Probabilistic population forecasting models: an overview

O. Almalik
oalmalik@hotmail.com

Population Research Centre, University of Groningen.
Groningen, 2006.

Abstract

This report tries to establish the relationship between the probabilistic models of population projection and the Bootstrap. The research question was 'Can the approximation of the probabilistic population projection models be viewed as an application of the Bootstrap? If yes, which simplifications or adjustments are made?'. The research objective was to get insight into whether the approximation of the probabilistic population projection models could be viewed as an application of the Bootstrap, and if that is the case which simplifications or adjustments were made. To answer the research question the probabilistic models of population forecasting were thoroughly described as explained in the literature. Four categories of these models were discussed; Ex-post analysis, aggregate time series, expert judgment, and stochastic modelling. RESULTS: one probabilistic model of population projection was found to implement the Bootstrap directly. Three probabilistic models of population projection were found to apply methods that bear similarity to the Bootstrap, making major simplifications in the process. CONCLUSIONS: there is very limited application of the Bootstrap in the probabilistic models of population forecasting. There are, however, possibilities for implementing the Bootstrap in the existing and future probabilistic models of population forecasting.

Table of contents

Glossary of symbols

$P(t)$	Population at time t
θ	Parameter
$\hat{\theta}$	Estimator
$\hat{\theta}(m)$	Bootstrap replication of $\hat{\theta}$
M	Number of Bootstrap samples/number of simulations
$\hat{F}$	Empirical distribution
$\hat{SE}(\hat{\theta})$	Bootstrap standard error
l	Last age group
$P(j,t)$	Population age j at time t
$s(j)$	Probability of survival from age j to age $j+1$
$f(j)$	Fertility rate age j
JO	Jump off
$W(t)$	Total population at time t
$m(j,t)$	Age-specific mortality rate
NI	Net immigration
$q(t)$	Age-structure
λ	Growth rate/eigenvalue
$n(j,t)$	Net maternity function
$B(t)$	Number of births
$\Psi_{Jo}(0)$	Covariance matrix of jump-off populations
$\Psi(t,u)$	Covariance matrix of populations at time t and time u
$\Psi_{ft}(t,u)$	Covariance matrix of fertility rates at time t and time u
$\Psi_{sv}(t,u)$	Covariance matrix of survival rates at time t and time u
$\sigma(i,j,t,u)$	Elements of $\Psi(t,u)$
$\sigma_{ft}(i,j,t,u)$	Elements of $\Psi_{ft}(t,u)$
$\sigma_{sv}(i,j,t,u)$	Elements of $\Psi_{sv}(t,u)$
ε_{JO}	Error term jump-off population
ε_{sv}	Error term survival rate
ε_{ft}	Error term fertility rate
ε_{mg}	Error term migration rate

1 Introduction

1.1 Introduction

Estimates of future population are called population forecasts or projections. (The words forecast and projection will be used synonymously in this paper. The use of the word projection, however, emphasizes the uncertainty of the assumptions made for the calculations (Alho and Spencer 1985).) Population forecasts are needed for many purposes including assessment of future environmental pressures, planning for manpower, education, and pension systems (Lee and Tuljapurkar 1994).

Traditionally, population forecasts are done using the cohort-component approach as described in Alho (1997). First, separate forecasts for the age-specific vital rates (fertility, mortality, and migration) are made. Second, letting $P(t)$ be the population at time t, these rates are applied successively to the jump-off population (the last time when data are available) through the bookkeeping equation

$$P(t+1) = P(t) + \text{births} - \text{deaths} + \text{in migration} - \text{out migration}$$

The method is called cohort-component because the calculations are made by birth cohort. According to Alho (1997) Whelpton specified target values for each vital rate at some future year, and used smooth curves to connect the jump-off values to the targets. This model can be expressed in a matrix form (see section 3.2).

The traditional way of handling uncertainty in population forecasts is the variants method dating back to Whelpton (Alho 1997). Whelpton dealt with the uncertainty of population forecasts by preparing three alternate forecast variants for mortality and fertility to cover future paths of development. This was done by conducting empirical analysis of cause-specific mortality and studying past trends in fertility in the United States and elsewhere. This procedure was complemented by detailed reasoning about the likelihood of future changes in the vital rates patterns. Whelpton's method is still the base of most official population forecasts nowadays (Alho 1997). Each variant is prepared using a scenario; a set of assumptions containing an assumed trajectory for fertility, another for mortality, and a third one for migration (Lee 1998). The United States Bureau of the Census applies the variants method by developing low (high) forecasts for each population subgroup by assuming that fertility, life expectancy, and immigration are low (high) (Alho and Spencer 1991).

There are many problems associated with the variants method (see e.g. Ahlburg and Lutz 1998, Alho and Spencer 1991, Lee 1998, Lee 1992, Lee 2003 for detailed discussion of the shortcomings), but here a selection is made. First, the choice of variants is arbitrary. Second, no probabilities can be attached to the high-low range of the projections. Third, the assumption of perfect correlation between forecasting errors of fertility and those of mortality (always high, medium, or low), along with the assumption that errors in fertility and mortality are perfectly correlated with one another are troublesome. Fourth, traditional methods are unable to convey uncertainty to dependency ratios; high fertility results in high denominator and low mortality in high numerator giving a too narrow interval. Fifth, the given variants' coverage is probabilistically inconsistent; e.g. a 95 percent probability for population size can not cover 95 percent probability for fertility since different fertility trajectories could lead to the same population size. Finally, the high-low trajectories ignore the fact that the demographic rates fluctuate or reverse trend (This fluctuation can be integrated in scenarios, however.).

These shortcomings of the variants methods led to attempts to develop probabilistic models for population projection (attaching probability distributions to population forecasts). According to Alho and Spencer (1991) and Alho (1997) the first probabilistic contribution to the assessment of error in demographic forecasts was made by Tornqvist in 1949. Since then there have many contributions to assess uncertainty in population forecasts (for an overview see e.g. Land 1986, Lee and Tuljapurkar 1994, Lee 1998, and Tuljapurkar et al. 2004). The probabilistic models for population forecasts can be divided into four categories (Lee 1998); ex-post evaluation, analysing amount of variability in past demographic time series, expert judgment, and stochastic modelling. This report tries to establish the link between the probabilistic models for population forecasting and the Bootstrap. The Bootstrap is a computer based statistical technique developed by Bradley Efron in 1979 to estimate the standard error of an estimate. It differs from the conventional methods in that no assumptions about the underlying distribution are needed. It is, however, an ad hoc method applicable only to the data set in question.

1.2 Research question and Research objective

This section states the research question and the research objective. The research question is 'Can the approximation of the probabilistic population projection models be viewed as an application of the Bootstrap? If yes, which simplifications or adjustments are made?'. The research objective is to get insight into whether the approximation of the probabilistic population projection models can be viewed as an application of the Bootstrap, and if that is the case which simplifications or adjustments are made.

1.3 Report outline

The report goes as follows. Chapter two discusses the Bootstrap. Data and methods are discussed in chapter three. Chapter four includes the results and findings, and chapter five the conclusions.

2 Theoretical Background

2.1 Introduction to the Bootstrap

The Bootstrap is one of the re-sampling techniques. These are procedures that resample from the original data set. Other re-sampling techniques are the Jackknife, permutation methods, and cross validation.
The Bootstrap has different applications. One application is the estimation of standard error. Another is making confidence intervals. The Bootstrap is also applied in regression analysis. Another area of application is time series analysis. The Bootstrap can also be used in density estimation.

2.2 The Bootstrap estimate of standard error and Bias

One of the main applications of the Bootstrap is the estimation of standard error of an estimate of a parameter. Given a sample of size n from independent identically distributed (i.i.d.) random variables, the Bootstrap uses the empirical distribution $\hat{F}$ to assess an estimate $\hat{\theta}$ of the population parameter θ. The empirical distribution $\hat{F}$ assigns probability $\frac{1}{n}$ to each of the n outcomes of the sample in question.
To introduce the symbols used consider the following example. An example of an estimate for a parameter is the sample mean $\overline{X}$ for the population mean μ, so that $\hat{\theta}$ is $\overline{X}$ and θ is μ. First, the sample mean is calculated from the original sample. Second, M independent Bootstrap samples, b_1, b_2, ..., b_M, each of size n, are drawn with replacement from the original sample. Calculation of the sample mean from each Bootstrap sample m produces the Bootstrap replication of $\hat{\theta}$, given by $\hat{\theta}(m)$; in this example $\hat{\theta}(m)$ is the mean of Bootstrap sample m. The standard deviation of the Bootstrap replications $\hat{SE}(\hat{\theta})$ is used as an estimate for the standard error of the original estimate $\hat{\theta}$, the mean of the original sample. The standard error of the Bootstrap replicate $\hat{\theta}(m)$ is denoted by $\hat{SE}(\hat{\theta}(m))$, which in this example will be given by $\frac{s(m)}{\sqrt{n}}$, where $s(m)$ is the standard deviation of the Bootstrap sample m.
To estimate the standard error the following algorithm is applied. First, M independent Bootstrap samples, b_1, b_2, ..., b_M, each of size n, are drawn with replacement from the original sample. Second, for each Bootstrap sample b_m, $m = 1,...,M$, the Bootstrap replication of $\hat{\theta}$ is calculated. Finally, the standard deviation of the M replications of $\hat{\theta}$ is used as an estimate for the standard error of $\hat{\theta}$. In other words, the standard error is estimated by

$$\hat{SE}(\hat{\theta}) = \sum_{m=1}^{M} \left(\hat{\theta}(m) - \bar{\theta}(.) \right)^2 / (M-1),$$

with $\hat{\theta}(m)$ referring to the calculation of the estimate $\hat{\theta}$ from the Bootstrap sample m, and $\bar{\theta}(.) = \sum_{m=1}^{M} \hat{\theta}(m) / M$.

The Bootstrap can also be used to estimate the bias (The bias is defined as the difference between the parameter and the expected value of the estimate.). The Bootstrap estimate of the bias is given by

$$\text{Bootstrap Bias} = \bar{\theta}(.) - \hat{\theta}$$

The estimated bias can be used to correct $\hat{\theta}$. The corrected (less biased) $\hat{\theta}$ is given by:

$$\hat{\theta}_{corrected} = \hat{\theta} - \text{Bootstrap Bias}$$

Efron and Tibshirani (1993) suggest rules of thumb based on experience to determine the number of Bootstrap samples needed. A good estimate of the standard error of the estimate is achieved by using 50 Bootstrap samples. In seldom cases more than 200 Bootstrap samples are needed.

2.3 The Bootstrap confidence intervals

The Bootstrap can be used to create confidence intervals by creating Bootstrap tables. These tables are used instead of conventional tables (e.g. normal distribution table, t-distribution table). The Bootstrap tables relieve the user from making assumptions about the underlying distribution. They are, unlike conventional tables however, only applicable to the given sample.
Three types of Bootstrap confidence intervals will be considered here. First, the Bootstrap-t interval will be described. Second, the percentile based Bootstrap intervals will be discussed. Third, the Bootstrap BC_a interval will be viewed.

2.3.1 Bootstrap-t interval

A generalization of the usual Student's t method is the Bootstrap-t procedure. It is particularly applicable to location statistics like the sample mean, but can not be trusted for more general problems.
The Bootstrap-t interval is created as follows. First, from the M Bootstrap samples b_1, ..., b_M, $m = 1,..., M$, the statistic $o(m)$ is calculated by

$$o(m) = \left(\hat{\theta}(m) - \hat{\theta} \right) / SE\left(\hat{\theta}(m) \right),$$

with $SE\left(\hat{\theta}(m)\right)$ is an estimated standard error of $\hat{\theta}(m)$ for the Bootstrap sample m ($SE\left(\hat{\theta}(m)\right)$ is the standard error of each Bootstrap estimate $\hat{\theta}(m)$. For example, if the estimate is the sample mean, then $\hat{\theta}(m)$ will be the mean calculated from each Bootstrap sample m. For each Bootstrap sample $SE\left(\hat{\theta}(m)\right)$ will be given by $\dfrac{s(m)}{\sqrt{n}}$, where $s(m)$ is the standard deviation of the Bootstrap sample m.). Second, $\hat{t}(\alpha)$ is determined so that

$$\#\left\{o(m) \le \hat{t}(\alpha)\right\}/\mathrm{M} = \alpha,$$

i.e. $\hat{t}(\alpha)$ is determined so that is the number of $o(m)$'s less than $\hat{t}(\alpha)$ divided by the number of Bootstrap samples M equals α.

Finally, the Bootstrap-t confidence interval is given by

$$\left\{\hat{\theta} - \hat{t}(1-\alpha).\hat{SE}(\hat{\theta}), \hat{\theta} + \hat{t}(\alpha).\hat{SE}(\hat{\theta})\right\},$$

with $\hat{SE}(\hat{\theta})$ an estimate of the standard error of $\hat{\theta}$, different from $\hat{SE}(\hat{\theta}(m))$.

2.3.2 The Bootstrap percentile interval

A second way of producing intervals is the Bootstrap percentile interval. Intervals produced using this method have the advantage of being transformation-respecting and range-preserving.

The Bootstrap percentile interval is calculated in the following way. First, given Bootstrap samples b_1, ..., b_M, $m = 1,...,\mathrm{M}$, the estimates $\hat{\theta}(m)$ are calculated from each Bootstrap sample. Second, the cumulative distribution function, $\hat{\mathrm{H}}$, of $\hat{\theta}(m)$ is determined. Finally, the $1 - 2\alpha$ percentile interval is defined by the α and the $1 - \alpha$ percentiles of $\hat{\mathrm{H}}$

$$(\hat{\theta}_{low}, \hat{\theta}_{high}) = \left\{\mathrm{H}^{-1}(\alpha), \hat{\mathrm{H}}^{-1}(1-\alpha)\right\},$$

with $\mathrm{H}^{-1}(\alpha) = \hat{\theta}^{\alpha}(m)$, the $100.\alpha$ percentile of the Bootstrap distribution of the $\hat{\theta}(m)$'s.

