مجلّه‌ی پژوهش‌های آماری ایران

J. Statist. Res. Iran **4** (2007): 227-238

دوره‌ی ۴، شماره‌ی ۲، پاییز و زمستان ۱۳۸۶، صص ۲۲۷-۲۳۸

پژوهشکده‌ی آمار

Predicting Population for Male of Rural Area in Bangladesh

Rafiqul Islam

University of Rajshahi

Abstract. In this paper the population for male of rural area in Bangladesh is predicted by using the geometric growth rate method. The predictions are computed in a three-steps procedure. In the first step, the predictions are computed using an exponential model estimated by Quasi-Newton method for the years 1974, 1981, 1991 and 2001 using the package STATISTICA. Using the cross-validation predictive power (CVPP) criterion and coefficient of determination, the shrinkage coefficient (λ) is constructed. The shrinkage coefficient determines the adequacy of the first step prediction. In the second step, these predicted values are used to estimate the growth rate for different age groups by using the geometric growth rate method. In the final step, considering the population for male of rural area in Bangladesh for the census year 2001 as the base population and using the estimated geometric growth rate of the second step estimation, the predictions for the male population of rural area of Bangladesh are estimated for the years 2002 through 2031 by applying geometric growth rate method.

Keywords. male rural population; exponential model; geometric growth rate method; cross-validity predictive power (CVPP); coefficient of determination; shrinkage; F-test.

1 Introduction

It is noted that Bangladesh is a rural predominant country. Moreover, Bangladesh is a poor country and mostly dependent on foreign aid and relief. Most of the people of Bangladesh reside in rural area of which more than fifty are male. Rural people of Bangladesh always struggle for survival. It should be mentioned here that rural families are much more dependent on male earnings than that of urban. Generally the rural people of Bangladesh are hard hitted financially in comparison to the urban people because of shortage of income generated by the activities in rural region. As a consequent, there is a trend of people, particularly for male population, moving from rural area to urban areas of Bangladesh for better welfare. Since the independence in 1971, the population of Bangladesh is rising rapidly with accelerated growth rate in accordance with its areas. It is noted that Bangladesh is an agrarian based country and in fact, it is happened largely in rural areas. Male people in the rural areas generally migrate to urban areas for better job and other good facilities of livelihood. As a consequence, the heavy influx of people in the urban areas is a burden to the government of Bangladesh for various reasons, e.g. lack of shelter, education, health, clean water, etc. For this reason, government of Bangladesh should gather proper information of the population in rural areas in near future, middle future, far future population so that it can take proper steps in income generating activities for the rural population especially male population. Consequently, the government of Bangladesh needs firm and rigid policy to accommodate its fast growing rural male population so that government can provide the necessary requirements to its citizens and hence the pressure of urbanization would be abated. That is why the stream of migration from rural to urban area would be slow down.

Since the pattern of age structure of the population differs for various demographic variables, from country to country, and from region to region. Islam et al. (2003) found that the age structure for male population of Bangladesh follows a modified negative exponential model. And the age structure for the population of both sexes and female of Bangladesh follows a modified negative exponential model (Islam, 2003 and 2005). It is noted that various types of model might be used to predict the population. Among them are linear model, geometric growth model, exponential growth model, modified exponential model, logistic model, Makeham model, Gompertz model, polynomial model, etc.

The objective, in this paper, is to provide proficient prediction of the population for male of rural areas in Bangladesh for the years 2002 to 2031 using the most recent data and the geometric growth rate method for different age groups by a three step procedures. These predictions can then be used for improvement of government's rural infrastructural policy decisions and job creating function.

This paper is organized as follows. The data sources of this study are presented in section 2. Section 3 describes the models, methodological issues and various tests. Empirical results and discussion of the results are reported in section 4. Finally, section 5 includes the conclusion of the paper.

