Dynamic Demographic Models and Parameter Identification: Simulations Based on Statistical Data

Ioannis Tzortzis and Charalambos D. Charalambous

***Abstract*— This paper is concerned with dynamical population models obtained from short and long-term changes in size and age composition due to demographic processes such as births, deaths, migration, etc. Both deterministic and stochastic models are presented. The parameters which are embedded in the models may be either unavailable or noisy, therefore system identification methods are invoked to estimate these parameters. The numerical results presented illustrate that the mathematical models can reproduce the data provided by the Statistic Department of the Republic of Cyprus, and that the unknown and noisy parameters postulated in the models are determined with high accuracy.**

I. INTRODUCTION

Human population of any country grows and shrinks over time, as a consequence of the variability in birth, death, immigration and emigration rates. By integrating these demographic processes and by studying human population variation in size and age composition, mathematical models can be derived which can be used to better understand both present population state and its future trends.

By partitioning the population into different groups (e.g., according to age), and taking into account the interaction of these groups, dynamic population models can be derived which provide information about the dynamic transitions from one age group of population to another and therefore obtain predicted estimates of population growth or decline. However, often the parameters embedded into the models are either unavailable or noisy, hence system identification methods should be invoked to estimate these parameters. The objectives of this paper are the following.

- To develop new Log-Normal Stochastic mathematical models, using existing Linear models.
- To identify and estimate missing and noisy data.

The first demographic model is proposed in 1798 by T. R. Malthus in [1]; it states that population grows exponentially and hence as time progresses it can grow to infinity. The Malthus model is modified to the logistic population model by P. F. Verhulst in 1838, in an attempt to ensure the existence of a limit to population growth. Since then much work has been done to develop dynamic population models, e.g., [2], [3], [4] and [5]. Recently, work in [6] and [7] describes optimum immigration and job creation policies, which they are applied to the Canadian population.

I. Tzortzis is with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus eep7ti2@ucy.ac.cy

C.D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus, and an Adjunct Professor with the School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada chadcha@ucy.ac.cy

In this paper the emphasis is to develop Log-Normal stochastic models and estimation techniques, and then use the data provided by the Statistic Department of Cyprus (CyStat) to find the parameters.

Following CyStat, the Cyprus population is divided into three main age groups. The first age group denoted by G_1 consists of children below the age of 14; the population in this age group at any time t is denoted by $x_1(t)$. The second age group denoted by G_2 consists of all members between the age of 15 to 64; the population is denoted by $x_2(t)$. The third and last age group denoted by G_3 consists of all members over the age of 65; the population is denoted by $x_3(t)$. The total population can be divided into less or more than three age groups. Here the partition is done according to CyStat, since CyStat is considered as the main source of accurate and up-to-date statistical and other form of information obtained from various censuses, surveys and studies, which are essential for the implementation of this work.

The paper is organized as follows. Section II defines population growth rate. Section III presents a linear deterministic model and its stochastic version as found in [7]. Next, a new Log-Normal model is developed. Section IV discusses two methods for identification and estimation of missing data. Section V presents various simulations obtained by testing the performance of the population models based on estimated parameters which are then compared to the actual data. Section **??** provides concluding remarks and comments for future work.

II. POPULATION GROWTH RATE

The rate of change of population over a period of time due to natural increase and net migration is known as Population Growth Rate (PGR). The growth rate takes into account all components of population growth, the number of births, deaths, immigration and emigration. Ordinarily, PGR is expressed as a percentage ratio and its value can be positive (population increases), negative (population declines), and zero (population neither grows nor declines). Consider an initial value of population $X_I(0)$ at a growth rate r per year. If r is evaluated once per year ($m = 1$), then the final value of population X_F after t years is given by

$$X_F(t) = X_I(0)(1 + r)^t \tag{1}$$

If r is evaluated m times per year, then the final value of population after t years is equal to

$$X_F(t) = X_I(0)\left(1 + \frac{r}{m}\right)^{mt} \tag{2}$$

The value of m, the times that the growth rate is evaluated in a period of one year, varies but more often takes one of the following values:

- $m = 2$: Semi-Annually
- $m = 4$: Quarterly
- $m = 12$: Monthly
- $m = 52$: Weekly
- $m = 365$: Daily

