Application of the Generalized Inverse Projection Technique to Madras's Population Data, 1866-1951: An Evaluation of Consistency among Demographic Data and Estimates[1]

Ryuichi Kaneko[2]

Osamu Saito[3]

Mihoko Takahama[4]

Extended Abstract

Introduction

This paper is an attempt to apply an inverse projection technique known as Generalized Inverse Projection (GIP) to data for a pre-independence Indian province. We have already tried to estimate annual series of TFR and e(0) for two pre-independence provinces, Madras and Punjab, by adopting the Brass relational Gompertz fertility model, the Brass logit life-table system and growth balance method for the period 1891-1951, whose tentative results seem reasonably robust (Saito et al. 2003). In this paper, we want to make a further methodological step: the application of GIP to the population data in one of the two provinces, Madras. In so doing, not just would we like to test the applicability of a new methodology, but also to demonstrate what kind of additional information could be obtained about historical populations by using such a sophisticated technique. In particular, we use the results to evaluate the data quality and the estimates previously obtained from the conventional methods. As a result, some inconsistencies in data and the previous estimates mainly due to incompleteness of the registration are detected. Some remedies are also briefly discussed.

Data Sources and Correction

Census populations and numbers of births and deaths in the area of Punjab and Madras over the projection period are used as targets to seek values of the unknown parameters in GIP. As is known, the Census of India is launched during the age of under the rule of Britain in 1871; it remains the modern manner and the regularity of every 10 years. Three types of correction for the census population are conducted; (1) correction on population age distributions by 5-year age group in the census 1891, 1901, 1911, 1921, 1931, (2) interpolation of population by 5-year age group into single year for the five censuses with the spline interpolation of the cumulative age distribution, and (3) correction on

[1] This research was supported by the Grant-in-Aid for Scientific Research 2001-2003 from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

[2] National Institute of Population and Social Security Research, Japan. e-male: r-kaneko@ipss.go.jp

[3] Institute of Economic Research, Hitotsubashi University, Tokyo. e-male: cr00100@srv.cc.hit-u.ac.jp

[4] Institute of Economic Research, Hitotsubashi University, Tokyo. e-male: mihoko@ier.hit-u.ac.jp

population age distributions by single year of age in the census 1951 through modified Mayer's blended method. While, the vital statistics such as births and deaths are taken from the Report on the Sanitary Administration, published on an annual basis since 1866 through 1950 with some missing years.

Simultaneous Estimation of the Demographics via Generalized Inverse Projection Approach: Method

We conducted simultaneous estimations of time series of the demographics for the colonial population in Madras state via the generalized inverse projection (GIP) approach. The GIP is a generic name of methods with which past age-structured population and vital rates are estimated from an appropriate set of demographic data by optimizing an objective function that measures goodness of fit of the projected results with the data as a whole. We employ this technique to reconstruct population histories of the province of Madras 7 districts in the colonial era and to check the consistency of our estimates for the population development.

The objective penalty function is constructed to represent total discrepancies between target data and the estimates of relevant demographic variables. The function, denoted by *F*, is to be minimized to obtain the best estimates that are as consistent with the inputs data as possible. *F* is defined in the present study as:

$$F = Av\left[\sum_{t\in T_{c1}}[(\hat{P}_t - P_t)/P_t]^2\right] + Av\left[\sum_{t\in T_{c2}}\sum_{a=0}^{w}[(\hat{P}_{a,t} - P_{a,t})/P_{a,t}]^2\right] + \kappa Av\left[\sum_{t\in T_d}[(\hat{D}_t - \tilde{D}_t)/\tilde{D}_t]^2\right] + \lambda Av\left[\sum_{t,t+1\in T_m}(\delta_{t+1} - \delta_t)^2\right] \quad (1)$$

where

P_t, $\hat{P}_t$: The observed and projected total population in year *t*,

$P_{a,t}$, $\hat{P}_{a,t}$: The observed and projected population in complete age *a* and year *t*,

$\tilde{D}_t$, $\hat{D}_t$: The objective and projected number of deaths in year *t*,

δ_t : The index of net-migration,

κ : Relative weight on consistency in number of deaths,

λ : Relative weight on the smoothness of annual variation of the index of net-migration.

