Applied Mathematics and Computation 224 (2013) 817–825

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Lotka–Volterra model parameter estimation using experiential data

P.H. Kloppers, J.C. Greeff *

Department of Mathematics and Statistics, Tshwane University of Technology, P/Bag X680, Pretoria 0001, South Africa

ARTICLE INFO

Keywords:
Systems modelling
Integral method
Log Integral method
Estimation of parameters
Linear regression

ABSTRACT

The purpose of mathematical models is, in principle, to assist management with decision-making processes in various fields of science. These models invariably include a number of parameters, of which the method of estimation is often in question when the models as managerial tools are evaluated. Advanced software packages involving artificial intelligence are available to estimate these parameters, and have been proven to be extremely effective. However, these techniques are expensive, require expert knowledge and often are not accessible to scientists. In this paper we discuss two alternative methods to estimate the parameters, using basic mathematical and statistical principles and a standard linear regression software package. The first method is referred to as the Integral method and the second as the Log Integral method. As an example, both these techniques are applied to determine the unknown parameters in a system of non-linear differential equations that describe an available set of data.

To illustrate the effectiveness of these simplistic methods, a three competing species Lotka–Volterra model is used as vehicle: For a given data set the model parameters obtained by an advanced artificial intelligence method are compared to the parameter values obtained when applying the proposed methods. The different approaches result in completely different estimates of the parameter values for the system, yet the solutions, using these different sets of parameters values, fit the raw data equally well. The Integral and Log Integral methods are accessible to researchers with little knowledge of mathematics, statistics or technology, and are not restricted by the nature of the problem, the number of equations or number of parameters involved. These methods are applicable to a broad spectrum of dynamical systems originating in the fields of environmental, physical, biological and social sciences.

1. Introduction

The general goal of modelling is to find a simple model that fits a data set within a predetermined error bound, while still allowing specific properties to be addressed. In this paper we investigate two simple, yet effective methods to estimate parameter values as encountered in classic Lotka–Volterra models. The parameters values describe, for example, growth, predation, competition or mutualism between interacting populations, and usually need to be quantified in order to apply the model as a managerial tool in a specific situation. The quest therefore is to quantify the parameter values of a model as accurately as possible in order to qualitatively solve the problem and fit the raw data as closely as possible.

* Corresponding author.
E-mail addresses: kloppersph@tut.ac.za, Quay.VanDerHoff@up.ac.za (P.H. Kloppers), greeffjc@tut.ac.za (J.C. Greeff).

0096-3003/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.08.093

Various methods to obtain this goal have been discussed in literature, for example Shatalov et al. [1] and Fedatov and Shatalov [2] initiate a mathematical approach of using direct integration to obtain goal functions and apply posterior quadrature rules to solve the unknown parameters. Michalakelis et al. [3] use artificial intelligence methods and advanced computer software to solve a non-linear system. In the field of analysis of linear regression models, basic rules were laid down by Searle [4], while Chatterjee and Hadi [5] illustrate these principles when applied to linear systems.

In this paper we propose two simple yet novel methods to estimate parameters, using numerical integration in conjunction with the statistical linear least squares regression technique for which computer software is generally available. The obtained parameters are then used to solve the non-linear model using Mathematica and the solutions obtained are compared to the solutions when using advanced algorithms developed in the area of artificial intelligence. The initial development of the Integral method is reported by Kloppers and Greeff [6], but this method has since been expanded and refined to include the Log Integral method and interpretation of obtained results.

An example of a generic Lotka–Volterra model describing the dynamics between three competing mobile telephone companies, based on a rich set of statistical data, is used to illustrate the practical application of the proposed methods. The problem is formulated in terms of a nonlinear dynamical system consisting of three first order differential equations, each containing linear and quadratic terms, similar to systems discussed by Bazykin [7], and Fay and Greeff [8]. The system is linear with regards to the unknown parameters that must be derived from available statistical data. When applying the proposed methods to the system, the unknown coefficients are estimated and the closeness of the respective solutions to the statistical data is discussed. The solutions are found to be sufficiently close to the data set, therefore the model may be considered as a managerial tool for predictive purposes. The proposed methods can easily be adapted to be applied to a variety of mathematical models often used to describe interaction of species in various fields of science.