Solving (2) in terms of r, gives the growth rate

$$r = m\left[\left(\frac{X_F(t)}{X_I(0)}\right)^{\frac{1}{mt}} - 1\right] \tag{3}$$

A growth rate that is declining does not necessarily mean that population is declining, since this rate acts on a large population base. It only indicates that the population is growing at a slower rate. From (2), taking the limit as m tends to infinity the final value of population is given by

$$X_F(t) = X_I(0)e^{rt} \tag{4}$$

where

$$e^{rt} = \lim_{m\to\infty}\left(1 + \frac{r}{m}\right)^{mt}$$

This is known as the continuous time model. In general, the continuous time model can be thought as being equivalent to the model of (2) but with growth rate evaluated daily.

The population growth model has many other applications besides calculating PGR. For example, in financial mathematics the future value of a sum of money invested for n years with interest rate r compounded once per annum is given by

$$FV_T = C_o(1 + r)^n \tag{5}$$

The interest rate can be expressed by

$$r = \left[\left(\frac{FV_T}{C_o}\right)^{\frac{1}{n}} - 1\right] \tag{6}$$

where FV_T denotes the future value of the money corresponding to an investment made today C_o, r denotes the annual interest rate, and n the period of time of the investment. Populations can grow at an exponential rate, just as compound interest accumulates in a bank account. The above observations will be used in subsequent sections to derive stochastic log-normal models that capture the evolution of various demographic data.

III. MATHEMATICAL MODELS

This section introduces mathematical models by considering different age groups. Linear Deterministic and Linear Stochastic, and Log-Normal Stochastic models are presented.

A. Linear Deterministic Model for Population

Below we describe the model introduced by [7] which is used in subsequent section to find the missing parameters from measurement data. Consider the population growth of age group G_1. The population in this age group increases, by births due to population of age group G_2 (while the effect of fertility rate of population in age groups G_1 and G_3 is consider negligible), and by the number of new immigrants. Similarly the population in age group G_1 decreases due to the number of deaths and emigrants and the number of 14 years old members passing from age group G_1 to group G_2. Thus the growth rate of population of age group G_1 is given by

$$\begin{aligned}\frac{\Delta x_1(t_i)}{\Delta t_i} &\triangleq \frac{x_1(t_i) - x_1(t_{i-1})}{(t_i - t_{i-1})}\\ \frac{\Delta x_1(t_i)}{\Delta t_i} &= (-d_1 - p_{12} - r_1 + \tau_1)x_1(t_{i-1})\\ &\quad + bx_2(t_{i-1})\end{aligned} \tag{7}$$

where, d_1 denotes the child mortality rate, p_{12} denotes the passing rate from age group G_1 to age group G_2, r_1 denotes the emigration rate, τ_1 denotes the immigration rate and b denotes the birth rate due to population of age group G_2 (while the effect of fertility rate of population in age groups G_1 and G_3 is consider negligible).

The growth rate of population of age group G_2 is given by

$$\begin{aligned}\frac{\Delta x_2(t_i)}{\Delta t_i} &\triangleq \frac{x_2(t_i) - x_2(t_{i-1})}{(t_i - t_{i-1})}\\ \frac{\Delta x_2(t_i)}{\Delta t_i} &= (-p_{23} - d_2 - r_2 + \tau_2)x_2(t_{i-1})\\ &\quad + p_{12}x_1(t_{i-1})\end{aligned} \tag{8}$$

where, p_{23} denotes the passing rate from age group G_2 to age group G_3, d_2 denotes the death rate, r_2 denotes the emigration rate and τ_2 denotes the immigration rate.

The growth rate of population of age group G_3 is given by

$$\begin{aligned}\frac{\Delta x_3(t_i)}{\Delta t_i} &\triangleq \frac{x_3(t_i) - x_3(t_{i-1})}{(t_i - t_{i-1})}\\ \frac{\Delta x_3(t_i)}{\Delta t_i} &= (-d_3 - r_3 + \tau_3)x_3(t_{i-1})\\ &\quad + p_{23}x_2(t_{i-1})\end{aligned} \tag{9}$$

where, d_3 denotes the death rate, r_3 denotes the emigration rate and τ_3 denotes the immigration rate.

