香港統計學會有限公司
Hong Kong Statistical Society Limited
c/o Department of Statistics & Actuarial Science,
The University of Hong Kong, Pokfulam Road,
Hong Kong

Bulletin
Volume 23 Number 2
September 2000

Editor's Foreword

As usual, we have three interesting articles in this issue. In the President's Forum, President Li again gives an updated narration on the development of the Society. In particular, we will soon be holding an EGM to discuss matters concerning the waiving of the word "Limited" from the name of the Society. In another article, Dr. Wai Chan, the associate editor, discusses the use of a combined weight matrix method in the covariance structure analysis. Finally, Mr. Y.H. Lee of the Census and Statistics Department introduces the method for revising the compilation of population estimates currently used by the Census and Statistics Department.

I hope you will enjoy reading the bulletin. In the News Section, you will find information on the upcoming 5th International Chinese Statistical Society Conference to be held at the University of Hong Kong.

Alan Wan

			Tel.	Fax.
Editor	:	Wan, Alan, CityU	2788 7146	2788 8560
Associate Editors	:	Chan, Wai, CUHK	2609 6241	2603 5019
		Chan, Ping-shing, CUHK	2609 7920	2603 5188
Secretary	:	Lam, John Hon-kwan, C&SD	2582 4899	2802 1101

CONTENTS

(Vol. 23/No. 2, September 2000)

President's Forum

Professor W.K. Li

I hope you all had a nice and enjoyable summer. Now that the summer is over and I am happy to report that the status of the Society is sound and healthy.

First, as you are aware we have been applying to the Companies' Registry for waiving the word "Limited" after the name of the Hong Kong Statistical Society. There has been some progress in this direction and we will be calling for an EGM on 5 December to make the necessary arrangements in the constitution. Once again we need members' support on this matter. Second, the Statistical Project Competition for Secondary School Students is now well underway. A briefing seminar and exhibition has been scheduled on November 4 at the City University of Hong Kong. Third, we are planning for an outing on January 7, though the destination is yet to be decided. I look forward to seeing you and your family in the outing. Fourth, we had a very well attended seminar delivered by Professor Jianqing Fan of the Chinese University of Hong Kong in August. Professor Fan has just been awarded the prestigious Committee of Presidents of Statistical Societies (COPSS) award this year. This is a great honour to the Hong Kong statistical community and I would like to congratulate Professor Fan on behalf of the Society.

Revision to the Method of Compiling Population Estimates of Hong Kong

Y.H. Lee
Census and Statistics Department

Introduction

In recent years, there has been a significant change in the residency and mobility patterns of the Hong Kong Population. The living and working boundaries of many Hong Kong people cover not only Hong Kong, but also the Mainland/Macau or even overseas. After carefully considering the residency and mobility patterns of the Hong Kong Population, international trend for the compilation of population estimates, statistical conceptualization of statistical indicators and development of information technology, the Census and Statistics Department (C&SD) has decided to replace the "extended de facto" method used previously by the "resident population" method for compiling population estimates. This change has been announced in August 2000, together with related data. This article explains the rationale for the change and introduces the concept of the "Hong Kong Resident Population".

Previous method of compiling population estimates

All along, the so-called "extended de facto" method has been used by C&SD to compile population estimates. Under a "de facto" concept, the population includes all persons who are in Hong Kong at the reference time-point, i.e. this method is equivalent to taking a 'snapshot' of the population at a reference time-point. The population includes Hong Kong Permanent Residents, Hong Kong Non-permanent Residents and visitors.

"Extended" relates to the fact that for a Hong Kong Permanent Resident, he/she will still be counted as part of the Hong Kong Population if, at the reference time-point, he/she is not in Hong Kong but temporarily in the Mainland or Macau.

The application of the "extended" compilation method is intended to avoid fluctuations in the "de facto" population figure around major public holidays when the movement of people between Hong Kong and Mainland/Macau is enormous. And in the past, most such departures to Mainland/Macau were of a temporary nature.

In practical terms, population

estimates are compiled as described below.

Population estimate at a "census moment" based on the "extended de facto" method

A population base is established at each population census (or by-census) moment. In addition to those who are physically in Hong Kong, also included are those Hong Kong residents who are working in Mainland/Macau and those who are usually living in Hong Kong but in Mainland/Macau for short trips at the census moment, and are reported by family members in Hong Kong as such.

Population estimate at any reference time-point in between census moments based on the "extended de facto" method

The population estimate at any reference time-point (t) is obtained by the following formula:

Population (at time-point *t*)
= Population (at census moment *c*, based on the concept of the "extended de facto" method)
+ births (from *c* to *t*)
– deaths (from *c* to *t*)
+ arrivals (from *c* to *t*)
– departures (from *c* to *t*)

and "arrivals – departures" excludes departures and arrivals of Hong Kong Permanent Residents to/from Mainland/Macau during the period from *c* to *t*. The implementation of this exclusion is to deal with the situation that a great deal of departures to Mainland/Macau shortly before the reference time-point may be related to short-term stays (though these would be matched with arrivals back in Hong Kong shortly after the time-point) and may consequently lead to highly fluctuating population estimates if they are all included.

Limitations of the "extended de facto" method

Since the residency and mobility patterns of the Hong Kong Population are undergoing change, the "extended de facto" method has gradually become less appropriate. The main reasons are :

(a) The trend for Hong Kong Permanent Residents to stay in the Mainland or Macau on a long-term basis for reasons of retirement, work or marriage is on the increase. Also, a considerable number of Hong Kong residents are staying overseas for work, touring or studies at the reference time-point. Applying the "de facto" method would lead to fluctuations in the population estimates and limit its ability to reflect the actual situation.