2. General interpretation of the Lotka–Volterra model

The well-known Lotka–Volterra system

$$\frac{dX_i}{dt} = X_i(a_{i0} + \sum a_{ij}X_j), \quad i,j = 1,2,\ldots,n,$$

is often used to describe the dynamics of n interacting species in a community. Rates of change in the population size of each of the n species are represented by $\frac{dX_i}{dt}$. The parameters a_{i0} describe the intrinsic population growth (in which case the sign of a_{i0} would be positive) or decline (sign of a_{i0} then negative) in the absence of the other species, while the parameters a_{ij} could be positive, negative or zero, and would reflect whether the species interact in terms of predation, competition, mutualism or not at all.

3. The Integral method

Consider a situation where three species x, y and z compete for available resources in a system. A Lotka–Volterra model for three competing species is generically represented by

$$\begin{aligned}\frac{dx}{dt} &= x(a_{10} + a_{11}x + a_{12}y + a_{13}z),\\ \frac{dy}{dt} &= y(a_{20} + a_{21}x + a_{22}y + a_{23}z),\\ \frac{dz}{dt} &= z(a_{30} + a_{31}x + a_{32}y + a_{33}z).\end{aligned} \tag{1}$$

We illustrate the proposed Integral method for estimating the coefficients $a_{10},\ a_{11}, a_{12}$ and a_{13} in the first differential equation in the above system only, since the coefficients in the second and third equations could be obtained in a similar manner. Therefore consider the simplified equation for species x, namely

$$\frac{dx}{dt} = x(a_0 + a_1x + a_2y + a_3z), \tag{2}$$

where x, y and z are functions of t. Integrating both sides of Eq. (2) with respect to t over the interval $[t_0, t_n]$ yields

$$\int_{t_0}^{t_n} \frac{dx}{dt}dt = x(t)|_{t_0}^{t_n} = \int_{t_0}^{t_n} x(t)(a_0 + a_1x(t) + a_2y(t) + a_3z(t))dt. \tag{3}$$

The time interval $[t_0, t_n]$ may be divided into n sub-intervals, each with unit length, as illustrated in Fig. 1.

For each of these intervals it follows that

$$x(t)|_{t_j}^{t_{j+1}} = \int_{t_j}^{t_{j+1}} x(t)(a_0 + a_1x(t) + a_2y(t) + a_3z(t))dt, \quad j = 0,1,2,\ldots,(n-1). \tag{4}$$

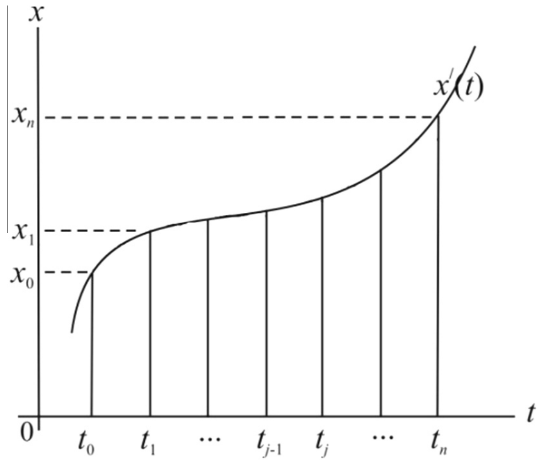

Fig. 1. Division of interval $[t_0, t_n]$ in n sub-intervals with unit length.

so that

$$x(t_{j+1}) - x(t_j) = a_0 \int_{t_j}^{t_{j+1}} x(t)dt + a_1 \int_{t_j}^{t_{j+1}} x^2(t)dt + a_2 \int_{t_j}^{t_{j+1}} x(t)y(t)dt + a_3 \int_{t_j}^{t_{j+1}} x(t)z(t)dt. \tag{5}$$