In addition, T_{c1}, T_{c2}, and T_d are respectively the sets of years in which data of total population, population by age, and number of deaths are available. T_m is the set of years for which smoothness of the net-migration index is assumed. *Av*[ ] stands for the average per unit data of the quantity within parentheses over all elements among a set for the summation. The averaging operation is to standardize weights of factors to obtain transparency of the quantity *F*, since number of elements for each factor considerably varies.

The first and second terms of *F* in the equation (1) represent discrepancies in populations at censuses. For the years $t \in T_{c1}$, only total populations are obtainable, while populations by age are available for the years $t \in T_{c2}$. In our data set T_{c1} =(1870, 1881, 1941), and T_{c2} =(1891, 1901, 1911, 1921, 1931, 1951). The third term of F is a factor to stand for inconsistency in mortality all over the projection period. It is expressed with discrepancies between projected and objective numbers of deaths. We introduced index of registration coverage c_t^d that designates proportion registered out of true number of deaths. The estimate of true number of deaths $\tilde{D}_t$ is, therefore, the observed number of deaths D_t divided by the coverage, or $\tilde{D}_t = D_t / c_t^d$. $\tilde{D}_t$ is the moving target of the estimates $\hat{D}_t$. The same modification applies to number of births.

Results

The tentative results of the projection for selected years are shown in Table 1. The projected total populations for the census years of 1891 and 1901 are somewhat smaller than the recorded data (Figure 1). The result implies that the recorded demographic data may be problematic in their consistency for the period before 1901. Examination of the population by age revealed that the underestimation in childhood and young adulthood causes the overall discrepancies. This may arise from two kinds of errors combined, a disproportionate under-count of infant deaths in the registration and an under-enumeration of infants at the censuses, which lead to smaller projected number of survivors in the subsequent ages.

Table 1 Projected Demographics for Population in Madras 7 States via GIP (Both Sexes)

Year	Population	Birth	Death	Migration	CBR	CDR	CNMR	e_0	IMR	Coverage (%)	
										Birth	Death
1866	8,204,737	212,858	216,694	-589	26.0	26.4	- 0.1	37.9	165.5		86.9
1871	8,479,487	281,134	226,109	5,300	33.0	26.6	0.6	39.0	142.7	67.4	74.2
1876	8,529,540	304,127	412,045	54,820	35.8	48.5	6.4	20.1	417.6	78.9	61.4
1881	8,638,711	339,872	215,427	115,876	38.8	24.6	13.2	41.6	128.5	76.3	72.5
1886	9,586,536	420,106	256,700	64,325	43.3	26.5	6.6	39.8	138.1	77.2	71.8
1891	10,727,762	446,343	362,099	10,235	41.4	33.6	0.9	31.6	221.6	78.9	77.8
1896	11,225,121	473,483	344,488	-641	41.9	30.5	- 0.1	33.7	197.7	76.1	73.4
1901	12,306,093	472,010	381,682	31,110	38.2	30.9	2.5	32.4	218.8	72.8	75.0
1906	12,521,042	549,123	443,096	-228,313	44.1	35.6	-18.3	28.3	252.4	71.9	76.5
1911	12,154,329	529,311	397,171	-183,760	43.6	32.7	-15.2	31.3	220.6	70.6	74.1
1916	12,182,349	555,360	379,238	-99,196	45.4	31.0	- 8.1	33.0	196.2	71.5	70.9
1921	12,226,468	460,850	300,433	140	37.4	24.4	0.0	38.4	153.5	73.2	73.2
1926	12,996,666	586,933	398,705	-1,458	44.8	30.5	- 0.1	33.2	193.4	78.3	74.6
1931	13,897,250	592,077	429,642	-2,099	42.4	30.7	- 0.2	32.8	206.5	84.7	76.1
1936	14,634,880	568,461	429,302	-1,368	38.7	29.2	- 0.1	32.8	203.8	90.1	74.9
1941	15,425,798	584,912	459,389	-4,661	37.8	29.7	- 0.3	32.0	213.2	96.4	74.8
1946	16,051,529	534,052	400,201	133,218	33.0	24.7	8.2	36.1	168.1	98.4	74.4
1951	17,308,557										

As for census populations in 1911 and after, the projected results match with the reported numbers fairly well both in total and by age. The fit in population size improves as it moves toward near past approaching the terminal census in 1951. The fits for year 1951 are visually presented Figure 2. This improvement of the fit indicates that the recorded demographic data are more consistent in the near past.

