A PROCEDURE FOR POINT AND INTERVAL POPULATION PROJECTION FOR SMALL AREAS.

Ramiro Coa Clemente[a]
Claudia Julia Horta[b]
Diana Oya Sawyer[c]

Abstract

When compared to populations in larger areas, those in smaller areas are prone to greater size and trend instability. Making these characteristics the main constraints in a population projection procedure, an application of the 'bootstrap' technique to stratified areas, by their population size and trend, is proposed to generate point and interval estimates.

The above procedure was followed to project the populations in 559 municipalities in São Paulo State, in Brazil. This resulted in a considerable increase in quality when compared to projections with no stratification. Point projections increased in their level of agreement with the census population in about 40%. Although interval projections had their power to capture the census population decreased by 13,5%, the interval amplitude narrowed an average of 55%.

1. Introduction

Population estimates and projections are increasingly demanded due to the administrative decentralization policy implemented in many Latin-American countries, which led to socioeconomic development plans and strategies for small geographical areas.

Point population estimates and projections are the most demanded by overall policy makers. Those estimates, however, tend to be uninformative, since they do not reflect their precision level. The confidence level of interval population projections might contribute more to a successful target-oriented policy than a point projection with an unknown level of confidence. Policy makers themselves might be more comfortable knowing the absolute error for the target population they are addressing.

a CEDEPLAR/UFMG, graduate student, rcoa@cedeplar.ufmg
b CEDEPLAR/UFMG, graduate student, cjhorta@cedeplar.ufmg.br
c CEDEPLAR/UFMG, full professor, dsawyer@cedeplar.ufmg.br

Population projections for small areas require that special attention be given to their demographic characteristics: i) population size variability, e.g In1970, in the Brazilian state of São Paulo, the populations of municipalities ranged from less than 1,000 to more than 300,000, the capital city's population of more than four million not included; ii) high sensitivity to large increments or decrements in very short time intervals. Small areas can easily decrease to half their initial population or twofold it in a five to ten-year period; iii) oscillating growth trends. Population can increase in one period and decrease in the next or vice-versa. Those characteristics are the main error sources in the usual small area population projections.

Many methods of population projection, specifically for small areas, have been suggested so far (see: DUCHESNE 1989, EBERSTEIN et al., 1984; HEIDE 1981, IRWIN 1977, JARDIM 1995, MADEIRA & SIMÕES 1972, NAKOSTEEN 1989, WALDVOGEL 1977,1989, among others), in an effort to overcome the aforementioned problems. Our main objective is to propose point and interval population projections and estimates for small areas, in accordance with the current method of Apportionment used by the Brazilian Institute of Geography and Statistics (IBGE), the official organization responsible for providing population projections in Brazil, using the municipalities in the state of São Paulo as a case study.

2. Population projection procedures

Firstly, an assumption is made regarding the composition of a specific area by a set of smaller units. Furthermore, it is assumed that a projected population is available for the larger unit. The projection for smaller areas is referred to the larger area according to the Apportionment method that can be summarized as (MADEIRA & SIMÕES 1972):

$$P_i(t) = A_i * P(t) + B_i \qquad (1)$$

$$Bi = P_i\,(t-10) - Ai * P(t) \qquad (2)$$

where:

$P_i(t)$: Projected population of the smaller area i, at time t;

P(t): Larger area's population (projected by an independent method, at time t;

$P_i(t-10)$: Smaller area census population at the last available census date (usually 10 years previous to the projection date);

Ai: Proportion representing the population increment (decrement) of the smaller in relation to the larger area, in a recent past;

B_i: Parameter of linear adjustment for the smaller area.

Under the restriction :

$$\sum_i Ai = 1, \quad \sum_i Bi = 0.$$

In this method, the relative share of each smaller area in the larger area's population increment will hold for the duration of the projection period. If this were a fact, all problems would be solved. The magnitude and direction of the shares can change within a short period of time, however.

One of many solutions is to assess the range of the future relative share variability for a specific unit, which, in our case, will be achieved through a bootstrap re-sampling (EFRON & TINSHIRANI 1998). From recently observed shares, an empirical distribution of the Apportionment equation's parameters can be estimated. The empirical distribution will then be the base for point and interval estimation for the projected smaller area's population. .

A stratification of smaller units prior to re-sampling is recommended for controlling the aforementioned demographic constraints. The stratification parameters should reduce the effects of: oscillating share trend (areas that contributed positively to growth in the larger area can contribute negatively in the following period); high variability in the share's magnitude among smaller units and variability of the share's magnitude in the same unit, along the projection period (an area with a reduced share in one period may substantially increase its share in the next period).