(b) According to the definition of "de facto" population, visitors to Hong Kong are included in the Hong

Kong Population. But strictly speaking, visitors are not part of the Hong Kong Population.

(c) The existing compilation method makes use of the difference between the number of arrivals and the number of departures. As those numbers are large, their difference may not be stable at times, especially around the commonly used reference time-points, e.g. end-June and end-December.

In regard to (a) above, the application of the extended de facto method leads to some over-estimation of the population because indeed there are people who go to Mainland/Macau for long-term stay or even for good (rather than just for short trips), and the trend is for this to intensify. It should be noted, though, that the population estimate released about 2 months after a reference time-point is, strictly speaking, *provisional* in nature.

This is because, when an ensuing population census (or by-census) is conducted and the new population base is established, those who have indeed gone to Mainland/Macau for long-term stay or for good would not be counted and for the population estimates in respect of the intervening reference time-points, a "retrojection" process will be carried out, with the number of those people deducted. In other words, the provisional population estimates are finalized after the next population census/by-census has been taken. (It has to be noted, though, that the population census/by-census performs the benchmarking function, such that the finalization process which involves revision of estimates is not solely due to the above factor.)

However, the finalization takes place years after the individual reference time-points. For example, the estimate for mid-1992 was "finalized" only in the latter part of 1996 when data from the 1996 By-census (March 1996) became available, and hence a time lag of slightly more than 4 years. (And for the estimate for mid-1993: a time lag of slightly more than 3 years; for mid-1994: 2 years; and so on.)

Previously, when the volume of traffic and the number of people going to Mainland/Macau for long-term stay (or for good) were not so large, the problem was limited as the extent of "revision" would also be relatively small. That is, the "provisional" estimates could be used in the interim (a rather long interim period though) without too much concern. The situation is, however, changing.

Merits of changing over to the "resident population" method for compiling population estimates

In the early days, many countries

and territories applied the “de facto” method for compiling population estimates. The main reason is that the compilation procedure is simple and the required data are easier to obtain. With statistical development and advancement, many countries and territories have gradually changed over to using the “resident population” method.

In changing over to using the “resident population” method for compiling population estimates of Hong Kong, the following considerations are relevant:

(a) The residency and mobility patterns of the Hong Kong Population

Recent studies have shown that there have been notable changes in the residency and mobility patterns of the Hong Kong Population. The application of the “resident population” method could better reflect the actual situation.

(b) International trend

Many countries and territories have changed from the “de facto” method to the “resident population” method for compiling population estimates. Hong Kong should also follow this international trend.

(c) Statistical conceptualization

Using the “resident population” concept to measure the population size of a place is actually more relevant from the statistical theory standpoint. Its previous non-application was mainly due to limitations of insufficient data. Also, in compiling various socio-economic statistical indicators (e.g. birth rate and death rate), the use of the “resident population” concept is generally more meaningful.

(d) Development of information technology

The computational capabilities of computers now available have enhanced significantly. Large volume of anonymized passenger movement data can be handled, thus enabling the computation of “resident population” figures at the reference time-point.

Revising the method for compiling population estimates

Concluding from the results of relevant studies, it is considered that Hong Kong has met the necessary requirements of applying a compilation method which is based on the concept of “resident population”. In doing so, Hong Kong follows the statistical practice of other advanced economies.

"Resident population" is a clear-cut concept in international statistical standard but the practical definitions adopted vary from place to place, as the residency and mobility patterns unique to each place need to be given adequate consideration. International statistical organizations have pointed out in particular that, owing to business and social development, the "mobility" of residents of certain countries/territories is rather high. In handling the population statistics of these countries/territories, the appropriate authorities should consider the situation in depth.

In the case of Hong Kong, our studies show that the "resident population" of Hong Kong (which we refer to as the "Hong Kong Resident Population") should be defined to include "Usual Residents" and "Mobile Residents".

In consideration of issues relating to concepts and availability of data, the technical definition of the "Usual Residents" would include two categories of people: (1) Hong Kong Permanent Residents who have stayed in Hong Kong for at least three months during the six months before or for at least three months during the six months after the reference time-point, regardless of whether they are in Hong Kong or not at the reference time-point; and (2) Hong Kong Non-permanent Residents who are in Hong Kong at the reference time-point.

As for "Mobile Residents", they are Hong Kong Permanent Residents who have stayed in Hong Kong for at least one month but less than three months during the six months before or for at least one month but less than three months during the six months after the reference time-point, regardless of whether they are in Hong Kong or not at the reference time-point.

The amount of time of stay in Hong Kong of "Mobile Residents" is less than that of the "Usual Residents". Nevertheless, the "Mobile Residents" have a close link with Hong Kong and most probably they have a regular residence in Hong Kong and utilize much of Hong Kong's facilities and services. In this regard, they should be considered as part of the Hong Kong Population.

Under the new method for compiling population estimates, a population census/by-census continues to serve the benchmarking function. Updating of the population estimates to any new reference time-point, however, does not depend on just immigration control-point data on the balance of total arrivals over departures in regard to *population movement*. Instead, in respect of Hong Kong Permanent Residents, arrival and departure records of individual persons are linked for statistical computation (*albeit* the records are anonymised and un-identifiable to the individuals). For Hong Kong

Non-permanent Residents, the balance of arrivals over departures is still used, since the record-linkage method on individuals cannot be applied to them given the form of their arrival/departure records. Also, they are grouped under "Usual Residents". This is because for the duration that they hold that status of "Non-permanent Resident" they can be expected to be usually staying in Hong Kong.