The overall population growth rate is obtained from (7), (8) and (9)

$$\frac{\Delta x_{total}(t_i)}{\Delta(t_i)} = \frac{\Delta x_1(t_i)}{\Delta t_i} + \frac{\Delta x_2(t_i)}{\Delta t_i} + \frac{\Delta x_3(t_i)}{\Delta t_i} \tag{10}$$

Define the state vector by

$$x \triangleq [x_1\ x_2\ x_3]^T \tag{11}$$

Taking the limit as $\max_i(t_i - t_{i-1}) \to 0$, the mathematical models (7), (8) and (9) can be represented by the state differential equation

$$\dot{x}(t) = A(t)x(t),\ x(0) = x_0,\ t \geq 0 \tag{12}$$

where the vector $\dot{x}(t)$ represent the population growth rates, the system matrix $A(t)$ represent the system parameters, and $x(t)$ denotes the population vector of the three age groups.

B. Linear Stochastic Model for Population

Next, we develop the stochastic versions of the deterministic models. The main reason for doing so is to account for the sources of uncertainty which affect the dynamic models. Some of these sources can be modeled by additive noise, due to errors occurring from a population survey, the unexpected changes in economy, i.e war, earthquake, illegal immigration etc. Assume that the number of births, deaths, emigrants, immigrants and the passing rate are subject to additive Gaussian noise uncertainty which has zero mean and certain variance. Then the contribution of the total uncertainty can be modeled by a single zero mean variance $\sigma^2(t)$ Gaussian noise $N(0;\sigma^2(t))$ at each instant. If we further assume the noise is white then each equation is modified by including an additional additive term. Thus the stochastic version of the deterministic model is now given by

$$\begin{aligned}
\dot{x}_1(t) &= (-d_1(t) - p_{12}(t) - r_1(t) + \tau_1(t))x_1(t) \\
&\quad + b(t)x_2(t) + \sigma_1(t)\dot{\omega}_1(t) \qquad (13)\\
\dot{x}_2(t) &= (-p_{23}(t) - d_2(t) - r_2(t) + \tau_2(t))x_2(t) \\
&\quad + p_{12}(t)x_1(t) + \sigma_2(t)\dot{\omega}_2(t) \qquad (14)\\
\dot{x}_3(t) &= (-d_3(t) - r_3(t) + \tau_3(t))x_3(t) \\
&\quad + p_{23}(t)x_2(t) + \sigma_3(t)\dot{\omega}_3(t) \qquad (15)
\end{aligned}$$

where, $\dot{\omega}_j(t)$ is $N(0;1)$ implying that $\sigma_j(t)\dot{\omega}_j(t)$ is $N(0;\sigma_j^2(t))$, $j = 1,2,3$ which are independent. The mathematical models (13), (14) and (15) can be represented in vector form

$$\dot{x}(t) = A(t)x(t) + \Sigma(t)\dot{w}(t),\ t \geq 0 \tag{16}$$

where the input vector is defined by $\dot{w} = [\dot{w}_1\ \dot{w}_2\ \dot{w}_3]^T$ and the matrix $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \sigma_3\}$. Model (16) although random does not ensure that its solution is always non-negative almost surely. The non-negativeness is desirable because population cannot take negative values.

C. Log-Normal Stochastic Model for Population

First, recall the exposition of Section II, in which the population is an exponential function. Using the fact that population is non-negative subject to some form of uncertainty, it can be modeled by Log-Normal Stochastic Differential Equations as follows. Assume that each parameter in (12) is subject to additive variation modeled by a random process. That is,

$$\begin{aligned}
b(t) &\to b(t) + \dot{\omega}_b(t)\\
d_j(t) &\to d_j(t) + \dot{\omega}_{d_j}(t)\\
\tau_j(t) &\to \tau_j(t) + \dot{\omega}_{\tau_j}(t)\\
r_j(t) &\to r_j(t) + \dot{\omega}_{r_j}(t)\\
p_{12}(t) &\to p_{12}(t) + \dot{\omega}_{p_{12}}(t)\\
p_{23}(t) &\to p_{23}(t) + \dot{\omega}_{p_{23}}(t)
\end{aligned}$$