3. The case study

The state of São Paulo was chosen for testing the proposed procedure, mainly because between 1970 and 1991 (test period), the state experienced intense migratory movements. The 1970 and 1980 censuses gave the baseline for calculation of each municipality's relative share, and the 1991 census was used to compare the projected population by sex and to assess its quality.

Among 571 municipalities, twelve were excluded, because: eleven had populations under 1,000 inhabitants and one had over 4 million. The remaining municipalities (559), were stratified by two of the three aforementioned criteria: share's trend and share's magnitude among municipalities.

Share's trend was evaluated in two consecutive 10-year periods: 1970-80 and 1980-90. Table 1 shows the classification of the trends in both periods, the (+) symbol representing municipalities with increasing population in the period and the (-) symbol representing decreasing population. Thus, municipalities increasing in both periods are represented by (+,+), which represents a relative stability.

Share's magnitudes were evaluated according to their degree of homogeneity, measured by their coefficient of variation. If that coefficient for a municipality group was between 50 and 90, the group was considered homogenous. Table 1 shows details of the 8 strata for male and female population.

Table1. Population characteristics of the municipality strata by sex, growth share trend and share's coefficient of variation (CV).
State of São Paulo, Brasil. 1970-1980.

	Growth	Male				Female				Number
	share	CV	Population in 1980			CV	Population in 1980			of
Stratum	Trend	by 100	mean	Minimum	maximum	by 100	mean	Minimum	maximum	Municipalities
1	(-,-)	87	3128	1139	11060	86	2919	1077	10379	81
2	(+,+)	79	5693	1336	20210	71	5369	1225	19714	154
3		55	11958	5039	19901	53	11368	4764	19083	74
4		87	40728	20190	99365	84	40209	20029	97639	79
5		53	186091	104011	328657	52	185712	100644	333584	13
6	(-,+)	69	5055	1102	17156	66	4798	1097	16378	72
7		107	3515	1031	19490	66	3318	1003	9962	64
8	(+,-)	64	4066	1361	8447	67	3811	1278	8045	22
Total		**338**	**14902**	**1031**	**328657**	**335**	**14556**	**1003**	**333584**	**559**

Source: Elaborated from de 1970 and 1980 Demographic Censuses

The non parametric bootstrap re-sampling was performed (1,000 replications) to estimate the empirical distribution of A_i. For each municipality the mean value of the distribution was taken as the bootstrap estimate of A_i and its respective 95% confidence interval was then obtained. The values of $P_i(t)$ were then estimates through equations (2) and (1).

4. Criteria for assessing quality of projection

The criteria for assessing the point projection's quality were the 90th and 95th percentiles of the distribution of the relative difference between the municipality projected population and the census population in 1991 (DR_i). Thus:

$DR_i = (PEM_i - PVM_i)/PVM_i$;

where,

PEM_i is the projected population, in 1991, for municipality i,

PVM_i is the census population, in 1991, for municipality i.

The interval projection's quality was evaluated by means of two indexes. The first refers to the capacity of the projected interval to encompass the census population, hence the capture index is the relative number of municipalities with census population within the boundaries of the 1991 projected interval. The second is the 90th and 95th percentile of the standardized amplitude of the confidence interval for municipality i (AI_i), defined as:

$AI_i = (LSE_i – LIE_i)/(LSE_i + LIE_i)$;

where,

LSE_i = Estimated upper bound population for municipality i, in 1990

LIE_i = Estimated lower bound population for municipality i, in 1990

5. Results and comments

Table 2 shows values of 90th and 95th percentiles of DRi, in four situations: plain Apportionment with previous stratification and no previous stratification and bootstrap estimates with previous stratification and no previous stratification.

Table 2: 90th and 95th percentiles (P90, P95) of the Maximum Relative Difference (DRi) between census and projected population by sex and type of projection. State of São Paulo, Brazil, 1991

	Number of Municipalities	Male		Female	
		P90	P95	P90	P95
No stratification, no Bootstrap	559	44	57	45	58
No stratification, with Bootstrap	559	39	46	40	47
With stratification, no Bootstrap	559	26	35	23	35
With stratification, with Bootstrap	559	27	34	27	35

Source: Projected population and 1991 Demographic Census

The projection with no previous stratification yields DRi (no bootstrap) of about 57% for men at the 95th percentile, the corresponding figure for women is 58%, meaning that the maximum relative difference between the census and estimated population for 95% of the municipalities is 57% for men and 58% for women. The advantage of previous stratification is very clear, both figures went down to 35%, getting to an improvement of 40% for both sexes. Table 2 also shows that there is no greater improvement by using bootstrap in point projections, further to previous stratification.