2.3.3 The BC$_a$ interval

Another Bootstrap interval is the Bootstrap BC_a interval. The Bootstrap BC_a is an improved version of the Bootstrap percentile interval, and has the advantage of being transformation respecting and accurate and is recommended for general use. The Bootstrap BC_a interval stands

for bias corrected and accelerated. It depends on two numbers; a (acceleration) and b (bias correction). The BC_a interval of $1-2\alpha$ coverage is given by

$$\mathrm{BC}_a : (\hat{\theta}_{low}, \hat{\theta}_{high}) = \left\{ \hat{\theta}^{\delta_1}(m), \hat{\theta}^{\delta_2}(m) \right\},$$

where

$$\delta_1 = \Phi\{b + (b + z_{score}(\alpha))/(1 - a(b + z_{score}(\alpha)))\},$$

and

$$\delta_2 = \Phi\{b + (b + z_{score}(1-\alpha))/(1 - a(b + z_{score}(1-\alpha)))\},$$

with $\Phi(.)$ is the standard normal cumulative distribution function, and $z_{score}(\alpha)$ is the $100.\alpha^{th}$ percentile point of a standard normal distribution, e.g. $\Phi(1.96) = 95\%$ and $z_{score}(95\%) = 1.96$. The acceleration a is calculated using the Jackknife method. The Jackknife method involves computing the estimate, $\hat{\theta}_i$, from the original sample with the ith value deleted. Then the acceleration a is given by

$$a = \sum_{i=1}^{n} (\hat{\theta}_{(.)} - \hat{\theta}_i)^3 / \sigma \left\{ \sum_{i=1}^{n} (\hat{\theta}_{(.)} - \hat{\theta}_i)^2 \right\}^{3/2},$$

with $\hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_i / n$. The quantity a is called the acceleration because it refers to the rate of

change of the standard error of $\hat{\theta}$ with respect to the true parameter value θ (Efron and Tibshirani 1993 acknowledge that it is not clear how the formula of a provides an estimate of the acceleration of the standard error, and refer to a paper written by Efron for discussion.).
The bias correction number, b, is obtained from the proportion of Bootstrap replications less than the original estimate. In other words, b is given by

$$b = \Phi^{-1}\left\{ \#(\hat{\theta}(m) < \hat{\theta})/\mathrm{M} \right\},$$

with $m = 1,...,\mathrm{M}$, and Φ^{-1} is the inverse function of a standard normal cumulative distribution function, e.g. $\Phi^{-1}(95\%) = 1.645$.

2.4 Bootstrap types

In this section three Bootstrap types are described. First, the parametric Bootstrap is discussed. Second, the Bayesian Bootstrap is described. This is followed by a discussion of the smoothed Bootstrap.

A substitute for the non-parametric Bootstrap methods discussed so far is the parametric Bootstrap. Instead of re-sampling from the empirical distribution, the population is assumed to have a parametric distribution. From the assumed parametric density estimate M Samples of size n are then drawn. From each of the M samples the Bootstrap replication $\hat{\theta}(m)$ of $\hat{\theta}$ is calculated, and the standard deviation of the M replications of $\hat{\theta}$ is used as an estimate for the standard error of $\hat{\theta}$.

The Bootstrap can be used after applying Bayesian methods. First, a sample $U_{[1]},...,U_{[n-1]}$ of size $n-1$ is drawn from the Uniform distribution $U(0,1)$. Second, the sample is organized in an ascending order. Third, $U_{[0]}$ and $U_{[n]}$ are defined by $U_{[0]} = 0$ and $U_{[n]} = 1$. Having the ordered values $U_{[0]},U_{[1]},...,U_{[n-1]},U_{[n]}$, the statistic v_i, $i = 1,...,n$, is defined by $v_i = U_{[i]} - U_{[i-1]}$. In other words, v_i is the difference between every two successive values of the ordered values $U_{[0]},U_{[1]},...,U_{[n-1]},U_{[n]}$. The last step is to select n observations with replacement from the original sample $x_1,...,x_n$, with x_i having probability v_i of being selected instead of $1/n$, i.e. x_1 is selected with probability v_1, x_1 is selected with probability v_2, en so on. A second set of Bayesian Bootstrap replication is created in the same way but using a new set of $n-1$ uniform random numbers and therefore, new v_i's. This process is repeated M times and M Bootstrap samples of size n are created.

The last Bootstrap type to be discussed here is the smoothed Bootstrap. This method replaces the empirical distribution, $\hat{F}$, by a smooth distribution based on a Kernel density estimate of F. Then re-sampling is done from the smoothed estimate.

2.5 Bootstrap applications

2.5.1 Bootstrap applications in regression analysis

Sections 2.5.1 and 2.5.2 discuss two applications of the Bootstrap. In section 2.5.1 the applications of the Bootstrap in regression analysis are described. The Bootstrap applications in time series analysis are viewed in section 2.5.2.

The Bootstrap can be applied in regression analysis in two ways. First, the Bootstrap can be applied to regression residuals. Second, the pair of dependent and independent variables can be bootstrapped.

Bootstrapping residuals is done in the following way. First, the regression model $y_i = x_i \hat{\beta} + \varepsilon_i$, $i = 1,...,n$, is fitted using Least Squares method or the Likelihood method.

Second, the residuals, $e_i = \hat{y}_i - x_i \overset{\wedge}{\beta}$, are calculated. Third, a random sample of size n is selected, with replacement, from $e_1,...,e_n$. Fourth, the sample from residuals, $\overset{*}{e}_1,...,\overset{*}{e}_n$, is used to generate a new response variable, $\overset{*}{y}_i = x_i \overset{\wedge}{\beta} + \overset{*}{e}_i$. Next, the procedure is repeated M times, creating M Bootstrap samples from the residuals and M newly created response variables, $\overset{*}{y}_i(m),\ i = 1,...,n,\ m = 1,...,M$. Finally, the Bootstrap Least Square estimate of β, $\overset{*}{\beta}$, is the one that minimizes the residual square error for the Bootstrap data. In other words, $\overset{*}{\beta}$ is given by:

$$\overset{*}{\beta} = \min_{\overset{*}{\beta}(m)} \sum_{i=1}^{n} (\overset{*}{y}_i(m) - x_i \overset{*}{\beta}(m))^2,$$

where

$$\overset{*}{\beta}(m) = \sum_{i=1}^{n} x_i \overset{*}{y}_i(m) / \sum_{i=1}^{n} x_i^2,\ m = 1,...,M.$$

Instead of bootstrapping the residuals, the pairs x_i and y_i can be bootstrapped. First, M samples of size n are drawn, with replacement, from $x_1,...,x_n$ and M random samples of size n are drawn, with replacement from $y_1,...,y_n$. (Random samples of size n of integers are drawn, with replacement, from $1,...,n$. The integers drawn will be the index of values chosen from the original sample. In other words, if integer j is chosen, the Bootstrap i^{th} observation $x_i(m)$ will be x_j.) Second, M bootstrapped pairs $\left\{ \overset{*}{x}_i(m), \overset{*}{y}_i(m) \right\},\ i = 1,...,n,\ m = 1,...,M$, are created, where $\overset{*}{x}_1(m),...,\overset{*}{x}_n(m)$ is the mth Bootstrap sample from $x_1,...,x_n$, and $\overset{*}{y}_1(m),...,\overset{*}{y}_n(m)$ is the mth bootstrapped sample from $y_1,...,y_n$. Next, the regression coefficient $\overset{*}{\beta}(m)$ for each two Bootstrap samples is calculated. Finally, the Bootstrap estimate of β is given by

$$\overset{*}{\beta} = \min_{\overset{*}{\beta}(m)} \sum_{i=1}^{n} (\overset{*}{y}_i(m) - \overset{*}{x}_i(m) \overset{*}{\beta}(m))^2$$

2.5.2 Bootstrap methods in time series analysis

Two methods of applying the Bootstrap in time series are considered in this section. First, the method of bootstrapping residuals is described. Second, the moving blocks Bootstrap method is discussed.

A procedure similar to bootstrapping residuals in regression can be used in time series analysis (see Appendix B section B.1 for description of time series analysis). The first step is to estimate β using either the Least Squares method or the Maximum Likelihood method. Next, residuals of

the model are calculated. The following step is to bootstrap residuals as was done in section 2.5.1, and using the residuals to create a new response variable. Then from each bootstrapped time series a replication of $\overset{\wedge}{\beta}$ is calculated using either Least Squares or Maximum Likelihood procedures. Finally, the standard error of the Bootstrap estimates of β is calculated.

The second Bootstrap method to be discussed here is the moving blocks method. First, a block length is chosen. Second, all contiguous blocks of the specified length are considered. Third, a sample with replacement from each block is drawn, and pasted together to form a time series. Finally, the Bootstrap replication of $\overset{\wedge}{\beta}$ is calculated for the Bootstrap sample. This procedure is repeated M times, and the standard error of the Bootstrap estimates of β is calculated.

3 Data and methods

3.1 Data sources

Data used in this paper were obtained from different sources (see References). The information on the Bootstrap was obtained from Efron and Tibshirani (1993) and Chernick (1999). The data on probabilistic models of population forecasting were mostly obtained from scholarly journals. Sources for time series analysis were Hamilton (1994) and Johnston and Dinardo (1997). Medhi (1994) and Ross (2000) were the sources for the branching Galton-Watson process.

3.2 The deterministic model and the Leslie matrix

In this section the deterministic model and the Leslie matrix are discussed. Given an age distribution of a population on a certain date, the deterministic model calculates an age distribution of the survivors and descendants of the original population at successive intervals of time. The individuals are supposed to be subjected to the same age-specific fertility and mortality rates.

Considering only the female population, and using constant age-specific fertility and mortality rates the deterministic model can be described as follows (Leslie 1945). The model can be expressed in $l+1$ linear equations, where l is the highest age considered in the life table distribution. Leslie defines $P(j,t)$ as the number of females alive in age group j to $j+1$ at time t, $s(j)$ as the probability that a female aged j to $j+1$ at time t will be alive in the age group $j+1$ to $j+2$ at time $t+1$ (survival probability), and $b(j)$ as the number of daughters born in the interval t to $t+1$ per female alive aged j to $j+1$ at time t, who will be alive in the age group $0-1$ at time $t+1$ (birth rate). Then the age distribution at the end of one unit's interval will be given by

$$\sum_{j=0}^{l} b(j)P(j,0) = P(0,1)$$
$$s(0)P(0,0) = P(1,1)$$
$$s(1)P(1,0) = P(2,1)$$
$$\vdots$$
$$s(l-1)P(l-1,0) = P(l,1)$$

In the highest age group $P(l,1)$ the age l can be as small as 85 years and as large as 110 years (Alho and Spencer 1991). In the former case $s(l)$ can be introduced to the calculation of $P(l,1)$ as the survival probability in the last age group (In this case the last equation will be given by: $s(l-1)P(l-1,0) + s(l)P(l,0) = P(l,1)$.). In the latter case $s(l)$ can be assumed to be zero, so that there is no survival in the last age group.

In matrix notation this can be expressed as $AP(0) = P(1)$, where $P(0)$ and $P(1)$ are column vectors giving the age distribution at $t=0$ and $t=1$ respectively, and the matrix A is given by

$$A = \begin{pmatrix} b(0) & . & . & . & b(l) \\ s(0) & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & s(l-1) & . \end{pmatrix}$$

The matrix A is a $(l+1) \times (l+1)$ matrix and is called the Leslie matrix. All elements are zero's, except those in the first row and in the sub diagonal below the principal diagonal. The $s(j)$ figures lie between zero and one, and the $b(j)$ figures are by definition positive quantities, with some possibly zero. Since $P(1) = AP(0)$, $P(2) = AP(1) = A^2 P(0)$, then $P(t) = A^t P(0)$. The $b(j)$'s and the $s(j)$'s figures are calculated using life table functions (see Keyfitz 1968b).

3.3 Ex-post analysis

The first category of probabilistic models of population projection to be discussed here is Ex-post analysis. This method transforms past projections and their realizations into confidence intervals for future population projections.

The method was pioneered in Keyfitz (1981). The author compares past forecasts with subsequent population performance. More specifically, he looks back to estimates for the 1970 population made in 1950, and compares actual and projected populations for 1970. This comparison provides a unit of information for the 1980-2000 population given that the same method of forecasting is used, and there are similar fluctuations of population in the past and the future.

Keyfitz's method can be described as follows. First, World countries are divided into three groups of 30 countries each; slow growing countries with annual growth rate of up to 1.8 percent, medium growing countries with annual growth rate between 1.85 and 2.6 percent, and fast growing countries with annual growth rate above 2.6. Three jump-off years are used; 1958, 1963 and 1968. Projection durations of 5, 10, 15 and 20 are used for 1958; 5, 10 and 15 for 1963; and 5 and 10 for 1968.

The next step is calculating error terms. For the jump-off year 1958, Keyfitz calculates for the group of slow growing countries, for the projection durations 5, 10, 15 and 20, the root mean square error (RMS), defined by:

$$RMS = \sqrt{\sum_{i=1}^{30} e^2_i / 30} \, ,$$

with e_i is the difference between the projected and realized population for country i, $i = 1,...,30$. The procedure is repeated for the group of medium growing countries, and that of the fast growing countries. The same method is applied to the three groups of countries for the jump-off year of 1963, for the projection durations of 5, 10 and 15 years. Finally, the procedure is repeated for the jump-off year 1968, for the projection durations of 5 and 10 years.

The author's next step is to combine all jump-off years and projection durations. The RMS is calculated for the slow growing countries. The same is done for the medium and fast growing countries. Keyfitz concluded with a RMS of 0.29 for slow growing countries, 0.48 for medium

growing countries, and 0.6 for fast growing countries. The author constructed 67 percent confidence intervals for the United States growth rate and the United States population in the year 2000.

Finally, the author calculated RMS for the world as a whole. After combining all countries in one category and considering all projection durations, Keyfitz introduced the one figure summary for the whole world, RMS of 0.48. The figure was rounded to 0.4, recognizing the declining trend in RMS from 1958 to 1968.

Another attempt to use observed errors in past projections was made by Stoto (1983). The author defines a statistic that takes into consideration the duration of projection and the population size. Then, the distribution of the statistic is studied in order to make a statement about future projection errors, i.e. confidence intervals.