2 Data Sources of this Study

To fulfill the objective of this paper the secondary quinquennial age data on the population for male of rural area of Bangladesh have been taken from various issues (Bangladesh Bureau of Statistics, 1977, 1984, 1994, and 2003) of Bangladesh population census. It is to be noted that the population census in Bangladesh occurs in every ten years distant and the latest census took place in 2001. Using the latest two census data of 1991 and 2001 the predictions for the years 2002 through 2031 are computed. The age group considered in this study are (0-4, 5-9, 10-14, 15-19, ..., 65-69, 70 and above). The population values which are counted in thousands are shown in Table 1.

3 Models, Methodological Issues, and Various Tests

3.1 Data Smoothing Technique

The data of the current study have been smoothed out to eliminate any anomalies from the age related population data. The method used to smooth the data is the "4253H twice" (Velleman, 1980). This is the default in the Minitab package – version 12.1. The smoothed data are then used to fit an exponential model for the years 1974, 1981, 1991, and 2001 separately. The smoothed data are given in Table 1.

3.2 The Models

Using the scattered plot of ages and smoothed age structure for male population in rural area of Bangladesh (Figure 1), it is observed that population is distributed with a negative exponential distribution with respect to ages. Therefore, a two-parameters negative exponential model is considered and the mathematical structure of the model is

$$y = e^{(-ax+b+u)},$$

where, x represents mean value of the age group; y represents male population; a, b are unknown parameters and u is the disturbance term of the model. Note

Table 1. Observed, smoothed, and predicted male population by age group in rural area of Bangladesh during 1974-2001 censuses.

Age group	1974 Observed population (000)	1974 Smoothed population (000)	1974 Predicted values (000)	1981 Observed population (000)	1981 Smoothed population (000)	1981 Predicted values (000)
0-4	5562	5741	5975	6494	6675	7001
5-9	6112	5486	4989	6248	6248	5788
10-14	4542	4542	4166	5279	5229	4785
15-19	2808	3234	3479	3385	3826	3956
20-24	2063	2380	2905	2503	2866	3270
25-29	2029	2042	2426	2532	2399	2703
30-34	1787	1879	2025	1964	2058	2235
35-39	1810	1755	1691	1908	1811	1848
40-44	1556	1556	1412	1560	1566	1527
45-49	1246	1317	1179	1326	1326	1263
50-54	1162	1110	985	1184	1130	1044
55-59	716	929	822	794	960	863
60-64	851	851	686	893	893	713
65+	1288	851	573	1479	893	590
Total	33532	33673	33313	37549	37880	37586

Age group	1991 Observed population (000)	1991 Smoothed population (000)	1991 Predicted values (000)	2001 Observed population (000)	2001 Smoothed population (000)	2001 Predicted values (000)
0-4	7399	7399	7507	6774	6814	7305
5-9	7541	6695	6274	7161	6650	6332
10-14	5542	5424	5244	6538	5987	5489
15-19	3508	4064	4383	4608	4874	4759
20-24	2978	3230	3663	3265	3868	4125
25-29	3220	2834	3062	3386	3337	3576
30-34	2508	2559	2559	3062	3113	3100
35-39	2463	2251	2139	3077	2879	2687
40-44	1857	1886	1788	2517	2497	2330
45-49	1520	1530	1494	1955	2030	2019
50-54	1307	1232	1249	1661	1577	1751
55-59	894	1032	1044	1025	1254	1518
60-64	1014	966	872	1228	1148	1316
65-69	530	966	729	659	1148	1140
70+	1146	966	609	1546	1148	989
Total	43427	43034	42616	48462	48324	48436

that these models have been estimated using the software STATISTICA. The predicted values of these models are also given in Table 1.

The most important properties of the negative exponential model are given below:

- The domain of the negative exponential model is real numbers;
- The range of this model is positive values, that is, it is always greater than zero;
- The graph of the negative exponential model is decreasing;
- The graph of this model is asymptotic to the x-axis when x approaches positive infinity;
- The graph is continuous and smooth.