The stratification effect is very clear, in Table 3, for interval projection. It is worthy of mention that those two criteria to assess interval projection: capture and confidence interval's range, have opposite effects. In other words, the narrower the range of the interval, the lesser is the power of capture. The judgement should be made by an optimization of both criteria in order to achieve the best capture from a narrower range.

Table 3. Capture index and 90th and 95th percentiles (P90, P95) of the maximum standardized amplitude of the 95% Confidence interval, by sex and type of population projection. State of São Paulo, Brazil, 1991.

Type of Projection	Number of Municipalities	Male			Female		
		Capture Index	Amplitude		Capture Index	Amplitude	
			P90	P95		P90	P95
No stratification, with Bootstrap	559	94,1	99	102	95	99	101
With stratification, with Bootstrap	559	84	41	45	80	41	46

Source: Projected population and 1991 Demographic Census

The absence of stratification leads to a degree of capture of 94% for the male population and 95% for the female population, which means that the projected interval had included the census population in 94% of the municipalities' male population. Albeit those promising figures of captures, the intervals are very wide. In 95% of the municipalities the interval's range indexes exceeded 100%, a useless outcome.

A previous stratification, however, reduced those indexes down to about 45%, and no great loss in the capture index was observed, it was reduced by 11% for the male population and by 16% for the female.

As for the quality of the projections by strata, there were no clear differences between stable share trends (+,+ or -,- type), and unstable share trends (+,- or -,+ type), especially when applying bootstrap. The quality indicators do not favor any specific group, for neither point nor interval projections (Tables 4 and 5). The explanation may lie in the fact that the stratification did not consider the variability of the share's magnitude, in the same municipality, along the projection period. In favor of this is the high positive correlation (0.78 for the male and 0.83 for the female population) between the point projection‘s DRi values and the variability of the participation magnitude in two periods (CRPi) shown in Table 4. This indicates that the greater the projection errors, the greater the shares’ variation between the 1970-80 and the 1980-90 periods.

Table 4: 90th and 95th percentiles (P90, P95) of the Maximum Relative Difference (DRi) between census and projected population and Maximum Relatice Difference between 1980-91 and 1970-1980 shares (CRP). by sex , stratum and type of rojection. State of São Paulo, Brazil , 1970-91.

Stratum parameters and Type of projection			Male			Female		
Growth Share Trend	Coefficient of * variation	Number of municipalities	P90	P95	CRP**	P90	P95	CRP**
No Bootstrap								
(-,-)	87	81	26	30	69	21	25	79
(+,+)	75	154	23	27	106	21	26	105
	54	74	19	22	55	20	22	50
	85	79	21	22	51	17	21	47
	53	13	11	12	31	11	12	27
(-,+)	68	72	33	40	152	40	47	161
	87	64	41	48	246	42	57	339
(+,-)	66	22	25	28	227	17	24	213
With Bootstrap								
(-,-)	87	81	17	20	-	15	17	-
(+,+)	75	154	32	45	-	33	46	-
	54	74	26	31	-	27	33	-
	85	79	29	34	-	29	32	-
	53	13	20	20	-	19	21	-
(-,+)	68	72	28	34	-	32	38	-
	87	64	26	31	-	31	34	-
(+,-)	66	22	24	26	-	23	24	-
Pearson correlation between P95 and CRP				**0,78**			**0,83**	

Source: Elaborated from de 1970, 1980 and 1991 Demographic Censuses

* average for males and females, by 100 .

Table 5. Capture index and 90th and 95th percentiles (P90, P95) of the maximum standardized amplitude of the 95% confidence interval, by stratum, sex and type of population projection. State of São Paulo, Brazil, 1991

Stratum parameters and Type of projection			Male			Female		
Growth Share Trend	Share's Coefficient * of variation	Number of municipalities	Capture Index	Amplitude P90	Amplitude P95	Capture Index	Amplitude P90	Amplitude P95
With Bootstrap								
(-,-)	87	81	91,4	43	52	84	33	40
(+,+)	75	154	92,2	43	48	92	44	49
	54	74	83,8	33	34	84	35	36
	85	79	92,4	45	45	92	46	47
	53	13	77,0	23	24	77	25	25
(-,+)	68	72	58,3	26	30	47	23	26
	87	64	82,8	36	39	67	30	32
(+,-)	66	22	63,6	41	61	73	57	88
No stratification, No Bootstrap		559	94,1	99	102	95	99	101

Source: Elaborated from de 1970, 1980 and 1991 Demographic Censuses

* average for males and females, by 100 .