Stoto's method can be described as follows. First, The growth rate, $\lambda(t)$, is assumed to be a function of time. Second, an exponential population growth is also assumed. Then, with $P(0)$ the population at time zero, the population at time t can be expressed as:

$$P(t) = P(0) \exp \int_0^t \lambda(t) dt,$$

and the average growth rate over the projection period is given by

$$\overline{\lambda} = 1/t \int_0^t \lambda(t) dt$$

This leads to the expression of $P(t)$ and $\overline{\lambda}$ as

$$P(t) = P(0) \exp(t\overline{\lambda})$$

and

$$\overline{\lambda} = \{\log(P(t)/P(0))\}/t$$

Defining $\hat{P}(t)$ as the projected population at time t, and $\hat{P}(0)$ as the estimated population at time zero (the true population at time zero might not be known), Stoto defines the following measure of error, e, as:

$$e = 100(\overline{\lambda}_{projected} - \overline{\lambda}_{actual}) = 100/t \log\{(\hat{P}(t)P(0)/\hat{P}(0)P(t)\}$$

Using US population projections made by the US Census Bureau, in jump-off years 1945, 1950, 1955, 1960, 1965, 1970 for target years 1950, 1955, 1960, 1965, 1970, 1975, Stoto studied the distribution of e. The author introduced dummy variables for the jump-off year, target year, and projection duration and conducted regression analysis of e to measure the contribution of each factor to the overall variability. Based on the analysis of variance it was found that the jump-off year effects were significant for developed and developing regions, but that the target year and duration effects were not significant. Consequently, it is concluded that the error is at least composed of two parts; a part depending on the jump-off year, and a random part.

In addition, using United Nations projections, Stoto studied the distribution of the jump-off error and the random term. The United Nations divides the world into 24 regions. Stoto introduces the

error measure, e_{ijk}, for each region i, $i = 1,...,24$, and each jump-off year j, $j = 1950, 1955, 1960, 1965$, and each duration k, $k = 5, 10, 15, 20$. The author defines the jump-off error, $\bar{e}_{ij}$, as the mean of e_{ijk}, for a specific region i, for the specific jump-off year j, over all durations k. Consequently, the random term will be given by:

$$\varepsilon_{ijk} = e_{ijk} - \bar{e}_{ij}$$

Stoto studied the distributions of both the jump-off error and the random term. The distribution of the jump-off error remained stable for the developed countries, while for the developing countries early projections were biased. The distribution of the random error was symmetric around zero, and had a roughly normal shape except for occasional outliers.

Next, the author estimates the error variance. He estimates the variance of e_{ijk} by:

$$\mathrm{var}(e_{ijk}) = \mathrm{var}(\bar{e}_{ij}) + \mathrm{var}(\varepsilon_{ijk}),$$

based on United Nations data for developed regions. (Another estimate of the standard deviation based on the United States data is referred to as the pessimistic estimate in Stoto (1983).) Finally, Stoto 's 95 percent confidence interval is constructed as follows:

$$\left\{ \mathrm{P}_{low}(t), \mathrm{P}_{high}(t) \right\} = \{ \mathrm{P}(0) \exp(T(\bar{\lambda}_{projected} - (1.96\sqrt{\mathrm{var}(e_{ijk})}))), \mathrm{P}(0) \exp(T(\bar{\lambda}_{projected} +$$
$$(1.96\sqrt{\mathrm{var}(e_{ijk})})))\}$$

Observed errors in past projections were also used by Keilman (1998). The author compared projected numbers for total population size, crude birth rate, crude death rate, and age structure in five-year groups for the period 1950-1990 with corresponding Ex-post observed numbers. The measure used was the percentage error, (PE), defined by:

$$PE = \left(\frac{Forecast - observation}{observation} \right) * 100$$

For results that are not size dependent, Keilman used mean error, with error defined as the forecast minus the observed value. Keilman's analysis was based on the United Nation medium variant. However, no projections were made using the measures calculated.

The methods in this sections use past forecasts and their realizations. Keyfitz (1981) uses the difference between the projected and realised populations, while Stoto (1983) uses the difference between the projected and realised growth rates. Keyfitz uses the root mean squared error (RMS), while Stoto calculates the variance of error observed after concluding that the error is composed of two parts. The authors then calculate confidence intervals for projected populations based on the calculated RMS and standard deviation.

3.4 Analysing past demographic time series

This category of population projection is concerned with the total population size. Age structure details are not dealt with. Two contributions will be discussed, that of Cohen (1986) and that of Pflaumer (1992).

In Cohen (1986) the author constructs confidence intervals using four different methods. The first two methods are model based, both using $\log(\lambda)$ and the variance derived by Heyde and Cohen (1985) (see section 3.6.1). The other two methods are empirical methods based on the distribution of error of forecasts. The forecasts for all four methods are done using methods described in Heyde and Cohen (1985).

The author starts with some definitions. Let $t = 1$ be the point in time of the earliest observed population size, $t = T$ the point in time of the last observed population size, and $t = \iota$ the point in time the projection is supposed to describe. Defining the total population at time t, $W(t)$, $W(t) = \log(e', P(t))$, with e' a vector of ones, the forecasts are given by:

$$W(\iota) = W(T) + (\iota - T)/(T - 1)\big(W(T) - W(1)\big)$$

Cohen bases his results on three assumptions. First, the projection matrices $A(t)$ form a stochastically stationary ergodic sequence that is uniformly mixing. Stochastically stationary means that the joint probability distribution of any finite number of the matrices A is invariant with respect to shifts in time. Furthermore, the sequence $A(t)$ is ergodic if for every event defined in terms of a finite number of matrices A, the average frequency of the event converges almost surely to the probability of the event. Uniformly mixing means that the vital rates in the matrix $A(t_1)$ at time t_1 approach independence of the rates in the matrix $A(t_2)$ at time t_2, as the times t_1 and t_2 become farther apart. The sequence of the projection matrices is uniformly mixing if the matrices are i.i.d. or they are finite in number and determined by an ergodic Markov Chain of arbitrary finite order. Second, the author assumes that there is a uniform upper bound on the largest vital rates that occur in the matrices, so that for an integer d any product of matrices has all of its elements positive with probability one. Finally, it is assumed that there is a constant, greater than one, such that ratio of the largest element in A to the smallest positive element in A is less than or equal to the constant, with probability one. (To check these assumptions, Cohen suggests that the log of the population size be plotted against time over a long span. If the assumptions are plausible, then the population size should fluctuate around a straight line. The absence of trends is consistent with stationarity.)

The first method for constructing a confidence interval can be described as follows. With sd is standard deviation (calculated by methods in Heyde and Cohen (1985)), the confidence interval is given by:

$$W(T) + (\iota - T)((W(T) - W(1))/(T - 1) \pm sd \min_{0 < q < \alpha}\{(\iota - T)(T - 1)^{1/2} z_{(\alpha - q)/[2(1 - q)]} +$$
$$(\iota - T)^{1/2} z_{q/2}\}$$

To calculate q, Cohen considers two events. Event 1 is that $|\log(\hat{\lambda}) - \log(\lambda)|$ and $\left|W(\iota) - W(T) - (\iota - T)\log(\hat{\lambda})\right|$ are bounded by multiples of the standard deviation. More specifically,

$$Event \ 1 = \left\{ \left| \log(\hat{\lambda}) - \log(\lambda) \right| < sd(T-1)^{1/2} z_{p/2} \right\} \ and$$

$$\left| W(\iota) - W(T) - (\iota - T)\log(\hat{\lambda}) \right| < sd(\iota - T)^{1/2} z_{q/2}$$

Event 2 is that $W(\iota)$ lies in the following interval

$$Event \ 2 = \left\{ W(\iota) \right\} \ is \ in \ the \ interval$$

$$W(T) + (\iota - T)(T-1)^{-1} \left(W(T) - W(1) \right) + \left\{ sd((\iota - T)(T-1)^{-1/2} z_{p/2} + sd((\iota - T))^{1/2} z_{q/2} \right\}$$

Cohen, noting that $Event \ 1$ implies $Event \ 2$, states that $\Pr(Event \ 2) \geq \Pr(Event \ 1)$, and one has asymptotically

$$\Pr(Event \ 1) \approx (1-p)(1-q), \tag{3.4.1}$$

because of normality and mixing. The author then sets $1 - \alpha = (1-p)(1-q)$, and solves for the value of p determined by each choice of q. Then the value of q that minimizes the width of the interval in $Event \ 2$ is chosen.

Cohen then moves to the second method of constructing confidence intervals. The author introduces the interval:

$$W(T) + (\iota - T)(T-1)^{-1} \left(W(T) - W(1) \right) \pm sd \left\{ (\iota - T)^2 (T-1)^{-1} + (\iota - T) \right\}^{1/2} z_{\alpha/2}$$

The term next to sd is derived in the following way. Let Y_1 and Y_2 be independent standard normal variables. An equivalent formulation of (3.4.1) is

$$W(\iota) \approx W(T) - (\iota - T) * \log(\lambda) + (\iota - T)(T-1)^{-1/2} sd * Y_1 + (\iota - T)^{1/2} sd * Y_2 \tag{3.4.2}$$

The term $(\iota - T)(T-1)^{-1/2} sd * Y_1$ represents the variation in $W(\iota)$ contributed by the possible deviation between $\log(\lambda)$ and $\log(\hat{\lambda})$, while the term $s^{1/2} sd * Y_2$ represents the variation in $W(\iota)$ contributed by the increment to population size between the launch and the target. Asymptotically these two terms are independent due to the mixing hypothesis. For every real numbers a_1 and a_2, and a standard normal variable Y, $a_1 Y_1 + a_2 Y_2$ has the same distribution as $(a_1^2 + a_2^2)^{1/2} Y$. Hence (3.4.2).

The third method of constructing confidence intervals considers the distribution of forecast errors. Cohen describes the distribution of signed forecast errors by the standard deviation. The author's justification for using the standard deviation is that the forecast errors, $\left\{ W(\iota) - \hat{W}(\iota) \right\}$, were found to look roughly normal.

The last method is based on Stoto's estimates of the standard deviation (See section 3.3). Cohen's last confidence interval is given by

$$W(T) + (t - T) * \log(\hat{\lambda}) \pm (t - T) * sd_{Stoto},$$

where sd_{Stoto} is Stoto's estimate of the standard deviation for developed regions. The estimate is multiplied by five in order to transform Stoto's time unit of one year to Cohen's time unit of five years.

The second approach to model total population sizes to be discussed here is that of Pflaumer (1992). The author applies ARIMA models to the United States population 1900-1988. The United States population size is determined every ten years, and the data for intervening years had to estimated. Two ARIMA models are proposed; the first one for the population size, the second one for the logarithm of the population size (see Appendix B section B.1 for a description of time series analysis).

The model for the population size can be described as follows. After differencing the series twice for stationarity, a suitable model was found to be an ARIMA (2,2,0) given by:

$$\nabla_2 P(t) = \alpha_0 + \alpha_1 \nabla_2 P(t-1) + \alpha_2 \nabla_2 P(t-2),$$

where

$$\nabla_2 P(t) = \big(P(t) - P(t-1)\big) - \big(P(t-1) - P(t-2)\big)$$

The parameters α_1 and α_2 were found to be significant.

For the logarithm of the population size the author proposes an ARIMA(1,1,0) model. Letting $\nabla \ln(P(t) = \ln(P(t)) - \ln(P(t-1))$, the model is given by:

$$\nabla \ln(P(t) = \beta_0 + \beta_1 \nabla \ln(P(t-1)),$$

where β_0 and β_1 were found to be significant.

Methods discussed in this section ignore the age-structure and deal only with population totals. Cohen (1986) uses methods discussed in Heyde and Cohen (1985) and methods based on forecast errors to construct confidence intervals for projected population totals. Pfaumer (1992), however, applies time series methods to population totals and the logarithm of population totals.

3.5 Expert judgement

The first model that incorporates expert judgement to be discussed here is that of Alho and Spencer (1985). The authors use a mix of statistical modelling and expert judgement. Realizing that error in population forecasts stems from errors in the jump-off population, and errors in prediction of future vital rates, they present different models for the jump-off population and for the vital rates.

The authors start with a model for the jump-off population of females in 1980. Based on earlier demographic analysis, Alho and Spencer recognize a 2.1 percent net undercount for black females, but consider a net overcount of white females apparent citing reasons related to failure to

account for many illegal aliens. The authors define the estimate P_{JO}, where JO stands for jump off or the last time in which data is available, as

$$P_{JO} = 1980 \text{ census figure of females} + (0.021 \times 1980 \text{ census figure of black females})$$

It is assumed that $\log(P_{JO}) \sim N(\mu, \sigma^2)$. The median, $\exp(\mu)$, equals the estimated true number of females. The variance, σ^2, is estimated assuming that the range of interval estimate of female illegal aliens (1.75 million, 3 million) has the length of a confidence interval for P_{JO} with a certain probability. For example, assuming that the range 1.25 millions is the length of a 95% confidence interval for P_{JO}, and using the quantiles of the standard normal distribution, the variance can be calculated from the following equation

$$1.25 = \exp\big(\log(P_{JO}) + 1.96\sigma\big) - \exp\big(\log(P_{JO}) - 1.96\sigma\big)$$

To derive estimates for each age the authors make two assumptions. First, it is assumed that the age-distribution for January 1980 is perfectly known. Second, the estimates for each age interval are perfectly correlated with variances σ^2.
Next the authors move to model fertility. They predict age-specific fertility, through application of approximately linear models. Alho and Spencer define the stochastic process $\{f(t) \mid t \in Z\}$, the logarithms of the age-specific fertility rates, as

$$f(t) = R(t) + \Gamma(t),$$

where $E(\Gamma(t)) = 0$, and $R(t)$ are assumed to be given by

$$R(t) = \sum_{i=1}^{k} \theta_i R_i + \Delta(t),$$

where $R_1(t), ..., R_k(t)$ are known functions, $\theta_1, ..., \theta_k$ are unknown parameters, and $\Delta(t)$ represents the bias of the ideal mean $R(t) - \Delta(t)$. (Alho and Spencer consider six age groups, i.e. $k = 6$.) The authors assume that the bias $\Delta(t)$ is independently generated of $\Gamma(t)$, and that it satisfies $E\big|\Delta\Delta'\big| \leq \sigma^2 \Psi$, where Ψ is some symmetric non-negative definite matrix given by

$$\Psi = \begin{bmatrix} \Psi_0 & \Psi_2 \\ \Psi'_2 & \Psi_1 \end{bmatrix}$$

Alho and Spencer use a diagonal Ψ with the ith diagonal element given by

$$\Psi(i,i) = 0.31 \mid i - T \mid$$

(For methods of getting the value 0.31 the authors refer to Alho's unpublished dissertation.)
The authors consider next the error process $\Gamma(t)$. The error process is assumed to be auto-regressive of order one. The AR(1) parameter is determined from the residuals of a Least Squares

regression of past data (see Appendix B section B.1.1). The auto-correlation is averaged for the six age groups by 0.71. The cross correlation between age groups r and s is given by $0.88^{|r-s|}$, *for* $r, s = 1,..., k$. The covariance of $\Gamma(t)$ is given by $\sigma^2\Sigma$, with Σ a known positive definite matrix given by

$$\Sigma = \begin{bmatrix} \Sigma_0 & \Sigma_2 \\ \Sigma'_2 & \Sigma_1 \end{bmatrix}$$