Below is an illustration of the situations of different Mobile Residents:

(1) Persons staying 5 to 6 days each week in the Mainland and staying regularly in Hong Kong during the weekends. As at end-1999, there were some 56 000 such persons. They are mainly those who work in the Mainland and return to Hong Kong to stay with their families during the weekends.

(2) Persons staying for a major proportion of time in Mainland/Macau and returning to Hong Kong from time to time. There were some 72 000 such persons at end-1999. Presumably they maintain residences in both Mainland/Macau and Hong Kong, and travel frequently between these places for business, work or family reasons.

(3) Elderly persons (aged 60 or over) staying for a major proportion of time in Mainland/Macau. There were some 7 000 such persons. Many of them have retired and settled in the Mainland but come back to Hong Kong frequently to visit children or relatives.

(4) Persons in school-attending age returning from overseas to stay in Hong Kong for several periods of time in a year. There were some 21 000 such persons. These are mainly Hong Kong students studying abroad and spending their vacations in Hong Kong.

(5) Besides, there are persons staying in overseas countries/territories but also staying frequently in Hong Kong for business, work or family reasons. There were some 32 000 such persons at end-1999.

Chart 1 provides a diagrammatic presentation of the residency/mobility profiles.

As at end-1999, there were a total of 188 000 "Mobile Residents". The above figures have been obtained through statistical processing of anonymized passenger movement data. The compilation procedures are conducted through computer processing, with all the personal data being anonymous and the identities kept strictly confidential.

Comparison of the old and the new population estimates

On the basis of the new method for compiling population estimates, the "Hong Kong Resident Population" as at end-1999 is 6.76 million, which is less than the previously released "extended de facto" population by 0.21 million (i.e. 3%). However, one should note that the previously released figure (6.97 million) includes 0.14 million visitors, who are not included in the figure of "Hong Kong Resident Population".

Table 1 presents the population estimates from 1996 to 1999, compiled based on the old and the new methods. Since the required data are not available for earlier dates, the estimates of "Hong Kong Resident Population" in the statistical series are only provided in respect of reference time-points from 1996 onwards. Table 2 compares sex-age composition of the population as at end-1999 between the old and the new methods.

Release of population estimates

The application of the "resident population" method requires information on the amount of time of stay in Hong Kong of residents during the six months before and after the reference time-point. It follows that the population estimates could not be obtained within one or two months after the reference time-point.

In order to provide relevant figures as early as possible for use, the C&SD will release "provisional estimates" for reference within two months after the reference time-point. Revised estimates will be released a further six months later.

That is to say, population estimates will be released every half-year, for the mid-year and year-end positions, by way of press release as in the past. To provide a concrete example, the provisional estimates for mid-2000 were released on 22 August 2000, while the revised estimates will be released in mid-February 2001.

The latest population estimates are shown in Table 3.

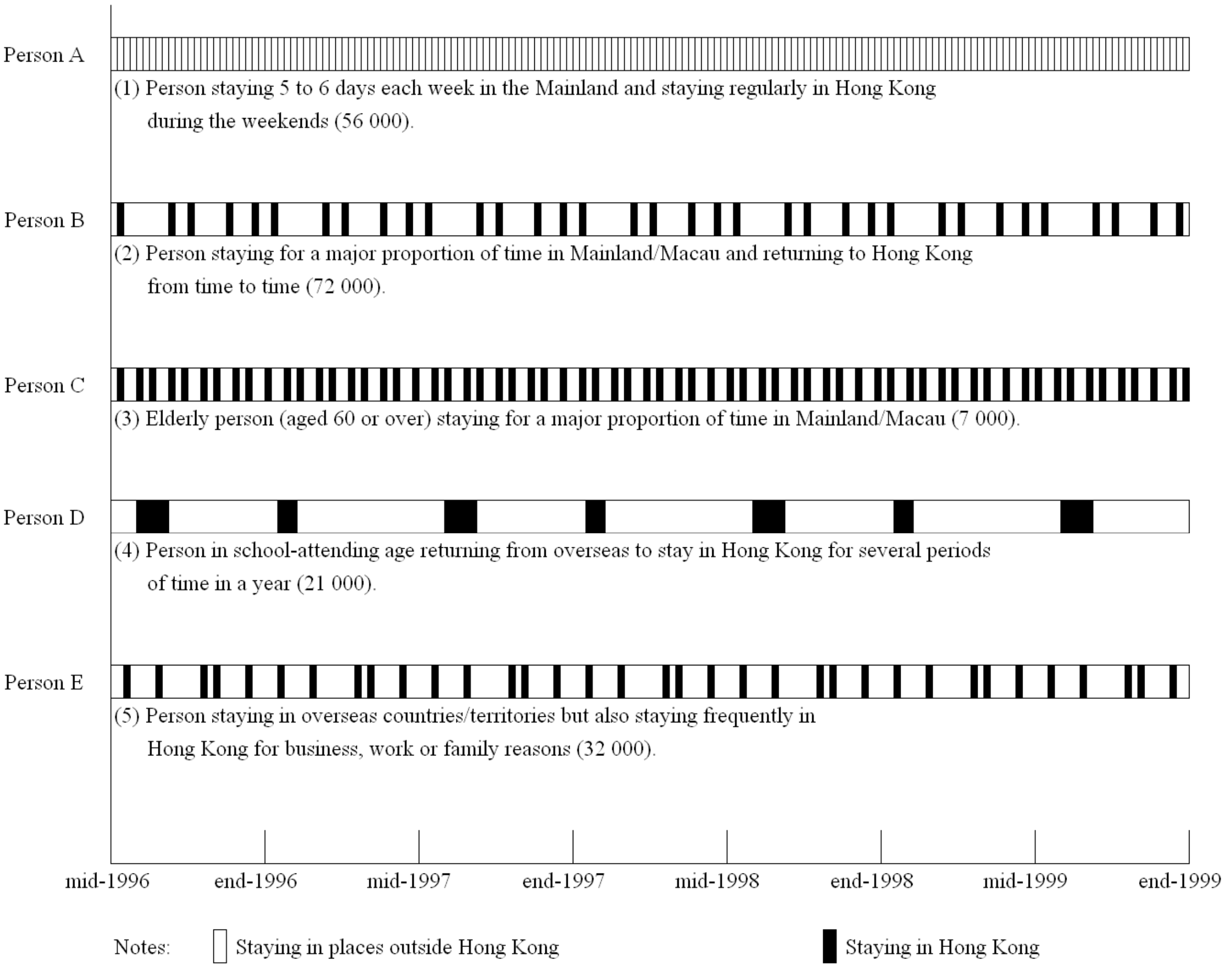
Chart 2 and Table 4 present data series in respect of population estimates and population growth rates for 1990–2000 for ease of reference.

