International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 331-340

Empirical Forecasting of Fertility Parameters of Bangladesh

Md. Rafiqul Islam
Dept. of Population Science and Human Resource Development
University of Rajshahi, Rajshahi, Bangladesh
E-mail: rafique_pops@yahoo.com

Md. Korban Ali
Dept. of Information & Communication Technology
Southeast University, Banani, Dhaka-1213, Bangladesh.

Md. Nurul Islam
Dept. of Statistics
University of Rajshahi, Rajshahi, Bangladesh.

Md. Atikur Rahman
Dept. of Population Science and Human Resource Development
University of Rajshahi, Rajshahi, Bangladesh

[Received May 4, 2005; Revised July 28, 2007; Accepted September 2, 2007]

Abstract

The aim of the present study is to forecast some fertility parameters of Bangladesh. A few mathematical time trend models have been fitted to time series data on crude birth rate (CBR), total fertility rate (TFR), gross reproduction rate (GRR) and net reproduction rate (NRR). CBR follows quadratic (i.e. parabolic) polynomial model while TFR, GRR and NRR follow simple linear regression model. Model validation technique such as cross-validity prediction power (CVPP), ρ_{cv}^2, is applied to these models to verify whether they are valid or not. It is seen that all the parameters of these fitted models are highly statistically significant. It is also found that all these fitted models are more than 88% stable and more than 90% of variance is explained.

Keywords and Phrases: Crude birth rate (CBR), Total fertility rate (TFR), Gross reproduction rate (GRR), Net reproduction rate (NRR), Polynomial model and Cross-validity prediction power (CVPP).

AMS Classification: 37M10.

1 Introduction

It is to be noted that a lot of works have been carried out for estimating the demographic parameters of Bangladesh. Some have used raw data to estimate the parameters and some have used indirect techniques to do these. These estimates suffer too much for their qualities of their raw data. We got different estimates for the parameters from different sources at various point of time. It is found that all the parameters are functions of time. Hence, we can have a time trend model which can provide us the estimates of the parameters with the changes of time. So an attempt has been made here to develop some time trend models for estimating the parameters CBR, TFR, GRR and NRR. The specific objectives of this study are:

i) to fit appropriate mathematical time trend models of CBR, TFR, GRR and NRR for Bangladesh,

ii) to apply cross-validity prediction power (CVPP), ρ^2_{cv}, to these models whether they are valid or not, and

iii) to forecast these fertility parameters of Bangladesh up to 2005.

2 Data and Methodology

2.1 Sources of Data

The secondary data of crude birth rate, total fertility rate, gross reproduction rate and net reproduction rate of Bangladesh have been taken from Statistical Year Book of Bangladesh (BBS, 1980, 1986, 1999 and 2001). These rates have been used to fit mathematical models.

Table 1: The CBR, TFR, GRR and NRR of Bangladesh During 1980-1998.

Years	CBR	TFR	GRR	NRR
1980	33.4	4.99	2.46	1.83
1981	34.6	5.04	2.49	1.84
1982	34.8	5.21	2.56	1.87
1983	35	5.07	2.48	1.82
1984	34.8	4.83	2.34	1.81
1985	34.6	4.71	2.2	1.79
1986	34.4	4.7	2.29	1.8
1987	33.3	4.42	2.14	1.69
1988	33.2	4.39	2.21	1.74
1989	33	4.35	2.1	1.72
1990	32.8	4.33	2.1	1.71
1991	31.6	4.24	2.06	1.7
1992	30.8	4.18	2.03	1.62
1993	28.8	3.84	2.01	1.57
1994	27.8	3.58	1.81	1.48
1995	26.5	3.45	1.68	1.46
1996	25.6	3.41	1.66	1.46
1997	21	3.1	1.52	1.37
1998	19.9	2.98	1.45	1.31

Source: (BBS, 1980, 1986, 1999 and 2001)

3 Methodology

3.1 Model Fitting

i) From the dotted plot of years and crude birth rate (CBR) (Figure 1), it is seen that CBR shows downward trend with time and a polynomial can be fitted. In this case, an nth degree polynomial model is treated and the model of the nth degree polynomial is

$$y = a_0 + \sum_{i=1}^{n} a_i \, x^i \, + u$$

(Montgomery, 1982), where, x is years; y is CBR; a_0 is the constant; a_i is the coefficient of x^i (i =1, 2, 3,..., n) and u is the error term of the model. Here we have to choose a suitable n for which the error sum of square is minimum. We can do this work using the software Econometric Views.

ii) The scattered plot of total fertility rate (TFR) and years is shown in Figure 2. It is observed that TFR is going downward with respect to time with some sort of random distortions. So a linear regression model of type

$$y = a_0 + a_1 x + u$$

can be fitted. Where, x represents the years; y represents TFR; a_0, a_1 are parameters and u is the disturbance term of the model.