Discussion and Conclusion

In our attempt, time series of demographic characteristics of Madras 7 districts over the period

1866-1951 are simultaneously estimated as a set by minimizing the single penalty value. First, we obtained time series of population by age and number of vital events in Madras 7 states through GIP. Then we use them to test the consistency among the projected estimates of some demographic variables. As a result, some inconsistencies are identified among the data. We found that massive omission of birth registration should be supposed to be consistent with the mortality estimates. Among mortality estimates themselves, it is suggested that slightly smaller number of infant deaths relative to the other ages should be registered in the period.

Figure 1 Projected and Census Population of Madras 7 States

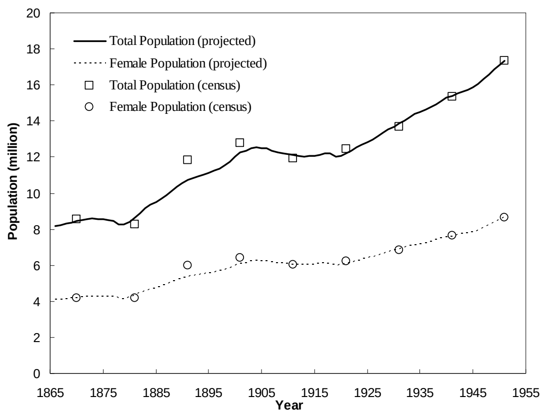
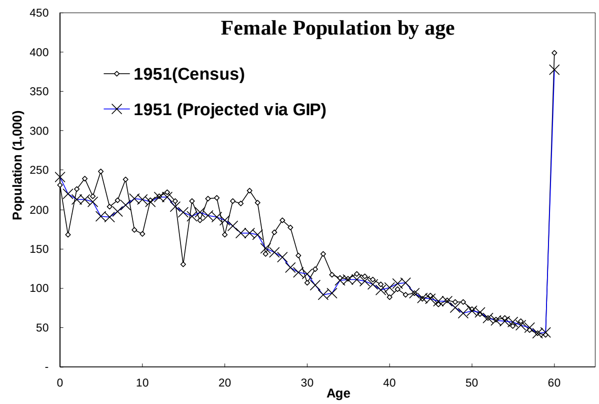


Figure 2 Projected and Census Population Age Distribution of Madras 7 States in 1951



In addition, the completeness of death registration should be much lower than the level of 75% for the period before 1891 so long as the life expectancy must not exceed the level of the following period. With this finding together, the relatively large discrepancies between the projected and input census populations in early period suggest that substantial inconsistency exists among data for the period. To the contrary, the projected results suggest that the completeness of death registration should have increased after 1932 possibly well above the level of 75%.

There are several limitations with the method employed in this study such as arbitrariness in choice of the coverage level of death statistics, which is inevitable for situation where incompleteness in both birth and death registration is presumed, though a reliable estimate even at one year improves the situation a lot. In spite of the methodological limitations and preliminary nature of this study, we found that GIP is a strong method to approach an amorphous projection practice in which only fragmented or under-reliable information on population process is available. Even though it does not produce a final estimate in an offhand manner when the input data are not internally consistent, it provides useful indication on loci of inconsistency among the information, leading to finer estimates of the population history.

Selected References

Oeppen, Jim. 1993b. "Generalized inverse projection." In: *Old and new methods in historical demography*, edited by David S. Reher and Roger Schofield. Oxford.

Saito, O., M. Takahama, and C. Yamamoto, 2003, "Notes on Estimation of Vital Rates in a Pre-independence Indian province," Working Paper No. 117, Institute of Comparative Economics Studies, Hosei University; Tokyo.