Concluding remarks

The degree of difficulty in compiling population estimates differs among countries/territories. In comparison with other places, statistics on birth and death are easily obtained in Hong Kong. However, it is not so easy with regard to residency and mobility. Since a resident is not required to state the time and purpose of staying in or being away from Hong Kong at arrival or departure, and the traffic flow is massive (over a hundred million movements a year), data handling is of a rather complex nature and detailed study and analysis are required to enable reliable figures to be compiled.

Persons interested in the subject can contact the Census & Demographic Statistics Branch of C&SD at 7/F, Kai Tak Multi-storey Carpark Building, 2 Concorde Road, Kowloon, Hong Kong.
(Telephone no. : 2716 8002,
E-mail : demo_1@censtatd.gov.hk.)

Chart 1 Examples of residency/mobility profile of Mobile Residents



Figures in brackets refer to the estimated number of persons of the respective categories as at end-1999

Table 1 Comparison of "Extended de facto Population" and "Hong Kong Resident Population" figures

Reference time-point	Extended de facto Population (Previously published estimates) (a)	Visitors (b)	Extended de facto Population (Excluding visitors) (c) = (a) – (b)	Hong Kong Resident Population		
				Total population (d) = (e) + (f)	*Among them:* Usual Residents (e)	Mobile Residents (f)
Mid-1996	6 311 000	*80 600*	*6 230 400*	6 484 300	*6 291 600*	*192 700*
End-1996	6 421 300	*121 600*	*6 299 700*	6 530 100	*6 336 000*	*194 100*
Mid-1997	6 502 100	*49 700*	*6 452 400*	6 564 200	*6 377 900*	*186 300*
End-1997	6 617 100	*105 100*	*6 512 000*	6 607 100	*6 415 100*	*192 000*
Mid-1998	6 687 200	*72 200*	*6 615 000*	6 645 600	*6 463 700*	*181 900*
End-1998	6 805 600	*119 700*	*6 685 900*	6 689 000	*6 500 700*	*188 300*
Mid-1999	6 843 000	*97 100*	*6 745 900*	6 720 700	*6 541 500*	*179 200*
End-1999	6 974 800	*141 400*	*6 833 400*	6 761 600	*6 573 400*	*188 200*

Table 2 Hong Kong Population by age and sex under the old and the new methods, end-1999

Age group	Extended de facto Population (Old method) (%)			HK Resident Population (New method) (%)		
	Male	Female	Total	Male	Female	Total
0–4	2.4	2.2	4.6	2.5	2.3	4.8
5–9	3.1	2.8	5.9	3.2	2.9	6.1
10–14	3.1	2.9	6.0	3.3	3.2	6.5
15–19	3.3	3.1	6.4	3.5	3.3	6.8
20–24	3.3	3.6	6.9	3.4	3.2	6.6
25–29	3.8	4.1	7.9	3.5	4.1	7.6
30–34	4.3	4.7	9.0	4.0	5.0	9.0
35–39	5.3	5.4	10.7	4.8	5.6	10.4
40–44	5.1	4.8	9.9	4.7	5.0	9.7
45–49	4.1	3.8	7.9	3.9	4.0	7.9
50–54	3.3	2.8	6.1	3.0	2.8	5.8
55–59	2.1	1.7	3.8	2.1	1.8	3.9
60–64	2.1	1.8	3.9	2.0	1.8	3.8
65–69	1.9	1.8	3.7	1.9	1.9	3.8
70–74	1.4	1.6	3.0	1.4	1.6	3.0
75–79	0.9	1.1	2.1	0.9	1.2	2.1
80–84	0.5	0.7	1.2	0.5	0.7	1.2
85+	0.3	0.6	0.9	0.3	0.6	0.9
Total	**50.3**	**49.7**	**100.0**	**48.9**	**51.1**	**100.0**