Any numerical method can be used to estimate each of the integrals on the right hand side of Eq. (5). Remembering that the intervals $[t_j, t_{j+1}]$ are of unit length and using the Trapezium rule, it follows that

$$a_0 \int_{t_j}^{t_{j+1}} x(t)dt \approx \frac{(t_{j+1} - t_j)}{2}\left[a_0(x(t_{j+1}) + x(t_j))\right] = \frac{a_0}{2}(x(t_{j+1}) + x(t_j)),$$

$$a_1 \int_{t_j}^{t_{j+1}} x^2(t)dt \approx \frac{(t_{j+1} - t_j)}{2}\left[a_1\left(x^2(t_{j+1}) + x^2(t_j)\right)\right] = \frac{a_1}{2}\left(x^2(t_{j+1}) + x^2(t_j)\right),$$

$$a_2 \int_{t_j}^{t_{j+1}} x(t)y(t)dt \approx \frac{(t_{j+1} - t_j)}{2}\left[a_2(x(t_{j+1})y(t_{j+1}) + x(t_j)y(t_j))\right] = \frac{a_2}{2}(x(t_{j+1})y(t_{j+1}) + x(t_j)y(t_j)),$$

and

$$a_3 \int_{t_j}^{t_{j+1}} x(t)z(t)dt \approx \frac{(t_{j+1} - t_j)}{2}\left[a_3(x(t_{j+1})z(t_{j+1}x) + (t_j)z(t_j))\right] = \frac{a_3}{2}(x(t_{j+1})z(t_{j+1}) + x(t_j)z(t_j)).$$

If statistical data for $x(t_j)$, $y(t_j)$ and $z(t_j)$ are available, the set of linear equations above can now be represented in matrix notation as

$$\begin{bmatrix} d_{1,0} \\ d_{2,1} \\ \cdot \\ \cdot \\ \cdot \\ d_{n,n-1} \end{bmatrix} = \begin{bmatrix} \bar{x}_{1,0} & \bar{x}^2_{1,0} & \bar{x}\bar{y}_{1,0} & \bar{x}\bar{z}_{1,0} \\ \bar{x}_{2,1} & \bar{x}^2_{2,1} & \bar{x}\bar{y}_{2,1} & \bar{x}\bar{z}_{2,1} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \bar{x}_{n,n-1} & \bar{x}^2_{n,n-1} & \bar{x}\bar{y}_{n,n-1} & \bar{x}\bar{z}_{n,n-1} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{or} \quad \underline{d} = X\underline{a}, \tag{6}$$

with $d_{j+1,j} = x(t_{j+1}) - x(t_j)$, $j = 0, 1, \ldots, (n-1)$. The matrix X contains the consecutive pair-wise means, namely

$$\bar{x}_{j+1,j} = (x(t_{j+1}) + x(t_j))/2,$$

$$\bar{x}^2_{j+1,j} = \left((x(t_{j+1}))^2 + (x(t_j))^2\right)/2,$$

$$\bar{x}\bar{y}_{j+1,j} = \left(x(t_{j+1})y(t_{j+1}) + x(t_j)y(t_j)\right)/2,$$

and

$$\bar{x}\bar{z}_{j+1,j} = (x(t_{j+1})z(t_{j+1}) + x(t_j)z(t_j))/2,$$

while the vector $\underline{a} = [a_0, a_1, a_2, a_3]'$ contains the unknown parameters to be determined.

The matrix X has n rows and four columns which are linearly independent, therefore matrix X is of full column rank (see Searle [4]). The estimated solution $\underline{\hat{a}}$ of System (6) is therefore a typical statistical non-intercept multiple regression problem and the least squares estimation of the unknown parameters can be calculated.

In order to solve the unknown parameters $\underline{a}$ in $\underline{d} = X\underline{a}$, consider the transpose of X, namely X', so that

$$X'\underline{d} = X'X\underline{a}.$$

Note that X is an $n \times 4$ matrix with $rank(X) = 4$, X' is a $4 \times n$ matrix with $rank(X') = 4$, thus $X'X$ is a 4×4 matrix with $rank(X'X) = 4$. Therefore $(X'X)$ is non-singular and its inverse $(X'X)^{-1}$ exists. The ordinary multiple least squares estimation $\underline{\hat{a}}$ of $\underline{a}$ is then given by

$$\underline{\hat{a}} = (X'X)^{-1}X'\underline{d}. \tag{7}$$

Any generally available statistical software package such as Microsoft Excel can be used to solve Eq. (7), since the values $d_{j+1,j}$, $\bar{x}_{j+1,j}$, $\bar{x}^2_{j+1,j}$, $\bar{x}\bar{y}_{j+1,j}$ and $\bar{x}\bar{z}_{j+1,j}$ can be tabulated for $j = 0, 1, \ldots, n-1$.