Armed with these assumptions, the authors proceed to calculate an estimate of logarithms of the age-specific fertility rates, $f(t)$. Assuming that $f(t)$ at $t = 1,..., T$ is observed, the authors define the vectors $f_0 = (f(1),..., f(T))$, $\Delta_0 = (\Delta(1),..., \Delta(T))$, and $\Gamma_0 = (\Gamma(1),...\Gamma(T))$. For predictions for the periods $t = T+1,..., T+s$, the following vectors are defined $f_1 = (f(T+1),..., f(\iota))$, $\Delta_1 = (\Delta(T+1),..., \Delta(\iota))$, and $\Gamma_1 = (\Gamma(T+1),...\Gamma(\iota))$. Furthermore, the design matrix X is defined by

$$X = \begin{bmatrix} (R_1(1) & . & . & . & R_k(T) \\ & . & . & & \\ . & & & . & \\ . & & & . & \\ R_k(1) & & & & R_k(T) \end{bmatrix}$$

This indicates that the model mentioned above can be formulated as

$$f_0 = X\Theta + \Delta_0 + \Gamma_0$$

To predict f_1 the authors use the linear Minimax mean squared error prediction[1] $\Lambda' f_0$ where

$$\Lambda = \Omega^{-1}X(X'\Omega_0 X)^{-1}(N - X'\Omega_0^{-1}\Omega_2) + \Omega_0^{-1}\Omega_2,$$

$\Omega_i = \Sigma_i + \Psi_i$, $(i = 0,2)$, and

$$N = \begin{bmatrix} (R_1(T+1) & . & . & . & R_k(\iota) \\ & . & . & & \\ . & & & . & \\ . & & & . & \\ R_k(T+1) & & & & R_k(\iota) \end{bmatrix}$$

---

[1] For details on the calculation of the linear Minimax mean squared error prediction, Alho and Spencer refer to Alho's unpublished PH.d dissertation.

For the calculation of the fertility rate Alho and Spencer incorporate Expert judgement. The authors give the United States Bureau of Census (USBC)'s forecasts a weight of 0.8. The Minimax estimate is given a weight of 0.2.

The authors' next step is to predict mortality. Alho and Spencer approximate the survival rate, $s(j,t)$, by $\exp(-m(j,t))$, where $m(j,t)$ is the age-specific mortality rate in age group j during year t. The variance of $\log(s(j,t))$ is approximated by $\mathrm{var}(m(j,t))$. To calculate $\mathrm{var}(m(j,t))$ a linear trend was extracted from the $m(j,t)$ series, and the residual variance was calculated.

The next model to be discussed is that of Pflaumer (1988). Letting $NI(t)$ represents the vector of net immigrants, the author uses the general representation

$$P(t+1) = AP(t) + NI(t)$$

The author starts with a set of assumptions. First, the total fertility rate (TFR) is a random variable, and the change in each age-specific rate is proportional to the change in TFR. Correlation exists between two succeeding TFR's, TFR(t) and TFR(t-1), so that TFR(t)-TFR(t-1)$<X(t)$, where $X(t)$ is random variable. Second, life expectancy is a random variable implying that survivorship rates of the matrix A are random. Third, $NI(t)$ is a random variable. For a random variable $Y(t)$ Pflaumer assumes a uniform distribution between the estimated upper and lower bounds with density:

$$f(y(t)) = \begin{cases} 1/2(med(t)-l(t)), & l(t) \leq y(t) < med(t), \\ 1/2(u(t)-med(t)), & med(t) \leq y(t) < u(t) \end{cases},$$

where $l(t)$ is the estimated lower limit, $med(t)$ is estimated median, $u(t)$ is the estimated upper limit, and $y(t)$ is a realization of $Y(t)$. The expectation and variance of $Y(t)$ are given by

$$E(Y(t)) = \frac{l(t)+med(t)+u(t)}{4},$$

$$Var(Y(t) = 1/6\{(l(t)^2)+l(t)med(t)+med(t)^2\}+(u(t)^2-u(t)med(t)+med(t)^2)-[l(t)+2med(t)+u(t)/4]^2$$

Next, the author proceeds with a simulation process. Fertility and mortality rates are found by selecting random numbers from the uniform distribution with specified lower, median and upper bounds. Another random number is selected from the uniform distribution with bounds on the random variable X. The bounds for fertility, mortality, and net immigration distributions were based on the assumptions of the United States Census Bureau (USCB) projections made in 1983. Bounds on $X(t)$ were obtained from past changes of the TFR. If the condition TFR(t)-TFR(t-1)$<X(t)$ is not satisfied, a new random number is selected.

The simulation process is executed in the following way. Using the random survival rates by year and age and sex, the population is carried forward. Each year a new birth cohort is added by applying random age-specific fertility rates to surviving female population. The population is projected to target year t. This procedure is repeated M times, resulting in M different population trajectories starting with year zero and ending with year t. The M observations of the

population at time t are then classified into a frequency distribution to get the expected value, the variance, and the confidence interval.

The simulation process used here is similar to the parametric Bootstrap. The parametric Bootstrap involves assuming a parametric distribution, in this case the uniform distribution. It also involves drawing at random from the assumed distribution, which is also done here. The random draws are used to project the population forward, and the procedure is repeated M times creating M projected population for the target year t. From the created distribution the standard deviation is calculated and confidence intervals are made.

Alho (1990) applies expert judgement to fertility and mortality rates. The author treats vital rates as a realization of random processes, producing high-low intervals with a given probability of covering the true size of an age-sex group in a given future year.

First, the author considers fertility rates. A transformation of the fertility rate in the year $(t, t+1)$, $f'(t)$ is studied. Next, letting L be a loss function, the volatility of $f'(t)$ is defined as

$$ E\big[L\big(f'(t) - f'(t-r) \mid f'(t-s)\big)\big], \ s \le r $$

Alho used the Finnish fertility in 1776-1976. The loss function $L(y) = |y|$ was used. Instead of using direct age-specific fertility rates, Alho uses their average $\bar{f}(t)$ defined by the TFR divided by the number of child-bearing ages (30). Furthermore, the author incorporates expert judgement by using a logit transformation, forcing the average fertility rate between a lower bound, $f_L(t)$, and an upper bound, $f_U(t)$, based on the observed minimum and maximum. With $f_L(t) = 0.025$, and $f_U(t) = 0.25$, $f'(t)$ is defined as

$$ f'(t) = \log\left\{\left(\bar{f}(t) - 0.025\right) \Big/ \left(\big(0.25 - \bar{f}(t)\big)\right)\right\} $$

Using a naive forecast f' for $f'(t)$, the present value, then $e(t) = f'(t) - f'$ is the prediction error, and

$$ \bar{f}(t) = \left\{\big(0.025 + 0.225\exp(f' + e(t))\big) \Big/ \big(1 + \exp(f' + e(t))\big)\right\} $$

The ratio of the number of true births to the point forecast (when $e(t) = 0$) is given by

$$ \frac{\left\{\big(0.025 + 0.225\exp(f' + e(t))\big) \big/ \big(1 + \exp(f' + e(t))\big)\right\}}{\left\{\big(0.025 + 0.225\exp(f')\big) \big/ \big(1 + \exp(f')\big)\right\}} \tag{3.5.1} $$

The empirical distribution of the volatility of $f'(t)$ is now used, after smoothing to reduce random errors, to create confidence intervals. For instance, after determining the 67th percentile of volatility, the $e(t)$ in 3.5.1 is replaced by the $\pm 67^{th}$ percentile to produce a 67% prediction interval for the number of births.

The use of the empirical distribution of the volatility of $f^{'}(t)$ to create confidence intervals for fertility rates resembles the Bootstrap. The resemblance stems from creating percentiles from the original sample. The omission of re-sampling, however, is a major simplification.

The author's next step is to forecast mortality. Alho suggests using the same volatility method used for fertility, using ARIMA models (see Appendix B section B.1) as naïve forecasts, to get error bounds for each age-specific mortality rate. To obtain a baseline forecast Alho used an ARIMA (0,1,1) model. Using jump-off years 1930-1981, 5, 10, and 15 years forecasts were made for both male and female mortality rates. The author used the logit transformation

$$m^{'}(t) = \log\{(m(t) - 0.002)/(0.03 - m(t))\},$$

where $m(t)$ is the age-standardized mortality rate, and 0.002 and 0.03 are the lower limit and the upper limit respectively, based on the minimum and the maximum of the standardized mortality rates. From the volatility of $m^{'}(t)$ error bounds for mortality rates were obtained.

Lee (1993), incorporating expert judgment, applies the Lee-Carter method (see section 3.6.2) to fertility. Define $c(j)$ and $d(j)$ as constants, and $f(t)$ as a time varying fertility index. Letting $f(j,t)$ be the collection of fertility rates for age j and time t, the $f(j,t)$ is given by

$$f(j,t) = c(j) + f(t)d(j) + e(j,t), \qquad (3.5.1)$$

where the $d(j)'s$ are standardized to sum to unity, and the $f(t)'s$ to sum to zero. This way the $c(j)$ will equal the average age-specific values, and c, the sum of $c(j)'s$ over j, will equal the average value of the TFR. Summing both sides of equation 3.5.1 over j results in

$$TFR(t) = c + f(t) + E(t),$$

where $E(t)$ is the sum across age of the $e(j,t)$. ($E(t)$ should be close to zero.)

For the estimation of the TFR Lee proposes two methods. The first method starts by estimating $c(j)$ and $d(j)$ (see section 3.6.2). Then $F(t) = f(t) + c$ is defined. To ensure non-negative fertility rates, Lee specifies lower and upper bounds on the TFR, denoted by Upper and Lower respectively. Then a transformed fertility index, $h(t)$, is defined by

$$h(t) = \ln\big((F(t) - Lower)/(Upper - F(t))\big)$$

This series can be modelled as a times series and forecasted. Next, the forecast of $F(t)$ is obtained by transforming back to $F(t)$ from $h(t)$ using

$$F(t) = \big(Upper.\exp(h(t) + Lower\big)/\big(1 + \exp(h(t))\big)$$

This ensures that the forecast and its probability interval will fall between Upper and Lower. Expert judgement is applied by taking the lower and upper limits to be zero and four respectively. Lee's second proposed method involves a convergence level for the TFR. Letting F^{*} be the level to which the forecast will converge in the long run, Lee defines the process $x(t) = F(t) - F^{*}$. Lee suggests modelling $x(t)$ as an ARMA model with no constant term, forecast ing $x(t)$ over

the desired horizon, and recovering the forecast of $F(t)$ by adding F^* to the forecast of $x(t)$.

Expert judgement is incorporated by taking F^* to be 1.85, which is the average of the USCB assumption of 1.8 and the Social Security Actuary's assumption of 1.9.

Lutz et al. (1996) applied expert judgment to population projection resulting in what is known as the random scenario approach. This method, like the preceding one, is based on a mix of expert judgement and stochastic modelling. Experts are asked to provide both a point forecast as well as high and low levels for fertility, mortality, and migration for 2030-2035. The consensus about ranges in 2030-2035 is thought to cover 90 percent of all future paths of the TFR, life expectancy at birth, and the interregional migration matrix.

To predict fertility rates, the authors proceed as follows. The high-low levels provided by experts are determined to constitute 90 percent confidence intervals for TFR. After assuming a normal distribution for TFR and defining D_k as the difference between the high and low levels at time k, the standard deviation could be estimated by $D_k / 3.29$ (Assuming normality the 90 percent confidence interval is given by: [Mean value $\pm 1.645\sigma$]. Assuming further that the high-low intervals provided by experts constitute 90 percent confidence intervals means that D_k is the difference between the confidence limits of the 90 percent confidence interval. In other words, $D_k = Mean\,value + 1.645\sigma - (Mean\,value - 1.645\sigma) = 3.29\sigma$, and therefore $\sigma = D_k / 3.29$.). Then z_i, the ith draw from a standard normal distribution, is used to create the ith random TFR in the following way

$$TFR_{ik} = u_k + z_i D_k / 3.29 \,,$$

where u_k is the central value for TFR determined by the experts for time k. This was done for the year 2000, 2030-2035, 2080-2085. TFR's at other dates were calculated using linear interpolation between two adjacent dates (see Appendix B section B.3). For the years after 2080-2085 the TFR's are assumed to remain at the 2080-2085 levels.

As for mortality, the authors apply a similar procedure. First, the low, central, and high levels of mortality are converted into values of life expectancy. Second, the same procedure applied to fertility is used to predict life expectancy.

Next, the authors proceed with predicting migration. Experts were asked to provide high values for migration forecasts. The central value was assumed to be half the high value, and low values assumed to be zero. The random migration flow for region r is then given by

$$MG_{rik} = u_r + z_i D_r / 3.29 \,,$$

where u_r is the central migration level for region r, z_i is the ith draw from a standard normal distribution, and D_r is the difference between the high and low migration levels for region r.

Lutz et al. then proceed as follows. The process of random drawing from a standard normal distribution was repeated 1000 times. Based on the random fertility, mortality and migration rates, 1000 independent cohort-component projections were performed for 2030-2035. The resulting 1000 independent simulation for each region resulted in a distribution of the resulting population for all age groups. From this distribution interesting results (e.g. the median, percentage above 60 years, 95 percent confidence intervals) can be drawn. (Lutz et al. (1997) applied the random scenario approach to the World's population.)