iii) The dotted plot of years and gross reproduction rate (GRR) are shown in Figure 3. It is also observed that GRR follows simple linear relation. Therefore, a simple linear regression model is fitted and the form of the model is found to be:

$$y = a_0 + a_1 x + u$$

where, x represents the years; y represents GRR; a_0, a_1 are parameters and u is the disturbance term of the model.

iv) Again, from the scattered plot of years and net reproduction rate (NRR) (Figure 4), it is observed that NRR can also be fitted by simple linear regression model. Therefore, the form of the model is found to be:

$$y = a_0 + a_1 x + u$$

where, x represents the years; y represents NRR; a_0, a_1 are parameters and u is the disturbance term of the model.

v) We are using time series data for the analysis. So, it would be frequent to deal with ARIMA (Autoregressive Integrated Moving Average) models of the type:

$$X_t = \alpha + \lambda t + \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j=1}^{q} \delta_j u_{t-j} + u_t$$

where t is the time trend component, $X_t = \Delta^d Y_t = \Delta^{d-1}(Y_t - Y_{t-1})$ and for d = 1 X_t is the first difference of the variable Y_t. This is an ARIMA (p, 1, q/C,t) model which includes intercept, trend, p autoregressive terms, q moving average terms, and differencing parameter d=1.

The value of d is determined by the stationarity criterion of the variable. Unit root test is applied in this case (MacKinnon, 1996). The order combination of p and q can be done using the significant effect of estimated parameters under a stationary estimation process (inverted roots) subject to the AIC (Akaike Information Criterion) is minimum (Gujarati, 1995; Pankratz, 1991; Cleary and Hay, 1980).

4 Model Validation

To check, how much these models are stable over the population, the restricted cross validity prediction power (RCVPP), ρ^2_{rcv}, is applied (Khan and Ali, 2003). Here

$$\rho^2_{rcv} = 1 - w(1 - R^2),\ R^2 \geq 1 - w^{-1},\ w = \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)},\ n > k+2$$

where, n is the number of cases, k is the number of explanatory variables in the model and the cross-validated R is the correlation between observed and predicted values of the dependent variable. Using the above statistics, it can be concluded that if the prediction equation is applied to many other samples from the same population, then $(\rho^2_{rcv} \times 100\%$ of the variance on the predicted variable would be explained by the model (Stevens, 1996).

5 Results and Discussion

Fitting polynomial of CBR on time a second degree polynomial of the form

$$\hat{y}_t = 35.89868 - 0.38937t^2$$

is obtained. The coefficient of determination $R^2 = 0.932$ and $\rho^2_{rcv} = 0.919$.

Again, fitting a simple regression of TFR on time, an equation

$$\hat{y}_t = 5.440351 - 0.118667t$$

is obtained. The R^2 of the model is 0.946 and ρ^2_{rcv} is 0.919.

Fitting a regression of GRR on time, it is found that the regression line is

$$\hat{y}_t = 2.652281 - 0.056860t$$

giving R^2=0.932 and $\rho^2_{rcv} = 0.919$.

Lastly, a regression line of NRR on time gives the relation

$$\hat{y}_t = 1.952982 - 0.029035t$$

with R^2=0.906 and ρ^2_{rcv}=0.882.

The fitted models with their computed RCVPP (ρ^2_{rcv}) and shrinkages corresponding to their R^2 have been summarized in Table 2. From this table, it is seen that all the fitted models posses the higher explanatory powers exceeding 90% and their shrinkages are 0.01258, .0099, 0.01256, and 0.01831, respectively. These imply that all these models are more than 90% stable (as explained by Stevens, 1996).
The information on time trend fitted models have been presented in Table 3 and Table 4. From Table 3, it is observed that all the variables are non-stationary and are stationary after first difference. Fitted ARIMA models for the variables have been summarized in Table 4. Models selected in Table 4 are subject to the significance of parameter estimates and to the minimum AIC levels. Comparing the shrinkage in Table 2 and Table 4 it can be authenticated that the fitted models in Table 2 are more stable than the fitted models in Table 4.
It should be mentioned here that the usual models, i.e. Gompertz model or Makeham model, exponential model, log-linear model and logistic model were also applied but seem to be worse fitted with respect to their shrinkages. Therefore, the results of these models were not shown here. Thus, we would like to accept the fitted models shown in Table 2 and would like to forecast using these fitted models.

Table 2: Estimated Cross-Validity Prediction Power, ρ^2_{cv} of the Predicted Equations of CBR, TFR, GRR and NRR of Bangladesh.