where, $j = 1,2,3$ denotes the number of the age group and $\dot{\omega}_b$, $\dot{\omega}_d$, $\dot{\omega}_\tau$, $\dot{\omega}_r$, $\dot{\omega}_{p_{12}}$, $\dot{\omega}_{p_{23}}$ are random processes. Letting

$$\begin{aligned}
\dot{\omega}_1 &\triangleq \dot{\omega}_{d_1} + \dot{\omega}_{p_{12}} + \dot{\omega}_{r_1} + \dot{\omega}_{\tau_1}\\
\dot{\omega}_2 &\triangleq \dot{\omega}_{p_{23}} + \dot{\omega}_{d_2} + \dot{\omega}_{r_2} + \dot{\omega}_{\tau_2}\\
\dot{\omega}_3 &\triangleq \dot{\omega}_{d_3} + \dot{\omega}_{r_3} + \dot{\omega}_{\tau_3}
\end{aligned}$$

Then from (12) we deduce the following modified SDE's in compact form

$$\begin{aligned}
\dot{x}_1(t) &= (-d_1(t) - p_{12}(t) - r_1(t) + \tau_1(t))x_1(t) + \\
&\quad b(t)x_2(t) + \dot{\omega}_b(t)x_2(t) + \dot{\omega}_1(t)x_1(t) \qquad (17)\\
\dot{x}_2(t) &= (-p_{23}(t) - d_2(t) - r_2(t) + \tau_2(t))x_2(t) + \\
&\quad p_{12}(t)x_1(t) + \dot{\omega}_{p_{12}}(t)x_1(t) + \dot{\omega}_2(t)x_2(t) \qquad (18)\\
\dot{x}_3(t) &= (-d_3(t) - r_3(t) + \tau_3(t))x_3(t) + \\
&\quad p_{23}(t)x_2(t) + \dot{\omega}_{p_{23}}(t)x_2(t) + \dot{\omega}_3(t)x_3(t) \qquad (19)
\end{aligned}$$

Notice that (17)-(19) are bilinear models in which the noise is multiplicative. Such models give rise to Log-Normal distributions.

D. Log-Normal Stochastic Model for Birth, Death, Immigration and Emigration

Using the fact that the number of births, deaths, immigrants, emigrants and the passing rates are non-negative subject to some form of uncertainty, these can be modeled by Log-Normal Stochastic Differential Equations as follows. A general modeling approach based on the analysis of III-C follows. Let $S_j : \Omega \times [0,T] \to \Re$, $j = 1,2..,m$ denote the number of births, deaths, emigrants, immigrants etc. Let $S = (S_1, ..., S_m)^T$, the mathematical model for $S(t)$ which is assumed to take the Log-Normal form written in incremental form

$$\begin{aligned}
dS_i(t) &= f_i(x(t), S(t))S_i(t)dt + g_i(x(t), S(t))\\
&\quad S_i(t)dv_i(t),\ S_i(0) = S_{0,i} \qquad (20)
\end{aligned}$$

where, $x : \Omega \times [0,T] \to \Re^n$ and $v : \Omega \times [0,T] \to \Re^k$ is another random process modeled by Brownian motion (e.g., dv_i is the Brownian motion increment). It can be shown that (20) yields solutions that have measure zero of taking negative values, provided f_i, g_i are chosen appropriately and $S_i(0) = S_{0,i} \geq 0$, almost surely. Consider the special case of (20) given by

$$dS_i(t) = Cx(t)S_i(t)dt + DS_i(t)dv_i(t),\ S_i(0) = S_{0,i} \tag{21}$$

where, C and D are matrices of appropriate dimensions. Here the process $x(t)$ which modulates the drift and diffusion

coefficients of (21) can be model by a Gaussian random process as follows. Suppose that $dy_i(t) \triangleq \frac{dS_i(t)}{S_i(t)}$, then $y_i(t)$ satisfies the equation

$$dy_i(t) = Cx(t)dt + Ddv_i(t) \quad (22)$$

Here, $y_i(t)$ represents percentage change; it corresponds to observations (noisy data obtained through population survey), while C, D are specific parameters to be determined shortly and $dv_i(t)$ is Brownian motion increment. The random process $x(t)$ is further modeled by a linear Gaussian stochastic differential equation given by

$$dx(t) = Ax(t)dt + Bdw(t),\ x_0 \sim N(m; \Sigma_0) \quad (23)$$

where, $A \in \Re^{n\times n}$, $B \in \Re^{n\times m}$ and $w(t)$ is vector of Brownian motion $N(0; I_{m\times m}t)$ and $(v(t),\ w(t))$ are independent of the initial state x_0.