The procedure of random drawing from a standard normal distribution is a partial application of the parametric Bootstrap. A parametric distribution is assumed, a standard normal distribution, and random draws are made from it. The random vital rates are then used to project the population forward. This procedure is repeated 1000 times. The resulting distribution of projected populations is then used to create confidence intervals.

To justify the assumption of the 90 percent coverage of the high-low interval, Lutz and Scherbov (1997) carried out sensitivity analysis. Two different distributions of population projection were obtained based on the assumption that the same interval covered 85 percent and 95 percent. Using t-tests, the hypothesis of equal means could not be rejected at any period.

Lutz and Scherbov also use an alternative method by introducing an auto-regressive component to the scenario. The authors introduce the scenario, $w(t)$, for the measure to be forecasted (i.e. TFR, change in life expectancy, or net migration) as

$$w(t) = v(t) + \varepsilon(t),$$

where $v(t)$ is a function that passes through the mean values of the fertility, mortality, and migration rates as defined by the experts, and $\varepsilon(t)$ is given by the auto-regressive process

$$\varepsilon(t) = \alpha \varepsilon(t-1) + \varepsilon'(t),$$

with $\varepsilon'(t) \sim N(0, \sigma^2_{\varepsilon'})$. Assuming that $\alpha = 0.8$ and that $\sigma^2_{w(t)} = \sigma^2(t)$, where $\sigma^2(t)$ is the variance of the variable considered (fertility, mortality or migration) derived according to the 90 percent range given by experts in year t, the authors calculated $\sigma^2_{\varepsilon'}$ using

$$\sigma^2_{w(t)} = \sigma^2_{\varepsilon'} \frac{(1-\alpha^{2t})}{(1-\alpha^2)}$$

Building on the aforementioned work in Lutz et al (1996, 1997) and Lutz and Scherbov (1997), Lutz et al. (2001) introduce a new model for forecasting the vital rates. The $w(t)$'s are given by

$$w(t) = \overline{w}(t) + \varepsilon(t),$$

where $\overline{w}(t)$ are assumed means for the fertility, change in life expectancy, and net migration, and $\varepsilon(t) \sim N(0, \sigma^2_{\varepsilon}(t))$, and the variance of $\varepsilon(t)$, $\sigma^2_{\varepsilon}(t)$, is supposed to depend on the rate considered. Lutz et al. use a moving average representation for $\varepsilon(t)$ given by

$$\varepsilon(t) = \sigma^2_{\varepsilon}(t) / \left(\sqrt{n} \sum_{i=t-n+1}^{t} z_{t-i} \right),$$

where the z_i's are the values of independent draws from a standard normal distribution, and n is the number of periods in the moving average scheme.

Lutz et al. (2001) also introduce a model for correlation. For each forecast year the authors generate correlated random numbers. Given K correlated states (regions or vital rates), a column

vector of the K correlated values, $r(t) = r_1(t),...,r_K(t)$, is generated. It is further assumed that $\sigma_{r_i}(t)$ is one, with $i = 1,...,K$. If Σ is the assumed variance-covariance matrix, then $\Sigma(t) = \Sigma_1(t),...,\Sigma_K(t)$ is calculated from the equation

$$\Sigma(t) = \Xi' r(t),$$

where Ξ is the Cholesky decomposition of Σ (Bay 1999, pp. 523).

Sanderson et al. (2003), using models developed in Lutz et al. (2001), suggest two methods for conditional probabilistic population forecasting. The first method is used to answer what-if kind of questions. The second method makes conditional probabilistic forecasts based on future jump-off years.

The first method can be described as follows. Projections were made for the period 2000-2050 based on methods in Lutz et al. (2001). A distribution for global population in 2005 was made from 2,000 simulated projections. The simulated projections were divided into those made using low fertility assumption (average fertility is less than 1.6), those using medium fertility assumption (average fertility between 1.6 and 1.8), and those using high fertility assumption (average fertility greater than 1.8). Each category was then reclassified based on whether the forecasts were made based on low life expectancy (less than 68 years), average life expectancy (between 68 and 71 years), or high life expectancy (above 71 years). Using these categorizations, the authors were able to answer some what-if kind of questions, e.g. what would be the effect on the world population size in 2050 of high fertility trends against low fertility trends over the coming decades combined with the medium range of uncertainty for future mortality?

The second method involves making population forecasts based on future jump-off years. Projections were first made for the year 2010 based on information in the year 2000. The simulated projections were then divided into two groups of 1000 each, depending on whether the simulated values were above or below the median of the distribution. Next, projections were made up to the year 2100 based on the projected population for the year 2010. For each year the projected values were divided into two groups the L group, projected values based on simulated values in 2010 that were below the median, and the H group, based on simulated values in 2010 that were above the median.

Methods mentioned in this section incorporate expert judgement in different ways. Alho and Spencer (1985) give experts forecasts a weight of 0.8, and the developed Minimax estimate a weight of 0.2. Pflaumer (1988) selects fertility, mortality and net migration rates from a uniform distribution with bounds based on the USCB projections. Using a logit transformation Alho (1990) forces the average fertility rate and the age standardized mortality rate between lower and upper bounds based on observed maximum and minimum. To ensure non-negative fertility rates Lee (1993) uses a fertility index after specifying lower and upper bounds on the TFR. Lee also proposes another method based on a convergence level for fertility determined by expert judgement. Lutz et al. (1996) assume that the high-low intervals provided by experts for fertility, mortality and migration cover 90 percent of all future paths and calculate a standard deviation based on this assumption using the so-called random scenario approach. Lutz et Scherbov (1997) apply the random scenario approach in combination with time series analysis to forecast fertility, mortality and migration rates. The random scenario approach is also used by Sanderson et al. (2003) to develop the conditional probabilistic population projections.

3.6 Stochastic models for forecasting vital rates

In this section stochastic models for forecasting vital rates are considered. First, methods explained in Sykes (1969) are described. Second, Lee's methods of using time series analysis to forecast births are explained. Third, the Lee-Carter method for forecasting mortality is discussed. Fourth, methods developed by Alho and Spencer (1991) and Alho (1992a, 1992b) will be viewed. Finally, Methods discussed in Bell (1997) are considered. The theories of population projection are relegated to Appendix A section A.5.

The first model to be discussed here is mentioned in Sykes (1969). Sykes introduces three stochastic versions of the discrete classical model of population projection. First, Sykes introduces the additive errors model. This model is given by

$$P(t+1) = AP(t) + S(t), \ t = 0,1\ldots, \tag{3.6.1}$$

where $S(t)$ are random vectors. Sykes assumes further that $E(S(t)) = 0$, and $E(S(t)S'(t)) = \Psi_{st}, s,t = 0,1\ldots$. Assuming that Ψ_{st} exists, and substituting repeatedly in equation 3.6.2.1, results in

$$P(t) = A^t P(0) + \sum_{i=1}^{t} A^i S_{t-i-1},$$

with $E(P(t)) = A^t P(0)$, and $var(P(t))$ is given by (see Appendix A section A.1)

$$Var(P(t)) = \sum_{i=1}^{t} \sum_{j=1}^{t} A^i \Psi_{t-i-1,t-j-1} A'^j$$

Sykes' next step is to create confidence intervals for $P(t)$. The author assumes that errors are uncorrelated and that they are normally distributed. In other words, $S(t) \sim N_l(0, \Psi)$. Consequently, $P(t) \sim N\left(A^t P(0), \sum_{i=1}^{t} A^i \Psi A'^i \right)$, since $P(t)$ is a linear combination of normal variates. Exact confidence intervals for $P(t)$ can then be calculated by considering $\left((P(t) - A^t P(0))' \sum_{i=1}^{t} A^i \Psi A'^i (P(t) - A^t P(0)) \right)$, which is distributed as χ_l^2. For the non-homogenous case, where A is a function of time, Sykes replaces A^i with the product of the A's, so that

$$E(P(t)) = \left(\prod_{k=0}^{t-1} A(k) \right) P(0)$$

and

$$Var(P(t)) = \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \left(\prod_{k=t-i}^{t-1} A(k) \right) \Psi_{t-1-i,t-1-j} \left(\prod_{k=t-j}^{t-1} A(k) \right)$$

Sykes' second method is a Branching process formulation. Individuals are assumed to give birth and die with probability $b(j)$ and $1-s(j)$ respectively. Each member of the jth age group at time t has a probability $s(j)$ of surviving to be a member of the $j+1$ group at time $t+1$, and probability $b(j)$ of contributing a member to the first group at time $t+1$. The probability of any two or more events is equal to the product of the probabilities of the respective component events.

Presenting the notation $P(j,t)$ for the jth component of the vector $P(t)$, the model provides for $P(j,t)$ binomial trials with probability of success $s(j)$, $j=1,...,l$, and, independently, for $P(j,t)$ binomial trials with probability of success $b(j)$, $j=1,...,l$. Sykes argues that the assumption of independent binomial trials is equivalent to assuming that the population growth process is a multi-type Branching process (See Appendix B section B.2). Moreover, the author allows an individual of type j to have children of types 1 (first age group), and type $j+1$, $j=1,...,l-1$ (i.e. allowing the individual to reproduce and survive). Individuals of type l are only allowed to have children of type 1. Note that the assumption of independence of individuals within and between types is a requirement of the Branching process.

More specifically, Sykes' Branching process model can be described as follows. The probability of having $P(j,t)$ individuals in group j is given by

$$\Pr[P(1,t),...,P(l,t)] = \Pr[P(t) = (P(1,t),...,P(l,t))],$$

and the P.G.F. (see Appendix B section B.2) is given by

$$G_t(r) = \sum_{P(1,t),...,P(l,t)} \Pr(P(1,t),...,P(l,t)) r_1^{P(1,t)}..r_l^{P(l,t)}$$

Sykes uses a normal approximation to the assumed Binomial distribution. The author defines the variable $Y(t+1)$, $Y(t+1) \sim N(0,I)$, by

$$Y(t+1) = [Var(P(t+1)\,|\,P(t))]^{-\frac{1}{2}}[P(t+1) - E(P(t+1)\,|\,P(t))],$$

where

$Var(P(t+1)\,|\,P(t)) = diag[Var(P(t+1)\,|\,P(t))]$ is a diagonal matrix whose non-zero elements are the component variances. Another variable, $Z(t+1)$, is defined by

$$Z(t+1) = Y'(t+1)Y(t+1) \sim \chi_l^2$$

Sykes then determines the two moments of the process. The expectation is given by (see Appendix A section A.1)

$$E(P(t+1)) = A^t P(0),$$

and the variance is given by

$$\mathrm{var}(\mathrm{P}(t)) = \sum_{i=0}^{t-1} \mathrm{A}^{t-1-i} C_i \mathrm{A}^{'t-1-i},$$

where

$$C_i = \sum_\alpha \sum_\beta a_{i\alpha} a_{i\beta} \left[\mathrm{cov}\left(\mathrm{P}_\alpha(i), \mathrm{P}_\beta(i)\right)\right].$$

These results are carried over to the non-homogenous case giving

$$E(\mathrm{P}(t)) = \prod_{l=0}^{t-1} \mathrm{A}(l)\mathrm{P}(0),$$

and

$$Var(\mathrm{P}(t)) = \sum_{l=0}^{t-1} \left(\prod_{i=t-l}^{t-1} \mathrm{A}^l\right) C(t-1-l) \left(\prod_{i=t-l}^{t-1} \mathrm{A}^l\right)'$$

Sykes' third approach is that of the random transition matrix. The author introduces the model

$$\mathrm{P}(t+1) = (\mathrm{A} + \mathrm{K}(t))\mathrm{P}(t), \; t = 0,1,\dots,$$

where $\mathrm{K}(t)$ is a sequence of independent $l \times l$ matrix random variables satisfying $E(\mathrm{K}(t)) = 0$, and $\mathrm{var}(\mathrm{K}(t)) = \Sigma$, with Σ a singular $l^2 \times l^2$ matrix. To find $E(\mathrm{P}(t+1))$ and $\mathrm{var}(\mathrm{P}(t+1))$ Sykes uses the conditional mean and the conditional variance formulas. The expectation and the variance are given by (see Appendix A section A.1)

$$E(\mathrm{P}(t)) = \mathrm{A}^t \mathrm{P}(0),$$

$$\mathrm{var}(\mathrm{P}(t)) = \sum_{l=0}^{t-1} \mathrm{A}^{t-1-l} \left\{ \sum_{v=1}^{p} \sum_{w=1}^{p} \mathrm{cov}\left(\mathrm{K}_{iw}, \mathrm{K}_{jw}\right)\left(Cov\left(\mathrm{A}_v(t), \mathrm{A}_w(t)\right) + \right.$$
$$\left. E\left(\mathrm{A}_v(t)\right)E\left(\mathrm{A}_w(t)\right)\right)\right\} \mathrm{A}^{'t-1-l}$$

These results carry over to the non-homogeneous case giving

$$E(\mathrm{P}(t)) = \left(\prod_{i=0}^{t-1} \mathrm{A}_i\right)\mathrm{P}(0),$$

and

$$\mathrm{var}(\mathrm{P}(t)) = \sum_{l=0}^{t-1} \left(\prod_{i=0}^{t-1-k} \mathrm{A}_i\right) \left(\sum_{v=1}^{p} \sum_{w=1}^{p} \mathrm{cov}\left(\mathrm{K}_{iw}, \mathrm{K}_{jw}\right)\{Cov\left(\mathrm{A}_v(t), \mathrm{A}_w(t)\right) + \right.$$
$$\left. E\left(\mathrm{A}_v(t)\right)E\left(\mathrm{A}_w(t)\right)\}\left(\prod_{i=0}^{t-1-k} \mathrm{A}_i'\right)\right.$$

Lee (1974) presents time series models for forecasting births to an age-structured population. The author suggests a white noise specification, a first order autoregressive model, a second order autoregressive model, and a random walk specification.