Models	n	k	R^2	ρ^2_{rcv}	Shrinkage
$\hat{y}_t = 35.89868 - 0.38937t^2$ (For CBR) prob.=(0.0000) (0.0000)	19	1	0.931707	0.91900	0.01258
$\hat{y}_t = 5.440351 - 0.118667t$ (For TFR) prob.=(0.0000) (0.0000)	19	1	0.946106	0.946106	0.0099
$\hat{y}_t = 2.652281 - 0.056860t$ (For GRR) prob.=(0.0000) (0.0000)	19	1	0.931827	0.919269	0.01256
$\hat{y}_t = 1.952982 - 0.029035t$ (For NRR) prob.=(0.0000) (0.0000)	19	1	0.90598	0.882289	0.01831

Here, t is the time trend component, and prob. indicates the significant level of estimated parameters.

Table 3: Unit Root Test.

Variables	Specification	DF-Value	MacKinnon Critical Value at 5% level	Stationary at
CBR	None	-2.288904	-1.9627	First Difference
TFR	None	-2.219726	-1.9627	First Difference
NRR	None	-3.174887	-1.9627	First Difference
GRR	None	-3.338522	-1.9627	First Difference

To study the residual analysis of these models the residuals have been plotted in the graph, which are shown in Figures 1 to 4. The forecasted values of these models have been presented in Table 5 and in Figures 5 to 8. From the table and figures, it is seen that all these fertility parameters are showing decreasing trend during 1999-2005.

It should be mentioned here that all these time trend models have been estimated using the software Econometric Views.

Table 4: Fitted ARIMA models.

Variables and models	Fitted Models	AIC	R^2	n (k)	RCVV ρ^2_{rcv}	Shrinkage
CBR ARIMA(0,1,1/C,t)	$\hat{X}_t = 0.888375 - 0.155136t - 0.946993\hat{u}_{t-1}$ Prob.=(0.00000) (0.0000) (0.0000) Inverted MA=0.95	-0.582655	0.709179	18 (2)	0.6324	0.0767
TFR ARIMA(1,1,1/C)	$\hat{X}_t = -0.143346 + 0.584681X_{t-1} - 0.989705\hat{u}_{t-1}$ Prob.=(0.00000) (0.0158) (0.0000) Inverted AR=0.58, Inverted MA=0.99	-4.186036	0.342328	17 (2)	0.0817	0.2606
GRR ARIMA(1,1,1/C)	$\hat{X}_t = -0.071317 + 0.586181X_{t-1} - 0.9899\hat{u}_{t-1}$ Prob.=(0.00000) (0.0376) (0.0000) Inverted AR=0.59 , Inverted MA=0.99	-4.898500	0.252542	17 (2)	$R^2 < 1 - w^{-1} = 0.2838$ and RCVV is equal to zero (Khan and Ali, 2003)	
NRR ARIMA(0,1,1/t)	$\hat{X}_t = -0.002907t - 0.989948\hat{u}_{t-1}$ Prob.=(0.00000) (0.0000) (0.0000) Inverted MA=0.99	-6.809671	0.523676	18 (2)	0.3488	0.1748

Here X_t is the first difference of variable (s), AIC is the Akaike Information Criterion, n is the sample size after adjusting the endpoint, and k is the number of predictors used in the model.

6 Conclusion

In the present study, crude birth rate of Bangladesh follows $2nd$ degree (i. e. quadratic) polynomial model in which the term containing t is absent. On the other hand, total fertility rate, gross reproduction rate and net reproduction rate of Bangladesh follow simple linear regression model. These models have the highest explanatory powers exceeding 90% with more than 90% stability. These time trend models have been provided forecasted values during 1999-2005.

References

1. BBS. (1980). Statistical Year Book of Bangladesh 1980.
2. BBS. (1986). Statistical Year Book of Bangladesh 1986.
3. BBS. (1999). Statistical Year Book of Bangladesh 1998.
4. BBS. (2001). Statistical Year Book of Bangladesh 1999
5. Gujarati, D. N. (1995) Basic Econometrics, 3rd ed., Mc Graw-Hill International.
6. Khan, M. A. R. and Ali, M. A. (2003) Restricted Cross Validity Predictive Power, Pakistan Journal of Statistics, Vol. 19, No. 2.
7. MacKinnon, J. G. (1996) Numerical Distribution Function for Unit Root and Cointegration Tests, Journal of Applied Econometrics, Vol.11, pp.601-618.
8. Mc Cleary, R. and Hay, R. A. (1980) Applied Time Series Analysis for the Social Science, Sega Publications, California.
9. Montgomery, Douglas C. and Peck, Elizabeth A. (1982). Introduction to Linear Regression Analysis, John Wiley and Sons, New York.
10. Pankratz, A. (1991) Forecasting with Dynamic Regression Models, John Wiley & Sons Inc., New York.
11. Stevens, J. (1996). Applied Multivariate Statistics for the Social Sciences, Third Edition, Lawrence Erlbaum Associates, Inc., Publishers, New Jersey.