Notice that the Log-Normal model (20) (together with (23)) is appropriate to model birth, death, immigration, emigration since the sample paths are non-negative almost surely. More importantly, model (20) is more general than the linear stochastic model.

IV. IDENTIFICATION AND ESTIMATION OF MISSING DATA

Based on the yearly data given by CyStat for the time period $1983 - 2004$, the parameters used for the modeling of the three population age groups are evaluated. Depending on the times when the growth rate of population needs to be evaluated, the annual parameters are converted into their corresponding values. Here, p_{12} stands for the yearly passing rate from G_1 to G_2 and p_{23} from G_2 to G_3. The coefficient b stands for the yearly birth rate, d_1, d_2, d_3 stand for the yearly mortality rates, τ_1, τ_2, τ_3 stand for the yearly immigration rates, and r_1, r_2, r_3 yearly emigration rates for the groups G_1, G_2, and G_3 respectively. For example consider the rate b for any given year, $i = 1, 2, 3, ...$ This is given by

$$b \equiv b(i) = \frac{\text{Number of Births during}(t_i - t_{i-1})}{x_2(t_{i-1})} \quad (24)$$

where, b represents the number of births for the period of time $(t_i - t_{i-1})$ in G_1 over the population of G_2 at the beginning of the time period. The fertility rate of the population in age groups G_1 and G_3 is considered as negligible. All the remaining rates are calculated in the same way. Often actual data provided by CyStat are either not available for a particular period of time or they are noisy. Hence system identification methods will be invoked to identify and estimate these unknown and noisy parameters.

A. Parameter Identification via the Theory of Calculus of Variations

One method is based on the Theory of Calculus of Variations and Pontryagin Minimum Principle to identify missing data [8]. According to the theory this identification problem can be formulated as an optimization problem. We need to estimate the yearly emigration, immigration, death and passing rates for the three age groups for a period of time, say $I \equiv [t_o, t_f]$. These eleven parameters which are function of time are to be identified. The state equation is now given by

$$\begin{aligned}\dot{x}_1(t) &= (-u_3(t) - u_4(t) - u_1(t) + u_2(t))x_1(t) \\ &\quad +b(t)x_2(t) \quad (25)\\ \dot{x}_2(t) &= (-u_8(t) - u_7(t) - u_5(t) + u_6(t))x_2(t) \\ &\quad +u_4(t)x_1(t) \quad (26)\\ \dot{x}_3(t) &= (-u_{11}(t) - u_9(t) + u_{10}(t))x_3(t) \\ &\quad +u_8(t)x_2(t) \quad (27)\end{aligned}$$

with the unknown parameters denoted by $u(t) \equiv [u_1(t), ..., u_{11}(t)]^T$. The parameters $u_1(t)$,$u_5(t)$ and $u_9(t)$ denote the emigration rates, $u_2(t)$, $u_6(t)$ and $u_{10}(t)$ denote the immigration rates, and $u_3(t)$, $u_7(t)$ and $u_{11}(t)$ denote the death rates for age groups G_1, G_2 and G_3 respectively. Finally, $u_4(t)$ denote the passing rate from age group G_1 to age group G_2 and $u_8(t)$ denote the passing rate from age group G_2 to age group G_3.