Table 3 Population estimates by components

Reference time-point	HK Resident Population	Usual Residents	Mobile Residents	From the previous reference time-point to this reference time-point			
				Births	Deaths	Natural increase#	Net Movement (Inflow *less* outflow)
Mid-1996	6 484 300	6 291 600	192 700	-	-	-	-
Mid-1997	6 564 200	6 377 900	186 300	62 700	30 600	32 100	47 900
Mid-1998	6 645 600	6 463 700	181 900	55 200	32 000	23 200	58 100
Mid-1999	6 720 700	6 541 500	179 200	52 100	32 300	19 800	55 300
Mid-2000*	6 782 100	6 593 400	188 700	51 400	33 800	17 600	43 800

Notes: # Natural increase = Births – Deaths
* Provisional figures

Chart 2 Hong Kong Population Series, 1990–2000

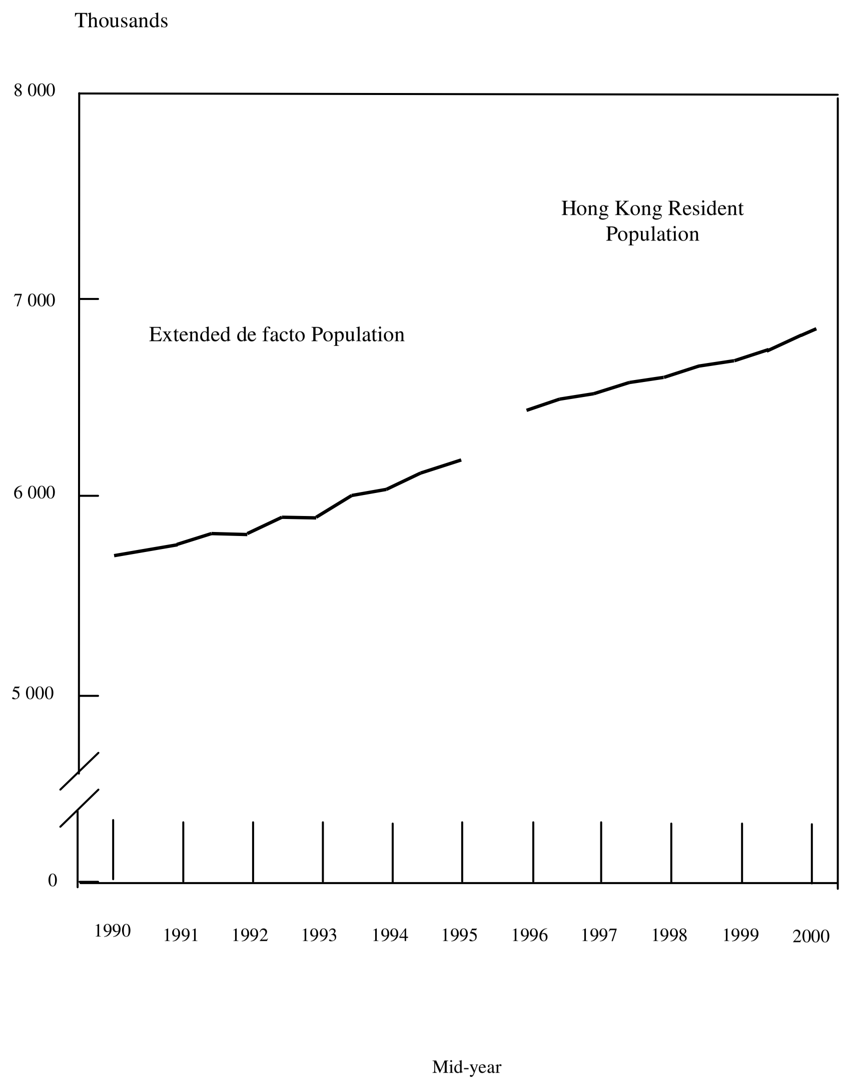


Table 4 Population growth since 1990

Reference time-point	Population estimate	Growth rate compared with the previous reference time-point
Mid-1989	5 686 200	-
Mid-1990	5 704 500	0.3%
Mid-1991	5 752 000	0.8%
Mid-1992	5 800 500	0.8%
Mid-1993	5 901 000	1.7%
Mid-1994	6 035 400	2.3%
Mid-1995	6 156 100	2.0%
Mid-1996	6 484 300#	
Mid-1997	6 564 200	1.2%
Mid-1998	6 645 600	1.2%
Mid-1999	6 720 700	1.1%
Mid-2000*	6 782 100	0.9%

Notes: * Provisional figures
Figure for mid-1996 is based on the new method and is not strictly comparable with the mid-1995 figure

On Using a Combined Weight Matrix in Covariance Structure Analysis

Shu-fai Cheung and Wai Chan
Department of Psychology
The Chinese University of Hong Kong

Introduction

Covariance structure analysis (CSA) has been widely used in social and behavioral sciences. Although it is technically sophisticated, its basic idea is indeed very general. Suppose we have a set of *p* variables and the covariance matrix of these *p* variables, Σ, is hypothesized as a function of a smaller number of more basic parameters, θ. The basic task in CSA is (1) to estimate the structural parameters θ, and (2) to evaluate the goodness of fit of the hypothesized model, $\Sigma = \Sigma(\theta)$.