The results so far have shown that, in average, a previous stratification has the potential to considerably increase the quality of point and interval projections. An improvement in quality is foreseen through the introduction of the third stratification criteria, missing in this exercise, namely the magnitude of the variation in the small area's share to the larger area growth. Furthermore, to better apprehend the variation in trends, the stratification over parameter in larger periods of time, possibly three intercensal periods, is recommended, especially for areas with high level of population mobility.

BIBLIOGRAPHY

DUCHESNE, L. Proyecciones de población por sexo y edad para áreas intermedias y menores: Método 'relación de cohortes'. In: Granados, M. P. (com). *Métodos para proyecciones subnacionales de poblacion*. Bogota: CELADE, 1989, p. 71-126.

EFRON, B & TIBSHIRANI, RJ. An introduction to the Bootstrap. Boca Raton, London, New York, Washington D.C, Chapman & Hall/CRC, 1998. Monographs on Statistics and Applied Probability 57.

EBERSTEIN IW; SEROW WJ; RIVES NW; MAYBERRY L; McLEOD KD; PARKER JR "*On the applicability of cohort-survival models to census tracts: Harris County, Texas, 1980*". Tallahassee, Fla, Florida State University, College of Social Sciences, Institute for Social Research, Center for the Study of Population, 1984. [24] p. Center for the Study of Population Working Paper no. 84-14

HEIDE H ; "Demographic distribution formulas: an aid to preparing population projections for small areas". The Hague, Netherlands, National Physical Planning Agency, 1981. 28 p.

IRWIN, R.; "Methods and data sources for population projections of small areas", In: *Population forecasting for small areas: an examination of the state of the art of small area forecasting as applied to the problems of planning*. Proceedings of the conference at Oak Ridge, Tennessee, May 5-6, 1975, sponsored by the Oak Ridge Associated Universities and by the Urban Research Prog Oak Ridge Associated Universities, Oak Ridge, Tenn., United States, 1977. :15-26.

IRWIN, R.; "Use of the cohort-component method in population projections for small areas"', In: *Population forecasting for small areas: an examination of the state of the art of small area forecasting as applied to the problems of planning*. Proceedings of the conference at Oak Ridge, Tennessee, May 5-6, 1975, sponsored by the Oak Ridge Associated Universities and by the Urban Research Prog Oak Ridge Associated Universities, Oak Ridge, Tenn., United States, 1977. :37-

48.

JARDIM TEIXEIRA M. de L., *Utilização de Variáveis Sintomáticas para estimar a distribuição espacial de populações : Aplicação aos municípios do Rio Grande do Sul.* – Dissertação de Mestrado – CEDEPLAR/FACE/UFMG – Belo Horizonte, 1995.

MADEIRA , J. L. & SIMÕES , C. C.. S. Estimativas preliminares da população urbana e rural, segundo as unidades da Federação, 1960/1980: por uma nova metodologia. *Revista Brasileira de Estatística*. Rio de Janeiro, v.33, n.129, p.3-11, 1972.

NAKOSTEEN R. A.; "Detailed population projections for small areas: the Massachusetts experience", *SOCIO-ECONOMIC PLANNING SCIENCES*. 1989; 23(3):125-38.

WALDVOGEL G., *Técnicas de Projeção populacional para o planejamento regional* – Cedeplar, Belo Horizonte, Estudos Cedeplar, 1977, N. 1 – 162 pags.

WALDVOGEL G. Demographic proportional parameters: an alternative in applying the "components method" to project the population of small areas. [Parametros demograficos proporcionales: una alternativa para aplicar el "metodo de los componentes", para proyectar la poblacion de areas pequenas. In: *Metodos para proyecciones subnacionales de poblacion*, [compiled by] Maria del Pilar Granados. Bogota, Colombia, Centro Latinoamericano de Demografia, 1989 Nov. :49-69. Serie 0l No. 42

Table1. Population characteristics of the municipalities strata by sex, growth share trend and share's coefficient of variation (CV).

State of São Paulo, Brasil. 1970-1980.