An admissible set of time-varying parameters u^* that causes the model to follow an admissible trajectory x^* over $[t_o, t_f]$ is the one that minimizes the performance measure $J(u)$. This performance must be selected in such a way, that the error between the population obtained by solving the dynamic model and the actual population obtained from census data, goes to zero. The performance measure is given by

$$J(u) = \int_{t_o}^{t_f} \frac{1}{2} \left\|x(t,u) - y(t)\right\|^2 dt \quad (28)$$

where, $x(t,u)$ denotes the population vector obtained by solving (25)-(27) corresponding to any arbitrary choice of the parameter $u(t)$. The vector, $y(t) = [y_1(t), y_2(t), y_3(t)]^T$ denotes the actual population obtained from census data for the period $[t_o, t_f]$. The first step is to form the Hamiltonian function

$$H(x,u,p,t) = \frac{1}{2} \left\|x(t,u) - y(t)\right\|^2 + p^T(a(x,u,t)) \quad (29)$$

where, p denote the Langrange multipliers used to convert the constraint problem into an unconstraint one. If $u^*(t)$ is the minimum of the cost functional (28) and $x^*(t)$ and $p^*(t)$ are the corresponding state and costate, then it is necessary that

$$\begin{aligned}\dot{x}^*(t) &= \frac{\partial H\left(x^*(t), u^*(t), p^*(t), t\right)}{\partial p} \quad (30)\\ \dot{p}^*(t) &= -\frac{\partial H\left(x^*(t), u^*(t), p^*(t), t\right)}{\partial x} \quad (31)\\ 0 &= \frac{\partial H\left(x^*(t), u^*(t), p^*(t), t\right)}{\partial u} \quad (32)\end{aligned}$$

Consider the case where the initial time t_o, the final time t_f and the initial state $x^*(t_o) = x_o$ are given and the final state $x(t_f)$ is free. The resulting boundary condition equations that must be satisfied are given by

$$p_1(t_f) = p_2(t_f) = p_3(t_f) = 0 \quad (33)$$

Using the state (30) and co-state equations (31), and the gradient vector (32) and by using the appropriate boundary condition equation according to (33), one can compute the best parameter $u(t)^o$ which minimizes the performance measure $J(u)$, using an algorithm for numerical solution of the optimization problem. The algorithm employed is the one found in [9].

B. System Identification via the Expectation-Maximization Algorithm

Note that (23) and the equation for $S(t)$ are generalizations of the material presented in Section II. They are able to capture the exponential growth subject to variability while the process $S(t)$ is positive with probability one, as it should be. For simplicity we discretize (22) and (23) and consider the discrete version

$$\begin{aligned} x_{t+1} &= Ax_t + Bw_t,\ x_0 \sim N(m; \Sigma_0) \\ y_t &= Cx_t + Dv_t \end{aligned} \tag{34}$$

Here $t = 0, 1, 2..$ are discrete time intervals, $x_t \in \Re^n$ is a state vector, $y_t \in \Re^d$ is a measurement vector, $w_t \in \Re^m$ is a state noise, and $v_t \in \Re^d$ is a measurement noise. The noises, $\{w_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$, are assumed to be independent zero mean Gaussian sequences, which are independent of the initial state x_0. The unknown system parameters $\theta = \{A, B, C, D\}$ are estimated using a sequence of observed samples of the data.

The Expectation Maximization (EM) algorithm is used to identify the state space model parameters A,B,C,D by using a bank of Kalman filters (KF) to find the Maximum Likelihood (ML) parameter estimates of θ, and the KF to estimate the states of the system. Let $P_\theta(x_0, x_1, ..., x_N, y_0, y_1, ..., y_N); \theta \in \Theta)$ denote a family of probability densities included by the system parameters θ. The EM algorithm is an iterative scheme for computing the ML estimate of the system parameters θ given the noisy data $Y_N = \{y_1, y_2, ..., y_N\}$. The EM algorithm employed is found in [10] and [11]. In the simulations we consider the following state-space model (since it is sufficient)

$$\begin{aligned} &A = \begin{bmatrix} 0 & 1 \\ \alpha_1 & \alpha_2 \end{bmatrix}, \quad B = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} \\ &C = [1\ 0], \quad D = [d] \end{aligned} \tag{35}$$

where the parameters $\{A, B, C, D\}$ are going to be estimated.

V. DISCUSSION OF SIMULATION RESULTS

In this section we present simulation results obtained through the solution of the problem of missing and noisy data. In particular, the results obtained via the Theory of Calculus of Variations, and the results obtained via the EM algorithm together with Kalman filtering, are discussed next.