When the distribution of the variables is multivariate normal, normal-theory (NT) methods such as maximum likelihood (ML) or generalized least squares (GLS) can be used to achieve the above two goals in CSA. Unfortunately, most social and behavioral data are not multivariate normal, so the NT methods can yield distorted results (see, e.g., Hu, Bentler, & Kano, 1992). In addition, Curran, West, and Finch (1996) showed that the performance of the ML method will be even worse when the variables are increasingly deviated from multivariate normality. Alternatively, Browne (1984) proposed the asymptotically distribution free (ADF) method for CSA which does not require the assumption of multivariate normality. While the ADF method is correct asymptotically, it can be extremely misleading when the method is applied to models with small to medium sample sizes (Muthen & Kaplan, 1985, 1992; Hu et al., 1992).

Technically speaking, the difference between the NT methods and the ADF method depends on the weight matrix they use. The NT methods compute a NT weight matrix (W_{NT}), which is a critical component for parameter estimation and model testing in CSA. Similarly, the ADF method computes an ADF weight matrix (W_{ADF}), which is influential in determining the performance of the method (e.g., Chan, Yung, & Bentler, 1995; Yung and Bentler, 1994).

It should be noticed that the NT

methods assume a jointly normal distribution for all the p variables, while the ADF method relaxes such an assumption completely. It is possible that in social sciences, we have p variables (x_1 , x_2, …, x_p) in which p_1 of them (x_1, …, x_{p_1}) are jointly normal and p_2 = p-p_1 of them (x_{p_1+1} , …, x_p) are not. Hence, using the NT methods under this situation will lead to an incorrect distribution assumption on x_{p_1+1} , …, x_p . On the other hand, using the ADF method will lead to waste of information because the knowledge about the distribution of x_1, …, x_{p_1} has not been used.

In this study, two combined estimation methods are proposed, namely Combined-A and Combined-B. First consider the Combined-A method. Following the above example, let x = $[x_1', x_2']$' be the $p \times 1$ vector which contains a normally distributed subvector x_1 ($p_1 \times 1$) and an arbitrarily distributed subvector x_2 ($p_2 \times 1$). When computing the elements of the weight matrix, the NT formulae will be used if the variables involved are all from x_1. If one or more variables are from x_2, then the ADF formulae will be used. As a result, the proposed method will depend on an optimal weight matrix,

$$W_A = \begin{bmatrix} W_{NT} & \\ W_{ADF1} & W_{ADF2} \end{bmatrix} \qquad (1)$$

where W_{NT} is the NT weight matrix with respect to x_1, W_{ADF1} is the ADF weight matrix between x_1 and x_2, and W_{ADF2} is the ADF weight matrix with respect to x_2. For example, assuming that $x_1' = (a_1, a_2)$ and $x_2' = (a_3, a_4)$. The re-arranged weight matrix has the following form:

$$W_A = \left[\begin{array}{ccc|cccc|ccc}
a_{1111} & & & & & & & & & \\
a_{1121} & a_{2121} & & & & & & & & \\
a_{1122} & a_{2122} & a_{2222} & & & & & & & \\
\hline
a_{1131} & a_{2131} & a_{2231} & a_{3131} & & & & & & \\
a_{1132} & a_{2132} & a_{2232} & a_{3132} & a_{3232} & & & & & \\
a_{1141} & a_{2141} & a_{2241} & a_{3141} & a_{3241} & a_{4141} & & & & \\
a_{1142} & a_{2142} & a_{2242} & a_{3142} & a_{3242} & a_{4142} & a_{4242} & & & \\
\hline
a_{1133} & a_{2133} & a_{2233} & a_{3133} & a_{3233} & a_{4133} & a_{4233} & a_{3333} & & \\
a_{1143} & a_{2143} & a_{2243} & a_{3143} & a_{3243} & a_{4143} & a_{4243} & a_{3343} & a_{4343} & \\
a_{1144} & a_{2144} & a_{2244} & a_{3144} & a_{3244} & a_{4144} & a_{4244} & a_{3344} & a_{4344} & a_{4444}
\end{array}\right] \qquad (2)$$

The upper left partition, which contains only variables from the normally distributed subvector x_1 , is computed using the NT theory. The lower right partition, which contains only variables from the arbitrarily distributed subvector x_2, is computed using the ADF theory. All other elements, which involve at least one variable from x_2, are also computed using the ADF theory.

The weight matrix of the Combined-B method is based on W_A. For elements involving only one variable form x_2, the ADF formula can be simplified. Assuming that only the i-th variable is nonnormal, the ADF weight matrix element is given by

$$a_{ijkl} = \frac{1}{N} \sum \left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right)$$

$$= \frac{1}{N} \left[\sum x_i \left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right) \right.$$

$$\left. - \bar{x}_i \sum \left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right) \right]$$

Because the *j*-th, *k*-th and *l*-th variables are normally distributed, the term $\sum \left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right)$ is zero at the population level. Deleting this term, the weight matrix element is given by

$$a_{ijkl} = \frac{1}{N} \sum x_i \left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right) \quad (3)$$