It should be mentioned here that other usual models such as linear, loglinear, semi-loglinear, logistic, quadratic, cubic, Gompertz, and Makeham models are also tried to be applied to fit this data aggregate. But, these models performed very poorly due to their shrinkage coefficients, coefficients of determination and the lack of significance of the model parameters for this data

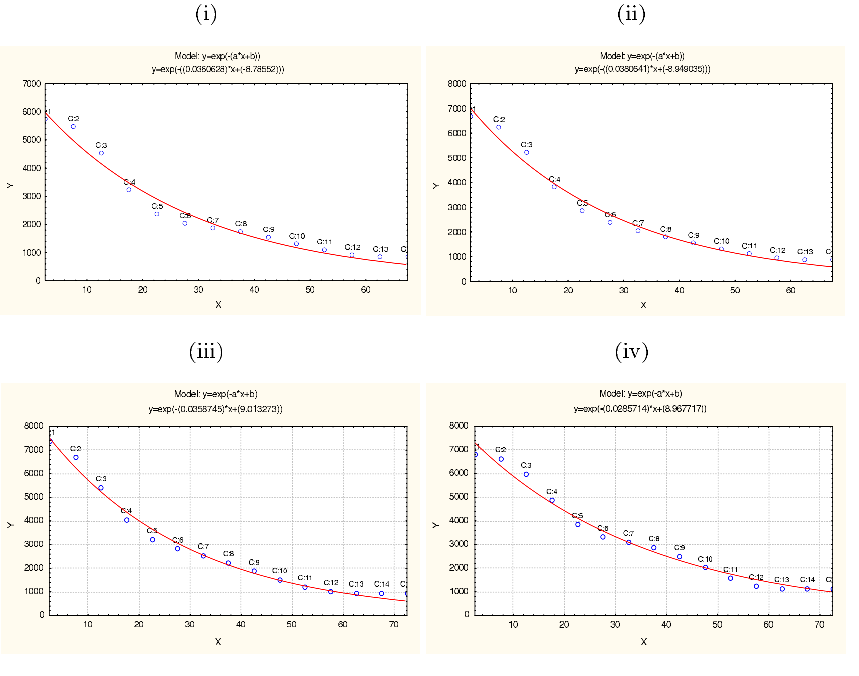

Figure 1. Observed and fitted male population in rural area of Bangladesh in (i) 1974, (ii) 1981, (iii) 1991, and (iv) 2001.

in this study. As a consequent, the findings of these models are not shown here.

3.3 Geometric Growth Rate Method

Geometric growth rate is estimated by using the following equation

$$\hat{P}_{t_2}^{m-m+5} = \hat{P}_{t_1}^{m-m+5}\{1 + r^{m-m+5}\}^{t_2-t_1}, \qquad (1)$$

where, $\hat{P}_{t_1}^{m-m+5}$ is the predicted initial population at time t_1 for the age interval m to $m+5$; $\hat{P}_{t_2}^{m-m+5}$ is the predicted terminal population at time t_2 for the age group m to $m+5$; r^{m-m+5} is the intercensal annual growth rate of the age interval m to $m+5$; $(t_2 - t_1)$ is the time distance between intercensal periods, and m is the lower limit of each age group. The intervals are counted as $[m, m+5)$.

The r^{m-m+5} is computed for different age groups from (1) as in the following.

$$r^{m-m+5} = \text{Antilog}\left[\left(\frac{1}{t_2 - t_1}\right)\log_e\left(\frac{\hat{P}_{t_2}^{m-m+5}}{\hat{P}_{t_1}^{m-m+5}}\right)\right] - 1. \qquad (2)$$

Years 1991 and 2001 are considered as the initial and the terminal populations respectively in estimating the age specific growth rates by equation (2).

For prediction purpose, the year 2001 census is treated as the base population. The intercensal annual growth rate, at different ages, during 1991-2001 is used in this study assuming fertility and mortality remain constant during the forecast period. Estimates of the intercensal annul geometric growth rate for different age groups are computed based on the first-phase predicted population for the census years 1991 and 2001.