A. Results via the Theory of Calculus of Variations

A comparison between the population obtained by the mathematical models based on the estimated parameters and the actual population reported by CyStat for the age groups G_1, G_2, G_3 and for the total population is depicted in Figure **??**, **??**, **??** and **??**, respectively. The dashed lines represent the actual population and solid lines the population generated via the mathematical models. Because of the scale of the graphs it appears that there is no difference between the actual and the model population. In fact, the maximum error over the time period for G_1 is equal to 76, for G_2 is equal to 642, for G_3 is equal to 332 and for the total population is equal to 1050. The methodology of system identification via the Theory of Calculus of Variations provides an effective tool for estimating the unknown parameters.

B. Results via the EM Algorithm together with Kalman Filtering

Here we present simulation results obtained by introducing the EM algorithm together with the Kalman filter, to estimate recursively the model parameters and states from noisy measurement data and mathematical models governed by Stochastic Differential equations in state-space form.

1) Implementation of the Gaussian Models: In order to account for the sources of uncertainty which affect the dynamic models we consider here the Linear Stochastic model, introduced in Section III-B, described by the stochastic differential equation (34). We let $y_t = \frac{S_i(t+1)-S_i(t)}{x_i(t)}, i = 1, .., 4$ denote the birth rate, the death rate of G_1, the immigration rate of G_2 and the emigration rate of G_3, respectively. Figure **??** - **??**, depict the comparison of a set of measurement data and their estimates using the Linear Stochastic models in which the parameters are identified by the EM algorithm. The dashed lines represent the parameter estimates and solid lines represent the observations.

2) Implementation of the Log-Normal Models: In order to account for the fact that the number of births, deaths, emigrants, immigrants and passings are non-negative subject to some form of uncertainty, we consider here the Log-Normal Stochastic model, introduced in Section III-C, and described by the stochastic differential equation (22) and (23). We let $dy_i(t) = \frac{dS_i(t)}{S_i(t)}, i = 1, .., 4$ denote the percentage change (PC) of births, of deaths of G_1, of immigrants of G_2, and the emigrants of G_3, respectively, where S_i are given by (21). The measurement data and the estimated values are plotted on the same graph as shown in Figure **??** - **??**. The dashed lines represent the parameter estimates and solid lines represent the observations.

Based on the simulation results, we can say that such estimators have good asymptotic properties when used in large sample statistical inference.

Note that there is fundamental difference between the Linear and the Log-Normal Stochastic models. The Linear model assumes that the data given by CyStat are precise hence free of any errors; on the other hand, the Log-Normal models assumes the data are indeed noisy. Another fundamental difference is that the Linear Stochastic models may give sample paths which can be negative at any given time while the Log-Normal models are not suffering from this disadvantage.

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

This paper describes a methodology of modeling the population growth rate of Cyprus. The mathematical models are useful since they show the dynamic transitions from one age group of population to another and therefore the short and long-term projections of population growth and decline in Cyprus. The Linear and Log-Normal Stochastic models also take into account some sources of uncertainty which might occur and also the fact that the parameters embedded in the models are non-negative, hence making the models more accurate and robust.

In case of unavailable or noisy parameters, parameter identification methods must be used to estimate these parameters. The proposed methodologies, the Theory of Calculus of Variations applied to the Linear Deterministic models and the EM algorithm combined with the Kalman Filter applied to the Linear and Log-Normal Stochastic models, can estimate with high accuracy these parameters.

The principal conclusion is that the implementation of such computer based models are essential for policy formulation, decision making and resource allocation by Government agencies and demographers.

B. Future Work

Having all the data available an analysis can be performed for short and long-term prediction. Identify resource allocation strategies which are important for government planning, by estimating the cost of several socio-economic programs. Formulate optimum immigration policies so that population growth can be sustained and therefore avoid population decline and ageing problems.