Following the above 4-variable example, the re-arranged weight matrix of the Combined-B method is of the form:

$$W_B = \left[\begin{array}{ccc|cccc|ccc}
a_{1111} & & & & & & & & & \\
a_{1121} & a_{2121} & & & & & & & & \\
a_{1122} & a_{2122} & a_{2222} & & & & & & & \\
\hline
\underline{a_{1131}} & \underline{a_{2131}} & \underline{a_{2231}} & a_{3131} & & & & & & \\
\underline{a_{1132}} & \underline{a_{2132}} & \underline{a_{2232}} & a_{3132} & a_{3232} & & & & & \\
\underline{a_{1141}} & \underline{a_{2141}} & \underline{a_{2241}} & a_{3141} & a_{3241} & a_{4141} & & & & \\
\underline{a_{1142}} & \underline{a_{2142}} & \underline{a_{2242}} & a_{3142} & a_{3242} & a_{4142} & a_{4242} & & & \\
\hline
a_{1133} & a_{2133} & a_{2233} & a_{3133} & a_{3233} & a_{4133} & a_{4233} & a_{3333} & & \\
a_{1143} & a_{2143} & a_{2243} & a_{3143} & a_{3243} & a_{4143} & a_{4243} & a_{3343} & a_{4343} & \\
a_{1144} & a_{2144} & a_{2244} & a_{3144} & a_{3244} & a_{4144} & a_{4244} & a_{3344} & a_{4344} & a_{4444}
\end{array}\right] \quad (4)$$

The only difference between W_A and W_B is the partition involving exactly one variable from the arbitrarily distributed subvector x_2. The underlined elements are computed by equation (3).

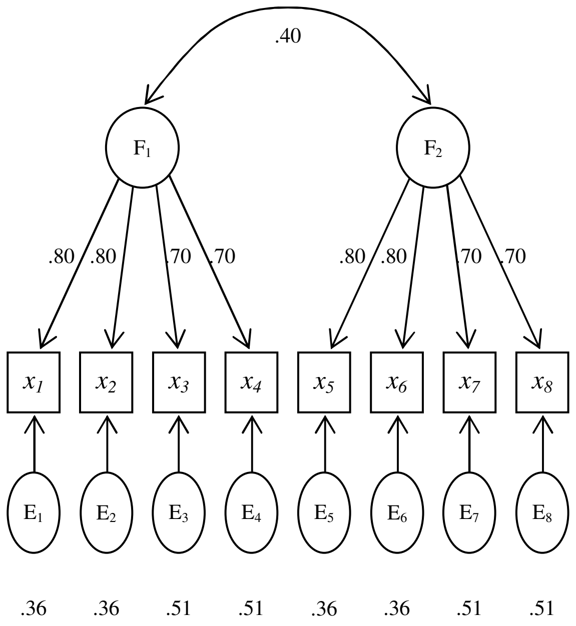
As far as we know, the Combined-A and Combined-B methods have never been considered in CSA. Hence, a Monte Carlo study was conducted to empirically investigate the performance of these two methods. It is believed that if part of the observed variables (say, x_1) are jointly normal, then the two proposed methods should behave better than the traditional ADF method because they have incorporated the distribution information of x_1 during the computation of the weight matrix. On the other hand, if part of the observed variables (say, x_2) are highly non-normal, then our methods should be at least as good as the NT methods because it does not wrongly assume normality of x_2 as the NT methods do.

Method

The Monte Carlo study was based on a two-factor confirmatory factor analysis (CFA) model with eight indicators (Figure 1). Each factor had four indicators loading on it, with loadings equal to .80, .80, .70, and .70. The two factors had a correlation of .40. The error variance was specified in a way such that the population variance of each indicator was equal to 1.00.

Three conditions of non-normality were studied. In the first condition, two (25%) of the indicators, x_7 and x_8, had non-normal error components. In the other two conditions, the numbers of non-normal indicators were four (50%, x_5 to x_8) and six (75%, x_2 to x_8), respectively. In this study, the non-normal errors were generated using the lognormal distribution. In

Figure 1. The CFA model used to generate the raw data



addition to the proportion of non-normal indicators, three different sample sizes were examined: 200, 500, and 1000. Consequently, the total number of conditions was nine (3 × 3).

For each condition, SAS/IML® (SAS Institute Inc., 1989) was used to generate 500 replications. For each replication, the correct two-factor model was fitted using four different methods. They were, namely, NT-GLS (normal theory generalized least square), ADF, Combined-A, and Combined-B. For the Combined-A and Combined-B method, the weight matrices W_A and W_B as depicted in equations (2) and (4) were used. Furthermore, it should be emphasized that for both Combined-A and Combined-B methods, we need to have correct information about which subset of variables is jointly normal and which subset is not in order to compute the weight matrix accurately. This undoubtedly requires a more systematic investigation about the distributions of the observed variables in real practice.

Results and Discussion

Table 1 shows the percentage of replications that successfully converged. It is noted that for a sample size of 200, the percentages of converged replications were exceptionally small for Combined-A and Combined-B methods when 25% of the variables are non-normal (24.2% and 24.4% respectively), as compared to the two traditional methods GLS and ADF. The major cause of nonconvergence was due to the occurrence of a non-positive definite weight matrix. This was probably due to mixing elements from different weight matrices that were based on different formulas and distribution assumptions. All subsequent analyses were based on successfully converged replications.

Table 1. Percentage of converged solutions out of 500 replications

Sample size	Proportions of nonnormal variables	GLS	ADF	Combined-A	Combined-B
200	25	100.0	98.4	24.2	24.4
	50	100.0	97.2	94.8	94.8
	75	100.0	95.4	72.8	71.4
500	25	100.0	100.0	65.8	65.2
	50	100.0	99.6	99.6	99.6
	75	100.0	98.8	91.2	90.0
1000	25	100.0	100.0	93.0	93.0
	50	100.0	100.0	100.0	100.0
	75	100.0	100.0	98.2	98.0

Table 2. Empirical rejection rates (in %) of Chi-square test based on 500 replications at $\alpha = .05$.