Lee starts with introducing models for the number of Births at time t, $B(t)$, and the net reproduction rate, $NRR(t)$. Defining $n(j,t)$ as the net maternity function (Keyfitz 1968a, pp. 100) for age j at time t, and $Upper$ as the upper limit of the female reproductive age, Lee introduces the following identity for $B(t)$

$$B(t) = \sum_{j=1}^{Upper} n(j,t)B(t-j) \tag{3.6.2}$$

Defining the net reproductive rate as $NRR(t) = \sum_{j=1}^{Upper} n(j,t)$, it is taken to be one based on the assumption of stationarity. Furthermore, the variable $x(j,t)$ is defined by

$$x(j,t) = n(j,t) - n(j),$$

where $n(j)$ is the expected value of $n(j,t)$. In addition, Lee defines $\varepsilon(t)$ as $\varepsilon(t) = NRR(t) - 1$. Letting B be the long run expected number of births, the variable $d(t)$ is defined by

$$d(t) = \frac{B(t) - B}{B}$$

Lee approximates $d(t)$ (see Appendix A section A.2) by the following AR process

$$d(t) = \sum_{j=1}^{Upper} n(j)d(t-j) + \varepsilon(t)$$

This AR process has a moving average representation of the form

$$d(t) = \sum_{i=1}^{\infty} \eta_i \delta(t-i),$$

where the first three terms of η_i's are given by $\eta_0 = 1$, $\eta_1 = \eta_0 n(1)$, $\eta_2 = \eta_0 n(2) + \eta_1 n(1)$.

Lee offers the following interpretation for the η_i. The η_i's correspond to the reproduction of a single birth η_0 after i periods, subject to the constant net maternity function. Because of the stationarity assumption, the η_i's converge to a constant value for large i, given by the $1/J$, where J is the mean age at child bearing.

The author starts with a white noise assumption for the errors. After observing a series of births and NRR's up to time t, the best prediction of birth at time $t+s$ is given by

$$\hat{d}(t+s) = \sum_{j=1}^{Upper} n(j)d(t+s-j),$$

where the values of $d(t+s-j)$ after t are replaced by their predicted value. The prediction error, $e(t+s) = d(t+s) - \hat{d}(t+s)$, can be written as (Johnston and DiNardo 1997, pp. 232)

$$e(t+s) = \sum_{i=1}^{s} \eta_{s-i}\varepsilon(t+i)$$

The prediction variance is given by

$$Var\big(e(t+s)\big) = \sigma_\varepsilon^2 \sum_{i=1}^{s} \eta_i^2$$

Confidence intervals can then be made for birth predictions using the prediction variance

$$Var(s) = \sigma_\varepsilon^2 \left[1 + (s-1)/J^2\right]$$

Lee's next approach relaxes the assumption of white noise disturbances. The disturbances are assumed to have a covariance structure instead. It is assumed that the *NRR* can be expressed as a linear function of lagged values of a white noise process. The model is given by

$$\varepsilon(t) = \sum_{i=1}^{n} \alpha(i)\varepsilon(t-i) + \tau(t),$$

where $\tau(t)$ is a white noise, and equivalently

$$\varepsilon(t) = \sum_{i=0}^{\infty} \beta(i)\tau(t-i)$$

The best predictor is given by

$$\overset{*}{\varepsilon}(t+s) = \sum_{i=1}^{n} \alpha(i)\varepsilon(t+s-i),$$

and the prediction error is given by $e(t+s) = \sum_{i=1}^{s} \beta(s-i)\tau(t+i)$, implying that

$$\varepsilon(t+s) = \hat{\varepsilon}(t+s) + \sum_{i=1}^{s} \beta(s-i)\tau(t+i)$$

Getting back to birth $d(t+s)$ can be written as

$$d(t+s) = \hat{d}(t+s) + \sum_{i=1}^{s} \eta_{s-i} \varepsilon(t+i)$$

$$= \hat{d}(t+s) + \sum_{i=1}^{s} \eta_{s-i} \left[\overset{\wedge}{\varepsilon}(t+i) + \sum_{i=1}^{s} \beta(s-i)\tau(t+i) \right],$$

where $\hat{d}(t+s)$ is the predictor of $d(t+s)$. The best predictor of $d(t+s)$ is given by

$$d(t+s) = \hat{d}(t+s) + \sum_{i=1}^{s} \eta_{s-i} \overset{\wedge}{\varepsilon}(t+i)$$

The prediction error will be given by

$$e(t+s) = \sum_{i=1}^{s} \eta_{s-i} \sum_{l=0}^{j} \beta(j-l)\tau(t+l)$$

Defining $v(i) = \sum_{l=0}^{i} \eta(j)\beta(i-l)$, the prediction error can be given by

$\sum_{i=1}^{s} v(s-i)\eta(t+i)$. The variance is given by

$$Var(s) = \sigma_\eta^{2} \sum_{i=0}^{s-1} (v(i))^2$$

Lee presents next a Markov process. The model is given by

$$\varepsilon(t) = \gamma\varepsilon(t-1) + \tau(t)$$

The predicted fertility is given by

$$\overset{\wedge}{\varepsilon}(t+s) = \gamma^s \varepsilon(t)$$

This indicates that

$$d(t+s) = \hat{d}(t+s) + \sum_{i=1}^{s} \eta_{s-i} \overset{\wedge}{\varepsilon}(t+s)$$

$$= \hat{d}(t+s) + \varepsilon(t)\sum_{i=1}^{s} \eta_{s-i}\gamma^i ,$$

and the variance is given by

$$Var(e(t+s)) = \sigma_\varepsilon{}^2 \left(\sum_{i=1}^{s} \alpha(s-i)\gamma^i \right)^2$$

$$= \sigma_\varepsilon{}^2 \left(\gamma^s + \frac{\gamma(1-\gamma^{s-1})}{(1-\gamma^s)J} \right)^2,$$

by taking $\alpha(0) = 0$, and $\alpha(i) = \dfrac{1}{J}$.

Finally, the author presents an AR(2) process and a random walk process. No properties for the AR(2) process were given in Lee's paper. The random walk model, however, was discussed in more details and is given by

$$\varepsilon(t) = \varepsilon(t-1) + \tau(t)$$

The best predictor is the last observed value. Since the base period level of *NRR* equals the expected value for all future periods, Lee takes $\varepsilon(t) = 0$, and $\overset{\wedge}{\varepsilon}(t+s) = 0$, and, therefore,

$$\overset{*}{d}(t+s) = \hat{d}(t+s)$$

Lee and Carter (1992) present the Lee-Carter method for forecasting mortality. Using the matrix of United States death rates 1933-1987, the authors made long run forecasts and confidence intervals for age-specific mortality.

The Lee-Carter model can be described as follows. Let $m(j,t)$ be the central death rate for age j at time t, $k(t)$ a time varying mortality index, and $c(j)$ and $d(j)$ as defined in section 3.5. Then $m(j,t)$ is modelled as

$$m(j,t) = \exp(c(j) + d(j)k(t) + \varepsilon(j,t)),$$

where $\varepsilon(j,t)$ has mean zero and variance $\sigma^2{}_\varepsilon$.

Lee and Carter suggest the following method for estimation. First, the $k(t)'s$ are chosen to sum to zero and the $d(j)'s$ to unity. Then $c(j) = \sum_{t=1}^{T} \ln(m(j,t))/T$, $k(t) \approx \sum_{j} \{\ln(m(j,t)) - c(j)\}$, and $d(j)$ is found by regressing $\ln(m(j,t)) - c(j)$ on $k(t)$, without a constant, separately for each age group.

Next, the authors present a model for $k(t)$. After calculating the mortality index $k(t)$, the authors model it by a random walk with a drift. Introducing the dummy variable flu for the 1918 influenza epidemic, $k(t)$ is given by

$$k(t) = k(t-1) - 0.365 + 5.24\,flu + \varepsilon_k(t)$$

The authors' next step is to make confidence intervals. Lee and Carter ignore all sources of error, except that of $k(t)$. Confidence intervals for life expectancy are constructed using confidence

intervals for $k(t)$. With $SE(k(t))$ the standard error of $k(t)$, the 95% confidence interval for the forecast of $m(j,t)$ is given by

$$Po\text{int } forecast \pm \exp\{1.96(d(j)SE(k(t)))\}$$

To justify ignoring all sources of error but that of $k(t)$, Lee and Carter carried out the following analysis. The true value of the forecast $\ln\left\{\hat{m}(j,t+s)\right\} = \hat{c}(j) + \hat{k}(t+s)\hat{d}(j)$ is given by

$$\ln(m(j,t+s)) = \left(\hat{c}(j) + \varepsilon_c(j)\right) + \left(\hat{k}(t+s) + \varepsilon_k(t+s)\right)\left(\hat{d}(j) + \varepsilon_d(j)\right) + \varepsilon(j,t+s)$$

The forecast error (FE) is given by

$$\varepsilon(j,t+s) + \varepsilon_c(j) + \varepsilon_k(t+s)\left(\hat{d}(j) + \varepsilon_d(j)\right) + \varepsilon_d(j)\hat{k}(t+s)$$

Assuming that error terms are independent, the variance of FE is given by

$$Var(FE) = var(\varepsilon(j,t+s)) + var(\varepsilon_c(j)) + \hat{d}^2(j)var(\varepsilon_k(t+s)) + var(\varepsilon_d(j))(\hat{k}^2(t+s) + var(\varepsilon_k(t+s))$$

The authors proceed with the estimation of the four variance terms. The term $var(\varepsilon(j,t+s))$ is estimated by the variance of the error of fitting age group j for the given data. The variance of the $c(j)'s$ was taken to be the variance of $\ln(m(j,t))/T$ where T is the number of observations on $m(j)$. For $var(\varepsilon_d(j))$ Lee and Carter applied the following procedure. From the residuals of the original fit of the model of the observed data, 400 samples with replacement were taken. The errors were then added to the matrix $\ln(m(j,t)) - \ln(m(j,.))$. For each of the 400 newly created matrices, $d(j)$ was re-estimated using the methods mentioned above. Finally, the variances and covariances were calculated

The procedure used for $var(\varepsilon_d(j))$ is a direct application of the Bootstrap. The residuals of the original fit are bootstrapped, and new data are created using the bootstrapped residuals. From these new data a new fit is made and $d(j)$ is re-estimated 400 times. The resulting distribution of $d(j)$ is used to calculate variances.

Alho and Spencer (1991) derived formulas for the propagation of error. Using a Taylor series expansion, formula's for variances and covariances of the first and second generations of births were developed.

The authors start with a set of definitions. Alho and Spencer define the population vector $P(t) = \{P(0,t),...,P(l,t)\}'$, where $P(j,t)$ is the size of the female population age j at time t,

and l is the last age group. They also define a $(l+1)(l+1)$ matrix $A(i,j,t)$, $i,j=0,...,l$. The elements $A(0,15,t),...,A(0,44,t)$ are the age-specific fertility rates of year t, and $A(1,0,t),...,A(l,l-1,t)$ are age-specific survival probabilities from age zero to age one, from age one to age two, and so on, during time t. The element $A(l,l,t)$ is defined as the average survival probability in the last age-group. All other elements of the matrix $A(i,j,t)$ are zero. This implies the classical Leslie model

$$P(t+1) = A(t+1)P(t)$$

Furthermore, the authors make the following set of definitions

$$p(j,t) = \log P(j,t)$$
$$ft(j,t) = \log A(0,j,t)$$
$$sv(j,t) = \log A(j+1,j,t)$$
$$sv(l,t) = \log A(l,l,t)$$

That means that $p(j,t)$ is the logarithm of the population at age j at time t, $ft(j,t)$ is the logarithm of age-specific fertility rates at time t given by $A(0,j,t)$, $sv(j,t)$ is the logarithm of the survival rates from age j to age $j+1$ at time t given by $A(j+1,j,t)$, and $sv(l,t)$ is the logarithm of the survival rate in the last age group l at time t given by $A(l,l,t)$.

Using these definitions, the authors introduce the following formulas for the survivors of the jump-off population

$$p(j,t) = p(j-t,0) + \sum_{m=o}^{t-1} sv(j-t+m,m)$$

for $1 \le t \le j \le l-1$. For $t \le l$, the formula becomes

$$p(l,t) = \log\left\{ \sum_{n=0}^{t} \exp\left[p(l-n,0) + \sum_{m=0}^{n-1} sv(l-n+m,m) + \sum_{h=n}^{t-1} sv(l,h) \right] \right\}$$

For $t = 1,...,16$, when there are no births to births, Alho and Spencer present the following formula for births

$$p(0,t) = \log\left\{ \sum_{j=15}^{44} \exp\left[p(j-t+1,0) + \sum_{n=0}^{t-2} sv(j-t+1+n,n) + ft(j,t-1) \right] \right\}$$

The formula for the survivors of the first 16 birth cohorts, $\max[0,t-16] \le j < t$, is given by

$$p(j,t) = p(0,t-j) + \sum_{n=0}^{j-1} sv(n,t-j+n)$$

Finally, Alho and Spencer present the formula for the second generation of births, $t = 17, ..., 32$,

$$p(0,t) = \log(\sum_{j=t-1}^{44} \exp(p(j-t+1,0) + \sum_{n=0}^{t-2} sv(j-t+1+n,n) + ft(j,t-1)) +$$

$$\sum_{j=15}^{t-2} \exp p(0,t-1-j) + \sum_{n=0}^{j-1} sv(n,t-1-j+n) + ft(j,t-1))$$

The first sum over j is for survivors of the jump-off population, and the other sum over j is for births to births.