Stratum	Growth share Trend	Male				Female				Number of Municipalities
		CV by 100	Population in 1980			CV by 100	Population in 1980			
			mean	minimum	maximum		mean	minimum	maximum	
1	(-,-)	87	3128	1139	11060	86	2919	1077	10379	81
2	(+,+)	79	5693	1336	20210	71	5369	1225	19714	154
3		55	11958	5039	19901	53	11368	4764	19083	74
4		87	40728	20190	99365	84	40209	20029	97639	79
5		53	186091	104011	328657	52	185712	100644	333584	13
6	(-,+)	69	5055	1102	17156	66	4798	1097	16378	72
7		107	3515	1031	19490	66	3318	1003	9962	64
8	(+,-)	64	4066	1361	8447	67	3811	1278	8045	22
Total		**338**	**14902**	**1031**	**328657**	**335**	**14556**	**1003**	**333584**	**559**

Source: Elaborated from de 1970 and 1980 Demographic Censuses

Table 2: 90th and 95th percentiles (P90, P95) of the Maximum Relative Difference (DRi) between census and projected population by sex and type of projection. State of São Paulo, Brazil, 1991

	Number of	Male		Female	
	Municipalities	P90	P95	P90	P95
No stratification, no Bootstrap	559	44	57	45	58
No stratification, with Bootstrap	559	39	46	40	47
With stratification, no Bootstrap	559	26	35	23	35
With stratification, with Bootstrap	559	27	34	27	35

Source: Projected population and 1991 Demographic Census

Table 3. Capture index and 90th and 95th percentiles (P90, P95) of the maximum standardized amplitude of the 95% confidence interval, by sex and type of population projection. State of São Paulo, Brazil, 1991

Type of Projection	Number of Municipalities	Male			Female		
		Capture Index	Amplitude		Capture Index	Amplitude	
			P90	P95		P90	P95
No stratification, with Bootstrap	559	94.1	99	102	95	99	101
With stratification, with Bootstrap	559	84	41	45	80	41	46

Source: Projected population and 1991 Demographic Census

Table 4: 90th and 95th percentiles (P90, P95) of the Maximum Relative Difference (DRi) between census and projected population and Maximum Relatice Difference between 1980-91 and 1970-1980 shares (CRP). by sex , stratum and type of projection. State of São Paulo, Brazil , 1970-91.

Stratum parameters and Type of projection		Number of municipalities	Male			Female		
Growth share Trend	Coefficient of * variation (municipality share on strata growth)		P90	P95	CRP**	P90	P95	CRP**
No Bootstrap								
(-,-)	87	81	26	30	69	21	25	79
(+,+)	75	154	23	27	106	21	26	105
	54	74	19	22	55	20	22	50
	85	79	21	22	51	17	21	47
	53	13	11	12	31	11	12	27
(-,+)	68	72	33	40	152	40	47	161
	87	64	41	48	246	42	57	339
(+,-)	66	22	25	28	227	17	24	213
With Bootstrap								
(-,-)	87	81	17	20	-	15	17	-
(+,+)	75	154	32	45	-	33	46	-
	54	74	26	31	-	27	33	-
	85	79	29	34	-	29	32	-
	53	13	20	20	-	19	21	-
(-,+)	68	72	28	34	-	32	38	-
	87	64	26	31	-	31	34	-
(+,-)	66	22	24	26	-	23	24	-
Pearson correlation between P95 and CRP				**0.78**			**0.83**	

Source: Elaborated from de 1970, 1980 and 1991 Demographic Censuses

* average for males and females, by 100 .

Table 5. Capture index and 90th and 95th percentiles (P90, P95) of the maximum standardized amplitude of the 95% confidence interval, by stratum, sex and type of population projection. State of São Paulo, Brazil, 1991

Stratum parameters and Type of projection			Male			Female		
Growth share Trend	Coefficient of * variation (municipality share on strata growth)	Number of municipalities	Capture Index	Amplitude P90	Amplitude P95	Capture Index	Amplitude P90	Amplitude P95
With Bootstrap								
(-,-)	87	81	91.4	43	52	84	33	40
(+,+)	75	154	92.2	43	48	92	44	49
	54	74	83.8	33	34	84	35	36
	85	79	92.4	45	45	92	46	47
	53	13	77	23	24	77	25	25
(-,+)	68	72	58.3	26	30	47	23	26
	87	64	82.8	36	39	67	30	32
(+,-)	66	22	63.6	41	61	73	57	88
No stratification, No Bootstrap		559	94.1	99	102	95	99	101

Source: Elaborated from de 1970, 1980 and 1991 Demographic Censuses

* average for males and females, by 100 .