3.4 Model Validation

Model assessment criteria used are the usual regression t-test, and R^2. For model validation, the cross-validation predictive power (CVPP) denoted by ρ_{cv}^2, is computed by $\rho_{cv}^2 = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1 - R^2)$, where n is the number of classes, k is the number of regressors in the model, and R^2 is the coefficient of determination in the first stage of estimation. The shrinkage coefficient of the model is equal to the absolute value of $\lambda = (\rho_{cv}^2 - R^2)$, (Stevens, 1996). Closer the value of λ tends to zero, then the prediction is better. Furthermore, the stability of R^2 of the model is equal to (1−shrinkage coefficient). The estimated CVPP's corresponding to their R^2 and information on model fittings are launched in Table 2.

Table 2. Information on model fittings and CVPP.

Model	n	K	R^2	ρ^2_{cv}	Shrinkage	Cal. F	Tab. F (at 1% level)
1	14	1	0.96992	0.961912	0.008008	303.0599	9.33 with (1, 12) d.f.
2	14	1	0.98160	0.976701	0.004899	640.1739	9.33 with (1, 12) d.f.
3	15	1	0.98772	0.984718	0.003002	1045.63	9.07 with (1, 13) d.f.
4	15	1	0.98344	0.9794	0.00404	772.0242	9.07 with (1, 13) d.f.

3.5 F-test

To investigate the overall measure of model significance as well as the significance of R^2 , the F-test is applied to the fitted model. The mathematical formula for F-test is given by

$$F = \frac{\frac{R^2}{p-1}}{\frac{1-R^2}{n-p}} \quad \text{with } (p-1, n-p) \text{ degrees of freedom (d.f.)},$$

where p is the number of parameters to be estimated in the model, n is the number of cases and R^2 is the coefficient of determination in the model (Gujarati, 1998).

4 Empirical Results and Discussion of the Results

Both the smoothed and the original data series are used to fit negative exponential model for the years 1974, 1981, 1991, and 2001 for all age groups. It is found that the smoothed data has better predictions. Consequently, the smoothed series have been used in the first phase prediction. The estimated models for the male population of rural area are given in the following forms:

(a) $y = \exp\{-0.03606x + 8.78552\}$ in 1974

$t(12)$-stats	15.08302	240.832
p-value	(0.00000)	(0.000)

with coefficient of determination $R^2 = 0.96992$ and $\rho^2_{\text{cv}} = 0.961912$.

(b) $y = \exp\{-0.03806x + 8.949\}$ in 1981

$t(12)$-stats	19.15078	308.453
p-value	(0.00000)	(0.000)

providing $R^2 = 0.98160$ and $\rho^2_{\text{cv}} = 0.976701$.

(c) $y = \exp\{-0.03587x + 9.0133\}$ in 1991

$t(13)$-stats	25.09937	342.8943
p-value	(0.00000)	(0.000)

giving proportion of variance explained $R^2 = 0.98772$ and ρ^2_{cv} is 0.984718.

(d) $y = \exp\{-0.02857x + 8.9677\}$ in 2001

$t(13)$-stats	22.78804	371.8038
p-value	(0.00000)	(0.000)

providing coefficient of determination $R^2 = 0.98344$ and $\rho^2_{\text{cv}} = 0.9794$.

The estimated CVPP, ρ^2_{cv}'s, corresponding to their R^2 are shown in Table 2. From this table it appears that all the fitted models (a)-(d) are highly cross-validated and their shrinkage coefficients are 0.008008, 0.004899, 0.003002 and 0.00404, respectively. These imply that the fitted models (a)-(d) stability is more than 96%, 97%, 98% , and 97%, respectively. Moreover, it is found that the parameters of the fitted models (a)-(d) are highly statistically significant with significance of variance explained. The stability of R^2 for these models is more than 99%.

The calculated values of F-test for the models (a)-(d) are 303.0599 with $(1, 12)$ d.f., 640.1739 with $(1, 12)$ d.f., 1045.63 with $(1, 13)$ d.f., and 772.0242 with $(1, 13)$ d.f., respectively, whereas the corresponding tabulated values of the models (a)-(b) are only 9.33 but for the models (c)-(d) is only 9.07 at 1% level of significance, respectively. Therefore, from these statistics it is seen that these models and their corresponding R^2's are highly statistically significant.