4. The Log Integral method

An improvement on the Integral method may be obtained if the number and complexity of the means in matrix X (which are represented by estimates) can be reduced. Again consider Eq. (2)

$$\frac{dx}{dt} = x(a_0 + a_1x + a_2y + a_3z),$$

where x, y and z are functions of t. Under the assumption that all $x_j > 0$, both sides of the latter equation can be divided by x, that is

$$\frac{1}{x}\frac{dx}{dt} = a_0 + a_1x + a_2y + a_3z, \tag{8}$$

Integrating both sides of Eq. (8) with respect to t over the interval $[t_0, t_n]$ yields

$$\int_{t_0}^{t_n} \frac{1}{x}\frac{dx(t)}{dt}dt = \ln x(t)|_{t_0}^{t_n} = \int_{t_0}^{t_n} (a_0 + a_1x(t) + a_2y(t) + a_3z(t))dt.$$

so that for each sub-interval

$$\ln x(t_{j+1}) - \ln x(t_j) = a_0 + a_1\int_{t_j}^{t_{j+1}} x(t)dt + a_2\int_{t_j}^{t_{j+1}} y(t)dt + a_3\int_{t_j}^{t_{j+1}} z(t)dt. \tag{9}$$

Using the Trapezium rule again, an estimation of Eq. (9) is

$$\ln x(t_{j+1}) - \ln x(t_j) \approx a_0 + a_1(x(t_{j+1}) + x(t_j))/2 + a_2(y(t_{j+1}) + y(t_j))/2 + a_3(z(t_{j+1}) + x(t_j))/2.$$

For a data set $x(t_j)$, $y(t_j)$ and $z(t_j)$ the set of linear equations to describe the above now is

$$\begin{bmatrix} l_{1,0} \\ l_{2,1} \\ \cdot \\ \cdot \\ \cdot \\ l_{n,n-1} \end{bmatrix} = \begin{bmatrix} 1 & \bar{x}_{1,0} & \bar{y}_{1,0} & \bar{z}_{1,0} \\ 1 & \bar{x}_{2,1} & \bar{y}_{2,1} & \bar{z}_{2,1} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 1 & \bar{x}_{n,n-1} & \bar{y}_{n,n-1} & \bar{z}_{n,n-1} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{or} \quad \underline{l} = X\underline{a}, \tag{10}$$

with

$$l_{j+1,j} = \ln x(t_{j+1}) - \ln x(t_j),$$

$$\bar{x}_{j+1,j} = (x(t_{j+1}) + x(t_j))/2,$$

$$\bar{y}_{j+1,j} = (y(t_{j+1}) + y(t_j))/2,$$

$$\bar{z}_{j+1,j} = (z(t_{j+1}) + z(t_j))/2, \quad j = 0, 1, 2, \ldots n-1.$$

This is a typical non-intercept multiple regression problem, and the vector $\underline{a}$ containing the unknown parameters of the system is estimated by

$$\underline{\hat{a}} = (X'X)^{-1}X'\underline{l}$$

Remark. Although the methods employed are referred to as the Integral and Log Integral methods, the resulting system contains consecutive differences of the variables (or logs thereof) on the left hand side, and means of linear (or non-linear) terms of the right hand side of the equations comprising the systems.

5. Example

It is a general problem in system analysis that reliable data sets are not freely available. However, Michalakelis et al. [3] published a rich set of historical data for three competitors within the market of cell phone telecommunications systems in Greece and propose a Lotka–Volterra competing species model to describe their dynamics. The purpose of their study is to project the possibility of ensuring a stable and healthy future competitive market equilibrium, based on available data with respect to the control of market shares in the cell phone telephony technology over the past 9 years. The competing cellular service providers' respective past and possible future capture and/or penetration of the existing market is under investigation. Data are available for the period 1995–2007, but since the second provider y only entered the market in 1998, the study on the three competing species is based on initial conditions as at the start of 1998. The data represent the percentage cut of the existing market controlled by each competitor at that point in time, as reflected in Table 1, and are used as vehicle to demonsrate the proposed methods.