Alho and Spencer introduce next randomness to their analysis. Random variables are distinguished by a ~. Notation given earlier is reserved for point forecasts. For the jump-off population it is assumed that $E\left(\tilde{p}(0)\right) = p(0)$ and $\text{cov}\left(\tilde{p}(0)\right) = \Psi_{Jo}(0)$, where $p(0)$ is a known vector and $\Psi_{Jo}(0)$ is a known covariance matrix. The authors further assume that there are predictions $ft(t)$ for $\tilde{ft}(t)$ such that $E\left(\tilde{ft}(t)\right) = ft(t)$, and $\text{cov}\left(\tilde{ft}(t), \tilde{ft}(u)\right) = \Psi_{ft}(t,u)$, where $ft(t)$ are known vectors and $\Psi_{ft}(t,u)$ are known covariance matrices, with elements $\sigma_{ft}(i,j,t,u)$. There are also predictions $sv(t)$ assumed for $\tilde{sv}(t)$ such that $E\left(\tilde{sv}(t)\right) = sv(t)$ and $\text{cov}\left(\tilde{sv}(t), \tilde{sv}(u)\right) = \Psi_{sv}(t,u)$, where $sv(t)$ are known vectors and $\Psi_{sv}(t,u)$ are known matrices with elements $\sigma_{sv}(i,j,t,u)$. It is further assumed that the variables $\tilde{p}(0)$, $\tilde{ft}(t)$ and $\tilde{sv}(t)$ are jointly normal, with jump-off population independent of vital rates, and fertility rates independent of survival rates. Furthermore, the authors assume that $E\left(\tilde{p}(t)\right) = p(t)$ and $\text{cov}\left(\tilde{p}(t), \tilde{p}(u)\right) = \Psi(t,u)$, where the elements of $\Psi(t,u)$ are denoted by $\sigma(i,j,t,u)$. This means that the distribution of $\tilde{p}(t)$ is approximated by a normal distribution for all t.

Alho and Spencer proceed with formulas for the covariances. The covariance between the prediction error of population size in ages i and j at time t, $1 \le t \le i \le j \le l$, is given by (see Appendix A, section A.3)

$$\text{cov}\left(\tilde{p}(i,t), \tilde{p}(j,t)\right) = \sigma(i,j,t) + \sum_{n=0}^{t-1} \sum_{m=0}^{t-1} \sigma_{sv}(i-t+m, j-t+n, m, n)$$

Alho and Spencer define $B(j,t)$ as the number of children born to a woman age j at time t. That means that $P(0,t) = B(15,t) + ... + B(44,t)$, where

$$B(j,t) = \exp\left[p(j,t-1) + ft(j,t-1)\right]$$

For the next covariance formula's Alho and Spencer use a Taylor series expansion (see Appendix A, section 1.2). The covariance between the jump-off populations at two different times is given by (see Appendix A, section A.3)

$$\sigma(0,0,t,u) = \left(\frac{1}{P(0,t)P(0,u)} \right) \sum_{i=15}^{44} \sum_{j=15}^{44} B(j,t)B(j,t)(\sigma(i-t+1,j-u+1) +$$

$$\sum_{m=0}^{t-2} \sum_{n=0}^{u-2} \sigma_{sv}(i-t+1+m,j-u+1+n,m,n) + \sigma_{ft}(i,j,t-1,u-1))$$

For the surviving births, $\max[0,t-16] \le j \le i < t$, $\mathrm{cov}\left(\tilde{p}(i,t), \tilde{p}(j,t) \right)$ is given by (see Appendix A, section A.3)

$$\mathrm{cov}\left(\tilde{p}(i), \tilde{p}(j) \right) \approx \sigma(0,0,t-i,t-j) + \sum_{m}^{i-1} \sum_{n}^{j-1} \sigma_{sv}(m,n,t-i+m,t-j+n)$$

Alho and Spencer finally consider the covariance between the surviving births in age i at time t and the survivors of the jump-off population at age j at time u ($u \le j$). This covariance is given by (see Appendix A, section A.3)

$$\mathrm{cov}\left(\tilde{p}(i,t), \tilde{p}(j,u) \right) = \sum_{n=0}^{i-1} \sum_{m=0}^{n-1} \sigma_{sv}(n, j-u+m,t-i+n,m)$$

In Alho (1992a) and Alho (1992b) the author presents formulas for the propagation of error in stochastic cohort-component population forecasts. More specifically, the author tries to identify the contribution of different vital processes to the forecast error of different ages. Building on earlier work (Alho and Spencer 1991), Alho makes a set of assumptions for the different sources of error, jump-off population, mortality, migration and fertility. Furthermore, the author assumes that the sources are independent.

Alho considers first the jump-off population. The author assumes that

$$\tilde{p}(j,0) = p(j,0) + \varepsilon_{JO}$$

where

$$\varepsilon_{JO} \sim \mathrm{N}(0,\sigma^2{}_{JO}),$$

and JO stands for jump-off.
For mortality it is assumed that

$$\tilde{sv}(j,t) = sv(j,t) + \varepsilon_{sv}(j,t)$$

Alho replaces $\varepsilon_{sv}(j,t)$ by its average value $\varepsilon_{sv}(t)$, and defines $\varepsilon_{sv}(t)$ by

$$\varepsilon_{sv}(t) = \sum_{q=0}^{t-1} e_{sv}(q),$$

with $e_{sv}(q) \sim N(0,\sigma_{sv}^{~2})$, i.i.d. for $q = 0,1,...$ Alho then assumes that the effect of migration on error is absorbed into $sv(j,t)$ by adding an error component. In other words $\tilde{s}(j,t)$ is given by

$$\tilde{sv}(j,t) = sv(j,t) + \varepsilon_{sv}(t) + \varepsilon_{mg}(t),$$

where $\varepsilon_{mg}(t) \sim N(0,\sigma_{mg}^{~2})$, i.i.d. for $t = 0,1,...$, are independent of $\varepsilon_{sv}(t)$. For fertility Alho assumes that

$$\tilde{ft}(j,t) = ft(j,t) + \varepsilon_{ft}(t)$$

where

$$\varepsilon_{ft}(t) = \sum_{q=0}^{t-1} e_{ft}(q),$$

with $e_{ft}(q) \sim N(0,\sigma_{ft}^{~2})$, i.i.d. for $q = 0,1,...$

Furthermore, Alho asserts that the births in the year $t-1$ that contribute to $P(0,t)$ should be subjected to migration and mortality during year $t-1$ in two ways. First, migration and mortality affect the number of women in child-bearing ages during the fraction of year before they give birth. Second, the children at age zero are subjected to migration and mortality before year t is reached. Alho deals with this issue by adding this uncertainty factor the fertility, taking

$$\tilde{ft}(j,t) = \tilde{ft}(j,t) + \varepsilon_{ft}(t) + \varepsilon_{sv}(t) + \varepsilon_{mg}(t)$$

In addition, Alho defines $V(t)$ as the error component due to mortality and migration used to get from jump-off year to forecast year t, by

$$V(t) = \sum_{k=0}^{t-1} \left(\varepsilon_{sv}(k) + \varepsilon_{mg}(k) \right)$$

Armed with these assumptions, the author moves to calculate the different variances. The author applies the methods developed in Alho and Spencer (1991) of using a Taylor series development. The covariance between births in the year t and in the year u, $1 \le t \le u \le 16$, is given by (see Appendix A section A.4)

$$\text{cov}\left(\tilde{p}(0,t), \tilde{p}(0,u) \right) \approx \sigma_{JO}^{~2} + t\sigma_{ft}^{~2} + t\sigma_{mg}^{~2} + \sigma_{sv}^{~2}(u+t+1)(t+1)t/6$$

That means that the variance is given by

$$\mathrm{var}\left(\tilde{p}(0,t)\right) \approx \sigma_{JO}^{2} + t\sigma_{ft}^{2} + t\sigma_{mg}^{2} + \sigma_{sv}^{2} g(t),$$

where $g(t) = (2t+1)(t+1)t/6$.

Next, Alho considers the second generation of births and their survival. That is when the births generated during the first 16 years contribute new births, $17 \le t \le 32$. The contribution of the jump-off population, mortality and migration to the uncertainty of $\tilde{p}(0,t)$ is $e_{JO} + V(t)$.

The contribution of fertility, however, consists of two parts. First, there is the direct contribution of fertility. Second, the women at the child-bearing ages who were born after the jump-off year contribute to the uncertainty of fertility. This contribution is given by

$$H(t) = \left(\frac{1}{P(0,t)}\right)\sum_{j=15}^{t-2} B(j,t)\varepsilon_{ft}(t-j-1)$$

Alho considers two types of covariances. First, for $1 \le t \le 16 \le u \le 32$, (see Appendix A section A.4)

$$\mathrm{cov}\left(\tilde{p}(0,t), \tilde{p}(0,u)\right) = \sigma_{JO}^{2} + t\sigma_{mg}^{2} + \sigma_{sv}^{2}(u+t+1)(t+1)t/6 + t\sigma_{ft}^{2} +$$
$$\frac{1}{P(0,t)}\sum_{j=15}^{u-2} B(j,u)\sigma_{ft}^{2}\min\{t,u-1-j\}$$

And for $17 \le t \le u \le 32$ this covariance is given by (see Appendix A section A.4)

$$\mathrm{cov}\left(\tilde{p}(0,t), \tilde{p}(0,u)\right) = \sigma_{JO}^{2} + t\sigma_{mg}^{2} + \sigma_{sv}^{2}(u+t+1)(t+1)t/6 + t\sigma_{ft}^{2} +$$
$$\frac{1}{P(0,u)}\sum_{j=15}^{u-2} B(j,u)\sigma_{ft}^{2}\{u-1-j\}+$$
$$\frac{1}{P(0,t)}\sum_{j=15}^{t-2} B(j,t)\sigma_{ft}^{2}\{t-1-j\}+$$
$$\frac{1}{P(0,t)}\frac{1}{P(0,u)}\sum_{j=15}^{t-2}\sum_{k=15}^{u-2} B(j,t)B(k,u)\sigma_{ft}^{2}\min\{t-1-j,u-1-k\}$$

That indicates that the variance is given by

$$\mathrm{var}\left(\tilde{p}(0,t)\right) = \sigma_{JO}^{2} + t\sigma_{mg}^{2} + \sigma_{sv}^{2}g(t) + t\sigma_{ft}^{2} + \frac{2}{P(0,t)}\sum_{j=15}^{t-2} B(j,t)\sigma_{ft}^{2}\{t-1-j\}+$$
$$\left(\frac{1}{P(0,t)}\right)^{2}\sum_{j=15}^{t-2}\sum_{k=15}^{t-2} B(j,t)B(k,t)\sigma_{ft}^{2}\min\{t-1-j,t-1-k\}$$

Alho finally presents formulas for the covariances for the third generation of births, i.e. the years $t = 33,...,48$. The formulas and their derivation are shown in Appendix A section A.4.

Bell (1997) discusses forecasting mortality and fertility rates using two general approaches. First, the author describes the curve fitting approach. Second, a principal component approach is discussed.

The curve fitting approach can be described as follows. Let $r(j,t)$ be an observed fertility or mortality rate or transformation of them for age $j = 1,...,l$ and year $t = 1,...,T$. The $r(j,t)$'s are approximated by a parametric curve, $h(j,\Theta_t)$, where Θ_t is a $m \times 1$ vector of curve parameters. The parametric curve $h(j,\Theta_t)$ is used for every year, but with different values of Θ_t every year. The curve is fitted to data separately for each year t by least squares by finding Θ_t that minimizes $\sum_i \left(r(i,t) - h(i,\Theta_t) \right)^2$. The calculated Θ_t are considered as observations on an m-dimensional multivariate time series and forecasted. The forecast $\overset{\wedge}{\Theta}_{T+s}$ of Θ_{T+s} for $s > 0$ produces forecasted curves $\hat{h}(j, \overset{\wedge}{\Theta}_{T+s})$, which are taken as forecasts for $r(j, T+s)$.

The author discusses next a principal component approach. A linear combination is defined by a $l \times v$ matrix C', $v \leq l$. Given C' the rate $r(t)$ is approximated each year by $C' \overset{\wedge}{\beta}(t)$, where $\overset{\wedge}{\beta}(t)$ is obtained by Least Squares regression of $r(t)$ on C'. The $\overset{\wedge}{\beta}(t)$'s are then viewed as a v-variate time series, which can be forecasted. The forecast of $r(t+s)$ is then given by $C' \overset{\wedge}{\beta}(t+s)$. The columns of C' are the first v principal component vectors of the sum of squares and cross products matrix of the data $\sum_t r(t) r'(t)$.

Section 3.6 discusses purely stochastic models for forecasting vital rates. Sykes uses stochastic versions of the Leslie matrix by introducing random vectors, assuming a Branching Galton-Watson process, and by introducing random matrices. Lee (1974) uses four different time series formulations to forecast births. Lee and Carter (1992) forecast mortality with the Lee-Carter method which incorporates regression models and time series analysis. Alho and Spencer (1991) provide variance formulas for the jump-off population and the vital rates using a Taylor series expansion. Alho (1992a,1992b) presents formulas for the variances of the first, second and third generations of births using a Taylor series expansion, after making assumptions about errors in vital rates. Bell describes methods for forecasting mortality and fertility rates based on the curve fitting approach and the principal component approach.

4 Results and findings

In this chapter the results and findings of the research are discussed. The chapter consists of two sections. Section one considers the direct applications of the Bootstrap, while section 2 describes the partial applications of the Bootstrap in probabilistic models of population projections.

4.1 Direct applications of the Bootstrap in probabilistic models of population projections

Lee and Carter (1992) make a direct application of the Bootstrap to estimate $\mathrm{var}(\varepsilon_d(j))$ (see section 3.6.2). First, the authors bootstrap the residuals of the original fit for the model of the observed data. Second, new data are created by adding the bootstrapped residuals to the matrix $\ln(m(j,t)) - \ln(m(j,.))$. From these new data a new fit is made and $d(j)$ is re-estimated. This procedure is repeated 400 times. The variance of $d(j)$ is calculated from the resulting distribution.

4.2 Indirect applications of the Bootstrap in probabilistic models of population projections

The method used in Pflaumer (1988) is a simulation process, but bears a resemblance to the parametric Bootstrap. A parametric distribution is used, the uniform distribution. Instead of drawing random samples from the distribution, random draws are made from the distribution assumed. The random draws are used to project the population forward. This procedure is repeated M times. The result is M projected populations at the target year. The distribution of the generated projected populations is used to produce confidence intervals.

In Lutz et al. (1996) the authors make an indirect application of the parametric Bootstrap. A standard normal distribution is assumed. Unlike the parametric Bootstrap, no samples are drawn from the distribution. The authors, however, make random draws from the standard normal distribution, and use the random draws to estimate the vital rates. The vital rates estimated are used to project the population. This procedure is repeated 1000 times creating a distribution of projected population. This distribution is used to construct confidence intervals.