REFERENCES

[1] T. R. Malthus, "An Essay on the Principle of Population," *Anthony Flew, Baltimore, Penguin Books*, 1798

[2] F. R. Sharpe, and A. J. Lotka, "A Problem in Age Distribution," *The London, Edinburgh and Duplin Philosophical Magazine and Journal of Science*, 21: 435-438, 1911

[3] P. H. Leslie, "On the Use of Matrices in Certain Population Mathematics," *Biometrika*, Vol. 33, p 183-212, 1945

[4] A. J. Lotka, "Application of Recurrent Series in Renewal Theory," *Annals of Mathematical Statistics*, Vol. 19, p190-206, 1948

[5] L. Cole, "The population Consequences of Life History Phenomenon," *Quarterly Review of Biology*, Vol. 19, p103-107, 1954

[6] N. U. Ahmed and M. A. Rahim, "Deterministic and Stochastic Dynamic Models for Demography," *Dynamic System and Application*, 10: 325-358, 2001

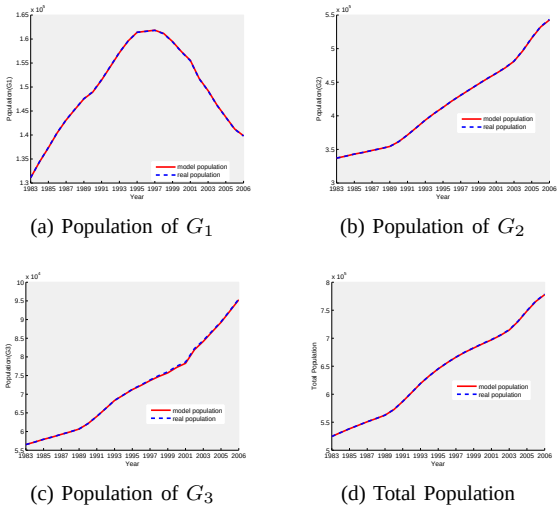
[7] N. U. Ahmed and H. Yongjuan, "Dynamic Model for Population Distribution and Optimum Immigration and Job Creation Policies," *Canadian Studies in Population*, 2: 261-295, 2007

[8] Donald E. Kirk, "Optimal Control Theory: An Introduction," *Prentice-Hall, Electrical Engineering Series*, 1970

[9] N. U. Ahmed, "A Simple Gradient Algorithm for Least Square Estimation of System Parameters," *International Journal of System Science*, 7: 637-677, 1976

[10] C. D. Charalambous, R. J. C. Baltitude, J. Zhang and Xin Li, "Modelling Wireless Fading Channels via Stochastic Differential Equations: Identification and Estimation Based on Measurements," *IEEE Transactions on Wireless Communications*, Vol. 7, No. 2, pp.434-439, 2008

[11] M.M. Olama, S.M. Djouadi and C.D. Charalambous, "Stochastic Differential Equations for Modeling, Estimation and Identification of Time-Varying Wireless Communication Channels," *EURASIP on Wireless Communications and Networking*, pages-23, 2008



(a) Population of G_1 (b) Population of G_2

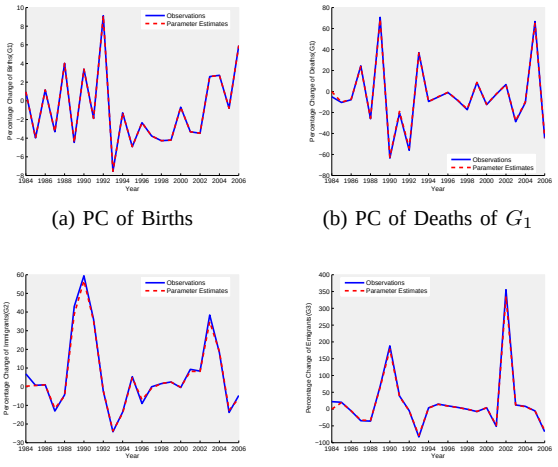
(c) Population of G_3 (d) Total Population

Fig. 1: Actual and model population using the Deterministic models and the Theory of Calculus of Variations

(a) Birth Rate (b) Death Rate of G_1

(c) Immigration Rate of G_2 (d) Emigration Rate of G_3

Fig. 2: Measured and estimated parameters using the Gaussian models and the EM algorithm together with KF



(a) PC of Births (b) PC of Deaths of G_1

(c) PC of Immigrants of G_2 (d) PC of Emigrants of G_3

Fig. 3: Measured and estimated parameters using the Log-Normal models and the EM algorithm together with KF