Sample size	Proportions of nonnormal variables	GLS	ADF	Combined-A	Combined-B
200	25	3.0	11.4	6.6	4.1
	50	3.8	10.7	9.7	9.7
	75	5.0	9.4	7.1	6.4
500	25	4.2	7.2	6.1	6.1
	50	4.0	7.0	6.6	6.8
	75	4.8	5.5	5.3	5.3
1000	25	4.6	6.6	6.7	6.7
	50	5.2	5.6	5.2	5.2
	75	5.6	4.2	4.1	4.1

Table 2 shows the rejection rate based on the traditional chi-square test, using a significance level of .05. In general, the rejection rate under the GLS method was close to the expected nominal level of 5%. When the sample size was 200, the ADF method over-rejected the true model seriously (9.4% to 11.4%). For larger sample sizes, the ADF method had a rejection rate close to 5%. Generally speaking, the ADF results reported here were consistent with previous research findings (e.g., Hu et al., 1992). The two proposed methods, Combined-A and Combined-B, had performance similar to the ADF method in most situations.

Next we examined the point estimates and standard errors of the model parameters. Since the pattern was similar across the various parameters, we arbitrarily report the results for the factor loading of the eighth variable (the path from F_2 to x_2 in Figure 1). Table 3 shows the means of the estimated parameter under different model conditions. The true value of the parameter was 0.70. The estimates of all four methods were in general close to the true value. Based on the Monte Carlo results, it was also found that the accuracy of the parameter estimates was not related to the proportion of non-normal variables at all.

Table 3. Mean of estimated parameter (true value = .70) out of 500 replications

Sample size	Proportions of nonnormal variables	GLS	ADF	Combined-A	Combined-B
200	25	0.695	0.695	0.689	0.690
	50	0.696	0.694	0.694	0.694
	75	0.696	0.695	0.698	0.698
500	25	0.697	0.697	0.695	0.695
	50	0.697	0.696	0.696	0.696
	75	0.697	0.696	0.695	0.695
1000	25	0.699	0.698	0.698	0.698
	50	0.698	0.698	0.699	0.698
	75	0.698	0.698	0.699	0.698

Table 4 shows the standard deviations (SD) of the parameter estimates in different conditions by different methods, as well as the mean estimated standard error (SE). First, the GLS estimates had the largest SD as compared to the other three methods in most conditions, except when the sample size is 1000. The SD of the GLS estimates is not related to the proportion of non-normal variables. Second, for sample sizes of 200 and 500, the robust SE of the GLS method was smaller than the SD consistently, suggesting that the standard errors of the GLS estimates were likely to be incorrect except for large sample size. Third, the other three methods had nearly identical SD and mean estimated SE. The mean SE was also smaller than its corresponding SD for these three methods. Moreover, the SD of the three methods have negligible association with the proportion of non-normal variables. Last, as expected, the SD and the mean SE decreased as the sample size increased in all conditions.

Table 4. Standard deviation of the parameter estimate and the mean estimated standard error (in bold and italic) out of 500 replications

Sample size	Proportions of nonnormal variables	GLS	ADF	Combined-A	Combined-B
200	25	0.068	0.062	0.067	0.069
		0.062	***0.053***	***0.054***	***0.054***
	50	0.068	0.060	0.059	0.059
		0.062	***0.052***	***0.052***	***0.052***
	75	0.067	0.061	0.060	0.060
		0.062	***0.052***	***0.052***	***0.052***
500	25	0.042	0.040	0.043	0.043
		0.040	***0.037***	***0.037***	***0.037***
	50	0.042	0.039	0.039	0.039
		0.040	***0.036***	***0.036***	***0.036***
	75	0.042	0.039	0.040	0.039
		0.040	***0.036***	***0.036***	***0.036***
1000	25	0.029	0.028	0.028	0.028
		0.029	***0.028***	***0.027***	***0.027***
	50	0.029	0.028	0.028	0.028
		0.029	***0.027***	***0.027***	***0.027***
	75	0.029	0.028	0.028	0.028
		0.029	***0.027***	***0.027***	***0.027***

Note: For the GLS method, the robust standard errors is displayed.

In conclusion, the proposed method of using a combined weight matrix failed to perform satisfactorily. Simulation results suggested that the proposed method behaved similarly to the ADF method in terms of consistency and efficiency of estimation, as well as rejection rate. More importantly, the proposed method suffered seriously from the problem of nonpositive definite weight matrix when the sample size was small (200) and the proportion of non-normal variable was either small (25%) or large (75%), as shown in Table 1. Hence, further effort is needed for developing other optimal methods that can outperform existing methods when some non-normal variables exist in a set of variables.

References

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 127-141.

Chan, W., Yung, Y.-F., & Bentler, P. M. (1995). A note on using an unbiased weight matrix in the ADF test statistic. *Multivariate Behavioral Research*, *30*, 453-459.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.

SAS Institute, Inc. (1989). *SAS/IML software: Usage and reference, Version 6, first edition*. Cary, NC: Author.

Yung, Y.-F., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, *47*, 63-84.

Acknowledgement

This project was supported by the Endowment Fund Research Grant from the United College at the Chinese University of Hong Kong.

News

1. Appointment

Department of Statistics, The Chinese University of Hong Kong

Professor Ngai Hang Chan from Carnegie Mellon University has joined the department as Chair Professor of Statistics and Director of the Risk Management Science Program.

Department of Management Sciences, The City University of Hong Kong

Dr. William Chung from the University of Waterloo has joined the department as an assistant professor.

Dr. Y.M. Lam from Queen's University at Kingston has joined the department as a visiting fellow.