The methods mentioned in Alho (1990) bear a small similarity to the Bootstrap. Using the Finnish fertility data in 1776-1976, the author calculates the volatility of fertility (see section 3.5). No re-sampling is made, but instead the author uses the data set itself to create percentiles. Those percentiles are then used to create confidence intervals.

5 Summary and conclusions

5.1 Summary

The report starts with explaining the Bootstrap. First, the Bootstrap estimate of standard error was discussed. Second, the Bootstrap confidence interval was described. Third, the Bootstrap t-interval, the Bootstrap percentile interval, and the Bootstrap BC_a interval were explained. Finally, Bootstrap types and Bootstrap applications in regression and time series analysis were viewed.

Before discussing the probabilistic models of population forecasts the data sources and the deterministic model were mentioned. The literature used in this paper was mentioned. Next, the deterministic model and the Leslie matrix were described.

The probabilistic models of population forecasts based on Ex-post analysis were described. First, the model explained in Keyfitz (1981) was described. This was followed by a description of the model mentioned in Stoto (1983).

The next step was the discussion of models for aggregate time series. First, the model discussed in Cohen (1986) was described. Second, the model discussed in Pflaumer (1992) was viewed.

Furthermore, the models of probabilistic population forecasting that are based on a mix of statistical modelling and expert judgement were viewed. First, the approximately linear models mentioned in Alho and Spencer (1985) were described. Second, Pflaumer's simulation procedures as explained in Pflaumer (1988) were discussed. Third, Alho's methods of modelling the vital rates incorporating expert judgement discussed in Alho (1990) were explained. Fourth, Lee's application of the Lee-Carter method mixed with expert judgement to forecast fertility as explained in Lee (1993) was described. Fifth, the random scenario model mentioned in Lutz et al. (1996,1997) was explained. Finally, the development of the random scenario model discussed in Lutz et al. (2001) and the conditional probability forecasting methods mentioned in Sanderson et al. (2003), were considered.

Finally, probabilistic models of population forecasts solely based on stochastic modelling were discussed. First, the theories of population projection mentioned in Cohen (1977), Heyde and Cohen (1985), and Tuljapurkar (1990) were described. Second, stochastic models for forecasting vital rates were discussed. These include Lee's ARIMA models for forecasting the number of births discussed in Lee (1994), the Lee-Carter method explained in Lee and Carter (1992), and the methods mentioned in Alho and Spencer (1991), and Alho (1992a, 1992b) for forecasting the vital rates. Finally, the curve fitting approach and the principal component approach mentioned in Bell (1997) were briefly discussed.

5.2 Conclusions

This report tries to establish whether the probabilistic models of population forecasts can be viewed as an application of the Bootstrap. This is done by describing the probabilistic models of population forecasts in detail, and noting when there is a resemblance with the Bootstrap.

The only direct application of the Bootstrap in the probabilistic models of population forecasts was found in the Lee-Carter method. Lee and Carter (1992) bootstrap the residuals of a regression model used to estimate the central death rate. The bootstrapped residuals are then used to re-estimate an age-varying constant. This procedure is repeated 400 times, and the variance of the age-varying constant is estimated from the resulting distribution. This variance formed part of the variance of the prediction error of the central death rate. This is a direct application of the Bootstrap to estimate the standard error.

Other probabilistic models of population forecasts apply methods bearing similarities to the Bootstrap, including the model mentioned in Pflaumer (1988). Although using a simulation process, Pflaumer makes random draws from an assumed uniform distribution for the vital rates. The random draws are then used to project the population forward to a specified target year. This procedure is repeated M times resulting in a distribution for the projected population at the target year. This distribution is then used to create confidence intervals.

In spite of not being a direct application of the Bootstrap, this method resembles the parametric Bootstrap. A parametric distribution is assumed, and a distribution is created from which confidence intervals are created. The simplification being not drawing random samples from the assumed parametric distribution

Another probabilistic model resembling the Bootstrap is the random scenario model mentioned in Lutz et al. (1996). The authors assume a standard normal distribution, and make random draws from the assumed distribution. These draws are used to estimate vital rates, and based on the estimated vital rates the population is projected. Repeating the procedure 1000 times, a distribution of the projected population is created, and confidence intervals are constructed from the this distribution.

This is not a direct application of the Bootstrap, but it bears similarities to the parametric Bootstrap. As in Pflaumer (1988) a parametric distribution is assumed without drawing random samples from it, and random draws from the distribution are used to estimate vital rates and to project the population.

The last probabilistic model for population forecasts found to resemble the Bootstrap is that mentioned in Alho (1990). After calculating the volatility of the Finnish fertility 1776-1976, the author uses the percentiles of the empirical distribution of the volatility to estimate the error in fertility forecasts. The estimated error is then used to create confidence intervals. Although there is no re-sampling done, the method resembles the Bootstrap. The resemblance originates from the use of percentiles to create confidence intervals.

A natural suggestion for new applications of Bootstrap methods in probabilistic models of population forecasting is to fully apply the Bootstrap in the models using methods bearing similarities to it. Instead of making random draws from the uniform distribution as in Pflaumer (1988), random samples can be drawn from the distribution and the mean rate calculated each time. The process can be repeated sufficient times, and a standard error of the rate can be calculated. After estimating the rate, e.g. by taking the mean of all means of Bootstrap samples, the estimated rate can be used, together with the standard error estimated to produce population projections along with confidence intervals. This same procedure may be applied to the random scenario approach as described in Lutz et al. (1996). In Alho (1990), the volatility estimates can be bootstrapped sufficient times and each time the mean or median is calculated. From the calculated means or medians a standard error can be estimated and used, along with the estimated rate, to create confidence intervals for fertility and, hence, for future populations.

Another idea would be to adapt Bootstrap methods when there is regression analysis or time series analysis of vital rates involved. When dealing with regression analysis bootstrapping residuals or bootstrapping pairs can be considered. If a time series is considered bootstrapping residuals or the moving blocks Bootstrap method could be used.

Bootstrap methods may further be incorporated in the probabilistic models for population forecasts based on Ex-post analysis. The projection errors can be bootstrapped and the mean error or the RMS (root mean squared error) can be calculated each time. The process can be repeated sufficient times and the standard error of the mean error or the RMS can be calculated. This standard error can be used as a forecasting error to construct confidence intervals for future population projections. The same procedure can be applied to Stoto's jump-off error and random error, as well to Keilman's percentage error.

References

Ahlburg D.A. & Lutz W. (1998) Introduction: the need to rethink approaches to population forecasts. In Lutz W., Vaupel J.W., and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol. 24, pp.1-14. New York, The United States.

Ahlburg D.A., Lutz W., Vaupel J.W. (1998) Ways to improve population forecasting: what should be done differently in the future. In Lutz W., Vaupel J.W., and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol. 24, pp. 191-198. New York, The United States.

Alho J. M. (1990) Stochastic methods in population forecasting. *International Journal of Forecasting*, 6(4), pp. 521-530. Amsterdam, the Netherlands.

Alho J. M (1992a) The magnitude of error due to different vital processes in population forecasts. *International Journal of forecasting*, 8(3), pp. 301-314. Amsterdam, the Netherlands.

Alho J. M (1992b) The contribution of different sources of error to the accuracy of national population forecasts in the U.S. SRD research report 92/07. Statistical Research Division, US Bureau of Census. Washington DC, the United States.

Alho J. M (1997) Scenarios, uncertainty and conditional forecasts of the world population. *Journal of The Royal Statistical Society* series A, 160(1), pp. 71-85. Oxford, the United Kingdom.

Alho J. M. & Spencer B. D. (1985) Uncertain population forecasting. Journal of the American Statistical Association, 80(390), pp. 306-314. Alexandria, Virginia, the United States.

Alho J. M. & Spencer B. D. (1990a) Error models for official mortality forecasts. *Journal of the American Statistical Association*. 85(411), pp. 609-616. Alexandria, Virginia, the United States.

Alho J. M. & Spencer B. D. (1990b) Effects of targets and aggregation on the propagation of error in mortality forecasts. *Mathematical Population Studies*, 2(3), pp.209-227. Philadelphia, the United States.

Alho J. M. & Spencer B. D. (1991) A population forecast as a database: Implementing the stochastic propagation of error. *Journal of Official Statistics*, 7(3), pp. 295-310. Örebro, Sweden.

Alho J. M. & Spencer B. D. (1997) The practical specification of the expected error of population forecasts. *Journal of Official Statistics*, 13(3), pp. 203-225. Örebro, Sweden.

Bay J.S. (1999) Fundamentals of linear state space systems. Mcgraw-Hill. Singapore.

Bell W. R. (1997) Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. Journal of Official Statistics, 13(3), pp. 279-303. Örebro, Sweden.

Billingsley P. (1995) Probability and measure. John Wiley and sons, Inc. New York, The United States

Chernick M. R. (1999) Bootstrap methods: A practitioner's guide. John Wiley and sons, Inc. New York, The United States

Cohen J. E. (1979) Ergodic theorems in Demography. *Bulletin of the American Mathematical society*, 1(2), pp. 275-295. Providence, Rhode Island, the United States.

Cohen J. E. (1986) Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography*, 23(1), pp. 105-126. Silver Spring, Maryland, the United States.

Efron B. & Tibshirani R. G. (1993) An introduction to the bootstrap. Chapman & Hall. New York, The United States.

Hamilton J. D. (1994) Time series analysis. Princeton University Press. New Jersey, the United States.

Heyde C. C. & Cohen J. E. (1985) Confidence intervals for demographic projections based on products of random matrices. *Theoretical Population Biology*, 27: pp. 120-153.

Johnston J. & Dinardo J. (1997) Econometric methods. International editions. McGraw-Hill. Singapore.

Keilman N. (1998) How accurate are the United Nations world population projections? In Lutz W., Vaupel J.W., and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol. 24, pp. 156-190. New York, The United States.

Keyfitz N. (1968a) Introduction to the mathematics of population. Addison-Wesley Publishing company, Inc. The United States.

Keyfitz N. (1968b) The population projection as a matrix operator. Demography, 1(1), pp. 56-73. Silver Spring, Maryland, the United States.

Keyfitz N. (1981) The limits of population forecasting. *Population and Development Review*, 7(4), pp. 579-593. New York, The United States.

King j. (2003) Conductance and rapidly mixing Markov Chains. University of British Columbia website. Vancouver, Canada. Internet: http://www.cs.ubc.ca/~king/academic/papers/conductance.pdf. Last visited on August 30, 2005.

Land C. K (1986) Methods for national population forecasts: a review. *Journal of the American statistical association*, 81(396): pp. 888-901. Alexandria, Virginia, the United States.

Lang S. (1984) Algebra. 2nd edition. Addison-Wesley Publishing Company, Inc. Canada.

Lee R. D (1974) Forecasting births in post-transition populations: stochastic renewal with serially correlated fertility. *Journal of the American Statistical Association*, 69(347), pp. 607-617. Alexandria, Virginia, the United States.

Lee R.D. (1990) Long-run global population forecasts: a critical appraisal. In Lutz W., Vaupel J.W., and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol.16, pp. 44-71. New York, The United States.

Lee R. D. (1992) Stochastic demographic forecasting. *International Journal of Forecasting*, 8(3), pp. 315-327. Amsterdam, the Netherlands.

Lee R. D (1993) Modelling and forecasting the time series of the US fertility: Age distribution, range and ultimate level. *International Journal of forecasting*, 9(2), pp. 187-202. Amsterdam, the Netherlands.

Lee R.D. (1998) Probabilistic approaches in population forecasting. In Lutz W., Vaupel J.W., and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol. 24, pp. 156-190. New York, The United States.

Lee R. D. (2004) Quantifying our ignorance: Stochastic forecasts of population and public budgets. Paper prepared for the Rand Summer Institute Gala celebration for the NIA Centres for the demography and Economic of Aging. Centre for the Economics and Demography of Aging. University of California, Berkeley. California, The United States.

Lee R. D. & Carter L. R. (1992) Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), pp. 650-670. Alexandria, Virginia, the United States.

Lee R. D & Tuljapurkar S (1994) Stochastic population forecasts for the United States: Beyond high, medium and low. *Journal of the American Statistical Association*, 89(428), pp. 1175-1189. Alexandria, Virginia, the United States.

Leslie P. H. (1945) On the use of matrices in certain population mathematics. *Biometrika*, 33, pp.183-212. Oxford, the United Kingdom.

Lutz W., Sandeson W., and Scherbov S. (1996) Probabilistic population projections based on expert opinion. In Lutz W. (ed.) ' the Future of population of the world: what can we assume today?'. Earthscan, London, the United Kingdom.

Lutz W., Scherbov S. (1997) Sensitivity analysis of expert-based probabilistic population projections in the case of Austria. International Institute for Applied Systems Analysis (IIASA). Interim report IR-97-048/August. Laxenburg. Austria.

Lutz W., Sandeson W., and Scherbov S. (1997) Doubling of world population unlikely. *Nature*, 387, pp. 803-805. New York, the United States.

Lutz W., Sanderson W.C., Scherbov S. (1998) Expert-based probabilistic population projections. In Lutz W., Vaupel J.W. and Ahlburg D.A. (Eds.) Frontiers of Population Forecasting, supplement to Population and Development Review, vol. 24, pp.139-155. New York, The United States.

Lutz W., Sanderson W. & Scherbov S. (2001) The end of world population growth. *Nature,* 412, pp. 543-545. New York, the United States.

Medhi J. (1994) Stochastic processes. 2nd edition. John Wiley. New York, the United States.

Pflaumer P. (1988) Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting*, 4(1), pp.135-142. Amsterdam, the Netherlands.

Pflaumer P. (1992) Forecasting US population totals with the Box-Jenkins approach. *International Journal of Forecasting*, 8(3), pp.329-338. Amsterdam, the Netherlands.

Ross S. M. (2000) Probability models. 7th edition. Academic Press. San Diego, the United States.

Stoto M. A. (1983) The accuracy of population projections. *Journal of the American Statistical Association*, 78(381), pp. 13-20. Alexandria, Virginia, the United States.

Sykes Z. M. (1969) Some stochastic versions of the matrix model for population dynamics. *Journal of the American Statistical Association*, 64(325), pp. 11-130. Alexandria, Virginia, the United States.

Tuljapurkar S. D (1990) Population dynamics in variable environments. Springer-Verlag, Berlin, Germany.