The model suggested by Michalakelis et al. [3] to describe this competition, is based on the principles as discussed in Eq. (1), where x, y and z represent the percentage of the market share captured by the three competitors respectively. They used an advanced artificial intelligence methods (referred to as the Advanced method in the next sections) to determine the unknown parameter values, resulting in

$$\begin{aligned} \frac{dx}{dt} &= x(0.45 - 0.6x - 0.2y - 0.66z), \\ \frac{dy}{dt} &= y(0.86 - 0.02x - 1.8y - 0.59z), \\ \frac{dz}{dt} &= z(0.2 - 0.06x - 0.13y - 0.5z). \end{aligned} \tag{11}$$

Table 1
Historical data on portion of available market shares captured by Competitor x, y and z respectively over the period 1998–2007.

Period	Competitor x	Competitor y	Competitor z
1998a	0.52	0.15	0.33
1998b	0.47	0.21	0.32
1999a	0.43	0.27	0.3
1999b	0.4	0.31	0.29
2000a	0.38	0.35	0.28
2000b	0.37	0.36	0.27
2001a	0.36	0.37	0.27
2001b	0.36	0.37	0.27
2002a	0.35	0.38	0.27
2002b	0.36	0.38	0.25
2003a	0.37	0.39	0.24
2003b	0.38	0.39	0.23
2004a	0.39	0.39	0.22
2004b	0.38	0.39	0.23
2005a	0.37	0.39	0.24
2005b	0.38	0.39	0.23
2006a	0.38	0.4	0.22
2006b	0.36	0.39	0.25
2007a	0.34	0.38	0.28

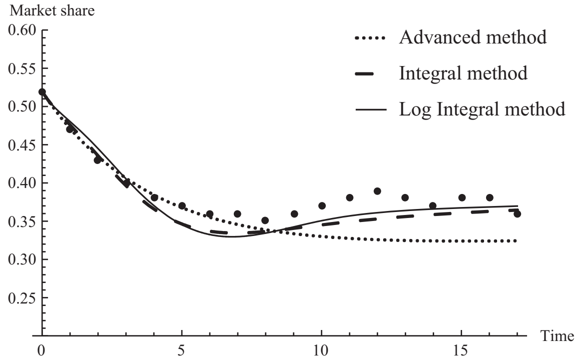

Fig. 2. The fit of the respective solution curves to the statistical data for the % market share for competitor x over the period 1998–2007. The big dots represent the true market share.

However, when applying the proposed Integral method, the resulting system is given by

$$\begin{aligned} \frac{dx}{dt} &= x(4.1492 - 4.2136x - 3.8654y - 4.5320z), \\ \frac{dy}{dt} &= y(0.9902 - 0.0360x - 2.0410y - 0.7621z), \\ \frac{dz}{dt} &= z(-2.6812 + 2.5223x + 2.8576y + 2.6392z), \end{aligned} \tag{12}$$

and for the Log Integral method, the system is

$$\begin{aligned} \frac{dx}{dt} &= x(4.3432 - 4.4069x - 4.0595y - 4.7270z), \\ \frac{dy}{dt} &= y(0.5734 + 0.3290x - 1.5848y - 0.3285z), \\ \frac{dz}{dt} &= z(-3.6170 + 3.4259x + 3.8169y + 3.5911z). \end{aligned} \tag{13}$$

Of interest are the notable differences between the sets of estimated parameter values in the respective models. This phenomenon illustrates the generally assumed fact that for an over-determined mathematical system many solutions exist. Furthermore, the original problem was interpreted as a typical competing species situation with intra-species competition a possibility, as indicated by inclusion of the terms containing x^2, y^2 and z^2. However, Systems (12) and (13) are examples of a classical one-predator–two-prey model with intra-species competition [6,9,10]. Note that species y is benefitting from the presence of species x, but not *vice versa*. This is referred to as one-sided mutualism between prey species in the presence of the predator z. With regards to the cellular service providers in Greece, this may be interpreted as follows: Firm y entered the market when firms x and z were already well established, but steadily captured its equal part of the existing market over time, benefitting from the one-sided mutualism with its more "abundantly available" competitor x, which shielded it successfully from the predator z.

