (CU) that is negatively associated with TFR is added in Model 1. It is found that the combination of, OBI(1) and CU is also good to estimate TFR than OBI(1) alone. This combination is represented by Model 2; it explains 93.7 percent variability in TFR. Further all the remaining variables (AGM, PBW, RS, BO3, and IMR) added one by one in Model 2 and try to develop new models, but the variability explained by model is not significantly more. Therefore a new variable i.e. proportion of childless women in last five year (PCL) is proposed to estimate TFR. That is represented by Model 3; it explains 92 percent variability in TFR. Now one more predictor CU is added in this model to predict TFR. This Model 4 explains about 94 percent variability in TFR. This model shows that if there is no childless women (PCL=0) and no users of contraceptive (CU=0) than on an average a woman have approximately 9 children in her reproductive period.

4.1 Model Validation of the Fitted Model

4.1.1 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 [28] 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.

4.1.2 Shrinkage and Stability of \(r^2\)

The standard adjustment made in the coefficient of determination to compensate for the subjective effects of further sampling, the shrinkage of the model is given as Shrinkage\(=|\rho_v^2 - r^2|\)

Where \(\rho_v^2\) is Cross Validity Prediction Power (CVPP) and \(r^2\) is the coefficient of determination of the model. The Stability of \(r^2\) of the model is equal to (1-Shrinkage) means lower Shrinkage provides more stability of the model.

5. RESULTS AND DISCUSSION

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 4 in which we have included contraceptive use (CU) as predictor variable along with proportion of childless women in last five years (PCL), but the stability of model is obtain as (1- Shrinkage) and the estimate of stability is maximum for Model 3 in which we have taken proportion of childless women in last five years (PCL) from the survey date. Therefore this model is more stable for the population than other models considered in this study. The contraceptive use (CU) has a significant effect on reduction of TFR once but in the recent time our findings indicates contraceptive use has no significant impact on reduction of TFR. The observed and estimated values of TFR for all the major states of India using four proposed regression models are given in Table 3. A critical review of the results

<table>
<thead>
<tr>
<th>Model</th>
<th>Mathematical form</th>
<th>(r^2)</th>
<th>Adjusted (r^2)</th>
<th>(\rho_v^2)</th>
<th>Stability of (r^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TFR = 0.427 + 0.068*OBI</td>
<td>0.915</td>
<td>0.910</td>
<td>0.892</td>
<td>0.977</td>
</tr>
<tr>
<td>2</td>
<td>TFR = 0.370 + 0.058<em>OBI + 0.008</em>CU</td>
<td>0.937</td>
<td>0.928</td>
<td>0.910</td>
<td>0.973</td>
</tr>
<tr>
<td>3</td>
<td>TFR = 10.112 – 0.108*PCL</td>
<td>0.922</td>
<td>0.917</td>
<td>0.904</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>TFR = 9.387 – 0.092<em>PCL – 0.008</em>CU</td>
<td>0.943</td>
<td>0.935</td>
<td>0.919</td>
<td>0.976</td>
</tr>
</tbody>
</table>
Many variables are considered in this study and explain as much variability as possible of TFR combination of predictors and allowed to those explain TFR and proposed best objective of this using indirect method of estimation. The This study has been done to estimate the TFR contains only two predictors and explained more estimation of TFR and the reason is that it the best regression model among these f reveals that all models are good and Model 4 is 3, and for ten states through Model 4. Table 3 through Model 2, for twelve states through Model 20) percent for Andhra Pradesh through Model 2, for Kerala and West Bengal through Model 3 and 2.05 2.12 (+3.56) 2.04 (-0.71) 2.16 (+5.52) 2.11 (+2.74) Jharkhand 2.55 2.61 (+2.19) 2.63 (+3.31) 2.64 (+3.47) 2.70 (+5.81) Karnataka 1.80 1.77 (-1.70) 1.83 (+1.61) 1.72 (-4.42) 1.82 (+1.34) Kerala 1.56 1.49 (-4.45) 1.58 (+1.33) 1.40 (-10.49) 1.54 (-1.40) Madhya Pradesh 2.32 2.27 (-2.04) 2.26 (-2.49) 2.30 (-0.71) 2.33 (+0.22) Maharashtra 1.87 1.91 (+2.26) 1.85 (-1.20) 1.93 (+2.97) 1.90 (+1.38) Odisha 2.05 2.27 (+10.86) 2.21 (+8.01) 2.20 (+7.10) 2.19 (+6.59) Punjab 1.62 1.72 (+5.86) 1.59 (-1.81) 1.60 (+1.14) 1.53 (-5.49) Rajasthan 2.40 2.42 (+0.64) 2.32 (-3.47) 2.39 (-0.42) 2.33 (-2.86) Tamil Nadu 1.70 1.64 (-3.52) 1.71 (+0.45) 1.62 (-4.52) 1.73 (+1.78) Telangana 1.78 1.78 (0.00) 1.78 (0.00) 1.86 (+4.54) 1.89 (+6.01) Uttar Pradesh 2.74 2.76 (+0.81) 2.73 (-0.39) 2.49 (-9.23) 2.53 (-7.74) West Bengal 1.77 1.91 (+7.65) 1.79 (+1.25) 2.03 (+14.89) 1.94 (+9.50) India 2.18 2.33 (+6.74) 2.29 (+5.09) 2.26 (+3.69) 2.27 (+4.16) Percent difference from actual TFR is given in the parenthesis. Value with positive sign indicates over estimated and with negative sign indicates under estimated values presented in Table 3 shows that the estimated values of TFR for all states are closer to observed values. The absolute percent difference between observed TFR and estimated TFR for Model 1 is less than 10 percent for all states except Andhra Pradesh & Odisha and even below 5 percent for twelve states. The absolute percent difference between observed TFR and estimated TFR is lies between (10 to 20) percent for Andhra Pradesh through Model 2, for Kerala and West Bengal through Model 3 and for Andhra Pradesh through Model 4. These differences are below 5 percent for fifteen states through Model 2, for twelve states through Model 3, and for ten states through Model 4. Table 3 reveals that all models are good and Model 4 is the best regression model among these for estimation of TFR and the reason is that it contains only two predictors and explained more than 94 percent variability in TFR.

6. CONCLUSIONS

This study has been done to estimate the TFR using indirect method of estimation. The objective of this study is to identify variables those explain TFR and proposed best combination of predictors and allowed to explain as much variability as possible of TFR. Many variables are considered in this study and it was found that proportion of women having birth in last five years (PWBL5Y) is a best predictor among existing predictors proposed. Its data is easy to collect and error is also unlikely. Weighted average of proportion of different birth orders (PBW) is also a good predictor. These three predictors is by far the better predictor through which TFR can be improved. After using CU the accuracy of above model is improved, therefore in this study we have suggested some other models. Hence, if one can wants to predict TFR, can use any of these according to availability of data.

CONSENT

As per international standard or university standard, participant’s written consent has been collected and preserved by the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Brass W. Methods of analysis and estimation. In The Demography of Tropical


© 2020 Tiwari et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/63113