The predicted values for the census years 1991 and 2001 are then used to estimate the geometric growth rates for each age group, which, in turns, are used to predict the population for male of rural area for the years 2002 to 2031 considering the observed population of 2001 census as base year population. The estimated geometric growth rate for each age group is provided in Table 3.

Table 3. Estimated geometric growth rate for male population of rural area in Bangladesh at each age group.

Age group	The estimated geometric growth rate
0-4	−0.002723971
5-9	0.000920627
10-14	0.004576593
15-19	0.008264377
20-24	0.011949181
25-29	0.015638680
30-34	0.019363645
35-39	0.023070797
40-44	0.026830691
45-49	0.030572554
50-54	0.034361556
55-59	0.038142874
60-64	0.042014927
65-69	0.045725582
70+	0.049682362

Using these estimated age specific growth rates and 2001 census observed population as the base, the population for male of rural area is predicted for the years 2002 through 2031 which is given in the Table 4.

Table 4. Predicted rural male population (in thousands) of Bangladesh during 2002-2021 (in thousands).

Age group	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
0-4	6756	6737	6719	6700	6682	6664	6646	6628	6610	6592
5-9	7168	7174	7181	7187	7194	7201	7207	7214	7221	7227
10-14	6568	6598	6628	6659	6689	6720	6750	6781	6812	6843
15-19	4646	4684	4723	4762	4802	4841	4881	4922	4962	5003
20-24	3304	3343	3383	3424	3465	3506	3548	3590	3633	3677
25-29	3439	3493	3547	3603	3659	3716	3775	3834	3893	3954
30-34	3121	3182	3243	3306	3370	3435	3502	3570	3639	3709
35-39	3148	3221	3295	3371	3449	3528	3610	3693	3778	3865
40-44	2585	2654	2725	2798	2873	2950	3030	3111	3194	3280
45-49	2015	2076	2140	2205	2273	2342	2414	2488	2564	2642
50-54	1718	1777	1838	1901	1967	2034	2104	2176	2251	2329
55-59	1064	1105	1147	1191	1236	1283	1332	1383	1436	1490
60-64	1280	1333	1389	1448	1509	1572	1638	1707	1779	1853
65-69	689	721	754	788	824	862	901	942	985	1031
70+	1623	1703	1788	1877	1970	2068	2171	2279	2392	2511
Total	49122	49802	50501	51220	51961	52723	53508	54317	55149	56007

Table 4 (continued). Predicted male rural population (in thousands) of Bangladesh during 2002-2021 (in thousands).

Age group	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
0-4	6574	6556	6538	6520	6502	6485	6467	6449	6432	6414
5-9	7234	7241	7247	7254	7261	7267	7274	7281	7287	7294
10-14	6875	6906	6938	6970	7001	7034	7066	7098	7131	7163
15-19	5045	5086	5128	5171	5213	5257	5300	5344	5388	5433
20-24	3721	3765	3810	3856	3902	3948	3996	4043	4092	4141
25-29	4016	4079	4143	4208	4273	4340	4408	4477	4547	4618
30-34	3781	3854	3929	4005	4083	4162	4242	4324	4408	4494
35-39	3954	4046	4139	4235	4332	4432	4534	4639	4746	4856
40-44	3368	3458	3551	3646	3744	3845	3948	4054	4163	4274
45-49	2723	2806	2892	2980	3071	3165	3262	3362	3464	3570
50-54	2409	2491	2577	2666	2757	2852	2950	3051	3156	3265
55-59	1547	1606	1668	1731	1797	1866	1937	2011	2087	2167
60-64	1931	2012	2097	2185	2277	2372	2472	2576	2684	2797
65-69	1078	1127	1178	1232	1289	1348	1409	1474	1541	1612
70+	2635	2766	2904	3048	3199	3358	3525	3700	3884	4077
Total	56890	57801	58739	59706	60703	61731	62790	63883	65011	66174