When graphically comparing the solutions of Systems (11)–(13) to the statistical data, it seems that all three sets of system solutions fit the raw data (dotted line) equally well. This is represented in Figs. 2–4. Mathematica® [11] was used to solve and plot the solution curves for the systems of differential equations.

From the figures above it seems that the solutions of the Advanced method [3] are less sensitive to variation of data values in consecutive time intervals. It may be assumed that this is due to multiple forward and backward smoothing techniques, using four-term moving averages in the development of the Advanced method, thus the under- and over-shooting of solution curves for functions x and z while the smoother set of raw data for y is traced more accurately. On the other hand, in the Integral and Log Integral methods only two-term moving averages are used, therefore a more representative estimation of the variation in raw data is obtained. In addition, in the Log Integral method the number of estimates of means, as well as the complexity of the products of means, has been reduced in order to ensure an even closer fit to the raw data than that obtained by the Integral method.

The future projections of these systems are very similar, indicating stability in the market over time, as illustrated in Figs. 5–7:

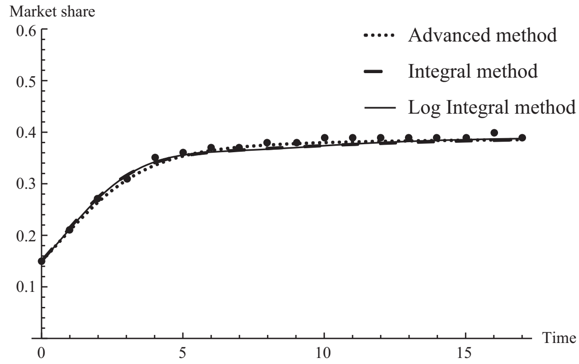

Fig. 3. The fit of the respective solution curves to the statistical data for the % market share for competitor y over the period 1998–2007. The big dots represent the true market share.

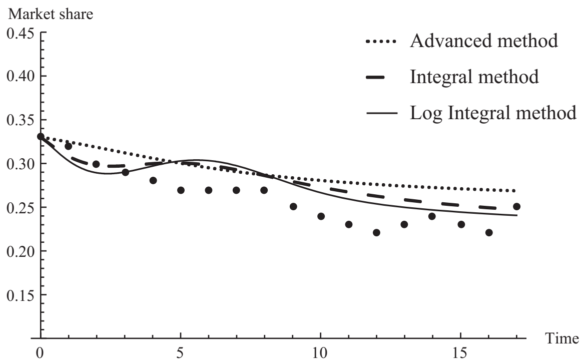

Fig. 4. The fit of the respective solution curves to the statistical data for the % market share for competitor z over the period 1998–2007. The big dots represent the true market share.

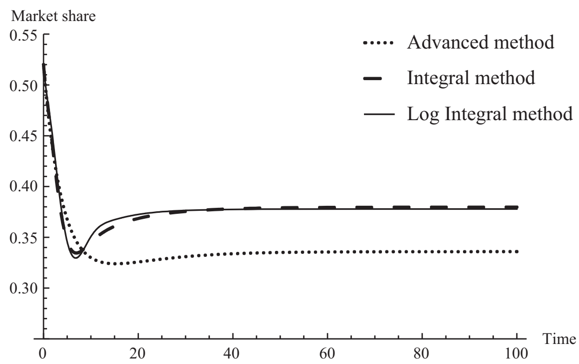

Fig. 5. Future projection of the market share for competitor x using the Advanced, Integral and Log Integral methods.

When comparing the mean square error of the Advanced method to those of the Integral and Log Integral methods with respect to competitor x, y and z respectively, it is clear that the solution of the Log Integral method fits the raw data of competitor x and z significantly more accurately than the Advanced method, and marginally better than the Integral method. For competitor y, having smooth raw data, the fit of the three methods differ marginally. To confirm these findings, a summary of mean square errors for all variables are given in Table 2.

It can therefore be assumed that, overall, the predictive qualities of the Integral method will be more reliable than that of the Advanced or Integral method for species x and z, and marginally less reliable for species y.