Age group	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
0-4	6397	6362	6362	6345	6327	6310	6293	6276	6259	6242
5-9	7301	7307	7314	7321	7328	7334	7341	7348	7355	7361
10-14	7196	7229	7262	7295	7329	7362	7396	7430	7464	7498
15-19	5477	5523	5568	5614	5661	5708	5755	5802	5850	5899
20-24	4190	4240	4291	4342	4394	4446	4500	4553	4608	4663
25-29	4690	4764	4838	4914	4991	5069	5148	5229	5310	5393
30-34	4581	4669	4760	4852	4946	5042	5139	5239	5340	5444
35-39	4968	5082	5199	5319	5442	5568	5696	5828	5962	6100
40-44	4389	4507	4628	4752	4879	5010	5145	5283	5424	5570
45-49	3680	3792	3908	4027	4151	4277	4408	4543	4682	4825
50-54	3377	3493	3613	3737	3865	3998	4135	4278	4425	4577
55-59	2250	2336	2425	2517	2613	2713	2816	2924	3035	3151
60-64	2914	3037	3164	3297	3436	3580	3731	3887	4051	4221
65-69	1685	1762	1843	1927	2015	2107	2204	2305	2410	2520
70+	4280	4492	4716	4950	5196	5454	5725	6009	6308	6621
Total	67374	68595	69891	71210	72572	73979	75432	76932	78482	80084

The rate at which the population for rural male is growing is a concern for the government of Bangladesh. The government of Bangladesh requires policies to provide the maximum welfare to its citizens of the rural areas. In essence, the government needs strategic plan to provide good health care, satisfactory accommodation, roads and highways, available educational institutions, and secured jobs to satisfy the minimum needs for the people of the rural area of Bangladesh so that the tendency of people from rural area to urban area would be abated. And henceforth, the pressure of urbanization would be reduced, that is expectation to our nation.

5 Conclusion

This study observed that the age pattern of the rural male population of Bangladesh follows a two-parameters negative exponential model for the census years 1974, 1981, 1991, and 2001. The smoothed data are then used to estimate the geometric growth rate following Malthusian law of population growth. This projected population is an early warning to the government of Bangladesh to take the matter serious in accommodating its rural citizens with maximum welfare in the future. Although the negative exponential models are used for the first phase predictions, there are a number of other models that can be applied to fit the data, e.g., logistic, quadratic, cubic, Gompertz, and Makeham models. However, these models performed poorly in accordance with their shrinkage coefficients in this study. Therefore, these were omitted from the analysis. The research can be further explored to investigate the district wise comparison of population growth of Bangladesh. In fact, the most densely populated districts e.g. Dhaka, Chittagong, Rajshahi, Sylhet, Barisal and Khulna are kept in our next research agenda for analysis.

References

Bangladesh Bureau of Statistics (1977). Population Census 1974, National Volume, Government of the People's Republic of Bangladesh, Dhaka.

Bangladesh Bureau of Statistics (1984). Population Census 1981, National Series, Government of the People's Republic of Bangladesh, Dhaka.

Bangladesh Bureau of Statistics (1994). Population Census 1991, National Series, Vol. 1, Government of the People's Republic of Bangladesh, Dhaka.

Bangladesh Bureau of Statistics (2003). Population Census 2001, National Report, Government of the People's Republic of Bangladesh, Dhaka.

Gujarati, D.N. (1998). *Basic Econometrics*, 3rd ed. McGraw Hill, New York.

Islam, R. (2003). Modeling of Demographic Parameters of Bangladesh An Empirical Forecasting, Unpublished Ph.D. Thesis, Rajshahi University.

Islam, R., Islam, N., Ali, A., and Mostofa, G. (2003). Construction of male life table from female widowed information of Bangladesh, *Int. J. Statist. Sci.* **2**, 69-82.

Islam, R. (2005). Construction of female life table from male widowed information of Bangladesh, *Pakistan J. Statist.* **21**, 275-284.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences*, 3rd ed. Lawrence Erlbaum, New Jersey.

Velleman, P.F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms, *J. Amer. Statist. Assoc.* **75**, 609-615.

Rafiqul Islam
Department of Population Science and
Human Resource Development,
University of Rajshahi,
Rajshahi - 6205,
Bangladesh.
e-mail: *rafique_pops@yahoo.com*