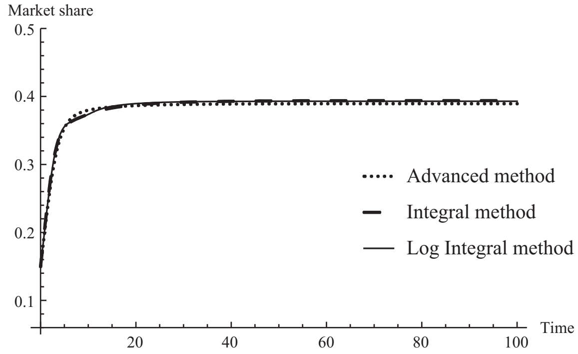

Fig. 6. Future projection of the market share for competitor *y* using the Advanced, Integral and Log Integral methods.

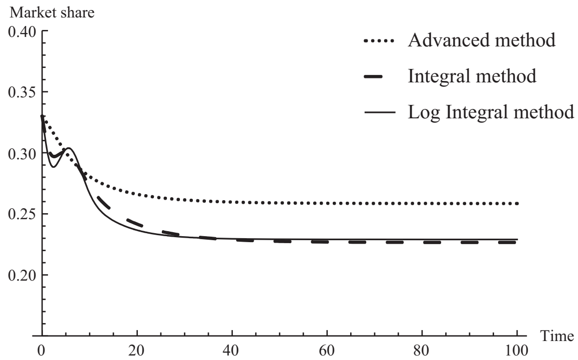

Fig. 7. Future projection of the market share for competitor *z* using the Advanced, Integral and Log Integral methods.

Table 2
Comparison of the mean square error of the advanced method, the Integral and Log Integral methods.

MSE	Competitor x	Competitor y	Competitor z
Advanced method	0.00119	0.00005	0.00103
Integral method	0.00036	0.00008	0.00062
Log Integral method	0.00035	0.00007	0.00054

6. Conclusion

In this paper we propose the simplistic Integral and Log Integral methods to estimate the unknown parameters used in systems of nonlinear differential equations. The methods involve elementary mathematical and statistical principles and are applicable to models describing problems encountered in any field of science, such as ecology, physics, biology, electronics etcetera. The application of the methods is therefore not restricted by the nature of the problem, the number of equations or number of parameters involved. Furthermore, the proposed methods are available to all researchers with only basic knowledge of mathematics and statistics and limited access to technology. Considering the cost and availability of advanced software packages in developing countries, the accuracy of parameter estimates and projective abilities of the Integral and Log Integral methods compare favourably to that of advance methods.

Acknowledgments

This research project was supported by the National Research Foundation, South Africa (Grant no. 2054454). The authors thank the Tshwane University of Technology for their support and the reviewers for their useful comments regarding the improvement of this manuscript.

References

[1] M. Shatalov, J.C. Greeff, I. Fedotov, S.V. Joubert, Parametric identification of the model with one predator and two prey species, in: TIME2008 International Conference Proceedings, 2008, pp. 101–110.
[2] M. Shatalov, I. Fedatov, On identification of dynamical system parameters from experiential data, Research Group in Mathematical Inequalities and Applications 10 (1) (2007) 1–9.
[3] C. Michalakelis, T.S. Sphicopoulos, D. Varoutas, Modelling competition in the telecommunications market based on the concepts of population biology, Transactions on Systems, Man and Cybernetics. Part C: Applications and Reviews 4 (2011) 200–210.
[4] S.R. Searle, Linear Models, Wiley & Sons Inc., 1971.
[5] S. Chatterjee, A.S. Hadi, Regression Analysis by Example, fourth ed., Wiley & Sons Inc., 2006.
[6] P.H. Kloppers, J.C. Greeff, Estimation of parameters in population models, in: Proceedings IASTED International Conference on Modelling and Simulation, 2010, pp. 87–91, ISBN: 978-088986862-6.
[7] A. Bazykin, Nonlinear Dynamics of Interacting Populations (Series in Neural Systems), World Scientific Pub Co Inc, 1998.
[8] T.H. Fay, J.C. Greeff, Lion, wildebeest and zebra: a three species model, Ecological Modelling 1 (2006) 237–244.
[9] J.D. Murray, Mathematical Biology, Springer-Verlag, New York, 1993.
[10] A.M. Starfield, A.L. Bleloch, Building Models for Conservation and Wildlife Management, Macmillan Publishing Company, Brisbane, 1986.
[11] S. Wolfram, The Mathematica® Book, third ed., Wolfram Media/Cambridge University Press, New York, 1998.